

Strapdown Inertial Navigation Technology – 2nd Edition

**David H. Titterton and
John L. Weston**

The Institution of Electrical Engineers

Copublished by:

The Institution of Electrical Engineers,
Michael Faraday House,
Six Hills Way, Stevenage,
Herts. SG1 2AY, United Kingdom

and

The American Institute of Aeronautics and Astronautics
1801 Alexander Bell Drive
Suite 500
Reston
VA 20191-4344
USA

© 2004: The Institution of Electrical Engineers

This publication is copyright under the Berne Convention and the Universal Copyright Convention. All rights reserved. Apart from any fair dealing for the purposes of research or private study, or criticism or review, as permitted under the Copyright, Designs and Patents Act, 1988, this publication may be reproduced, stored or transmitted, in any forms or by any means, only with the prior permission in writing of the publishers, or in the case of reprographic reproduction in accordance with the terms of licences issued by the Copyright Licensing Agency. Inquiries concerning reproduction outside those terms should be sent to the IEE at the address above.

While the authors and the publishers believe that the information and guidance given in this work are correct, all parties must rely upon their own skill and judgment when making use of them. Neither the authors nor the publishers assume any liability to anyone for any loss or damage caused by any error or omission in the work, whether such error or omission is the result of negligence or any other cause. Any and all such liability is disclaimed.

The moral right of the authors to be identified as authors of this work have been asserted by them in accordance with the Copyright, Designs and Patents Act 1988.

British Library Cataloguing in Publication Data

Titterton, D. H. (David H.)

Strapdown inertial navigation technology. – 2nd ed. – (Radar, sonar, navigations & avionics)

1. Inertial navigation (Aeronautics) 2. Inertial navigation systems

I. Title II. Weston, J. L. (John L.) III. Institution of Electrical Engineers

629.1'351

ISBN 0 86341 358 7

Typeset in India by Newgen Imaging Systems (P) Ltd., Chennai, India
Printed in the UK by MPG Books Limited, Bodmin, Cornwall

Preface

The use of inertial sensors and inertial navigation has developed rapidly in the recent past, owing to a number of very significant technological advances. The rapid development of micro electromechanical sensors and superior computer performance has provided the stimulus for many new applications. These new sensors and enhanced computer power, as well as state-of-the-art computational techniques have been applied to the traditional applications to give enhanced performance in a miniature system. These traditional applications are:

- navigation of any mobile platform or vehicle;
- autopilots for guidance and control of aircraft, missiles, ships and land vehicles;
- control of gimbals and other structures.

The availability of small, cheap and reliable sensors during the last five years has led to many novel applications being investigated. Many of these new applications are either established or are currently being developed. Examples include:

- surveying underground pipelines during drilling operations;
- active control in high-performance vehicles;
- passive missile roll control;
- calibration and measurement;
- personal transportation.

The recent past has seen the design and demonstration of highly accurate sensors such as ring laser gyroscopes and cold atom sensors. These devices continue to be developed and they provide a unique approach to measurement of motion and apparent anomalies. As a consequence, these devices are being applied to an increasing number of novel applications, including geodesy and fundamental physical studies.

It is not realistic to ignore satellite-based navigation in this type of text book, despite it not really being a 'strapdown' technique. The advances with this form of navigation has had a dramatic impact on navigation methods and techniques, and consequently, on the range of applications now using navigational methods. In particular this technique has been used in conjunction with other sensors to produce a highly effective and cost-effective navigation technique commonly known as integrated navigation. In this context satellite-based navigation is considered along with

some applications. Moreover, a number of other non-inertial navigational aides are reviewed.

The aim of this book is to provide a clear and concise description of the physical principles of inertial navigation; there is also a more detailed treatise covering recent developments in inertial sensor technology and the techniques for implementing such systems. This includes a discussion of the state-of-the-art of so-called MEMS devices and other novel approaches to sense angular and linear motion.

It is intended that the book should provide an up to date guide to the techniques of inertial navigation, which will be of interest to both the practising engineer and the post-graduate student. The text describes a range of technologies and evaluation techniques to enable informed judgements to be made about the suitability of competing technologies and sensors. Data are provided to give an indication of the range of performance that can be achieved from both component devices and systems.

There is a detailed description of the techniques that may be used to evaluate different technologies, covering a review of testing, characterisation and calibration methods used to ensure optimum performance is achieved from the sensors and the system. Illustrated examples are given to highlight the interaction between competing effects and their impact on performance.

These methods and techniques are drawn together in a detailed design example, which illustrates approaches for defining and analysing the problem, deriving an appropriate specification and designing a solution. The design example also considers the computational requirements, as well as interfaces and evaluation techniques. This study should be of particular interest to technologists tasked with making a system perform to a specification as it illustrates the potential interactions and the compromises that have to be made in the inevitable trade-off between parameters and performance during the design process.

This edition has a chapter describing modern and unusual application of inertial sensors and techniques. The aim is to inspire engineers and technologists to greater innovation with the vast array of technology, methods and techniques that are available to them. This chapter also includes a brief design example to illustrate the issues concerning systems for operation in a hostile environment.

This technology uses many specialised terms and expressions as well as jargon. A comprehensive glossary of terms with an explanation or definition is included to aid the understanding of the subject. Additionally, appendices are included to give further development of techniques and concepts related to strapdown inertial navigation.

It would not be possible to complete a project with this scope without the help of many people, whom we thank. We are particularly grateful to the following for their help and encouragement:

Mr S.K. Davison, Dr P. Groves, Dr K. Brunsen of QinetiQ

Mr E. Whitley, Dr R. Fountain and Mr P.G. Moore of BAE Systems

Mr J.I.R. Owen and Mr P. Marland of DSTL

Mr G. Baker of Systron-Donner Inertial Division of BEI Technologies, Inc.

Mr C.R. Milne, formerly of RAE Farnborough

Mr A.D. King formerly of Ferranti and latterly of GEC Avionics

Mr W.D. Coskren of The Charles Stark Draper Laboratories Inc.

Mr G. Vigurs of SULA Systems, UK

Dr P. Rodney, Mr D.P. McRobbie of Sperry-Sun, Halliburton Energy Services Group

Dr Milan Horemuz of the Royal Institute of Technology, Sweden

Dr R. Ebert and Dr P. Lützmann of FGAN-FOM, Germany

Prof. G.E. Stedman of the University of Canterbury, New Zealand

Prof. E.U. Schreiber of Fundamentalstation Wettzell in Bavaria, Germany, Technical University of Munich

Mr B. Waggoner of NSWC, Crane, Indiana, USA

Mr D. Goetze, SEG, Germany

Contents

Preface	xv
1. Introduction	1
1.1 Navigation	1
1.2 Inertial Navigation	2
1.3 Strapdown Technology	3
1.4 Layout of the Book	4
2. Fundamental Principles and Historical Developments of Inertial Navigation	7
2.1 Basic Concepts	7
2.2 Summary	11
2.3 Historical Developments	11
2.4 The Modern-day Inertial Navigation System	14
2.5 Trends in Inertial Sensor Development	15
3. Basic Principles of Strapdown Inertial Navigation Systems	17
3.1 Introduction	17
3.2 A Simple Two-dimensional Strapdown Navigation System	17
3.3 Reference Frames	21
3.4 Three-dimensional Strapdown Navigation System – General Analysis	22
3.4.1 Navigation with Respect to a Fixed Frame ...	22
3.4.2 Navigation with Respect to a Rotating Frame	24
3.4.3 The Choice of Reference Frame	24

3.4.4	Resolution of Accelerometer Measurements	24
3.4.5	System Example	25
3.5	Strapdown System Mechanisations	25
3.5.1	Inertial Frame Mechanisation	26
3.5.2	Earth Frame Mechanisation	28
3.5.3	Local Geographic Navigation Frame Mechanisation	31
3.5.4	Wander Azimuth Navigation Frame Mechanisation	34
3.5.5	Summary of Strapdown System Mechanisations	36
3.6	Strapdown Attitude Representations	36
3.6.1	Introductory Remarks	36
3.6.2	Direction Cosine Matrix	39
3.6.3	Euler Angles	40
3.6.4	Quaternions	42
3.6.5	Relationships between Direction Cosines, Euler Angles and Quaternions	45
3.7	Detailed Navigation Equations	47
3.7.1	Navigation Equations Expressed in Component Form	47
3.7.2	The Shape of the Earth	49
3.7.3	Datum Reference Models	51
3.7.4	Variation of Gravitational Attraction over the Earth	55
4.	Gyroscope Technology 1	59
4.1	Introduction	59
4.2	Conventional Sensors	60
4.2.1	Introduction	60
4.2.2	Fundamental Principles	60

4.2.3	Components of a Mechanical Gyroscope	68
4.2.4	Sensor Errors	71
4.2.5	Rate-integrating Gyroscope	73
4.2.6	Dynamically Tuned Gyroscope	77
4.2.7	Flex Gyroscope	81
4.3	Rate Sensors	84
4.3.1	Dual-axis Rate Transducer (DART)	84
4.3.2	Magnetohydrodynamic Sensor	86
4.4	Vibratory Gyroscopes	88
4.4.1	Introduction	88
4.4.2	Vibrating Wine Glass Sensor	89
4.4.3	Hemispherical Resonator Gyroscope	91
4.4.4	Vibrating Disc Sensor	93
4.4.5	Tuning Fork Sensor	94
4.4.6	Quartz Rate Sensor	94
4.4.7	Silicon Sensor	96
4.4.8	Vibrating Wire Rate Sensor	98
4.4.9	General Characteristics of Vibratory Sensors	99
4.5	Cryogenic Devices	100
4.5.1	Nuclear Magnetic Resonance Gyroscope	100
4.5.2	SARDIN	103
4.6	Electrostatically Suspended Gyroscope	103
4.7	Other Devices for Sensing Angular Motion	105
4.7.1	Fluidic (Flueric) Sensors	105
4.7.2	Fluxgate Magnetometers	107
4.7.3	The Transmission Line Gyroscope	112
5.	Gyroscope Technology 2	115
5.1	Optical Sensors	115
5.1.1	Introduction	115

5.1.2	Fundamental Principles	116
5.1.3	Ring Laser Gyroscope	118
5.1.4	Three-axis Ring Laser Gyroscope Configuration	126
5.1.5	Fibre Optic Gyroscope	126
5.1.6	Photonic Crystal Optical Fibre Gyroscope	137
5.1.7	Fibre Optic Ring Resonator Gyroscope	140
5.1.8	Ring Resonator Gyroscope	142
5.1.9	Integrated Optical Gyroscope	143
5.2	Cold Atom Sensors	143
5.2.1	Introduction	143
5.2.2	Rotation Sensing	144
5.2.3	Measurement of Acceleration	145
5.2.4	Gravity Gradiometer	146
5.3	Summary of Gyroscope Technology	148
6.	Accelerometer and Multi-sensor Technology	153
6.1	Introduction	153
6.2	The Measurement of Translational Motion	153
6.3	Mechanical Sensors	155
6.3.1	Introduction	155
6.3.2	Principles of Operation	155
6.3.3	Sensor Errors	156
6.3.4	Force-feedback Pendulous Accelerometer	157
6.3.5	Pendulous Accelerometer Hinge Elements	159
6.3.6	Two-axes Force-feedback Accelerometer	160
6.3.7	Open-loop Accelerometers	161

6.4	Solid-state Accelerometers	161
6.4.1	Vibratory Devices	162
6.4.2	Surface Acoustic Wave Accelerometer	163
6.4.3	Silicon Sensors	165
6.4.4	Fibre Optic Accelerometer	168
6.4.5	Optical Accelerometers	173
6.4.6	Other Acceleration Sensors	173
6.5	Multi-functional Sensors	174
6.5.1	Introduction	174
6.5.2	Rotating Devices	174
6.5.3	Vibratory Multi-sensor	178
6.5.4	Mass Unbalanced Gyroscope	179
6.6	Angular Accelerometers	182
6.6.1	Liquid Rotor Angular Accelerometer	183
6.6.2	Gas Rotor Angular Accelerometer	184
6.7	Inclinometers	185
6.8	Summary of Accelerometer and Multi-sensor Technology	186
7.	MEMS Inertial Sensors	189
7.1	Introduction	189
7.2	Silicon Processing	192
7.3	MEMS Gyroscope Technology	193
7.3.1	Introduction	193
7.3.2	Tuning Fork MEMS Gyroscopes	195
7.3.3	Resonant Ring MEMS Gyroscopes	202
7.4	MEMS Accelerometer Technology	205
7.4.1	Introduction	205
7.4.2	Pendulous Mass MEMS Accelerometers	206
7.4.3	Resonant MEMS Accelerometers	207
7.4.4	Tunnelling MEMS Accelerometers	209

x Contents

7.4.5	Electrostatically Levitated MEMS Accelerometers	210
7.4.6	Dithered Accelerometers	212
7.5	MOEMS	212
7.6	Multi-axis/Rotating Structures	212
7.7	MEMS Based Inertial Measurement Units	213
7.7.1	Silicon IMU	213
7.7.2	Quartz IMU	214
7.8	System Integration	215
7.9	Summary	216
8.	Testing, Calibration and Compensation	219
8.1	Introduction	219
8.2	Testing Philosophy	220
8.3	Test Equipment	221
8.4	Data-logging Equipment	222
8.5	Gyroscope Testing	223
8.5.1	Stability Tests – Multi-position Tests	223
8.5.2	Rate Transfer Tests	226
8.5.3	Thermal Tests	231
8.5.4	Oscillating Rate Table Tests	233
8.5.5	Magnetic Sensitivity Tests	233
8.5.6	Centrifuge Tests	235
8.5.7	Shock Tests	237
8.5.8	Vibration Tests	238
8.5.9	Combination Tests	241
8.5.10	Ageing and Storage Tests	242
8.6	Accelerometer Testing	242
8.6.1	Multi-position Tests	244
8.6.2	Long-term Stability	244
8.6.3	Thermal Tests	246

8.6.4	Magnetic Sensitivity Tests	246
8.6.5	Centrifuge Tests	247
8.6.6	Shock Tests	250
8.6.7	Vibration Tests	250
8.6.8	Combination Tests	251
8.6.9	Ageing and Storage Tests	252
8.7	Calibration and Error Compensation	253
8.7.1	Introduction	253
8.7.2	Gyroscope Error Compensation	254
8.7.3	Accelerometer Error Compensation	254
8.7.4	Further Comments on Error Compensation	255
8.8	Testing of Inertial Navigation Systems	255
8.9	Hardware in the Loop Tests	259
9.	Strapdown System Technology	263
9.1	Introduction	263
9.2	The Components of a Strapdown Navigation System	263
9.3	The Instrument Cluster	264
9.3.1	Orthogonal Sensor Configurations	264
9.3.2	Skewed Sensor Configurations	265
9.3.3	A Skewed Sensor Configuration Using Dual-axis Gyroscopes	266
9.3.4	Redundant Sensor Configurations	268
9.4	Instrument Electronics	269
9.5	The Attitude Computer	271
9.6	The Navigation Computer	272
9.7	Power Conditioning	274
9.8	Anti-vibration Mounts	274
9.9	Concluding Remarks	274

10. Inertial Navigation System Alignment	277
10.1 Introduction	277
10.2 Basic Principles	278
10.2.1 Alignment on a Fixed Platform	278
10.2.2 Alignment on a Moving Platform	280
10.3 Alignment on the Ground	282
10.3.1 Introduction	282
10.3.2 Ground Alignment Methods	283
10.3.3 Northfinding Techniques	287
10.4 In-flight Alignment	289
10.4.1 Introduction	289
10.4.2 Sources of Error	289
10.4.3 In-flight Alignment Methods	289
10.5 Alignment at Sea	300
10.5.1 Introduction	300
10.5.2 Sources of Error	300
10.5.3 Shipboard Alignment Methods	301
11. Strapdown Navigation System Computation	309
11.1 Introduction	309
11.2 Attitude Computation	310
11.2.1 Direction Cosine Algorithms	311
11.2.2 Rotation Angle Computation	315
11.2.3 Rotation Vector Compensation	316
11.2.4 Body and Navigation Frame Rotations	318
11.2.5 Quaternion Algorithms	319
11.2.6 Orthogonalisation and Normalisation Algorithms	322
11.2.7 The Choice of Attitude Representation	324

11.3	Acceleration Vector Transformation Algorithm	324
11.3.1	Acceleration Vector Transformation Using Direction Cosines	325
11.3.2	Rotation Correction	326
11.3.3	Dynamic Correction	328
11.3.4	Acceleration Vector Transformation Using Quaternions	329
11.4	Navigation Algorithm	329
11.5	Summary	332
12.	Generalised System Performance Analysis	335
12.1	Introduction	335
12.2	Propagation of Errors in a Two-dimensional Strapdown Navigation System	336
12.2.1	Navigation in a Non-rotating Reference Frame	336
12.2.2	Navigation in a Rotating Reference Frame ..	337
12.2.3	The Schuler Pendulum	339
12.2.4	Propagation of Errors in a Schuler Tuned System	340
12.2.5	Discussion of Results	341
12.3	General Error Equations	342
12.3.1	Derivation of Error Equations	342
12.3.2	Discussion	346
12.4	Analytical Assessment	350
12.4.1	Single Channel Error Model	350
12.4.2	Derivation of Single Channel Error Propagation Equations	352
12.4.3	Single-channel Error Propagation Examples	358
12.5	Assessment by Simulation	360
12.5.1	Introductory Remarks	360

12.5.2	Error Modelling	361
12.5.3	Simulation Techniques	363
12.6	Motion Dependence of Strapdown System Performance	365
12.6.1	Manoeuvre-dependent Error Terms	366
12.6.2	Vibration Dependent Error Terms	368
12.7	Summary	374
13.	Integrated Navigation Systems	377
13.1	Introduction	377
13.2	Basic Principles	378
13.3	External Navigation Aids	379
13.3.1	Radio Navigation Aids	379
13.3.2	Satellite Navigation Aids	384
13.3.3	Star Trackers	391
13.3.4	Surface Radar Trackers	393
13.4	On-board Measurements	394
13.4.1	Doppler Radar	394
13.4.2	Magnetic Measurements	395
13.4.3	Altimeters	396
13.4.4	Terrain Referenced Navigation	397
13.4.5	Scene Matching	398
13.4.6	Continuous Visual Navigation	399
13.5	System Integration	401
13.6	Application of Kalman Filtering to Aided Inertial Navigation Systems	402
13.6.1	Introduction	402
13.6.2	Design Example of Aiding	403
13.7	INS-GPS Integration	409
13.7.1	Uncoupled Systems	411
13.7.2	Loosely Coupled Integration	412

13.7.3	Tightly Coupled Integration	413
13.7.4	Deep Integration	415
13.7.5	Concluding Remarks	416
13.7.6	INS Aiding of GPS Signal Tracking	416
13.8	Multi-sensor Integrated Navigation	417
13.9	Summary	418
14.	Design Example	421
14.1	Introduction	421
14.2	Background to the Requirement	422
14.3	The Navigation System Requirement	423
14.3.1	Navigation Data Required	423
14.3.2	Operating and Storage Environment	423
14.3.3	Performance	424
14.3.4	System Reaction Time	425
14.3.5	Physical Characteristics	425
14.4	Why Choose Strapdown Inertial Navigation?	426
14.5	Navigation System Design and Analysis Process	426
14.5.1	Introduction	426
14.5.2	Choice of System Mechanisation	427
14.5.3	Error Budget Calculations	428
14.5.4	System Alignment	433
14.5.5	Choice of Inertial Instruments	434
14.5.6	Computational Requirements	436
14.5.7	Electrical and Mechanical Interfaces	437
14.6	Testing, Calibration and Compensation Requirements	438
14.7	Performance Enhancement by Aiding	438
14.8	Concluding Remarks	439

15. Alternative Applications of IN Sensors and Systems	441
15.1 Introduction	441
15.2 Borehole Surveying	442
15.2.1 Introduction	442
15.2.2 Historical Background	443
15.2.3 Inertial Survey System	445
15.2.4 System Design Requirements	446
15.2.5 System Design Issues	447
15.2.6 System Calibration and Test	451
15.2.7 Concluding Remarks	452
15.3 Ship's Inertial Navigation Systems (SINS)	453
15.3.1 NATO SINS	454
15.4 Vehicle Stabilisation and Control	456
15.4.1 Autopilots	456
15.4.2 Passive Missile Roll Control (Rollerons)	462
15.4.3 Intelligent Transport Systems – Automotive Applications	464
15.4.4 Intelligent Transport Systems – Trains	467
15.4.5 Personal Transport	467
15.5 Equipment Stabilisation	469
15.5.1 Aero-flexure Compensation	470
15.5.2 Laser Beam Director	475
15.5.3 Laser Radar	479
15.5.4 Seeker-head Stabilisation	482
15.5.5 Sightline Stabilisation	487
15.5.6 Relative Angular Alignment	491
15.5.7 Calibration and Measurement	493
15.6 Geodetic and Geophysical Measurements and Observation of Fundamental Physical Phenomena	495

15.7 Other Applications	499
15.7.1 Moving-map Displays	499
15.7.2 Safety and Arming Units	502
15.7.3 Aircraft Ejection Seats	503
15.7.4 Agricultural Survey	505
15.7.5 Artillery Pointing	505
15.7.6 Other Unusual Applications	507
15.8 Concluding Remarks	508
Appendices	
Appendix A: Kalman Filtering	511
Appendix B: Inertial Navigation System Error Budgets	519
Appendix C: Inertial System Configurations	523
Appendix D: Comparison of GPS and GLONASS Satellite Navigation Systems	529
List of Symbols	535
Glossary of Principal Terms	539
Index	549

Chapter 1

Introduction

1.1 Navigation

Navigation is a very ancient skill or art which has become a complex science. It is essentially about travel and finding the way from one place to another and there are a variety of means by which this may be achieved [1].

Perhaps one of the simplest forms of navigation is the following of directions, or instructions. For example, a person wishing to travel to a given destination from where they are at the moment may be instructed: turn right at the next junction, turn left at the ‘Rose and Crown’, keep to the right of a given landmark, . . . , it will be in front of you! Clearly, this method of navigation relies on the observation and recognition of known features or fixed objects in our surroundings and moving between them. In technical narratives, the locations of these features are often referred to as ‘way-points’.

An extension of this process is navigation by following a map. In this case, the navigator will determine his or her position by observation of geographical features such as roads, rivers, hills and valleys which are shown on the map. These features may be defined on the map with respect to a grid system or ‘reference frame’. For example, positions of terrain features are often defined with respect to the Earth’s Equator and the Greenwich meridian by their latitude and longitude. Hence the navigator is able to determine his or her position in that reference frame. As will become clear later in the text, the use of reference frames is fundamental to the process of navigation.

As an alternative method, the navigator may choose to observe other objects or naturally occurring phenomena to determine his or her position. An ancient and well-established technique is to take sightings of certain of the fixed stars to which the navigator can relate his or her position. The fixed stars effectively define a reference frame which is fixed in space. Such a reference is commonly referred to as an ‘inertial’ reference frame and star sightings enable an observer to determine his or her position with respect to that frame. Given knowledge of the motion of the Earth and the time of the observation, the navigator is able to use the celestial measurements to define his or her position on the surface of the Earth. Navigation systems of this type, which rely upon observation of the outside world, are known as ‘position fixing’ systems.

One of the principal impediments to accurate navigation at sea was the lack of an accurate time reference to determine longitude. Latitude could be determined from the celestial bodies, provided they could be observed; however, the other reference was dependent on the knowledge of the time of the observation. During the eighteenth century, the astronomer royal in the United Kingdom commissioned a number of studies to solve the ‘longitude problem’. The longitude problem was eventually solved by John Harrison, a working class joiner from Lincolnshire with little formal education, who invented an accurate chronometer capable of measuring time on board ship to extraordinary accuracy [2].

An alternative approach is to use the principle of ‘dead reckoning’ by which present position may be calculated from knowledge of initial position and measurements of speed and direction. The process of dead reckoning is performed by taking the last known position and the time at which it was obtained, and noting the average speed and heading since that time, and the current time. The speed must be resolved through the heading angle to give velocity components north and east. Each is then multiplied by the time which has elapsed since the last position was obtained to give the change in position. Finally, the position changes are summed with initial position to obtain present position.

An equivalent process may be conducted using inertial sensors – gyroscopes and accelerometers – to sense rotational and translational motion with respect to an inertial reference frame. This is known as inertial navigation.

1.2 Inertial navigation

The operation of inertial navigation systems depends upon the laws of classical mechanics as formulated by Sir Isaac Newton. Newton’s laws tell us that the motion of a body will continue uniformly in a straight line unless disturbed by an external force acting on the body. The laws also tell us that this force will produce a proportional acceleration of the body. Given the ability to measure that acceleration, it would be possible to calculate the change in velocity and position by performing successive mathematical integrations of the acceleration with respect to time. Acceleration can be determined using a device known as an accelerometer. An inertial navigation system usually contains three such devices, each of which is capable of detecting acceleration in a single direction. The accelerometers are commonly mounted with their sensitive axes perpendicular to one another, i.e. mutually perpendicular.

In order to navigate with respect to our inertial reference frame, it is necessary to keep track of the direction in which the accelerometers are pointing. Rotational motion of the body with respect to the inertial reference frame may be sensed using gyroscopic sensors and used to determine the orientation of the accelerometers at all times. Given this information, it is possible to resolve the accelerations into the reference frame before the integration process takes place.

Hence, inertial navigation is the process whereby the measurements provided by gyroscopes and accelerometers are used to determine the position of the vehicle in which they are installed. By combining the two sets of measurements, it is possible

to define the translational motion of the vehicle within the inertial reference frame and so to calculate its position within it.

Unlike many other types of navigation system, inertial systems are entirely self-contained within the vehicle, in the sense that they are not dependent on the transmission of signals from the vehicle or reception from an external source. However, inertial navigation systems do rely upon the availability of accurate knowledge of vehicle position at the start of navigation. The inertial measurements are then used to obtain estimates of changes in position which take place thereafter.

1.3 Strapdown technology

Whilst the underlying principles of operation are common to all types of inertial navigation system, their implementation may take a variety of different forms. The original applications of inertial navigation technology used stable platform techniques.¹ In such systems, the inertial sensors are mounted on a stable platform and are mechanically isolated from the rotational motion of the vehicle. Platform systems are still in common use, particularly for those applications requiring very accurate estimates of navigation data, such as ships and submarines.

Modern systems have removed most of the mechanical complexity of platform systems by having the sensors attached rigidly, or ‘strapped down’, to the body of the host vehicle. The potential benefits of this approach are lower cost, reduced size and greater reliability compared with equivalent platform systems. As a result, small, light weight and accurate inertial navigation systems may now be fitted to small guided missiles, for instance. The major penalties incurred are a substantial increase in computing complexity and the need to use sensors capable of measuring much higher rates of turn. However, recent advances in computer technology combined with the development of suitable sensors have allowed such designs to become a reality.

Inertial navigation systems of this type, usually referred to as ‘strapdown’ inertial navigation systems, are the subject of this book. Whilst there are many books which described the older and well-established platform technology, no similar book exists which deals explicitly with strapdown systems. It was this fact which provided the primary motivation for this publication.

This text describes the basic concepts of inertial navigation and the technological developments which have led to modern strapdown systems. It is intended to provide an introduction to the subject of strapdown inertial navigation which may be read at various levels by both suppliers of inertial sensors and systems and customers for such products and so encourage a more effective two-way dialogue.

By selective reading, the engineer new to the subject may obtain a background understanding of the subject. For those needing to become more closely involved in the various aspects of strapdown system technology, the text provides a more

¹ A major advance occurred in 1953 with the demonstration of the feasibility of all-inertial navigation in flight trials with a system called SPIRE (Space Inertial Reference Equipment) which was 5 ft in diameter and weighed 2700 lb.

extensive description of system configurations, an appreciation of strapdown inertial sensors and computational requirements and an awareness of techniques which may be used to analyse and assess the performance of such systems. References are provided for those seeking more detailed information on different aspects of the subject.

Strapdown inertial navigation systems rely on complex technology and many technology specific terms and jargon are in common usage. Such terminology is defined in the glossary of terms.

Where appropriate, mathematical descriptions of the physical principles and processes involved are presented. The reader new to the subject, who perhaps wishes to gain an appreciation of physical principles without dwelling on the mathematical details of the processes involved, may merely wish to take note of the results of the more mathematical sections, or possibly to skip over these aspects altogether.

1.4 Layout of the book

Chapter 2 introduces the underlying concepts of strapdown inertial navigation systems with the aid of simplified examples, and culminates in the definition of the basic functions which must be implemented within such a system. It is shown how the measurements of rotational and translational motion are fundamental to the operation of an inertial navigation system. There follows a brief review of the historical developments which have led to the current state of development of strapdown inertial navigation systems. This is accompanied by an outline discussion of system applications.

The way in which the measurements of rotational and translational motion are combined to form an inertial navigation system are addressed more fully in Chapter 3. This chapter deals at some length with attitude computation and the concept of the navigation equation, both of which are fundamental to the operation of strapdown inertial navigation systems. In addition, a number of possible system configurations are described.

Gyroscope and accelerometer technologies are discussed in some detail in Chapters 4–7. This part of the text provides descriptions of the various instrument types currently in use and some which are likely to become available within the foreseeable future. These include conventional angular momentum gyroscopes, optical rate sensors such as the ring laser gyroscope and the fibre optic gyroscope, pendulous force-feedback accelerometers, solid-state devices and cold-atom sensors, as well as multisensors. The text covers mechanical and electronic aspects of the instruments, measurement accuracy, mathematical descriptions and applications. Chapter 7 is devoted to the description and performance of micro-machined electromechanical systems (MEMS) sensor technology which has, in recent years, found broad application in modern navigation and stabilisation systems.

The testing, calibration and compensation of inertial sensors and systems is addressed in Chapter 8. Predictable errors can be corrected or compensated from observation of performance by a process which involves the implementation of

algorithms which are as close as possible to the inverse of the classical sensor error models.

Chapter 9 describes the basic building blocks which combine to form a strapdown inertial system, drawing attention to alternative mechanisations.

A vital factor in the achievement of accurate navigation is the initialisation of the inertial navigation system before the commencement of navigation, prior to take-off in the case of an aircraft navigation system, for instance. This process involves the accurate determination of the position, velocity and attitude of the vehicle navigation system with respect to the chosen reference frame, and is usually referred to as inertial navigation system alignment. Such alignment may have to be undertaken in a vehicle which is moving, as in the case of the in-flight alignment of an airborne inertial navigation system. The difficulties of achieving an accurate alignment in various vehicle applications are highlighted, and techniques for alleviating such problems are described in Chapter 10.

The computer processing of the gyroscope and accelerometer measurements which must be carried out in order to complete the task of navigation is examined in Chapter 11. Computational algorithms are discussed in some detail during the course of this chapter.

Techniques for the analysis of inertial navigation system performance are presented in Chapter 12 to enable the designer to assess system performance. Attention is drawn to a number of errors which are of particular concern in strapdown systems and the use of simulation methods to assess system performance is highlighted.

It is common practice for many applications to combine the outputs of an inertial navigation system with some external measurement data to achieve an overall improvement in navigation accuracy. For example, independent position fixes may be used to aid an inertial navigation system, and so enhance navigation performance beyond that which may be obtained using either the position fixing system or the inertial navigator in isolation. Possible navigation aids are discussed and techniques are presented for mixing inertial and external measurement data to form a so-called integrated navigation system in Chapter 13.

Chapter 14 draws together much of the preceding text through the discussion of a design example. Because the background of the authors is predominantly in the field of guided missile systems, this part of the book is directed at such an application. The design example will be of particular interest to the engineer wishing to specify a system to meet a given requirement and to assess its potential performance.

Finally, Chapter 15 describes a number of applications of inertial navigation systems for both military and civil roles covering systems that operate at sea, in the air, as well as on and below the ground. This chapter aims to describe the particular problems encountered in attempting to design navigation and stabilisation systems fit for the broad range of roles in which this technology has been applied.

The appendices provide descriptions of Kalman filtering techniques, inertial navigation error budgets, inertial system configurations and a comparison of GPS and GLONASS satellite navigation systems. A glossary of principal terms used is given at the end of the book.

References

- 1 ANDERSON, E.W.: 'The principles of navigation' (Hollis and Carter, 1966)
- 2 SOBEL, D.: 'Longitude: the true story of a lone genius who solved the greatest scientific problem of his time' (Penguin Books, 1996)

Chapter 2

Fundamental principles and historical developments of inertial navigation

2.1 Basic concepts

The basic concepts of inertial navigation are outlined here with the aid of some simple examples.

A simple one-dimensional example of navigation involves the determination of the position of a train which is moving along a track between two locations on a perfectly flat plane. It is possible to determine the instantaneous speed of the train and the distance it has travelled from a known starting point by using measurements of its acceleration along the track. Sensors called accelerometers provide such information about their own movement. If an accelerometer is fixed in the train, it will provide information about the acceleration of the train. The time integral of the acceleration measurement provides a continuous estimate of the instantaneous speed of the train, provided its initial speed was known. A second integration yields the distance travelled with respect to a known starting point. The accelerometer together with a computer, or other suitable device capable of integration, therefore constitutes a simple one-dimensional navigation system.

In general, a navigation system is required to provide an indication of the position of a vehicle with respect to a known grid system or reference frame. For instance, it may be required to determine the location of a vehicle in terms of x and y coordinates in a Cartesian reference frame. Considering again the example of a train moving along a track, as depicted in Figure 2.1, it is now necessary to determine the position of the train with respect to the coordinate reference frame shown in the figure.

Given the knowledge of the train's acceleration along the track, and the angle which the track makes with the reference frame, the x and y coordinate positions may be determined. This may be accomplished by resolving the measured acceleration in the reference frame to give x and y components, and by suitable integration of the resolved signals to yield the velocity and position of the train in reference axes.

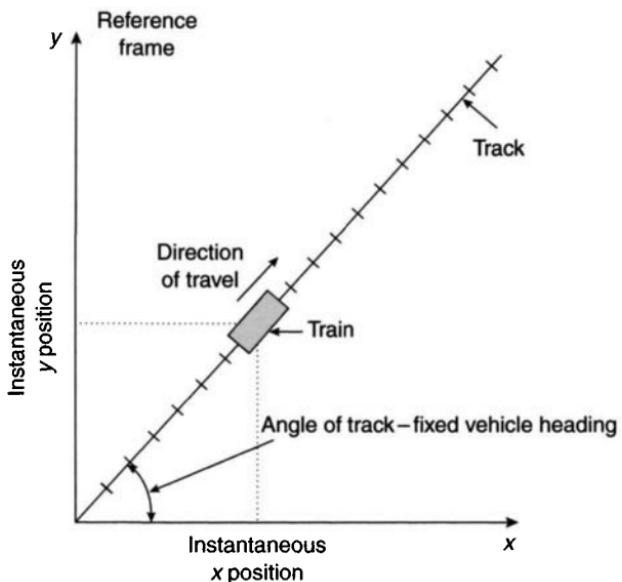


Figure 2.1 One-dimensional navigation

In this simple case, the angle of the track defines the heading of the train with respect to the reference frame.

With the more general situation illustrated in Figure 2.2, where the track curves, it is necessary to detect continuously the translational motion of the train in two directions and changes in its direction of travel, that is, to detect the rotations of the train about the perpendicular to the plane of motion as the train moves along the track.

Two accelerometers are now required to detect the translational motion in perpendicular directions along and perpendicular to the track. One sensor suitable for the measurement of the rotational motion is a gyroscope. Depending on the form of construction of this sensor, it may be used to provide either a direct measure of the train's heading with respect to the reference frame, or a measurement of the turn rate of the train. In the latter case, the angular orientation of the train may be calculated by the integration of this measurement, provided the angle is known at the start of navigation. Given such information, it is possible to relate the measurements of acceleration, which are obtained in an axis set which is fixed in the train, to the reference frame. The instantaneous measurements of acceleration may therefore be resolved in the reference frame and integrated with respect to time to determine the instantaneous velocity and position of the vehicle with respect to that frame.

Clearly then, it is possible to construct a simple, two-dimensional, navigation system using a gyroscope, two accelerometers and a computer. In practice, the inertial sensors may be mounted on a platform which is stabilised in space, and hence isolated from the rotation of the vehicle, or mounted directly on to the vehicle to form a strapdown system. The measurements are processed in the computer to provide continuous estimates of the position, speed and the direction of travel or heading of

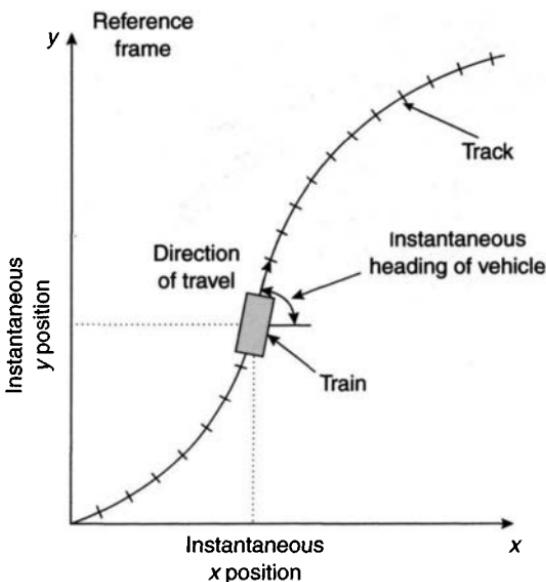


Figure 2.2 Two-dimensional navigation

the train. It must be stressed that inertial navigation is fundamentally dependent on an accurate knowledge of position, speed and heading being available prior to the start of navigation. This is because it uses dead reckoning which relies for its operation on the updating of the system's previous estimates of these navigational quantities, commencing with the initial values input to the system at the start of navigation.

It will be apparent from the preceding discussion that successful navigation of a vehicle can be achieved by using the properties of suitable sensors mounted in the vehicle. In general, it is required to determine a vehicle's position with respect to a three-dimensional reference frame. Consequently, if single-axis sensors are used, three gyroscopes will be required to provide measurements of vehicle turn rates about three separate axes, whilst three accelerometers provide the components of acceleration which the vehicle experiences along these axes. For convenience and accuracy, the three axes are usually chosen to be mutually perpendicular.

In most applications, the axis set defined by the sensitive axes of the inertial sensors is made coincident with the axes of the vehicle, or body, in which the sensors are mounted, usually referred to as the body axis set. The measurements provided by the gyroscopes are used to determine the attitude and heading of the body with respect to the reference frame in which it is required to navigate. The attitude and heading information is then used to resolve the accelerometer measurements into the reference frame. The resolved accelerations may then be integrated twice to obtain vehicle velocity and position in the reference frame.

Gyroscopes provide measurements of changes in vehicle attitude or its turn rate with respect to inertial space. Accelerometers, however, are unable to separate the total acceleration of the vehicle, the acceleration with respect to inertial space, from

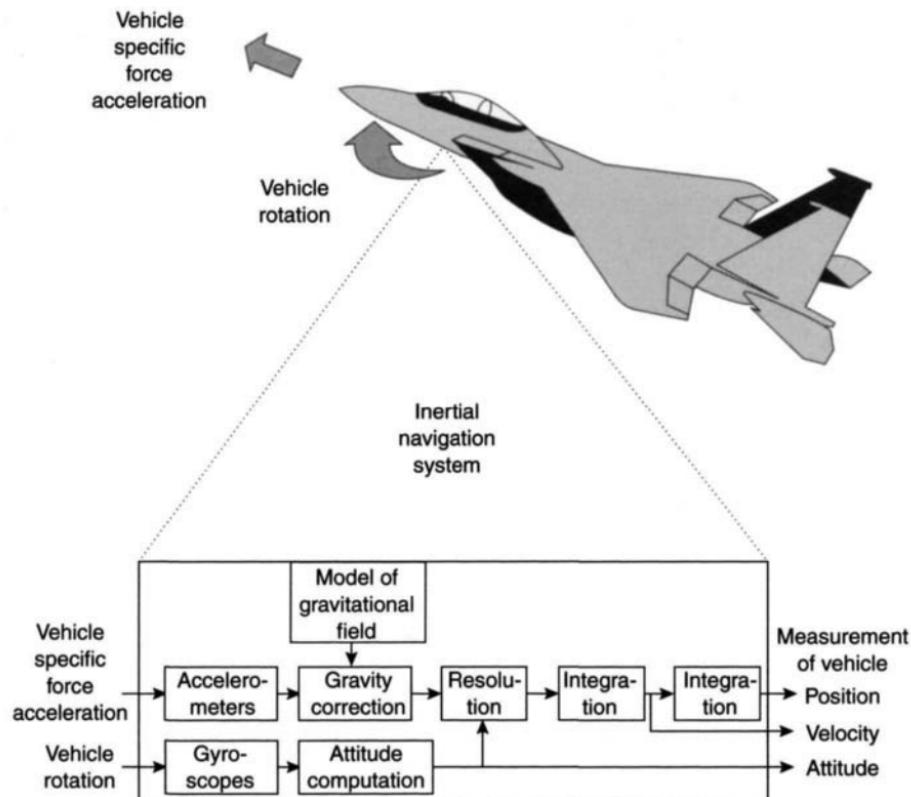


Figure 2.3 Functional components of an inertial navigation system

that caused by the presence of a gravitational field. These sensors do in fact provide measurements of the difference between the true acceleration in space and the acceleration due to gravity.¹ This quantity is the non-gravitational force per unit mass exerted on the instrument, referred to in this text for brevity as 'specific force' [1].

Hence, the measurements provided by the accelerometers, especially when close to a large body such as the Earth, must be combined with knowledge of the gravitational field of that body in order to determine the acceleration of the vehicle with respect to inertial space. Using this information, vehicle acceleration relative to the body may be derived.

The navigational function is therefore fulfilled by combining the measurements of vehicle rotation and specific force with knowledge of the gravitational field to compute estimates of attitude, velocity and position with respect to a pre-defined reference frame. A schematic representation of such an inertial navigation system is shown in Figure 2.3.

¹ Algebraically, the sum of the acceleration with respect to inertial space and the acceleration due to gravitational attraction.

2.2 Summary

It follows from the introductory discussion that the essential functions which an inertial navigation system must perform may be defined as follows:

- determination of the angular motion of a vehicle using gyroscopic sensors, from which its attitude relative to a reference frame may be derived;
- measure specific force using accelerometers;
- resolve the specific force measurements into the reference frame using the knowledge of attitude derived from the information provided by the gyroscopes;
- evaluate the force resulting from the gravitational field – the gravitational attraction of the Earth in the case of systems operating in the vicinity of the Earth;
- integrate the resolved specific force measurements to obtain estimates of the velocity and position of the vehicle.

The later chapters describe the principles of inertial navigation in some depth and provide detailed information on system mechanisations, inertial sensor technology, computational aspects (including algorithms), design analysis and applications of such systems. However, prior to this, it is instructive to have a brief review of the historical developments which have led to the current state of development of present day inertial navigation systems and their technology.

2.3 Historical developments

From the earliest times, people have moved from one place to another by finding or ‘knowing’ their way; this skill has required some form of navigation. There is an oblique reference to inertial navigation in the Bible [2]. Generally, as in the case of the biblical reference, the earliest applications were on land. Then as the desire developed to explore farther afield, instruments were developed for marine applications. More recently, there have been significant developments in inertial sensors, and systems for inertial navigation on land, in the air, on or under the oceans as well as in space to the planets and beyond.

Our earliest ancestors travelled in search of food, usually on land. As they developed, they crossed rivers generally using landmarks, that is, navigation by observation. Further development of position fixing techniques saw the Polynesians cross the Pacific Ocean about two millennia ago using their understanding of celestial bodies and landmarks. These techniques can only be used in clear weather conditions. During the thirteenth century, the Chinese discovered the properties of lodestone and applied the principles of magnetism to fabricate a compass. They used this instrument to navigate successfully across the south China Sea. This device could be used irrespective of visibility but was difficult to use in rough weather. The other significant development to help the long distance traveller was the sextant, which enabled position fixes to be made accurately on land.

In the seventeenth century, Sir Isaac Newton defined the laws of mechanics and gravitation, which are the fundamental principles on which inertial navigation is based. Despite this, it was to be about another two centuries before the inertial sensors were developed that would enable the demonstration of inertial navigation techniques. However, in the early eighteenth century, there were several significant developments; Serson demonstrating a stabilised sextant [3] and Harrison devising an accurate chronometer, the former development enabling sightings to be taken of celestial objects without reference to the horizon and the latter enabling an accurate determination of longitude. These instruments, when used with charts and reference tables of location of celestial bodies, enabled accurate navigation to be achieved, provided the objects were visible.

Foucault is generally credited with the discovery of the gyroscopic effect in 1852. He was certainly the first to use the word. There were others, such as Bohneberger, Johnson and Lemerle, developing similar instruments. All of these people were investigating the rotational motion of the Earth and the demonstration of rotational dynamics. They were using the ability of the spin axis of a rotating disc to remain fixed in space. Later in the nineteenth century, many fine gyroscopic instruments were made. In addition, there were various ingenious applications of the gyroscopic principle in heavy equipment such as the grinding mill.

A significant discovery was made in 1890 by Professor G.H. Bryan concerning the ringing of hollow cylinders, a phenomenon later applied to solid-state gyroscopes.

The early years of the twentieth century saw the development of the gyrocompass for the provision of a directional reference. The basic principle of this instrument is the indication of true north by establishing the equilibrium between the effect of its pendulosity and the angular momentum of the rotating base carrying the compass. Initially, this instrument was sensitive to acceleration. Professor Max Schuler produced an instrument with a vertical erection system enabling an accurate vertical reference to be defined [4]. This instrument was tuned to the undamped natural period defined by $2\pi\sqrt{R/g}$, approximately equal to 84 minutes, where R is the radius of the Earth and g is the acceleration caused by the Earth's gravitational field. Later, this technique became known as 'Schuler tuning' [5], a phrase originated by Dr Walter Wrigley of MIT. This ingenious method produced a directional instrument insensitive to acceleration for use at sea. Elmer and Lawrence Sperry improved the design of the gyrocompass with further refinements by Brown and Perry. These instruments provided the first steps towards all-weather, autonomous navigation. The Sperry brothers were also at the forefront of the application of the gyroscopic effect to control and guidance in the early twentieth century. They produced navigation and autopilot equipment for use in aircraft and gyroscopes for use in torpedoes.

Rate of turn indicators, artificial horizons and directional gyroscopes for aircraft were being produced in the 1920s. At a similar time, side-slip sensors were being developed, early open-loop accelerometers, and Schuler was demonstrating a north-seeking device for land use giving an accuracy of 22 seconds of arc! There was significant progress during the early part of the twentieth century with the development of stable platforms for fire control systems for guns on ships and the identification of the concept for an inertial navigation system. Boykow identified the use of accelerometers and gyroscopes to produce a full inertial navigation system.

However, at this stage, the quality of the inertial sensors was not suitable for the production and demonstration of such a system.

World War II saw the demonstration of the principles of inertial guidance in the V1 and V2 rockets by German scientists, a prime step forward being the use of a system with feedback leading to accurate guidance. At this time there was much activity in various parts of the world devising new types of inertial sensors, improving their accuracy and, in 1949, the first publication suggesting the concept of the strapdown technique for navigation.

The pace of development and innovation quickened in the 1950s with many significant developments for seaborne and airborne applications. More accurate sensors were produced, with the accuracy of the gyroscope being increased substantially. The error in such sensors was reduced from about $15^\circ/\text{hour}$ to about $0.01^\circ/\text{hour}$, Professor Charles Stark Draper and his coworkers at MIT being largely responsible for many technical advances with the demonstration of the floated rate-integrating gyroscope [6]. It was also during the 1950s that the principle of force-feedback was applied to the proof mass in an accelerometer to produce an accurate acceleration sensing instrument.

The early part of the 1950s saw the fabrication of a stabilised platform inertial navigation system followed by the first crossing of the United States of America by an aircraft using full inertial navigation. Inertial navigation systems became standard equipment in military aircraft, ships and submarines during the 1960s, all of these applications using the so-called stable platform technology. This era saw further significant developments with increases in the accuracy of sensors, the miniaturisation of these devices and the start of ring laser gyroscope development. Major projects of this period in which inertial system technology was applied were the ballistic missile programmes and the exploration of space.

Similar progress has taken place in the last two decades; one major advance being the application of the micro-computer and development of gyroscopes with large dynamic ranges enabling the strapdown principle to be realised. This has enabled the size and complexity of the inertial navigation system to be reduced significantly for very many applications. The use of novel methods has enabled small, reliable, rugged and accurate inertial sensors to be produced that are relatively inexpensive, thus enabling a very wide range of diverse applications as discussed below. This period has also seen significant advances in the development of solid-state sensors such as optical fibre gyroscopes and silicon accelerometers.

The development of inertial navigation systems in recent years has been characterised by the gradual move from stable platform to strapdown technology as indicated in Figure 2.4. The figure gives an indication of the increasing application of strapdown systems which has resulted from advances in gyroscope technology. Milestones in this continuing development have occurred as a result of the development of the miniature rate-integrating gyroscope, the dynamically tuned gyroscope and more recently, ring laser and fibre optic rate sensors and vibratory gyroscopes, all of which are described in Chapters 4 and 5. MEMS sensors have provided an exciting development that should expand the range of applications of inertial navigation.

Strapdown systems are becoming widely used for aircraft and guided missile applications. More recently this technology has been applied to ship and submarine

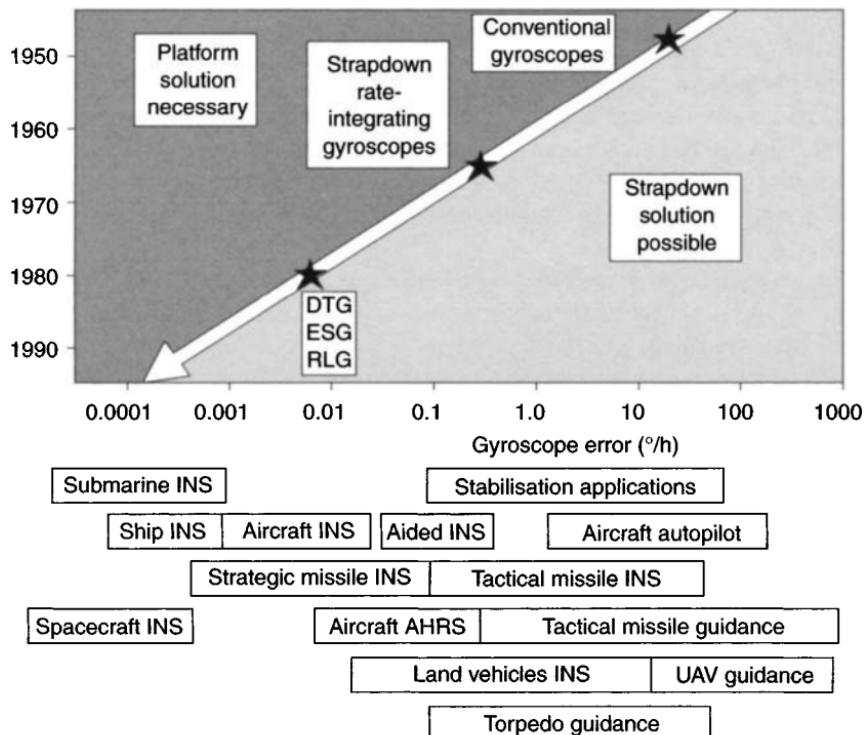


Figure 2.4 Strapdown sensor development and some applications

applications as indicated in Figure 2.4. The diagram shows other applications for strapdown technology, the accuracy required from the gyroscopes being related to the position and size of the box in which it is mentioned.

2.4 The modern-day inertial navigation system

From the preceding section, it is clear that the range of applications in which inertial navigation systems can and are being used is very extensive, covering navigation of ships, aircraft, tactical and strategic missiles and spacecraft. In addition, there are some more novel applications in the field of robotics, active suspension in racing or high performance motor cars and for surveying underground wells and pipelines.

Such diverse applications call for navigation systems having a very broad range of performance capabilities, as well as large differences in the periods of time over which they will be required to provide navigation data. For instance, tactical missile applications may require inertial navigation and guidance to an accuracy of a few hundred metres for periods of minutes or even a few seconds, whilst other airborne systems are required to operate for several hours whilst maintaining knowledge of aircraft position to an accuracy of one or two nautical miles or better. In the cases

of marine or space applications, such systems may be required to provide navigation data to similar accuracy over periods of weeks, months or even longer in the case of interplanetary exploration. One extreme example is the Voyager spacecraft which has been navigating through the solar system and beyond for more than 25 years.

Although the basic principles of inertial navigation systems do not change from one application to another, it will come as no surprise to find that the accuracy of the inertial sensors and the precision to which the associated computation must be carried out varies dramatically over the broad range of applications indicated earlier. It follows therefore that the instrument technologies and the techniques used for the implementation of the navigation function in such diverse applications also vary greatly. Part of the function of this text is to provide some insight into the methods and technologies appropriate to some of the different types of inertial system application outlined earlier.

2.5 Trends in inertial sensor development

A number of categories of inertial sensor have made significant progress over the last decade. The development of micro-machined electromechanical systems (MEMS) devices has been spectacular and the performance that can be achieved is approaching inertial grade, so that is likely to be readily achieved in the near term. The development of micro-optical machined electromechanical systems (MOEMS) is expected to provide very high performance sensors in the medium term.

Progress with refinement of the fibre optical gyroscope is likely to continue so it should start to replace the more expensive ring laser gyroscope in the near term.

New techniques, such as the cold atom interferometers, are being researched for highly specialised investigations and very precise measurements, but these approaches are a long way from maturity.

The big drive from the systems applications viewpoint will be further reduction in cost and complexity, with a corresponding leap in reliability.

Inertial sensor performance is discussed in Chapters 4–7.

References

- 1 BRITTING, K.: 'Inertial navigation system analysis' (Wiley Interscience, New York, 1971)
- 2 The Bible, Amos 7:7–9
- 3 SORG, H.W.: 'From Serson to Draper – two centuries of gyroscopic development', *Journal of Institute of Navigation*, 1976–77, **23** (4)
- 4 SCHULER, M.: 'Die Störung von Pendel-und Kreiselapparaten durch die Beschleunigung der Fahrzeuges', *Physikalische Zeitschrift B*, 1923, **24**
- 5 WRIGLEY, W., HOLLISTER, W.M., and DENHARD, W.G.: 'Gyroscopic theory, design and instrumentation' (MIT Press, 1969)
- 6 DRAPER, C., WRIGLEY, W., and HOVORKA, J.: 'Inertial guidance' (Pergamon Press, 1960)

Chapter 3

Basic principles of strapdown inertial navigation systems

3.1 Introduction

The previous chapter has provided some insight into the basic measurements that are necessary for inertial navigation. For the purposes of the ensuing discussion, it is assumed that measurements of specific force and angular rate are available along and about axes which are mutually perpendicular. Attention is focused on how these measurements are combined and processed to enable navigation to take place.

3.2 A simple two-dimensional strapdown navigation system

We begin this chapter by describing a simplified two-dimensional strapdown navigation system. Although functionally identical to the full three-dimensional system discussed later, the computational processes which must be implemented to perform the navigation task in two dimensions are much simplified compared with a full strapdown system. Therefore, through this introductory discussion, it is hoped to provide the reader with an appreciation of the basic processing tasks which must be implemented in a strapdown system without becoming too deeply involved in the intricacies and complexities of the full system computational tasks.

For the purposes of this discussion, it is assumed that a system is required to navigate a vehicle which is constrained to move in a single plane. A two-dimensional strapdown system capable of fulfilling this particular navigation task was introduced very briefly in Chapter 2 and is shown diagrammatically in Figure 3.1.

The system contains two accelerometers and a single axis rate gyroscope, all of which are attached rigidly to the body of the vehicle. The vehicle body is represented, in the figure, by the block on which the instruments shown are mounted. The sensitive axes of the accelerometers, indicated by the directions of the arrows in the diagram, are at right angles to one another and aligned with the body axes of the

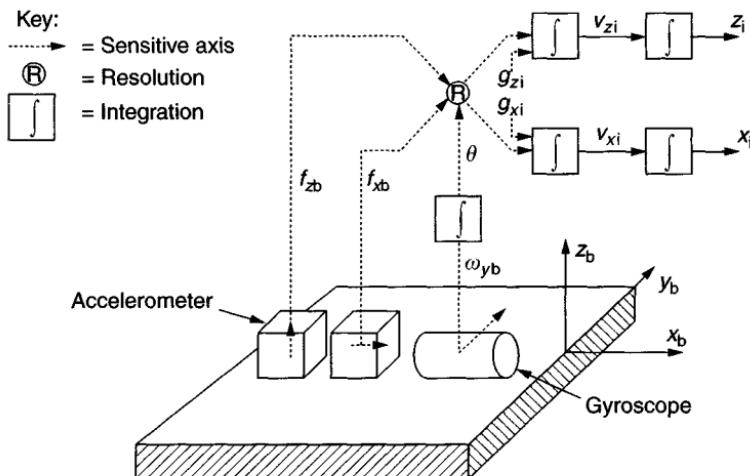


Figure 3.1 Two-dimensional strapdown inertial navigation system

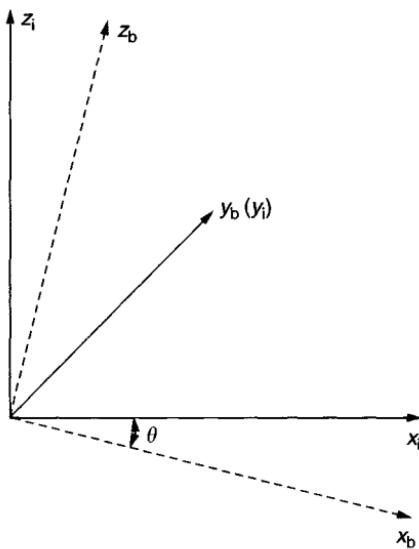


Figure 3.2 Reference frames for two-dimensional navigation

vehicle in the plane of motion; they are denoted as the x_b and z_b axes. The gyroscope is mounted with its sensitive axis orthogonal to both accelerometer axes allowing it to detect rotations about an axis perpendicular to the plane of motion; the y_b axis. It is assumed that navigation is required to take place with respect to a space-fixed reference frame denoted by the axes x_i and z_i . The reference and body axis sets are shown in Figure 3.2, where θ represents the angular displacement between the body and reference frames.

$$\begin{aligned}\dot{\theta} &= \omega_{yb} \\ f_{xi} &= f_{xb} \cos \theta + f_{zb} \sin \theta \\ f_{zi} &= -f_{xb} \sin \theta + f_{zb} \cos \theta \\ \dot{v}_{xi} &= f_{xi} + g_{xi} \\ \dot{v}_{zi} &= f_{zi} + g_{zi} \\ \dot{x}_i &= v_{xi} \\ \dot{z}_i &= v_{zi}\end{aligned}$$

Figure 3.3 Two-dimensional strapdown navigation system equations

Referring now to Figure 3.1, body attitude, θ , is computed by integrating the measured angular rate, ω_{yb} , with respect to time. This information is then used to resolve the measurements of specific force, f_{xb} and f_{zb} , into the reference frame. A gravity model, stored in the computer, is assumed to provide estimates of the gravity components in the reference frame, g_{xi} and g_{zi} . These quantities are combined with the resolved measurements of specific force, f_{xi} and f_{zi} , to determine true accelerations, denoted by \dot{v}_{xi} and \dot{v}_{zi} . These derivatives are subsequently integrated twice to obtain estimates of vehicle velocity and position. The full set of equations which must be solved are given in Figure 3.3.

Having defined the basic functions which must be implemented in a strapdown inertial navigation system, consideration is now given to the application of the two-dimensional system, described above, for navigation in a rotating reference frame. For instance, consider the situation where it is required to navigate a vehicle moving in a meridian plane around the Earth, as depicted in Figure 3.4. Hence, we are concerned here with a system which is operating in the vertical plane alone. Such a system would be required to provide estimates of velocity with respect to the Earth, position along the meridian and height above the Earth.

Whilst the system mechanisation as described could be used to determine such information, this would entail a further transformation of the velocity and position, derived in space fixed coordinates, to a geographic frame. An alternative and often used approach is to navigate directly in a local geographic reference frame, defined in this simplified case by the direction of the local vertical at the current location of the vehicle. In order to provide the required navigation information, it now becomes necessary to keep track of vehicle attitude with respect to the local geographic frame denoted by the axes x and z . This information can be extracted by differencing the successive gyroscopic measurements of body turn rate with respect to inertial space, and the current estimate of the turn rate of the reference frame with respect to inertial space. For a vehicle moving at a velocity, v_x , in a single plane around a perfectly spherical Earth of radius R_0 , this rate is given by $v_x/(R_0 + z)$ where z is the height of the vehicle above the surface of the Earth. This is often referred to as the transport rate.

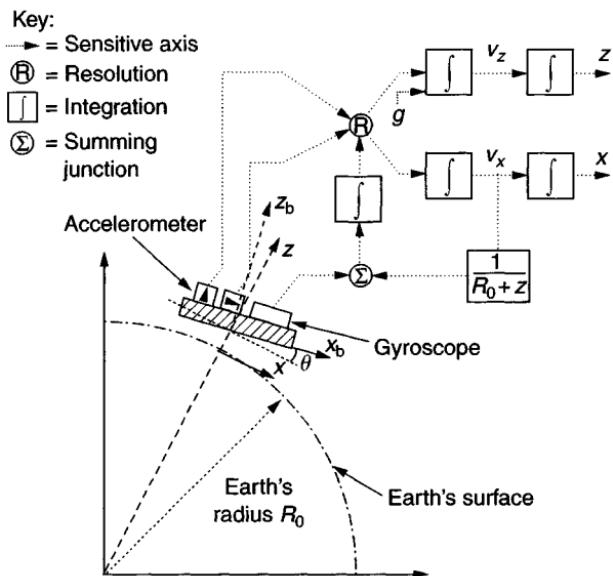


Figure 3.4 Two-dimensional strapdown inertial system for navigation in a rotating reference frame

$$\begin{aligned}
 \dot{\theta} &= \omega_{yb} - v_x / (R_0 + z) \\
 f_x &= f_{xb} \cos \theta + f_{zb} \sin \theta \\
 f_z &= -f_{xb} \sin \theta + f_{zb} \cos \theta \\
 \dot{v}_x &= f_x + v_x v_z / (R_0 + z) \\
 \dot{v}_z &= f_z + g - v_x^2 / (R_0 + z) \\
 \dot{x} &= v_x \\
 \dot{z} &= v_z
 \end{aligned}$$

Figure 3.5 Simplified two-dimensional strapdown system equations for navigation in a rotating reference frame

Figure 3.4 shows a modified two-dimensional strapdown system for navigation in the moving reference frame. As shown in the figure, an estimate of the turn rate of the reference frame is derived using the estimated component of horizontal velocity.

The equations which must be solved in this system are given in Figure 3.5.

Comparison with the equations given in Figure 3.3, relating to navigation with respect to a space-fixed axis set, reveals the following differences. The attitude computation is modified to take account of the turn rate of the local vertical reference frame

as described above. Consequently, the equation in θ is modified by the subtraction of the term $v_x/(R_0 + z)$ in Figure 3.4. The terms $v_x v_z/(R_0 + z)$ and $v_x^2/(R_0 + z)$ which appear in the velocity equations are included to take account of the additional forces acting as the system moves around the Earth (Coriolis forces, see Section 3.4). The gravity term (g) appears only in the v_z equation as it is assumed that the Earth's gravitational acceleration acts precisely in the direction of the local vertical.

This section has outlined the basic form of the computing tasks to be implemented in a strapdown navigation system using a much simplified two-dimensional representation. In the remainder of this chapter the extension of this simple strapdown system to three dimensions is described in some detail. It will be appreciated that this entails a substantial increase in the complexity of the computing tasks involved. In particular, attitude information in three dimensions can no longer be obtained by a simple integration of the measured turn rates.

3.3 Reference frames

Fundamental to the process of inertial navigation is the precise definition of a number of Cartesian co-ordinate reference frames. Each frame is an orthogonal, right-handed, co-ordinate frame or axis set.

For navigation over the Earth, it is necessary to define axis sets which allow the inertial measurements to be related to the cardinal directions of the Earth, that is, frames which have a physical significance when attempting to navigate in the vicinity of the Earth. Therefore, it is customary to consider an inertial reference frame which is stationary with respect to the fixed stars, the origin of which is located at the centre of the Earth. Such a reference frame is shown in Figure 3.6, together with an Earth-fixed reference frame and a local geographic navigation frame defined for the purposes of terrestrial inertial navigation.

The following co-ordinate frames are used in the text:

The inertial frame (i-frame) has its origin at the centre of the Earth and axes which are non-rotating with respect to the fixed stars, defined by the axes Ox_i , Oy_i , Oz_i , with Oz_i coincident with the Earth's polar axis (which is assumed to be invariant in direction).

The Earth frame (e-frame) has its origin at the centre of the Earth and axes which are fixed with respect to the Earth, defined by the axes Ox_e , Oy_e , Oz_e with Oz_e along the Earth's polar axis. The axis Ox_e lies along the intersection of the plane of the Greenwich meridian with the Earth's equatorial plane. The Earth frame rotates, with respect to the inertial frame, at a rate Ω about the axis Oz_i .

The navigation frame (n-frame) is a local geographic frame which has its origin at the location of the navigation system, point P, and axes aligned with the directions of north, east and the local vertical (down). The turn rate of the navigation frame, with respect to the Earth-fixed frame, ω_{en} , is governed by the motion of the point P with respect to the Earth. This is often referred to as the transport rate.

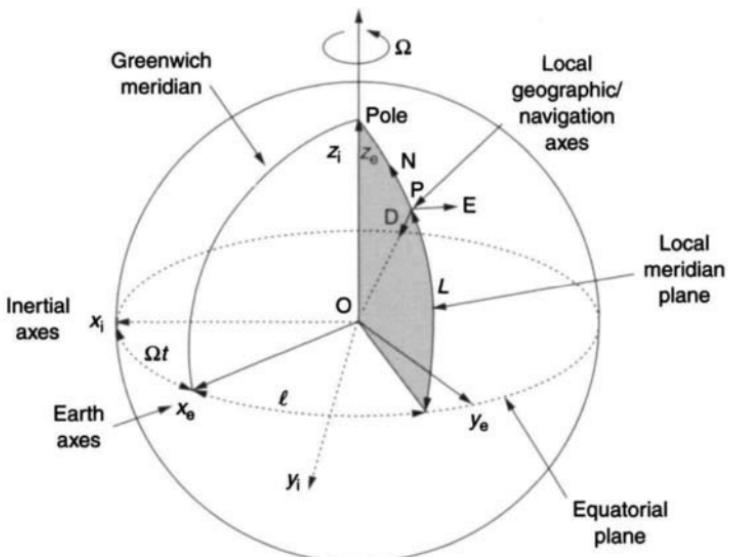


Figure 3.6 Frames of reference

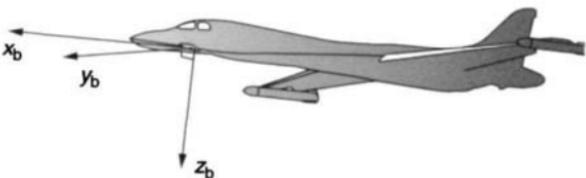


Figure 3.7 Illustration of a body reference frame

The *wander azimuth frame* (*w-frame*) may be used to avoid the singularities in the computation which occur at the poles of the navigation frame. Like the navigation frame, it is locally level but is rotated through the *wander angle* about the local vertical. Its use is described in Section 3.5.

The *body frame* (*b-frame*), depicted in Figure 3.7, is an orthogonal axis set which is aligned with the roll, pitch and yaw axes of the vehicle in which the navigation system is installed.

3.4 Three-dimensional strapdown navigation system – general analysis

3.4.1 Navigation with respect to a fixed frame

Consider the situation where it is required to navigate with respect to a fixed, or non-accelerating, and non-rotating set of axes. The measured components of specific

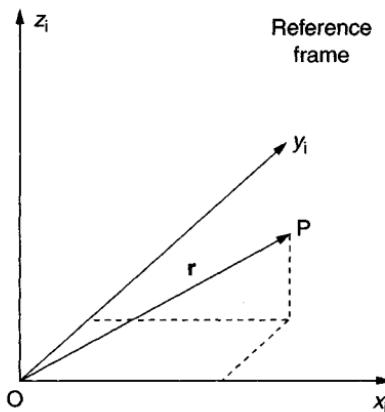


Figure 3.8 Position vector with respect to reference frame

force and estimates of the gravitational field are summed to determine components of acceleration with respect to a space-fixed reference frame. These quantities can then be integrated twice, giving estimates of velocity and position in that frame.

This process may be expressed mathematically in the following manner.¹ Let \mathbf{r} represent the position vector of the point P with respect to O, the origin of the reference frame shown in Figure 3.8.

The acceleration of P with respect to a space-fixed axis set, termed the i-frame and denoted by the subscript i, is defined by:

$$\mathbf{a}_i = \left. \frac{d^2 \mathbf{r}}{dt^2} \right|_i \quad (3.1)$$

A triad of perfect accelerometers will provide a measure of the specific force (\mathbf{f}) acting at point P where

$$\mathbf{f} = \left. \frac{d^2 \mathbf{r}}{dt^2} \right|_i - \mathbf{g} \quad (3.2)$$

in which \mathbf{g} is the mass attraction gravitation vector.

Rearranging eqn. (3.2) yields the following equation:

$$\left. \frac{d^2 \mathbf{r}}{dt^2} \right|_i = \mathbf{f} + \mathbf{g} \quad (3.3)$$

This is called the navigation equation since, with suitable integration, it yields the navigational quantities of velocity and position. The first integral gives the velocity

¹ Vector and matrix notation is widely used throughout the text for the mathematical representation of strapdown inertial system processes. This notation is adopted both in the interests of brevity and to be consistent with other texts on the subject. Vector and matrix quantities are written in boldface type.

of point P with respect to the i-frame, viz.

$$\mathbf{v}_i = \frac{d\mathbf{r}}{dt} \Big|_i \quad (3.4)$$

whilst a second integration gives its position in that frame.

3.4.2 Navigation with respect to a rotating frame

In practice, one often needs to derive estimates of a vehicle's velocity and position with respect to a rotating reference frame, as when navigating in the vicinity of the Earth. In this situation, additional apparent forces will be acting which are functions of reference frame motion. This results in a revised form of the navigation equation which may be integrated to determine the ground speed of the vehicle, \mathbf{v}_e , directly. Alternatively, \mathbf{v}_e may be computed from the inertial velocity, \mathbf{v}_i , using the theorem of Coriolis, as follows,

$$\mathbf{v}_e = \frac{d\mathbf{r}}{dt} \Big|_e = \mathbf{v}_i - \boldsymbol{\omega}_{ie} \times \mathbf{r} \quad (3.5)$$

where $\boldsymbol{\omega}_{ie} = [0 \ 0 \ \Omega]^T$ is the turn rate of the Earth frame with respect to the i-frame and \times denotes a vector cross product.

Revised forms of the navigation equation suitable for navigation with respect to the Earth are the subject of Section 3.5.

3.4.3 The choice of reference frame

The navigation equation, eqn. (3.3), may be solved in any one of a number of reference frames. If the Earth frame is chosen, for example, then the solution of the navigation equation will provide estimates of velocity with respect to either the inertial frame or the Earth frame, expressed in Earth coordinates, denoted \mathbf{v}_i^e and \mathbf{v}_e^e , respectively.²

In Section 3.5, a number of different strapdown system mechanisations for navigating with respect to the Earth are described. In each case, it will be shown that the navigation equation is expressed in a different manner depending on the choice of reference frame.

3.4.4 Resolution of accelerometer measurements

The accelerometers usually provide a measurement of specific force in a body fixed axis set, denoted \mathbf{f}^b . In order to navigate, it is necessary to resolve the components of the specific force in the chosen reference frame. In the event that the inertial frame is selected, this may be achieved by pre-multiplying the vector quantity \mathbf{f}^b by the direction cosine matrix, \mathbf{C}_b^i , using,

$$\mathbf{f}^i = \mathbf{C}_b^i \mathbf{f}^b \quad (3.6)$$

² Superscripts attached to vector quantities denote the axis set in which the vector quantity coordinates are expressed.

where \mathbf{C}_b^i is a 3×3 matrix which defines the attitude of the body frame with respect to the i-frame. The direction cosine matrix \mathbf{C}_b^i may be calculated from the angular rate measurements provided by the gyroscopes using the following equation:

$$\dot{\mathbf{C}}_b^i = \mathbf{C}_b^i \boldsymbol{\Omega}_{ib}^b \quad (3.7)$$

where $\boldsymbol{\Omega}_{ib}^b$ is the skew symmetric matrix:

$$\boldsymbol{\Omega}_{ib}^b = \begin{bmatrix} 0 & -r & q \\ r & 0 & -p \\ -q & p & 0 \end{bmatrix} \quad (3.8)$$

This matrix is formed from the elements of the vector $\omega_{ib}^b = [p \ q \ r]^T$ which represents the turn rate of the body with respect to the i-frame as measured by the gyroscopes. Equation (3.7) is derived in Section 3.6.

The attitude of the body with respect to the chosen reference frame, which is required to resolve the specific force measurements into the reference frame, may be defined in a number of different ways. For the purposes of the discussion of navigation system mechanisations in this and the following section, the direction cosine method will be adopted. Direction cosines and some alternative attitude representations are described in some detail in Section 3.6.

3.4.5 System example

Consider the situation in which it is required to navigate with respect to inertial space and the solution of the navigation takes place in the i-frame. Equation (3.3) may be expressed in i-frame coordinates as follows:

$$\frac{d^2\mathbf{r}}{dt^2}\Big|_i = \mathbf{f}^i + \mathbf{g}^i = \mathbf{C}_b^i \mathbf{f}^b + \mathbf{g}^i \quad (3.9)$$

It is clear from the preceding discussion that the integration of the navigation equation involves the use of information from both the gyroscopes and the accelerometers contained within the inertial navigation system. A block diagram representation of the resulting navigation system is given in Figure 3.9.

The diagram displays the main functions to be implemented within a strapdown navigation system; the processing of the rate measurements to generate body attitude, the resolution of the specific force measurements into the inertial reference frame, gravity compensation and the integration of the resulting acceleration estimates to determine velocity and position.

3.5 Strapdown system mechanisations

Attention is focused here on inertial systems which may be used to navigate in the vicinity of the Earth. It has been shown in Section 3.4 how estimates of position and velocity are derived by integrating a navigation equation of the form given in

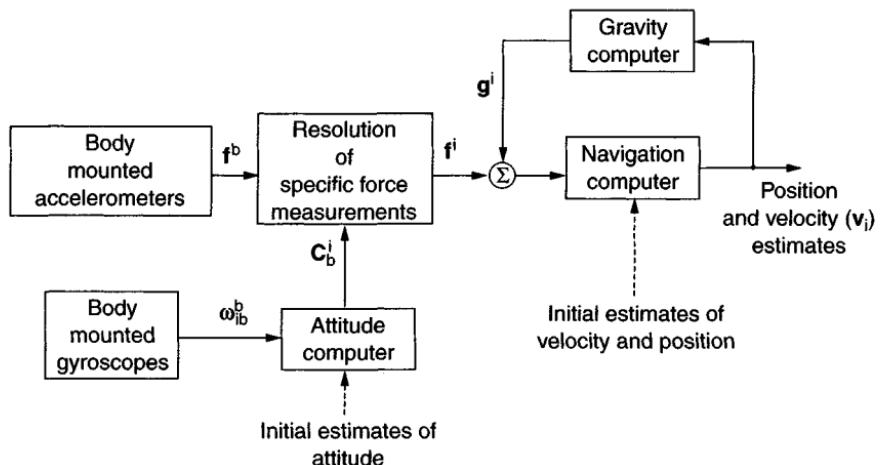


Figure 3.9 Strapdown inertial navigation system

eqn. (3.3). In systems of the type described later, in which it is required to derive estimates of vehicle velocity and position with respect to an Earth fixed frame, additional apparent forces will be acting which are functions of the reference frame motion. In this section, further forms of the navigation equation are derived, corresponding to different choices of reference frame [1].

The resulting system mechanisations are described together with their applications. As will become apparent, the variations in the mechanisations described here are in the strapdown computational algorithms and not in the arrangement of the sensors or the mechanical layout of the system.

3.5.1 Inertial frame mechanisation

In this system, it is required to calculate vehicle speed with respect to the Earth, the ground speed, in inertial axes, denoted by the symbol v_e^i . This may be accomplished by expressing the navigation equation (eqn. (3.3)) in inertial axes and deriving an expression for $\frac{d^2\mathbf{r}}{dt^2}|_e$ in terms of ground speed and its time derivatives with respect to the inertial frame.

Inertial velocity may be expressed in terms of ground speed using the Coriolis equation, viz.

$$\frac{d\mathbf{r}}{dt}\Big|_i = \frac{d\mathbf{r}}{dt}\Big|_e + \boldsymbol{\omega}_{ie} \times \mathbf{r} \quad (3.10)$$

Differentiating this expression and writing $\frac{d\mathbf{r}}{dt}|_e = \mathbf{v}_e$, we have,

$$\frac{d^2\mathbf{r}}{dt^2}\Big|_i = \frac{d\mathbf{v}_e}{dt}\Big|_i + \frac{d}{dt}[\boldsymbol{\omega}_{ie} \times \mathbf{r}]\Big|_i \quad (3.11)$$

Applying the Coriolis equation in the form of eqn. (3.10) to the second term in eqn. (3.11) gives:

$$\frac{d^2\mathbf{r}}{dt^2} \Big|_i = \frac{d\mathbf{v}_e}{dt} \Big|_i + \boldsymbol{\omega}_{ie} \times \mathbf{v}_e + \boldsymbol{\omega}_{ie} \times [\boldsymbol{\omega}_{ie} \times \mathbf{r}] \quad (3.12)$$

In generating the above equation, it is assumed that the turn rate of the Earth is constant, hence $\frac{d\boldsymbol{\omega}_{ie}}{dt} = 0$.

Combining eqns. (3.3) and (3.12) and rearranging yields:

$$\frac{d\mathbf{v}_e}{dt} \Big|_i = \mathbf{f} - \boldsymbol{\omega}_{ie} \times \mathbf{v}_e - \boldsymbol{\omega}_{ie} \times [\boldsymbol{\omega}_{ie} \times \mathbf{r}] + \mathbf{g} \quad (3.13)$$

In this equation, \mathbf{f} represents the specific force acceleration to which the navigation system is subjected, while $\boldsymbol{\omega}_{ie} \times \mathbf{v}_e$ is the acceleration caused by its velocity over the surface of a rotating Earth, usually referred to as the Coriolis acceleration. The term $\boldsymbol{\omega}_{ie} \times [\boldsymbol{\omega}_{ie} \times \mathbf{r}]$, in eqn. (3.13), defines the centripetal acceleration experienced by the system owing to the rotation of the Earth, and is not separately distinguishable from the gravitational acceleration which arises through mass attraction, \mathbf{g} . The sum of the accelerations caused by the mass attraction force and the centripetal force constitutes what is known as the local gravity vector, the vector to which a 'plumb bob' would align itself when held above the Earth (Figure 3.10). This is denoted here by the symbol \mathbf{g}_l , that is:

$$\mathbf{g}_l = \mathbf{g} - \boldsymbol{\omega}_{ie} \times [\boldsymbol{\omega}_{ie} \times \mathbf{r}] \quad (3.14)$$

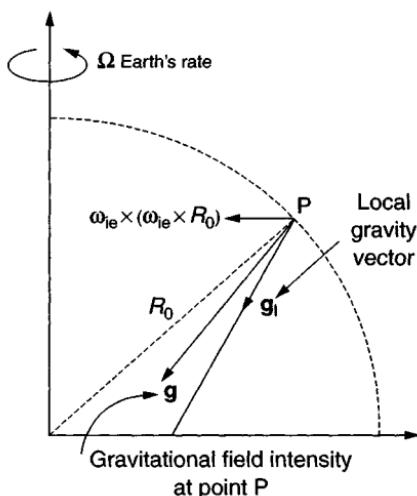


Figure 3.10 Diagram showing the components of the gravitational field

Combining eqns. (3.13) and (3.14) gives the following form of the navigation equation:

$$\frac{d\mathbf{v}_e}{dt} \Big|_i = \mathbf{f} - \boldsymbol{\omega}_{ie} \times \mathbf{v}_e + \mathbf{g}_l \quad (3.15)$$

This equation may be expressed in inertial axes, as follows, using the superscript notation mentioned earlier.

$$\dot{\mathbf{v}}_e^i = \mathbf{f}^i - \boldsymbol{\omega}_{ie}^i \times \mathbf{v}_e^i + \mathbf{g}_l^i \quad (3.16)$$

The measurements of specific force provided by the accelerometers are in body axes, as denoted by the vector quantity \mathbf{f}^b . In order to set up the navigation eqn. (3.16), the accelerometer outputs must be resolved into inertial axes to give \mathbf{f}^i . This may be achieved by pre-multiplying the measurement vector \mathbf{f}^b by the direction cosine matrix \mathbf{C}_b^i as described in Section 3.4.4 (eqn. (3.6)). Given knowledge of the attitude of the body at the start of navigation, the matrix \mathbf{C}_b^i is updated using eqns. (3.7) and (3.8) based on measurements of the body rates with respect to the i-frame which may be expressed as follows:

$$\boldsymbol{\omega}_{ib}^b = [p \quad q \quad r]^T \quad (3.17)$$

Substituting for \mathbf{f}^i from eqn. (3.6) in eqn. (3.16) gives the following form of the navigation equation:

$$\dot{\mathbf{v}}_e^i = \mathbf{C}_b^i \mathbf{f}^b - \boldsymbol{\omega}_{ie}^i \times \mathbf{v}_e^i + \mathbf{g}_l^i \quad (3.18)$$

The final term in this equation represents the local gravity vector expressed in the inertial frame.

A block diagram representation of the resulting inertial frame mechanisation is shown in Figure 3.11.

3.5.2 Earth frame mechanisation

In this system, ground speed is expressed in an Earth-fixed co-ordinate frame to give \mathbf{v}_e^e . It follows from the Coriolis equation, that the rate of change of \mathbf{v}_e , with respect to Earth axes, may be expressed in terms of its rate of change in inertial axes using:

$$\frac{d\mathbf{v}_e}{dt} \Big|_e = \frac{d\mathbf{v}_e}{dt} \Big|_i - \boldsymbol{\omega}_{ie} \times \mathbf{v}_e \quad (3.19)$$

Substituting for $\frac{d\mathbf{v}_e}{dt} \Big|_i$ from eqn. (3.15), we have:

$$\frac{d\mathbf{v}_e}{dt} \Big|_e = \mathbf{f} - 2\boldsymbol{\omega}_{ie} \times \mathbf{v}_e + \mathbf{g}_l \quad (3.20)$$

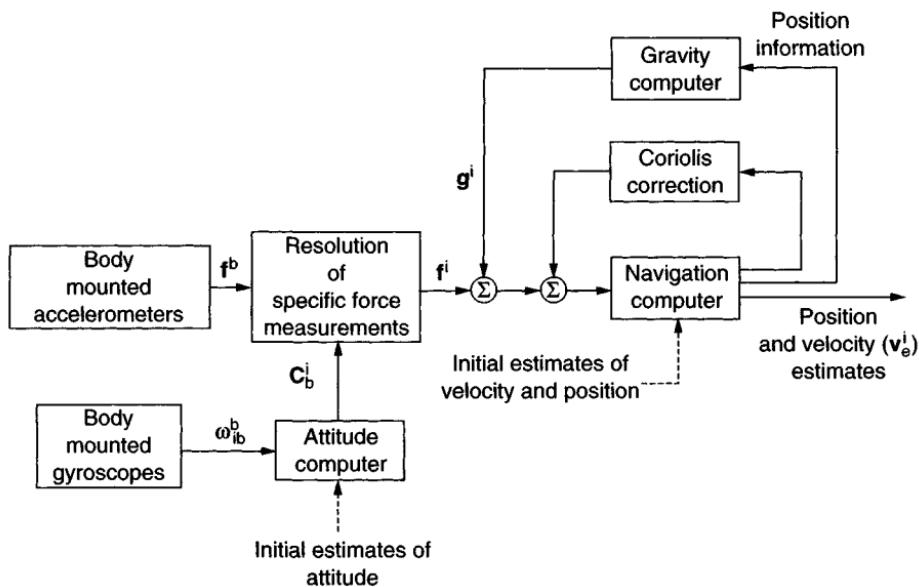


Figure 3.11 Strapdown inertial navigation system – inertial frame mechanisation

This may be expressed in Earth axes as follows:

$$\dot{v}_e^e = \mathbf{C}_b^e f^b - 2\omega_{ie}^e \times v_e^e + g_e^i \quad (3.21)$$

where \mathbf{C}_b^e is the direction cosine matrix used to transform the measured specific force vector into Earth axes. This matrix propagates in accordance with the following equation:

$$\dot{\mathbf{C}}_b^e = \mathbf{C}_b^e \Omega_{eb}^b \quad (3.22)$$

where Ω_{eb}^b is the skew symmetric form of ω_{eb}^b , the body rate with respect to the Earth-fixed frame. This is derived by differencing the measured body rates, ω_{ib}^b , and estimates of the components of Earth's rate, ω_{ie} , expressed in body axes as follows:

$$\omega_{eb}^b = \omega_{ib}^b - \mathbf{C}_e^b \omega_{ie}^e \quad (3.23)$$

in which $\mathbf{C}_e^b = \mathbf{C}_b^{eT}$, the transpose of the matrix \mathbf{C}_b^e .

A block diagram representation of the Earth frame mechanisation is shown in Figure 3.12.

A variation on this system may be used when it is required to navigate over relatively short distances, with respect to a fixed point on the Earth. A mechanisation of this type may be used for a tactical missile application in which navigation is required with respect to a ground based tracking station. In such a system,

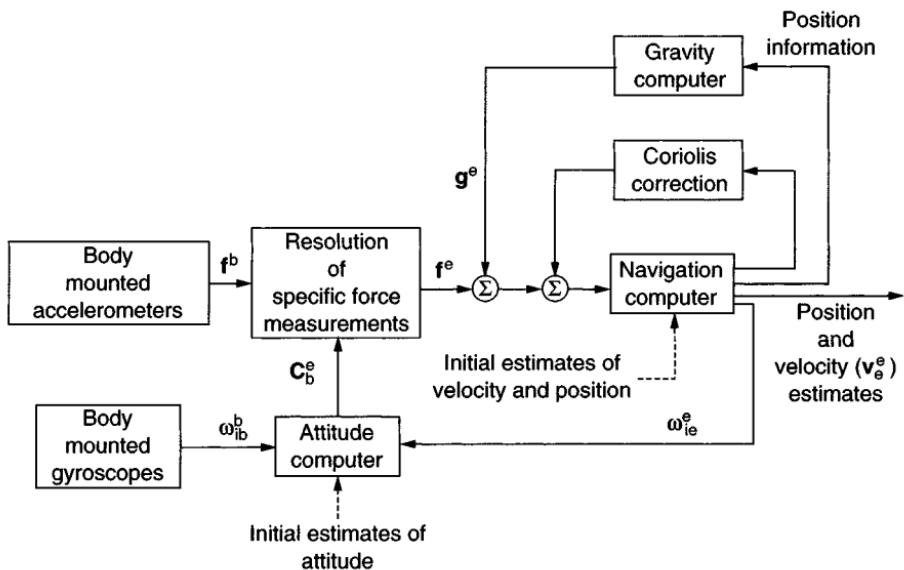


Figure 3.12 Strapdown inertial navigation system – Earth frame mechanisation

target tracking information provided by the ground station may need to be combined with the outputs of an on-board inertial navigation system to provide missile mid-course guidance commands. In order that the missile may operate in harmony with the ground systems, all information must be provided in a common frame of reference.

In this situation, an Earth-fixed reference frame may be defined, the origin of which is located at the tracking station, its axes aligned with the local vertical and a plane which is tangential to the Earth's surface, as illustrated in Figure 3.13.

For very short term navigation, as required for some tactical missile applications, further simplifications to this system mechanisation may be permitted. For instance, where the navigation period is short, typically 10 minutes or less, the effects of the rotation of the Earth on the attitude computation process can sometimes be ignored, and Coriolis corrections are no longer essential in the velocity equation to give sufficiently accurate navigation. In this situation, attitude is computed solely as a function of the turn rates measured by the gyroscopes, and eqn. (3.21) reduces to the following:

$$\dot{v}_c^e = C_b^e f^b + g^e \quad (3.24)$$

It is stressed, that such simplifications can only be allowed in cases where the navigation errors, induced by the omission of Earth rate and Coriolis terms, lie within the error bounds in which the navigation system is required to operate. This situation arises when the permitted gyroscopic errors are in excess of the rotation rate of the Earth, and allowable accelerometer biases are in excess of the acceleration errors introduced by ignoring the Coriolis forces.

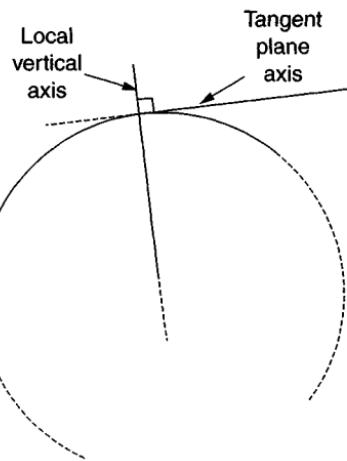


Figure 3.13 Tangent plane axis set

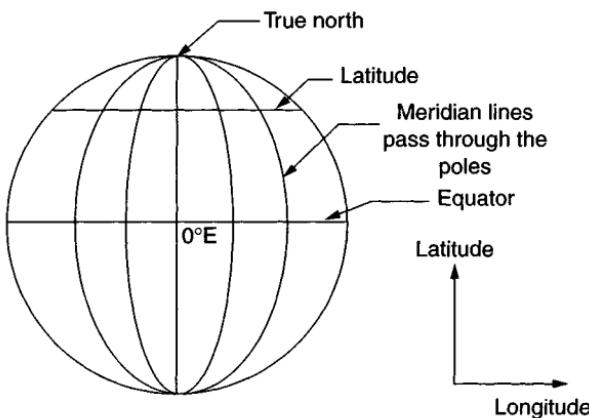


Figure 3.14 Geographic co-ordinate system

3.5.3 Local geographic navigation frame mechanisation

In order to navigate over large distances around the Earth, navigation information is most commonly required in the local geographic or navigation axis set described earlier. Position on the Earth may be specified in terms of latitude (degrees north or south of a datum) and longitude (degrees east or west of a datum). Figure 3.14 shows this geographic co-ordinate system on a globe. Lines of constant latitude and longitude are called parallels and meridians, respectively.

Navigation data are expressed in terms of north and east velocity components, latitude, longitude and height above the Earth. Whilst such information can be

computed using the position estimates provided by the inertial or Earth frame mechanisations described before, this involves a further transformation of the vector quantities v_e^i or v_e^e . Further, difficulties arise in representing the Earth's gravitational field precisely in a computer. For these reasons, the navigation frame mechanisation, described here, is often used when navigating around the Earth.

In this mechanisation, ground speed is expressed in navigation coordinates to give v_e^n . The rate of change of v_e^n with respect to navigation axes may be expressed in terms of its rate of change in inertial axes as follows:

$$\frac{dv_e}{dt} \Big|_n = \frac{dv_e}{dt} \Big|_i - [\omega_{ie} + \omega_{en}] \times v_e \quad (3.25)$$

Substituting for $\frac{dv_e}{dt} \Big|_i$, from eqn. (3.15), we have:

$$\frac{dv_e}{dt} \Big|_n = f - [2\omega_{ie} + \omega_{en}] \times v_e + g_i \quad (3.26)$$

This may be expressed in navigation axes as follows:

$$\dot{v}_e^n = C_b^n f^b - [2\omega_{ie}^n + \omega_{en}^n] \times v_e^n + g_i^n \quad (3.27)$$

where C_b^n is a direction cosine matrix used to transform the measured specific force vector into navigation axes. This matrix propagates in accordance with the following equation.

$$\dot{C}_b^n = C_b^n \Omega_{nb}^b \quad (3.28)$$

where Ω_{nb}^b is the skew symmetric form of ω_{nb}^b , the body rate with respect to the navigation frame. This is derived by differencing the measured body rates, ω_{ib}^b , and estimates of the components of navigation frame rate, ω_{in} . The latter term is obtained by summing the Earth's rate with respect to the inertial frame and the turn rate of the navigation frame with respect to the Earth, that is, $\omega_{in} = \omega_{ie} + \omega_{en}$. Therefore,

$$\omega_{nb}^b = \omega_{ib}^b - C_n^b [\omega_{ie}^n + \omega_{en}^n] \quad (3.29)$$

A block diagram representation of the navigation frame mechanisation is shown in Figure 3.15.

It is instructive to consider the physical significance of the various terms in the navigation equation (3.27). From this equation, it can be seen that the rate of change of the velocity, with respect to the surface of the Earth, is made up of the following terms:

1. The specific force acting on the vehicle, as measured by a triad of accelerometers mounted within it.
2. A correction for the acceleration caused by the vehicle's velocity over the surface of a rotating Earth, usually referred to as the Coriolis acceleration. The effect in two dimensions is illustrated in Figure 3.16. As the point P moves away

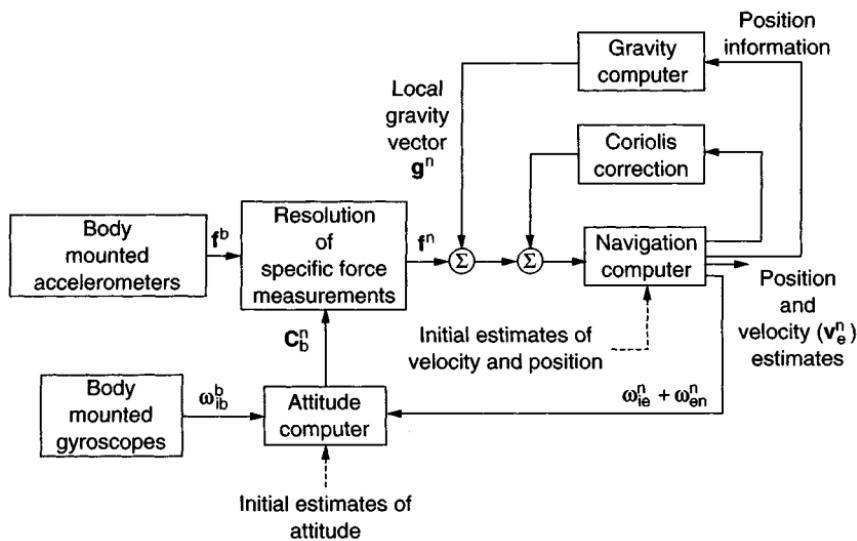


Figure 3.15 Strapdown inertial navigation system – local geographic navigation frame mechanisation

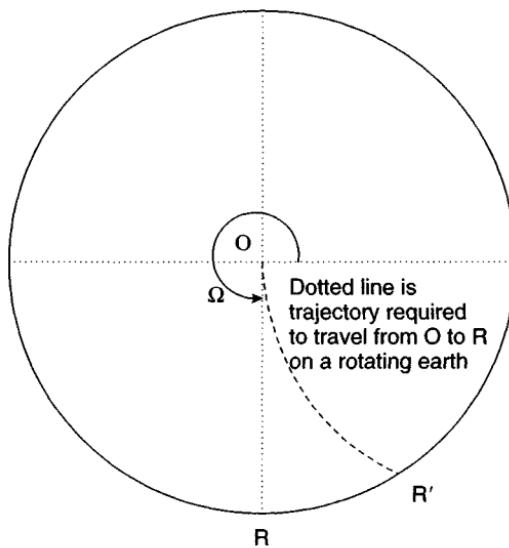


Figure 3.16 Illustration of the effect of Coriolis acceleration

from the axis of rotation, it traces out a curve in space as a result of the Earth's rotation.

3. A correction for the centripetal acceleration of the vehicle, resulting from its motion over the Earth's surface. For instance, a vehicle moving due east over

the surface of the Earth will trace out a circular path with respect to inertial axes. To follow this path, the vehicle is subject to a force acting towards the centre of the Earth of magnitude equal to the product of its mass, its linear velocity and its turn rate with respect to the Earth.

- Compensation for the apparent gravitational force acting on the vehicle. This includes the gravitational force caused by the mass attraction of the Earth, and the centripetal acceleration of the vehicle resulting from the rotation of the Earth. The latter term arises even if the vehicle is stationary with respect to the Earth, since the path which it follows in space is circular.

A simple example serves to illustrate the importance of the Coriolis effect. Consider a vehicle launched from the north pole with the intention of flying to New York city. The vehicle is assumed to travel at an average speed of 3600 miles/h. During the flight, of approximately 1 h, the Earth will have rotated by about 15° , a distance of approximately 900 miles at the latitude of New York. Consequently, if no Coriolis correction was made to the on-board inertial guidance system during the course of the flight, the vehicle would arrive in the Chicago area rather than New York as originally intended.

3.5.4 Wander azimuth navigation frame mechanisation

In the local geographic navigation frame mechanisation described in the previous section, the n-frame is required to rotate continuously as the system moves over the surface of the Earth in order to keep its x -axis parallel to true north. In order to achieve this condition worldwide, the n-frame must rotate at much greater rates about its z -axis as the navigation system moves over the surface of the Earth in the polar regions, compared to the rates required at lower latitudes. This effect is illustrated in Figure 3.17 which shows a polar view of a near polar crossing. It should be clear from the diagram that the rate at which the local geographic navigation frame must rotate about its z -axis in order to maintain the x -axis pointing at the pole becomes very large, the heading direction slewing rapidly through 180° when moving past the pole. In the most extreme case, a direct crossing of the pole, the turn rate becomes infinite when passing over the pole.

The effect is illustrated mathematically as follows. The turn rate of the navigation frame, the transport rate, may be expressed in component form as:

$$\omega_{en}^n = \begin{bmatrix} v_E & -v_N & -v_E \tan L \\ R_0 + h & R_0 + h & R_0 + h \end{bmatrix}^T \quad (3.30)$$

where v_N is the north velocity, v_E the east velocity, R_0 the radius of the Earth, L the latitude and h the height above ground.

It will be seen that the third component of the transport rate becomes indeterminate at the geographic poles.

One way of avoiding the singularity, and so providing a navigation system with world-wide capability, is to adopt a wander azimuth mechanisation in which the z -component of ω_{en}^n is set to zero. A wander axis system is a locally level frame

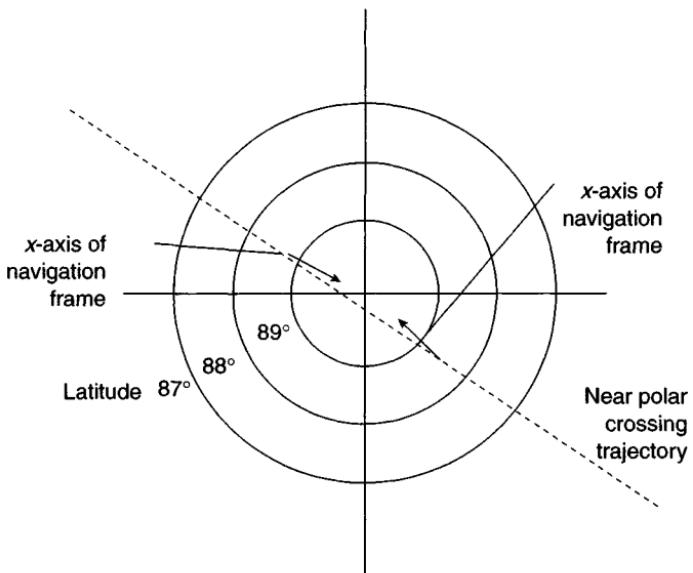


Figure 3.17 Geographic reference singularity at pole crossings

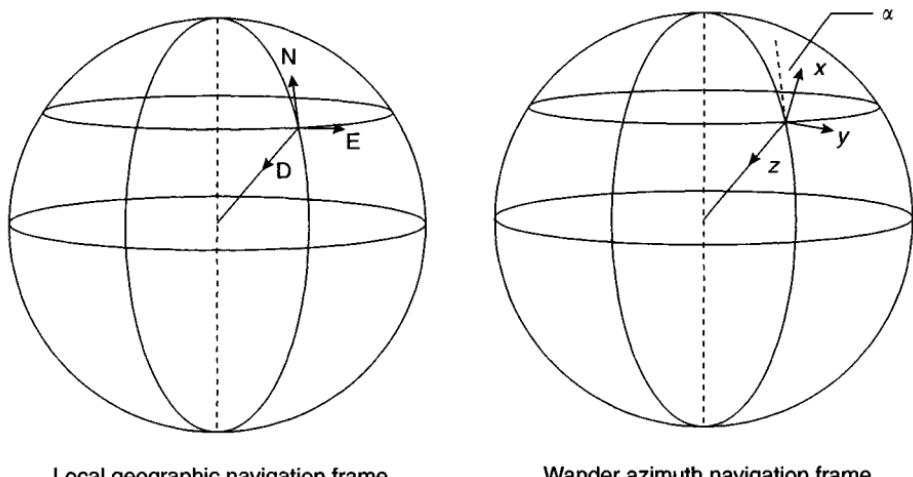


Figure 3.18 Illustration of wander azimuth frame

which moves over the Earth's surface with the vehicle, as depicted in Figure 3.18. However, as the name implies, the azimuth angle between true north and the x -axis of the wander axis frame varies with vehicle position on the Earth. This variation is chosen in order to avoid discontinuities in the orientation of the wander

frame with respect to the Earth as the vehicle passes over either the north or south poles.

A navigation equation for a wander azimuth system, which is similar in form to eqn. (3.27), may be constructed as follows:

$$\dot{\mathbf{v}}_e^w = \mathbf{C}_b^w \mathbf{f}^b - [2\mathbf{C}_e^w \boldsymbol{\omega}_e^e + \boldsymbol{\omega}_{ew}^w] \times \mathbf{v}_e^w + \mathbf{g}_i^w \quad (3.31)$$

This equation is integrated to generate estimates of vehicle ground speed in the wander azimuth frame, \mathbf{v}_e^w . This is then used to generate the turn rate of the wander frame with respect to the Earth, $\boldsymbol{\omega}_{ew}^w$. The direction cosine matrix which relates the wander frame to the Earth frame, \mathbf{C}_e^w , may be updated using the equation

$$\dot{\mathbf{C}}_e^w = \mathbf{C}_e^w \boldsymbol{\Omega}_{ew}^w \quad (3.32)$$

where $\boldsymbol{\Omega}_{ew}^w$ is a skew symmetric matrix formed from the elements of the angular rate vector $\boldsymbol{\omega}_{ew}^w$. This process is implemented iteratively and enables any singularities to be avoided. Further details concerning wander azimuth systems and the mechanisations described earlier appear in Reference 1.

3.5.5 Summary of strapdown system mechanisations

This section has provided outline descriptions of a number of possible strapdown inertial navigation system mechanisations. Further details are given in Reference 1. The choice of mechanisation is dependent on the application. Whilst any of the schemes described may be used for navigation close to the Earth, the local geographic navigation frame mechanisation is commonly employed for navigation over large distances. The wander azimuth system provides a world-wide navigation capability. These mechanisations provide navigation data in terms of north and east velocity, latitude and longitude and allow a relatively simple gravity model to be used. For navigation over shorter distances, an Earth fixed reference system may be applicable.

3.6 Strapdown attitude representations

3.6.1 Introductory remarks

Consider now ways in which a set of strapdown gyroscopic sensors may be used to instrument a reference co-ordinate frame within a vehicle which is free to rotate about any direction. The attitude of the vehicle with respect to the designated reference frame may be stored as a set of numbers in a computer within the vehicle. The stored attitude is updated as the vehicle rotates using the measurements of turn rate provided by the gyroscopes.

The co-ordinate frames referred to during the course of the discussion which follows are orthogonal, right-handed axis sets in which positive rotations about each axis are taken to be in a clockwise direction looking along the axis from the origin, as indicated in the Figure 3.19. A negative rotation acts in an opposite sense, that is, in an anti-clockwise direction. This convention is used throughout this book.

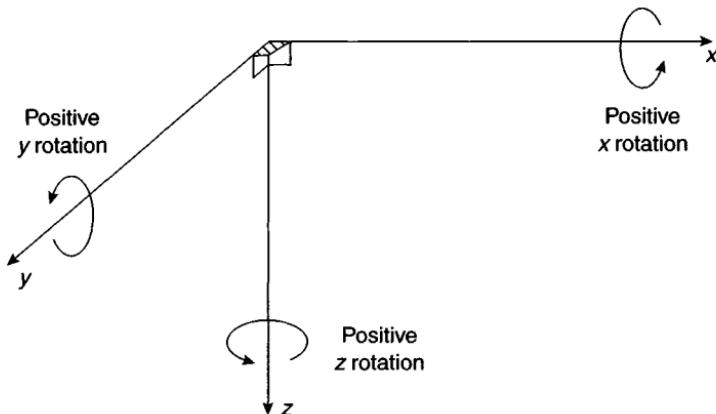


Figure 3.19 Definition of axis rotations

It is important to remember that the change in attitude of a body, which is subjected to a series of rotations about different axes, is not only a function of the angles through which it rotates about each of those axes, but the order in which the rotations occur. The illustration given in Figure 3.20, although somewhat extreme, shows quite clearly that the order in which a sequence of rotations occurs is most important.

Rotations are defined here with respect to the orthogonal right-handed axis set, Oxyz, indicated in the figure. The sequence of rotations shown in the left half of the figure is made up of a 90° pitch, or y-axis rotation, followed by a 90° yaw, or z-axis rotation, and a further pitch rotation of -90° . On completion of this sequence of turns, it can be seen that a net rotation of 90° about the roll (x) axis has taken place. In the right hand figure, the order of the rotations has been reversed. Although the body still ends up with its roll axis aligned in the original direction, it is seen that a net roll rotation of -90° has now taken place. Hence, individual axis rotations are said to be non-commutative. It is clear that failure to take account of the order in which rotations arise can lead to a substantial error in the computed attitude.

Various mathematical representations can be used to define the attitude of a body with respect to a co-ordinate reference frame. The parameters associated with each method may be stored within a computer and updated as the vehicle rotates using the measurements of turn rate provided by the strapdown gyroscopes. Three attitude representations are described here, namely:

1. *Direction cosines.* The direction cosine matrix, introduced in Section 3.5, is a 3×3 matrix, the columns of which represent unit vectors in body axes projected along the reference axes.
2. *Euler angles.* A transformation from one co-ordinate frame to another is defined by three successive rotations about different axes taken in turn. The Euler angle representation is perhaps one of the simplest techniques in terms of physical

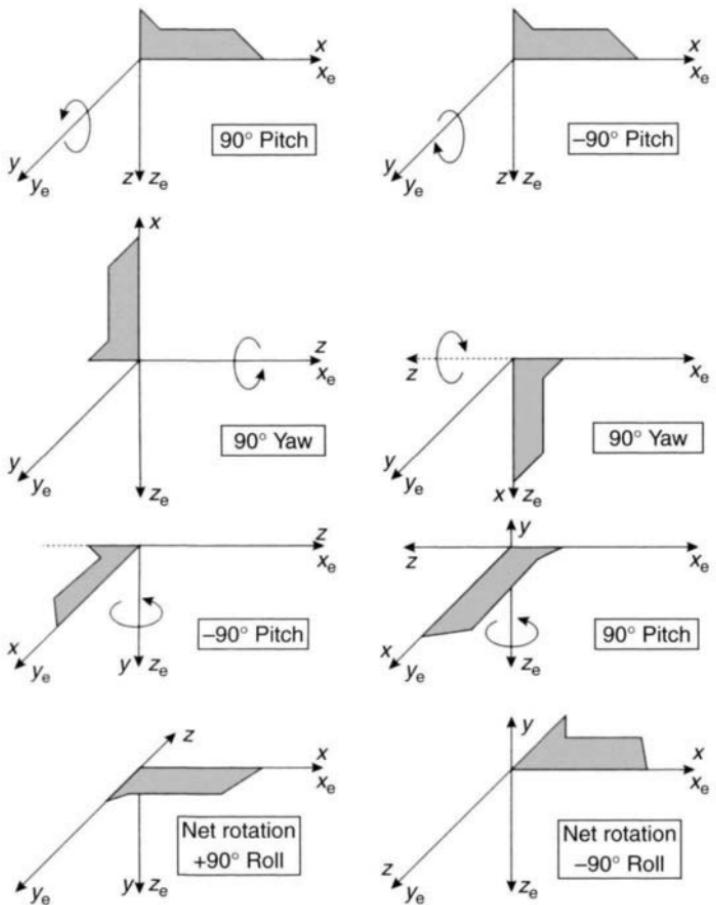


Figure 3.20 Illustration of effect of order of body rotations

appreciation. The three angles correspond to the angles which would be measured between a set of mechanical gimbals,³ which is supporting a stable element, where the axes of the stable element represent the reference frame, and with the body being attached via a bearing to the outer gimbal.

3. *Quaternions.* The quaternion attitude representation allows a transformation from one co-ordinate frame to another to be effected by a single rotation about a vector defined in the reference frame. The quaternion is a four-element vector representation, the elements of which are functions of the orientation of this vector and the magnitude of the rotation.

³ A gimbal is a rigid mechanical frame which is free to rotate about a single-axis to isolate it from angular motion in that direction. A stable platform can be isolated from body motion if supported by three such frames with their axes of rotation nominally orthogonal to each other.

In the following sections, each of these attitude representations is described in detail.

3.6.2 Direction cosine matrix

3.6.2.1 Introduction

The direction cosine matrix, denoted here by the symbol \mathbf{C}_b^n , is a 3×3 matrix, the columns of which represent unit vectors in body axes projected along the reference axes. \mathbf{C}_b^n is written here in component form as follows:

$$\mathbf{C}_b^n = \begin{bmatrix} c_{11} & c_{12} & c_{13} \\ c_{21} & c_{22} & c_{23} \\ c_{31} & c_{32} & c_{33} \end{bmatrix} \quad (3.33)$$

The element in the i th row and the j th column represents the cosine of the angle between the i -axis of the reference frame and the j -axis of the body frame.

3.6.2.2 Use of direction cosine matrix for vector transformation

A vector quantity defined in body axes, \mathbf{r}^b , may be expressed in reference axes by pre-multiplying the vector by the direction cosine matrix as follows:

$$\mathbf{r}^n = \mathbf{C}_b^n \mathbf{r}^b \quad (3.34)$$

3.6.2.3 Propagation of direction cosine matrix with time

The rate of change of \mathbf{C}_b^n with time is given by:

$$\dot{\mathbf{C}}_b^n = \lim_{\delta t \rightarrow 0} \frac{\delta \mathbf{C}_b^n}{\delta t} = \lim_{\delta t \rightarrow 0} \frac{\mathbf{C}_b^n(t + \delta t) - \mathbf{C}_b^n(t)}{\delta t} \quad (3.35)$$

where $\mathbf{C}_b^n(t)$ and $\mathbf{C}_b^n(t + \delta t)$ represent the direction cosine matrix at times t and $t + \delta t$, respectively. $\mathbf{C}_b^n(t + \delta t)$ can be written as the product of two matrices as follows:

$$\mathbf{C}_b^n(t + \delta t) = \mathbf{C}_b^n(t) \mathbf{A}(t) \quad (3.36)$$

where $\mathbf{A}(t)$ is a direction cosine matrix which relates the b-frame at time t to the b-frame at time $t + \delta t$. For small angle rotations, $\mathbf{A}(t)$ may be written as follows:

$$\mathbf{A}(t) = [\mathbf{I} + \delta\Psi] \quad (3.37)$$

where \mathbf{I} is a 3×3 identity matrix and

$$\delta\Psi = \begin{bmatrix} 0 & -\delta\psi & \delta\theta \\ \delta\psi & 0 & -\delta\phi \\ -\delta\theta & \delta\phi & 0 \end{bmatrix} \quad (3.38)$$

in which $\delta\psi$, $\delta\theta$ and $\delta\phi$ are the small rotation angles through which the b-frame has rotated over the time interval δt about its yaw, pitch and roll axes, respectively.

In the limit as δt approaches zero, small angle approximations are valid and the order of the rotations becomes unimportant.

Substituting for $\mathbf{C}_b^n(t + \delta t)$ in eqn. (3.35) we obtain:

$$\dot{\mathbf{C}}_b^n = \mathbf{C}_b^n \lim_{\delta t \rightarrow 0} \frac{\delta \Psi}{\delta t} \quad (3.39)$$

In the limit as $\delta t \rightarrow 0$, $\delta \Psi / \delta t$ is the skew symmetric form of the angular rate vector $\boldsymbol{\omega}_{nb}^b = [\omega_x \ \omega_y \ \omega_z]^T$, which represents the turn rate of the b-frame with respect to the n-frame expressed in body axes, that is,

$$\lim_{\delta t \rightarrow 0} \frac{\delta \Psi}{\delta t} = \boldsymbol{\Omega}_{nb}^b \quad (3.40)$$

Substituting in eqn. (3.39) gives:

$$\dot{\mathbf{C}}_b^n = \mathbf{C}_b^n \boldsymbol{\Omega}_{nb}^b \quad (3.41)$$

where

$$\boldsymbol{\Omega}_{nb}^b = \begin{bmatrix} 0 & -\omega_z & \omega_y \\ \omega_z & 0 & -\omega_x \\ -\omega_y & \omega_x & 0 \end{bmatrix} \quad (3.42)$$

An equation of the form of eqn. (3.41) may be solved within a computer in a strapdown inertial navigation system to keep track of body attitude with respect to the chosen reference frame. It may be expressed in component form as follows:

$$\begin{aligned} \dot{c}_{11} &= c_{12}\omega_z - c_{13}\omega_y & \dot{c}_{12} &= c_{13}\omega_x - c_{11}\omega_z & \dot{c}_{13} &= c_{11}\omega_y - c_{12}\omega_x \\ \dot{c}_{21} &= c_{22}\omega_z - c_{23}\omega_y & \dot{c}_{22} &= c_{23}\omega_x - c_{21}\omega_z & \dot{c}_{23} &= c_{21}\omega_y - c_{22}\omega_x \\ \dot{c}_{31} &= c_{32}\omega_z - c_{33}\omega_y & \dot{c}_{32} &= c_{33}\omega_x - c_{31}\omega_z & \dot{c}_{33} &= c_{31}\omega_y - c_{32}\omega_x \end{aligned} \quad (3.43)$$

3.6.3 Euler angles

3.6.3.1 Introduction

A transformation from one co-ordinate frame to another can be carried out as three successive rotations about different axes. For instance, a transformation from reference axes to a new co-ordinate frame may be expressed as follows:

- rotate through angle ψ about reference z -axis
- rotate through angle θ about new y -axis
- rotate through angle ϕ about new x -axis

where ψ , θ and ϕ are referred to as the Euler rotation angles. This type of representation is popular because of the physical significance of the Euler angles which correspond to the angles which would be measured by angular pick-offs between a set of three gimbals in a stable platform inertial navigation system.

3.6.3.2 Use of Euler angles for vector transformation

The three rotations may be expressed mathematically as three separate direction cosine matrices as defined below:

$$\text{rotation } \psi \text{ about } z\text{-axis, } \mathbf{C}_1 = \begin{bmatrix} \cos \psi & \sin \psi & 0 \\ -\sin \psi & \cos \psi & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad (3.44)$$

$$\text{rotation } \theta \text{ about } y\text{-axis, } \mathbf{C}_2 = \begin{bmatrix} \cos \theta & 0 & -\sin \theta \\ 0 & 1 & 0 \\ \sin \theta & 0 & \cos \theta \end{bmatrix} \quad (3.45)$$

$$\text{rotation } \phi \text{ about } x\text{-axis, } \mathbf{C}_3 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos \phi & \sin \phi \\ 0 & -\sin \phi & \cos \phi \end{bmatrix} \quad (3.46)$$

Thus, a transformation from reference to body axes may be expressed as the product of these three separate transformations as follows:

$$\mathbf{C}_n^b = \mathbf{C}_3 \mathbf{C}_2 \mathbf{C}_1 \quad (3.47)$$

Similarly, the inverse transformation from body to reference axes is given by:

$$\mathbf{C}_b^n = \mathbf{C}_b^{bT} = \mathbf{C}_1^T \mathbf{C}_2^T \mathbf{C}_3^T \quad (3.48)$$

$$\begin{aligned} \mathbf{C}_b^n &= \begin{bmatrix} \cos \psi & -\sin \psi & 0 \\ \sin \psi & \cos \psi & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \cos \theta & 0 & \sin \theta \\ 0 & 1 & 0 \\ -\sin \theta & 0 & \cos \theta \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos \phi & -\sin \phi \\ 0 & \sin \phi & \cos \phi \end{bmatrix} \\ &= \begin{bmatrix} \cos \theta \cos \psi & -\cos \phi \sin \psi & \sin \phi \sin \psi \\ \cos \theta \sin \psi & \cos \phi \cos \psi & -\sin \phi \cos \psi \\ -\sin \theta & \sin \phi \cos \theta & \cos \phi \cos \theta \end{bmatrix} \quad (3.49) \end{aligned}$$

This is the direction cosine matrix given by eqn. (3.33) expressed in terms of Euler angles.

For small angle rotations, $\sin \phi \rightarrow \phi$, $\sin \theta \rightarrow \theta$, $\sin \psi \rightarrow \psi$ and the cosines of these angles approach unity. Making these substitutions in eqn. (3.49) and ignoring products of angles which also become small, the direction cosine matrix expressed in terms of the Euler rotations reduces approximately to the skew symmetric form shown below:

$$\mathbf{C}_b^n \approx \begin{bmatrix} 1 & -\psi & \theta \\ \psi & 1 & -\phi \\ -\theta & \phi & 1 \end{bmatrix} \quad (3.50)$$

This form of matrix is used in Chapter 11 to represent the small change in attitude which occurs between successive updates in the real time computation of body attitude, and in Chapters 10 and 12 to represent the error in the estimated direction cosine matrix.

3.6.3.3 Propagation of Euler angles with time

Following the gimbal analogy mentioned earlier, ϕ , θ and ψ are the gimbal angles and $\dot{\phi}$, $\dot{\theta}$ and $\dot{\psi}$ are the gimbal rates. The gimbal rates are related to the body rates, ω_x , ω_y and ω_z as follows:

$$\begin{bmatrix} \omega_x \\ \omega_y \\ \omega_z \end{bmatrix} = \begin{bmatrix} \dot{\phi} \\ 0 \\ 0 \end{bmatrix} + \mathbf{C}_3 \begin{bmatrix} 0 \\ \dot{\theta} \\ 0 \end{bmatrix} + \mathbf{C}_3 \mathbf{C}_2 \begin{bmatrix} 0 \\ 0 \\ \dot{\psi} \end{bmatrix} \quad (3.51)$$

This equation can be rearranged and expressed in component form as follows:

$$\begin{aligned} \dot{\phi} &= (\omega_y \sin \phi + \omega_z \cos \phi) \tan \theta + \omega_x \\ \dot{\theta} &= \omega_y \cos \phi - \omega_z \sin \phi \\ \dot{\psi} &= (\omega_y \sin \phi + \omega_z \cos \phi) \sec \theta \end{aligned} \quad (3.52)$$

Equations of this form may be solved in a strapdown system to update the Euler rotations of the body with respect to the chosen reference frame. However, their use is limited since the solution of the $\dot{\phi}$ and $\dot{\psi}$ equations become indeterminate when $\theta = \pm 90^\circ$.

3.6.4 Quaternions

3.6.4.1 Introduction

The quaternion attitude representation is a four-parameter representation based on the idea that a transformation from one co-ordinate frame to another may be effected by a single rotation about a vector μ defined with respect to the reference frame. The quaternion, denoted here by the symbol \mathbf{q} , is a four element vector, the elements of which are functions of this vector and the magnitude of the rotation:

$$\mathbf{q} = \begin{bmatrix} a \\ b \\ c \\ d \end{bmatrix} = \begin{bmatrix} \cos(\mu/2) \\ (\mu_x/\mu) \sin(\mu/2) \\ (\mu_y/\mu) \sin(\mu/2) \\ (\mu_z/\mu) \sin(\mu/2) \end{bmatrix} \quad (3.53)$$

where μ_x , μ_y , μ_z are the components of the angle vector μ and μ the magnitude of μ .

The magnitude and direction of μ are defined in order that the reference frame may be rotated into coincidence with the body frame by rotating about μ through an angle μ .

A quaternion with components a , b , c and d may also be expressed as a four-parameter complex number with a real component a , and three imaginary components,

b, c and d , as follows:

$$\mathbf{q} = a + \mathbf{i}b + \mathbf{j}c + \mathbf{k}d \quad (3.54)$$

This is an extension of the more usual two parameter complex number form with one real component and one imaginary component, $x = a + \mathbf{i}b$, with which the reader is more likely to be familiar.

The product of two quaternions, $\mathbf{q} = a + \mathbf{i}b + \mathbf{j}c + \mathbf{k}d$ and $\mathbf{p} = e + \mathbf{i}f + \mathbf{j}g + \mathbf{k}h$ may then be derived as shown below applying the usual rules for products of complex numbers, viz:

$$\mathbf{i} \cdot \mathbf{i} = -1 \quad \mathbf{i} \cdot \mathbf{j} = \mathbf{k} \quad \mathbf{j} \cdot \mathbf{i} = -\mathbf{k} \quad \dots \text{etc.}$$

Hence,

$$\begin{aligned} \mathbf{q} \cdot \mathbf{p} &= (a + \mathbf{i}b + \mathbf{j}c + \mathbf{k}d)(e + \mathbf{i}f + \mathbf{j}g + \mathbf{k}h) \\ &= ea - bf - cg - dh + (af + be + ch - dg)\mathbf{i} \\ &\quad + (ag + ce - bh + df)\mathbf{j} + (ah + de + bg - cf)\mathbf{k} \end{aligned} \quad (3.55)$$

Alternatively, the quaternion product may be expressed in matrix form as:

$$\mathbf{q} \cdot \mathbf{p} = \begin{bmatrix} a & -b & -c & -d \\ b & a & -d & c \\ c & d & a & -b \\ d & -c & b & a \end{bmatrix} \begin{bmatrix} e \\ f \\ g \\ h \end{bmatrix} \quad (3.56)$$

3.6.4.2 Use of quaternion for vector transformation

A vector quantity defined in body axes, \mathbf{r}^b , may be expressed in reference axes as \mathbf{r}^n using the quaternion directly. First define a quaternion, $\mathbf{r}^{b'}$, in which the complex components are set equal to the components of \mathbf{r}^b , and with a zero scalar component, that is, if:

$$\mathbf{r}^b = \mathbf{i}x + \mathbf{j}y + \mathbf{k}z$$

$$\mathbf{r}^{b'} = 0 + \mathbf{i}x + \mathbf{j}y + \mathbf{k}z$$

This is expressed in reference axes as \mathbf{r}^n' using:

$$\mathbf{r}^n' = \mathbf{q}\mathbf{r}^{b'}\mathbf{q}^* \quad (3.57)$$

where $\mathbf{q}^* = (a - \mathbf{i}b - \mathbf{j}c - \mathbf{k}d)$, the complex conjugate of \mathbf{q} .

Hence,

$$\begin{aligned} \mathbf{r}^n' &= (a + \mathbf{i}b + \mathbf{j}c + \mathbf{k}d)(0 + \mathbf{i}x + \mathbf{j}y + \mathbf{k}z)(a - \mathbf{i}b - \mathbf{j}c - \mathbf{k}d) \\ &= 0 + \{(a^2 + b^2 - c^2 - d^2)x + 2(bc - ad)y + 2(bd + ac)z\}\mathbf{i} \\ &\quad + \{2(bc + ad)x + (a^2 - b^2 + c^2 - d^2)y + 2(cd - ab)z\}\mathbf{j} \\ &\quad + \{2(bd - ac)x + 2(cd + ab)y + (a^2 - b^2 - c^2 + d^2)z\}\mathbf{k} \end{aligned} \quad (3.58)$$

Alternatively, \mathbf{r}^n' may be expressed in matrix form as follows:

$$\mathbf{r}^{n'} = \mathbf{C}' \mathbf{r}^b'$$

where

$$\mathbf{C}' = \begin{bmatrix} 0 & 0 \\ 0 & \mathbf{C} \end{bmatrix} \quad \mathbf{r}^b' = \begin{bmatrix} 0 \\ \mathbf{r}^b \end{bmatrix}$$

and

$$\mathbf{C} = \begin{bmatrix} (a^2 + b^2 - c^2 - d^2) & 2(bc - ad) & 2(bd + ac) \\ 2(bc + ad) & (a^2 - b^2 + c^2 - d^2) & 2(cd - ab) \\ 2(bd - ac) & 2(cd + ab) & (a^2 - b^2 - c^2 + d^2) \end{bmatrix} \quad (3.59)$$

which is equivalent to writing:

$$\mathbf{r}^n = \mathbf{C} \mathbf{r}^b$$

Comparison with eqn. (3.34) reveals that \mathbf{C} is equivalent to the direction cosine matrix \mathbf{C}_b^n .

3.6.4.3 Propagation of quaternion with time

The quaternion, \mathbf{q} , propagates in accordance with the following equation:

$$\dot{\mathbf{q}} = 0.5\mathbf{q} \cdot \mathbf{p}_{nb}^b \quad (3.60)$$

This equation may be expressed in matrix form as a function of the components of \mathbf{q} and $\mathbf{p}_{nb}^b = [0, \omega_{nb}^b]^T$ as follows:

$$\mathbf{q} = \begin{bmatrix} \dot{a} \\ \dot{b} \\ \dot{c} \\ \dot{d} \end{bmatrix} = 0.5 \begin{bmatrix} a & -b & -c & -d \\ b & a & -d & c \\ c & d & a & -b \\ d & -c & b & a \end{bmatrix} \begin{bmatrix} 0 \\ \omega_x \\ \omega_y \\ \omega_z \end{bmatrix} \quad (3.61)$$

that is,

$$\begin{aligned} \dot{a} &= -0.5(b\omega_x + c\omega_y + d\omega_z) \\ \dot{b} &= 0.5(a\omega_x - d\omega_y + c\omega_z) \\ \dot{c} &= 0.5(d\omega_x + a\omega_y - b\omega_z) \\ \dot{d} &= -0.5(c\omega_x - b\omega_y - a\omega_z) \end{aligned} \quad (3.62)$$

Equations of this form may be solved in a strapdown navigation system to keep track of the quaternion parameters which define body orientation. The quaternion parameters may then be used to compute an equivalent direction cosine matrix, or used directly to transform the measured specific force vector into the chosen reference frame (see eqn. (3.57)).

3.6.5 Relationships between direction cosines, Euler angles and quaternions

As shown in the preceding sections, the direction cosines may be expressed in terms of Euler angles or quaternions, viz:

$$\begin{aligned}
 \mathbf{C}_b^n &= \begin{bmatrix} c_{11} & c_{12} & c_{13} \\ c_{21} & c_{22} & c_{23} \\ c_{31} & c_{32} & c_{33} \end{bmatrix} \\
 &= \begin{bmatrix} \cos \theta \cos \psi & -\cos \phi \sin \psi & \sin \phi \sin \psi \\ \cos \theta \sin \psi & +\sin \phi \sin \theta \cos \psi & +\cos \phi \sin \theta \cos \psi \\ -\sin \theta & \sin \phi \cos \theta & \cos \phi \cos \theta \end{bmatrix} \\
 &= \begin{bmatrix} (a^2 + b^2 - c^2 - d^2) & 2(bc - ad) & 2(bd + ac) \\ 2(bc + ad) & (a^2 - b^2 + c^2 - d^2) & 2(cd - ab) \\ 2(bd - ac) & 2(cd + ab) & (a^2 - b^2 - c^2 + d^2) \end{bmatrix} \tag{3.63}
 \end{aligned}$$

By comparing the elements of the above equations, the quaternion elements may be expressed directly in terms of Euler angles or direction cosines. Similarly, the Euler angles may be written in terms of direction cosines or quaternions. Some of these relationships are summarised in the following sections.

3.6.5.1 Quaternions expressed in terms of direction cosines

For small angular displacements, the quaternion parameters may be derived using the following relationships:

$$\begin{aligned}
 a &= \frac{1}{2}(1 + c_{11} + c_{22} + c_{33})^{1/2} \\
 b &= \frac{1}{4a}(c_{32} - c_{23}) \\
 c &= \frac{1}{4a}(c_{13} - c_{31}) \\
 d &= \frac{1}{4a}(c_{21} - c_{12})
 \end{aligned} \tag{3.64}$$

A more comprehensive algorithm for the extraction of quaternion parameters from the direction cosines, which takes account of the relative magnitudes of the direction cosine elements, is described by Shepperd [2].

3.6.5.2 Quaternions expressed in terms of Euler angles

$$\begin{aligned} a &= \cos \frac{\phi}{2} \cos \frac{\theta}{2} \cos \frac{\psi}{2} + \sin \frac{\phi}{2} \sin \frac{\theta}{2} \sin \frac{\psi}{2} \\ b &= \sin \frac{\phi}{2} \cos \frac{\theta}{2} \cos \frac{\psi}{2} - \cos \frac{\phi}{2} \sin \frac{\theta}{2} \sin \frac{\psi}{2} \\ c &= \cos \frac{\phi}{2} \sin \frac{\theta}{2} \cos \frac{\psi}{2} + \sin \frac{\phi}{2} \cos \frac{\theta}{2} \sin \frac{\psi}{2} \\ d &= \cos \frac{\phi}{2} \cos \frac{\theta}{2} \sin \frac{\psi}{2} + \sin \frac{\phi}{2} \sin \frac{\theta}{2} \cos \frac{\psi}{2} \end{aligned} \quad (3.65)$$

3.6.5.3 Euler angles expressed in terms of direction cosines

The Euler angles may be derived directly from the direction cosines as described below. For conditions where θ is not equal to 90° the Euler angles can be determined using

$$\begin{aligned} \phi &= \arctan \left[\frac{c_{32}}{c_{33}} \right] \\ \theta &= \arcsin [-c_{31}] \\ \psi &= \arctan \left[\frac{c_{21}}{c_{11}} \right] \end{aligned} \quad (3.66)$$

For situations in which θ approaches $\pi/2$ radians, the equations in ϕ and ψ become indeterminate because the numerator and the denominator approach zero simultaneously. Under such conditions, alternative solutions for ϕ and ψ are sought based upon other elements of the direction cosine matrix. This difficulty may be overcome by using the direction cosine elements c_{12} , c_{13} , c_{22} and c_{23} , which do not appear in eqn. (3.66), to derive the following relationships:

$$\begin{aligned} c_{23} + c_{12} &= (\sin \theta - 1) \sin(\psi + \phi) \\ c_{13} - c_{22} &= (\sin \theta - 1) \cos(\psi + \phi) \\ c_{23} - c_{12} &= (\sin \theta + 1) \sin(\psi - \phi) \\ c_{13} + c_{22} &= (\sin \theta + 1) \cos(\psi - \phi) \end{aligned} \quad (3.67)$$

For θ near $+\pi/2$:

$$\psi - \phi = \arctan \left[\frac{c_{23} - c_{12}}{c_{13} + c_{22}} \right]$$

For θ near $-\pi/2$:

$$\psi + \phi = \arctan \left[\frac{c_{23} + c_{12}}{c_{13} - c_{22}} \right] \quad (3.68)$$

Equations (3.67) and (3.68) provide values for the sum and difference of ϕ and ψ under conditions where θ approaches $\pi/2$. Separate solutions for ϕ and ψ cannot be

obtained when $\theta = +\pi/2$ because both become measures of angle about parallel axes (about the vertical), that is, a degree of rotational freedom is lost. This is equivalent to the ‘gimbal lock’ (or nadir) condition which arises with a set of mechanical gimbals when the pitch, or inner, gimbal is rotated through 90° .

When θ approaches $+\pi/2$, either ϕ or ψ may be selected arbitrarily to satisfy some other condition while the unspecified angle is chosen to satisfy eqn. (3.68). To avoid ‘jumps’ in the values of ϕ or ψ between successive calculations when θ is in the region of $+\pi/2$, one approach would be to ‘freeze’ one angle, ϕ , for instance, at its current value and to calculate ψ in accordance with eqn. (3.68). At the next iteration, ψ would be frozen and ϕ determined using eqn. (3.68). This process of updating ϕ or ψ alone at successive iterations would continue until θ is no longer in the region of $+\pi/2$.

3.7 Detailed navigation equations

3.7.1 Navigation equations expressed in component form

For a terrestrial navigation system operating in the local geographic reference frame, it has been shown (Section 3.5.3) that the navigation equation may be expressed as follows:

$$\dot{\mathbf{v}}_e^n = \mathbf{f}^n - (2\omega_{ie}^n + \omega_{en}^n) \times \mathbf{v}_e^n + \mathbf{g}_l^n \quad (3.69)$$

where, \mathbf{v}_e^n represents velocity with respect to the Earth expressed in the local geographic frame defined by the directions of true north, east and the local vertical, in component form:

$$\mathbf{v}_e^n = [v_N \quad v_E \quad v_D]^T \quad (3.70)$$

\mathbf{f}^n is the specific force vector as measured by a triad of accelerometers and resolved into the local geographic reference frame;

$$\mathbf{f}^n = [f_N \quad f_E \quad f_D]^T \quad (3.71)$$

ω_{ie}^n is the turn rate of the Earth expressed in the local geographic frame;

$$\omega_{ie}^n = [\Omega \cos L \quad 0 \quad -\Omega \sin L]^T \quad (3.72)$$

ω_{en}^n represents the turn rate of the local geographic frame with respect to the Earth-fixed frame; the transport rate. This quantity may be expressed in terms of the rate of change of latitude and longitude as follows:

$$\omega_{en}^n = [\dot{\ell} \cos L \quad -\dot{L} \quad -\dot{\ell} \sin L]^T \quad (3.73)$$

Writing $\dot{\ell} = v_E/(R_0 + h) \cos L$ and $\dot{L} = v_N/(R_0 + h)$ yields:

$$\omega_{en}^n = \left[\frac{v_E}{R_0 + h} \quad -\frac{v_N}{R_0 + h} \quad -\frac{v_E \tan L}{R_0 + h} \right]^T \quad (3.74)$$

where R_0 is the radius of the Earth and h is the height above the surface of the Earth.

\mathbf{g}_l^n is the local gravity vector which includes the combined effects of the mass attraction of the Earth (\mathbf{g}) and the centripetal acceleration caused by the Earth's rotation ($\omega_{ie} \times \omega_{ie} \times \mathbf{R}$). Hence, we may write

$$\mathbf{g}_l^n = \mathbf{g} - \omega_{ie} \times \omega_{ie} \times \mathbf{R} = \mathbf{g} - \frac{\Omega^2(R_0 + h)}{2} \begin{pmatrix} \sin 2L \\ 0 \\ 1 + \cos 2L \end{pmatrix} \quad (3.75)$$

The navigation equation may be expressed in component form as follows:

$$\begin{aligned} \dot{v}_N &= f_N - v_E(2\Omega + \dot{\ell}) \sin L + v_D \dot{L} + \xi g \\ &= f_N - 2\Omega v_E \sin L + \frac{v_N v_D - v_E^2 \tan L}{R_0 + h} + \xi g \end{aligned} \quad (3.76)$$

$$\begin{aligned} \dot{v}_E &= f_E + v_N(2\Omega + \dot{\ell}) \sin L + \dot{v}_D(2\Omega + \dot{\ell}) \cos L - \eta g \\ &= f_E + 2\Omega(v_N \sin L + v_D \cos L) + \frac{v_E}{R_0 + h}(v_D + v_N \tan L) - \eta g \end{aligned} \quad (3.77)$$

$$\dot{v}_D = f_D - v_E(2\Omega + \dot{\ell}) \cos L - v_N \dot{L} + g = f_D - 2\Omega v_E \cos L - \frac{v_E^2 + v_N^2}{R_0 + h} + g \quad (3.78)$$

where ξ and η represent angular deflections in the direction of the local gravity vector with respect to the local vertical owing to gravity anomalies, as discussed in Section 3.7.4.

Latitude, longitude and height above the surface of the Earth are given by:

$$\dot{L} = \frac{v_N}{R_0 + h} \quad (3.79)$$

$$\dot{\ell} = \frac{v_E \sec L}{R_0 + h} \quad (3.80)$$

$$\dot{h} = -v_D \quad (3.81)$$

It is assumed, in the equations given above, that the Earth is perfectly spherical in shape. Additionally, it is assumed that there is no variation in the Earth's gravitational field with changes in the position of the navigation system on the Earth or its height above the surface of the Earth.

The modifications which must be applied to the navigation equations in order to take account of the errors introduced by these assumptions and so permit accurate navigation over the surface of the Earth are summarised briefly in the following sections. The reader requiring a more detailed analysis of these effects is referred to the texts by Britting [3] and Steiler and Winter [4] in which such aspects are discussed in detail.

3.7.2 The shape of the Earth

It is clear from the preceding analysis that, in order to determine position on the Earth using inertial measurements, it is necessary to make some assumptions regarding the shape of the Earth. The spherical model assumed so far is not sufficiently representative for very accurate navigation. Owing to the slight flattening of the Earth at the poles, it is customary to model the Earth as a reference ellipsoid which approximates more closely to the true geometry. Terrestrial navigation involves the determination of velocity and position relative to a navigational grid which is based on the reference ellipsoid; see illustration in Figure 3.21.

In accordance with this model, the following parameters may be defined:

$$\begin{aligned} \text{the length of the semi-major axis, } & R \\ \text{the length of the semi-minor axis, } & r = R(1 - f) \\ \text{the flattening of the ellipsoid, } & f = (R - r)/R \\ \text{the major eccentricity of the ellipsoid, } & e = [f(2 - f)]^{1/2} \end{aligned} \quad (3.82)$$

By modelling the Earth in accordance with a reference ellipsoid as defined here, a meridian radius of curvature (R_N) and a transverse radius of curvature (R_E) may be derived in accordance with the following equations:

$$R_N = \frac{R(1 - e^2)}{(1 - e^2 \sin^2 L)^{3/2}} \quad (3.83)$$

$$R_E = \frac{R}{(1 - e^2 \sin^2 L)^{1/2}} \quad (3.84)$$

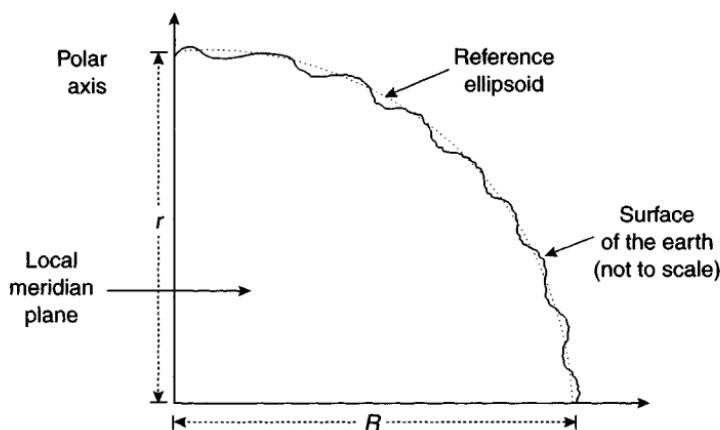


Figure 3.21 Reference ellipsoid

The rates of change of latitude and longitude may then be expressed in terms of a R_N and R_E as follows:

$$\dot{L} = \frac{v_N}{R_N + h} \quad (3.85)$$

$$\dot{\ell} = \frac{v_E \sec L}{R_E + h} \quad (3.86)$$

The mean radius of curvature used in the earlier equations, $R_0 = (R_E R_N)^{1/2}$. The flattening of the Earth at the poles gives rise to a difference of approximately 20 km between the mean radius used for a spherical earth model and the measured polar radius of approximately 20 km.

Similarly, the transport rate now takes the following form:

$$\omega_{en}^n = \begin{bmatrix} \frac{v_E}{R_E + h} & \frac{-v_N}{R_N + h} & \frac{-v_E \tan L}{R_N + h} \end{bmatrix}^T \quad (3.87)$$

Further discussion regarding the shape of the Earth and the choice of datum reference frames appears in the following section.

With the aid of Figure 3.22, the distinction is drawn here between geocentric and geodetic latitude (see following section).

Geocentric latitude at a point on the surface of the Earth is the angle between the equatorial plane and a line passing through centre of Earth and the surface location point. Geodetic latitude at a point on the surface of the Earth is the angle between the equatorial plane and a line normal to the reference ellipsoid which passes through the point.

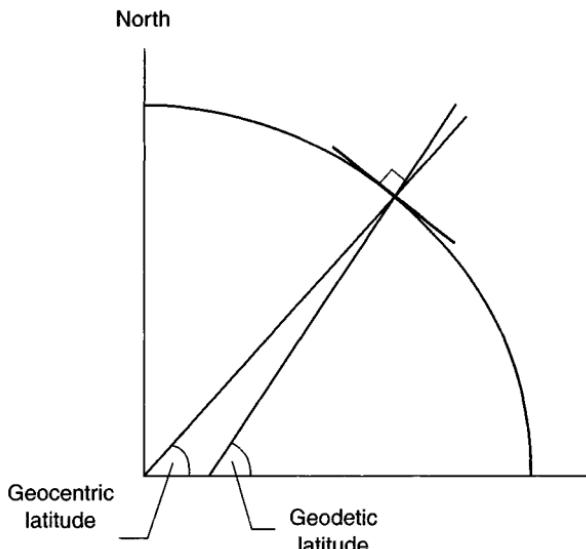


Figure 3.22 Geocentric and geodetic latitude

3.7.3 Datum reference models

The surface of the Earth is highly irregular in shape and can be modelled in various ways.

Topographic models represent the physical shape of the Earth and the mean level of the oceans.

Geodetic models yield a surface which is perpendicular to the local gravity vector at all points; an ‘equipotential surface’. The resulting shape is referred to as a geoid.

Geodesy is the name given to the study of the size and shape of the Earth, and the term geodetic navigation is used here to refer to navigation which takes proper account of this shape. Aspects of geodesy and the use of geodetic datum points for mapping, surveying and navigation are discussed in the following paragraphs. For a more detailed discussion of the subject and a full definition of the terminology, the reader is referred to the standard text book by Bomford [5].

Local variations in gravity, caused by variations in the Earth’s core and surface materials, cause the surface of the geoid, the gravity surface, to be irregular. Whilst it is much smoother than the physical surface of the Earth, as represented by a topographic model, it is too irregular to be used as a surface in which to specify spatial coordinates. For terrestrial navigation, a geometrical shape that approximates closely to the geoid model is used; an ellipsoid, which in this context is a three-dimensional (3-D) shape formed by rotating an ellipse about its minor axis. The term oblate spheroid is sometimes used in place of ellipsoid.

The term geodetic datum is used to define the ellipsoid and its positional relationship with respect to the solid Earth. In combination with an axis definition, a geodetic datum defines a 3-D geographic co-ordinate system, the dimensions being geodetic latitude and longitude and ellipsoidal height (height above the surface of the ellipsoid).

In practice, vertical position is not defined with respect to the surface of the ellipsoid because this surface offers no physical reference point for measurement. The geoid, which corresponds approximately to mean sea level, offers a much more convenient vertical reference. For this reason, height above mean sea level is most commonly used. For land based surveys [6], the reference level used as a zero datum is defined by mean sea level at a selected coastal location, or an average value of mean sea level at several locations, over a specified period of time. Land surveys should reference the vertical datum chosen. In the United Kingdom, the vertical datum used is Ordnance Datum Newlyn (ODN) and in the United States, the North American Vertical Datum of 1988 (NAVD88).

It is possible to define a geodetic datum which approximates to the shape of the Earth over the entire globe. The figures given in Figure 3.23 were defined for such a datum by the World Geodetic System Committee in 1984, the WGS-84 model [7]. The value of Earth’s rate is discussed in Figure 3.24.

Many geodetic datum points used for mapping, surveying and navigation are defined to provide a more precise fit over a restricted geographical area, the

Length of the semi-major axis,	R	= 6378137.0 m
Length of the semi-minor axis,	$r = R(1-f)$	= 6356752.3142 m
Flattening of the ellipsoid,	$f = (R-r)/R$	= 1/298.257223563
Major eccentricity of the ellipsoid,	$e = [f(2-f)]^{1/2}$	= 0.0818191908426
Earth's rate (see Figure 3.24),	Ω	$= 7.292115 \times 10^{-5} \text{ rad/s}$ (15.041067°/h)

Figure 3.23 WGS-84 model⁴

The duration of a solar day is 24h, the time taken between successive rotations for an Earth-fixed object to point directly at the sun. The Sidereal day represents the time taken for the Earth to rotate to the same orientation in space and is of slightly shorter duration than the solar day, 23h, 56min, 4.1s. The Earth rotates through one geometric revolution each Sidereal day, not in 24h, which accounts for the slightly strange value of Earth's rate.

$$\begin{aligned}\text{Solar day} &= 24 \text{ h} \\ \text{Sidereal day} &= 23 \text{ h } 56 \text{ min } 4.1 \text{ s} \\ \text{Earth's day} &= 1 \text{ revolution/Sidereal day} \\ &= 7.292115 \times 10^{-5} \text{ rad/s} \\ &= 15.041067^{\circ}/\text{h}\end{aligned}$$

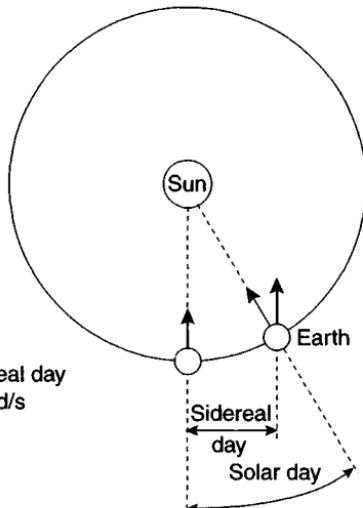


Figure 3.24 The solar day and the sidereal day

UK Ordnance Survey (OSGB 1936), the British National Grid for example. Regional datum points such as this have proliferated over time with the result that their areas of application may overlap. As a consequence, it is necessary when referring to a positional location on the Earth in terms of its latitude and longitude, to specify also

⁴ The former Soviet Union devised a similar model, SGS-90 or PZ-90, discussed in Appendix D.

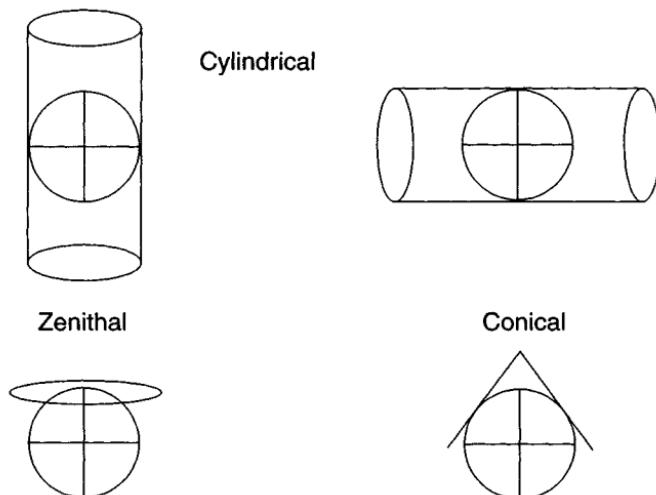


Figure 3.25 Examples of orthomorphic projection schemes used for mapping

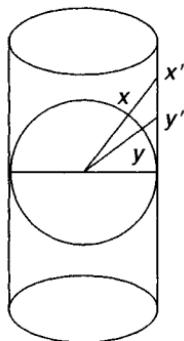


Figure 3.26 Standard Mercator projection

the corresponding geodetic datum or geographic co-ordinate reference. Contrary to common belief, the coordinates alone do not adequately define a particular location.

Lines of constant latitude and longitude are curved in three dimensions, but may be represented on a plane by means of a projection. The resulting rectangular co-ordinate system on the plane is called a grid. Various projections of the Earth's surface into two dimensions have been used, using a geodetic reference ellipsoid as the basis for projection.

A flat grid system can be obtained by projecting the reference ellipsoid onto a cylindrical, conical or flat shape as indicated in Figure 3.25. It is noted that the x and y axes must be orthomorphic, that is, of equal scale.

The Standard Mercator projection, illustrated in Figure 3.26, is generated by placing a cylinder over the Earth so that the contact point is around the equator. A point

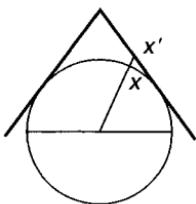


Figure 3.27 Lambert conical projection

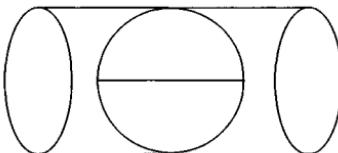


Figure 3.28 Universal Transverse Mercator projection

on the Earth is projected onto the inside of the cylinder by taking a line from the centre of the Earth through the point (e.g. x to x' , y to y' as shown in Figure 3.26). When all points have been projected onto the cylinder, the cylinder is unwrapped and laid flat. This type of projection produces the most commonly observed map of the world.

The Mercator is a useful projection for navigation because a bearing direction on the globe is very similar to a direction on the projection. However, distances and areas will be distorted on small scale maps. Mercator and other projections do not preserve scale or area. For instance, with Standard Mercator, distance along the equator is represented exactly, but distance at higher latitude is magnified. As a result, Greenland appears to be the same size as South America, whereas it is actually about one-third of the size.

Various projection techniques are used to overcome the effects of distortion in regional maps where it is required to have a map or rectangular grid system that is appropriate to the locality of interest. For example, the Lambert conical projection, illustrated in Figure 3.27, provides an accurate representation of the area around the point of contact between the cone and the reference ellipsoid. Alternatively, the use of a cylinder or cone that cuts the reference ellipsoid in two places close to the area of interest may be used. This allows a reduction in the distortion to be achieved adjacent to and between the points of contact, providing two horizontal parallels and minimum vertical distortion.

One of the most commonly used projections is the Universal Transverse Mercator (UTM) projection. This uses the same principle as the Standard Mercator, with the exception that the cylinder is rotated through 90° so that the contact point of the cylinder with the Earth is along a meridian line; see Figure 3.28. To minimise distortion and preserve accuracy, the technique is employed of using only a small strip on either side of a designated central meridian. The UTM projection is used worldwide.

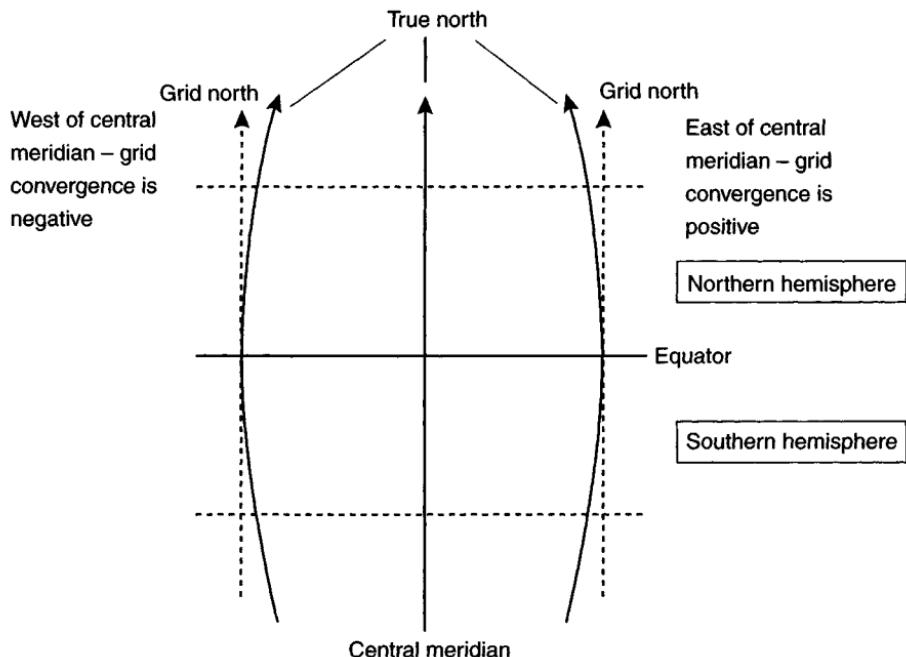


Figure 3.29 Universal Transverse Mercator grid zone

The UTM system uses the Transverse Mercator in separate zones each 6° wide. UTM zones are identified alphanumerically; numbering from 1 to 60, starting with 1 in the 180°W – 174°W zone and increasing eastwards to 60 in the 174°E – 180°E zone. Hence grid zone 32 is from 6° to 12° longitude. The UTM grid is further subdivided into blocks of 8° of latitude from -80° to $+80^{\circ}$ which are identified by the use of the letters C to X, excluding I and O (e.g. 64°S – 56°S is E).

Any UTM grid zone can be represented as shown in Figure 3.29 where the central meridian is an odd number and a multiple of 3. The grid north and the true north will only be coincident along the central meridian and on the equator. At all other points there is a difference referred to as ‘convergence’. It is noted that to the true north is west of grid north to the right of the central meridian and true north is east of grid north to the left of the central meridian.

3.7.4 Variation of gravitational attraction over the Earth

As described earlier, accelerometers provide measurements of the difference between the acceleration with respect to inertial space and the gravitational attraction acting at the location of the navigation system. In order to extract the precise estimates of true acceleration needed for very accurate navigation in the vicinity of the Earth, it is necessary to model accurately the Earth’s gravitational field. This of course is also true for navigation close to any other body with a gravitational field.

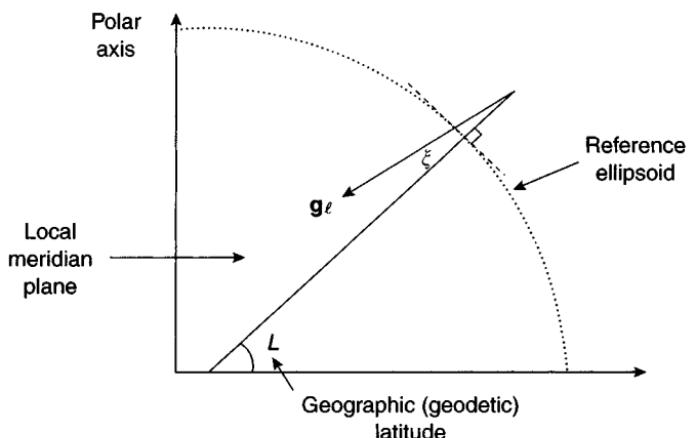


Figure 3.30 Deflection of local vertical owing to gravity anomalies

It is assumed in the earlier derivation of the navigation equation that the gravity vector acts vertically downwards, that is, normal to the referenced ellipsoid. In practice, both the magnitude and the direction of the gravity vector vary with position on the Earth's surface and height above it. Variations occur because of the variation between the mass attraction of the Earth and gravity vector; the centrifugal acceleration being a function of latitude. In addition, gravity varies with position on the Earth because of the inhomogenous mass distribution of the Earth. Such deviations in the magnitude and direction of the gravity vector from the calculated values are known as gravity anomalies.

Mathematical representations of the Earth's gravitational field are discussed in some depth by Britting [3]. The deflection of the local gravity vector from the vertical may be expressed as angular deviations about the north and east axes of the local geographic frame as follows:

$$\mathbf{g}_l = [\xi g, -\eta g, g]^T \quad (3.88)$$

where ξ is the meridian deflection and η is the deflection perpendicular to the meridian. The deflection in the meridian plane is illustrated in Figure 3.30.

The resulting deviation of the vertical over the surface of the Earth varies by up to 30 arc s.

The precise knowledge of the gravity vector becomes important for certain high accuracy applications, such as for marine navigation where the deflection of the vertical becomes an important factor. Exact knowledge of the magnitude of gravity is also vital for the testing of very precise accelerometers, that is, sensors having a measurement bias of less than $10^{-5} g$. Similarly, it is important for surveying and gravity gradiometry, where attempts are made to measure the gravity vector very accurately.

Various international models for the variation of gravity with latitude are given in the literature. Steiler and Winter [4] give the following expressions for the variation

of the magnitude of the gravity vector with latitude at sea level ($h = 0$) and its rate of change with height above ground:

$$g(0) = 9.780318(1 + 5.3024 \times 10^{-3} \sin^2 L - 5.9 \times 10^{-6} \sin^2 2L) \text{ m/s}^2 \quad (3.89)$$

$$\frac{dg(0)}{dh} = -0.0000030877(1 - 1.39 \times 10^{-3} \sin^2 L) \text{ m/s}^2/\text{m} \quad (3.90)$$

For many applications, precise knowledge of gravity is not required and it is sufficient to assume that the variation of gravity with altitude is as follows:

$$g(h) = \frac{g(0)}{(1 + h/R_0)^2} \quad (3.91)$$

where $g(0)$ is derived from eqn. (3.89).

References

- 1 WRAY, G.L., and FLYNN, D.J.: 'An assessment of various solutions of the navigation equation for a strapdown inertial system'. Technical report 79017, Royal Aircraft Establishment, January 1979
- 2 SHEPPERD, S.W.: 'Quaternion from rotation matrix', *AIAA Journal of Guidance and Control*, 1978, **1** (3)
- 3 BRITTING, K.: 'Inertial navigation system analysis' (Wiley Interscience, New York, 1971)
- 4 STEILER, B., and WINTER, H.: 'AGARD flight test instrumentation volume 15 on gyroscopic instruments and their application to flight testing'. AGARD-AG-160-VOL.15, September 1982
- 5 BOMFORD, G.: 'Geodesy' (Oxford University Press Inc., New York City, 1985, 4th edn.)
- 6 WILLIAMSON, H.S., and WILSON, H.F.: 'Directional drilling and earth curvature', *SPE Drill. & Completions*, 2000, **15** (1)
- 7 'Department of Defence World Geodetic System 1984: Its definition and relationship with local geodetic systems'. DMA TR 8350.2, September 1987

Chapter 4

Gyroscope technology 1

4.1 Introduction

Gyroscopes are used in various applications to sense either the angle turned through by a vehicle or structure (displacement gyroscopes) or, more commonly, its angular rate of turn about some defined axis (rate gyroscopes). The sensors are used in a variety of roles such as:

- stabilisation,
- autopilot feedback,
- flight path sensor or platform stabilisation,
- navigation.

It is possible with modern gyroscopes for a single sensor to fulfil each of the above tasks, but often two or more separate clusters of sensors are used.

The most basic and the original form of gyroscopes makes use of the inertial properties of a wheel or rotor spinning at high speed. Many people are familiar with the child's toy which has a heavy metal rotor supported by a pair of gimbals [1]. When the rotor is spun at high speed, the rotor axis continues to point in the same direction despite the gimbals being rotated. This is a crude example of a mechanical, or conventional, displacement gyroscope.

Examples of mechanical spinning wheel gyroscopes used in strapdown applications are the single-axis rate-integrating gyroscope and twin axis 'tuned' or flex gyroscopes. An alternative class designation for gyroscopes that cannot be categorised in this way, is not surprisingly called unconventional sensors, some of which are solid-state devices. The very broad and expanding class of unconventional sensors includes devices such as:

- Rate transducers which include mercury sphere and magneto-hydrodynamic sensors;
- Vibratory gyroscopes;
- Nuclear magnetic resonance (NMR) gyroscopes;

- Electrostatic gyroscopes (ESGs);
- Optical rate sensors which include ring laser gyroscopes (RLGs) and fibre optic gyroscopes (FOGs);
- Micro-machined electromechanical system (MEMS) gyroscopes.

Whilst many of the sensors in this class are strictly angular rate sensors and not gyroscopes in the sense that they do not rely on the dynamical properties of rotating bodies, it has become accepted that all such devices be referred to as gyroscopes since they all provide measurements of body rotation.

In this chapter, some conventional sensors are described followed by sections which outline the principles of operation and performance of some of the other gyroscope technologies noted above. Finally, a brief mention is made of other forms of instrument or novel techniques that may be used to sense rotational motion. Optical and MEMS gyroscope technologies are discussed separately in Chapters 5 and 7, respectively.

Throughout this and the later chapters on gyroscope technology, emphasis is placed on those sensors which are used, or have the potential to be used, in strapdown inertial systems. It is for this reason that both optical and MEMS sensors are described in some detail. Advances in interferometric fibre optic gyroscope (IFOG) technology are leading to the wider application of these devices, whilst MEMS sensors are seen very much as the technology of the future with wide application in strapdown systems. Details of fabrication of various types of gyroscopes can be found in References 2 and 3.

4.2 Conventional sensors

4.2.1 Introduction

Conventional gyroscopes make use of the inertial properties of a wheel or rotor spinning at high speed [2, 3]. A spinning wheel tends to maintain the direction of its spin axis in space by virtue of its angular momentum vector, the product of its inertia and spin speed, and so defines a reference direction. The development of the mechanical gyroscope owes much to the excellent work of Professor C.S. Draper and his co-workers, at the Massachusetts Institute of Technology. The performance which may be achieved using gyroscopes of this type varies from the precision devices with error rates of less than $0.001^\circ/\text{h}$, to less accurate sensors with error rates of tens of degrees per hour. Many devices of this type have been developed for strapdown applications, being able to measure angular rates up to about $500^\circ/\text{s}$. Some designs are very rugged, having characteristics which allow them to operate in harsh environments such as guided weapons.

4.2.2 Fundamental principles

There are several phenomena on which the operation of the conventional spinning mass gyroscope depends, namely gyroscopic inertia, angular momentum and

precession. In the case of two-degrees-of-freedom gyroscopes, there are also the phenomena of nutation, gimbal lock and tumbling. These are considered in turn in the following sections.

4.2.2.1 Gyroscopic inertia

Gyroscopic inertia is fundamental to the operation of all spinning mass gyroscopes, as it defines a direction in space that remains fixed in the inertial reference frame, that is, fixed in relation to a system of coordinates which do not accelerate with respect to the 'fixed stars'. The establishment of a fixed direction enables rotation to be detected, by making reference to this fixed direction. The rotation of an inertial element generates an angular momentum vector which is coincident with the axis of spin of the rotor or 'wheel'. It is the direction of this vector which remains fixed in space, given perfection in the construction of the gyroscope.

A practical reference instrument may be designed by having the rotor supported in a set of frames or gimbals which are free to rotate with respect to one another as shown in Figure 4.1. This is an external gimbal type gyroscope. The orientation of the case of the instrument with respect to the direction of the spin axis may be measured with angle pick-off devices mounted on the gimbals.

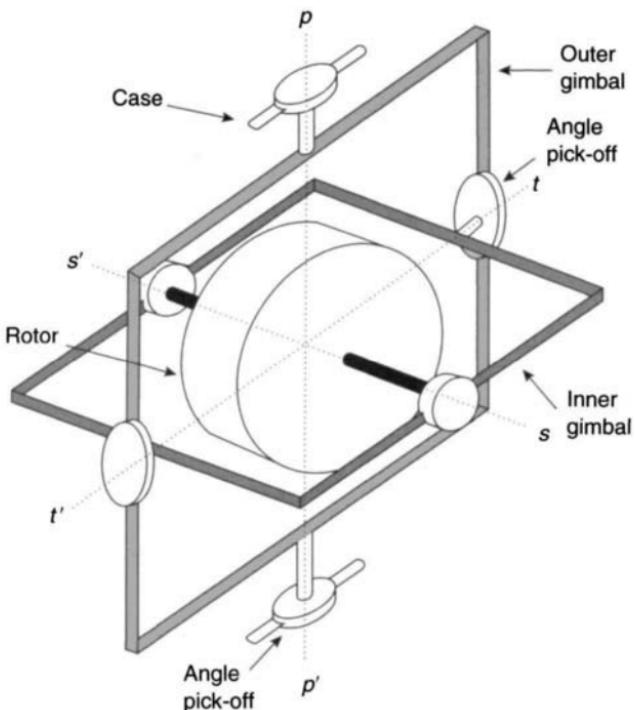


Figure 4.1 Schematic diagram of a two-axis gyroscope

4.2.2.2 *Angular momentum*

The angular momentum (H) of a rotating body is the product of its moment of inertia (I) and its angular velocity (ω_s) referred to the same axis of rotational motion, that is,

$$H = I\omega_s \quad (4.1)$$

where I is the sum of the products of the mass elements that make up the rotor and the square of their distances from the given axis.

Angular momentum is defined by the distribution of mass on a rotor as well as by its angular velocity. For many applications, the angular momentum is chosen to be very high, so that the undesired torques that can act on a rotor and cause errors are virtually insignificant. This of course, given good design and fabrication techniques, results in a gyroscope with little movement of the direction of the spin axis. Any undesired movement of the direction of the spin axis is usually referred to as ‘drift’. Clearly, one technique for producing a very high angular momentum is to have the majority of the mass of the rotor at its edge owing to the dependence of the moment of inertia on the square of the distance of its mass element from the centre of rotation.

Careful consideration must be given to the value of the angular momentum selected for a gyroscope to be used in a given application. The choice of a very high angular momentum should result in negligible drift, but there could be some considerable penalties. The gyroscope would almost certainly be relatively large and heavy, and it may take many seconds, if not minutes, for the rotor to reach its operating speed. Further, when used in a strapdown mode, the associated control system may not be capable of recording, or ‘capturing’, angular rates beyond a few tens of degrees per second. Hence, many compromises have to be made when selecting a gyroscope for a given application.

4.2.2.3 *Precession*

Because the motion of a spinning mass occurs in a way that does not coincide with ‘common sense’ expectations, it has acquired a confusing aura of mystery. Some simple explanation may help.

First, think of the gyroscope rotor mounted in bearings in a gimbal, as shown in Figure 4.2. The gimbal axis system has one axis through the bearing axes, ss' , and two mutually orthogonal axes through the centre of mass of the rotor, tt' and pp' .

Spin is the rotation of the gyroscope rotor relative to the gimbal.

Precession is the rotation of the gimbal, relative to inertial space. In the case of a freely spinning body, such as the Earth (or the rotor of an electrostatic gyroscope, see Section 4.7), there is not a material frame with spin bearings. In this case, the precession must be considered to be that of the axis system which an imaginary gimbal would have – one axis through the north and south poles, and two mutually orthogonal in the plane of the Equator.

Consider now the disc shown in Figure 4.2 spinning about the axis ss' . If the disc is acted upon by a couple, that is, a torque, the torque being about the axis tt' ,

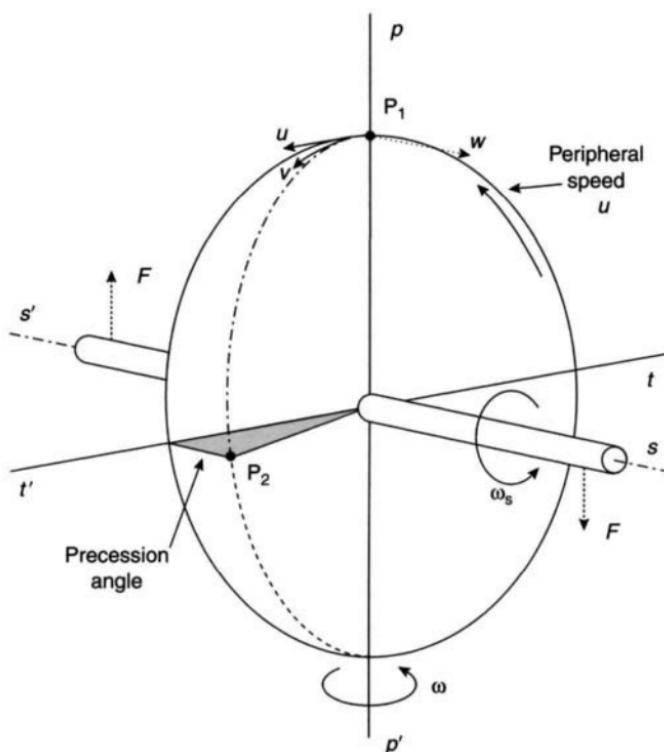


Figure 4.2 Simple explanation of precession

the spin axis of the disc will be forced to turn about the axis pp' . This turning is the precession. Note that the precession axis, pp' , is orthogonal to the torque axis, tt' . It is the unexpectedness of this result which causes confusion. However, if Newton's laws are applied carefully, the result can be explained, both qualitatively and quantitatively.

The disc is spinning about the axis ss' in an anti-clockwise direction looking from s to s' . Suppose that the disc is rigid with all the mass in the rim and that the rim has a peripheral speed u . Consider an element of mass at the highest point, P_1 . Apply an impulsive couple FF , as shown in the figure in an anti-clockwise direction looking from t to t' .

The instantaneous velocity of the mass is changed by adding the velocity w in the same sense as the couple FF . The resultant velocity, v , is now in a different direction. It is noted that the other elements of the rim change their velocities in proportion to their distance from the axis tt' . After the disc has spun through 90° , the element of mass arrives at the point P_2 , which is not in the expected line tt' , but in a plane which has precessed about the axis pp' .

This simple picture indicates how the spinning disc reacts to the impulsive couple, and shows the axis and sense of the precession. Of course, the process is normally continuous, not impulsive. The dynamics can be analysed using co-ordinate geometry

and applying Newton's laws. The result agrees with eqn. (4.5) which is arrived at using vectors.

The particles making up a spinning body undergo:

1. accelerations caused by accelerations of the centre of mass of the body;
2. centripetal accelerations caused by the spinning of the body;
3. Coriolis accelerations as a result of the precession of the body.

The Coriolis accelerations are simply the additional accelerations experienced by a mass moving relative to an axis system when that axis system is itself rotating in inertial space. The precession torque is simply the torque necessary to produce the sum of the particle masses times their Coriolis accelerations.

Mathematical description of precession

Consider a heavy spinning disc, as shown in cross section in Figure 4.3, with angular momentum H defined by the vector OA, that is, $H\mathbf{a}$, where \mathbf{a} is a unit vector.

From Newton's first law, applied to angular motion, the angular momentum vector \mathbf{H} remains constant unless the disc is acted upon by the torque. Let us suppose that a torque T is applied to the disc which causes it to precess at a rate ω ($= \omega_c$, where \mathbf{c} is also a unit vector) about an axis which will lie in the plane of the disc and may be taken to be normal to the plane of the paper. Over a short period of time δt the disc will have precessed through an angle $\omega\delta t$ about \mathbf{c} , and the angular momentum vector will have changed to OB, that is, to $(H + \delta H)\mathbf{b}$, where $\mathbf{b} = \mathbf{a} + \omega\delta t(\mathbf{c} \times \mathbf{a})$.

The change in angular momentum over this time is represented by the vector AB and may be expressed as:

$$\begin{aligned}\delta H &= (H + \delta H)\mathbf{b} - H\mathbf{a} \\ &= H(\mathbf{b} - \mathbf{a}) + \delta H\mathbf{b}\end{aligned}$$

that is,

$$\delta H = H\omega\delta t(\mathbf{c} \times \mathbf{a}) + \delta H\mathbf{b} \quad (4.2)$$

Thus, in the limit, as $\delta t \rightarrow 0$, the rate of change of angular momentum is given by

$$\frac{dH}{dt} = H\omega(\mathbf{c} \times \mathbf{a}) + \frac{dH}{dt}\mathbf{b}$$

that is,

$$\frac{dH}{dt} = \omega \times \mathbf{H} + \frac{dH}{dt}\mathbf{b} \quad (4.3)$$

From Newton's second law, the rate of change of angular momentum is equal to the torque \mathbf{T} applied to the body, hence

$$\mathbf{T} = \omega \times \mathbf{H} + \frac{dH}{dt}\mathbf{b} \quad (4.4)$$

Thus, the component of the torque which is along the spin axis \mathbf{b} gives rise to an acceleration in the spin rate. In a practical gyroscope, this is normally negligible and countered by the effect of the spin motor. The component normal to the spin axis

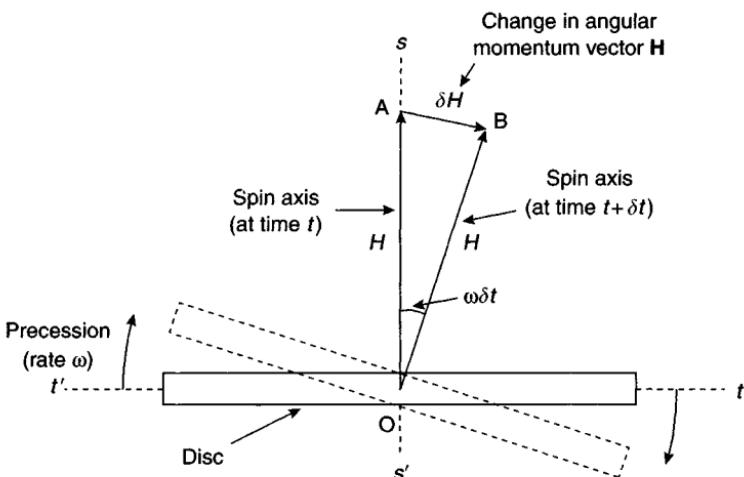


Figure 4.3 Illustration of precession

gives rise to a precession ω which is normal to both the torque and the spin axes, and from inspection of the figure, the direction of the precession is such as to try to align the spin axis with the torque axis.

Neglecting the component along the spin axis, in vector terms we may write

$$\mathbf{T} = \omega \times \mathbf{H} \quad (4.5)$$

and in magnitude terms

$$T = \omega H \quad (4.6)$$

This is sometimes known as the law of gyroscopes.

The application of the precession principle

The principle of precession can be exploited to provide a very accurate measure of angular rotation or rotation rate. Since a spinning wheel, or rotor, will only precess if a torque is applied to it, a rotor suspended in an instrument case by gimbals will maintain its spin axis in a constant direction in space. Changes in the angles of the gimbals will then reflect any changes in orientation of the case with reference to the spin axis direction.

Alternatively, if controlled torques are applied to the rotor to keep its spin axis aligned with a direction defined by the case of the instrument, then the measurement of these torques will provide measurements of the angular velocity of the instrument, and hence of the angular velocity of any body to which the instrument is attached.

Note that when a torque is applied to the rotor, which responds by precessing, then there is an equal but opposite reaction torque from the rotor to the application mechanism. However, if precession is prevented, as when the supporting gimbal hits a stop, then the reaction torque disappears and the rotor and gimbal act as a non-gyroscopic body about this axis.

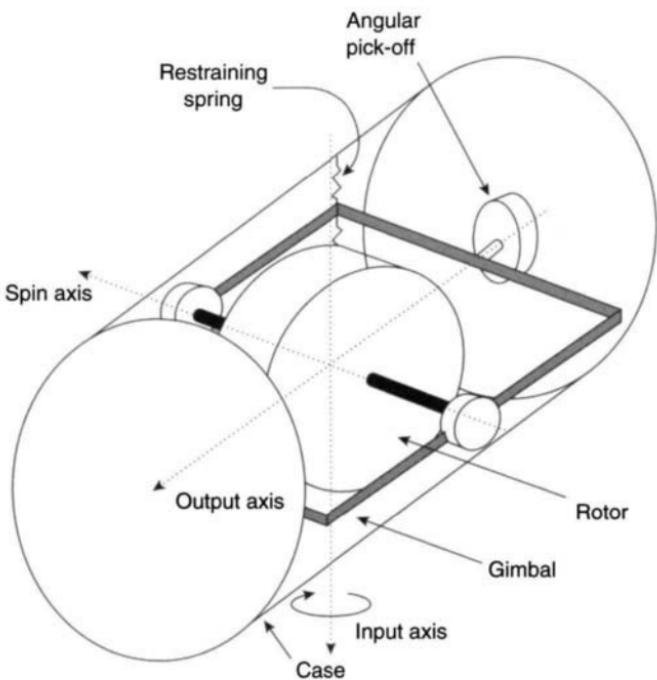


Figure 4.4 A single-axis gyroscope

In the single-axis rate gyroscope, shown schematically in Figure 4.4, the gyroscope's rotor is supported by a single gimbal whose axis is normal to the spin axis. The gimbal is restrained about its axis by a spring attached to the case, and there is an angular pick-off which measures the displacement of the gimbal about its axis from a 'null', or zero, position.

In the case where the instrument is rotating about the input axis and the rotor is not precessing at the same rate, the difference in rates will result in elastic compression of the gimbal pivots. This gives a torque on the gimbal (and a reaction torque on the case) which is applied to the rotor about the input axis. This torque causes the rotor to precess initially about the output axis, about which it is free to turn. The resulting displacement about the output axis causes a torque to be produced about this axis by the restraining spring.

The spring torque on the gimbal and the rotor about the output axis results in the rotor precessing about the input axis until, in the steady state, it is precessing at the same rate as the case is turning, with the deflection of the restraining spring providing just that amount of torque needed to keep the case and rotor in alignment.

Assuming that the restraining spring is linear, the deflection of the gimbal is proportional to the torque required to keep the rotor precessing with the case, and so to the turn rate of the case.

In practice, it is quite difficult to measure angular displacements accurately without resorting to sophisticated and consequently expensive equipment. However,

it is quite easy to measure accurately a fixed or defined position, particularly a zero deflection or 'null' position. Hence, if the spin axis of a rotor is made to precess back to the 'null' position by the application of a suitable torque, there is potential for very accurate angular measurement, provided that the torque required to null the deflection can be generated and measured.

This is achieved in practice by replacing the restraining spring with an electro-magnetic torque generator that produces a torque to cause precession of the rotor in a direction opposite to that caused by rotation of the case about its input axis. The current required can be measured very accurately by simple techniques, and when the system is 'balanced' this current is directly proportional to the applied angular rate. This technique is commonly called nulling, and is fundamental to the use of strapdown techniques. Hence, application of the reverse of the precession principle enables very accurate measurements to be made of the angular displacement or rate of turn of the case of the rate gyroscope.

A conventional single-axis gyroscope can be considered to have three orthogonal axes: its rotor or spin axis, an input precession axis and an output or torque axis. That is, torque is always applied about the output axis to cause precession about the input axis to keep it in alignment with the case. In the case of the so-called two- or dual-axis gyroscopes, these sensors have a spin axis with two orthogonal input axes. In this case, angular motion of the case of the sensor with respect to the rotor is sensed by pick-off angle sensors on the gimbals, as indicated in Figure 4.1, and each gimbal is provided with a torque mechanism.

Even the most accurate of gyroscopes will appear to drift, or have their spin axis precess. This is because the angular momentum vector is fixed with respect to space axes, not the co-ordinate system defined by the Earth. Hence, for some orientations on Earth, it is necessary to apply corrective torques to precess the gyroscope if it is to be used as an Earth reference.

4.2.2.4 Nutation

This is a natural phenomenon that occurs with so-called two-degrees-of-freedom gyroscopes, such as those in which the rotor is supported by a gimbal structure. Nutation is simply a wobbling of the spin axis of the rotor. It is a self-sustaining oscillation which physically represents a continuous transfer of energy from one degree of freedom to the other and back again. In contrast to precession, this motion does not need any external torques to sustain it. This motion has a natural frequency ω_R , commonly known as the nutation frequency, given by:

$$\omega_R = \frac{H}{\sqrt{I_{ig} I_{og}}} \quad (4.7)$$

where H is the angular momentum of the rotor; I_{ig} the moment of inertia of the rotor and inner gimbal about the inner axis and I_{og} is the moment of inertia of the rotor, inner gimbal and outer gimbal about the outer axis.

In a frictionless system, nutation would be self-perpetuating. However, friction in the gimbal bearings or deliberately applied viscous drag damps out this undesirable

motion. Energy dissipation varies in proportion to the nutation frequency. Therefore, in order to minimise the occurrence of nutation, it is required to increase ω_R . It is usual for the rotor to have as large an angular momentum as possible, combined with gimbals having low moments of inertia. This is achieved through the use of light but stiff materials such as beryllium (alloy) in the construction of gimbals.

4.2.2.5 *Gimbal lock*

Gimbal lock is an effect which prevents a two-degrees-of-freedom gyroscope having 360° of freedom about both its inner and outer gimbal axes. Gimbal lock occurs when the spin axis of the rotor coincides with the outer gimbal axis owing to a 90° rotation about the inner axis. At this point, the gyroscope loses one degree of freedom. Application of motion about an axis perpendicular to the plane containing the outer gimbal causes the outer gimbal to spin. Once this spinning motion has begun, the spin axis of the rotor and the axis of the outer gimbal remain permanently coaxial. The only method of separating them is to stop the rotation of the inertial element to allow the two axes to be reset. This undesirable effect is prevented by using mechanical stops to limit the motion of the inner gimbal. These stops usually permit up to $\pm 85^\circ$ of motion by the inner gimbal. Gimbal lock can also occur in stable platforms with three gimbals.

4.2.2.6 *Tumbling*

Tumbling is a consequence of using mechanical stops to prevent gimbal lock. This phenomenon occurs when the inner gimbal hits one of the mechanical stops. This causes the outer gimbal to turn through 180° about its own axis. This motion of the outer gimbal is known as tumbling. Once tumbling occurs, the reference is lost.

4.2.3 *Components of a mechanical gyroscope*

The basic components of mechanical gyroscopes are as follows:

- 1 *The instrument case:* The case in which the other elements are housed and which provides the structure by which the instrument is mounted in a vehicle.
- 2 *The rotor or inertial element:* This is essentially a flywheel rotated at high angular velocity. The rotor usually has the majority of its mass at the outer edge, as the moment of inertia is the sum of the products of the individual masses (m_i) and the square of their distance (r_i) from the axis of rotation, $\sum m_i r_i^2$. This enables a high angular momentum (H) to be achieved for a given angular velocity, since H is the product of moment of inertia (I) and angular velocity (ω_s) as described in the previous section. This approach also allows a high angular momentum to be achieved with the lowest overall mass. A low rotor mass is desirable to minimise vibration and shock effects.
- 3 *Gimbals:* These are support frames on which the rotor or another gimbal is mounted to isolate the rotor from rotational motion, by allowing freedom of angular movement of these frames about the rotor. In the case of a two-gimbal

sensor, the axes of rotation of the two gimbals and the rotor are arranged to be mutually orthogonal as shown in Figure 4.1.

- 4 *Pick-off*: This device is used to detect relative motion between the rotor and the gimbals or, in some cases, between the rotor and the instrument case. The pick-off produces an electrical signal, indicating the direction and amplitude of the motion from a reference position. There are three basic forms of pick-off technology commonly used with mechanical gyroscopes operating in torque re-balance mode:

- moving coil – using a small receiver coil and an a.c. excitation coil, so that any relative movement between the two modifies the flux sensed by the receiver coil;
- variable reluctance – the excitation and receiver coils are fixed to the case of the gyroscope, with a soft iron assembly attached to the moving component so that it is in the flux return path between the excitation and receiver coils. Motion of the soft iron components causes a change in its orientation in the excitation field, thereby modifying the return flux to the receiver coil;
- capacitive – there is a stationary plate close to the rotor, or moving component, whilst the rotor acts as the other plate of the capacitor. Movement of the rotor about its input axis, or axes for a two axis sensor, causes a change in separation between the two plates of the capacitor and hence there is a change in capacitance.

Open-loop gyroscopes, such as simple rate sensors, often use a potentiometer to sense angular displacement of a gimbal. Generally, this form of sensor is not used for navigation purposes.

- 5 *Torque motor or electromagnetic torquer*: When a gyroscope is used in a closed loop or torque re-balance mode, it is necessary to generate a torque on the rotor in order to return the rotor to the ‘null’, or zero, position. This is achieved using a torque generator, which usually takes one of two common forms:

- Permanent magnet – this type relies on the interaction between the field generated by a permanent magnet and that of an electromagnetic coil. Particularly with single-axis sensors, a coil ‘cup’ is fixed to the moving element and the permanent magnet attached to the case. This has several advantages such as reducing the sensitivity to external magnetic fields and allowing the magnet to be outside of the flotation fluid. However, it does require a pair of flexible leads to the coil which can generate error torques. In general, as a result of other constraints, dynamically tuned gyroscopes have the opposite configuration with the permanent magnet fixed to the moving element.
- Electromagnet – a soft iron component is attached to the sensing element and a coil is fixed to the case. When a current is applied to the coil, a magnetic field is produced that interacts with the soft iron producing a torque on the sensing element.

- 6 *Re-balance loop*: This is the term given to the electronic circuitry that receives and uses the signals from the pick-off assembly. It interprets these signals in terms of the current required in the torquer coils to return the inertial element to its ‘null’

position. The re-balance loop electronics can either be analogue or digital. In the case of the analogue re-balance loop, a continuously variable current is passed through the coil to return the inertial element to its ‘null’ position. When there is no displacement, then there is no current flow. A digital re-balance system generates precision current pulses of particular duration to force the inertial element back to its ‘null’ position. With some implementations, pulses of equal amplitude but opposite sign are passed into the torque generator even when there is no displacement of the rotor. Imbalance in the number of pulses applied in each direction gives rise to a net torque.

- 7 *Spin motor:* This is the motor used to rotate the inertial element and give it the angular momentum that is vital for the operation of the mechanical gyroscope. Usually the spin motor is either a hysteresis motor or an inductive device. Some gyroscopes that have a short run-time use a blast of air or a small explosive charge to spin the inertial element, and for cheap and crude applications, a d.c. electric motor may be used.
- 8 *Float:* Rate-integrating gyroscopes, as discussed below in Section 4.2.5, have their rotor and spin motor sealed in a can that is immersed in a fluid to reduce the load on the gimbal bearings. This can, with its encapsulated components, is known as the float. Careful choice of the flotation fluid can reduce the load of the rotor assembly on the gimbal to zero. In such a design, bellows are used to compensate for changes in the volume of the fluid when the temperature inside the case of the gyroscope changes. The centre of buoyancy is arranged to be close to the centre of gravity of the float and along the output axis in order to minimise acceleration sensitive errors. There is further consideration of the effect of this on sensor performance in Section 4.2.4.
- 9 *Flotation fluid:* This is the fluid in the gyroscope that gives buoyancy to the float in a floated rate-integrating gyroscope. It also provides damping of the motion of the float which gives rise to the integration function for the single-axis rate-integrating gyroscope.
- 10 *Bearings:* For the spin axis of the rotor, most gyroscopes use ball bearings in a race with a retainer, and are chosen to have low noise characteristics. This form of bearing needs a suitable lubricant with the following characteristics:
 - it should not separate into solid and liquid components;
 - it should have small or negligible change in viscosity over the temperature range of the sensor;
 - it should not leak out from the bearing;
 - it should retain its physical and chemical properties for the required shelf-life of the gyroscope.

Lubricants can severely limit the environmental performance and shelf-life potential of a sensor. An alternative form of bearing that overcomes the well-known problems of ‘ball bearings’ is the gas bearing. This form of bearing can be either self-acting or externally pumped. In the former case, the bearing draws a gas, usually air, into a series of grooves which, owing to the viscosity of the fluid, supports the structure in the other part of the bearing. In the latter case, the fluid is pumped into the grooves to support the structure. The drawbacks with

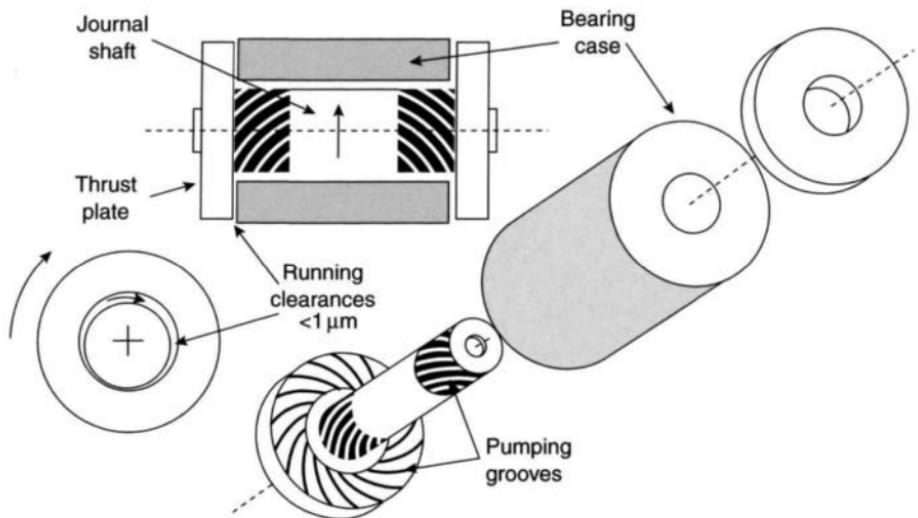


Figure 4.5 Self-pumping gas bearing

gas bearings are the need for very tight tolerances in manufacture, of the order of a micrometre or better, and the use of very hard materials such as boron carbide as the two bearing surfaces touch and rub during both starting and stopping of the rotor. However, they are very low noise bearings and can last a very long time, particularly if the bearing runs continuously. A schematic diagram of a self-pumping gas bearing is shown in Figure 4.5.

4.2.4 Sensor errors

All gyroscopic sensors are subject to errors which limit the accuracy to which the angle of rotation or applied turn rate can be measured. Spurious and undesired torques, caused by design limitations and constructional deficiencies, act on the rotors of all mechanical gyroscopes. These imperfections give rise to precession of the rotor, which manifests itself as a 'drift' in the reference direction defined by the spin axis of the rotor. In a free gyroscope, that is, one which measures angular displacements from a given direction, it is customary to describe the performance in terms of an angular drift rate. For a restrained gyroscope, that is, one operating in a nulling or re-balance loop mode to provide a measure of angular rate, any unwanted torques act to produce a 'bias' on the measurement of angular rate.

The terms 'drift' and 'bias' are commonly used interchangeably. In this book, we reserve the term 'drift' for the motion of the spin axis in a free gyroscope, whereas 'bias' is used with 'nulled' sensors. In practice, the way in which the errors are quoted often depends on the accuracy band of the sensor rather than whether the gyroscope is used with its spin axis fixed in space or restrained in some way.

The major sources of error which arise in mechanical gyroscopes are itemised overleaf. Further details relating to specific types of gyroscope will be given later

in the chapter where the physical effects which give rise to each type of error are discussed in more detail.

Fixed bias: This refers to the sensor output which is present even in the absence of an applied input rotation. It may be a consequence of a variety of effects, including residual torques from flexible leads within the sensor, spurious magnetic fields and temperature gradients, which produce such biases. The size of the bias is independent of any motion to which the gyroscope may be subjected and is sometimes referred to as the acceleration (or g)-independent bias. It is usually expressed in units of degrees per hour ($^{\circ}/\text{h}$), or for the less accurate sensors in degrees per second ($^{\circ}/\text{s}$).

Acceleration-dependent bias (g -dependent bias): Biases which are proportional to the magnitude of the applied acceleration. Such errors arise in spinning mass gyroscopes as a result of mass unbalance in the rotor suspension, that is, non-coincidence of the rotor centre of gravity and the centre of the suspension mechanism. The relationship between these components of bias and the applied acceleration can be expressed by means of coefficients having units of $^{\circ}/\text{h}/\text{g}$. In general, such terms relate accelerations in each of the principal axes of the gyroscope, that is, accelerations which act both along and orthogonal to the sensitive axis of the sensor, to errors in the measurement of turn rate. In the presence of a steady acceleration, a fixed bias in the measured rate occurs.

Anisoelastic bias (g^2 -dependent bias): Biases which are proportional to the product of acceleration along orthogonal pairs of axes. Such biases arise in spinning mass gyroscopes because the gyroscope rotor suspension structure, particularly the bearings, has finite compliances which are unequal in different directions. The anisoelastic coefficients have units of $^{\circ}/\text{h}/\text{g}^2$.

Anisoinertia errors: Such errors arise in spinning mass gyroscopes and introduce biases owing to inequalities in gyroscope moments of inertia about different axes. Anisoinertia is frequency sensitive if the rotor is driven by a hysteresis motor. This is a consequence of the elastic coupling between the magnetic ring on the rotor and the rotating magnetic field. The resulting biases are proportional to the product of angular rates applied about pairs of orthogonal axes. The anisoinertia coefficients may be expressed in units of $^{\circ}/\text{h}/(\text{rad}/\text{s})^2$.

Scale-factor errors: Errors in the ratio relating the change in the output signal to a change in the input rate which is to be measured. Scale-factor error is commonly expressed as a ratio of output error to input rate, in parts per million (ppm), or as a percentage figure for the lower performance class of sensor. Additional errors arise as a result of scale-factor non-linearity and scale-factor asymmetry. Scale-factor non-linearity refers to the systematic deviations from the least-squares straight line or non-linear function fitted to the measurements, which relates the output signal to the applied angular rate. The latter term includes differences in the magnitude of the output signal for equal rotations of the sensor in opposite directions. In spinning mass gyroscopes, scale-factor non-linearity relates to thermal changes that result in changes of the magnetic flux.

Cross-coupling errors: Erroneous gyroscope outputs resulting from gyroscope sensitivity to turn rates about axes normal to the input axis. Such errors arise through non-orthogonality of the sensor axes and may also be expressed as parts per million or a percentage of the applied angular rate.

Angular acceleration sensitivity: This error is also known as the gyroscopic inertial error. All mechanical gyroscopes are sensitive to angular acceleration owing to the inertia of the rotor. Such errors become important in wide bandwidth applications. This error increases with increasing frequency of input motion. Hence, it is necessary to compensate for this error if accuracy is to be preserved. A detailed analysis is given by Edwards in Reference 4 for both the rate-integrating gyroscope and the dynamically tuned gyroscope, which are described in Sections 4.2.5 and 4.2.6.

It is important to realise that each of the errors described will, in general, include some or all of the following components:

- fixed or repeatable terms;
- temperature induced variations;
- switch-on to switch-on variations;
- in-run variations.

For instance, the measurement of angular rate provided by a gyroscope will include:

- (i) a bias component which is predictable and is present each time the sensor is switched on and can therefore be corrected;
- (ii) a temperature-dependent bias component which can be corrected with suitable calibration;
- (iii) a random bias which varies from gyroscope switch-on to switch-on but is constant for any one run;
- (iv) an in-run random bias which varies throughout a run; the precise form of this error varies from one type of sensor to another.

The fixed components of error, and to a large extent the temperature induced variations, can be corrected to leave residual errors attributable to switch-on to switch-on variation and in-run effects, that is, the random effects caused by instabilities within the gyroscope. Assuming that the systematic errors are compensated, it is mainly the switch-on to switch-on and in-run variations which influence the performance of the inertial system in which the sensors are installed. Compensation techniques are discussed further in Chapter 8.

A number of different types of mechanical gyroscope of interest in strapdown applications are now described.

4.2.5 Rate-integrating gyroscope

4.2.5.1 Introductory remarks

The design of this type of mechanical gyroscope was conceived in the late 1950s for use on stabilised platforms, the early examples appearing at the start of the 1960s.

This basic concept is capable of achieving a wide spectrum of performance from a very small gyroscope that fits into a cylinder of diameter 25 mm (1 in.) and length 50 mm (2 in.). Typically, the drift performance of the miniature versions of this type of sensor is in the 1–10°/h class, although substantially better than 0.01°/h can be achieved with the larger ‘top of the range’ sensors. The smaller sensors are able to measure turn rates typically of the order of 400°/s or better. This type of sensor has found many different applications as a result of this wide spectrum of performance, including navigation systems in aircraft, ships and guided weapons.

4.2.5.2 Detailed description of sensor

A rate-integrating gyroscope has one input axis and so it is known as a single-axis gyroscope. Besides the case, it has three main component parts, as illustrated in Figure 4.6:

- the float, which contains the rotor and its motor. It is supported in precision bearings to allow rotation about an axis perpendicular to the spin axis of the rotor;
- the angle pick-off which senses rotation of the float assembly;
- the torque motor, which is used to apply precise torques to return the float to its ‘null’ position.

These components are sealed into a case and the small gap between the float and the case is filled with a highly viscous liquid. This liquid provides some support for the float in its bearings, thus reducing undesired torques, and, in some very

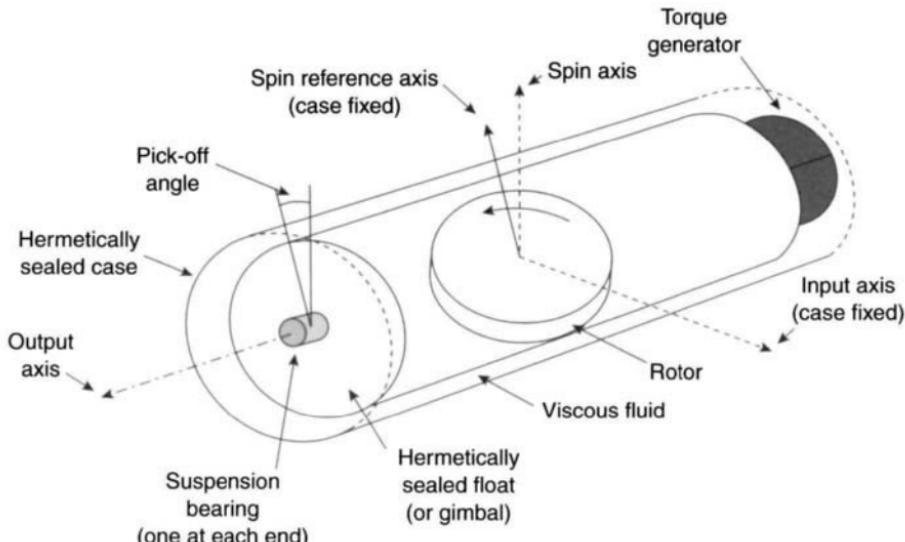


Figure 4.6 Single degree-of-freedom rate-integrating gyroscope

particular instances, it provides total buoyancy. The flotation fluid also provides viscous damping between the float and the case. Electrical signals and power are transmitted between the case and the float via delicate flexible (flex) leads.

When an angular rate is applied about the input axis, the float develops a precessional rate about the output axis shown in the figure. As a result of the damping fluid which supports the float, the output axis rate gives rise to a viscous torque about the output axis. This torque causes the float to precess about the input axis at the input rate and so follow the case rotation. The output axis rate therefore becomes proportional to the input rate. The gyroscope operates in this manner, as a precision rate-integrating gyroscope. In other words, the output which is sensed by the pick-off, is proportional to the integral of the input axis rate, that is, to the change in input angle.

If an additional torque is applied electrically via the torque motor, the pick-off angle rate becomes proportional to the difference between the input rate and the precessional rate induced by the torque motor. Hence, the pick-off angle becomes proportional to the integral of the difference between the input and torque motor rates. For strapdown operation, the pick-off angle is 'nulled' by feeding back the pick-off output to the torque motor. In this situation, the time integral of the difference between the input and torque motor rates becomes zero. It follows that the current applied to the torque motor to maintain the 'null' position is proportional to the applied input rate. This gyroscope is used as a closed loop sensor as this leads to a better definition of the input axis and more accurate measurement of rotation.

Figure 4.7 shows the components of a rate-integrating gyroscope in more detail.

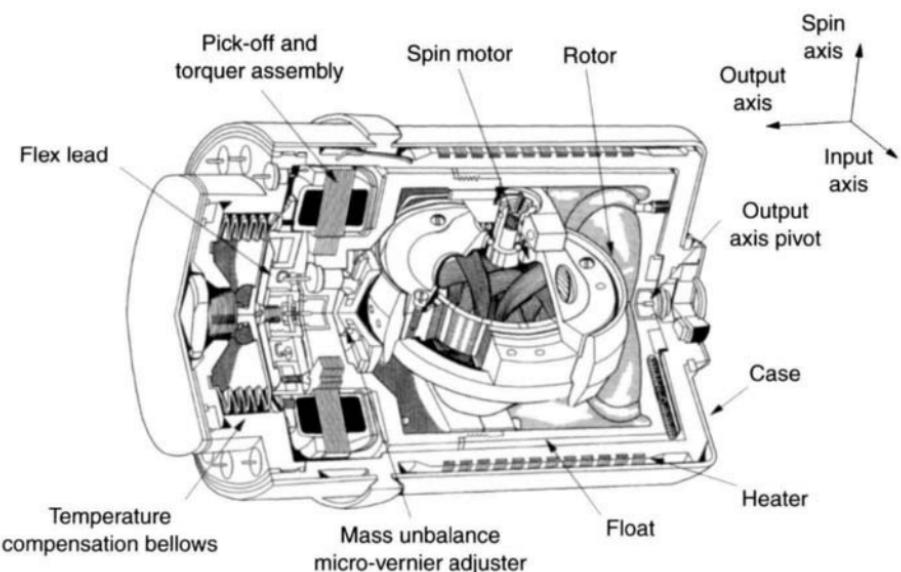


Figure 4.7 Rate-integrating gyroscope

4.2.5.3 Sources of error

The major error processes that influence the performance of this type of gyroscope are shown outlined below:

g-insensitive bias, resulting from a variety of causes which include residual flex lead torques, thermal gradients across the sensor which result in fluid flow around the float assembly and pivot stiction.

g-sensitive bias caused by:

- mass unbalance of the float relative to the pivots of the gimbal along the spin motor axis – principally the result of rotor movement along the spin axis caused by spin motor bearing compliance.
- mass unbalance of the float along the input axis.

anisoelastic bias, which results from unequal compliance of the gyroscope's float assembly along the input and spin axes.

scale-factor error, caused by imperfections and temperature fluctuations in the pick-off and nulling components, which may be expressed as the sum of a 'fixed' error and a set of non-linear components.

cross-coupling, which arises through imperfections in the construction of the sensor.

zero-mean random bias, caused by instabilities in the gyroscope which have short correlation times, variations in pivot friction and random movements of the rotor along the spin axis, for instance.

This sensor is intended to measure angular rates, but unfortunately it is also sensitive to linear and angular accelerations and vibrations and these can give rise to errors in measurements. Careful shielding is required to eliminate errors resulting from stray magnetic fields interacting with the torque generator. Changes in temperature alter the characteristics of the magnetic materials within the sensor. Without at least approximate compensation, these changes in temperature give rise to scale factor errors. Generally, heating effects in conjunction with magnetic imperfections give rise to first, second and third order scale-factor errors. The significant error sources are usually systematic and can be readily corrected.

The angular rate measurement ($\tilde{\omega}_x$) provided by a rate-integrating gyroscope may be expressed in terms of the true input rate and the error terms as follows:

$$\tilde{\omega}_x = (1 + S_x)\omega_x + M_y\omega_y + M_z\omega_z + B_{fx} + B_{gx}a_x + B_{gz}a_z + B_{axz}a_xa_z + n_x \quad (4.8)$$

where ω_x is the turn rate of the gyroscope about its input axis; ω_y and ω_z are the turn rates of the gyroscope about its output and spin axes, respectively; a_x and a_z are the accelerations of the gyroscope along its input and spin axes, respectively. B_{fx} is the *g-insensitive bias*, B_{gx} , B_{gz} are the *g-sensitive bias coefficients*, B_{axz} is the *anisoelastic bias coefficient*, n_x is the *zero-mean random bias*, M_y , M_z are the *cross-coupling coefficients* and S_x is the *scale-factor error* which may be expressed as a polynomial in ω_x to represent scale factor non-linearities.

4.2.5.4 Typical performance characteristics

Typical 1σ values for the major error sources are:

g -Independent bias	0.05–10°/h
g -Dependent/mass unbalance bias	1–10°/h/g
Anisoelastic bias	1–2°/h/g ²
Scale-factor error (uncompensated temperature effects)	up to 400 ppm/°C
Scale-factor non-linearities (at high rotation rates)	0.01–0.1%
Bandwidth	up to 60 Hz
Maximum input rate	up to 400°/s

In certain applications, other systematic error effects may become important, but generally, those given above are dominant.

4.2.6 Dynamically tuned gyroscope

4.2.6.1 Introductory remarks

This sensor is sometimes also called the tuned rotor gyroscope, or dry tuned gyroscope. It has two input axes which are mutually orthogonal and which lie in a plane which is perpendicular to the spin axis of the gyroscope. Work to demonstrate this form of technology was underway at the Royal Aircraft Establishment, Farnborough (now the DSTI and QinetiQ) by Philpot and Mitchell [5] during the late 1940s. Although demonstration of the tuning phenomenon took place in the early 1950s, it is only since the 1970s that this type of gyroscope has been fully developed. The original concept was developed for stabilised platform applications, but has been applied to strapdown systems since the mid- to late 1970s in many types of vehicle.

Generally, the performance of these gyroscopes is very similar to that achieved by the rate-integrating gyroscope. Miniature instruments of this type developed for strapdown applications are typically about 30 mm in diameter and 50 mm in length. Sub-miniature devices have also been produced, with some slight degradation in performance, which are about 20 mm by 25 mm. These gyroscopes have found many applications similar to the floated rate-integrating gyroscope.

4.2.6.2 Detailed description of sensor

The sensor consists of three major sub-assemblies as indicated in Figure 4.8:

- (i) the body block, which consists of the spin motor and angle pick-off arrangement;
- (ii) the rotor assembly, which also includes the torque generator magnets and the Hooke's joint suspension;
- (iii) the case and torque generator coil assembly.

The rotor is connected to the drive shaft by a pair of flexure hinges to an inner gimbal ring. This inner 'gimbal' is also connected to the drive shaft by a pair of

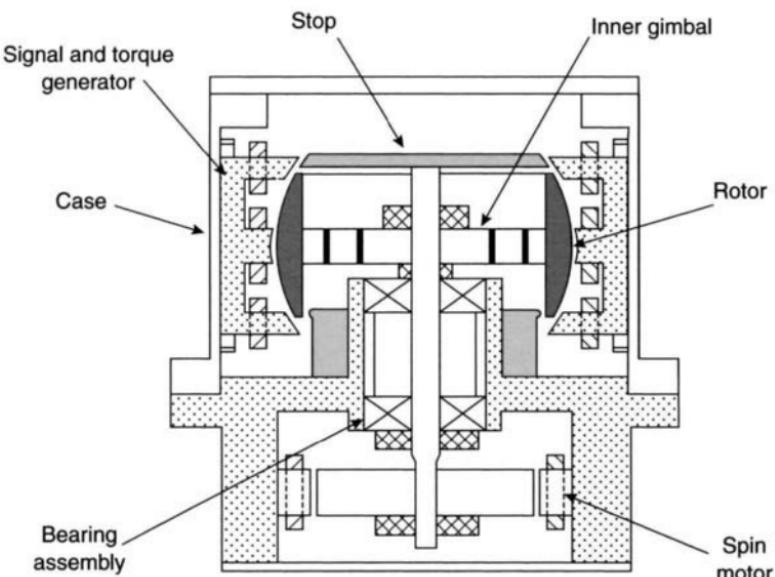


Figure 4.8 Typical tuned rotor gyroscope configuration

flexure hinges, the two axes of freedom being orthogonal as illustrated schematically in Figure 4.9. This is often called a Hooke's joint or a Cardan joint and allows torsional flexibility. This is an internal type of gimbal and is far more compact than the external gimbal shown in Figure 4.1. At the other end of the drive shaft is a synchronous motor.

Rotation of the gimbal causes a reaction at the rotor that is equivalent to a negative torsional spring stiffness. This effect occurs when the angular momentum of the shaft does not coincide with that of the rotor, the angular momentum of the gimbal jumping between that of the shaft and the rotor, at twice the speed of the rotor. Thus careful selection of the torsional stiffness of the gimbal components and the rotational speed of the rotor, allows the rotor suspension to have a net zero spring stiffness at a particular rotor speed, known as the tuned speed. Under these conditions, the rotor is decoupled from the motion of the rest of the sensor and hence is 'free'. In practice, this condition is usually adjusted or trimmed by the use of screws set into the inner gimbal ring that allow minor changes in the mass properties of the gimbal.

Normally, the decoupling of the rotor is not complete or perfect and residual elastic restraints restrict the useful angular range of movement of the rotor. Therefore, the sensor is usually used in a torque re-balance mode allowing only very small deflections of the rotor. Deflections of the rotor are sensed about two orthogonal axes, and are directly proportional to the motion of the gyroscope case about the respective axes in inertial space.

A figure of merit [6] is sometimes used for describing the quality of a dynamically tuned gyroscope. The figure of merit relates the inertias of the rotor to the inertias of

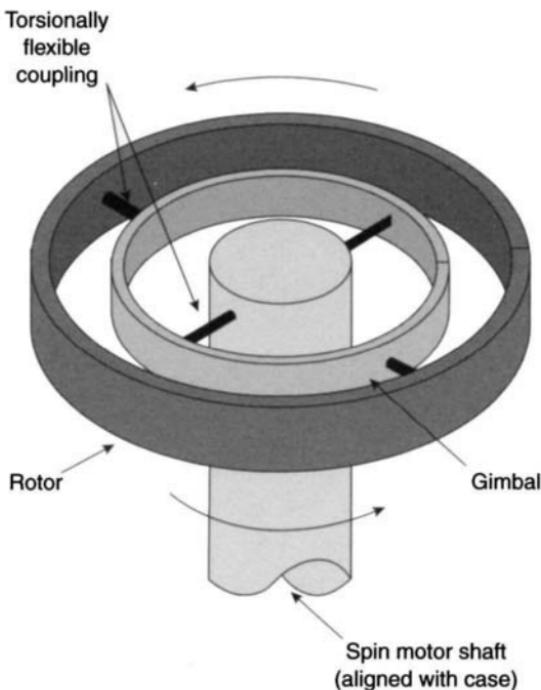


Figure 4.9 Dynamically tuned gyroscope rotor and drive shaft assembly

the gimbal, given as

$$\text{figure of merit} = \frac{C}{I_g + J_g - K_g}$$

where C is the spin inertia of the rotor, I_g , J_g are the gimbal transverse inertias and K_g is the gimbal polar inertia.

Typical values for figure of merit for a moderate performance instrument are in the region of 50.

4.2.6.3 Sources of error

It will be noticed that the errors are of similar form to those given for the rate-integrating gyroscope errors.

g-Insensitive bias, principally the result of stray internally generated magnetic fields which interact with the torque motor magnet mounted on the rotor plus re-balance loop biases. The effects of tuning errors and gimbal damping are often included in this error.

g-Sensitive bias, caused by mass unbalance of the rotor assembly and geometrical imperfections in the torsional elements. The flexures can also generate a torque when loaded axially, leading to an acceleration sensitive bias about the axis opposite

to that axis along which the acceleration is acting. This is known as quadrature mass unbalance.

Anisoelastic bias, which results from unequal compliance of the rotor assembly in the x -, y - and z -directions.

Anisoinertia bias, results from differences in rotor inertias in the x -, y - and z -directions, and is frequency sensitive.

Scale-factor errors, mainly caused by thermally induced changes in magnets and coils used in the re-balance system.

Zero-mean random bias, caused, for example, by error torques resulting from changes in spin motor-shaft orientation owing to variations in the bearing pre-load.

As in the case of the rate-integrating gyroscope, this sensor is sensitive to linear and angular accelerations, vibratory motion, stray magnetic fields and temperature changes, all of which give rise to errors in measurements. This type of sensor is sensitive to vibrations at integer multiples of the spin speed, not only at the spin frequency, as in the single degree of freedom gyroscope, but also vibrations at twice this frequency. Vibration about the input axis interacts with the gimbal angular momentum, and is rectified to give a fixed bias.

The angular rate measurements provided by the sensor ($\tilde{\omega}_x$ and $\tilde{\omega}_y$) may be expressed mathematically as follows:

$$\begin{aligned}\tilde{\omega}_x &= (1 + S_x)\omega_x + M_y\omega_y + M_z\omega_z + B_{fx} + B_{gx}a_x + B_{gy}a_y + B_{axz}a_xa_z + n_x \\ \tilde{\omega}_y &= (1 + S_y)\omega_y + M_x\omega_x + M_z\omega_z + B_{fy} + B_{gy}a_y - B_{gx}a_x + B_{ayz}a_ya_z + n_y\end{aligned}\quad (4.9)$$

where ω_x and ω_y are the turn rates of the gyroscope about its input axes, a_x and a_y are the accelerations along its input axes and a_z is the acceleration along its spin axis. B_{fx} , B_{fy} are the g -insensitive bias coefficients, B_{gx} , B_{gy} are the g -sensitive bias coefficients, B_{axz} , B_{ayz} are the anisoelastic bias coefficients, n_x , n_y represent the zero-mean random bias, S_x , S_y are the scale-factor errors and M_x , M_y , M_z denote the cross-coupling coefficients.

4.2.6.4 Typical performance characteristics

Typical values for the significant error sources and performance parameters are given below:

<i>g</i> -Independent bias	0.05–10°/h
<i>g</i> -Dependent/mass unbalance bias	1.00–10°/h/g
Anisoelastic bias	0.1–0.5°/h/g ²
Scale-factor error (uncompensated temperature effects)	up to 400 ppm/°C
Scale-factor non-linearities (at high rotation rates)	0.01–0.1%
Bandwidth	up to 100 Hz
Maximum input rate	up to 1000°/s

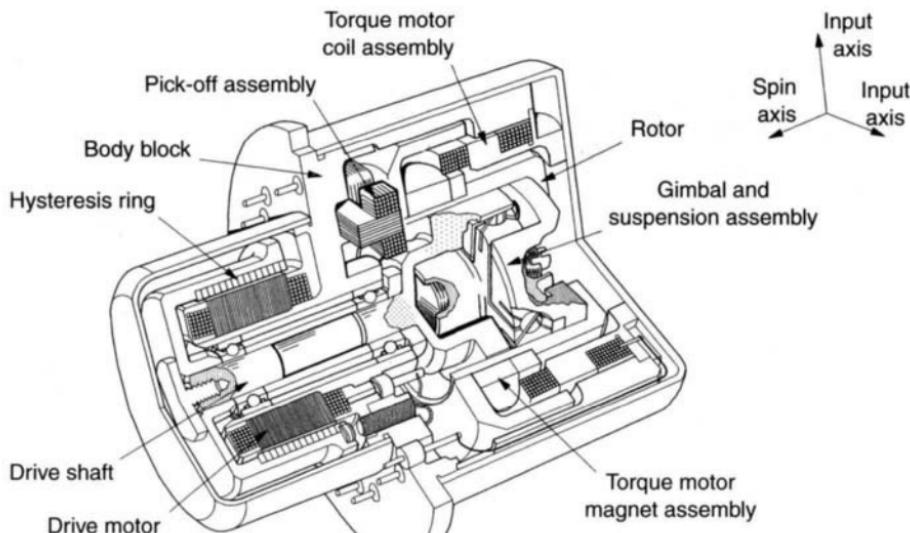


Figure 4.10 Sectional diagram of a dynamically tuned gyroscope

The dynamically tuned gyroscope offers a number of significant advantages for many applications, when compared with the rate-integrating gyroscope. These are usually quoted as fewer parts, a fluid free suspension, no flex lead torques, simplified spin motor bearing design and a fast warm up characteristic. Of course, it offers the ability to measure angular motion about two axes, and additionally, its construction allows the sensor either to be re-worked more easily, or to have its performance optimised before final sealing of the case. One potential drawback is its susceptibility to disturbances and oscillations at the tuned frequency and harmonics of this frequency. Its suspension is analogous to a mass on a spring. For this reason, careful design is required to ensure that mechanical resonances do not interact with the suspension and destroy it. For reliable performance in a harsh environment, careful design of the suspension and mounting is crucial. The rate-integrating gyroscope is generally more resilient in this type of environment owing to its inherently rugged design. Figure 4.10 shows a typical arrangement of the various components of a dynamically tuned gyroscope.

Miniature instruments of this type developed for strapdown applications are typically about 30 mm in diameter and 50 mm in length. Sub-miniature devices have also been produced which are about 20 mm by 25 mm in diameter (see Figure 4.11).

4.2.7 Flex gyroscope

4.2.7.1 Introductory remarks

This sensor bears a close resemblance to the dynamically tuned gyroscope and operates in a similar manner, as the rotor acts as a free inertial element. It also has two



Figure 4.11 Photograph of a modern dynamically tuned gyroscope (published courtesy of Northrop Grumman Corporation, Litton Systems)

sensitive input axes. Development of this inertial instrument has progressed dramatically since the mid-1970s. The form of construction allows a very small instrument to be made, typically about 20 mm in diameter and 30 mm in length. These sensors have found many applications in aerospace and industrial applications.

4.2.7.2 Detailed description of sensor

The major difference in construction between the flex gyroscope and the dynamically tuned gyroscope is that the flex device does not have a Hooke's joint type of flexure pivot arrangement but has a flexible pivot where the drive shaft is reduced in diameter, as shown in Figure 4.12.

The rotor is attached to the main shaft usually using a spider and strut arrangement. Flexible joint torques arising from this form of suspension are compensated by small permanent magnets attached to the rim of the rotor which attract a set of high permeability screws mounted on a plate attached to the shaft. This use of magnetic forces to balance the flex pivot torques has the effect of decoupling the rotor from the drive shaft, as depicted in Figure 4.12. Generally, magnetic shielding is crucial with this sensor to ensure effective decoupling of the rotor. A schematic diagram of such a sensor is given in Figure 4.13.

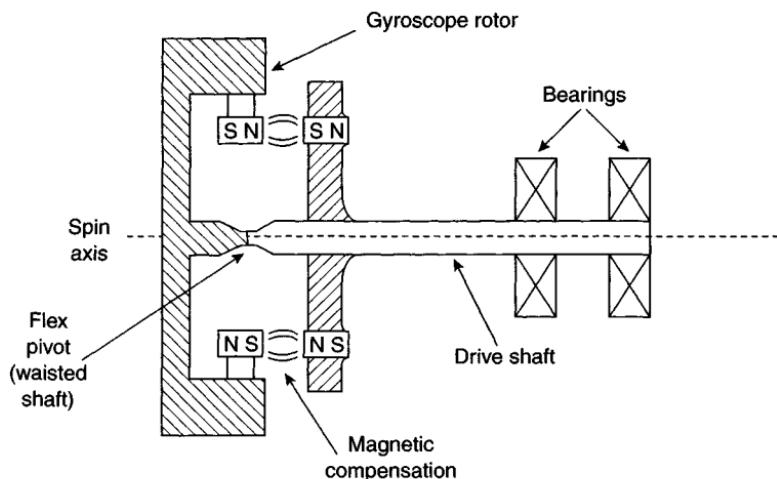


Figure 4.12 Shaft assembly of a flex gyroscope

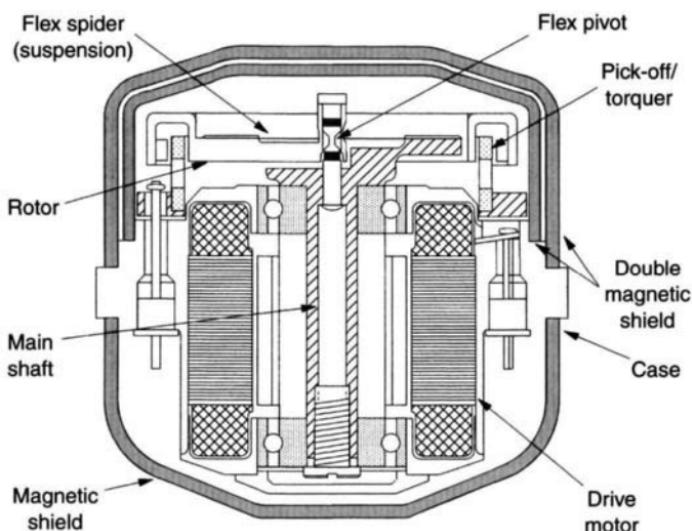


Figure 4.13 Flex gyroscope

4.2.7.3 Sources of error

The error mechanisms associated with this sensor are very similar to the dynamically tuned gyroscope described above and will not be repeated here. The outputs may be expressed mathematically in the same form as those for the dynamically tuned gyroscope in eqn. (4.9). Use of magnetic tuning allows the option of running the rotor at different speeds to fulfil different needs or applications. Additionally, this type of suspension gives very good resilience to vibratory inputs.

4.2.7.4 Typical performance characteristics

Typical values for the significant error sources and performance parameters are given below:

<i>g</i> -Independent bias	1–50°/h
<i>g</i> -Dependent/mass unbalance bias	1–10°/h/ <i>g</i>
Anisoelastic bias	0.05–0.25°/h/ <i>g</i> ²
Scale-factor error (uncompensated temperature effects)	up to 400 ppm/°C
Scale-factor non-linearities (at high rotation rates)	0.01–0.1%
Bandwidth	up to 100 Hz
Maximum input rate	>500°/s

It can be seen that the error parameters are very similar to those quoted for the dynamically tuned gyroscope in Section 4.2.6.4. Typically, the drift performance of such a device is in the range 1–50°/h with the capability to capture rotation rates up to at least 500°/s. Additionally, the anisoelasticity is often slightly smaller, typically by a factor of 2–5.

4.3 Rate sensors

There is a class of mechanical sensors designed to sense angular rate using various physical phenomena which are suitable for use in some strapdown applications. Such devices resemble conventional gyroscopes in that they make use of the principles of gyroscopic inertia and precession described in Section 4.2.2. They are suitable for some lower accuracy strapdown applications, particularly those that do not require navigational data, but stabilisation. These devices tend to be rugged and to be capable of measuring rotation rates up to about 500°/s with typical drift accuracies of a few hundred degrees per hour. A number of devices of this type are discussed in the following sections.

4.3.1 Dual-axis rate transducer (DART)

4.3.1.1 Introductory remarks

Development of this type of gyroscope started in the United States during the 1960s. It has, as its name implies, the ability to sense angular rate about two orthogonal axes. Its basic performance is certainly sub-inertial, typically having a drift in the region of 0.5°/s or less. Its size is somewhat smaller than the rate-integrating gyroscope being about 18 mm in diameter and 40 mm in length.

4.3.1.2 Detailed description of sensor

The inertial element in this form of gyroscope is a sphere of heavy liquid, such as mercury, contained in a spherical cavity. This cavity is rotated at high speed about an axis along the case in order to give high angular momentum to the fluid sphere. There

4.2.7.4 Typical performance characteristics

Typical values for the significant error sources and performance parameters are given below:

<i>g</i> -Independent bias	1–50°/h
<i>g</i> -Dependent/mass unbalance bias	1–10°/h/ <i>g</i>
Anisoelastic bias	0.05–0.25°/h/ <i>g</i> ²
Scale-factor error (uncompensated temperature effects)	up to 400 ppm/°C
Scale-factor non-linearities (at high rotation rates)	0.01–0.1%
Bandwidth	up to 100 Hz
Maximum input rate	>500°/s

It can be seen that the error parameters are very similar to those quoted for the dynamically tuned gyroscope in Section 4.2.6.4. Typically, the drift performance of such a device is in the range 1–50°/h with the capability to capture rotation rates up to at least 500°/s. Additionally, the anisoelasticity is often slightly smaller, typically by a factor of 2–5.

4.3 Rate sensors

There is a class of mechanical sensors designed to sense angular rate using various physical phenomena which are suitable for use in some strapdown applications. Such devices resemble conventional gyroscopes in that they make use of the principles of gyroscopic inertia and precession described in Section 4.2.2. They are suitable for some lower accuracy strapdown applications, particularly those that do not require navigational data, but stabilisation. These devices tend to be rugged and to be capable of measuring rotation rates up to about 500°/s with typical drift accuracies of a few hundred degrees per hour. A number of devices of this type are discussed in the following sections.

4.3.1 Dual-axis rate transducer (DART)

4.3.1.1 Introductory remarks

Development of this type of gyroscope started in the United States during the 1960s. It has, as its name implies, the ability to sense angular rate about two orthogonal axes. Its basic performance is certainly sub-inertial, typically having a drift in the region of 0.5°/s or less. Its size is somewhat smaller than the rate-integrating gyroscope being about 18 mm in diameter and 40 mm in length.

4.3.1.2 Detailed description of sensor

The inertial element in this form of gyroscope is a sphere of heavy liquid, such as mercury, contained in a spherical cavity. This cavity is rotated at high speed about an axis along the case in order to give high angular momentum to the fluid sphere. There

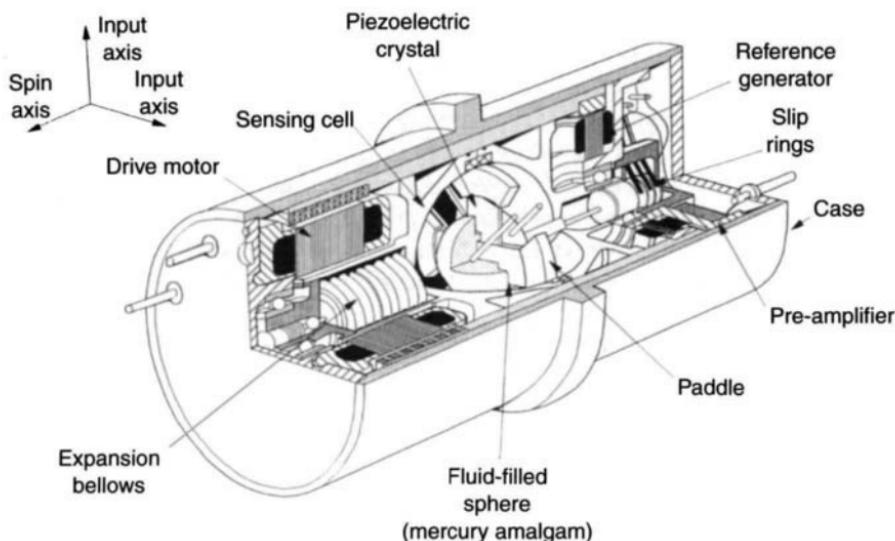


Figure 4.14 Dual-axis rate transducer

is an assembly of paddles, rigidly mounted to the inside of this spherical cavity. These paddles have piezoelectric crystals attached to them as shown in Figure 4.14. The instrument is sensitive to angular rates of the case about two orthogonal axes normal to the spin axis.

A simplified explanation of the operation of the complex dynamical interaction of this sensor is as follows. As the case of the sensor is rotated about either of its two sensitive axes, the spin axis of the mercury tends to lag behind that of the spherical cavity which moves with the rotation of the case. As a result of viscous coupling, a torque is applied to the rotating sphere of fluid in such a way as to make it precess at the input rate. This motion of the fluid causes a deflection of the paddles within the spherical cavity, bending the piezoelectric crystals and generating an a.c. electric signal which is proportional to the applied angular rate. The phase of this signal relative to the reference generator on the rotor shaft gives the axis of the applied rate.

4.3.1.3 Typical performance characteristics

Typical values for the significant error sources and performance parameters are as follows:

g -Independent bias including temperature effects	0.1–0.4°/s
g -Dependent bias	0.03–0.05°/s/g
g^2 -Dependent bias	~0.005°/s/g ²
Scale-factor temperature sensitivity over operating temperature	~5%
Scale-factor non-linearity	~0.5% of maximum rate
Bandwidth	>80 Hz
Maximum input rate	up to 800°/s

This form of sensor is very rugged owing to the form of its fabrication. Its error processes tend to be similar to those of the dynamically tuned gyroscope. The temperature sensitivity is quite a complex function and can be difficult to correct exactly. In general, accuracy is usually somewhat less than that of the rate-integrating and dynamically tuned gyroscopes, so it is not usually used for inertial navigation applications. However, it does have many applications such as seeker stabilisation and provision of signals for autopilot feedback.

Derivatives of this sensor have also been produced which do not use any liquid within the sphere. The accuracy of such devices is somewhat less than the mercury filled device.

4.3.2 Magnetohydrodynamic sensor

4.3.2.1 Introductory remarks

The development of this dual-axis rate sensor also has its origins in the United States and has taken place in parallel with the development of the dual-axis rate transducer described earlier. It is of similar size to the dual-axis rate transducer and has comparable performance capability, the g -insensitive bias being in the region of $0.05\text{--}0.5^\circ/\text{s}$.

4.3.2.2 Detailed description of sensor

This device does not rely upon the angular momentum of a spinning mass, but uses a rotating angular accelerometer to sense angular rates about two mutually perpendicular axes of the sensor. The rotating angular accelerometer acts as an integrator and provides an electrical signal directly proportional to the applied angular rate.

The sensor consists of the angular accelerometer and a synchronous motor, as illustrated in Figure 4.15. A slip ring assembly is required to access the electrical signals produced by the rotation of the angular accelerometer. The case is usually a high permeability alloy that provides the necessary magnetic shielding.

The principle by which the sensor operates is as follows. When an angular accelerometer is rotated at a constant rate (ω_a) about an axis perpendicular to its sensitive axis, and a steady rotation rate (ω_i) is applied about an axis perpendicular to this axis of rotation, then the instantaneous angular rate (ω_o) about the input axis of the angular accelerometer is given by:

$$\omega_o = \omega_i \sin \omega_a t$$

Hence, the angular acceleration is:

$$\dot{\omega}_o = \omega_a \omega_i \cos \omega_a t$$

Consequently, the input rate is changed to a time-varying angular acceleration. The rotating angular accelerometer produces an alternating signal, the amplitude of which is directly proportional to the applied angular rate whilst the frequency is equal to the rate of turn of the angular accelerometer. This output signal can be resolved to give the applied angular rate about two orthogonal axes, both of which are mutually perpendicular to the axis of rotation of the angular accelerometer.

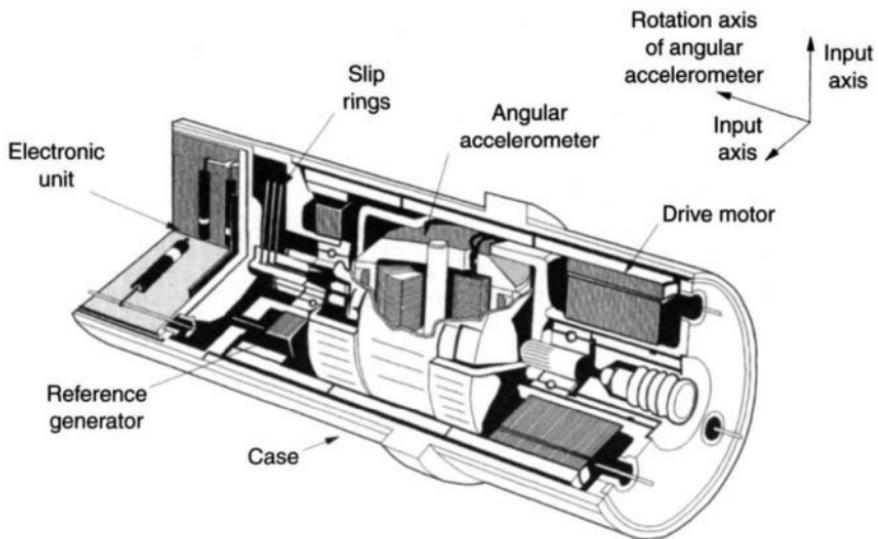


Figure 4.15 Magnetohydrodynamic sensor

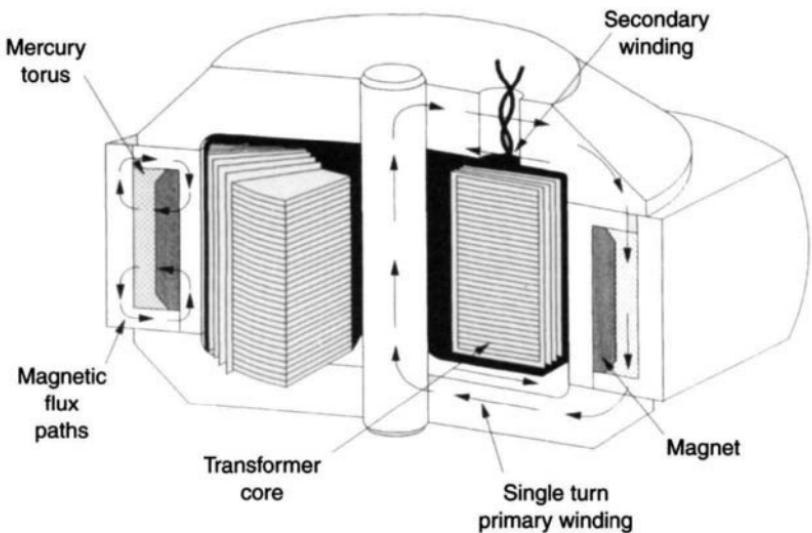


Figure 4.16 Magnetohydrodynamic active element

A diagrammatic representation of the angular accelerometer arrangement within the sensor is shown in Figure 4.16.

The angular accelerometer usually has an annular ring of mercury between the radially oriented permanent magnet and the magnetic case, which provides a path

for the magnetic field. The presence of an input rate results in relative motion of the magnetic field with respect to the torus of mercury. As a result of the magnetodynamic effect, the motion of the magnetic field produces a voltage gradient across the mercury, mutually at right angles to the motion and the magnetic field. The presence of the transformer windings, as shown in Figure 4.16, results in a voltage appearing in the secondary winding.

4.3.2.3 Typical performance characteristics

The performance figures for the magnetohydrodynamic (MHD) sensor are typically as shown below:

<i>g</i> -Independent bias including temperature effects	0.05–0.5°/s
<i>g</i> -Dependent bias	~0.05°/s/g
<i>g</i> ² -Dependent bias	~0.001°/s/g ²
Scale-factor temperature sensitivity over operating temperature	~4%
Scale-factor non-linearity (at maximum rotation rate)	~0.1% maximum rate
Bandwidth	100 Hz
Maximum input rate	up to 400°/s

This sensor is very rugged and capable of surviving in very harsh environments. The performance capability appears to be that of a good rate sensor and is particularly suited for stabilisation applications. The error equation used to define performance may be expressed in a form similar to that used for the conventional mechanical gyroscopes as discussed in Section 4.2.

4.4 Vibratory gyroscopes

4.4.1 Introduction

The origins of this type of gyroscope may possibly be considered to be in the middle of the nineteenth century. Foucault demonstrated that a vibrating rod would maintain its plane of vibration whilst it was being rotated in a lathe. Later that century, Bryan [7] demonstrated that angular rate sensing, as well as linear acceleration sensing, could be achieved using this principle.

It was during the 1950s that work started to develop this principle of a vibrating element to sense angular rate, the majority of the effort being in the United States. The vibrating element has taken various forms such as a string, a hollow cylinder, a rod, a tuning fork, a beam and a hemispherical dome. This form of gyroscope occurs in nature as laterae in flying insects. One of the earliest forms of gyroscope using a vibrating element was produced by the Sperry Gyroscope Company. It was based on the tuning fork principle and was known as the gyrotron.

The basic principle of operation of such sensors is that the vibratory motion of part of the instrument creates an oscillatory linear velocity. If the sensor is rotated about an axis orthogonal to this velocity, a Coriolis acceleration is induced. This acceleration modifies the motion of the vibrating element and provided that this can be detected, it will indicate the magnitude of the applied rotation.

The most common design technology for these sensors has generally used a stable quartz resonator with piezoelectric driver circuits. Some designs have produced sensors with small biases, in the region of $0.01^\circ/\text{h}$. However, the smaller sensors have tended to produce biases in the region of $0.1\text{--}1^\circ/\text{s}$. Typical limitations for this type of technology for use in inertial navigation systems have been high drift rates, resonator time constants and sensitivity to environmental effects, particularly temperature changes and vibratory motion. However, these sensors can be made to be extremely rugged, including the capability of withstanding applied accelerations of many tens of thousands of ' g '.

These sensors are usually quite small, usually with a diameter of somewhat less than 15 mm and a length of about 25 mm. Others are significantly smaller than this and are packaged in rectangular cases. These sensors have been used in many applications, particularly to provide feedback for stabilisation or angular position measurement tasks.

As there are many similarities in the performance characteristics of vibratory gyroscopes, such aspects are covered in a single section following general descriptions of the operating principles for different types of design.

4.4.2 *Vibrating wine glass sensor*

This sensor is synonymous with the vibrating cylinder and vibrating dome gyroscopes. These devices usually have three basic components inside a sealed case:

- (i) A resonant body in the shape of a hemisphere or a cylinder with a high Q factor to maintain a stable resonance. It is often made from ceramic, quartz or steel.
- (ii) A forcing or driving mechanism, commonly made from a piezoelectric material.
- (iii) A pick-off device to sense the modified motion, also usually a piezoelectric device.

These components are shown schematically in Figure 4.17a together with the resonant vibration patterns in the static and rotating cases in Figure 4.17b.

The resonant body, usually a hemisphere or cylinder, is forced to vibrate at its resonant frequency by four equally spaced piezoelectric 'driving crystals' that are firmly attached to its circumference. One 'opposite' pair of crystals is driven with an oscillatory signal to distort the resonant body so that modes appear in the distortion pattern on its circumference. The other pair of crystals are used as feedback sensors to control the nodes in the induced motion. When the cylinder is stationary, the nodes in the vibratory motion are positioned exactly between the driving crystals, the anti-node axes A and B being shown in the figure. If the resonant body is rotated at an angular rate about an axis orthogonal to the plane containing the vibratory motion of this body, the pattern of vibration is modified by the Coriolis acceleration.

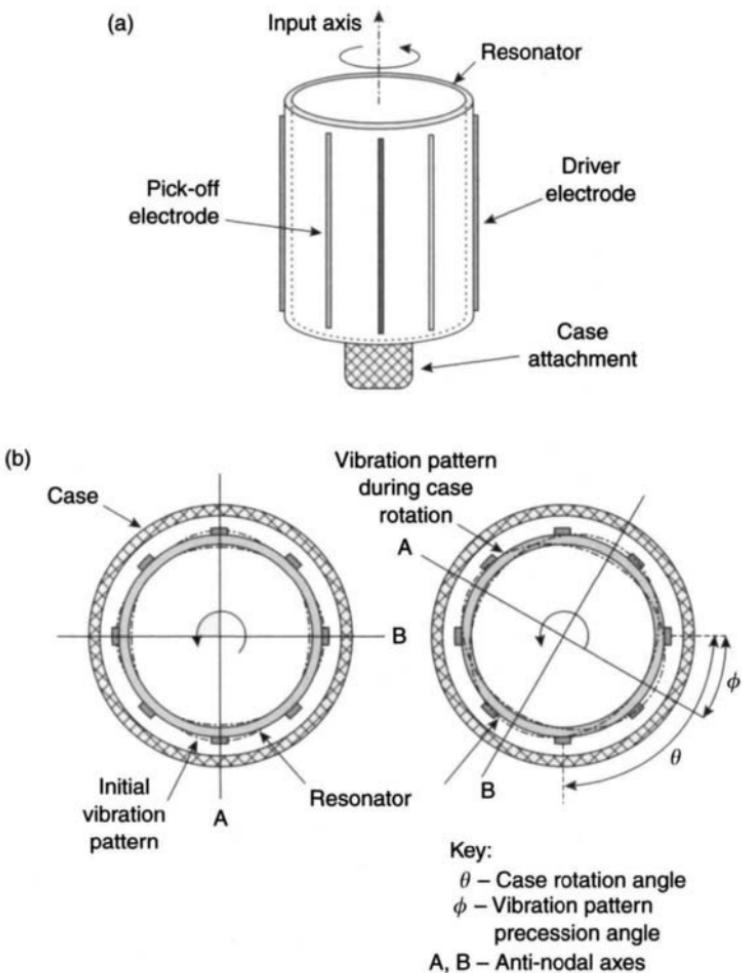


Figure 4.17 (a) Vibratory gyroscope schematic diagram. (b) Resonator vibration patterns

The effect is to add a tangential force to the vibratory force along the diameter of the resonant body. Consequently, there is a change in the motion at the points mid-way between the 'driving crystals' as the vibration pattern has moved through an angle ϕ relative to the case. Hence, the pick-off transducer crystals now sense movement of the resonant body, the amplitude of displacement being directly proportional to the applied rotation rate.

By demodulating the signal from the pick-off transducers, with respect to the waveform used to power the driving crystals that vibrate the cylinder, a d.c. signal is produced. Its magnitude is proportional to the applied rotation rate and its sign indicates the sense of rotation. The second pair of piezoelectric crystals, which are nominally at the nodal positions, can be used to modify the vibration characteristics of

the cylinder in order to enhance the bandwidth of this sensor. These crystals are driven by a feedback signal derived from the signal produced by the pick-off transducers.

An alternative configuration of this form of sensor is to fabricate the resonant body from a ceramic material and then deposit metal electrodes on to the ceramic. This design has some advantages in terms of reliability as the wires can be attached at points of zero movement. An alternative method of making the cylinder vibrate is to use a magnetically driven ferromagnetic cylinder. Capacitive pick-offs can be used thus reducing the damping of the resonance produced by attaching leads to the resonator.

It is very important that the vibrating shape, such as the shell of the hemispherical resonator gyroscope, is machined to have a wall thickness that is as uniform as possible and that it is then dynamically balanced to compensate for material inhomogeneities and machining errors. A non-uniform shell is not sensitive to small rotations as the nodes do not move when the case is rotated at low rates about the input axis.

Sensors of this type can be very rugged and have been demonstrated to withstand accelerations or shocks well in excess of 20,000g. Additionally, these devices can be activated very rapidly, but great care is required in the choice of resonant material to achieve a form of temperature sensitivity that does not mask its rate sensitivity. This form of device generally does not show any significant acceleration sensitivity as such a response only results from deformation of the resonant body.

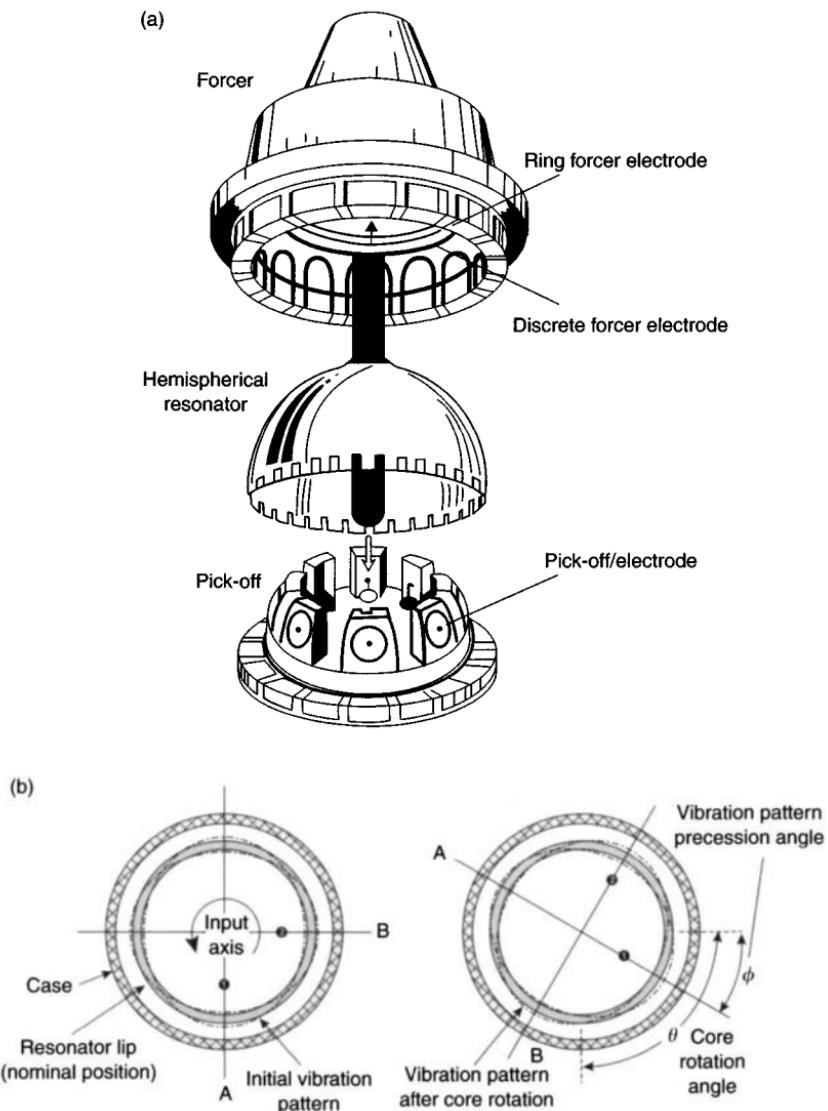
Vibrating wine glass sensors can be operated as either open-loop or closed loop devices. In the open-loop configuration, the electric signal merely increases as the angular rate increases. When used in a closed loop configuration, the second set of crystals are used to null any displacement sensed by the pick-off crystals, this secondary drive signal being proportional to the detected rate. This latter technique leads to a far more linear relationship between the output signal and the input stimulus.

4.4.3 Hemispherical resonator gyroscope

Whilst most vibrating wine glass gyroscopes sensors are relatively low accuracy devices having biases in the region of 0.1–1°/s, an exception to this is the hemispherical resonator gyroscope (HRG). An inertial grade HRG was developed originally by Delco (now the Northrop Grumman Corporation, Litton Systems) in the 1980s, primarily for space applications, incorporating a 58 mm resonator. This sensor has a bias stability in the region of 0.01°/h. In addition, the device is characterised by its excellent scale-factor accuracy (less than 1 ppm uncertainty) and low random walk ($0.0008^{\circ}/\sqrt{h}$) [8].

The base material used in the construction of the HRG is quartz. As with other sensors of this type, it comprises three main elements:

- a quartz forcer to induce and sustain a standing wave in the resonator;
- a quartz resonator with a high Q factor ($\sim 10^7$);
- a quartz pick-off which senses the locations of nodes and anti-nodes in the standing wave pattern of the resonator.



*Figure 4.18 (a) Major components of a hemispherical resonator gyroscope.
 (b) Principle of operation
 (Published courtesy of Northrop Grumman Corporation, Litton Systems)*

The HRG uses a capacitive electrostatic charge between metal-coated surfaces on the quartz components to sustain the standing wave and to sense its position. An exploded view of such a device is shown in Figure 4.18a.

Figure 4.18b illustrates the operating principle of the HRG. When the vibrating element is stationary, the nodes in the vibratory motion are positioned exactly between

the driving crystals, the anti-node axes A and B being shown in the figure. If the resonant body is rotated at an angular rate about an axis orthogonal to the plane containing the vibratory motion of this body, the pattern of vibration is modified by the Coriolis acceleration. The effect is to add a tangential force to the vibratory force along the diameter of the resonant body. Consequently, there is a change in the motion at the points mid-way between the driving crystals as the vibration pattern has moved through an angle ϕ relative to the case. Hence, the pick-off transducer crystals now sense movement of the resonant body, the amplitude of displacement being directly proportional to the applied rotation rate.

In recent years, attempts have been made to develop a scaled-down version of the HRG with bore hole survey applications in mind, where measurement systems designed to operate during the drilling process are frequently subjected to high levels of mechanical shock and vibration. This development set out to capitalise on the projected reliability of this type of sensor, and its expected resistance to shock and vibration. Thus far, this work has failed to yield a sensor of the required performance, at an acceptable cost. Work is continuing in Russia at the present time based on an HRG containing a 30 mm diameter resonator [9, 10].

4.4.4 Vibrating disc sensor

An alternative configuration has been developed by British Aerospace based on a planar metal disc [11]. The resonator is formed from a metal alloy disc, which is machined to form a ring that is supported by rigid spokes, as shown in Figure 4.19.

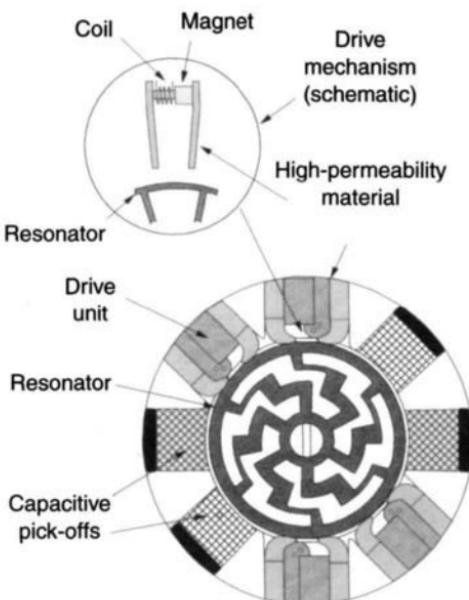


Figure 4.19 A section through a vibrating disc gyroscope

This ring is forced into resonant sinusoidal oscillation in the plane of the ring, using an alternating magnetic field, creating distortions in the shape of the ring. This motion of the ring is detected using capacitive techniques to measure the distance between a fixed plate and the edge of the ring.

The operation of this type of gyroscope is identical to the sensor described in the previous section. The vibration pattern remains fixed with respect to the ring whilst the sensor is stationary. However, the position of the nodes and anti-nodes of this vibratory motion are displaced through an angle when the sensor is rotated about an axis perpendicular to the plane containing the resonator. The magnitude of the angular displacement of the vibration pattern is proportional to the applied angular rate, and is measured using the capacitive pick-offs arranged around the edge of the resonator.

It has been suggested that the performance of this configuration is superior to the performance produced by the resonant cylinder sensor owing to the improved stability properties of the metal alloy used.

4.4.5 Tuning fork sensor

This form of device is very similar to the wine glass sensor described earlier. The sensing element is two vibrational structures mounted in parallel on a single base, each structure having a mass positioned at the end of a flexible beam. When the two structures are excited to vibrate in opposition, the effect is analogous to the motion of the tines of a tuning fork. When rotated about an axis parallel to the length of the beams, the effect of the Coriolis acceleration is to produce a torque couple about this input axis. The torque is oscillatory and is in phase with the tine mass velocity. The amplitude of the oscillation is proportional to the applied rate.

A schematic diagram showing the principle of operation of the tuning fork sensor is given in Figure 4.20.

Two specific problems delayed the development of this type of sensor:

- (i) variation in the bending and torsional elastic moduli of materials with temperature;
- (ii) bias instabilities caused by the lateral displacement of the tine mass centres.

Use of crystalline quartz tine forks has alleviated many of these problems.

A typical implementation of this form of technology is to use a pair of piezoelectric vibrating beams, each pair consisting of two piezoelectric 'bender' elements mounted end to end. The element that is firmly attached to the base is driven resonantly so that the second element swings but does not bend. This element senses the angular motion. When there is angular motion about the sensitive axis of this 'tuning fork', there is a momentum transfer to the perpendicular plane as a result of the Coriolis acceleration. This sensing element now bends as a consequence of this momentum transfer and an electrical signal is produced that is proportional to the applied angular rate.

4.4.6 Quartz rate sensor

The quartz rate sensor (QRS) is a direct application of the tuning fork principle. It is a single degree of freedom, open-loop, solid-state sensor. In this device, quartz is

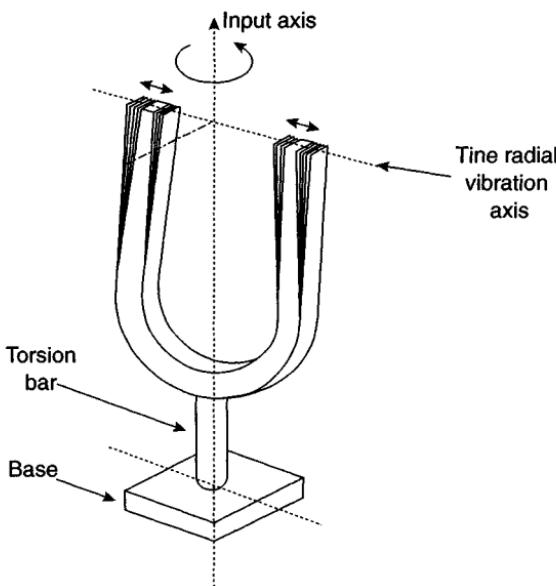


Figure 4.20 Principle of a tuning fork sensor

formed into an 'H' fork configuration, where one pair of tines has an array of electrodes. These tines are driven at their resonant frequency of about 10 kHz.

When the sensor is rotated at a given rate about the input axis, a Coriolis torque is produced, which oscillates in phase with the tine mass velocity. This torque produces a 'walking' motion of the pick-off tines, perpendicular to the vibrating plane of the driven tines. The time-varying displacement of the tines, which is proportional to the applied rate, is detected with a capacitive sensor. It is vital that the mount is strong so that it supports the quartz element, but sufficiently isolated in order to maximise the Coriolis coupling torque into the pick-off tines. Drive and pick-off signals are routed through the mount. A general arrangement of this sensor is shown in Figure 4.21.

This sensor, like other solid-state devices, can have various rate sensitivities and the full scale output can be modified by changing the electronic gain control. These parameters are functions of the signal processor which controls the input range and the signal bandwidth. Additionally, the design of the vibrating fork can have an almost infinite combination of size, thickness and electrode pattern, enabling flexibility of performance to be achieved.

Since its introduction, the performance of the QRS has improved dramatically as a result of enhancements in the design of the sensing element, improvements in the manufacturing process for the sensing element, as well as improvements in the signal processing and calibration techniques used. This has been accompanied by a reduction in the size of the tuning fork element; the sensors are micromachined using photolithographic processes, and are at the forefront of MEMS technology, as discussed in Chapter 7. These advances have resulted in gyroscopic sensors with an in-run bias of $1^\circ/\text{h}$.

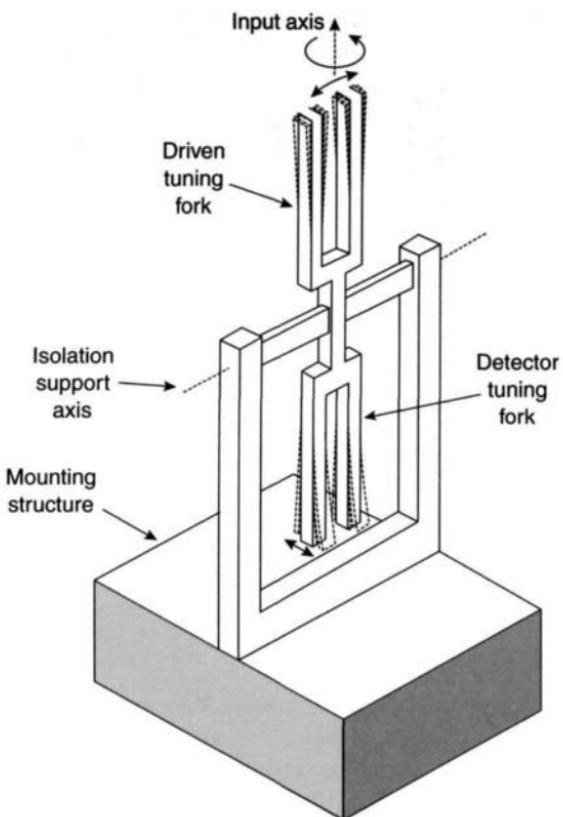


Figure 4.21 Principle of operation of a quartz rate sensor

4.4.7 Silicon sensor

The material silicon has many properties that make it suitable for the fabrication of very small components and intricate monolithic devices. It is inexpensive, very elastic, non-magnetic, it has a high strength to weight ratio and possesses excellent electrical properties allowing component formation from diffusion or surface deposition. Additionally, it can be electrically or chemically etched to very precise tolerances, of the order of micrometres.

A team at The Charles Stark Draper Laboratories, Inc. [12] has used chemical etching techniques to make a very small gyroscope from a wafer of single crystal silicon. The sensor does not have any continuously rotating parts, but part of its structure is vibrated at very high frequency. A schematic representation of this sensor is shown in Figure 4.22.

The sensor comprises a double gimbal structure with a vertical member electroplated with gold mounted on the inner gimbal. The gimbals are each supported by a set of orthogonal 'flexure pivots' as indicated in the figure. These pivots allow each

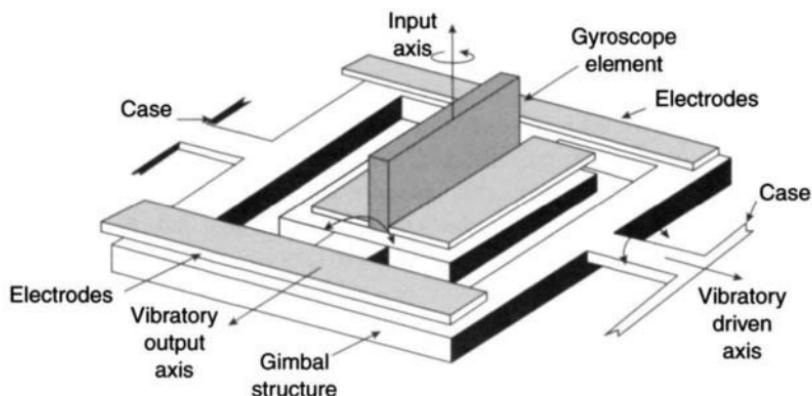


Figure 4.22 Silicon gyroscope

gimbal a small amount of torsional freedom about their respective support axes whilst remaining rigid in other directions. The outer gimbal is forced to oscillate through a small angle by applying an oscillatory electrostatic torque using pairs of electrodes mounted above and below the outer gimbal structure. When the structure is rotated about an axis perpendicular to the plane of the sensor, as indicated in the figure, the inner gimbal also starts to oscillate. The inner gimbal vibrates at the same frequency as the outer gimbal, but with an amplitude proportional to the applied angular rate. This motion is sensed electrostatically by a pair of bridging electrodes. The sensitivity of the device is determined largely by the geometrical arrangement of the structure.

In order to achieve high sensitivity and accuracy, the gyroscope is operated in a closed loop re-balance mode. The inner gimbal is torqued electrostatically in order to maintain it at a 'null' position, the torquer drive signal being proportional to the angular displacement sensed by the electrostatic pick-off, and hence to the applied angular rate. The pick-off and re-balance signals pass through the same electrodes, but use different frequencies. This method of operation allows the gyroscope to tolerate variations in the frequency of the vibratory motion. It also allows the amplitude of the vibratory motion to be increased, without cross-coupling interactions becoming unacceptably large, thus enabling an increase in the signal to noise ratio of the output signal. The electronic loops which control the operation of the gyroscope also allow compensation for imperfections in the device fabrication and changes in temperature to be applied.

The gyroscope is packaged in a sealed case to maintain a vacuum. This enables a high Q (quality factor) to be achieved in the resonant structure, enhancing further the sensitivity of the device. It is anticipated that this type of device should be capable of achieving an in-run drift performance of better than $100^\circ/\text{h}$, and be capable of measuring very high rotation rates. The device can be substantially less than 1 mm long and can be used with silicon accelerometers (discussed in Section 6.4.3) to make a very small inertial measurement unit.

4.4.7.1 Silicon vibrating disc sensor

It is possible that the resonator for the vibrating disc gyroscope outlined in Section 4.4.4 could also be manufactured from silicon. This material possesses many characteristics that would be ideal for this type of component: low cost, elasticity and high strength to weight ratio for example. However, if silicon were to be used, it would be necessary to use an alternative technique to force the structure to vibrate, such as the use of piezoelectric devices. One clear advantage that silicon could offer would be a size reduction, probably with reduced cost of manufacture.

4.4.8 Vibrating wire rate sensor

This device, also known as the vibrating string gyroscope, has three fundamental components within its case, as shown in Figure 4.23:

1. the vibrating element in the form of a taut conductor;
2. the drive magnet;
3. the signal or pick-off magnet.

The principle of operation is similar to that of Foucault's pendulum. If a wire or string is oscillating in a plane and the supports of the vibrating element turn through an angle, then the plane of vibration remains fixed in space despite the fact that the string rotates with its supports.

An alternating current at a selected drive frequency is applied to the wire between the points A and B, as indicated in Figure 4.23. The interaction between the magnetic field around the wire and that of the drive magnet sets up a standing wave vibration of the wire. A second magnet, the signal magnet, is arranged with its magnetic field at right angles to the drive magnet as indicated in the figure.

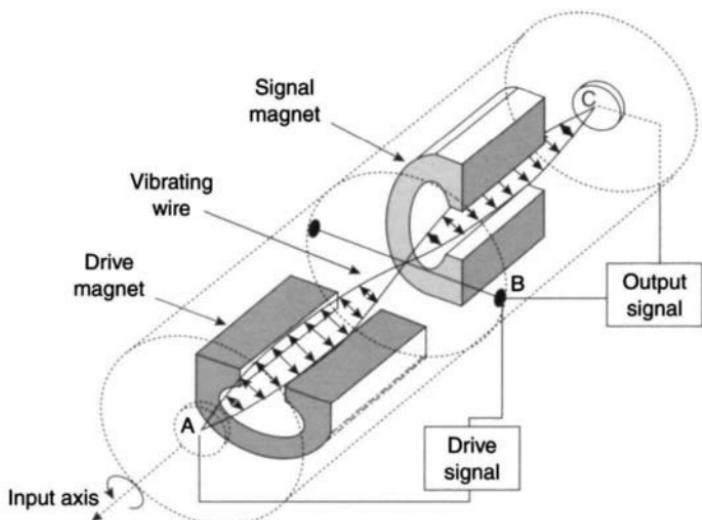


Figure 4.23 Illustration of a vibrating wire rate sensor

Consider now the effects of rotations about an axis that passes between the points A and C. When the gyroscope is not being rotated about this axis, the vibration of the wire between the poles of the signal magnet will not induce any change in the current in the wire. Thus, when the signal emerging from the point C is compared with the applied signal in a suitable demodulator, there will be no resultant signal. In the situation where the device is rotated about the axis AC, a rotation of the signal magnet with respect to the plane of vibration will arise. This causes the signal magnet to modify the current flowing in the wire and thus to modulate the carrier. Comparison of the drive and output signals now yields a resultant signal which is a measure of the applied angular rate about the axis AC.

It is usual to choose the natural frequency of the wire and the drive frequency to be in the region of 20 kHz or more so that the vibrations are well above those that are likely to be produced by environmental vibrations. This prevents synchronous vibration of the sensor's case along the pick-off axis being interpreted as an input rotation.

4.4.9 General characteristics of vibratory sensors

All vibrating sensors tend to have a very short reaction time, that is, rapid start-up capability, and some designs are very rugged. Significant sources of error with these devices are their sensitivity to changes in ambient temperature and the potential for cross talk between similar sensors mounted on the same structure. Careful design can minimise these effects and the errors they introduce into the output signal. These devices are usually termed solid-state sensors and offer good shelf-life and good dormancy characteristics as they do not have bearings, lubricants or any other fluid within their case. Good reliability is possible because of the need for only one bonded joint and the power leads can be connected, with suitable design, at a point which does not move. The form of design also ensures low power consumption.

These types of sensor are subject to biases and scale-factor errors equivalent to those which arise in conventional gyroscopes. Typical performance of the miniature vibratory sensors is not compatible with the requirements of inertial navigation systems, but have much to offer for control and stabilisation processes.

The performance range of such miniature devices is as follows:

<i>g</i> -Independent bias including temperature effects	0.1–1°/s
<i>g</i> -Dependent bias	0.01–0.05°/s/g
Scale-factor temperature sensitivity over operating temperature	0.01–0.05%/ $^{\circ}$ C
Scale-factor non-linearity (at maximum rotation rate)	0.03–0.3%
Bandwidth	60–500 Hz
Shock resistance	>25 000g

In addition, such sensors can be sensitive to vibration although, with careful design, such effects can be minimised. The error equation used to define performance

may be expressed in a form similar to that used for the conventional mechanical gyroscopes, as discussed in Section 4.2.

4.5 Cryogenic devices

4.5.1 Nuclear magnetic resonance gyroscope

Investigation started in the 1960s into the application of the phenomenon of nuclear magnetic resonance (NMR) to the sensing of angular rotation. The NMR gyroscope has many attractions, particularly as it will not have any moving parts. Its performance will be governed by the characteristics of the atomic material and will not demand the ultimate in accuracy from precision engineering techniques. Hence, in theory, it offers the prospect of a gyroscope with no limit to either its dynamic range or linearity and therefore, potentially, an ideal sensor for use in strapdown inertial navigation.

NMR [13] is a physical effect arising from the interaction between the nuclei of certain elements and an external magnetic field. Generally, nuclei possess spin angular momentum and, associated with it, a magnetic dipole moment. In the presence of a magnetic field, H , the spinning nuclei are subjected to a torque which results in a precession of the nuclear spin axis about the direction of the magnetic field. This is known as the Larmor precession and has a characteristic angular frequency, ω_L , given by the relation:

$$\omega_L = \gamma H \quad (4.10)$$

where γ is the ratio of the magnetic dipole moment to the angular momentum, known as the gyromagnetic ratio peculiar to any species of nuclei.

When an angular rate, Ω is applied to a cell containing the precessing nuclei, then the readout mechanism is in a rotating axis frame, resulting in an apparent change in the precessing frequency of the atoms. This rotation of the cell is equivalent to applying a torque to an inertial element in a conventional gyroscope as the precessing nuclei act as an inertial element. Thus, the observed precessional frequency, ω_{obs} , becomes:

$$\omega_{\text{obs}} = \omega_L + \Omega \quad (4.11)$$

or

$$\omega_{\text{obs}} = \gamma H + \Omega$$

Therefore, the determination of the applied rate Ω is dependent on establishing a constant magnetic field at the sensing element and the measurement of a nuclear precessional frequency.

Several techniques [14, 15] employing optical pumping and optical readout systems have been investigated to allow the small frequency shift induced by the rotation of the sensor to be detected. Optical pumping techniques will transfer an assembly of spins from an equilibrium to a non-equilibrium state. Light of the right frequency directed along the direction of the magnetic field will excite magnetic substances

so that the chances of observing the transition of spins from one state to another is enhanced. Transitions are brought about by applying a weak oscillating field at 90° to the static field, H . When the frequency of the oscillating field is close to the Larmor frequency of the excited spins, the orientation of the nuclear spins is reversed. This effect can be detected by circularly polarised light directed at 90° to the direction of the static field. The orientation of the magnetic moments affects the plane of polarisation of this light and hence, after passing through an analyser, the resulting intensity modulation may be picked up by a photodetector.

Techniques for compensating for instability of the external field H involve the measurement of the Larmor frequencies of two magnetic substances contained within one sample cell.

The Northrop Grumman Corporation (formerly Litton Industries) produced a design as shown in Figure 4.24. The NMR cell contains rubidium vapour, krypton and xenon. These gaseous materials are used as they have suitably long

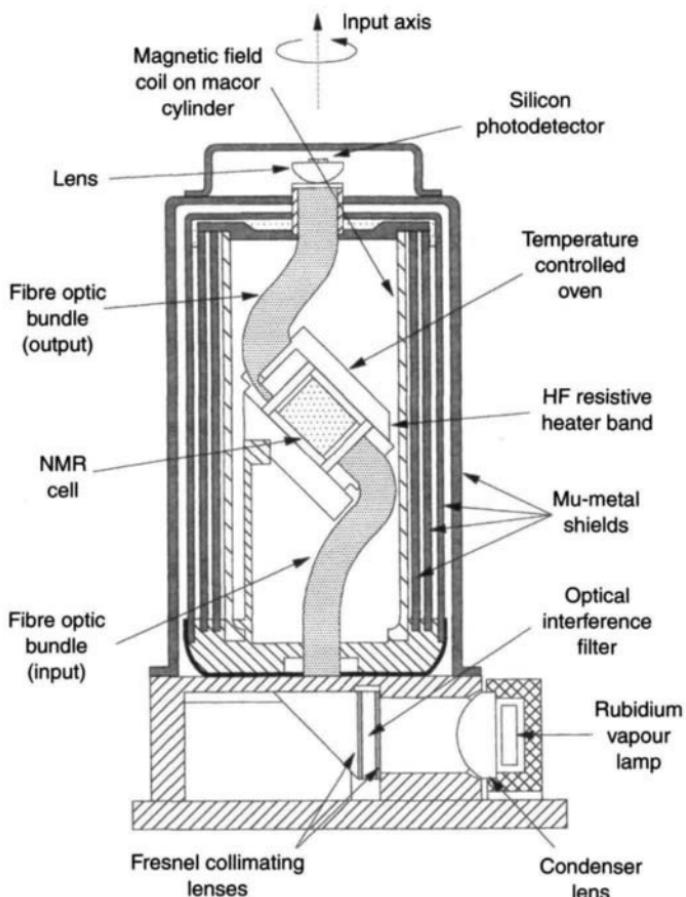


Figure 4.24 Experimental nuclear magnetic resonance gyroscope configuration

relaxation times enabling good sensitivity to be achieved from the very accurate determination of the observed precession frequencies. This design employs one light beam for both pumping and detection.

Clearly, a practical device requires a very uniform and constant magnetic field over the active sample of material. One way, and possibly the only practical technique, is to apply the unique properties of superconductors, such as the cryogenic Meissner effect [16, 17]. A cylinder made of superconducting material operating at a temperature below its critical temperature, will prevent magnetic flux from entering or leaving the central space. Hence, not only will this material shield the nuclear material, it will also trap the flux within the space as its temperature cools below the critical temperature of the superconductor.

One method of detecting the Larmor precession of the nuclear material placed inside the superconducting tube is to use a superconducting quantum interference device (SQUID) magnetometer [17]. In fact, these devices were used during the early development of cryogenic NMR gyroscopes to detect the free induction decay of a sample substance in a superconducting cylinder. The sample substance used was ^3He which, being gaseous at cryogenic temperatures, has the advantage of long relaxation times. This technique was unsuccessful because of the very poor signal to noise characteristics.

The cryogenic NMR gyroscope, as shown in Figure 4.25 [18], has many attractions, particularly the solid-state nature of the construction and the extremely low drift that is possible, as low as arc seconds per annum. This accuracy and the anticipated size of a few litres suggest the most likely application to be ship's strapdown navigation systems.

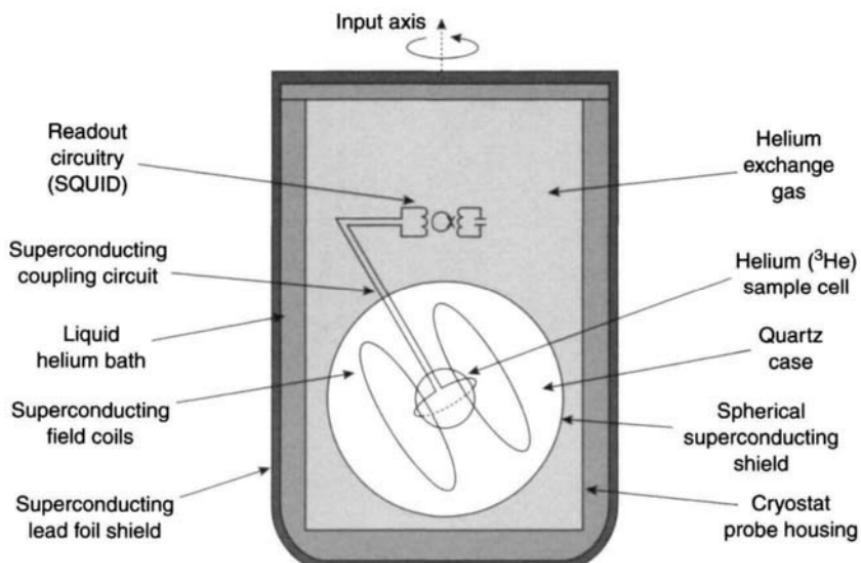


Figure 4.25 Cryogenic nuclear magnetic resonance gyroscope

In the 1970s, the NMR gyroscope and the ring laser gyroscope (discussed in Chapter 5) were looked upon as rival technologies for higher accuracy navigation applications. Although the manufacture of a viable sensor based on NMR technology appeared feasible, its development was overshadowed by massive investment in ring laser technology throughout the 1970s and 1980s, the latter being considered the more promising technology at that time. As far as the authors are aware, there has been little in the way of research effort and resources directed towards the NMR sensor in recent years and, as a result, the full potential of such devices has never been realised; the major thrust for the ‘super-high’ accuracy devices now appears to lie focused on cold-atom sensors (discussed in Section 5.2).

4.5.2 SARDIN

Another example of a rotation sensor which exploits superconductivity has been proposed by Brady [19]. This sensor is known by the acronym SARDIN (Superconducting Absolute Rate Determining INstrument). The device is basically a superconducting cylindrical capacitor, the behaviour of which is governed by the principle that a closed superconducting ring always keeps the amount of flux linking it at a constant value. It does this by generating supercurrents which flow in the ring for an indefinite length of time encountering no resistance to their motion.

Consider two concentric cylinders of superconducting material, one cylinder being raised to a potential V relative to the other. If the assembly is rotated around its central axis with angular velocity Ω , then the moving charges constitute currents which give rise to a net magnetic flux in the region between the two cylinders. As the magnetic flux inside a loop of superconductor cannot change, this field must be backed off by further supercurrents (I) induced in each cylinder. A rigorous mathematical treatment of this effect [20] gives the governing equation:

$$I = \frac{C_0 V r \Omega}{\epsilon} \quad (4.12)$$

where C_0 is the capacitance per unit length of the assembly, r is the mean radius and ϵ is the permittivity of the dielectric between the two cylinders.

With a voltage of 225 V applied to such a device, having a mean radius of 2 cm, coupled to a SQUID magnetometer and an ammeter, Brady reported an angular rate sensitivity of 1 rad/s.

Following initial attempts to produce a superconducting rate sensor of this type in the early 1980s, development was not pursued because of the difficulties encountered in detecting the very small output signals generated. However, the development of higher temperature superconductors coupled with enhancements in signal processing techniques could possibly make this a viable technique in the future.

4.6 Electrostatically suspended gyroscope

Work started to develop this class of gyroscope, also known as the ESG, during the 1950s in the United States. The aim was to suspend a rotating sphere, known as

a gyroscopic ball, by means of an electric field in an evacuated cavity. This form of suspension eliminates the undesirable characteristics of conventional gyroscopes such as bearings, flotation (in a liquid) or suspension.

The principle of operation is that a precision sphere, usually made of beryllium, is initially spun to very high speed whilst being suspended electrostatically in a hard vacuum. This sphere is then allowed to coast, its spin axis maintaining a fixed direction in inertial space provided the sphere and suspension do not have any imperfections which can introduce torques on the sphere. Any angular movement of the case of the gyroscope in space can be determined optically or electrically by reference to this spin axis. Navigation may be achieved by determining the apparent change of this reference axis with respect to a local reference direction such as the local vertical. A schematic diagram of the ESG is shown in Figure 4.26.

The development of the ESG was very successful, producing very accurate sensors capable of achieving drifts of the order of $0.0001^\circ/\text{h}$. The sensors were developed primarily for use on stable platforms, giving navigation accuracies of the order of 0.1 nautical miles per hour. However, they are capable of being used in a strapdown application, particularly in a benign environment.

The major difficulties with this sensor have been associated with:

- manufacturing the inertial element, that is, the ball, to sufficient accuracy to eliminate mass unbalance;
- producing a readout device that does not disturb the motion of the sphere;
- producing a design that avoids ‘grounding the ball’ or dropping the ball as it is sometimes referred to;
- avoiding self-destruction if the power supply fails.

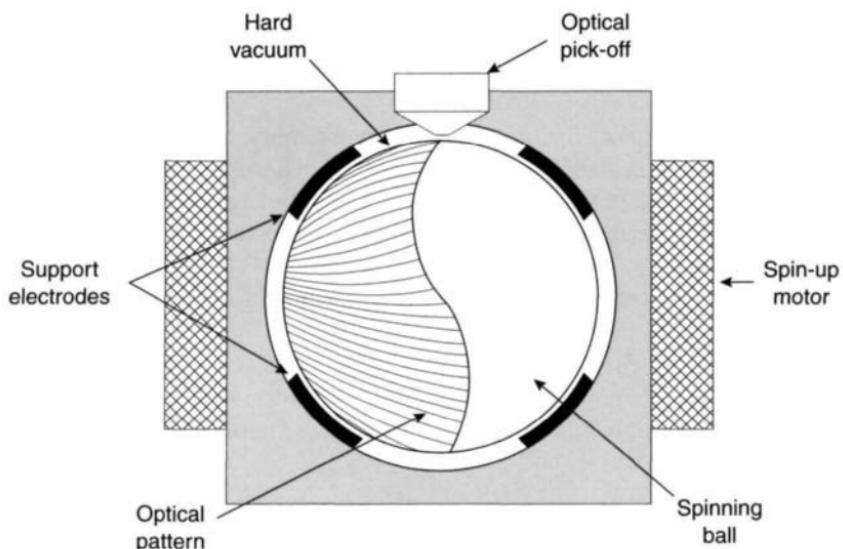


Figure 4.26 Electrostatically supported gyroscope

Errors in the production of the spherical inertial element lead to oscillations, nutations and whirls when the sphere is rotated at the very high speed which is necessary, typically of the order of 150 000 rpm. These undesired excursions are not very well damped because of the high vacuum inside the sensor. Additionally, the sensor requires special techniques to provide shock and vibration protection as the gyroscope does not have any inherent mechanism to damp out linear disturbances of the sphere. The addition of the anti-shock and anti-vibration mounts add to the size of the sensor.

The ESG is a very high accuracy sensor and has been used for many decades in specialised applications in aircraft, ships and submarines. This sensor has also been suggested for detecting acceleration. Since the support electrodes provide the only forces to accelerate the sphere to keep it moving with its case, measurement of these forces provide a measure of acceleration.

Unfortunately, despite being a very simple concept, the design is complex and the gyroscope is large and expensive. However, it is one of the most accurate conventional gyroscopes ever to be designed and produced.

A similar sensor was produced by Rockwell in the United States known as the gas-bearing free rotor gyroscope. This sensor was based on a spherical bearing, like a ball and socket arrangement, using a self-acting gas-lubricated bearing to support the ball. However, this sensor differs from the ESG as it only has limited angular freedom and is not really suitable for strapdown applications.

4.7 Other devices for sensing angular motion

There are a number of other devices that are either used or could be developed to sense angular motion. They are considered separately as they do not fit easily in any of the above classes of gyroscope, and are generally not the 'prime' angular motion sensor used in a strapdown navigation system. Included in this category are fluidic sensors and fluxgate magnetometers, both of which offer well-established technology suitable for various applications. The sensors tend to be small, reliable and rugged. Fluidic devices are often used for providing short-term references, as in stabilisation tasks, whereas magnetometers are generally used for long term aiding of systems.

4.7.1 Fluidic (*flueric*) sensors

This term has several meanings, such as sensors which make use of the flow of fluid for either the propulsion of the rotor or its support. Another meaning, and of more direct relevance to the immediate consideration here, is the use of a fluid for the sensing of angular motion. Sometimes, this class of angular motion sensor is known by the term flueric; the term fluidic being reserved to describe those sensors that use a fluid either for support or for powering the rotor. The text will concentrate on the use of a fluid for the sensing of angular motion, that is, flueric sensors.

Development of this type of sensor started in the 1960s. They appear to offer an interesting alternative to the electromechanical instruments. Despite much effort,

it has not proved possible to make this type of sensor suitable directly for inertial navigation applications. The major difficulty is achieving adequate stability, resolution and insensitivity to environmental effects, particularly temperature changes. However, this class of sensor has found many applications, including flight control, stabilisation and in flight safety systems.

One form of sensor is a flueric gyroscope. In this device, there is a spherical cavity with porous walls and a rotating mass or swirl of gas within the cavity. When the sensor case is rotated, the direction of the swirl of gas remains fixed and the displacement can be detected by monitoring pressure changes. Figure 4.27 depicts the form of this sensor.

Another form of this class of instrument has a continuous laminar flow of gas from an orifice which impinges on a pair of hot wire detectors. When angular motion is applied about axes orthogonal to the gas flow, the gas jet appears to be deflected laterally relative to the case. This results in a differential cooling of the hot wires, with a consequential change in resistance which is detected using a bridge circuit. The output signal is proportional to the applied angular rate. This form of sensor is shown schematically in Figure 4.28. It tends to be sensitive to temperature gradients, acceleration, vibration and shock.

These sensors may either use fluid bled from an engine, such as a jet engine's efflux, or may be pumped in a closed cycle.

Such sensors provide an inexpensive short-term reference with turn rate measurement accuracies of around 1% of the applied rate. Generally, they also show significant sensitivity to most environmental features, particularly temperature changes. However, some designs are very rugged and reliable and are capable of operating over a wide temperature range.

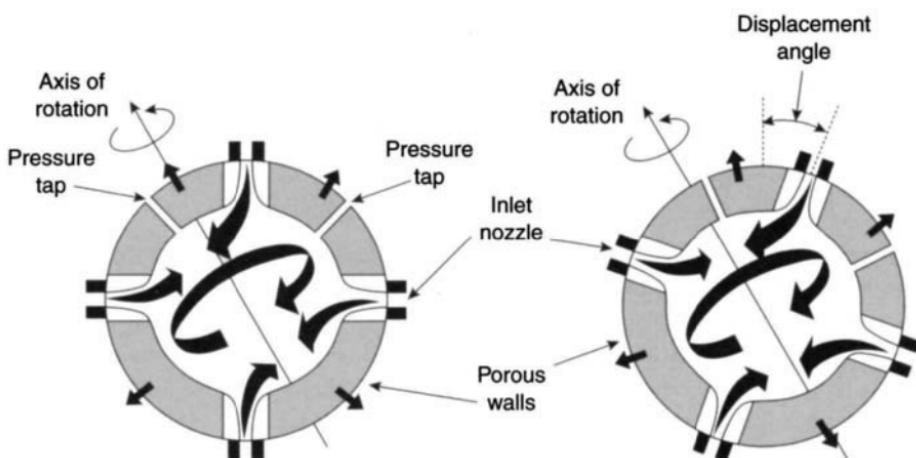


Figure 4.27 Schematic diagram of a flueric attitude sensor

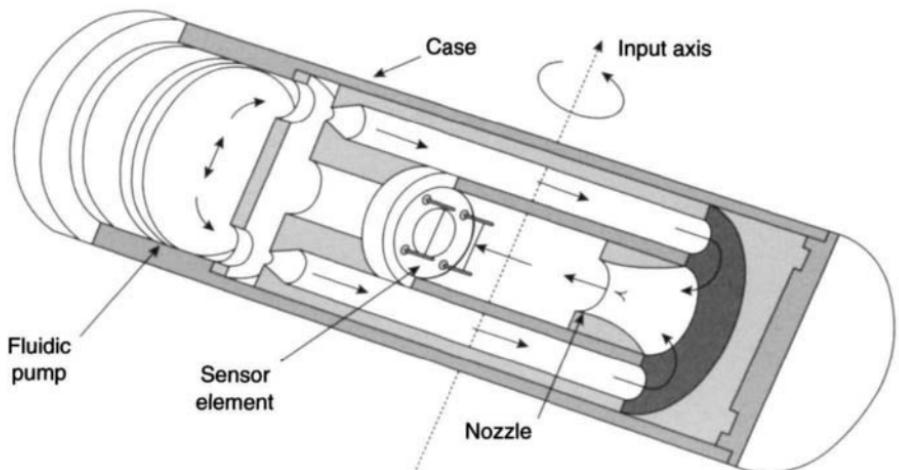


Figure 4.28 Gas jet sensor

4.7.2 Fluxgate magnetometers

It was during the 1950s that an airborne magnetometer was developed at the Royal Aircraft Establishment, Farnborough (now DSTI and QinetiQ) for attitude measurement. Since that time, there have been many developments of this type of attitude sensor [21, 22]. It is usually based on fluxgate elements. A magnetometer has three such magnetic sensing elements that are mounted mutually orthogonal to each other. These axes are usually arranged to be aligned with the principal axes of the vehicle. This configuration enables the attitude of the vehicle to be determined with respect to the Earth's magnetic vector. The magnetometer alone can not give an unequivocal measurement of a vehicle's attitude. Measurements made with such a sensor define the angle between the Earth's magnetic field and a particular axis of the vehicle. However, this axis can lie anywhere on the surface of a cone of semi-angle equal to that angle about the magnetic vector. Hence, an additional measurement is required to determine attitude with respect to another reference such as the gravity vector.

The basic operation of a fluxgate magnetometer is similar to the operation of an electrical transformer. However, the magnitude of the excitation signal is chosen to drive the core, linking the excitation and pick-up coils, into saturation on alternate peaks of the excitation signal. The high permeability of the core magnifies the effect of the changing magnetic field (H) generated from the changing current in the excitation coil. A simple fluxgate magnetometer element is shown in Figure 4.29.

The magnetic induction (B) in the core increases as the applied magnetic field from the excitation coil increases up to a maximum value (B_{\max}), determined by the core material. As the magnetic field increases beyond the saturating value (H_{sat}) the induction remains constant. This is illustrated in the B - H curves shown in Figure 4.30. This effect is reversible, so as the magnetic field decreases below the saturation value the induction decreases until it reaches its negative saturation ($-B_{\max}$).

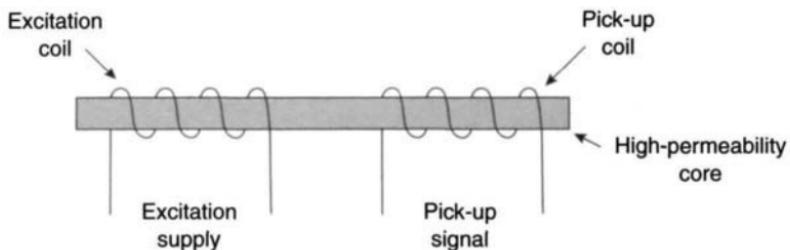


Figure 4.29 A simple fluxgate magnetometer

A voltage signal is only induced in the pick-up coil whilst the magnetic induction in the core is changing. Consequently, there is no signal in the pick-up coil during periods of induction saturation. The figure also shows the induction waveforms which result when a square wave signal is applied to the excitation coil with and without an external magnetic field acting along the core. In addition, the figure includes the resulting voltage waveform across the pick-off coil and the 'driving' magnetic field profile.

In the absence of an external magnetic field, the positive and negative excursions in the voltage appearing across the pick-off coil are of equal magnitude, as shown in the left-hand side of Figure 4.30, and there is no net output from the device. In the presence of an external magnetic field, this field acts either to aid or to impede the field generated by the alternating current in the excitation coil. As a result, the core is saturated more rapidly and remains in saturation slightly longer for one cycle of the excitation, the cycle in which the excitation and external fields are acting in the same direction. The converse is true when these fields are in opposition. As shown in the right-hand side of the figure, the induction waveform is no longer symmetrical and a modified voltage waveform appears across the pick-off coil. The amount of the advance and delay in the voltage waveform is proportional to the strength of the external magnetic field.

The change in the voltage waveform characteristic is not easy to measure in the simple scheme described above. A preferred arrangement which enables a signal proportional to the size of the external magnetic field to be extracted with relative ease is shown in Figure 4.31.

In this device, a pair of cores are used with the excitation coils wound in series opposition. In this case, the magnetic induction in the two cores cancel out in the absence of an external magnetic field. However, when an external field is present, the magnetic induction waveforms for the two cores are modified in the manner described above. The resulting voltage waveforms for each of the cores are shown in the bottom of Figure 4.31. The pick-up coil is wound so that it sums these two contributions giving rise to the sum channel waveform shown in the figure. This signal can be applied to a low pass filter and a steady output produced indicating the magnitude of the external magnetic field.

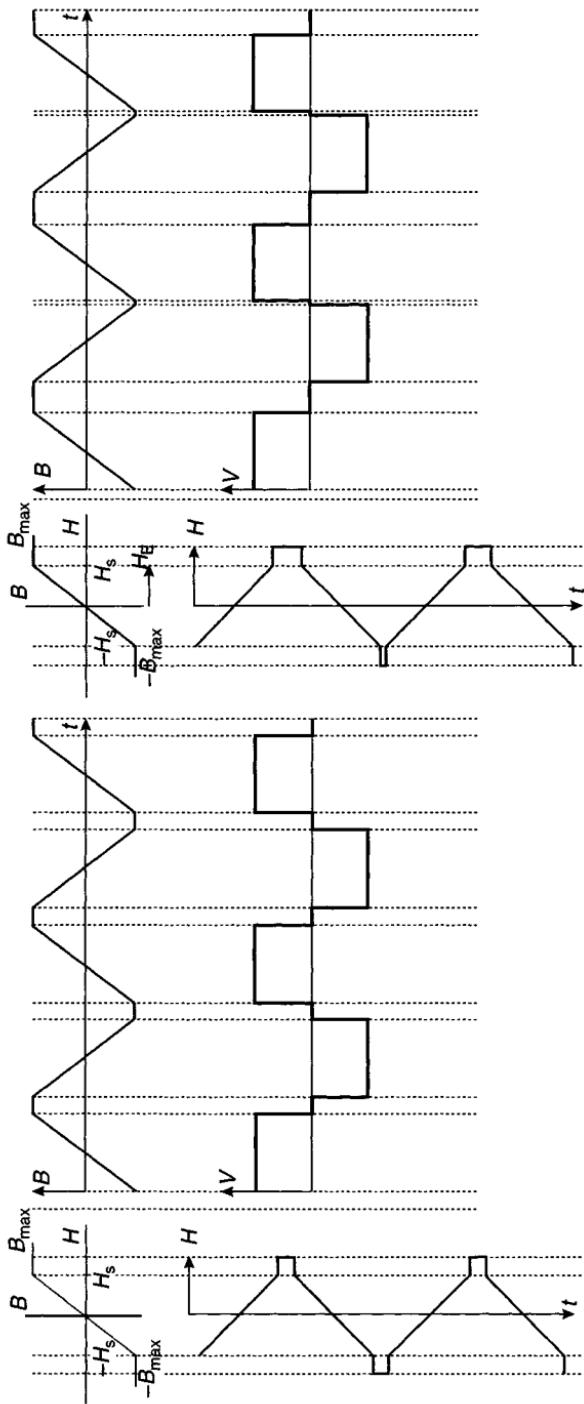


Figure 4.30 Magnetometer waveforms

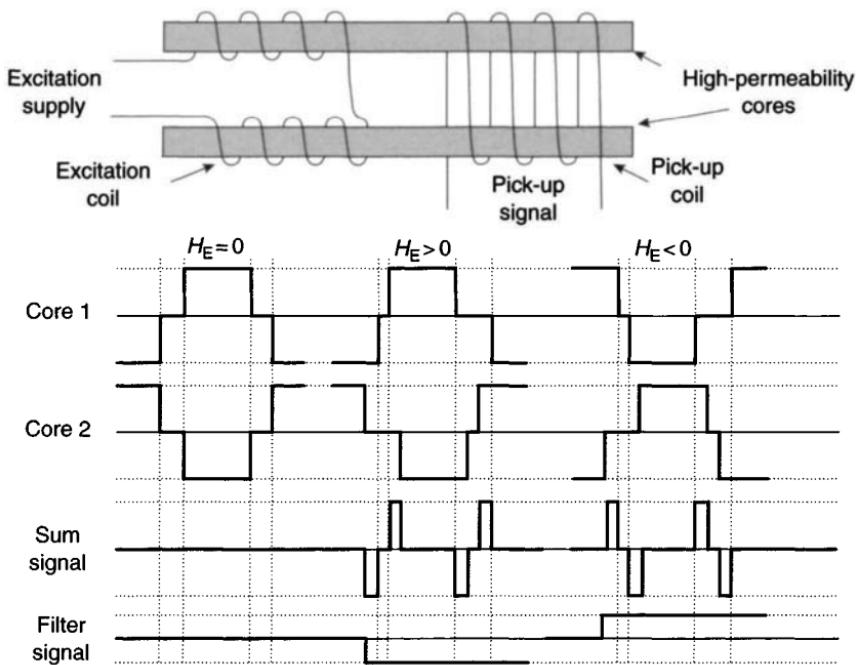


Figure 4.31 Twin-core fluxgate magnetometer and waveforms

This analysis for the two core configuration is dependent on the use of perfectly matched cores and windings. This fundamental requirement for accurate measurement of the magnetic field is very difficult to achieve in a practical device. A common solution to this problem is to use a circular toroidal core over wound with an excitation coil. The excitation coil is energised with an alternating current signal having a frequency of 1–1.5 kHz. The pick-up coil is a single coil wound over the toroid. A second coil can be wound over the toroid at right angles to the pick-up coil, and this can be used as a nulling device. This magnetometer arrangement is shown in Figure 4.32.

When operated in a closed loop mode, commonly known as the closed loop second harmonic mode, a current is passed through the second over-wound coil to null the effects of the detected magnetic flux. It is the amplitude of this current which is used to give a measurement of ambient magnetic field strength. The output from this coil is filtered to select only the second harmonic frequency, hence avoiding saturation of the amplifiers by unwanted frequencies.

The major problem with this form of instrument is that it has poor angular resolution capability, typically about 0.1° , and sensitivity to magnetic anomalies. However, this latter sensitivity can be valuable provided the magnetic anomaly in the region of operation is stable and well charted. A more serious limitation is often posed by the effects of the structure of the vehicle in which the sensor is mounted. The vehicle is likely to have ferromagnetic materials present as well as changing fields associated

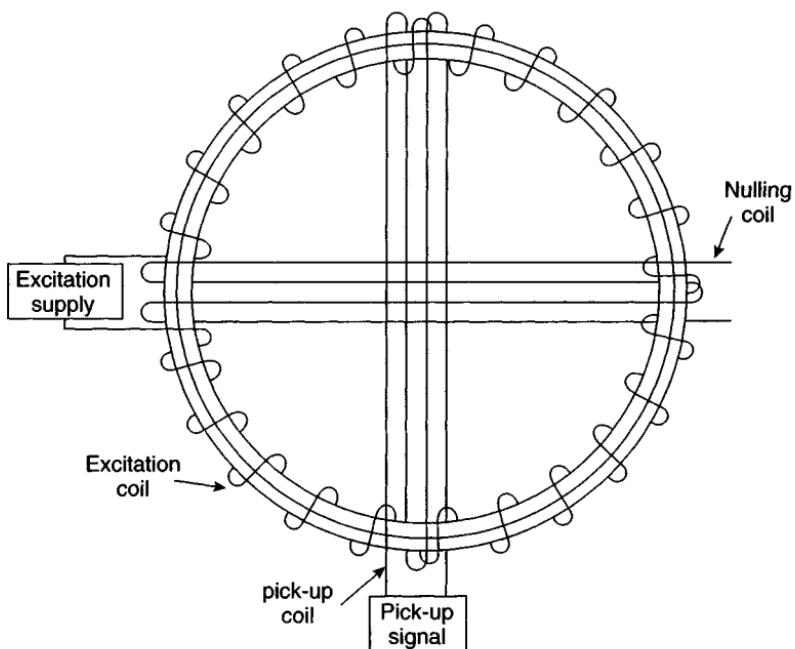


Figure 4.32 Toroidal fluxgate magnetometer

with power supplies and other instrumentation. Hence, such extraneous magnetic fields are likely to determine the precision with which the attitude of the vehicle can be determined, with respect to the magnetic field vector of the Earth. Therefore, these sensors are usually mounted at the extremities of a vehicle, as far away as possible from the source of extraneous fields. In some applications, it is possible to route wiring so as to cancel induced fields or else calibrate the sensor in its host vehicle.

A further potential problem arises in a strapdown configuration in a rolling airframe travelling at latitudes where the dip angle of the Earth's magnetic field is large. The roll motion will produce the effect of an alternating field on the magnetometer. An error will result unless the frequency response of the magnetometer is adequate to 'follow' this rate of change of field intensity. In some circumstances it may be necessary to incorporate another sensor to monitor the rapid roll rate in order to compensate the output of the magnetometer.

Generally, fluxgate magnetometers are cheap, small, rugged and reliable, and are capable of working over a very wide temperature range. They are commonly used in low cost systems to provide an attitude reference in pilotless aircraft for example, or in more sophisticated reference systems to provide a long term attitude reference. In this case, the data supplied by the magnetometer supplement the estimates made by other sensors, such as gyroscopes, enabling the attitude error to be bounded.

The dimensions of the high-permeability core can be modified to give the desired response characteristics. For example, shorter length cores give greater linearity,

whereas long thin cores provide a higher output for a given drive power, hence greater sensitivity.

4.7.3 *The transmission line gyroscope*

The use of the inertia of an electromagnetic field as the rotation sensing element has been demonstrated many times, in applications such as the ring laser gyroscope and the fibre gyroscope, and is discussed in the following chapter. Consideration has also been given by some researchers to the use of non-propagating fields such as the electrostatic gyroscope.

Analysis has also been undertaken by Forder [23] of general relativistic effects which may be observed in a closed ring of parallel wire or coaxial transmission line, appropriately energised, with either static or propagating electromagnetic fields. Three distinct effects can be identified, viz. electrostatic, magnetostatic and electromagnetic, allowing this type of device to be operated as an angular motion sensor in a number of different modes. It is predicted that the changes in line voltage and current detected in a circular loop of transmission line of radius R and line impedance Z rotated at an angular rate Ω are given by:

$$\Delta V = -(ZI)\Omega R/c$$

or

$$\Delta V = -(V/Z)\Omega R/c \quad (4.13)$$

where V is the charging potential, I is the line current and c is the speed of light.

Clearly, one of the major difficulties with the implementation of such devices is the detection of the very small changes in potential and current predicted above. However, practical forms of such sensors based on the above predictions may become feasible and practical in the future as various technologies advance.

References

- 1 CASE, W.: 'The gyroscope: an elementary discussion of a child's toy', *American Journal of Physics*, 1977, **45** (11)
- 2 COUSINS, F.W.: 'The anatomy of the gyroscope', in HOLLINGTON, J.H. (Ed.): 'Agardograph 313, Parts I, II and III' (AGARD, 1990)
- 3 SAVET, P.H. (Ed.): 'Gyroscopes: theory and design' (McGraw-Hill, 1961)
- 4 EDWARDS, C.S.: 'Inertial measurement units – building blocks into the 1990s'. DGON proceedings, *Gyro Technology Symposium*, Stuttgart, 1986
- 5 PHILPOT, J.StL., and MITCHELL, J.H.: GB Patent 599826, 1948
- 6 CRAIG, R.J.G.: 'Theory of operation of an elastically supported tuned gyroscope', *IEEE Transactions on Aerospace and Electronic Systems*, 1972, **AES-8** (3), pp. 280–297
- 7 BRYAN, G.H.: 'On the beats in the vibrations of a revolving cylinder or bell', *Proceedings Cambridge Philosophical Society*, 1890, **7** (101)

- 8 LOPER, E.J., and LYNCH, D.D.: 'Hemispherical resonator gyro status report and test results'. National Technical Meetings, Institute of Navigation, San Diego, CA, January 1984
- 9 LOPER, E.J., LYNCH, D.D., and STEVENSON, K.M.: 'Projected performance of smaller hemispherical resonator gyros'. *Position, Location and Navigation Symposium Plans 86*, Las Vegas, NV, November 1986
- 10 BODUNOV, B.P., LOPATIN, V.M., BODUNOV, S.B., and KOVSHOV, G.N.: 'Gyroinclinometer for surveying during the drilling process'. DGON proceedings, *Gyro Technology Symposium*, Stuttgart, 1999
- 11 JOHNSON, B., and LONGDEN, I.M.: 'Vibrating structure gyroscopes and their applications'. DGON proceedings, *Gyro Technology Symposium*, Stuttgart, 1994
- 12 ELWELL, J.: 'Progress on micromechanical inertial instruments'. DGON proceedings, *Gyro Technology Symposium*, Stuttgart, 1991
- 13 RUSHWORTH, F.A., and TUNSTALL, D.P.: 'Nuclear magnetic resonance' (Gordon and Breach, 1978)
- 14 KARWACKI, F.A.: 'Nuclear magnetic resonance gyro development', *Navigation Journal of The Institute of Navigation*, 1980, **27** (1)
- 15 KANEGBURG, E.: 'A nuclear magnetic resonance (NMR) gyroscope with optical magnetometer detection', SPIE 157, *Laser Inertial Rotation Sensors*, 1978, 73
- 16 ILLINGWORTH: 'Dictionary of physics' (Penguin, 1990)
- 17 POTTS, S.P., and PRESTON, J.: 'A cryogenic nuclear magnetic resonance gyroscope', *Journal of Institute of Navigation*, 1981, **34** (1), pp. 19–37
- 18 SHAW, G.L., and TABER, M.A.: 'The ^3He gyro for an all cryogenic inertial measurement unit'. DGON proceedings, *Gyro Technology Symposium*, Stuttgart, 1983
- 19 BRADY, R.M.: 'A superconduction gyroscope with no moving parts', *IEEE Transactions*, 1981, **MAG-17**, pp. 861–2
- 20 POTTS, S.P., and CREWE, P.P.: ASWE Memorandum XTN 82010, 1981
- 21 NOBLE, R.: 'Fluxgate magnetometry', *Electronics World and Wireless World*, 1991, **97** (Pt 1666), p. 726
- 22 HINE, A.: 'Magnetic compasses and magnetometers' (Adam Hilger, 1968)
- 23 FORDER, P.W.: 'General relativistic electromagnetic effects in transmission line gyroscopes', *Classical Quantum Gravity*, 1986, **3**, p. 1125

Chapter 5

Gyroscope technology 2

5.1 Optical sensors

5.1.1 Introduction

This term is applied to those classes of gyroscope which use the properties of electromagnetic radiation to sense rotation. Such devices often use the visible wavelengths, but it is also possible to operate in the near infrared. Some mechanical gyroscopes use optical angle pick-off sensors, but for this discussion they are not classed as optical gyroscopes. Optical gyroscopes use an interferometer or interferometric methods to sense angular motion. In effect, it is possible to consider the electromagnetic radiation as the inertial element of these sensors.

It was during the late nineteenth century that Michelson pioneered work with optical interferometers, although his goal was not to produce an optical gyroscope. In 1913, the Sagnac effect was reported [1] and this is the fundamental principle on which optical gyroscopes are based. When light travels in opposite directions (clockwise and anti-clockwise) around an enclosed ring, differences arise in the apparent optical length of the two paths when the ring is rotated about an axis orthogonal to the plane containing the ring. In 1925, this concept was applied by Michelson and Gale [2] in Chicago using a ring gyroscope with a perimeter of over one mile. By sending ordinary light through evacuated water pipes, they were able to detect the shift produced by the rotation of the Earth.

Further impetus to produce an optical sensor resulted from the demonstration of a laser by Maiman in 1960 [3]. These devices produce a well collimated and highly monochromatic source of electromagnetic energy between the ultraviolet and far infrared part of the spectrum, the wavelength being determined by the laser medium. In 1963, the first ring laser was demonstrated by workers at Sperry Gyroscope [4]. This marked the beginning of the development of the ring laser gyroscope. About a decade later the fibre optic gyroscope was first demonstrated [5].

Clearly, the history of the development of optical gyroscopes is far more recent than the history of the mechanical sensors. Further impetus for the development of

these sensors, besides the development of the laser, was the interest in the application of strapdown technology and the desire to capitalise on the benefits anticipated from the use of solid-state inertial sensors. One of the main difficulties in the application of strapdown technology from the point of view of gyroscope performance was the lack of adequate dynamic range of the mechanical sensors for the more accurate applications. Initial estimates of performance suggested that the ring laser gyroscope could provide the solution.

The spectrum of performance of optical gyroscopes ranges from the very accurate with bias of less than $0.001^\circ/\text{h}$, usually ring lasers, to tens of degrees per hour, often from the simpler fibre optic gyroscopes. Hence the range covered by the optical devices is very similar to that covered by the mechanical gyroscopes. Generally, all the types of optical gyroscopes are suitable for various strapdown applications, depending of course on the demanded accuracy of the system.

It appears that the application of optics to the sensing of angular rate can offer a number of advantages over the use of the well-established mechanical technology. Some of the advantages often cited are listed below:

1. wide dynamic range;
2. instant start-up;
3. digital output;
4. output independent of some environmental conditions (acceleration, vibration or shock);
5. high rate capability;
6. easy self-test;
7. system design flexibility;
8. extended running life.

5.1.2 Fundamental principles

Optical gyroscopes rely upon the detection of an effective path length difference between two counter-propagating beams of light in a closed path. The mathematical development given here shows how this path difference arises in the presence of an applied turn rate about an axis perpendicular to the plane containing the light path.

Consider a perfect stationary circular interferometer with the light constrained to travel around the circumference of a circle of radius R as shown in Figure 5.1. Light enters the 'ring' at the point X, where there is a beam splitter which directs two beams of light in opposite directions around the complete ring, these beams recombining later at the beam splitter. The transit time (t), the time for the light to make one complete pass around the ring whilst the ring is stationary, is identical for both beams and is given by:

$$t = \frac{2\pi R}{c} \quad (5.1)$$

where c is the velocity of light, which is considered to be invariant.

However, when the interferometer is rotated with angular velocity Ω , the time for each light beam to pass around the circumference is modified. This is because of the

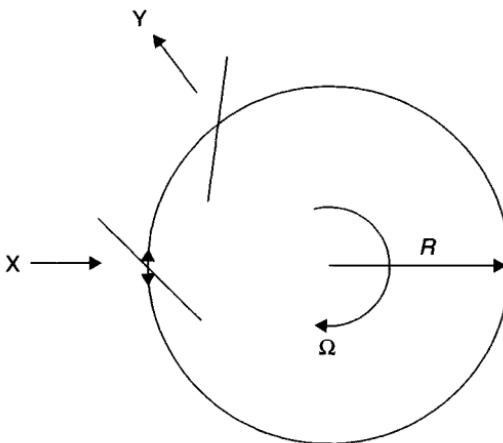


Figure 5.1 Circular rotating (Sagnac) interferometer

motion of the beam splitter during the time taken for the light to pass around the ring. As shown in the figure, the beam splitter will have moved to position Y. Therefore, light travelling in a clockwise direction will have to travel further than the distance travelled when stationary. The converse is also true for the anticlockwise beam. More generally, with respect to inertial space, light travelling with the direction of rotation must travel further than when the interferometer is stationary. Light travelling against the direction of rotation will have its path length reduced when compared with the stationary condition. Hence, the single pass transit time for the two beams is given by the following equations.

$$\text{Clockwise path, } t_1 = \frac{2\pi R + \Delta L_+}{c} \quad (5.2)$$

$$\text{Anti-clockwise path, } t_2 = \frac{2\pi R - \Delta L_-}{c}$$

Now, $\Delta L_+ = R\Omega t_1$ and $\Delta L_- = R\Omega t_2$ are the increment and decrement in the path length respectively. As reported by Aronowitz [6], this can also be interpreted as the velocity of light being different for the two counter-propagating beams and the path length being invariant.

From the above equations, the difference in transit time, Δt , is given by:

$$\Delta t = t_1 - t_2 = 2\pi R \left[\frac{1}{c - \Omega R} - \frac{1}{c + \Omega R} \right] \quad (5.3)$$

To first order approximation, this becomes:

$$\Delta t = \frac{4\pi R^2 \Omega}{c^2} \quad (5.4)$$

The optical path length difference $\Delta L = c\Delta t$, and may therefore be expressed as:

$$\Delta L = \frac{4\pi R^2 \Omega}{c} \quad (5.5)$$

The area (A) enclosed by the path length is πR^2 . Hence, the above equation may be rewritten as follows:

$$\Delta L = \frac{4A\Omega}{c} \quad (5.6)$$

Aronowitz [6] gives a more rigorous equation for the difference in closed path transit time for counter-propagating light beams on a rotating frame. This is based on the loss of synchronisation between a clock travelling on a rotating reference frame compared with one on a stationary reference frame. The conclusion from this study is that the optical path difference, $4A\Omega/c$, is independent of the position of the axis of rotation. As noted by Aronowitz, measurement of the optical path difference enables an observer, located on a rotating reference frame, to measure the so-called absolute rotation of his reference frame.

The various optical sensors described below rely on generating a path difference in an interferometer, the major differences being in how the light is generated and how the path difference is 'observed'.

5.1.3 Ring laser gyroscope

5.1.3.1 Introductory remarks

As stated above, serious development started in the early 1960s. The first ring laser gyroscopes were large and somewhat delicate. Substantial investment has led to the production of very compact devices that produce extremely low bias, of $0.001^\circ/\text{h}$ or better. Typical path lengths for the accurate sensors are about 300 mm. Very small laser gyroscopes have also been produced with a path length of about 50 mm. Generally, these small sensors have a bias in the region of $5^\circ/\text{h}$. Currently, the more accurate gyroscopes are used in strapdown navigation systems in commercial aircraft as well as in military fixed and rotating wing aircraft.

5.1.3.2 Principle of operation

Operation of a ring laser gyroscope relies on the fact that an optical frequency oscillator can be assembled as a laser using three or more mirrors to form a continuous light path. Typically, three mirrors are used to form a triangular shaped light path. If a light beam is generated at any point on this path, it can travel around the closed path, being reflected in turn from each mirror, to arrive back at its starting point. Sustained optical oscillation occurs when the returned beam is in phase with the outgoing beam. Two such travelling wave laser beams are formed independently, one moving in a clockwise direction and the other in an anti-clockwise direction.

When the sensor is stationary in inertial space, both beams have the same optical frequency. However, when the sensor is rotated about the axis perpendicular to the plane containing the light beams, changes occur in the optical path lengths of the

two beams. The frequency of each beam changes to maintain the resonant condition required for laser action such that the frequency of the beam with the longer path length decreases whilst the frequency of the other beam increases. This path difference is very small, no more than 1 nm, thus a source with high spectral purity and stability, such as a helium–neon gas laser, is required to make the laser gyroscope concept feasible.

Maintenance of laser action requires a constant phase at a given mirror surface after every round trip in order to maintain the resonant condition.

Hence, if L_a is the anti-clockwise path length and L_c the clockwise path length, then the resonant condition is given by:

$$\begin{aligned} L_a &= p\lambda_a \\ L_c &= p\lambda_c \end{aligned} \quad (5.7)$$

where p is the mode number, typically of the order of a million, and λ_a and λ_c are the two wavelengths of laser energy. When this interferometer is rotated at a rate Ω , these path lengths differ and are given by:

$$\begin{aligned} L_a &= p\lambda_a = L + \frac{2A\Omega}{c} \\ L_c &= p\lambda_c = L - \frac{2A\Omega}{c} \end{aligned} \quad (5.8)$$

where L is the perimeter length and the path difference $\Delta L = 4A\Omega/c$.

Now if v_a and v_c are the optical frequencies of the two beams, $v_a\lambda_a = v_c\lambda_c = c$. Substituting for wavelength in the above equations, we have,

$$v_a = \frac{cp}{L_a} \quad \text{and} \quad v_c = \frac{cp}{L_c} \quad (5.9)$$

Hence, small changes in path length result in small changes in frequency, Δv , given by the relation,

$$\frac{\Delta v}{v} = \frac{\Delta L}{L} \quad (5.10)$$

Substituting for ΔL in this equation, this beat frequency can be expressed as:

$$\Delta v = \frac{4A\Omega}{cL} \quad v = \frac{4A\Omega}{L\lambda} \quad (5.11)$$

where

$$\lambda = \frac{\lambda_a + \lambda_c}{2} \quad \text{and} \quad v = \frac{v_a + v_c}{2} \quad (5.12)$$

It follows from eqn. (5.11) that the turn rate (Ω) may be determined from the frequency difference (Δv) generated in its presence. The scale-factor of the sensor is directly proportional to the area (A) enclosed by the optical path. Changes in A result from variations in the cavity length. Use of active laser gain control and cavity path length control usually contain these excursions to a few parts per million or less for most designs.

Substitution of typical values into eqn. 5.11 shows the beat frequency to be from a few hertz up to megahertz. This beat frequency can be detected, even for very slow

rotation rates. As noted by Aronowitz, thermal and mechanical instabilities can cause frequency variations in the individual beams that are far greater than the rotational beat frequency, as can be deduced from eqns. (5.9)–(5.11). The successful operation of this type of sensor is achieved since both beams occupy the same laser cavity and therefore are subject to identical perturbations.

In order to detect the rotational motion, a small amount of light from each beam is allowed to ‘escape’ through one of the mirrors, known as the output mirror, and the two beams are combined using a prism to form an interference pattern on a set of photo-diodes. The frequency difference between the two beams causes the interference fringes to move across the detectors at a rate equal to the difference in frequency between the two beams which is also proportional to the rotational motion.

Hence, the movement of a single fringe past the detector corresponds to an incremental rotation, $\Delta\theta$, where,

$$\Delta\theta = \frac{\lambda L}{4A} \quad (5.13)$$

This equation can be used to determine the sensitivity of a ring laser gyroscope. For a device having an equilateral triangular path of total length L , the area is given by:

$$A = \frac{1}{2} \left(\frac{L}{3} \right)^2 \sin 60^\circ \quad (5.14)$$

Substituting in the preceding equation, it can be shown that the sensitivity is inversely proportional to the path length, viz.

$$\Delta\theta = \frac{3\sqrt{3}\lambda}{L} \quad (5.15)$$

For instance, considering a helium–neon laser in which the optical wavelength is 0.6328 μm (632.8 nm), the sensitivity of a 30 cm path length device is 2.25 arcs per fringe. For the smallest ring caser gyroscope, with a 50 mm path length, the sensitivity is 13.5 acrs per fringe.

5.1.3.3 *The lock-in phenomenon*

At very low rotation rates, the two laser beams in the cavity cease to oscillate at different frequencies and assume the same frequency, or ‘lock together’. The interference pattern does not change so there is no output signal. This phenomenon of frequency synchronisation is illustrated in Figure 5.2 and is known as the lock-in condition, or simply lock-in. It is analogous to the mutual coupling common in electronic oscillators working in close proximity at similar frequencies. In the optical case, it is caused by the radiation of one laser beam being scattered into the other beam causing the host mode to change frequency towards that of the back scattered energy, with the consequence of both beams being shifted to the same frequency. There are many sources of back scattering, but careful design and the use of very high quality mirrors allows the effect to be minimised and the lock-in condition is restricted to a very narrow zone close to the zero rotation rate.

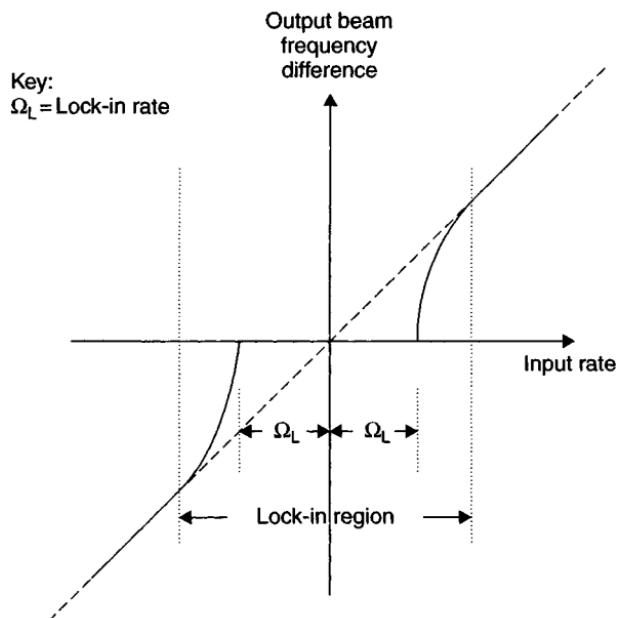


Figure 5.2 Laser gyroscope input/output characteristic

Alleviation of lock-in

One of the most common methods used to alleviate the lock-in problem is the use of mechanical oscillation. Mechanical dithering consists of applying angular vibrations to the entire cavity at high frequency but at low amplitude and through small angles, thereby avoiding low frequency outputs. Through the use of a so-called large dither product (dither frequency multiplied by the amplitude), having a high frequency motion but small displacement, very little time is spent by the sensor in the lock-in region, hence greater accuracy is achieved through missing fewer pulses.

The dither frequency has a random frequency component superimposed on it which randomises slightly the motion of the cavity. The result of this randomisation is that the motion has a randomised rate noise rather than a mean bias which would be produced by a sinusoidal motion of the cavity block [7]. This motion produces a random walk in angle which appears on the output of the sensor.

The use of mechanical dither causes an increase in size, weight and complexity. It is necessary to subtract the dither motion from the gyroscope's output and this may be accomplished either optically or electronically. Any difference between the actual and compensated output is termed dither spill over which leads to a scale-factor error.

Another technique that is currently being applied is called magnetic mirror biasing. This electro-optical technique uses a non-reciprocal magneto-optical effect (the transverse Kerr effect [8]). One of the highly reflective mirrors has a magnetic coating on its top surface. The magnetic coating, when saturated by an applied magnetic field, causes a difference in phase delay between the two counter-propagating

laser beams, biasing the frequencies away from the lock-in zone. In order to prevent any drifts in bias voltage being interpreted as a rotation rate, it is necessary to switch between two bias points so any drifts average to zero. A potential disadvantage with magnetic mirrors is the introduction of higher cavity losses which may exclude it from high accuracy applications. However, it is a genuine solid-state sensor [9] which is smaller and less complex than the mechanically dithered ring laser gyroscope.

Multi-oscillator concepts have been demonstrated [10–12] where more than a single pair of beams propagate in the same cavity, usually four beams in a square configuration. Independent lasing of left and right polarised modes are propagated in each direction in the cavity, giving a total of four modes. Avoiding the phenomenon of lock-in still applies, so it is necessary to bias the modes away from this zone. The reciprocal splitting between the right-hand and left-hand circularly polarised modes can be several hundred megahertz, achieved by using a quartz retarder plate, or a non-planar cavity configuration. The real difficulty is to achieve adequate biasing of the direction dependent (non-reciprocal) modes, that is, the clockwise and anti-clockwise right-hand circularly polarised modes, and similarly with the opposite handed modes.

Three methods have shown promise: use of a Faraday rotation element in the laser cavity; a mirror with a magneto-optic coating that uses either the polar or transverse Kerr effect; application of a Zeeman field to the discharge to induce a frequency change. Sometimes the ring laser gyroscope that has a Faraday cell in its cavity is called the differential laser gyroscope or DILAG; the other name is the four-frequency gyroscope. One distinct advantage that this form of gyroscope has over the conventional ring laser gyroscope is an enhanced scale-factor; giving a sensitivity that is twice that of the equivalent sized conventional laser gyroscope.

5.1.3.4 Detailed description of the sensor

The more commonly used ring laser gyroscope configuration which uses three mirrors is illustrated in Figure 5.3. Successful configurations using four mirrors have also been produced.

The major components of the gyroscope, as shown in Figure 5.3, are:

1. The laser block—formed in a low expansion ceramic glass such as Zerodur or Cervit. Contained within it is the lasing medium which is usually a mixture of helium and two isotopes of neon to enable the two modes to propagate without competition.
2. The optical components—usually just the mirrors and photo-detectors, but an optical biasing element may also be included in the laser cavity. There are two types of mirror; the partially reflecting (partially transmitting) ‘output’ mirrors, as mentioned above, and the mirror with a very high reflection coefficient. All of the mirrors are multi-layer dielectric stacks of alternate layers of materials with a different refractive index, deposited on extremely high quality polished substrates to give very low back scatter. One of the mirrors is attached to a piezoelectric device so that it can be moved in and out to maintain a constant path length (at the resonant condition) as the temperature changes. The other mirrors are

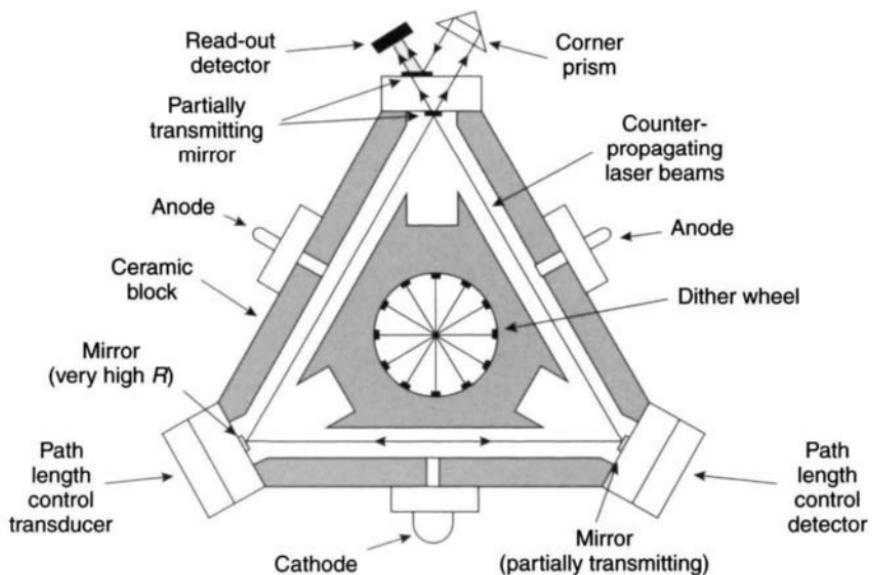


Figure 5.3 Schematic diagram of a ring laser gyroscope

firmly bonded directly to the laser block. A variety of techniques are used including optical contacting, using soft metal seals such as gold and indium, and an alternative bonding technique known as frit sealing.

3. The non-optical components—the usual configuration is to have one cathode and two anodes which produce a discharge when a high voltage is applied to these electrodes. This discharge then provides the source of excitation for the laser action. The usual laser wavelengths that are used are either the red line at 632.8 nm or the 1.152 μm line in the infrared part of the electromagnetic spectrum.
4. The biasing mechanism required to overcome the lock-in phenomenon described in detail above. A bias can be applied by various techniques such as mechanical dither, magnetic mirror or the use of optical elements within the laser cavity.

A photograph of a mechanically dithered ring laser gyroscope is shown in Figure 5.4.

The primary disadvantage of this technology is the precision engineering that is required to make and polish the faces of the laser block and the high technology required to produce the mirrors. This tends to make the cost of the sensor quite high, although techniques for reducing this are being sought. A further anticipated problem is the potential for helium to leak out of the cavity through one of the many seals. Radio frequency pumping of the laser cavity has been demonstrated and is a method of reducing the number of components fitted into the block and hence reducing the number of orifices and seals through which this gas can leak.

Mirror quality assessment prior to assembly of the sensor is crucial to the performance and yield achieved in production of these devices. It is usual to evaluate

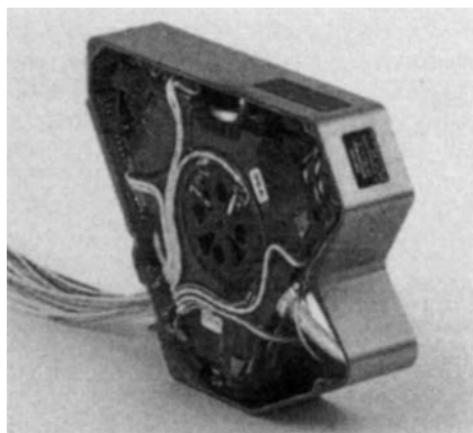


Figure 5.4 Mechanically dithered ring laser gyroscope

scatter, loss, surface quality and flatness. The two mirror parameters that are most closely related to sensor performance are the scatter and loss. Deterioration of the mirrors is minimised by operating the gyroscope at the lowest possible internal laser intensity consistent with reliable performance.

More recent developments use an optical arrangement that has four independent beams in the cavity. Suitable comparison between the frequencies of the different beams enables the lock-in phenomenon to be eliminated without the use of mechanical motion (dither). The Northrop-Grumman (formerly Litton Industries) optically biased, four-beam ring laser gyroscope, termed the Zero-Lock™ Laser Gyroscope, is shown in Figure 5.5.

5.1.3.5 Sources of error

There are three types of error which are characteristic of a ring laser gyroscope:

1. the lock-in phenomenon considered in detail above;
2. null shift, where the input/output characteristic does not pass through the origin so that the sensor records some counts from the detector even when stationary;
3. scale-factor changes resulting from mode pulling effects.

A null shift error arises when one of the laser beams experiences some difference in its optical path when compared with the other laser beam. Hence, the use of a split discharge and the balancing of the discharge currents in the two discharges in order to make the laser cavity as isotropic or reciprocal as possible. Similarly, the sensor is usually shielded from stray magnetic fields in order to minimise any unwanted magneto-optic effects, particularly in the mirrors.

Mode pulling effects, considered by Lamb [13], give rise to dispersion effects, normal or anomalous. Any changes in the dispersive effects of the laser medium can

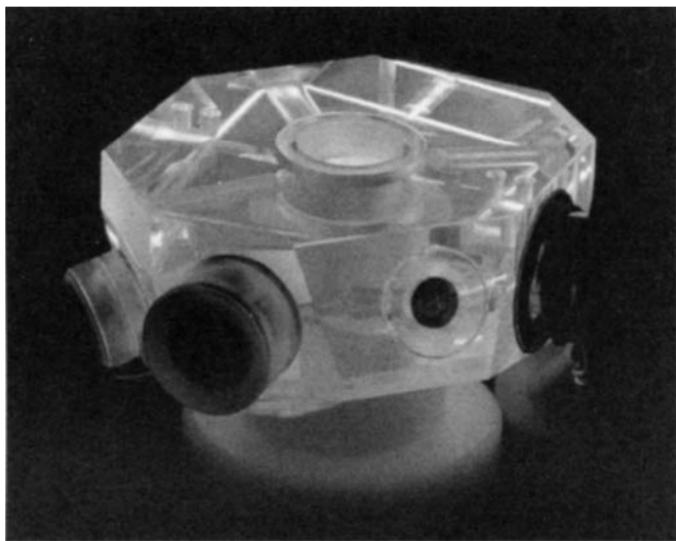


Figure 5.5 Zero-Lock™ Laser Gyroscope (published courtesy of Northrop Grumman Corporation, Litton Systems)

give rise to instabilities and continuous changes to the scale-factor. All of these errors are considered in detail by Aronowitz [6].

The stability of the sensing axis is also a key parameter which influences system performance. This is defined by the plane containing the laser beams which can move owing to disturbances in the laser block and movements of the beam induced by non-parallel motion of the cavity path length control mirror.

The output of a ring laser gyroscope ($\tilde{\omega}_x$) may be expressed mathematically in terms of the input rate ω_x and the rates about the axes which lie in the lasing plane (ω_y and ω_z) as:

$$\tilde{\omega}_x = (1 + S_x)\omega_x + M_y\omega_y + M_z\omega_z + B_x + n_x \quad (5.16)$$

where S_x is the scale-factor error, M_y , M_z are the misalignments of the gyroscope lasing plane with respect to the nominal input axis, B_x is the fixed bias and n_x is the random bias error.

The random bias term includes the random walk error referred to earlier which gives rise to a root-mean-square magnitude of angular output which grows with the square root of time. Whilst present to some extent in mechanical gyroscopes, the effect is generally an order of magnitude larger in optical sensors. In a mechanically dithered ring laser gyroscope, this error is largely caused by the random phase angle error introduced as the input rate passes through the lock-in region. An additional noise term gives rise to a bounded error and is the result of scale-factor errors in the mechanism used to eliminate lock-in.

5.1.3.6 Typical performance characteristics

With careful design, this form of gyroscope does not exhibit any significant acceleration or vibration sensitivity. The typical range of performance that can be achieved from these devices is as follows:

g -Independent drift (bias)	<0.001–10°/h
g -Sensitive bias	Usually insignificant for most applications
g^2 -Sensitive bias	Usually insignificant for most applications
Scale-factor errors	Few parts per million to 0.01% (of maximum rotation rate)
Bandwidth	>200 Hz (can be made very large)
Maximum input rate	Several thousand degrees per second
Random walk	0.001–0.01°/h

Hence, the key parameters are the bias repeatability, random noise, scale-factor repeatability and sensing axis stability.

5.1.4 Three-axis ring laser gyroscope configuration

Various schemes have been proposed and implemented to produce a single sensor with three sensitive axes using ring laser technology. Such devices are commonly called triads. These configurations are generally based on the use of three mutually orthogonal square laser cavities within a single cubic block. This arrangement enables each mirror to be shared by two of the laser cavities so only six mirrors are required for this device [14]. Similarly, the cathode is shared between discharges. The use of mechanical dither, applied about a body diagonal of the laser block, enables a bias to be applied simultaneously to each of the individual sensors within the monolith, and hence alleviate the lock-in problem for each of the axes.

Use of such a configuration can be very attractive for a number of applications as it provides great stability between the axes. The cost and complexity can also be reduced compared with the use of three independent sensors by using one dither mechanism, one discharge circuit and reducing the number of mirrors required through sharing.

The major disadvantage of such a system is the care needed and difficulty in machining the laser block to the necessary accuracy as well as avoiding damage during production, that is, achieving a high yield. Additionally, a single fault could mean that all three axes of angular motion information is lost.

A schematic diagram of a triad is shown in Figure 5.6.

5.1.5 Fibre optic gyroscope

5.1.5.1 Introductory remarks

Work at the US Naval Research Laboratories in the late 1960s suggested that multiple circulations of a Sagnac interferometer may give sufficient sensitivity to enable angular rotation to be detected and measured. By the mid-1970s, research progressed

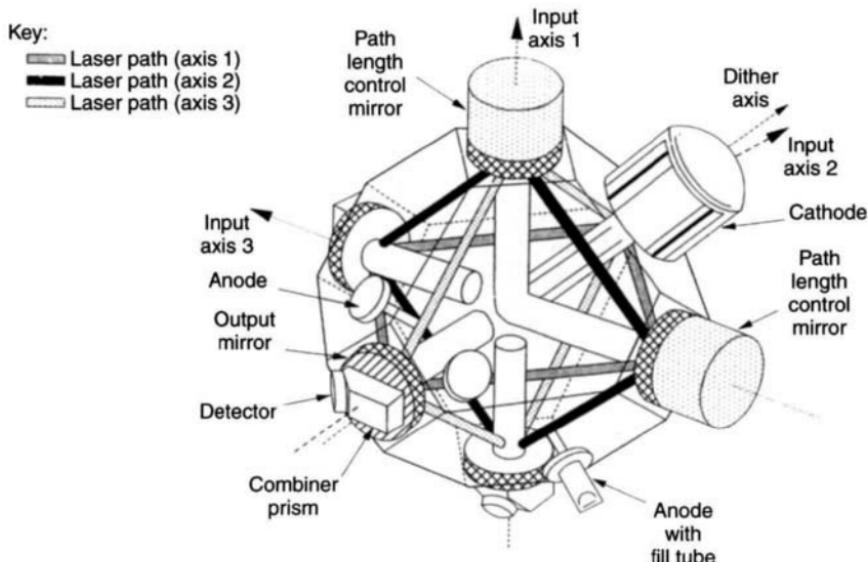


Figure 5.6 Schematic diagram of a triad

significantly on the use of passive interferometric techniques to sense angular motion, by applying optical fibre technology to form the light path [5]. This approach was seen as offering a far cheaper alternative to the ring laser technique, as the need to machine and polish surfaces to fractions of an optical wavelength would not be required. However, it was recognised that this approach was thought unlikely to produce a sensor with true high performance inertial performance characteristics, that is, drift values in the region of $0.01^\circ/\text{h}$ or better. Application of modern technology and technique has enabled this goal to be achieved.

In contrast with the ring laser technology, the fibre optic gyroscope senses angular motion by detecting the phase difference between the two beams passing round the light path in opposite directions. The gyroscope can be constructed as a genuine solid-state sensor, even in a closed loop mode, by the use of integrated optical components (chips). Use of this technology means that this type of inertial instrument can be very compact. However, extreme care and good design is required with the necessary fibre connections to avoid failure in harsh environments. Currently this is the subject of various research projects in many parts of the world [15].

These sensors have found many applications, particularly in the robotics and automobile industries. Aerospace applications are developing, especially for stabilisation and inertial navigation.

5.1.5.2 Principle of operation

Operation of the fibre optic gyroscope is dependent on the formation of a Sagnac interferometer [16]. In its simplest form, light from a broad band source is split into two beams that propagate in opposite directions around an optical fibre coil.

These two beams are combined at a second beam splitter to form an interference pattern where the resultant intensity is observed using a photo-detector. When the interferometer is stationary, the path length of the two counter-rotating beams is identical so there is no phase difference resulting in maximum amplitude. However, when the fibre coil is rotated about an axis normal to the fibre coil, the light travelling in the same direction as the rotation travels slightly further than the light travelling in the opposing direction. The resulting phase difference results in a change in amplitude of the interference pattern formed when the two beams are recombined.

For a rotating fibre gyroscope with a single turn of fibre, the phase difference ($\Delta\Phi$) between the counter-propagating beams of light may be expressed in terms of the path difference (ΔL) generated when the device rotates as:

$$\Delta\Phi = 2\pi \frac{\Delta L}{\lambda} \quad (5.17)$$

Substituting for ΔL from eqn. (5.6) gives:

$$\Delta\Phi = \frac{8\pi A\Omega}{c\lambda} \quad (5.18)$$

where A is the area enclosed by the fibre coil, Ω is the applied rotation rate and c is the velocity of light.

For a coil of N turns, this becomes:

$$\Delta\Phi = \frac{8\pi AN\Omega}{c\lambda} \quad (5.19)$$

This may be expressed in terms of the length of the fibre ($L = 2\pi RN$) as:

$$\Delta\Phi = \frac{4\pi RL\Omega}{c\lambda} \quad (5.20)$$

Consider a coil of radius 40 mm containing a 100 m length of fibre. If the optical wavelength is 850 nm, the phase differences which occur for rotation rates of (a) 15°/h and (b) 500°/s are as follows:

- (a) $\Delta\Phi = 0.0008^\circ$
- (b) $\Delta\Phi = 98.6^\circ$

Clearly, if a sensor is going to be capable of detecting Earth's rate or comparable rotations, a high level of dimensional stability will be necessary. Hence, light travelling one way around the fibre coil must travel exactly the same path as the light which travels in the opposite direction, that is, reciprocity must be maintained.

Comparing this equation with eqn. 5.11 for the ring laser gyroscope, the difference in sensitivity between the two sensors is obvious owing to the occurrence of the velocity of light in the denominator of the above equation. Hence, it is necessary to measure minute phase shifts to achieve high performance, which is a non-trivial task, but application of modern techniques enables this measurement accuracy to be achieved.

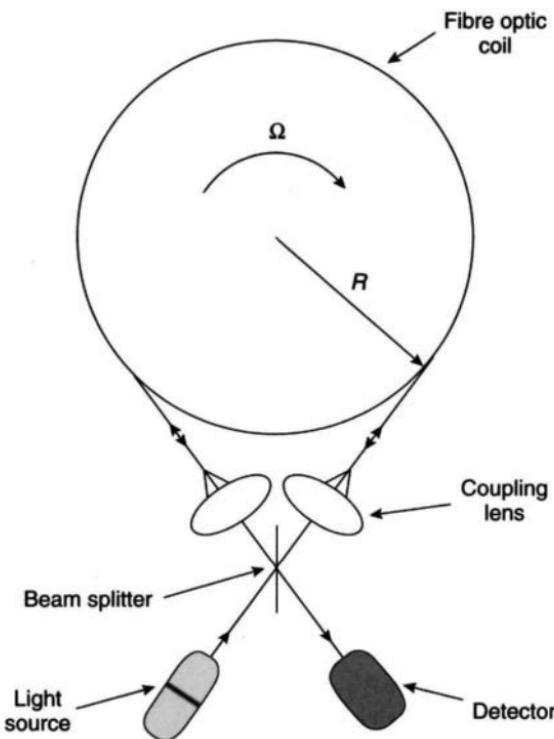


Figure 5.7 Open-loop fibre optic gyroscope

5.1.5.3 Detailed description of the sensor

The fundamental optical components of a fibre optical gyroscope, as illustrated schematically in Figure 5.7, are:

1. A light source, usually a broad band source with a coherence length chosen to minimise the scattering effects within the fibre.
2. Couplers to link energy into and out of the fibre. It is usual to use 3 dB couplers so they act as beam splitters.
3. The fibre coil, the angular motion sensing element. As a small single coil is unlikely to provide sufficient sensitivity, multiple turns are used. Depending on the desired sensitivity, high birefringence mono-mode or polarisation maintaining fibre may be used.
4. The detector, a photo-diode used to detect the changes in the fringe pattern.

The non-optical components include the former on which the fibre coil is wound and the electronic components.

The fibre optic gyroscope can be operated in either an open-loop mode or a closed loop mode [16–18]. When it is used in the simple open-loop configuration, it is particularly sensitive to any non-reciprocal effects, consequently reducing the sensitivity of the device.

Open-loop operation

A scheme was devised that ensured that both of the returning waves had propagated along the identical path, but in opposite directions. This was achieved by projecting polarised light into the interferometer through a single mode waveguide, such as a mono-mode optical fibre, and observing the returning interference wave which had been filtered by the same waveguide prior to detection. This arrangement is known as the reciprocal or minimum configuration gyroscope, and is shown schematically in Figure 5.8.

The returning light beams from the fibre coil are combined at the second beam splitter and emerge from the so-called reciprocal port. These two beams are exactly

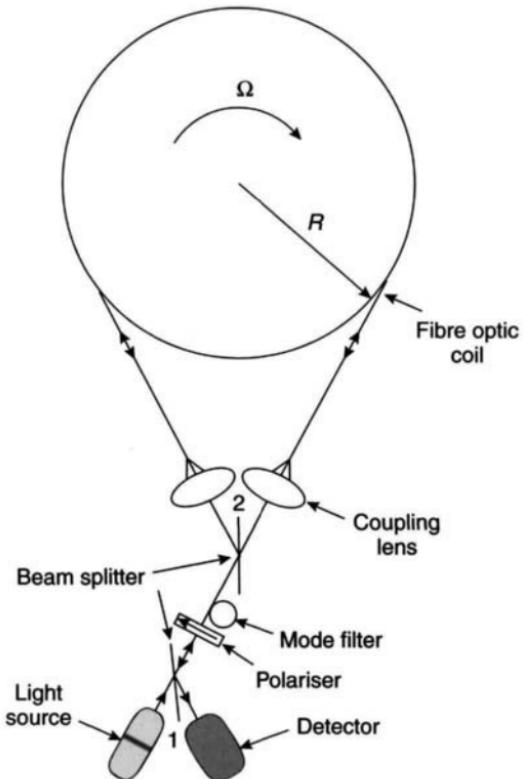


Figure 5.8 Reciprocal configuration fibre gyroscope and detector response

in phase when the fibre coil is at rest, but the resultant intensity varies sinusoidally with the rate of angular rotation of the coil. The major disadvantage of this form of fibre gyroscope is the lack of sensitivity at small applied input rates, owing to the co-sinusoidal shape of the fringe pattern, as shown in Figure 5.8.

It is possible to modify the fringe pattern to enhance the sensitivity at low rotation rates by incorporating a phase modulator at one end of the fibre coil. This modulator acts like a delay line. It is operated asymmetrically to give a phase dither of $\pm\pi/2$, which appears at the detector at twice the modulation frequency. Consequently, the gyroscope (detector) output is now biased to give its greatest sensitivity at and around small rotation rates. However, the response is still sinusoidal. This is illustrated in Figure 5.9 which also shows a schematic diagram of a phase biased fibre gyroscope.

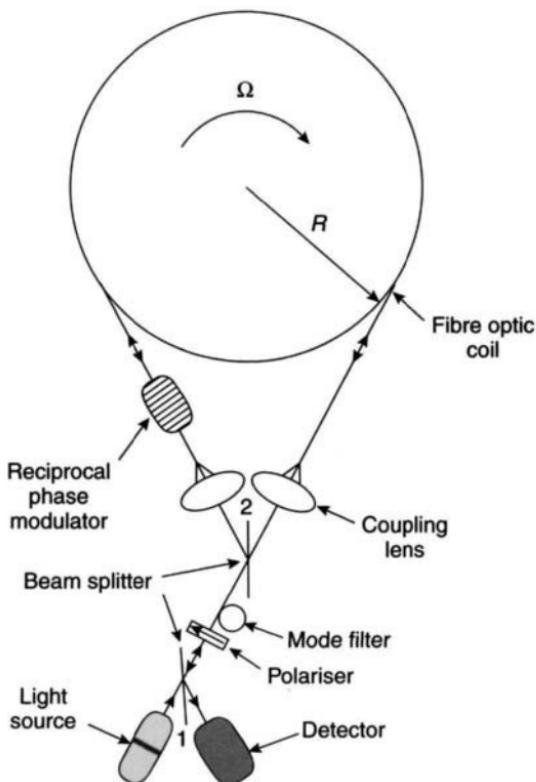


Figure 5.9 Phase biased fibre gyroscope and detector response

The phase modulator can be made by winding a few turns of optical fibre round a piezoelectric cylinder. A square wave signal may be applied to this cylinder to make it change shape and consequently to modulate the optical path length of the fibre coil.

Closed loop operation

There are many applications that require good accuracy over a wide angular rate range, possibly up to hundreds of degrees per second, and not merely at or close to 'zero rate'. It is generally desirable that the scale-factor linking the perceived angular motion to the actual motion should have good stability, be linear and be independent of the returning optical power. This can be achieved using a closed loop signal processing approach. Two techniques, or architectures, have been demonstrated, namely phase nulling and frequency nulling.

In the case of phase nulling, a second phase modulator is added to the 'other end' of the fibre coil, as shown in Figure 5.10. It is operated at twice the frequency of the dither (biasing) modulator and is used to 'back-off' or null the effects of the Sagnac induced phase shifts caused by the angular motion of the fibre coil. The open-loop signal is used as an error signal to generate an additional phase difference ($\Delta\phi_n$) that has an opposite sign to the rotation induced phase difference. Consequently, the total phase difference is arranged to be at or very close to the 'zero' value, so that the system is operated about the point of greatest sensitivity. The value of the additional feedback ($\Delta\phi_n$) is used as a measurement of angular rate. It has a linear response with good stability, since it is independent of the power of the returning optical signal and the gain of the detection system. However, the accuracy of the scale-factor does depend on the stability of the source wavelength and the geometric stability of the sensing coil.

An alternative technique for achieving closed loop operation is through the use of a frequency shift generated by an acousto-optical modulator, or Bragg cell, placed at one end of the sensor coil. This frequency shift is used to produce a differential phase shift to null that caused by the Sagnac effect [15]. By varying the voltage applied to the Bragg cell, the frequency shift it induces can be varied. Hence, the voltage that needs to be applied to the Bragg cell to null the detector's output is directly proportional to the applied angular motion.

The frequency shift given to the light by the Bragg cell is chosen so that the sensor is operated in the region where it is most sensitive to low rotation rates, as described above. This frequency shift produced by the Bragg cell is dithered about the centre frequency using a square wave modulator and the intensities of the light beams incident on the detector are monitored. When the sensor coil is rotated about its input axis, a small phase shift is introduced between the two beams giving rise to a mismatch between the intensities of the two beams on the detector. Consequently, the output signal from the detector will be modulated at the dither frequency. A phase sensitive detection system is used to deduce the amount that the centre frequency of the Bragg cell needs to be altered in order to return the intensities to their original matched state, thereby nulling the sensor. This change in the voltage applied to the Bragg cell is directly proportional to the applied angular motion.

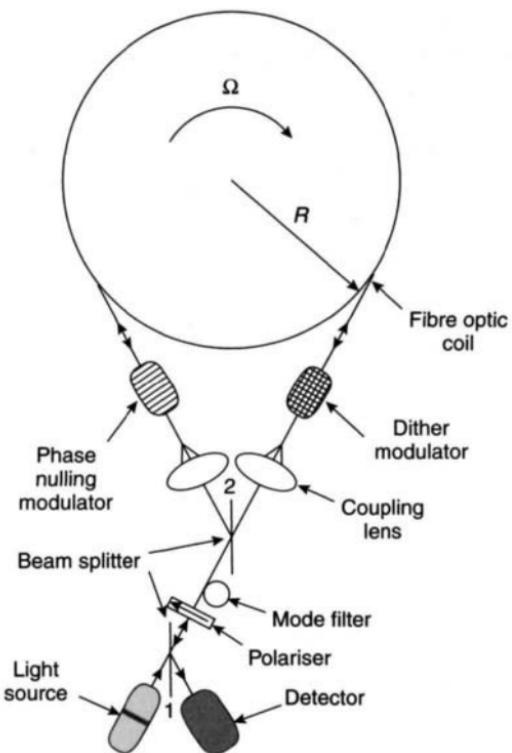


Figure 5.10 Closed loop (phase nulled) fibre gyroscope and detector response

The principal drawback of this architecture is the generation of a suitable bandwidth of frequencies in the modulator. This can be achieved by using two acousto-optical modulators at opposite ends of the sensor coil and dithering their frequencies about a centre frequency. An alternative approach is to use two modulators in opposition at one end of the fibre coil. With both arrangements, great care is required to achieve satisfactory mechanical stability of the whole assembly.

One technique that is currently used is the so-called phase ramp, serrodyne modulation [16], which relies on the fact that a frequency can be considered to be a time derivative of phase. In practice, a sawtooth waveform is used to modulate the applied phase shift, with a very fast 'flyback' at the reset positions. An alternative method that alleviates the 'flyback' problem is the digital phase ramp. In this case, 'phase steps' are generated with a duration equal to the group delay difference in time between

the long and short paths that connect the phase modulator and the beam splitter. These ‘phase steps’ and the resets can be synchronised with a square-wave biasing modulation. The use of digital logic enables this technique to be implemented easily.

One of the current developments of the fibre optic gyroscope is the demonstration of the so-called integrated fibre optic gyroscope. In this device, all the bulk optical, or fibre, components are replaced with components fused into a lithium niobate substrate [16]; it is used to produce the beam splitter or couplers, optical waveguides and the necessary modulation to the light required to measure the rotation rate accurately. Fibre ‘leads’ are used to connect the source and detector to this so-called integrated optics ‘chip’. This form of gyroscope has the potential to be compact, rugged and have a long shelf-life whilst retaining the advantages of optical sensors listed earlier in Section 5.1. Currently these sensors are about 70–80 mm in diameter; a size which is a compromise between producing a compact sensor and avoiding excessive biases induced by the strain in the fibre when bent into a small radius of curvature. Photographs of a commercially available fibre optic inertial measurement unit are shown in Figure 5.11.

As perceived, the use of a reciprocal path in optical sensors like the fibre optic gyroscope should be free from errors associated with the environment, such as temperature, acceleration and vibration. However, regrettably these devices do exhibit some sensitivity to these effects. Fortunately, through careful design, particularly the winding of the sensor coil, these sensitivities can be minimised. These effects are considered below.

5.1.5.4 Sources of error

Changes in ambient temperature can cause a bias or drift to be observed because of a multitude of effects within the sensor, temperature gradients within the sensor being a particular problem. As the ambient temperature changes, the source wavelength changes and the sensitivity of the sensor is inversely proportional to the source wavelength. Temperature changes result in variation of the refractive index of the optical fibre leading to changes in the modulation. If possible, the expansion coefficient of the fibre and the coil former should be well matched, or else differential stresses are induced by thermal expansion which result in measurement errors. Thermal changes also alter the size of the coil resulting in changes in the scale-factor of the gyroscope.

A bias occurs when there is a time-dependent thermal gradient along the optical fibre as a result of a temperature gradient across the coil. This results in a non-reciprocity occurring when corresponding wave fronts of the counter-rotating beams cross the same region at different times. This is known as the Shupe effect [19]. Anti-Shupe windings have been devised so that parts of the optical fibre that are equal distances from the centre of the coil are adjacent.

When an acceleration is applied to a coil, it can result in the distortion of the coil producing a change in the scale-factor of the gyroscope. Distortions also change the birefringence of the fibre and hence, the bias of the sensor. Additionally, distortions can lead to changes in the direction of the sensitive axis and, in the case of a three-axis configuration, changes in the relative orientation or alignment of the three input axes.

(a)



(b)

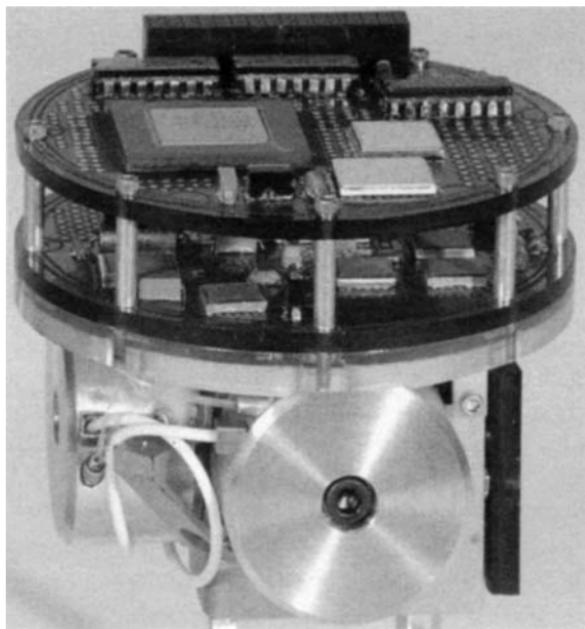


Figure 5.11 Inertial measurement unit containing fibre optic gyroscopes (published courtesy of Northrop Grumman Corporation, Litton Systems)

Application of vibratory motion to a coil of fibre, or length of fibre, can lead to a distortion of the coil and the fibre, depending on the amplitude of the input motion. As discussed earlier, this leads to errors in the angular motion measurements made by the sensor.

The presence of stray magnetic fields also can have an adverse effect owing to interaction with the non-optical components. Magnetic fields can also produce changes in the state of polarisation of the light in the optical fibre through the Faraday effect. These sensitivities lead to a bias in the output signal from the device. The use of magnetic shielding can minimise this effect.

The bias and drift in the output signal is a consequence mainly of birefringence in the optical fibre and propagation of cladding modes, as well as polarisation modulation of the optical signals. The problem is essentially one of many modes existing within the fibre, each with a different phase and with a lack of coherence, so fading occurs. One method of overcoming spurious coherent effects, including scattering within the fibre, is to use a low coherence source, i.e. one with a short coherence length. A superluminescent diode fulfils this criterion. Use of this source also minimises the bias generated by the Kerr electro-optic effect, caused by changes in the refractive index of the optical medium through variations in the power of the two counter-propagating beams.

5.1.5.5 Typical performance characteristics

Typical performance parameters for fibre optic gyroscopes are given below:

g -Independent bias	0.5–50°/h
g -Sensitive bias	~1°/h/g
g^2 -Sensitive bias	~0.1°/h/g ²
(g -Dependent biases can be made negligible with good design but often show some sensitivity)	
Scale-factor errors	0.05–0.5%
Bandwidth	>100 Hz
Maximum input rate	>1000°/s

Many techniques have been developed to reduce the sensitivity of these sensors to the environment in which they operate and it is anticipated that significant progress will be demonstrated in the near future. It is worth noting that these ‘problems’ are far less severe than those which occurred with the mechanical gyroscopes 30 years earlier!

5.1.5.6 Recent developments in fibre optic gyroscope technology

Fibre optical gyroscope technology has developed significantly over the last 5–10 years. The closed loop device interferometric fibre optical gyroscope (IFOG) has seen very substantial increases in performance owing to:

- the developments in the super luminescent diode technology;
- enhanced optical fibres with lower optical loss, lower scatter and greater uniformity;

- the perfection of the functions of the integrated circuit chip, such as the polarisation of the light, the splitting of the beam into the clockwise and counter-clockwise beams, the recombination and control of the frequency shifting process;
- the development of microprocessor technology that can be combined within the sensor to give real-time accurate compensation of systematic errors.

Modern technology enables each sensor to be calibrated and appropriate compensation techniques applied, particularly for the effects of changes in temperature. Moreover, it allows the same architecture to be used in different performance sensors, by reducing the length of optical fibre needed for the lower performance devices. However, the modulation frequency still has to be tracked to the actual length of fibre to give the optimum performance for the chosen fibre length.

The measurement accuracy of these IFOGs is now approaching, if not comparable with, that of a standard inertial-quality ring laser gyroscope. The IFOG devices have yet to supersede the ring laser gyroscope in the high performance applications owing to the industrial investment in this technology base. However, as the cost of the IFOG technology decreases they are very likely to replace ring laser gyroscopes. A high-performance IFOG can have the following performance characteristics.

Parameter	Value
Bias stability	<0.0003°/h
Random walk	<0.00008°/√h
Scale-factor error	<0.5 ppm

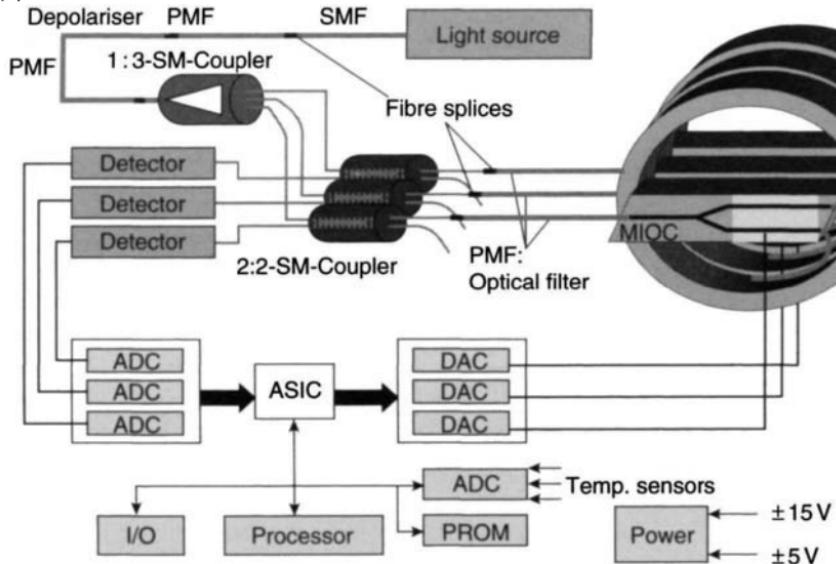
IFOG technology has been used in many lower-grade applications. Examples include: unmanned air vehicles, unmanned underwater vehicles, many types of stabilisation, gyrocompasses, and attitude and heading reference systems.

The major thrust is cost and size reduction. A typical example is the development of a fibre optical gyroscope triad system, a system being produced by the Litef company in Germany. In this system the cost of the super luminescent diode was identified as a major cost driver in any FOG-based system. The solution was to use a single super-luminescent diode and share the light with the three fibre coils via their integrated optical chips. This class of optical device is reported to have a bias stability in the 0.05°/h class. Changes to the coil design enables lower-grade performance to be achieved, but still maintaining a very high-quality linear scale-factor. A schematic of this arrangement is shown in Figure 5.12.

5.1.6 Photonic crystal optical fibre gyroscope

New optical fibre technology [20] is being developed that gives superior light-guiding properties compared with traditional step-index fibres. These new optical fibres have a periodic array of holes in the structure that provide very high-quality light-guiding

(a)



(b)

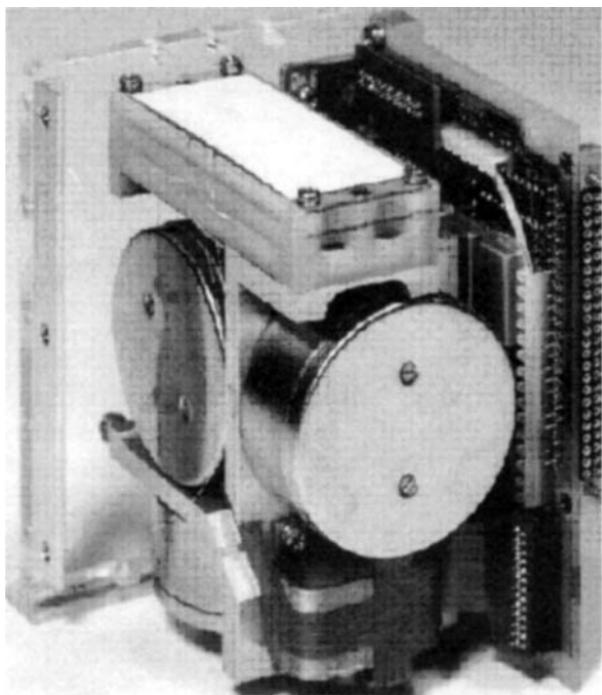


Figure 5.12 (a) Fibre optical gyroscope triad structure. (b) Sensor block (published courtesy of Litef GmbH)

properties with very low optical losses. In fact they are sometimes known as holey fibres and come in two forms depending on the construction of the core:

- photonic crystal fibres have a solid core, to transmit the light, that is surrounded by an array of holes that give a reduced refractive index to the surrounding medium;
- photonic band gap devices with a void or defect in the centre of the core for the transmission of light, so the light passes along a hollow core. The array of holes creates a band gap analogue to a semiconductor allowing the light to propagate in only certain parts of the structure, trapping it in others.

These classes of fibre provide very tight mode confinement of the propagating light, and single mode propagation is possible over many wavelengths. Additionally, the polarisation maintaining versions of these fibres have demonstrated ten times the birefringence of the conventional fibres.

The tight mode confinement results in much smaller bend losses so tighter coils can be produced leading to much smaller packages. Additionally, the smaller cladding structure also allows small devices to be made without impeding performance.

It is possible to embody dispersion compensation into these fibres to reduce the effects of spectral distortion of the propagating light on the performance of the sensor. Additionally, the micro-structured fibres could be used with light having a wavelength in the 1–2 μm waveband.

Figure 5.13 shows two types of micro-structured optical fibre.

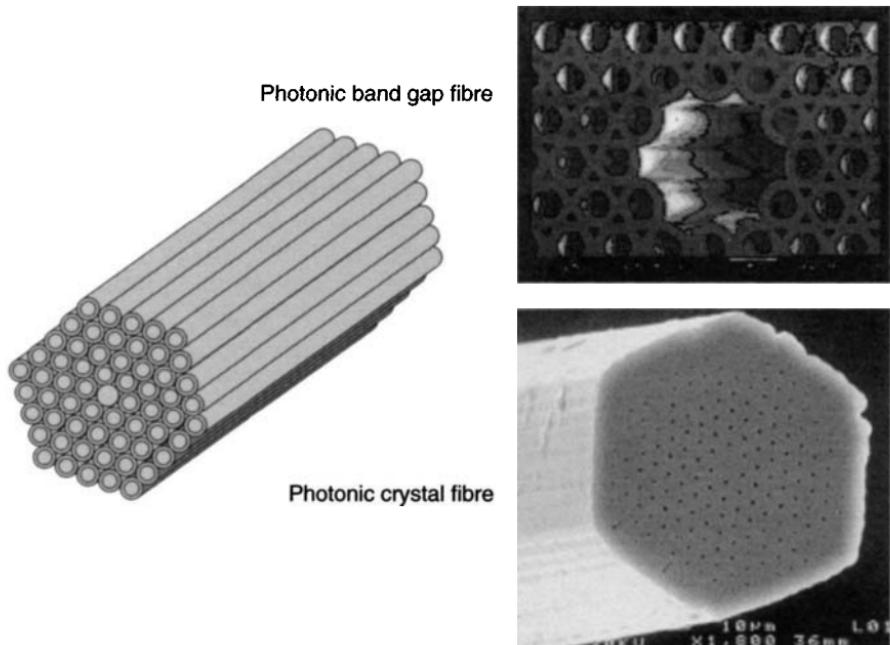


Figure 5.13 Micro-structured fibres

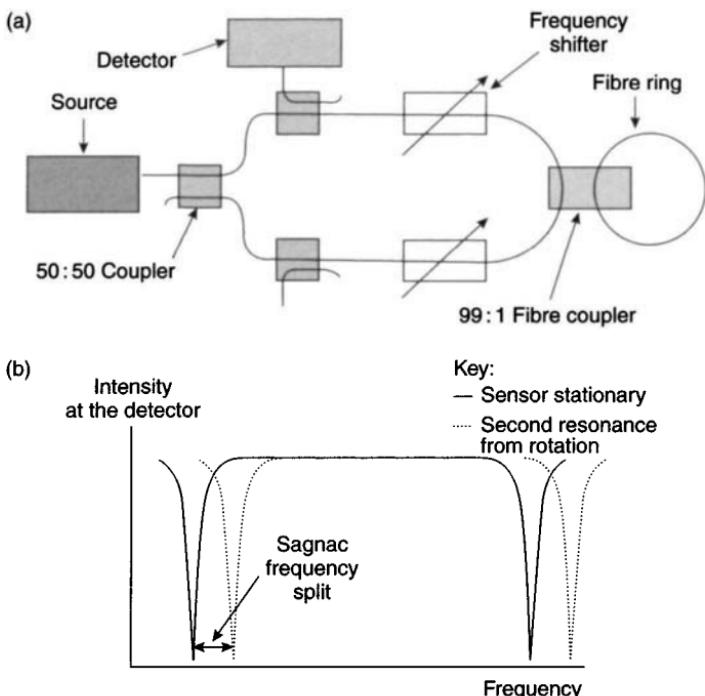


Figure 5.14 Fibre optic ring resonator gyroscope (a) and intensity waveform (b)

5.1.7 Fibre optic ring resonator gyroscope

This is yet another implementation of the Sagnac effect and could be considered to be a development of the fibre optic gyroscope. Work to develop this concept started at various institutions in the early 1980s, much of the pioneering work being led by Dr Ezikiel at the Massachusetts Institute of Technology.

This type of gyroscope uses a recirculating ring resonant cavity to enhance the phase differences induced by the Sagnac effect when the sensor is rotated. A ring cavity is formed by joining the ends of a short optical fibre cable, typically 10 m or less, to a coupler, as shown in Figure 5.14a. The light enters and leaves the ring cavity through this coupler.

The principle of operation of the ring cavity is very similar to that of a Fabry–Perot interferometer [16], but with multiple interference between the recirculating waves and the light being introduced to the cavity. Resonance can be achieved in the ring either by stretching the fibre or by sweeping the frequency of the light, thus resulting in constructive interference occurring in the ring and hence maximum light intensity.

At resonance, there is a dip in the light intensity reflected at the output port resulting from the majority of the light entering the fibre ring and then being lost by scattering after many recirculations in the ring cavity. Figure 5.14b shows the form of the intensity waveform at the detector. The width of the resonance is set by the finesse of the resonator [21]. This parameter is inversely proportional to the coupling ratio into

the ring. Hence, it is necessary to have a high coupling ratio. The quality of the cavity, and its sensitivity, is defined by the finesse, which is the ratio of the free spectral range to the linewidth of the resonance. The resonant condition is observed by detecting the energy reflected by the coupler linking the fibre ring to the source and detector.

When the sensor is stationary, the resonant frequencies of the two counter-propagating beams are identical. When the ring is rotated about its sensitive axis, the resonant frequencies of the two counter-propagating beams differ. This results in a shift in the positions of the resonances and hence, in the light intensity from the output port.

Equation (5.19), derived earlier, for the fibre optic gyroscope gives the phase difference for the two beams. Because the two counter-propagating beams experience different loop paths, there is a different resonance frequency (ν_d) generated in each direction. It is given in terms of the radius of the coil by substitution for $A (= \pi R^2)$ and $L (= 2\pi R)$ in eqn. (5.11). Hence:

$$\nu_d = \frac{2R\Omega}{\lambda} \quad (5.21)$$

where R is the radius of the coil, Ω is the applied rate and λ is the wavelength.

Clearly, the number of turns of fibre does not influence its sensitivity, as this is dependent on the radius of the loop for a given wavelength and rotation rate. Hence, given practical considerations, the length of optical fibre required is very much less, typically of the order of 10 m.

This sensor is used in a feedback mode so the frequency of the two counter-propagating beams is shifted when the sensor is rotated, thus keeping the two beams at resonance simultaneously even under rotation. Sinusoidal phase modulation with frequency modulation can be applied to the two counter-propagating beams. The difference between the two modulations required to maintain resonance in the two beams is proportional to the applied rotation rate.

The fundamental requirement for the light energy used in this sensor is that it has a narrow bandwidth and a high coherence. Hence, a light source is needed that is very stable in order to maintain the narrow width resonance in the ring. Currently, the lasers that fulfil this requirement are expensive, but with the development of quantum well lasers the price should reduce. Careful design is required to avoid incoherent back-scattering from one beam to the other and the consequent onset of the familiar 'lock-in' problem. It is usual to control the polarisation of the light before it enters the ring to avoid intensity fading.

As in the case of the fibre optic gyroscope, the bulk optical components can be replaced with integrated optical materials using guided waves. Hence this, together with the very short length of fibre, typically 10 m or less, can lead to a very compact sensor.

Currently, this sensor is being developed by a number of companies but definite performance and error data are not available. However, it is expected to have a similar performance to the fibre gyroscope; data for that sensor are given in Section 5.1.5.5. Generally, it is considered to be more sensitive than the fibre gyroscope, possibly by a factor of 3 [22].

5.1.7.1 *Ring resonator rotation rate sensors – non-linear mode*

Currently the measurement of rotation rate by the exploitation of non-linear effects, induced in the energy in the ring resonator, has been investigated. These non-linear effects occur when there is only a small loss, of the order of 5 per cent, in the ring, leading to about a 20-fold increase in power. The non-linear interactions are usually divided into two groups, intensity dependent refractive index effects and scattering phenomena. In the case of the fibre optic gyroscope, these non-linear effects are considered detrimental as they degrade performance.

The non-linear refractive index effect makes use of the Kerr effect [8]. The electric field of the optical energy propagating through a medium produces a refractive index change. Now, when a ring resonator, operating in a resonant condition, is rotated, the intensity in one direction decreases whilst the intensity in the other direction decreases. Additionally, the refractive index in each direction also changes, amplifying the intensity change. This amplification can be varied by changing both the parameters of the resonant ring and the Kerr material, which, in turn, improves the sensitivity of the resonator.

The propagation of a high power optical beam through a dense medium produces phonons; energy is transferred to these phonons resulting in a change in frequency of the beam. This results in a scattered optical wave with a modified frequency. The two major types of non-linear scattering are called Brillouin and Raman [23], and are threshold effects; Brillouin occurs at lower power levels. Brillouin scattering has a characteristic linewidth which requires a source with a coherence length shorter than the life time of a phonon. Consequently, a coherent source produces Brillouin scattering in preference to Raman.

It has been suggested [24] that a ring laser, exhibiting Brillouin scattering, pumped from both directions will produce a frequency difference between Brillouin scattered energy that is directly proportional to the rotation rate. This sensor should produce a signal that is similar to that produced by a ring laser gyroscope, that is, a rotation rate dependent frequency. However, there should not be any lock-in as the two scatter mechanisms should be independent.

A sensor has also been proposed based on the Raman scatter phenomena using high power pulses from a mode locked laser [25]. It appears that the major drawback will be the size of the sensor. Currently, there are no reports indicating that any of these sensors have been demonstrated.

5.1.8 *Ring resonator gyroscope*

This sensor is very similar to the fibre optic ring resonator, but the fibre ring is replaced with an optical waveguide which can be etched into a suitable substrate [26]. The principle of operation is very similar to that described for the fibre optic ring resonator. Typical rings are about 50 mm in diameter. Hence, when the problems associated with scattering and coupling can be overcome, the prospect of a gyroscope on a 'chip' can become a reality, along with all the advantages of optical sensors.

This type of sensor has been developed by Northrop in the United States [7] and is known as the micro-optic gyroscope (MOG). Rugged sensors with a

diameter of about 25 mm have been produced with performance in the 1–100°/h category.

5.1.9 Integrated optical gyroscope

There have been developments in this highly desirable sensor, as it is a ‘gyroscope on a chip’. The sensing element is an optical waveguide on a substrate with the light travelling in opposite directions. The relative position of the resonance within the ring is a measure of the applied optical rotation rate about an axis that is perpendicular to the ring. The fundamental operation of this device is described more fully in Section 5.1.7.

These integrated optical gyroscopes are created on wafers and combine micro-machined electromechanical system processes and integrated optical fabrication processes. Speciality glasses are formed by highly specialised techniques such as flame hydrolysed deposition, radio frequency and reactive sputtering. The flame hydrolysed deposition allows the refractive index of the medium to be controlled so that the glass can be doped with rare earth laser ions, such as neodymium, ytterbium or erbium, creating lasers and laser amplifiers within the chip.

The definition of the waveguide is still quite a challenge, as vertical sidewalls with very low roughness are required to minimise losses. Clearly, the etching process is potentially the limiting process in the fabrication as there are likely to be various compositions of glass with varying dopant concentrations.

Currently, the performance goal is to meet applications requiring performance in the 0.1–1°/h regime. This class of sensor offers significant size and mass reduction, of the order of a factor of 20 compared with conventional fibre gyroscopes. Additionally, the power consumption is likely to be reduced by a factor of 5 or 6 along with a significantly lower cost owing to the reduced number of optical components.

5.2 Cold atom sensors

5.2.1 Introduction

This is an approach or technique for measuring inertial properties; such devices are also known as atom interferometer devices and owe much to the excellent work of Professor Chu of Stanford University. The technique is in the early stages of practical development and offers the prospect of a route to the most accurate accelerometers, gyroscopes, precision clocks and gravity gradiometers therefore performance enhancements of several orders of magnitude are predicted. If successful devices can be devised then, given precise knowledge of the gravity field, there is the prospect of a sub 10 m/h navigation capability without using global positioning system (GPS) or other external aiding techniques.

The sensor relies on the super cooling of an atom or molecule by techniques such as laser cooling and uses the de Broglie wavelength of an atom, which is about 3×10^4 smaller than the wavelength of visible light. The physical principle relies on

diameter of about 25 mm have been produced with performance in the 1–100°/h category.

5.1.9 Integrated optical gyroscope

There have been developments in this highly desirable sensor, as it is a ‘gyroscope on a chip’. The sensing element is an optical waveguide on a substrate with the light travelling in opposite directions. The relative position of the resonance within the ring is a measure of the applied optical rotation rate about an axis that is perpendicular to the ring. The fundamental operation of this device is described more fully in Section 5.1.7.

These integrated optical gyroscopes are created on wafers and combine micro-machined electromechanical system processes and integrated optical fabrication processes. Speciality glasses are formed by highly specialised techniques such as flame hydrolysed deposition, radio frequency and reactive sputtering. The flame hydrolysed deposition allows the refractive index of the medium to be controlled so that the glass can be doped with rare earth laser ions, such as neodymium, ytterbium or erbium, creating lasers and laser amplifiers within the chip.

The definition of the waveguide is still quite a challenge, as vertical sidewalls with very low roughness are required to minimise losses. Clearly, the etching process is potentially the limiting process in the fabrication as there are likely to be various compositions of glass with varying dopant concentrations.

Currently, the performance goal is to meet applications requiring performance in the 0.1–1°/h regime. This class of sensor offers significant size and mass reduction, of the order of a factor of 20 compared with conventional fibre gyroscopes. Additionally, the power consumption is likely to be reduced by a factor of 5 or 6 along with a significantly lower cost owing to the reduced number of optical components.

5.2 Cold atom sensors

5.2.1 Introduction

This is an approach or technique for measuring inertial properties; such devices are also known as atom interferometer devices and owe much to the excellent work of Professor Chu of Stanford University. The technique is in the early stages of practical development and offers the prospect of a route to the most accurate accelerometers, gyroscopes, precision clocks and gravity gradiometers therefore performance enhancements of several orders of magnitude are predicted. If successful devices can be devised then, given precise knowledge of the gravity field, there is the prospect of a sub 10 m/h navigation capability without using global positioning system (GPS) or other external aiding techniques.

The sensor relies on the super cooling of an atom or molecule by techniques such as laser cooling and uses the de Broglie wavelength of an atom, which is about 3×10^4 smaller than the wavelength of visible light. The physical principle relies on

the fact that atoms in any medium have mass and internal structure and that atomic interferometers are extremely sensitive. This method has some similarities with the NMR gyroscope described in Section 4.5.1, or use could be made of the Sagnac effect to detect the angular rotations experienced by the sensor through variation in atomic motion round a closed path.

In research studies to date, devices have used incoherent atoms propagating in free space. However, in the future it may be feasible to use coherent Bose–Einstein condensates propagating in a guiding medium, achieved by using laser-cooling techniques.

Laser cooling of atoms and atom-trapping techniques are being used in a number of applications in many scientific areas. An example is atom interferometry, where an atom is placed into a superposition of two or more spatially separated atomic states, which will interfere with each other if they are subsequently brought back together.

This technique has enabled atom interferometers to be created and have now been established as very sensitive techniques for the detection and measurement of inertial forces. These sensors have been used to make measurements of gravity gradients, local gravity and rotations [27–29]. These devices are still being developed in the laboratory and use laser beams to cool, manipulate and control the associated atomic wave packets.

One possible application for this technology is the observation of very small changes in the gravity vector to detect changes in the density of the Earth's surface. This could be used in prospecting for oil deposits or for monitoring the development of underground facilities.

5.2.2 *Rotation sensing*

Very accurate and stable measurement of rotation has been achieved [28] using an atom interferometer based on the principle of having two counter-propagating atomic beams, as used in the Sagnac effect with ring laser gyroscopes. This device uses stimulated Raman transitions, to manipulate atomic wave packets of caesium atoms. The device uses counter-propagating high-flux atomic beams of caesium to form two interferometers with opposite Sagnac phase shifts that share key components, such as the Raman atomic state manipulation laser beams. Subtraction of the interferometric signals from the two separate atomic beams allows the common-mode rejection of spurious noise sources and various systematic effects.

In this sensor the interferometer is arranged in a Mach–Zehnder configuration [30], in which caesium atoms from a thermal atomic beam are arranged to be the interfering particles. The laser light pulses act as beam splitters and mirrors for the atoms to form a closed cavity, allowing the two counter-propagating beams to travel round their defined closed paths. The light pulses are used to put the atoms in a superposition of two states, corresponding to different spatial trajectories that separate and recombine enclosing a given area of the Sagnac device, as shown in Figure 5.15.

The two counter-propagating caesium atomic beams are transversally laser-cooled and optically pumped before entering a 2 m long magnetically shielded interaction region of the vacuum chamber. Three sets of light pulses, from two photon-stimulated Raman transitions, serve to divide, deflect and recombine the atomic wave packets.

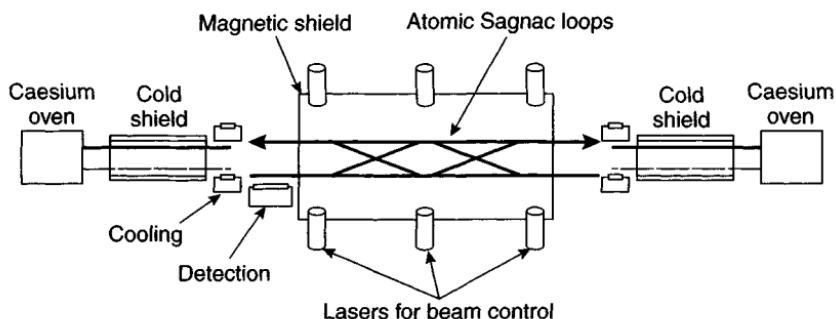


Figure 5.15 Cold atom sensor – Sagnac loop

The Interference signal is observed by detecting the number of atoms in the $F = 4$ ground state by resonant fluorescence.

Results from the device indicate a short-term stability in the region of $3 \times 10^{-9} \text{ rad/s}/\sqrt{\text{Hz}}$.

5.2.3 Measurement of acceleration

An atom interferometer, based on an atomic fountain of laser-cooled caesium atoms and using laser light to ‘form’ the atom optical components, has been used to make a very accurate measurement of ‘g’. This atom interferometer uses optical pulses of light to stimulate transitions between two internal states of the atoms that initially drift apart and then recombine after a second pulse of light encourages a second transition in the internal states of the atoms, which complements the initial transitions. A further illumination pulse with the appropriate phase relative to the atomic phase encourages a further complementary transition.

The phase difference between the two parts of the interferometer is a function of the phases of the laser at various positions and times in the cavity at the start of the optical pulse, the frequency of the light and the optical paths in the interferometer. In this interferometer the frequency of the light is changed in a phase-continuous way, so that it remains resonant with the transitions as the atoms accelerate under the influence of gravity. As a consequence the phase difference between the two paths in the interferometer is proportional to the gravitational attraction.

In the atomic fountain, shown in Figure 5.16, caesium atoms are extracted from a low-pressure background vapour and loaded into a magneto-optical trap during a 600 ms period. The magnetic fields are turned off and the captured atoms are launched into the atomic fountain of this sensor using a specialised technique, known as moving polarisation gradient optical molasses. During this period further cooling of the ‘launched’ atoms occurs, using resonant techniques and in the final stage of the launch the laser intensities are reduced to zero in 400 μs , so that the atoms are adiabatically cooled.

The launched atoms are subjected to a series of pulses that place the atoms in a specific internal state with an effective vertical temperature of 10 nK. This low

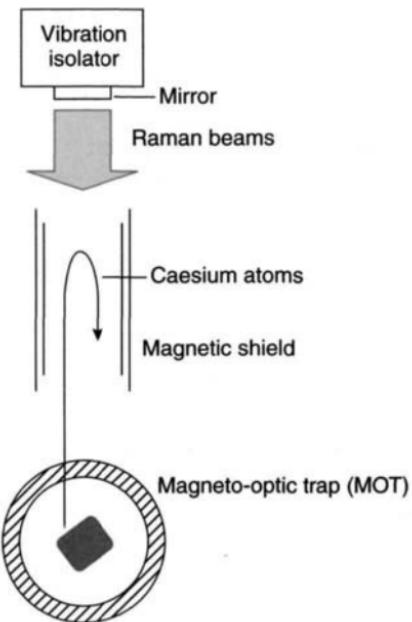


Figure 5.16 Atom interferometer – overview of experimental set-up

velocity spread leads to a high fringe contrast over a period of about 150 ms. The interferometer measurement occurs in a magnetically shielded region.

This type of device is capable of measuring ' g ' to better than a part per billion accuracy.

5.2.4 Gravity gradiometer

Gravity gradiometers are particularly useful sensors for detecting changes in the gravity vector close to the surface of the Earth. The changes may indicate a number of features that may be present in the Earth's shell owing to local changes in the density. The local changes to the gravity vector may be caused by many reasons; it may indicate a number of features, such as the position of oil reserves, mineral deposits, voids or other phenomena.

An atom interferometry technique has been used to create a gravity gradiometer (Figure 5.17), using two laser-cooled and trapped sources of caesium atoms, as described above and a pair of vertically propagating laser beams. The device is arranged so that two independent measurements of acceleration can be made using the two vertically separated ensembles of caesium atoms in freefall, under the influence of gravity.

The caesium atoms are launched into a vertical trajectory from the magneto-optical trap and conditioned to be in a particular internal state using optical and microwave techniques. These atoms are then suitable for interacting with the gravity vector and

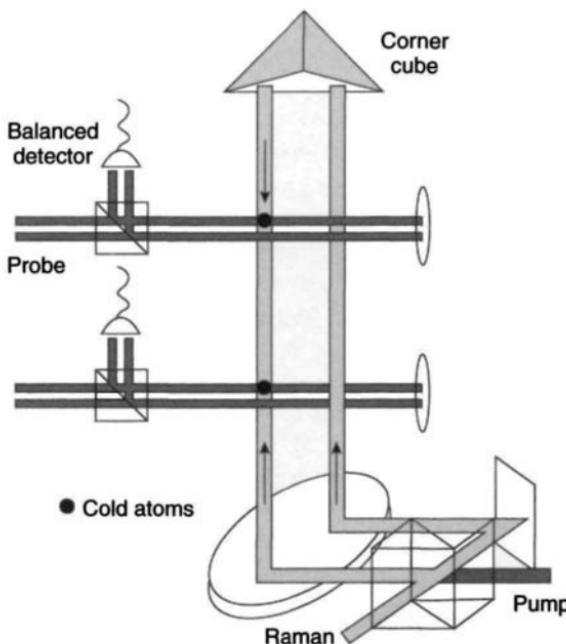


Figure 5.17 Cold atom gradiometer

then changes in the atomic states due to gravitational acceleration can be detected in the interferometer due to gravitational acceleration.

This cold atom device is particularly sensitive as it uses atoms as proof masses, rather than macroscopic objects. This eliminates variability from device to device and provides insensitivity to many environmental perturbations, such as temperature gradients and magnetic fields. Moreover, the two absolute accelerometers used in this device are operated in a differential mode to give enhanced performance.

A light-pulse atom interference method is used to measure the acceleration of each atomic ensemble with respect to a reference frame defined by the wave fronts of the interrogating optical fields. The two simultaneous measurements of the effects of gravity on the pair of vertically separated sensors are made with respect to the same set of Raman laser fields. This is achieved by a simultaneous measurement of the fraction of atoms excited by the laser pulse sequence at the two positions where the gravity vector is measured. The differential acceleration is given by the differential phase shift between the upper and lower atomic ensembles, and this difference in phase shift is proportional to the difference in the mean value of ' g ' measured at the two parts of the sensor.

The difference between the measured acceleration experienced by each atom ensemble, divided by their separation, is a measure of the in-line component of the gravity-gradient tensor. Accelerations of the common reference frame, defined by the optical field wave fronts, are rejected as a common mode in the differencing process.

The gradiometer references its calibration to the wavelength of the measurement laser, which is locked to an atomic spectral line, thus providing absolute accuracy and long-term stability. The propagation axes of these laser beams are aligned to pass through both ensembles of atoms. Moreover, the sensitive axis is defined by the Raman propagation vector, and the measurement of acceleration is referenced to only one retro reflector, consequently, the two accelerometers may be separated by some distance without destroying the common mode vibration-rejection features. Increasing the separation between the two accelerometers linearly increases the sensitivity to gravity gradients and gives insensitivity to near-field perturbations.

High signal-to-noise ratio is vital to give high interferometer sensitivity. The device uses a balanced modulation transfer technique to reduce laser-induced detection noise and differentiate cold atoms from thermal background atoms [31]. Additionally, each interference fringe is recorded a number of times in consecutive cycles of the observation, typically 15. The number of points per scan is kept small to reduce sensitivity to long-term drifts in signal amplitude and fringe contrast. The predominant noise source in the observations is atom shot noise; this is the Poissonian fluctuation that occurs from detecting atoms in coherent superposition states.

The differential performance is of the order of $4 \times 10^{-9} g/\sqrt{\text{Hz}}$.

Other ‘atomic techniques’ are being investigated for this sensor, such as:

- Interferometric devices using diffraction in the Raman–Nath regime [32], using short intense pulses of light applied to the atomic ensemble.
- Large area interferometers using adiabatic transfer of momentum [33] have been used in proof-of-principle experiments, in this case the atoms are put into coherent superpositions of two states using a microwave pulse.
- Use of the a.c. Josephson effect in arrays of Bose–Einstein condensed atoms [34], where condensate atoms tunnel from an array of vertically spaced lattice sites and atoms tunnelling from different sites subsequently interfere. The resulting interference pattern is a periodic train of atomic pulses whose frequency is a function of the strength of the gravitational potential.

Direct gradient measurements can be made using multiple-loop Raman-pulse based interferometers such as a double loop or a figure-of-eight loop [29], by modifying the sequence of phase pulses applied to the ensemble.

5.3 Summary of gyroscope technology

There are many types or classes of sensor that can be used to sense or detect angular motion. Many of these devices have been considered in the foregoing text, particularly those that are used currently, or could be applied in the future, in strapdown applications. These instruments range from the conventional mechanical gyroscopes, using a rotating mass, to the unconventional, using atomic spin.

A great deal of effort has, and still is, being expended to develop the so-called novel technology, the aim being to produce a ‘gyroscope’ on a chip. New technology used in industry, such as robotics, is helping to sustain this effort. However, this is also

being applied to the conventional technology and is helping to keep the mechanical gyroscope competitive.

The range of accuracy that can be achieved from the spectrum of rotation sensing devices spans many orders of magnitude. Some sensors have a bias of less than 0.0001°/h, whilst others are in the 1°/s class or worse. Most sensors show some unwanted sensitivity to the environment in which they operate. The goal of much research is to reduce these sensitivities or improve the ruggedness of the instruments as many, particularly some high-precision devices, are quite sensitive to vibratory motion.

A summary of typical performance characteristics for a range of sensors suitable for strapdown application is given in the following table.¹

Characteristic	RIG	DTG	Flex gyroscope	DART/ MHD	Vibratory gyro	RLG	FOG
<i>g</i> -Independent bias (°/h)	0.05–10	0.05–10	1–50	360–1800	360–1800	0.001–10	0.5–50
<i>g</i> -Dependent bias (°/h/g)	1–10	0.01–10	1–10	180	36–180	0	<1
Anisoelastic bias (°/h/g ²)	1–2	0.1–0.5	0.05–0.25	18–40	18	0	<0.1
Scale-factor non-linearity (%)	0.01–0.1	0.01–0.1	0.01–0.1	0.5–0.1	0.2–0.3	5–100	0.05–0.5
Bandwidth (Hz)	60	100	100	100/80	500	>200	>100
Maximum input rate (°/s)	>400	1000	>500	800/400	>1000	>1000	>1000
Shock resistance	Moderate	Moderate	Moderate	Moderate	>25 000g	Good	Good

In general, a significant amount of precision engineering and high technology is required to produce a device that is functional. As the accuracy required from the sensor increases, so does the precision and the size needed to fulfil the requirement, although this is not universally true. In some of the recent research programmes, an effort has been made to alleviate the need for ultra high precision for the high accuracy instruments. Usually, however, this has led to the need to apply very high technology, such as superconductivity, which imposes its own demands such as cooling to cryogenic temperatures.

The performance of developed inertial sensors for near-term applications requiring angular motion data is shown in Figure 5.18. The figure shows the

¹ These are typical values applicable over the range of parameters stated. In many cases, the values given could be improved. However, it is not normally possible to have all the best case values in a single unit, particularly for the conventional sensors. These values are only for general indicative purposes.

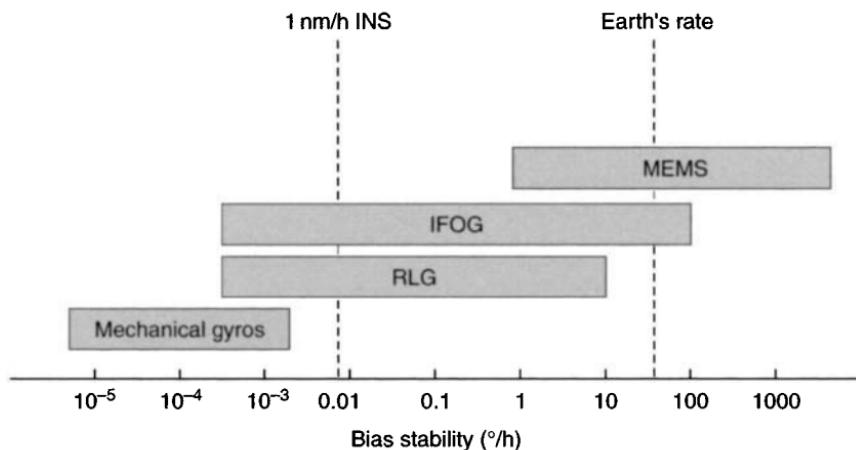


Figure 5.18 Near-term gyroscope performance summary

anticipated division between the micro-machined electromechanical system (MEMS, see Chapter 7), optical (ring laser gyroscope (RLG) and IFOG) and mechanical gyroscopes covering the low, medium and high accuracy system applications respectively. Despite advances in optical sensors, the high performance applications (10^{-4} – $10^{-5} \text{ }^{\circ}/\text{h}$) remain the regime of the mechanical gyroscope. For the mid-range applications requiring very high scale-factor stability, the ring laser gyroscope is the sensor of choice. In the longer term, it is expected that MEMS gyroscope performance will continue to improve and find increasing application in higher performance systems.

References

- 1 SAGNAC, G.: ‘L’ether lumineux demonstre par l’effet du vent relatif d’ether dans un interferometre en rotation uniforme’, *Comptes Rendus de l’Academie des Science*, 1913, **157**, p. 708; ‘Sur la preuve de la realite de l’ether lumineux par l’experience de l’interferograph tournant’, *Comptes Rendus de l’Academie des Science*, 1913, **157**, p. 1410
- 2 MICHELSON, A.H., and GALE, H.G.: ‘The effect of the Earth’s rotation on the velocity of light’, *Astrophysical Journal*, 1925, **61**, p. 140
- 3 MAIMAN, T.H.: *Nature*, 1960, **187**, p. 493; *Physical Review*, 1961, **123**, p. 1145; ‘Optical and microwave experiments in ruby’, *Physics Review Letters*, 1960, **4**, p. 564
- 4 MACEK, W.M., and DAVIS, D.T.M.: ‘Rotation rate sensing with travelling wave ring lasers’, *Applied Physics Letters*, 1963, **2**, p. 67
- 5 VALI, V., and SHORTHILL, R.W.: ‘Fibre ring interferometer’, *Applied Optics*, 1976, **15**, p. 1099
- 6 ARONOWITZ, F.: in ROSS, M. (Ed.): ‘Laser applications vol 1 – The laser gyro’ (Academic Press, 1971)

- 7 LAWRENCE, A.W.: 'Modern inertial technology – navigation, guidance and control' (Springer-Verlag, 1998, 2nd edn.)
- 8 YOUNG, M.: 'Optics and lasers' (Springer-Verlag, 1992)
- 9 PERLMUTTER, M.S., BRESMAN, J.M., and PERKINS, H.A.: 'A low cost tactical ring laser gyroscope with no moving parts'. DGON Proceedings, *Gyro Technology Symposium*, Stuttgart, 1990
- 10 BRECHMAN, J., COOK, N., and LYSOBURY, D.: 'Differential laser gyro development', *Journal of the Institute of Navigation*, 1977, **24** (2), p. 153
- 11 DICKINSON, M.R., and KING, T.A.: 'Polarisation frequency splitting in non-planar ring laser resonators', *Journal of Modern Optics*, 1987, **34** (8), p. 1045
- 12 KNIPE, C., ANDREWS, D.A., and KING, T.A.: 'Mode interactions in four-frequency degeneracy-lifted ring lasers', *Journal of Modern Optics*, 1988, **35** (3)
- 13 LAMB, Jr., W.E.: 'Theory of an optical maser', *Physical Review A*, 1964, **134**, p. 1429
- 14 SIMMS, G.J.: 'Ring laser gyroscopes'. US Patent 4 407 583, 1983; UK Patent GB 2076213B4, 1980; 'A triaxial laser gyroscope', *Proceedings Institute of Mechanical Engineers, Mechanical Technology of Inertial Devices*, 1987, C56/87
- 15 SMITH, R.B.: 'Fibre optic gyroscopes 1991 – a bibliography of published literature', SPIE, 1585 *Fibre Optic Gyros: 15th Anniversary Conference*, 1991, p. 464
- 16 LEFEVRE, H.: 'The fibre-optic gyroscope' (Artec House, 1993)
- 17 BURNS, W.: 'Optical fibre rotation sensing' (Academic Press, 1994)
- 18 CULSHAW, B., and GILES, I.P.: 'Fibre optic gyroscopes', *Journal Physics E: Scientific Instruments*, 1983, **16**, p. 5
- 19 SHUPE, D.N.: 'Thermally induced non-reciprocity in the fibre-optic interferometer', *Applied Optics*, 1980, **19** (5), p. 654
- 20 YABLANOVIĆ, E.: 'Inhibited spontaneous emission in solid-state physics and electronics', *Physical Review Letters*, 1987, **58** (20), pp. 2059–2062
- 21 MALVERN, A.R.: 'The optical fibre ring resonator gyroscope'. Symposium proceedings, 'Technology of inertial sensors and systems', Institution of Mechanical Engineers, 1991
- 22 GILES, I.P.: 'Optical fibre ring resonator gyroscopes'. Symposium proceedings, *Lasers and fibre optic gyros*, Royal Aeronautical Society, 1987
- 23 MILLS, D.L.: 'Non-linear optics' (Springer-Verlag, 1991)
- 24 THOMAS, P.J., VAN DRIEL, H.H., and STEGEMAN, G.I.A.: 'Possibility of using an optical fibre Brillouin ring laser for inertial sensing', *Applied Optics*, 1980, **19**, p. 12
- 25 NAKAZAWA, M.: 'Synchronously pumped fibre Raman gyroscope', *Optics Letters*, 1985, **10**, p. 4
- 26 BERNARD, W., ENGLERT, R., et al.: 'Waveguide ring resonators in glass for optical gyros'. DGON Proceedings, *Gyro Technology Symposium*, Stuttgart, 1986
- 27 PETERS, A., et al.: 'Measurement of gravitational acceleration by dropping atoms', *Nature*, 1999, **400**, pp. 849–852

- 28 GUSTAVSON, T.L., *et al.*: 'A dual atomic beam matter-wave gyroscope', *Proceedings SPIE*, 1998, **3270**, pp. 62–69
- 29 McGUIRK, J.M., *et al.*: 'Sensitive absolute gradiometry using atom interferometry', *Physics*, 2001
- 30 FRANCON, M.: 'Optical interferometry' (Academic Press, 1966)
- 31 McGUIRK, J.M., *et al.*: 'Low noise detection of ultra cold atoms', *Optics Letters*, 2001, **26** (6), pp. 364–366
- 32 CAHN, S.B., *et al.*: 'Time domain de Broglie wave interferometry', *Physical Review Letters*, 1997, **79** (5), pp. 784–787
- 33 FEATONBY, P.D., *et al.*: 'Separated-path Ramsey atom interferometry', *Physical Review Letters*, 1998, **81** (3), pp. 495–499
- 34 ANDERSON, B.P., and KASEVICH, M.A.: 'Macroscopic quantum interference from atomic tunnel arrays', *Science*, 1998, **282** (5394), pp. 1686–1689

Chapter 6

Accelerometer and multi-sensor technology

6.1 Introduction

As described in Chapter 1, inertial navigation relies upon the measurement of acceleration which can be integrated successively to provide estimates of changes in velocity and position. Measurements of acceleration are used in preference to direct measurements of velocity or position because velocity and position measurements require an external reference whilst acceleration can be measured internally.

The form of construction of devices which may be used to sense acceleration may be classified as either mechanical or ‘solid-state’. The technology of mechanical sensors is well established [1, 2] and devices capable of sensing acceleration over a wide accuracy range, from 50 milli-*g* down to a few micro-*gs* and to a similar level of resolution, are currently available. There have been significant advances in the development of solid-state sensors in recent years, particularly with silicon technology.

The concept of using a single instrument to measure acceleration and angular motion has been the subject of research for a number of decades and during the 1980s was developed by a number of institutions and companies. This device has become known as the multi-sensor and has tended to be based on either vibratory technology or gyroscopic mass unbalance technology. Evaluation of this technology has generally shown it to be capable of providing estimates of linear acceleration and angular motion compatible with sub-inertial¹ navigation applications.

6.2 The measurement of translational motion

The translational acceleration of a rigid body, resulting from the forces acting upon it, is described by Newton’s second law of motion. A force *F* acting on a body of mass

¹ The term sub-inertial is sometimes used when describing system performance for short duration navigation systems. Typically, sub-inertial systems use gyroscopes and accelerometers with measurement biases of the order of $1^\circ/\text{h}$ and 1 milli-*g* (1σ), respectively.

m causes the body to accelerate with respect to inertial space. This acceleration (a) is given by:

$$F = ma \quad (6.1)$$

Whilst it is not practical to determine the acceleration of a vehicle by measuring the total force acting upon it, it is possible to measure the force acting on a small mass contained within the vehicle which is constrained to move with the vehicle. The small mass, known as a proof or seismic mass, forms part of an instrument called an accelerometer. In its simplest form, the accelerometer contains a proof mass connected via a spring to the case of the instrument as shown in Figure 6.1.

When the case of the instrument is subjected to an acceleration along its sensitive axis, as indicated in the figure, the proof mass tends to resist the change in movement owing to its own inertia. As a result, the mass is displaced with respect to the body. Under steady state conditions, the force acting on the mass will be balanced by the tension in the spring, the net extension of the spring providing a measure of the applied force, which is proportional to the acceleration.

The total force (F) acting on a mass (m) in space may be represented by the equation:

$$F = ma = mf + mg \quad (6.2)$$

where f is the acceleration produced by forces other than the gravitational field. In the case of a unit mass, $F = a = f + g$. The acceleration (a) may be expressed as the total force per unit mass. An accelerometer is insensitive to the gravitational acceleration (g) and thus provides an output proportional to the non-gravitational force per unit mass (f) to which the sensor is subjected along its sensitive axis. As described in Chapter 2, this is referred to as the specific force exerted on the sensor.

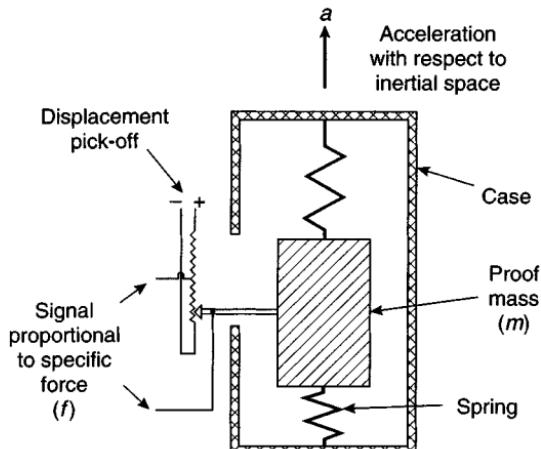


Figure 6.1 A simple accelerometer

Taking the case of an accelerometer which is falling freely within a gravitational field, the case and the proof mass will fall together with the result that there will be no net extension of the spring. Hence, the output of the instrument will remain at zero. In this situation, the acceleration of the instrument with respect to an inertially fixed set of axes, $a = g$ and the specific force is zero in accordance with the above equation. Conversely, in the situation where the instrument is held stationary, $a = 0$, the accelerometer will measure the force acting to stop it from falling. Following from eqn. (6.2), this force, $mf = -mg$, is the specific force required to offset the effect of gravitational attraction. It is clear therefore, that knowledge of the gravitational field is essential to enable the measurement provided by the accelerometer to be related to the inertial acceleration.

Many mechanical devices commonly used in present day inertial navigation systems for the measurement of specific force operate in a manner analogous to the simple spring and mass accelerometer described above.

In order to carry out the full navigation function, information is required about the translational motion along three axes, as described in Chapter 3. Commonly, three single-axis accelerometers are used to provide independent measurements of specific force, although multi-axis instruments can be used. It is common practice to mount the three accelerometers with their sensitive axes mutually orthogonal, although such a configuration is not essential, as will be discussed later.

The various principles of operation and performance of current accelerometer technology are reviewed in the following sections, covering both the mechanical and solid-state instruments. Later in the chapter, multi-sensors and angular accelerometers are reviewed in a similar way. Linear accelerometers may also be used to measure rotational motion [3]. However, owing to the need for very accurate measurements as well as precise sequencing and timing of the measurements, this technique is rarely used.

6.3 Mechanical sensors

6.3.1 Introduction

This is the broad division of sensors primarily described in Section 6.2 as mass-spring type devices. These sensors have been developed over many decades. Different construction techniques have been identified for use in different environments. Compact and reliable devices giving high accuracy and wide dynamic range have been produced in large quantities. The most precise force-feedback instruments are capable of measuring specific force very accurately, typically with resolutions of micro-g, or better. This class of mechanical sensors are used in both inertial and sub-inertial applications.

6.3.2 Principles of operation

As in the case of gyroscopes, accelerometers may be operated in either open or closed loop configurations. The basic principle of construction of an open-loop device is

as follows. A proof mass is suspended in a case and confined to a zero position by means of a spring. Additionally, damping is applied to give this mass and spring system a realistic response corresponding to a proper dynamic transfer function. When the accelerations are applied to the case of the sensor, the proof mass is deflected with respect to its zero or ‘null’ position and the resultant spring force provides the necessary acceleration of the proof mass to move it with the case. For a single-axis sensor, the displacement of the proof mass with respect to its ‘null’ position within the case is proportional to the specific force applied along its input, or sensitive, axis.

A more accurate version of this type of sensor is obtained by nulling the displacement of the pendulum, since ‘null’ positions can be measured more accurately than displacements. With a closed loop accelerometer, the spring is replaced by an electromagnetic device that produces a force on the proof mass to maintain it at its ‘null’ position. Usually, a pair of coils is mounted on the proof mass within a strong magnetic field. When a deflection is sensed, an electric current is passed through the coils in order to produce a force to return the proof mass to its ‘null’ position. The magnitude of the current in the coils is proportional to the specific force sensed along the input axis. The force-feedback type is far more accurate than the open-loop devices and is currently the type most commonly used in inertial navigation systems.

6.3.3 *Sensor errors*

All accelerometers are subject to errors which limit the accuracy to which the applied specific force can be measured. The major sources of error which arise in mechanical accelerometers are listed below. Further details relating to specific types of accelerometer will be given later in this chapter where the physical effects which give rise to each type of error are discussed more fully.

Fixed bias: This is a bias or displacement from zero on the measurement of specific force which is present when the applied acceleration is zero. The size of the bias is independent of any motion to which the accelerometer may be subjected and is usually expressed in units of milli-*g* or micro-*g* depending on the precision of the device involved.

Scale-factor errors: Errors in the ratio of a change in the output signal to a change in the input acceleration which is to be measured. Scale-factor errors may be expressed as percentages of the measured full scale quantity or simply as a ratio; parts per million (ppm) being commonly used. Scale-factor non-linearity refers to the systematic deviations from the least-squares straight line, or other fitted function, which relates the output signal to the applied acceleration.

Cross-coupling errors: Erroneous accelerometer outputs resulting from accelerometer sensitivity to accelerations applied normal to the input axis. Such errors arise as a result of manufacturing imperfections which give rise to non-orthogonality of the sensor axes. Cross-coupling is often expressed as a percentage of the applied acceleration.

Vibro-pendulous errors: Dynamic cross-coupling in pendulous accelerometers arises owing to angular displacement of the pendulum which gives rise to a rectified output when subjected to vibratory motion. This type of error can arise in any

pendulous accelerometer depending on the phasing between the vibration and the pendulum displacement. The magnitude of the resulting error is maximised when the vibration acts in a plane normal to the pivot axis at 45° to the sensitive axis and when the pendulum displacement is in phase with the vibration. This error may be expressed in units of g/g^2 .

As in the case of gyroscopic sensors, repeatability errors, temperature dependent errors, switch-on to switch-on variations and in-run errors arise in sensors of this type. Even with careful calibration, the residual errors caused by the unpredictable error components will always be present, restricting the accuracy of inertial system performance.

6.3.4 Force-feedback pendulous accelerometer

6.3.4.1 Detailed description of sensor

These devices are also known as restrained pendulum accelerometers. The main components of such a sensor are:

1. A pendulum, which has a proof mass attached to it or as an integral part of it.
2. A suspension mechanism or hinge element. This flexible member attaches the pendulum to the case and is usually either a flexible hinge or a pivot type arrangement.
3. A pick-off device to sense motion of the pendulum. It may use optical, inductive or capacitive techniques. The optical system may be very simple, a detector measuring the change in transmittance of a light beam through a slit in the pendulum. The inductive system involves measuring the differential current in coils fixed to the case interacting with a plate on the pendulum, which affects the mutual inductance of the coils. This system measures the relative position of the pendulum between the pick-off coils and not the 'null' position. In the case of a capacitive system, movement of the pendulum causes a change in capacitance between the faces of the pendulum and two electrodes in close proximity to the pendulum. This change is sensed using a bridge circuit.
4. A force re-balance mechanism to oppose any detected movement of the pendulum. This component usually takes the form of two identical poles of two magnets arranged centrally about the proof mass and a pair of coils mounted symmetrically on the pendulum. A current flowing in the coils generates an electromagnetic restoring force. This component is often referred to as the torquer.
5. The various components are usually hermetically sealed in a case. The case may be filled with a low viscosity oil to give resistance to shock and vibratory forces in both its quiescent and active states. Alternatively, the case may be filled with a dry gas such as air.

Such a sensor is shown schematically in Figure 6.2.

Displacement of the pendulum, which occurs in the presence of an applied acceleration, is sensed by the pick-off. In the most simple devices, this displacement provides a direct measure of the applied acceleration. Generally however, a device of

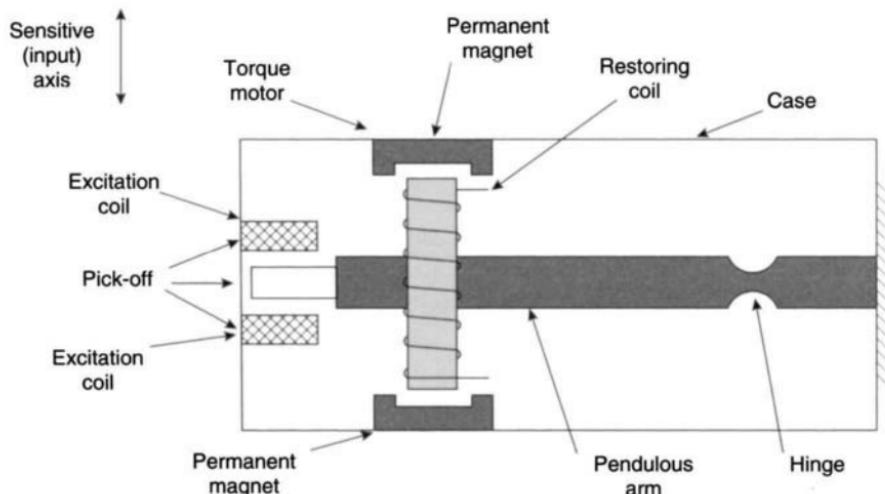


Figure 6.2 Force-feedback pendulous accelerometer

this type operates with an electronic re-balance loop to feed the pick-off signal back to the torquer. The electromagnetic force, produced by the torquer, acts to offset any displacement of the pendulum and maintain the pick-off output at zero. The current in the torquer coil is proportional to the applied acceleration. Operating the sensor in this mode means that the hinge is not under any bending stress.

6.3.4.2 Sources of error

Accelerometers of this type are capable of very high performance with good linearity, small biases and with a dynamic range in the region of 10^4 – 10^5 . This is a dimensionless quantity obtained by dividing the maximum acceleration which the sensor can measure by its resolution. The dominant sources of error are as follows:

measurement bias which arises as a result of residual spring torques and ‘null’ shift in the electrical pick-off device used;

scale-factor error, principally caused by temperature effects and non-ideal behaviour of components;

cross axis coupling which gives rise to a measurement bias when the sensor is under ‘g’ loading in the direction of the hinge axis or the pendulum axis, the latter being essentially a hinge interaction effect;

vibro-pendulous error which can give rise to a measurement bias under certain conditions when the sensor is subject to vibration along the sensitive and pendulum axes simultaneously;

random bias caused by instabilities within the sensor assembly.

Further errors occur in the measurements provided by pendulous accelerometers, such as those resulting from hysteresis effects, non-repeatability of bias and higher order scale-factor errors. Changes in the characteristics of the permanent magnets may

also change the scale-factor by a process known as ageing. This may be corrected by periodic recalibration.

The measurement provided by such sensor (\tilde{a}_x) may be expressed in terms of the applied acceleration acting along its sensitive axis (a_x) and the accelerations acting along the pendulum and hinge axes, a_y and a_z , respectively, by the equation:

$$\tilde{a}_x = (1 + S_x)a_x + M_y a_y + M_z a_z + B_f + B_v a_x a_y + n_x \quad (6.3)$$

where S_x is the scale-factor error, usually expressed in polynomial form to include non-linear effects, M_y , M_z are the cross axis coupling factors, B_f is the measurement bias, B_v is the vibro-pendulous error coefficient and n_x is the random bias.

6.3.4.3 Typical performance characteristics

Typical performance figures for the moderate accuracy sensors are as follows:

Input range	up to $\pm 100g$
Scale-factor stability	$\sim 0.1\%$
Scale-factor non-linearity	$\sim 0.05\%$ of full scale
Fixed bias	$0.0001g$ – $0.01g$
Bias repeatability	$0.001g$ – $0.03g$
Bias temperature coefficient	$\sim 0.001g/\text{ }^{\circ}\text{C}$
Hysteresis	$<0.002g$
Threshold	$\sim 0.00001g$
Bandwidth	up to 400 Hz

Most of these figures are improved significantly with the very high accuracy accelerometers. Biases as low as a few micro-g can be achieved with very high precision sensors, whereas those likely to experience high accelerations in very dynamic environments usually have a bias of a few milli-g.

6.3.5 Pendulous accelerometer hinge elements

The hinge element of a pendulous accelerometer is the component that enables the proof mass to move in one plane normal to the hinge axis. It must be stiff normal to the hinge line to maintain the mechanical stability of the hinge relative to the case under conditions of dynamic loading. However, it must be flexible about the hinge line and must minimise unpredictable spring restraint torques that cannot be distinguished from applied accelerations. The hinge must not be overstressed by either shock acceleration or vibratory motion. It must also return to its 'null' position exactly when the proof mass is displaced, in order to give the sensor good bias stability. Hinge elements exist that enable the proof mass to move in two orthogonal directions. These are essentially a complex combination of two single-axis elements as described in Section 6.3.6.

The two basic forms of hinge elements are flexures and pivots, there being several variations of each type.

6.3.5.1 *Flexure hinges*

The materials used to form the hinge are selected for their low mechanical hysteresis in order to minimise unpredictable spring torque errors. Hysteresis effects are minimised by choosing the hinge dimensions so that hinge stresses under dynamic forces, and pendulum movement, are well below the yield stress for the hinge material. A material that is commonly used is the alloy beryllium–copper since, because of the high ratio of its yield stress to its Young's modulus [4], it is capable of sustaining a large deflection without exceeding its yield stress. Fused quartz is another very suitable material. Some designs have both the pendulum and the hinge etched from a quartz substrate.

The main advantages of flexure hinges are that they exhibit very low static friction so offer almost infinite resolution and very low threshold. However, these hinges have a significant temperature dependent bias that requires calibration and compensation for the most accurate applications. Additionally, these hinges can be susceptible to damage from shock accelerations and also demand very tight tolerance, typically in the region of a micrometer, if the desired flexure compliance is to be attained.

6.3.5.2 *Jewelled pivot hinges*

This form of hinge supports the pendulum between a pair of spring-loaded synthetic jewel assemblies. The spring loading provides three-dimensional shock protection. These hinges have very small temperature dependent bias characteristics. However, stiction, under very quiescent conditions, can limit the resolution and wear of the pivots in very harsh vibratory environments can be a problem. This can be partially alleviated by the use of very hard materials on the bearing surface such as silicon carbide.

There are many applications, particularly those requiring a low maximum acceleration capability (20g or less), where jewel and pivot hinges can offer a cheaper and more sensitive instrument. However, with higher accelerations and high vibratory environments, flexure hinges tend to provide enhanced performance.

6.3.6 *Two-axes force-feedback accelerometer*

This form of instrument has many applications including some of the most demanding; such as ship's inertial navigation systems. This sensor has a pendulum which has freedom to swing about two orthogonal axes. Like the single-axis device, described earlier, it is restrained to its 'null' position by electrically energised coils working in a permanent magnetic field.

Clearly, it is necessary to have a hinge that constrains the pendulum to deflect about these two orthogonal axes, albeit by very small angles. Typically, the pendulum is attached rigidly to a plate at its top end, which is attached by two weak leaf springs to another plate. This second plate is attached to the case by means of a second pair of similar hinges, which are mounted at 90° to the first pair of hinges. Motion of the pendulum is often damped by filling the case with silicone fluid.

The principle of operation is identical to that described above for the single-axis sensor. Its performance is similar to that which can be obtained using the higher grade single-axis devices.

6.3.7 Open-loop accelerometers

A common form of this sensor is the mass–spring device of the type described in Section 6.2. Generally, these instruments are less stable and less accurate than the closed loop accelerometers. Undesirable characteristics inherent in open-loop accelerometers are sensitivity to supply voltage variations, non-linearity of the displacement caused by the applied acceleration and high thermal coefficients of bias and scale-factor. Consequently, they are generally inappropriate for most inertial navigation applications and therefore the mechanical variant will not be discussed further. Currently, however, an optical open-loop pendulous fibre optic accelerometer is being developed and the principle of operation will be discussed in the following section.

6.3.7.1 Optical fibre accelerometer

The fundamental principle of operation of this sensor is identical to the mechanical device. The major difference essentially lies in the form of the pick-off mechanism and pendulous mechanism which allows accelerations about two axes to be sensed. Optical fibres have excellent mechanical strength and elastic modulus characteristics, and additionally have negligible thermal expansion over the normal operating temperature range of inertial sensors, but need to be selected carefully to have isoelastic properties.

In this sensor, the pendulum is a length of fibre cable with a proof mass attached, together with a micro-lens at the bottom of the fibre and light from a solid-state laser coupled into the top. When an acceleration is applied to the case along any axis normal to the fibre, the bottom is deflected. Its displacement is sensed and measured by means of the laser light passing through the optical fibre and being focused on to a two dimensional photo-sensitive array. A suitable array is a charge coupled imaging device (CCID) which can provide both x and y coordinates of the displacement. A schematic diagram of a fibre optic pendulous accelerometer is shown in Figure 6.3.

Several parameters determine the range of this sensor, viz.

- size of proof mass;
- diameter of the optical fibre;
- length of optical fibre;
- height of the suspension point above the photo-sensitive detectors;
- size of the photo-sensitive array.

Currently, the accuracy is limited by the pixel density of the photo-sensitive array. Performance data are not currently available.

6.4 Solid-state accelerometers

During recent years, there has been intensive research effort to investigate various phenomena that could be used to produce a solid-state accelerometer. Various devices have been demonstrated, with surface acoustic wave, silicon and quartz devices being

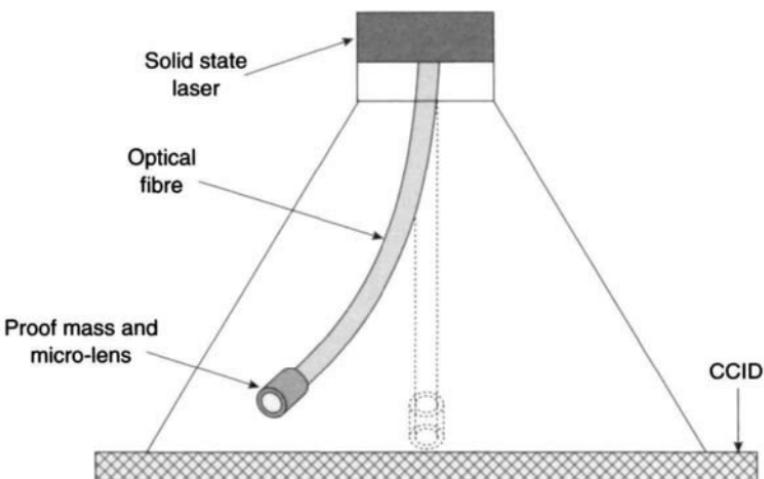


Figure 6.3 Pendulous fibre optic accelerometer

most successful. These sensors are small, rugged, reliable and offer the characteristics needed for strapdown applications.

Many of the devices discussed below have been the subject of research studies into the concepts only and have not been developed for particular applications as far as the authors are aware. In these cases, performance figures are not given.

6.4.1 Vibratory devices

These are open-loop devices which use quartz crystal technology. A common configuration uses a pair of quartz crystal beams mounted symmetrically back-to-back, each supporting a 'proof mass' pendulum. A schematic representation of such a device is shown in Figure 6.4.

Each beam is made to vibrate at its own resonant frequency. In the absence of any acceleration along the axis sensitive to acceleration, both beams vibrate at the same resonant frequency. However, when an acceleration is applied along the sensitive axis, one beam experiences compression whilst the other is stretched, or under tension, owing to the inertial reaction of the proof mass. The result is that the beam in compression experiences a decrease in frequency, whereas the beam in tension has an increase in frequency. The difference in frequency is measured and this is directly proportional to the applied acceleration.

Some of the errors often associated with this type of technology can be minimised by careful design. The symmetrical arrangement of the beams produces a cancellation of several errors that exist if only one beam is used. Error effects that are usually alleviated, or even eliminated, by this design include variations in nominal beam frequency owing to temperature changes and ageing of the quartz, asymmetrical scale-factor non-linearities, anisoinertia errors and vibro-pendulous effects.

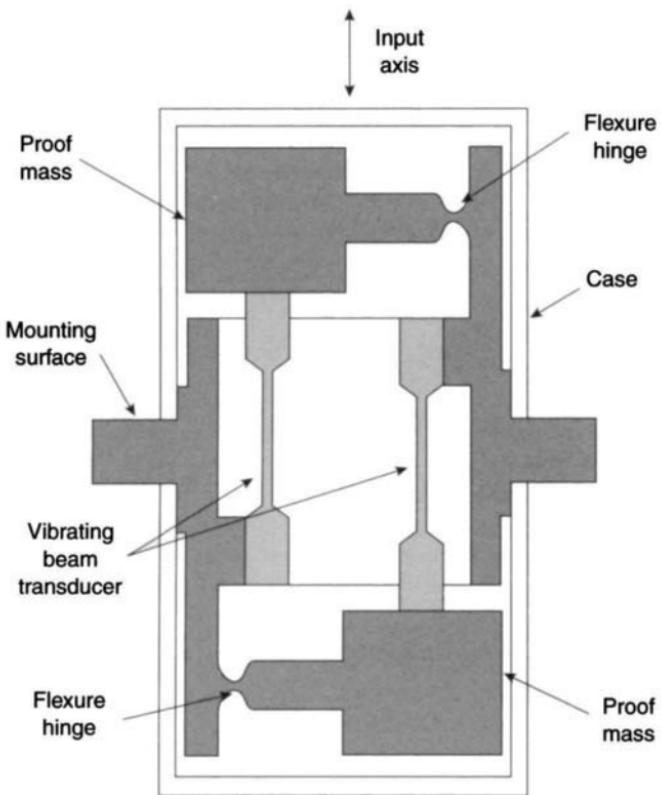


Figure 6.4 Vibrating beam accelerometer

Typical performance data are shown below:

Input range	$\pm 200g$
Scale-factor stability	$\sim 100 \text{ ppm}$
Scale-factor non-linearity	$\sim 0.05\%$ of full scale
Bias	$\sim 0.1\text{--}1 \text{ milli-}g$
Threshold	$< 10 \text{ micro-}g$
Bandwidth	$> 100 \text{ Hz}$

6.4.2 Surface acoustic wave accelerometer

This sensor is an open-loop instrument that has a surface acoustic wave resonator electrode pattern on the surface of a piezoelectric quartz cantilever beam [5, 6]. This beam is rigidly fixed at one end to the 'case' of the structure but is free to move at the other end, where a proof mass is rigidly attached, as shown in Figure 6.5. A surface

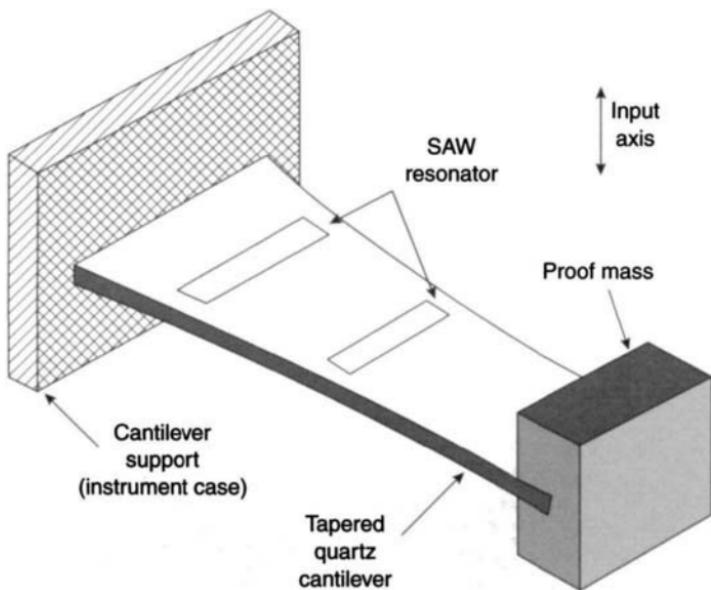


Figure 6.5 Surface acoustic wave accelerometer

acoustic wave train [7] is generated by use of the positive feedback between a pair of the metal electrode inter-digital arrays, its wavelength being determined by the separation of the metal electrodes, often called fingers.

When an acceleration is applied normal to the plane containing the beam, the inertial reaction of the assembly causes the beam to bend. When the surface of the beam is subjected to an applied strain, as occurs when the beam bends, the frequency of the surface acoustic wave changes in proportion to the applied strain. Comparison of this change with the reference frequency provides a direct measure of the acceleration applied along the sensitive axis.

The effects of temperature and other effects of a temporal nature can be minimised by generating the reference frequency from a second oscillator on the same beam. Lock-in type effects are prevented by ensuring that this reference signal is at a slightly different frequency from that used as the 'sensitive' frequency.

Typical performance data are shown below:

Input range	$\pm 100g$
Scale-factor stability	0.1–0.5%
Scale-factor non-linearity	<0.1%
Bias	<0.5 milli-g
Threshold	1–10 micro-g
Bandwidth	~400 Hz

6.4.3 Silicon sensors

Over the last decade or so, there have been research studies directed to fabricating accelerometers from silicon [8, 9]. As a material, silicon has many advantages over other materials [10]. It is inexpensive, very elastic, non-magnetic, it has a high strength to weight ratio and possesses excellent electrical properties allowing component formation from diffusion or surface deposition. Additionally, it can be electrically or chemically etched to very precise tolerances, of the order of micrometres.

In one concept, micro-machining techniques were used to form cantilevered beams of silicon dioxide over shallow cavities etched in silicon. The end of the cantilever beam was gold plated to provide the proof mass and hence increase the sensitivity of the instrument. The cantilever was metal plated along its top surface to form one plate of a capacitor, the silicon substrate forming the other plate of the capacitor, as illustrated in Figure 6.6. This form of accelerometer can be operated

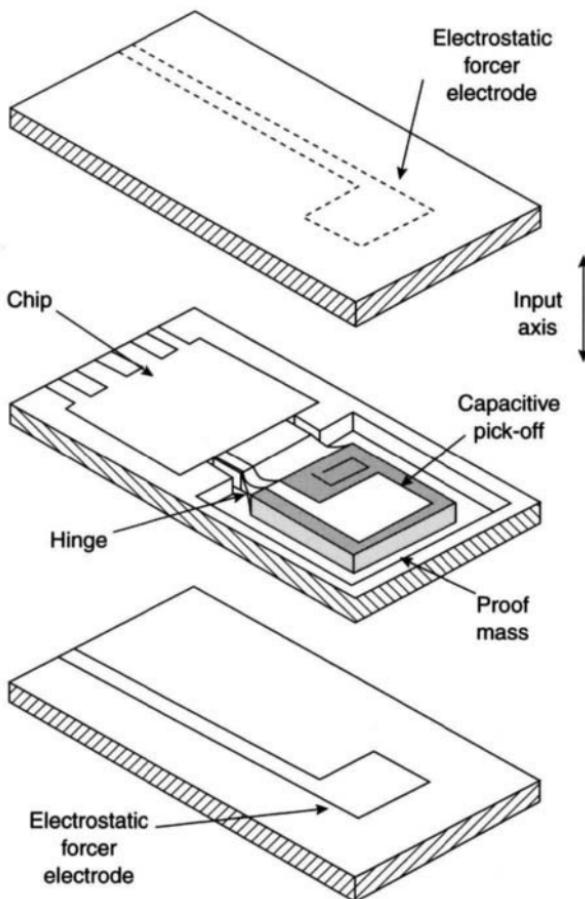


Figure 6.6 Silicon accelerometer

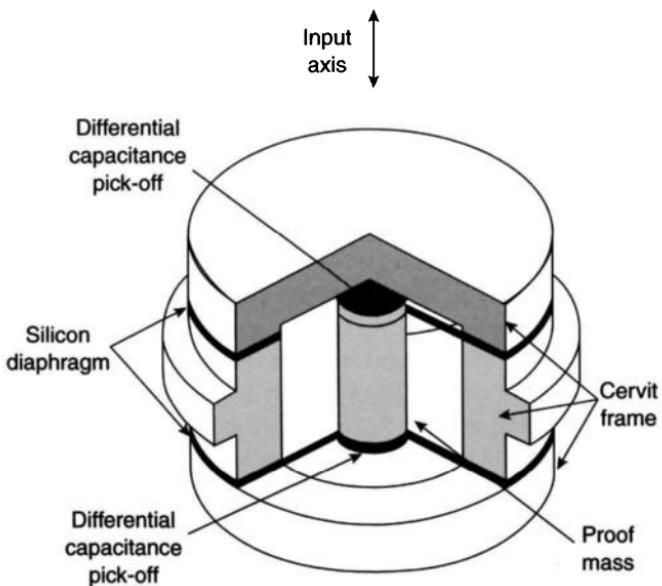


Figure 6.7 Monolithic accelerometer

in either an open-loop mode or as a closed loop device. In the open-loop mode, the capacitance between a pair of metal plates changes with the deflection of the cantilever, that is, the applied acceleration. In the closed loop mode, as shown in Figure 6.6, a pair of electrodes are used to null any deflections of the cantilever. Use of the closed loop mode increases its sensitivity. Although such devices tend not to be very accurate, they are very small and quite rugged.

A monolithic accelerometer was developed in the United States during the early 1980s. A cylindrical proof mass was supported by single crystal silicon diaphragm discs which were hinged on a cervit frame, as shown in Figure 6.7. This instrument was operated open-loop, using a differential capacitive pick-off on each end to detect motion of the proof mass when subjected to an applied acceleration. The materials were chosen to provide a thermally stable path. The major problem areas with this instrument have centred around difficulties machining the materials, achieving adequate scale-factor linearity and bonding the components together. Currently, performance data are not available for this sensor.

Another form of silicon accelerometer that is currently under development has frequency sensitive resonant tie bars integrally attached to a silicon seismic mass. These tie bars are maintained at mechanical resonance, typically vibrating at frequencies between 40 and 100 kHz depending on the configuration. When an acceleration is applied along the sensitive axis, movement of the seismic mass induces a strain in the tie bars resulting in a change in frequency of the order of tens of hertz for each applied unit g . This change in frequency is reasonably detectable. A conceptual diagram of this sensor is shown in Figure 6.8.

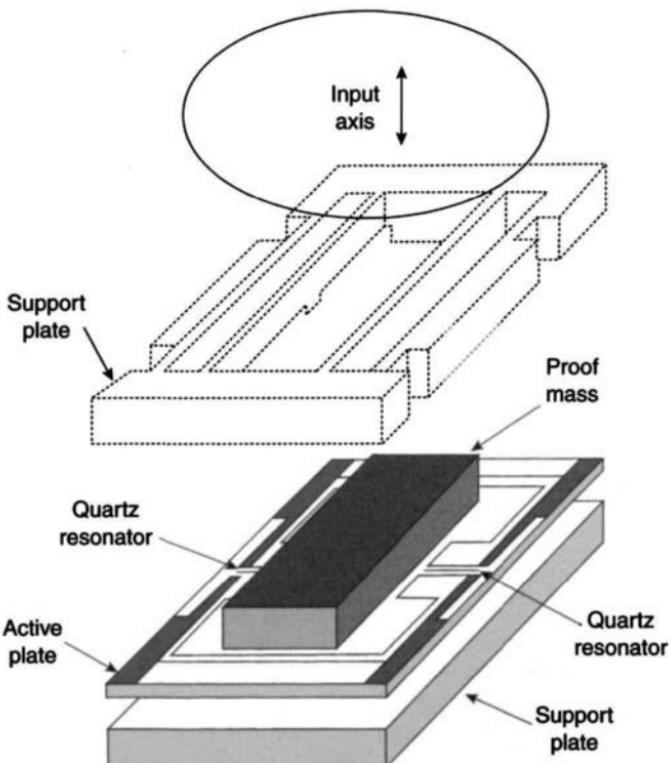


Figure 6.8 Resonant silicon accelerometer

Typical performance parameters are:

Input range	$\pm 100g$
Scale-factor stability	0.5–2%
Scale-factor non-linearity	0.1–0.4%
Bias (with compensation)	<25 milli-g
Threshold	1–10 micro-g
Bandwidth	~400 Hz

Work in the United Kingdom has investigated a thermal excitation method as an alternative to the use of piezoelectric transducers for excitation of the proof mass. This thermal excitation technique is achieved by depositing a form of bimetallic strip on the tie bars, which is used in place of the piezoelectric transducer.

A bimetallic element is formed on a tie bar by the deposition of a resistor on the top surface of a tie bar. Application of a potential difference to this resistive load produces localised heating on the top surface of the tie bar. Consequently, there is an expansion of the hot surface with respect to the cooler surface which causes the tie bar

to bend. If an alternating potential is applied to this resistive load, then the localised heating will be periodic and the top surface of the tie bar will expand and contract with respect to the lower surface, depending on the heating cycle of the resistive material. The frequency of the applied current is chosen to be synchronous with one of the natural resonant frequencies of the tie bars. As a result of this periodic bending of the tie bars, the proof mass is forced to oscillate as described above for the piezoelectric excitation technique.

A second resistor is located on each of the driving tie bars and is used as a detector to sense the oscillation frequency. This is then used as the feedback signal to modify the frequency of the applied alternating current. The drive and control electronics can also be formed in the silicon material. Quality factors in excess of 1000 have been demonstrated with such designs.

Variation in the heating effect produced by the resistive material on the tie bars is achieved by applying a suitable bias in combination with the alternating drive current. Consequently, the variation in the polarity of this applied potential allows the heating effect of the resistive material to be modulated at the frequency of this applied potential.

The main motivation for the development of this excitation technique was that an all silicon sensor could be developed. Several techniques exist for the deposition of the resistive heating elements on to the tie bars. Examples include direct diffusion doping or polysilicon deposition. Similar techniques can be used to form the detector.

6.4.4 Fibre optic accelerometer

The use of fibre optical elements is very attractive for many applications as the fibre optical waveguide is immune to electromagnetic interference. One form of fibre optic sensor has already been described, as it is very similar in operation to the pendulous accelerometers; the fibre optics merely providing an alternative form of readout. Other forms of fibre optic accelerometer rely on some physical change in a component which can be sensed using electromagnetic radiation.

Ensuring that these changes are linear functions of acceleration in known directions remains a difficult development problem, although the use of fibre technologies gives a very sensitive position readout.

6.4.4.1 Mach-Zehnder interferometric accelerometer

A Mach-Zehnder interferometer [11] uses either one or two optical fibres attached to an inertial mass as its sensitive element. When an acceleration is applied along the axis of the optical fibre, this will produce a small change in length which is proportional to the applied acceleration. The change in length can be detected by interferometric techniques similar to those described for the fibre optic gyroscope. The use of two optical fibres allows each fibre to form an arm of the interferometer and the use of nulling techniques enables greater sensitivity to be achieved, along with compensation for temperature changes in the fibres. Additionally, it is necessary to constrain the proof mass to move only along the sensitive axis of the instrument.

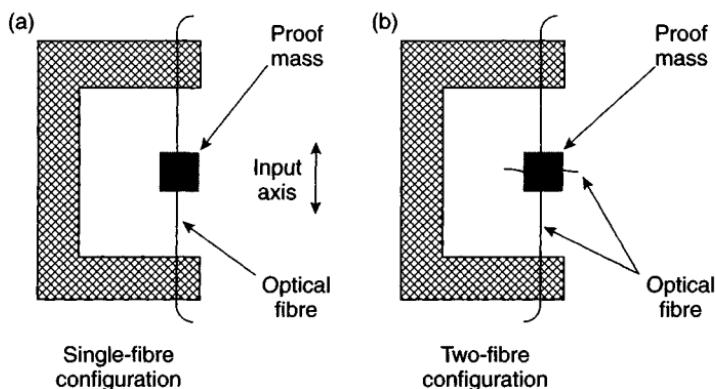


Figure 6.9 Sensitive elements of a Mach-Zehnder interferometric accelerometer

Schematic illustrations of the sensitive elements of two possible configurations are shown in Figure 6.9.

A very sensitive sensing element for accelerometers can be produced by winding a fibre optic coil around a compliant former, such as a rubber cylinder. When an acceleration is applied to the sensing element it changes dimensions and hence produces a phase change in the interferometer which is proportional to the applied acceleration. The sensitivity of the device is proportional to the number of optical fibre turns on the cylinder. Maximum sensitivity can be achieved by operating the device in a feedback mode, as shown in Figure 6.10. The intensities of the two light beams in the interferometer are detected separately and compared in a differential amplifier. The output signal from this component can then be used to 'drive' a piezoelectric device to null the phase change introduced by the distortion of the sensing element. The output of the differential amplifier is proportional to the applied acceleration. Again, it is necessary to constrain the movement of the element to be only along the sensitive axis of the device. Other technological concerns are the longer term stability of the compliant component and the effect of the different thermal expansion coefficients.

6.4.4.2 Vibrating fibre optic accelerometer

A short length of single mode optical fibre is fastened and tensioned between two pivot points in a rigid structure. This structure is vibrated so that the optical fibre oscillates at its fundamental frequency. In the absence of any applied acceleration, the displacements are symmetrical and the maximum stretch occurs at the maximum displacement with relaxation as it passes the centre line. Light passing through this optical fibre is phase modulated at $2f$, and at higher-order even harmonics of f , where f is the fundamental frequency. However, when the sensitive element is subjected to an acceleration parallel to the plane containing the oscillation, the displacement of the fibre will now be asymmetrical.

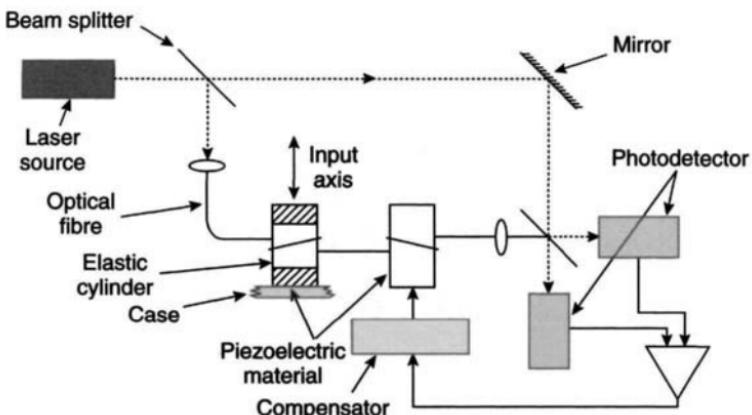


Figure 6.10 Interference accelerometer

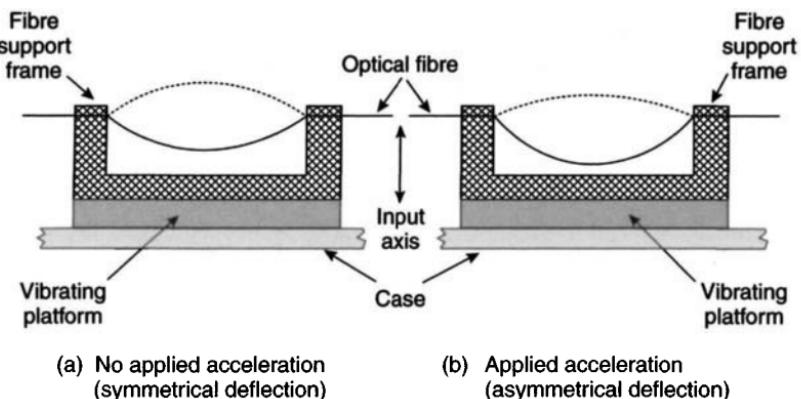


Figure 6.11 Oscillating modes of a vibrating fibre accelerometer

Light passing through the optical fibre will now be phase modulated at f and at the odd harmonics of f . The first and odd harmonic phase modulation has an amplitude proportional to the applied acceleration, and its phase relative to the drive signal will depend on the sense of the applied acceleration. Again, fibre optic interferometric techniques are used to sense the phase changes. Care is necessary in the choice of fundamental frequency and the design to reduce the effects of orthogonal acceleration sensitivities and environmental vibratory motion. The displacement of the fibre is shown schematically under the conditions of no acceleration and applied acceleration in Figure 6.11.

It may be possible to produce an amplitude modulation system by using 'lossy' multi-mode optical fibre, which is optimised for micro-bending losses, described in Figure 6.11. In this case, light which is guided along the vibrating optical

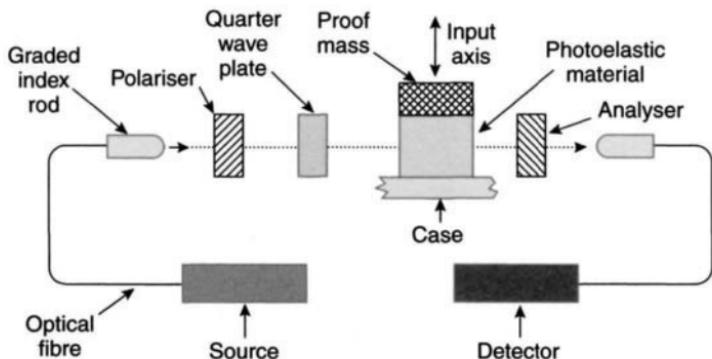


Figure 6.12 Photo-elastic accelerometer

fibre, is coupled into the cladding surrounding the optical core at the support points. This occurs as a consequence of the bending of the fibre decreasing the barrier between the core and cladding modes. Such a system would not need to use interferometry to determine the magnitude of the applied acceleration. This is because this technique converts the device from a phase modulator to an amplitude modulator.

6.4.4.3 Photo-elastic fibre optic accelerometers

The sensitive element in this device is a birefringent material [11]. Suitably polarised light is coupled into the sensitive element using multi-mode optical fibre. When an acceleration is applied to the photo-elastic material the transmission of light through it changes and the change is proportional to the applied acceleration. Research is continuing with this form of sensor. A schematic diagram of an engineering concept is shown in Figure 6.12.

6.4.4.4 Bragg grating fibre accelerometer

Research work in the United States and in Europe, at the Microelectronic Centre of Denmark, has demonstrated an accelerometer containing a Bragg grating in an optical waveguide. The centre wavelength of a Bragg grating is determined by the characteristics of the grating but can be changed by changes in the temperature, strain and pressure applied to the grating [12, 13]. Thus, an optical waveguide containing a Bragg grating is distorted when an acceleration is applied along the waveguide, the wavelength of light transmitted along the waveguide changes. This change in wavelength is proportional to the applied acceleration. The change is small but can be detected using a fibre interferometer [14].

The effect of the applied acceleration can be enhanced if the fibre waveguide is rigidly bonded to a proof mass. The general layout of this sensor is shown in the schematic diagram in Figure 6.13a. Care is necessary to ensure that the proof mass and fibre move in the direction of the applied acceleration and do not deflect

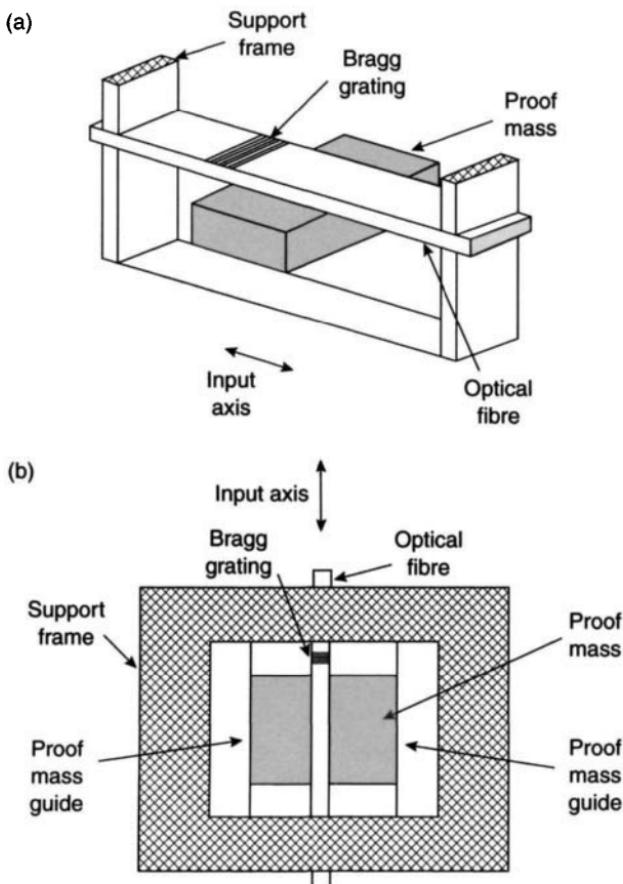


Figure 6.13 Bragg grating fibre accelerometer. (a) Schematic layout. (b) Section view

when cross-axis accelerations are applied. This is accomplished using 'guides' as shown in Figure 6.13b. Clearly, care is necessary to ensure that movement of the proof mass is not impeded by these 'guides'. Performance data are not available, but initial measurements suggest sensitivities of these devices to be in the micro-g regime.

6.4.4.5 Combined fibre optic sensors

The use of similar materials such as solid-state lasers, photo-detectors, optical fibres and common techniques in the fabrication of the sensors, suggests that there is plenty of scope for producing integrated devices, enabling both angular rate and linear acceleration to be sensed in a single device. The operation of the individual aspects of each sensor has already been dealt with in each appropriate section and will not be repeated here. The major problems are associated with the integration of the

individual components and the sharing of components. Additionally, it is necessary to isolate particular processes, such as modulation frequencies, in order that effects can be identified uniquely.

6.4.5 Optical accelerometers

It appears that there have been relatively small developments in this class of sensor. The value of optical readout techniques is well recognised, particularly with respect to enhanced sensitivity leading to greater resolution and accuracy. This class of device may provide resolution in the nano-g range and be valuable for detecting seismic disturbances or gravity gradients, but of course other features of the accelerometer must be compatible with this aspect of performance, particularly the noise in the output signal. This class of performance is comparable with the MEMS tunnelling devices (Section 7.4.4), and so may compete for similar applications.

There is some continuing research involving fibre optical accelerometers and fibre Bragg devices; the physical principles of these devices are reviewed in Section 6.4.4.

Measurement of applied acceleration has been demonstrated using optical micro-spheres. In this case the light coupled into an optically resonant micro-sphere changes as the sphere moves towards a waveguide.

6.4.6 Other acceleration sensors

Many physical effects have been exploited over the last half century or more in an attempt to measure acceleration. For completeness, two other interesting concepts known to the authors are discussed below. Generally, these programmes are not active, but either or both could become active if there is a significant change in a relevant technology.

6.4.6.1 Solid-state ferroelectric accelerometer

Attempts have been made to use the piezo-optic and dielectric properties of ferromagnetic materials. It was hoped to measure the magnitude of the applied acceleration as a function of the strain or pressure induced in a thin fibre of this material. However, technological limitations in the past prevented the feasibility of the device being demonstrated.

6.4.6.2 Solution electrolytic accelerometer

This is a solid-state ion device, making use of a shift in ions in a solution owing to the application of an acceleration. This motion causes a resultant change in the potential in the electrolyte and this potential change was found to be proportional to the applied acceleration, with good linearity. However, the electrolyte is, by its very nature thermally sensitive. This device was originally developed as part of the German missile programme during World War II.

6.5 Multi-functional sensors

6.5.1 Introduction

Previous discussions in this book have indicated that elements that are vibrating change their frequency of resonance when rotated or accelerated. Alternatively, cantilevered piezoelectric materials can be used as transducers by measuring the change in electrical charge across a ‘crystal’ when it is deflected by an applied force. These principles have been applied by mounting several such ‘elements’ at particular orientations with respect to each other. This enables one sensor to produce information about both the applied acceleration along, and the rotation rate about, an axis. These instruments are often called multi-sensors.

Multi-sensors are not confined to bending cantilever or vibrating beam technologies. It will be recalled that a mechanical gyroscope, with a mass unbalance in its rotor support, will drift when subjected to an applied acceleration about an appropriate axis. This phenomenon can be applied using a cluster of three two-axis sensors, suitably oriented with appropriate known mass unbalance, to produce an inertial measurement unit that will provide information on both linear acceleration and rotation sensed about three reference axes.

The use of multi-sensors offers the distinct advantage of reducing the number of inertial instruments required to measure the rotation and linear motion of a vehicle. Only three instruments are required for some sensor types to give full inertial data in three axes. However, the information is generally mixed in each axis and needs to be separated at some particular frequency, usually the spin frequency of the assembly. There can also be problems in achieving satisfactory or compatible performance from both accelerometer and gyroscopic channels of a multi-sensor for some applications. An additional problem area can also be cross-coupling between the different channels, although careful design can minimise this effect.

6.5.2 Rotating devices

Research began on these devices in the United States and in Britain in the late 1970s. Such devices operate by detecting the change of dynamic input, or force, which is applied to a piezoelectric transducer. Such a device is mounted on a cantilever as shown in Figure 6.14 and can be attached rigidly to a rotating element. This transducer produces an alternating electric signal proportional to the applied input.

The two principal components of a rotating multi-sensor are:

1. a rotating assembly;
2. piezoelectric transducers.

Additionally, it is necessary to have a set of slip-rings to transfer the electrical signals from the transducers to the electrical connecting pins on the case.

Piezoelectric accelerometers do not have very low threshold, nor do they have good day-to-day stability characteristics. These deficiencies can be reduced if the sensor is designed to operate at only a set frequency, and then use synchronous demodulation to remove d.c. uncertainties.

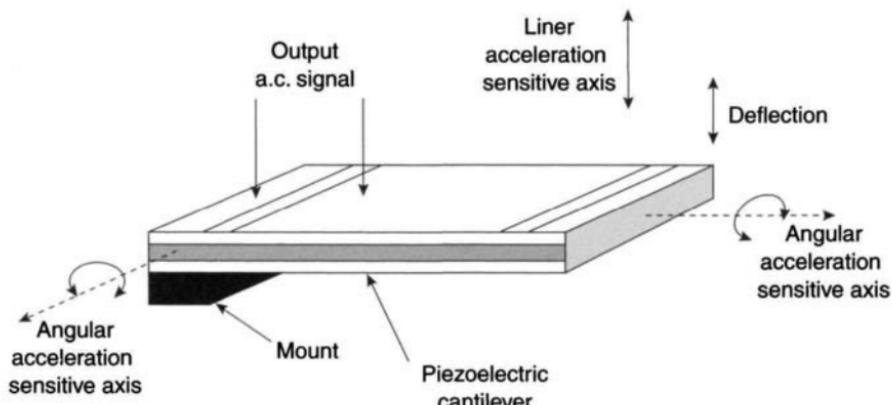


Figure 6.14 Piezoelectric accelerometer

Typically, there are four piezoelectric transducers mounted rigidly as cantilevers to the rotating assembly. The transducers are mounted in pairs, each pair being orthogonal to each other and orthogonal to the spin axis of the rotating member as shown in Figure 6.15. The construction and mounting is such that one set of transducers generates signals which are proportional to angular rate, whilst the other set produces signals proportional to linear acceleration.

Each transducer is made of layers of piezoelectric ceramic. These are ferroelectric materials, which are non-symmetric crystals and have a built-in electric dipole. In its polarised form, any stress applied to the structure results in a variation of the dipole moments causing a voltage to appear across the electrodes. Thus the material can convert mechanical energy to electrical signals and vice versa.

When a stationary cantilevered beam is oriented so that an applied acceleration deflects the beam, then a steady voltage results. However, if the beam is now rotated about an axis, so that when it has rotated by 180° the direction of the deflection is in the opposite direction, then the signal will have the reverse sense. In the case of continuous rotation of the beam, the output signal voltage is a sinusoid, with a frequency equal to the rotation frequency, the peak voltage being proportional to the applied acceleration.

As noted above, there are usually two identical cantilevered beams mounted on the rotating assembly, with their flexing axes co-linear with the spin axis. In this orientation, these transducers sense linear acceleration in the plane perpendicular to the spin axis. These rotating beams produce a suppressed carrier modulated spin frequency signal, with a peak amplitude proportional to the amplitude of the applied acceleration. The amplitude is a maximum when these transducers sense the total applied acceleration, and becomes a minimum when rotated through 90°. The principle of detection of angular rate is based on the gyroscopic behaviour of an elastically restrained body which is rotated about an axis. Usually, the cantilevered transducers are mounted 180° apart on the rotating assembly, for common mode rejection and also

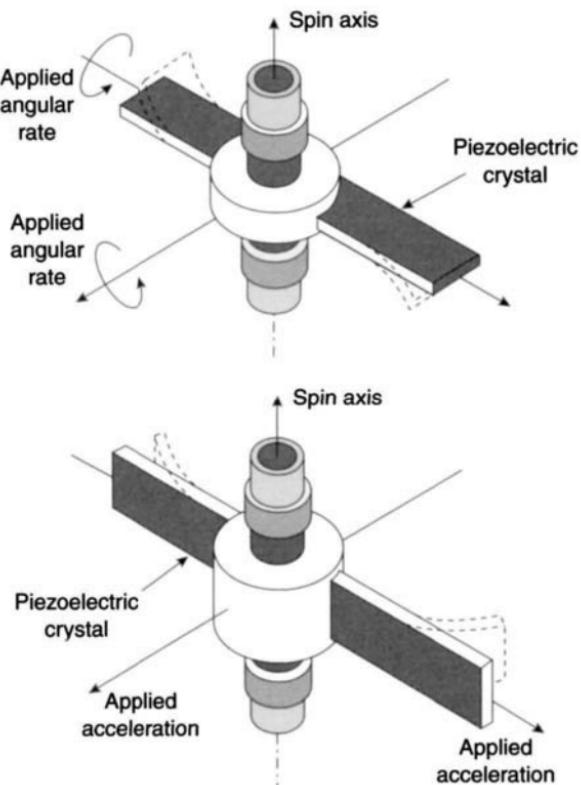


Figure 6.15 Principle of operation of a rotating multi-sensor

signal enhancement. As mounted, these transducer elements act both as the inertial members and the restraining springs.

When an angular rotation rate is applied to the spinning assembly, about an axis orthogonal to the spin axis, the angular momentum of the rotating transducers generates a forcing function, bending the transducer. This forcing function is a suppressed carrier signal modulated at the spin frequency of the mounting assembly. A plot of the amplitude of the signal generated is a sinusoid, its magnitude being proportional to the magnitude of the applied rate. The phase of the output is such that the maximum signal occurs when the sensing transducer is co-linear with the applied input. The minimum voltage occurs one quarter of a rotation away from the maximum signal.

Clearly, such an instrument can sense accelerations applied along two axes in the plane perpendicular to the spin axis. Similarly, it can sense angular rotation rates applied along two axes in the plane orthogonal to the spin axis. Therefore only two sensors are required to be mounted so that their spin axes are not parallel to enable three axes of angular rate data and three axes of linear acceleration data to be generated. The redundant information generated along and about the fourth axis is available for system checks.

These sensors are open-loop devices and consequently tend to have poorer scale-factor characteristics when compared with closed loop sensors such as the floated rate integrating gyroscope. Additionally, the scale-factor can change as the piezoelectric crystals age. However, these open-loop devices do not consume extra power, and consequently do not liberate heat, when measuring high rates of rotation.

These sensors are capable of very accurate measurement of angular rotation rates and linear acceleration. To achieve the high accuracies required for inertial navigation purposes, careful calibration and characterisation is necessary and temperature compensation is usually vital. Currently, the very accurate instruments tend to be quite large; up to 150 mm long by 35 mm diameter. This is offset by the fact that this single sensor provides four of the six measurements required by a navigation system.

Careful choice of certain components is necessary in order to contain certain error sources. Use of 'low noise' bearings minimises the noise coupled into the crystals producing a background signal. Noise generated by the slip rings is synchronous with the piezoelectric signals and consequently represents an acceleration or angular rate error. The measurement bandwidth of the sensor is dependent on the spin speed of the rotating assembly. Consequently, as it requires at least two readings per cycle to define a sinusoid, the rotating assembly must have a spin frequency at least twice that of the measurement bandwidth. For high bandwidth applications, this can lead to significant generation of bearing noise giving rise to potential saturation of the electronic measurement system.

Variations in temperature can also lead to changes in bearing generated noise as a result of variations in the bearing characteristics such as internal loading and viscosity changes to the lubricant. Additionally, variations in the temperature can alter the reference electronics used to resolve the acceleration and angular signals. This appears like a scale-factor error.

Typical performance data are given below:

<i>Gyroscope:</i>	
Maximum input rate	300–400°/s
<i>g</i> -Independent bias	1–10°/h
<i>g</i> -Dependent/mass unbalance bias	5–10°/h/g
Anisoelastic bias	0.1–0.2°/h/g ²
Scale-factor stability	0.1–2%
Scale-factor non-linearity	0.03–0.1%
Bandwidth	60–100 Hz
<i>Accelerometer:</i>	
Input range	up to ±100g
Scale-factor stability	0.1–2%
Scale-factor non-linearity	0.03–0.1%
Bias	1–10 milli-g
Threshold	1–10 micro-g
Bandwidth	>70 Hz

One concept that was the subject of a recent research programme was based on the use of rotating surface acoustic wave accelerometers mounted on a common shaft, instead of piezoelectric sensors. Angular motion and linear acceleration of the sensor's case are sensed by mounting pairs of surface acoustic wave accelerometers as cantilevers on a body that is rotated at a constant speed. When angular motion or linear acceleration is applied, the cantilevers are deflected owing to the various physical effects described earlier for the piezoelectric based sensor. The output signals are also generated using similar techniques.

The use of the surface acoustic wave elements offered several advantages as various effects, such as temperature induced biases, can be compensated. These effects can be compensated by the use of two surface acoustic wave oscillators on the same cantilever, as described in Section 6.4.2. Additionally, a digital output can be generated directly on the element and passed through the slip rings allowing the effect of noise on small signals to be eliminated. However, work on this sensor appears to be dormant.

6.5.3 Vibratory multi-sensor

Research into this form of sensor has been most active in the United States. It is a single-axis device based on vibrating sensor technology enabling measurements to be made of both angular rotation rates and linear acceleration. The sensor measures continuous angular rates by the oscillating Coriolis acceleration induced on an accelerometer which is being vibrated, and forces subject to oscillatory linear velocity. Typically, the accelerometer in this device uses silicon solid-state technology.

The principal components of this instrument are:

1. an accelerometer;
2. a vibrator (vibratory platform).

The principle of operation is essentially that of a pendulous accelerometer which is vibrated along its hinge or pivot axis (Figure 6.16). As a result of this vibration, an oscillatory linear velocity is imparted to the pendulum. Consequently, the accelerometer will sense a Coriolis acceleration at the frequency of vibration proportional to the angular motion applied about the axis of the pendulum. Additionally, the accelerometer will measure any linear acceleration applied along its input axis. The electrical signal generated by such a sensor will have a d.c. value proportional to the applied linear acceleration and an a.c. signal at the vibration excitation frequency. The latter signal can be demodulated to produce a signal which is proportional to the applied rotation rate about the axis of the pendulum.

In a practical device, an accelerometer using vibrating beam technology is excited by piezoelectric crystals. Typically, two accelerometers are vibrated 180° out of phase to give common mode rejection, thus preventing random inputs at the vibrating frequency from corrupting measurements. This type of sensor has many advantages through the elimination of rotating elements and bearings, but has the disadvantage of small signal-to-noise ratio for the angular rate measurement. Schemes have been devised to mount three accelerometers at various orientations on a plate, enabling

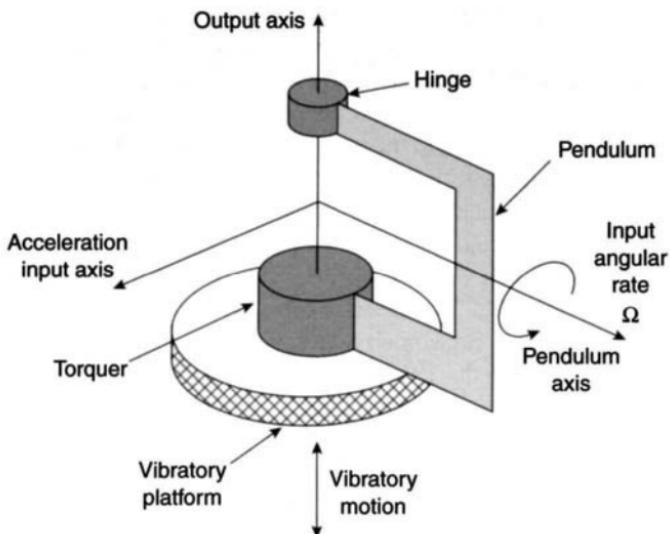


Figure 6.16 Principle of operation of a vibratory multi-sensor

them to be vibrated, or dithered, about the body diagonal of the reference axis set. Using one common activation axis has several advantages, such as the elimination of cross-talk and aliasing between the sensors. It also allows some common electronic circuits to be used resulting in a very compact three axis inertial measurement unit.

Anticipated performance parameters for this form of multi-sensor are as follows:

<i>Gyroscope:</i>	
Maximum input rate	$\pm 1000^\circ/\text{s}$
<i>g</i> -Independent bias	5–10°/h
Scale-factor stability	~0.1%
Scale-factor non-linearity	~0.05%
Bandwidth	>100 Hz
<i>Accelerometer:</i>	
Input range	up to $\pm 200g$
Scale-factor stability	~0.05%
Scale-factor non-linearity	<0.1%
Threshold	~10 micro- <i>g</i>
Bandwidth	>100 Hz

6.5.4 Mass unbalanced gyroscope

Angular momentum gyroscopes, such as the floated rate integrating gyroscope and the dynamically tuned gyroscope, are precision instruments requiring careful

assembly to achieve the levels of performance normally required for most applications. During manufacture, particular care must be taken to ensure that the spinning rotor, or rate integrating gyroscope rotor/float combination, is balanced accurately. In the presence of a linear acceleration normal to the float axis in the case of the rate integrating gyroscope, or normal to the rotor spin axis in the dynamically tuned gyroscope, any mass unbalance will induce a torque, causing the rotor to precess, and so produce an erroneous rate measurement. However, by introducing a known amount of mass unbalance into gyroscopes of this type, it is possible to obtain a measure of the acceleration to which the instrument is subjected, in addition to the angular rates which it is sensing. Much of the pioneering work on this form of sensor was undertaken in Germany and the United States, but more recently in France [15].

This concept dates from the 1950s and indeed, instruments have been manufactured based on this principle for many years. The Honeywell precision integrating gyroscopic accelerometer (PIGA) is one such device. This type of device was also developed in the United Kingdom by Ferranti, now part of BAE Systems. The design of the PIGA was based on a single-axis floated gyroscope, in which the rotor was made pendulous with respect to the output axis of the instrument. Generally, this form of sensor was intended for use on stable platforms.

More recently, attention has focused on a development of the dynamically tuned gyroscope, which has the centre of suspension of its rotor displaced slightly with respect to its centre of gravity.

The displacement occurs along the motor drive shaft, in a manner that causes accelerations applied perpendicular to the drive shaft, that is, parallel to the input axes, to produce torques which cause precession of the rotor. This is shown schematically in Figure 6.17. As with the conventional dynamically tuned gyroscope, the device operates in a torque re-balance mode. However, in this case the pick-off outputs are fed back to null the precession caused by both the input rates and the applied accelerations.

A perfect two-axes gyroscope operating in a torque re-balance mode has the steady state relationship between the input rates, ω_x and ω_y , and the applied torquer moments, M_x and M_y , given by:

$$\begin{aligned}\omega_x &= -\frac{M_y}{H} \\ \omega_y &= \frac{M_x}{H}\end{aligned}\tag{6.4}$$

where H is the angular momentum of the rotor. If, in addition to the torquer moments, moments act as a result of the unbalance in the rotor suspension, the above equations take the form shown below, where the unbalance torques are proportional to the applied accelerations, a_x and a_y ,

$$\begin{aligned}\omega_x &= -\frac{M_y + Ba_x}{H} \\ \omega_y &= \frac{M_x - Ba_y}{H}\end{aligned}\tag{6.5}$$

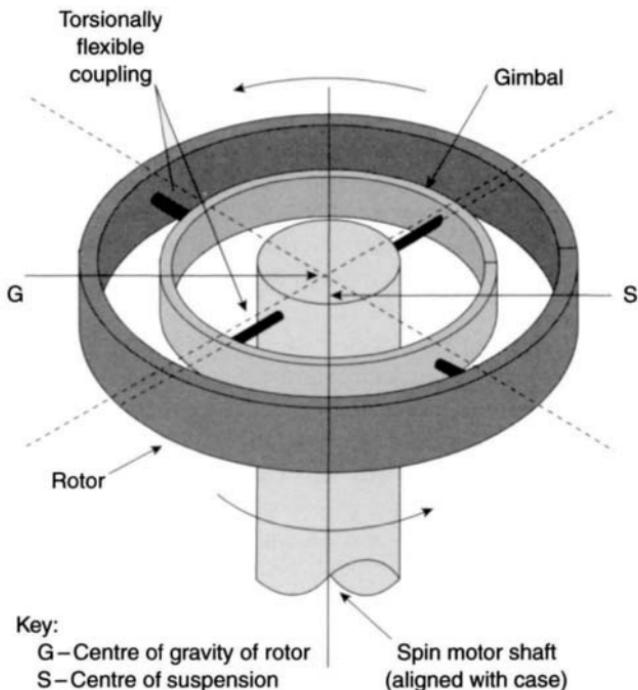


Figure 6.17 Mass unbalanced gyroscope rotor configuration

The factor B is a function of the displacement between the rotor centre of gravity and its centre of suspension, and the inertia of the rotor about an axis perpendicular to its spin axis. Rearranging eqn. (6.4) and writing $B/H = b$, the measurements, m_1 and m_2 , provided by a single sensor may be expressed as follows:

$$\begin{aligned} m_1 &= -\frac{M_y}{H} = \omega_x + ba_x \\ m_2 &= \frac{M_x}{H} = \omega_y + ba_y \end{aligned} \quad (6.6)$$

Thus, a single gyroscope can provide a weighted sum of the turn rate about, and the acceleration along, each input axis. The constant b is referred to as the mass unbalance coefficient and may be expressed in units of $^{\circ}/s/g$. The choice of b depends on many factors including the required measurement range of the gyroscope and the motion of the vehicle in which it is to be installed.

By combining three mass unbalanced gyroscopes of this type in an inertial measurement unit, it is possible to obtain estimates of angular rates and linear accelerations in three mutually orthogonal directions provided:

- the mass unbalance coefficient is different for each gyroscope;
- the spin axes of the gyroscopes are not co-planar.

A conventional strapdown system using dynamically tuned gyroscopes would require two such gyroscopes and three accelerometers. It is postulated therefore, that the three accelerometers may be replaced by one additional gyroscope to form a mass unbalanced 'navigation' system. Such a system has the advantage of using identical re-balance loop electronics for all sensors, and uses fully the information from each input axis. Potential disadvantages include reduced dynamic range compared with the conventional system, some additional computing complexity associated with the extraction of separate angular rate and linear acceleration estimates. Additionally, there is the possibility of additional dynamic cross-coupling between these quantities.

Analysis of the effects of bias, scale-factor and cross-coupling errors in a mass unbalanced system, reveals that each error term produces inaccuracies in the estimates of both angular rate and linear acceleration, the latter being functions of the mass unbalance coefficient, b . For example, take the case of an orthogonal system in which the mass unbalance coefficients for two of the gyroscopes are equal, but of opposite sign, and zero for the third gyroscope. The general form of the rate and acceleration estimation errors, $\delta\omega$ and $\delta\mathbf{a}$, in such a system is illustrated by the following matrix form:

$$\begin{bmatrix} \delta\omega \\ \dots \\ \delta\mathbf{a} \end{bmatrix} = \begin{bmatrix} \mathbf{B}_f \\ \dots \\ \frac{\mathbf{B}_f}{b} \end{bmatrix} + \begin{bmatrix} \mathbf{S} & \vdots & b\mathbf{S} + \mathbf{B}_g \\ \dots & \vdots & \dots \\ \frac{\mathbf{S}}{b} & \vdots & \mathbf{S} + \frac{\mathbf{B}_g}{b} \end{bmatrix} \begin{bmatrix} \omega \\ \dots \\ \mathbf{a} \end{bmatrix} \quad (6.7)$$

where \mathbf{B}_f is the fixed bias, \mathbf{B}_g is the g -dependent bias or uncertainty in the mass unbalance coefficient, \mathbf{S} is the matrix containing scale-factor errors and cross-coupling terms, \mathbf{a} is the applied linear acceleration and ω is the applied turn rate.

Preliminary tests of this type of device carried out under laboratory conditions [15] suggest that measurements of turn rate and acceleration can be derived to an accuracy of significantly less than $100^\circ/\text{h}$ and less than 10 milli- g , respectively, and that the scale-factor error for such a device would be less than 10^{-3} . Errors in the acceleration measurements may be deduced, to a large extent, from the axial unbalance factor, which is assumed here to be in the region of $5^\circ/\text{h}/g$. In addition to the usual bias and scale-factor errors which arise when separate gyroscope and accelerometers are used, some additional cross-coupling between the rate and acceleration estimates arises in a mass unbalanced system.

6.6 Angular accelerometers

This form of inertial sensor provides a means for sensing angular motion. Traditionally, non-gyroscopic angular motion sensors have used a balanced mass suspended in bearings which generated a torque proportional to the applied angular acceleration. When the mass is constrained by a spring, the angular displacement is a measure of the angular acceleration.

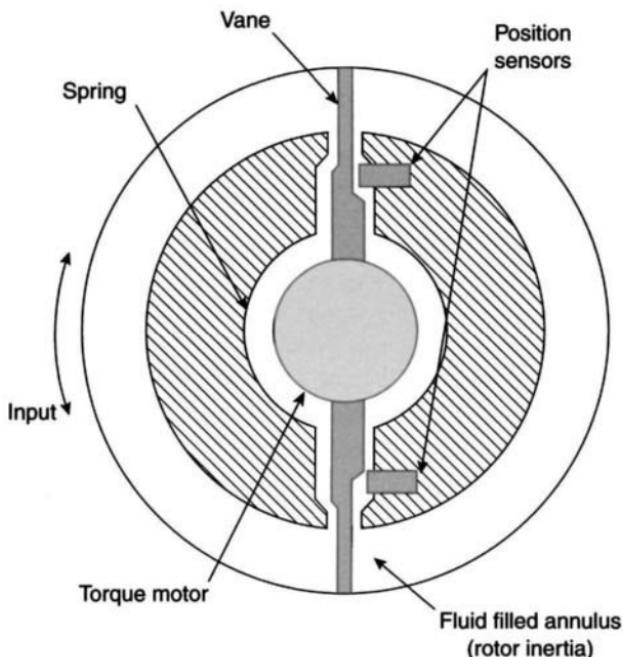


Figure 6.18 Fluid rotor angular accelerometer

There have been significant developments in the technology used in angular accelerometers. Small, compact, rugged and accurate sensors can now be produced, and have been applied to several applications. The devices may be operated in a closed or open-loop mode, depending on the configuration.

6.6.1 Liquid rotor angular accelerometer

Recently, progress in the development of this type of instrument in the United States has enabled it to evolve from a heavy, and often fragile, device to a small lightweight sensor. The modern devices have almost instantaneous readiness, reduced power consumption, enhanced ruggedness and increased sensitivity when compared with the older designs and have eliminated rotating elements.

A schematic diagram of a liquid rotor angular accelerometer is shown in Figure 6.18.

The fluid-filled sensor has an annular tube containing a liquid such as silicone oil, or a high density liquid of the type used in rate-integrating gyroscopes. This liquid forms the seismic or proof mass in the sensor. The annular tube is blocked by a disc connected to a galvanometer movement, supported by jewel and pivot bearings. This arrangement forms a servoed torque generator.

Application of an angular acceleration about the axis of the annular tube, would accelerate this tube leaving the inertial mass behind. However, the disc causes the fluid to move with the case, with a consequential reaction at the disc. This motion

is sensed by the position-sensing mechanism and provides feedback to the galvanometer torquer, which provides the torques necessary to accelerate the fluid with the case. The magnitude of this feedback signal is directly proportional to the angular acceleration acting about the input axis.

This form of system provides a good deal of flexibility as the electronic gain can be set to generate full scale deflections for any given displacement of the disc. Various other parameters, such as the cross section, diameter and sensing area of the tube, as well as the density of the fluid in this tube, may be selected individually to provide the desired frequency response and sensitivity.

One common problem with this sensor is the effects of change in temperature and thermal gradients across the sensor which can produce non-linear responses and sensitivity to linear acceleration. Changes in ambient temperature can be corrected using a volume compensator. However, careful design and thermal screening are necessary to avoid thermal gradients across the instruments.

Typical performance parameters are given below:

Input acceleration	up to 50 rad/s ²
Scale-factor linearity	~0.1% of full range
Bias	~0.001 rad/s ²
Bias temperature coefficient	~0.0005 rad/s ² /°C
Threshold	~0.005% of full range
Bandwidth	up to 60 Hz

6.6.2 Gas rotor angular accelerometer

This design of angular accelerometer has some similarities with the instrument which uses a liquid rotor. However, in this case, a high density gas at a high pressure is contained in a single tube, its end being attached to a pressure sensor. Generally, this device is operated in an open-loop mode.

When angular motion is applied about an axis perpendicular to the plane containing the gas filled tube, there is relative motion between the tube and the gas. This motion is sensed using the pressure sensor which forms a barrier across the tube and prevents free flow around it.

The pressure generated by the gas rotor is usually quite small, typically in the range 10–100 Pa. Hence the pressure sensor must have high sensitivity. Additionally, it must impart a stiffness to the system to produce the necessary dynamic characteristics of the instrument. A pressure sensor with an electrically conductive membrane positioned between two circular electrodes, is one possible simple design that could be used to detect the gas motion. In this case, when a differential pressure is applied across the membrane, resulting from the motion of the gas, its displacement results in a differential capacitance change between the membrane and the electrodes. Alternative pressure sensors giving greater accuracy can also be used, although they are usually more expensive.

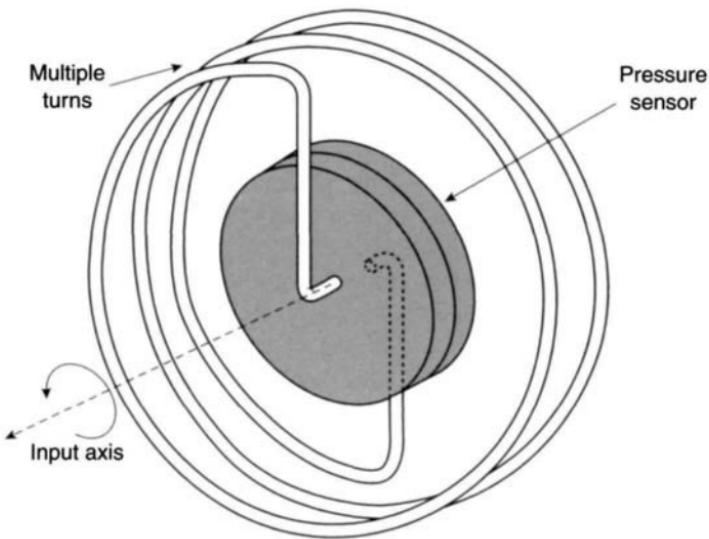


Figure 6.19 Gas rotor sensor

The tube containing the gas can be formed into any shape. However, a helix is the most common. The pressure generated is proportional to the mean radius of the helix, the number of turns, the density of the gas and, of course, the applied acceleration. Successful designs, with diameter of about 40 mm and a few tens of turns have been demonstrated. A constriction in the tube was necessary to provide damping of the motion of the gas.

A schematic diagram of a gas rotor angular accelerometer is shown in Figure 6.19.

As in the case of the liquid rotor, temperature gradients across the sensor must be avoided. Additionally, careful screening of leads and components is necessary to avoid stray capacitances corrupting the output signals.

Inherently, this design is very robust and offers a long operational life and low cost. However, there does not appear to be any current development activity of this type of sensor, in contrast with the status of the liquid rotor devices.

6.7 Inclinometers

An inclinometer is a gravity reference device capable of sensing tilt. The instrument is basically a special implementation of a linear accelerometer with low maximum acceleration capability. The accelerometer output is usually processed to give a d.c. voltage directly proportional to the angle of tilt. Typical applications of the inclinometer are platform levelling for target acquisition systems and fire control systems, and in inertial component testing.

6.8 Summary of accelerometer and multi-sensor technology

Many different types of inertial sensors can be used for sensing and measuring the magnitude of an 'accelerating' force. These sensors are of many different types and design. The review has included the mechanical sensors, using the classical pendulum principle, to the modern solid-state devices. Generally, all these instruments are suitable for strapdown applications and in such an environment will give accuracies ranging from tens of micro-gravitational acceleration (micro-*g*) to fractions of a 'g'.

The mechanical accelerometers come in various forms, with a selection of materials and designs for the pendulum's hinge mechanism. These sensors may be fluid filled in order to improve the damping of the motion of the pendulum. The pendulum may be constrained to very small displacements, through the use of force-feedback techniques, in order to achieve high accuracy. Alternatively, the sensor may be operated in an open-loop mode.

Solid-state technology offers various techniques that may be applied to enable small, reliable and relatively inexpensive instruments to be produced. A variety of techniques have been reviewed, including the use of optical fibres, vibratory devices, surface acoustic wave devices and the use of silicon materials. These sensors are generally operated in an open-loop mode, but some designs are amenable to the use of closed loop techniques. In the case of the closed loop mode, the displacement of the 'proof mass' is generally not returned to its 'null' position. Instead, the sensor operates by nulling an observed effect, such as a frequency change or a modified resonant condition.

A summary of typical performance characteristics² for a range of accelerometers is given in the following table.

Characteristic	Accelerometer type				
	Force-feedback pendulous	Vibrating fibre optic	Vibrating quartz	SAW	Silicon
Input range (<i>g</i>)	±100	±20	±200	±100	±100
Scale-factor stability (%)	0.1	0.001	0.01	0.1–0.5	0.5–2
Scale-factor non-linearity (% full scale)	0.05	0.05	0.05	<0.1	0.1–0.4
Fixed bias (milli- <i>g</i>)	0.1–10	1	0.1–1	<0.5	<25
Threshold (micro- <i>g</i>)	10	1	<10	1–10	1–10
Bandwidth (Hz)	400	100	400	400	400

² These are typical values applicable over the range of parameters stated. In many cases, the values given could be improved. However, it is not normally possible to have all the best case values in a single unit. These values are only for general indicative purposes.

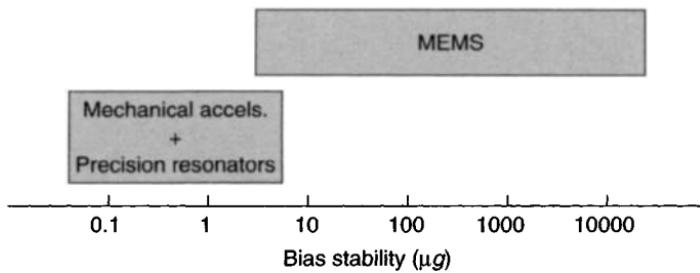


Figure 6.20 Near-term accelerometer performance summary

It is noted that substantially higher performance can be achieved using force-feedback devices. Precision devices capable of detecting accelerations as small as a few micro-g have been made with scale-factor stabilities of $10^{-5}\%$. However, such instruments are not normally designed to measure accelerations of $\pm 100\text{g}$.

Since the 1980s, there have been significant developments in the performance of so-called multi-sensors enabling a simple instrument to sense both linear and angular motion along and about two axes. These sensors offer significant potential for many applications in the future.

Finally, there has been progress in the state of the art of the manufacture of angular accelerometers, which can offer an alternative to the use of gyroscopes for some applications. The use of a fluid ring rotor to sense the applied motion has enabled small, sensitive, rugged and reliable angular accelerometers to be produced.

The performance of developed inertial sensors for near-term applications requiring acceleration measurements is shown in Figure 6.20.

It is believed that applications calling for high accuracy accelerometers will continue to incorporate mechanical sensors, with some use of resonant devices. In other application areas, MEMS sensors (see Chapter 7) are expected to become most widely used.

References

- 1 McLAREN, I.: 'Open and closed loop accelerometers'. AGARD, Agardograph No. 160, 6, 1974
- 2 SMITHSON, T.G.: 'A review of the mechanical design and development of a high performance accelerometer'. Proceedings *I.Mech.E., Mechanical Technology of Inertial Sensors*, C49/87, 1987
- 3 SCHULER, A.R., GRAMMATIKOS, A., and FEGBY, K.A.: 'Measuring rotational motion with linear accelerometers', *IEEE Transactions on Aerospace and Electronic Systems*, 1967, AES-3 (3), p. 465
- 4 NEWMAN, F.H., and SEARLE, V.H.L.: 'The general properties of matter' (Edward Arnold, 1963)
- 5 PARKER, D.F., and MAUGIN, G.A. (Eds.): 'Recent developments in surface acoustic waves' (Springer Verlag, 1988)

- 6 ACHENBACH, J.D.: 'Wave propagation in electric solids' (North-Holland, 1973)
- 7 INTERNATIONAL SPECIALIST SEMINAR: 'Component performance and system applications of surface acoustic wave devices'. Proceedings, *Institution of Electrical and Electronic Engineers*, 1973
- 8 BARTH, P.W.: 'Silicon sensors meet integrated circuits', *IEE Spectrum*, 1981, **9**, p. 33
- 9 MIDDLEHOEK, S., and AUDET, S.A.: 'Silicon sensors' (Academic Press, 1989)
- 10 PETERSEN, K.E.: 'Silicon as a mechanical material'. *Proceedings IEEE*, 1982, **70** (5), pp. 420–457
- 11 YOUNG, M.: 'Optics and lasers' (Springer Verlag, 1992)
- 12 LUCKOSZ, W.: 'Integrated optical nanomechanical devices as modulators, switches, and as tunable frequency filters, and as acoustical sensors'. *SPIE*, 1973, *Integrated Optics and Microstructures*, 1992
- 13 MOREY, W.W.: 'Distributed fibre grating sensors'. Proceedings *OFS '90*, 1990, pp. 285–289
- 14 KERSEY, A.D., BERKOFF, T.A., and MORSEY, W.W.: 'Fibre-grating based strain sensor with phase sensitive detection'. Proceedings 1st European Conference on *Smart structures and materials*, 1992, pp. 61–62
- 15 MICHELIN, J.L., and MASSON, P.: 'Strapdown inertial systems for tactical missiles using mass unbalanced two-axis rate gyros'. AGARD Lecture Series No. 133, 1984

Chapter 7

MEMS inertial sensors

7.1 Introduction

New applications that have demanded low-cost sensors for providing measurements of acceleration and angular motion have provided a major incentive for the development of micro-machined electromechanical system (MEMS) sensors. One typical modern application is the use of inertial sensors in transport, such as motor cars; an indication of the current extent of the use of these devices in this application and examples of many others are given in Chapter 15.

MEMS devices are one of the most exciting developments in inertial sensors in the last 25 years. These devices overcome many of the features that have impeded the adoption of inertial systems by many potential applications, especially where cost, size and power consumption have been governing parameters. Many efforts involving value engineering and automation have been applied to the design and manufacture of conventional inertial instruments, with significant success; however, the cost has remained high. The major reasons centre on:

- high parts count;
- requirement for many parts with high-precision tolerances;
- intricate and precision assembly techniques;
- accurate testing, characterisation and calibration.

The use of silicon as the base material in the manufacture of the components offers a radical approach and overcomes many of the issues considered above for conventional mechanical sensors. This is summarised in Figure 7.1. Additionally, they can be considered as offering a sensing device on a chip with the very real prospect of a precision inertial measurement unit sensor on a chip in the near future. This holds the opportunity to make an inertial navigation system (with GPS aiding) costing less than \$1000.

MEMS sensor technology makes direct use of the chemical etching and batch processing techniques used by the electronics integrated circuit industry. Precision

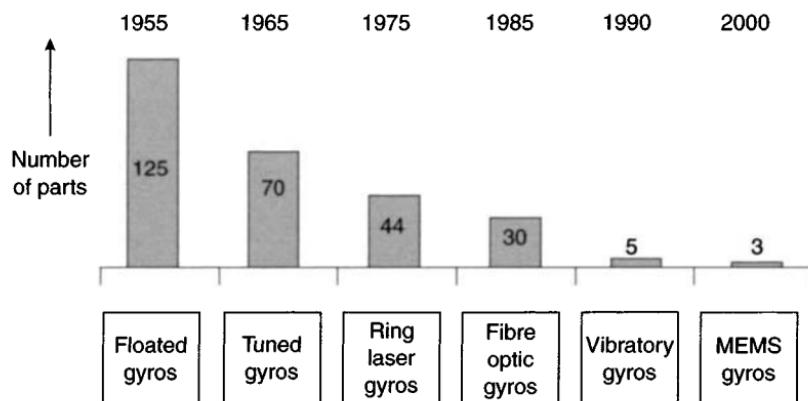


Figure 7.1 Reduction in part count

techniques developed by this industry for ‘machining silicon’ have been adapted to make very small mechanical structures using silicon or quartz. Particular advances have been obtained through the use of plane-wall etching. The properties of the resulting solid-state sensors, viz.

- small size;
- low weight;
- rugged construction;
- low power consumption;
- short start-up time;
- inexpensive to produce (in high volume);
- high reliability;
- low maintenance;
- compatible with operation in hostile environments;

provide the engineer with a level of design flexibility beyond anything that has preceded these developments. The consequence has been a proliferation of applications, both military and commercial, where such devices may be used; some of these are discussed in Chapter 15.

The properties of silicon and the ease of perfecting high-fidelity components was a key breakthrough in the transition of this sensor technology from a research device to a practical, mass-volume sensor. However, the reduction in size of the sensing elements brings with it challenges for attaining good measurement performance and high resolution. In general, reductions in size give rise to decreases in sensitivity/scale-factor and increases in noise. In addition, there are thermal sensitivity concerns; for example, the change in the Young’s modulus of silicon with temperature is $\sim 100 \text{ ppm}/^\circ\text{C}$.

Despite these limitations, low-cost MEMS gyroscopes and accelerometers demonstrating performance approaching $1^\circ/\text{h}$ and 50–100 micro-g, respectively, are expected to become readily available within the next few years. One of the major

reasons for the enhanced performance of MEMS devices is the ability to undertake complex compensation of the systematic errors exhibited by these sensors, which can now be accomplished in real time. Complex or deep compensation methods rely on a very detailed knowledge of both the error mechanisms and the fundamental characteristics of each sensor type and have to be embodied into the design of the Kalman filter to give IMU-level compensation.

Moreover, recent research has given rise to an enhanced understanding of the physical behaviour of the sensing element technology and the fundamental interaction with the supporting electronics used by these devices. These advances, along with sophisticated compensation techniques, have seen a dramatic increase in the measurement accuracy. Over the last decade techniques have been demonstrated that will enable the measurement accuracy of the best quality devices to approach those of inertial-grade sensors, namely:

- angular rate measurement accuracy of $0.01^\circ/\text{h}$ with MEMS gyroscopes;
- specific force measurement to better than 1 milli-*g* from MEMS accelerometers.

Hence, there is the real prospect that such devices with enhanced performance will displace ring laser and fibre optic gyroscopes in many tactical applications in a similar time scale.

Initial MEMS sensor developments focused on the generation of miniature accelerometers, the system and performance requirements of which were driven by the demands of the automobile industry. As a result, MEMS accelerometer technology was the first to achieve a level of technical maturity, with a significant number of sensors now being commercially available.

In contrast, there was less of a stimulus for the development of similar angular measurement sensors, so commercially available devices in this class were slow to emerge initially. More recently, MEMS gyroscopic sensor development has been the subject of significant investment and development effort in industry and research institutions, both in Europe and the United States, culminating in the wide availability of lower cost/lower performance sensors in recent years. Research effort has concentrated, and continues to be focused, on both the physics of the devices as well as the refinement of the batch processing techniques required for their cost-effective manufacture leading to a high batch yield.

There are a number of reasons for the phenomenal increase in the performance of the sensors during the last decade of the twentieth century; particularly plane-wall etch. This has been a direct result of enhanced knowledge of the effects of geometry of the structure and its size, as well as the electronics and packaging, on the ultimate performance and reliability of the devices. Moreover, there has been a significant investment in developing techniques for integrating all of the sensors on to a single chip, which led to many system and performance benefits.

Hence these micro-miniature devices appear to be the sensors of the future. This is because they are based on a solid-state architecture and with careful design have few components and therefore have the potential to become the 'Holy Grail' for inertial sensor technology. Consequently, inertial sensor technology developments in recent

years have been concentrated almost exclusively on the development and perfection of MEMS devices.

In the more distant future it is likely that further enhancements in performance could show a similar trend. The miniature sensors of the future may well be based on micro-opto-electromechanical systems (MOEMS). Currently, the technology required to make a true MOEMS device with an optical read-out has not been reported.

This chapter sets out to summarise the advances in MEMS technology relating to sensors designed to measure both angular rotation and linear acceleration. The physical principles of the operation of these devices are considered along with an attempt to predict the future technical performance of these devices.

7.2 Silicon processing

The use of silicon to make the sensing elements is attractive, as it is a proven route to cost reduction through use of techniques developed in the semiconductor electronics industry for wafer processing. This industry has established very effective robust methods for high-yield and high-volume production that lead to precision low-cost components. Moreover, the mechanical properties of crystalline silicon are interesting as it has a fracture limit of 7 GPa, which exceeds that of many steel alloys; additionally, its density is quite low at 2390 kg/m^3 . Hence, crystalline silicon is very robust and ideal for this application.

A range of manufacturing techniques has been devised and perfected over the last three or four decades. These have concentrated on rapid and accurate etching approaches that avoid mechanical methods such as sawing, cleaving or filing. The modern methods lead to components with flat surfaces and an absence of undercutting. Efficient chemical 'machining' methods are very attractive, as they enable stress-free components to be produced rapidly. Moreover, the technique is capable of producing multiple components with identical dimensions to a high tolerance and having identical characteristics.

Silicon processing enables a robust design process to be devised. It is possible to reduce the fabrication process to three mask levels. For example, a $100 \mu\text{m}$ wafer can have a deposited oxide layer patterned with mask level 1, with a metal layer sputter deposited and patterned with mask level 2. Finally, photo-resist is spun and patterned with mask level 3 giving the shape of the resonator structure that is then generated through a dry deep-trench etch technique. On completion of the etching process the photo-resist is removed leaving a wafer of resonators.

The silicon wafer of resonators is anodically bonded to a pre-shaped glass wafer and then diced to give individual sensor elements. This design technique avoids the need to leave small gaps between the resonator and the surrounding material required for conventional processing; consequently the problems associated with stiction are mitigated. Additionally with some designs, such as the ring resonator where the vibratory motion is in one plane, all of the silicon processing is planar, thus avoiding the necessity for multi-layer processing.

Trade-off studies have shown that optimal gyroscopic performance is achieved with sensing elements having a thickness in the region of $50\text{--}100 \mu\text{m}$. It is considered

that continued evolution of the advanced micro-machining processes are required to build thicker and more three-dimensional parts that have less critical tolerances on the entire structure in order to reduce the cost of these sensors.

Another suitable material used as an alternative to crystalline silicon is quartz, and its application as the base material for MEMS inertial sensing devices is discussed later in the chapter.

7.3 MEMS gyroscope technology

7.3.1 Introduction

MEMS gyroscopes operate on a very similar principle to that already described for vibrating gyroscopes, in Chapter 4. However, it will be described again here, but in the context of a MEMS gyroscope.

MEMS gyroscopes are non-rotating devices and use the Coriolis acceleration effect on a vibrating proof mass(es) to detect inertial angular rotation. Thus, these sensors rely on the detection of the force acting on a mass that is subject to linear vibratory motion in a frame of reference which is rotating about an axis perpendicular to the axis of linear motion. The resulting force, the Coriolis force, acts in a direction, that is perpendicular to both the axis of vibration and the axis about which the rotation is applied. This is shown in Figure 7.2.

Whilst there are many practical sensor configurations based upon this principle, they fall generally into one of the three categories described below, and discussed previously in Chapter 4:

- simple oscillators;
- balanced oscillators (tuning fork gyroscope);
- shell resonators (wine glass, cylinder, ring oscillators).

Figure 7.3 provides a schematic illustration of these three angular sensors.

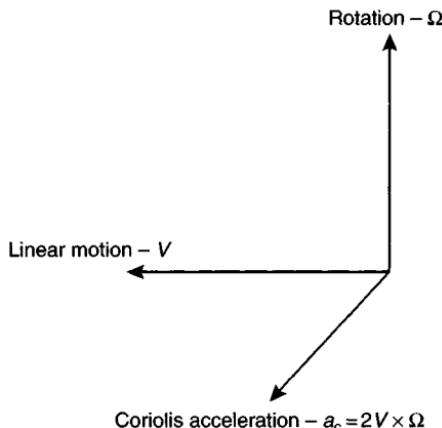


Figure 7.2 Generation of Coriolis force

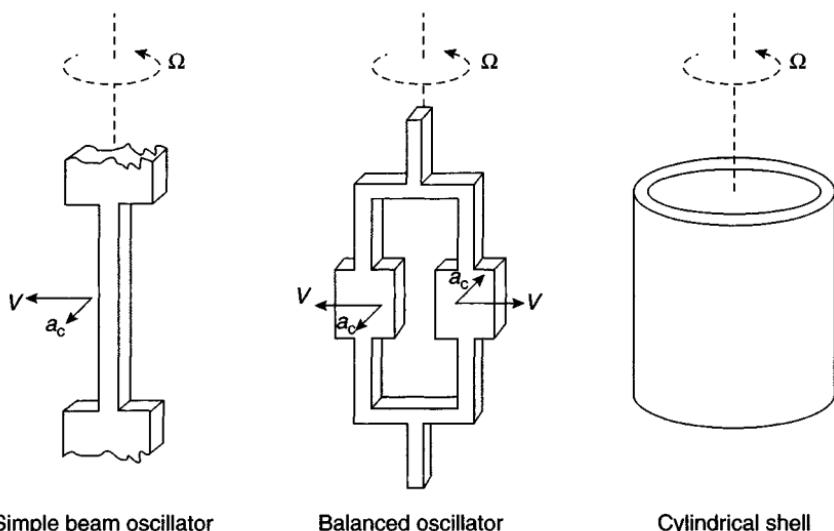


Figure 7.3 Classes of MEMS oscillator

Simple oscillators form the most basic class of vibratory gyroscopes; devices based on an oscillator that can be modelled as a single vibrating mass. The major drawbacks in this type of design arise through mechanical asymmetry, resulting in additional coupling between the axis of vibration and the output or measurement axis, and sensitivity to external vibration and interaction of the vibrating element with the structure on which the instrument is mounted. The latter effect is attributable to the reaction forces exerted by the case of the sensor on the vibrating member. One of the more successful sensor designs, which fall into this category, is the vibrating wire sensor, as described in Chapter 4 (Section 4.4.8).

The problems associated with the sensitivity simple oscillators to external vibration can be largely overcome by using a balanced oscillator. A device of this type that has received considerable attention over the years is the tuning fork gyroscope (TFG), and a number of MEMS sensors based upon this principle have been produced using either quartz or silicon in their construction. As discussed later in the chapter, it is the developments based around the TFG principle that have resulted in significant performance improvements in recent years.

The third category for this type of sensor is the vibrating shell or ring device, which is symmetric about the axis of rotation. Sensors that fall into this category include the wine glass gyroscope and devices based on a vibrating cylinder or ring. Such devices are relatively easy to manufacture to a high level of accuracy.

MEMS devices have been implemented based upon a vibrating ring structure. For such a ring, the modes of flexural vibration may be classified as in-plane or out-of-plane modes. They occur in degenerate pairs at a mutual angle of $90/p$ degrees, where p is the number of modal diameters. Existing devices measure angular rate about an axis perpendicular to the plane of the ring using the Coriolis coupling between

the in-plane modes. One mode is maintained at constant amplitude of vibration. The carrier mode may be excited by a number of methods including electromagnetic, electrostatic or piezoelectric means.

The vibrational behaviour of this type of sensor is considerably more complex than that of the simpler single mass or balanced oscillator devices outlined above. However, such devices are also less susceptible to the effects of applied vibration and the interactions between the vibrating element and the structure in which it is mounted.

As indicated above, the vibratory sensor types, which have been the focus of recent MEMS developments, have been based upon the principles of the tuning fork gyroscope and the vibrating ring device. Some examples of developments in this field of technology are described in the following sections.

7.3.2 Tuning fork MEMS gyroscopes

7.3.2.1 Silicon sensors

The focus for much of the development of advanced performance silicon tuning fork gyroscopes has been, and continues to be, The Charles Stark Draper Laboratories, Inc. in the United States. In 1992, a team of engineers succeeded in implementing a Coriolis vibratory gyroscope using MEMS technology employing silicon wafer photolithographic and chemical etching processes adapted from the electronics industry. Since that time, as part of a continuing development process, gyroscopes have been produced with an in-run bias stability in the region 3–10°/h over military operating temperatures, falling to tenths of a degree per hour in a temperature controlled environment with in-depth compensation techniques. Further dramatic improvements in performance are anticipated over the next five years.

Whilst much of this development has been directed towards military applications, driving the quest for higher performance devices, it has also been licenced for commercial exploitation. The technology lends itself to high volume, low cost production (\$10 per axis) for more modest performance applications. The largest user in the near term is the automobile industry, where applications include gyroscopes for anti-skid braking, steering control, roll detection and map navigation displays.

Principles of operation

The Draper micro electromechanical gyroscope, shown in Figure 7.4, consists of a silicon structure suspended above a glass substrate. The silicon structure contains two masses suspended by a sequence of beams that are anchored to the substrate at specific points. These two masses are made to oscillate 180° out of phase through the application of voltages to the outer comb motor drives. This feature has led to the designation of the sensor as a tuning fork gyroscope; see Section 4.4.4.

The application of an angular rate about the input axis, which is perpendicular to the velocity vector of the masses, gives rise to a Coriolis force that acts to push the masses in and out of the plane of oscillation. Because the instantaneous velocity vectors of the relative masses are equal and opposite, anti-parallel motion is induced in response to the Coriolis force. The resultant motion is measured by capacitor

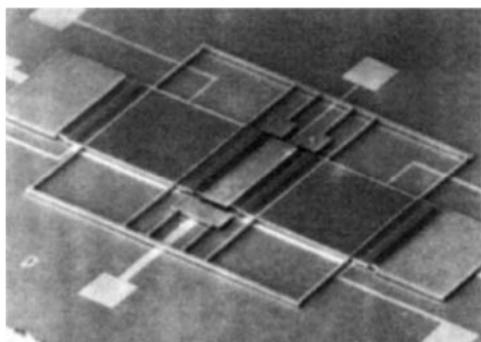


Figure 7.4 MEMS comb-drive tuning fork gyroscope (© The Charles Stark Draper Laboratory, Inc. All rights reserved. Reprinted with permission)

plates above and below each of the two masses, providing a signal proportional to the applied input rate. The operating elements of the gyroscope are shown schematically in Figure 7.5.

A simplified schematic is shown in Figure 7.6 to illustrate further the principle of operation. This diagram indicates the directions of vibration of the proof mass when the device is rotated about its input axis.

This device, also known as an in-plane sensing device,¹ is fabricated from a single crystal of silicon on Pyrex, the size of the proof mass element being $1000\text{ }\mu\text{m} \times 1000\text{ }\mu\text{m} \times 20\text{ }\mu\text{m}$. The operating frequency is $\sim 12\text{ kHz}$, and the amplitude of the motion imparted to the proof mass is $10\text{ }\mu\text{m}$ (peak). The sensor drive is electrostatic and a capacitive pick-off is employed to detect the output motion of the proof mass.

The following values are given to put into context the precision of the measurement task that needs to be undertaken in a MEMS gyroscope. For a typical device, a 1 rad/s input rate results in a Coriolis force of approximately $9 \times 10^{-8}\text{ N}$, a peak motion along the sense axis of 10^{-9} m , a 3 aF ($\text{aF} = \text{attofarad}; 1\text{ aF} = 10^{-18}\text{ F}$) peak change in capacitance and a charge generation of $15\,000\text{--}65\,000$ electrons.

As with all gyroscopes, the design of the electronics used to drive and control the sensors is a critical sub-system. This ‘component’ requires careful optimisation to give stable and repeatable output, in order to achieve good and reliable performance. For the particular device described here, the drive motor loop and the sense electronics are particularly important. The motor loop comprises a self-drive oscillator that provides voltages to generate an electrostatic force to induce motion of each proof mass. This motion is sensed and controlled to sustain a constant drive amplitude. Detection of the Coriolis induced motion is carried out using a highly

¹ Care must be taken in defining what is meant by ‘in-plane’ and ‘out-of-plane’ devices. ‘In-plane’ refers here to the orientation of the sensitive axis of the sensor, which in this case is parallel to the plane of the sensing element. It is noted that the resulting Coriolis motion of the sensitive element is perpendicular to the plane of the device (out-of-plane).

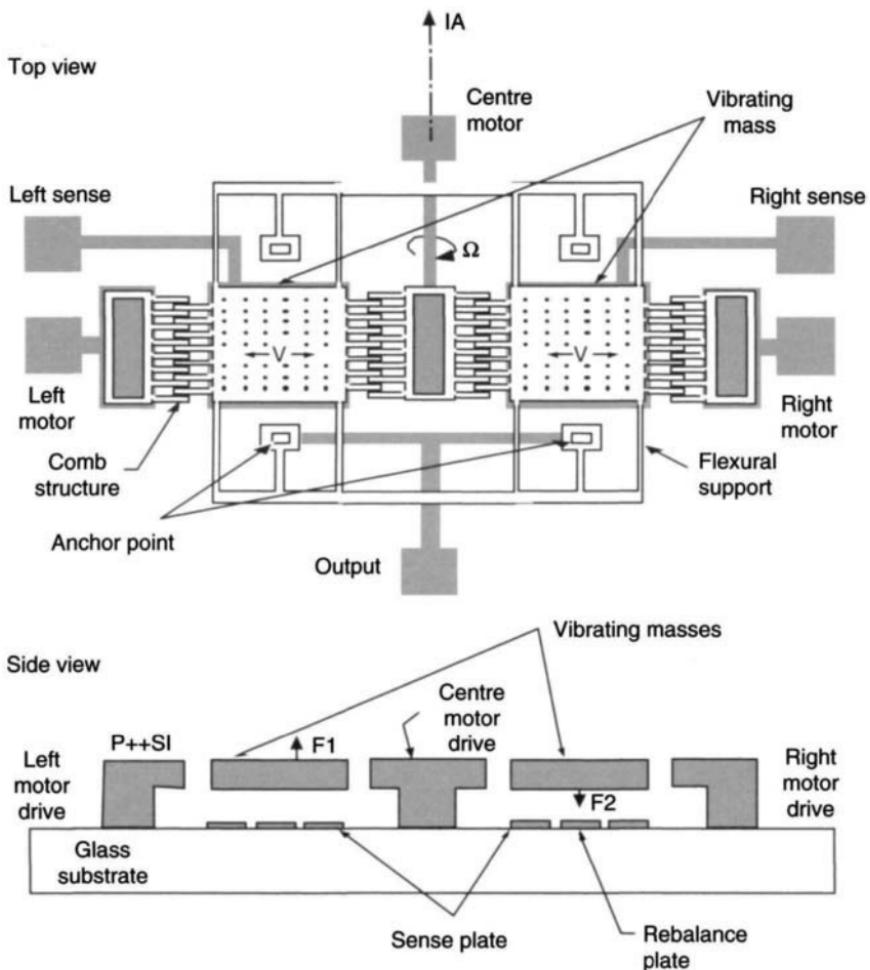


Figure 7.5 Operating elements of MEMS tuning fork gyroscope (© The Charles Stark Draper Laboratory, Inc. All rights reserved. Reprinted with permission)

sensitive capacitive pick-off. A pre-amplifier detects the charge generated through changes in proof-mass displacement. Existing devices operate in an open-loop mode, the output of the pick-off electronics providing a signal proportional to the input rate. A block diagram representation of the MEMS gyroscope electronics is given in Figure 7.7.

Sensor fabrication and packaging

Sensors are manufactured by the dissolved wafer process, illustrated in Figure 7.8. The first step (Mask 1) involves the etching of recesses in a doped-silicon wafer, which defines the height of the silicon above the glass substrate and the gap spacing

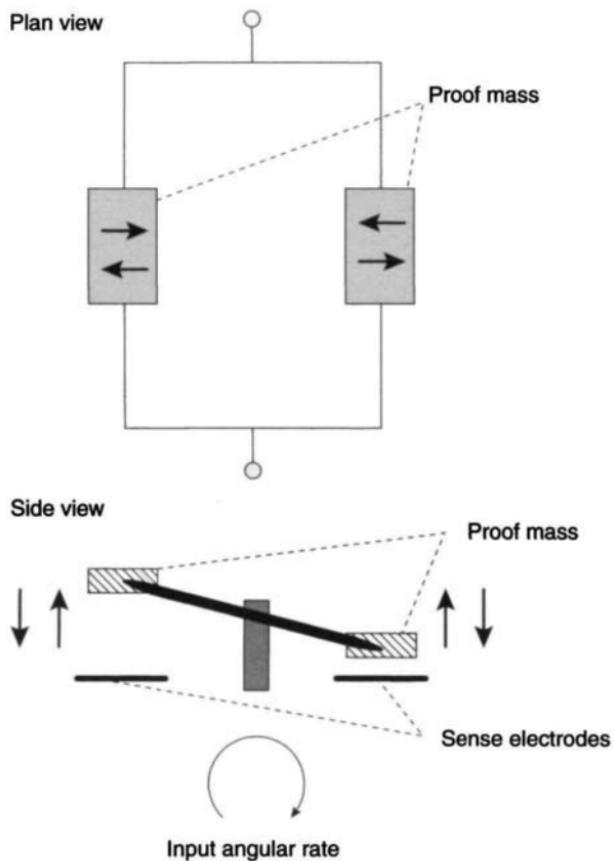


Figure 7.6 Motion of proof mass elements of a MEMS tuning fork gyroscope

for the capacitive sensing plates. A boron diffusion process follows, which defines the thickness of the structure. The pattern features are defined by Mask 2 and are then micro-machined using a reactive ion etching process. The glass wafer is processed separately, Mask 3 defining the glass recess and metal electrode pattern. The silicon wafer is then inverted and bonded to the glass wafer. This is followed by the final etching process to dissolve the un-doped silicon and leave the free-standing device. Hundreds of sensors are made on a single wafer.

The device is hermetically sealed in a package, which has a vacuum maintained in it to ensure the high-Q resonance is achieved to enable the desired operational characteristics. Leadless chip carriers (LCCs) with braze-sealed lids have been used in pilot production. Sensor chips are installed in packages using compression bonding and wire bond interconnections. The sensor package is then aligned mechanically on a printed circuit board containing the electronics. The resulting LCC is 0.25 in.² by 0.1 in. high.

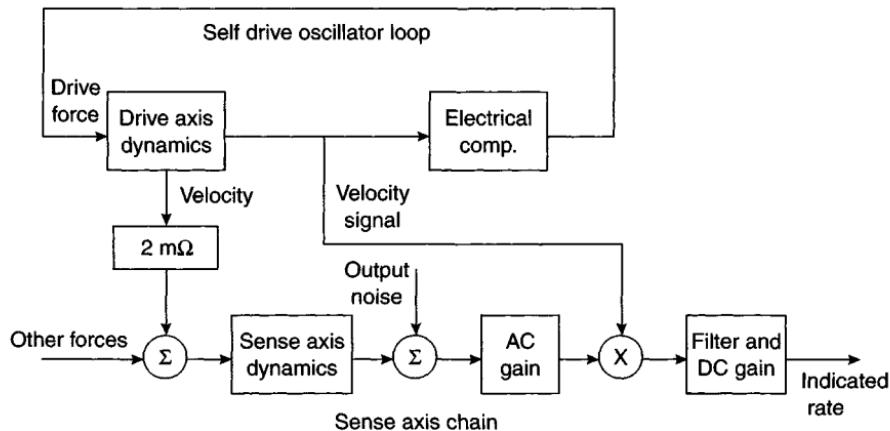


Figure 7.7 MEMS gyroscope electronics (© The Charles Stark Draper Laboratory, Inc. All rights reserved. Reprinted with permission)

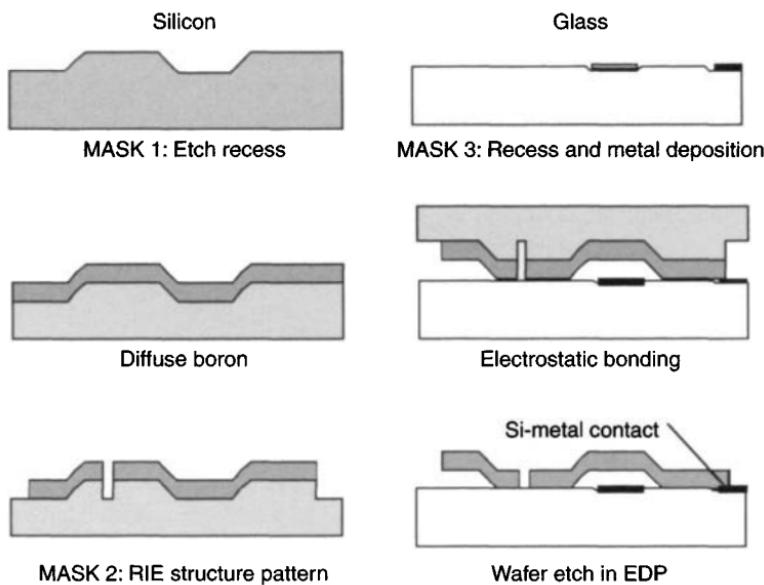


Figure 7.8 Dissolved wafer, silicon on glass process for gyro manufacture (© The Charles Stark Draper Laboratory, Inc. All rights reserved. Reprinted with permission)

Recent advances in the reactive ion etching process used in the manufacture of MEMS gyroscopes have resulted in designs incorporating silicon proof masses with increased component thickness up to 50 µm. The additional thickness of the silicon proof mass is expected to yield greater stability and improved device performance. In addition, recent devices include both upper and lower glass-sense plates to give

increased signal-to-noise ratio. Further details concerning the fabrication of these devices and their drive electronics can be found in the literature [1–3].

Performance

The MEMS gyroscope described here is capable of measuring rotation rates of several thousand degrees per second, whilst having the resolution to detect small fractions of a degree per hour. Typical performance figures are tabulated in Table 7.1.

It is stressed that current MEMS sensors are heavily reliant on pre-run characterisation followed by calibration to remove turn-on errors and so achieve high performance.

As with other solid-state technologies, such instruments are able to withstand high levels of mechanical shock and vibration, and hence are applicable for operation in hostile environments. MEMS gyroscopes and accelerometers have already found application in guided-munitions' applications where they must withstand and operate in the presence of launch accelerations of many thousands of gs. This feature combined with their capability to provide precise rotational measurements, opens up many new areas of application, many of which require in-depth investigation.

Since this is an emerging technology, it is felt to be appropriate to give some indication of performance expectations, as indicated by the performance goals given in Table 7.1. Gyroscopic performance improvements are expected to come about through the use of thicker and larger component parts; increased thickness of the silicon proof-mass in gyroscopes. This has become possible as a result of recent improvements in reactive ion etching techniques, allowing straight sidewall and flatness tolerances to be met. This in turn leads to reduced susceptibility to fabrication tolerances, higher scale-factor and greater sensor performance stability. Although higher performance

Table 7.1 MEMS tuning fork gyroscope performance figures

	Current sensors	Performance goals	Comments
Operating range ($^{\circ}/\text{s}$)	100–6000	100–6000	Selectable
Turn-on bias stability ($^{\circ}/\text{h}$)	10–150	<1	All environments
In-run bias stability ($^{\circ}/\text{h}$)	3–30 0.3–10	<1 <0.1	–40 to 85°C 5°C temperature range
Turn-on scale-factor stability (ppm)	500–1500	<100	All environments
In-run scale-factor stability (ppm)	300–1500 100–300	<100 <10	–40 to 85°C 5°C temperature range
Angle random walk ($^{\circ}/\sqrt{\text{h}}$)	0.01–0.3	0.01–0.03	Lower random walk at lower maximum input rate
<i>g</i> -Sensitivity ($^{\circ}/\text{h/g}$)	10	0.5	

devices will be physically larger as a result of the developments outlined above, these changes will yield sensors that are still quite small in size and capable of fulfilling size requirements for the types of application in which they are expected to be used.

7.3.2.2 Quartz sensors

A number of MEMS gyroscopes use quartz as the base material for the sensing element. The use of piezoelectric quartz material simplifies the sensing element, resulting in a reliable and durable sensor that is stable over both temperature and time.

An example of this type of device is the quartz rate sensor (QRS) produced by the Systron Donner Inertial Division of BEI Technologies, Inc. [4], and described earlier in Section 4.4.5. A picture of this device, which is based around an H-shaped quartz crystal, is shown in Figure 7.9. MEMS versions of these devices are in high volume production for automobile applications, as well as uses in platform stabilisation and smart munitions; a high-*g* version has been produced for the latter application.

This device comprises a pair of coupled tuning forks, the drive tines and the pick-up tines, along with their support flexures and frames that are batch fabricated from thin wafers of single-crystal piezoelectric quartz. The piezoelectric drive tines are driven by an oscillator to vibrate at a precisely defined amplitude, causing the tines to move toward and away from one another at a high frequency. This vibration causes the drive fork to become sensitive to angular rate about an axis parallel to its tines, defining the true input axis of the sensor.

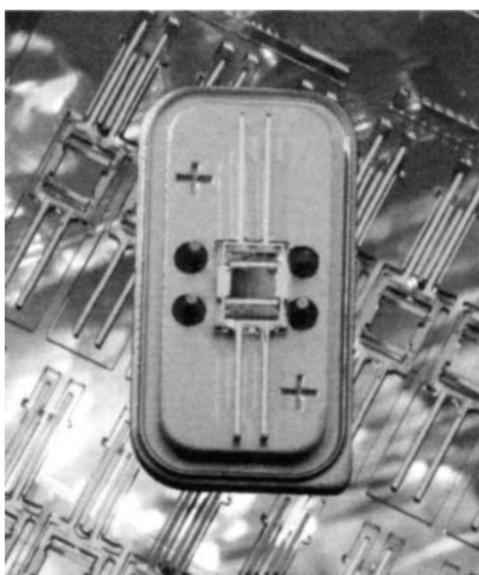


Figure 7.9 Quartz rate sensor (Courtesy of BEI Systron Donner Inertial Division)

Vibration of the drive tines causes them to act like the arms of a spinning ice skater, where moving them in causes the skater's spin rate to increase, and moving them out causes a decrease in rate. In the presence of tine vibration, an applied rotation rate about an axis parallel to the tines gives rise to a torque about the sensitive axis of the device that varies sinusoidally at the frequency of oscillation of the drive tines. The pick-up tines respond to this oscillating torque by moving up and down, out of the plane of the fork assembly, at the frequency of the drive tines. The motion of the pick-up tines is sensed, giving rise to an alternating electrical signal that is demodulated to produce a d.c. output proportional to the applied turn rate.

A further sensor that uses a quartz element is the Sagem Quapason gyroscope which has four quartz tines extending upwards from a common base. The advantage of this device is its ability to reduce unwanted cross-coupling from drive to sense channels [5].

7.3.3 *Resonant ring MEMS gyroscopes*

Vibrating ring structures have been successfully used to detect turn rates applied about an axis that is perpendicular to the plane of the ring using Coriolis force coupling between in-plane displacements. Such sensors have an advantage in that the ring structure maintains the drive and sense vibrational energy in a single plane. However, such devices do suffer from the drawback of having a relatively low vibrating mass, and hence exhibit a low scale-factor.

The initial work on this class of gyroscope involved the design of a vibrating element based on a gyroscope with a piezoceramic cylinder to sense the angular motion. The technology evolved into a device with a metal ring or disc resonator as the sensing element of the device, as shown in Figure 4.17 and described in Section 4.4.3. The design evolved through the use of silicon instead of metal [6–9].

An example of a MEMS implementation of this type of technology is the BAE Systems' silicon vibrating structure (SiVSG), which comprises a silicon ring supported by eight spokes that are radially compliant [10], as shown in Figure 7.10.

The silicon gyroscope creates and sustains a resonance in the structure through a combination of currents flowing through the metallic tracks on the surface of the ring and a magnetic field perpendicular to the plane of the sensor. This is analogous to an electric motor for the drive and a generator for the pick-off. The diameter of the ring is 6 mm, and the fundamental vibration mode of interest occurs at 14.5 kHz. The sensor chip is anodically bonded to a supporting glass structure, which has its thermal coefficient of expansion matched to that of silicon.

The sensing element has eight identical conducting loops that follow a similar path along metallic tracks on the surface of the silicon. These activation and pick-off circuits pass from the bond pad, along the top of one support leg, around one-eighth of the silicon ring, along the length on the next supporting leg and back to the bond pad. Each supporting leg has three conductors running along its length, one from each adjacent loop and between them a single conductor to minimise capacitive coupling. A ground plane is also created in this device by connecting to the silicon substrate.

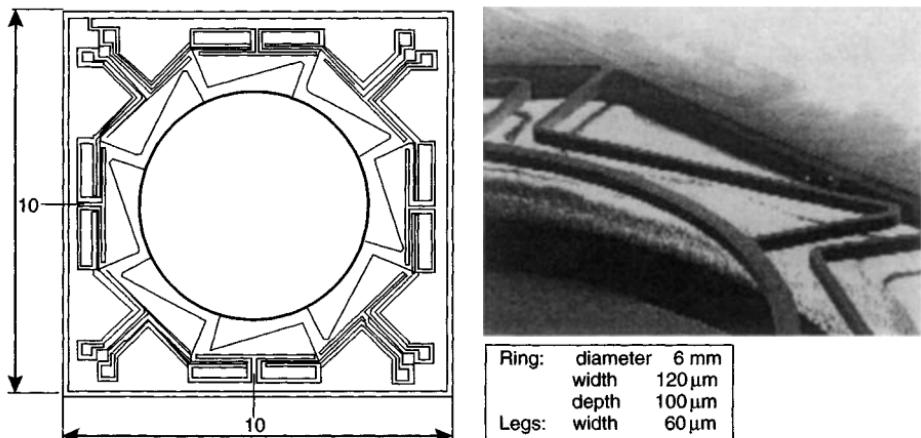


Figure 7.10 Silicon vibrating structure gyroscope sensitive element (Courtesy of BAE Systems)

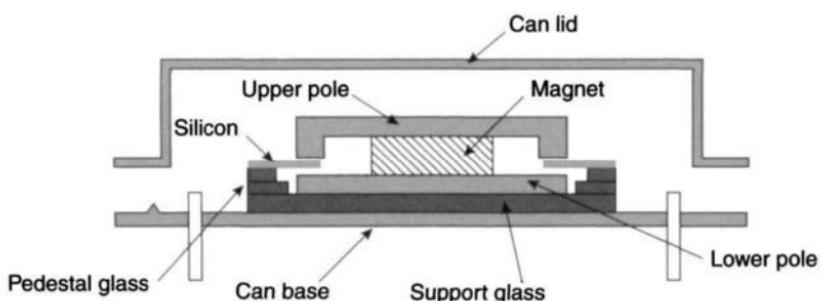


Figure 7.11 Cross-section of MEMS sensor (Courtesy of BAE Systems)

These conduction paths and a magnetic field, which is perpendicular to the plane of the silicon ring, form the drive and pick-off elements of the sensor. As usual, these transducers are arranged in pairs, 180° apart. The four transducers are the primary mode drive, the secondary mode drive, the primary mode pick-off and the secondary mode pick-off. A samarium cobalt magnet and a magnetic circuit, with two pole pieces, provide the magnetic field, which is confined to the size of the ring. The magnetic circuit maximises the magnetic field at the ring. A schematic of the layout is shown in Figure 7.11.

The sensing element has been designed so that the lowest order vibration mode is in excess of 5 kHz; this is whole-ring motion with respect to the mount. Consequently, this type of vibratory sensor is insensitive to the normal bandwidth of ambient vibratory motions experienced in many applications. Moreover, the design of the planar-ring structure is such that all of the vibratory motion is in one plane, whilst

being subjected to angular motion so there is no coupling of vibrations from one crystal plane to another. Hence, performance parameters such as frequency, modal-frequency split and Q are very stable over a wide temperature range.

The sensing element is operated in a closed loop mode to give high performance. The primary vibration mode is controlled by the primary drive loop to ensure that the vibration mode is at resonance; there is also an automatic gain control loop to control the magnitude of the displacement. When the sensor is subjected to angular disturbance, energy is coupled from the primary mode to the secondary mode. This motion is sensed by the secondary pick-off and this signal provides a measure of angular rate, that is, this would be an open-loop device architecture. However, in this design the secondary drive loop is used to null this motion, and this drive current provides the measure of the allied angular rate.

This form of feedback control aims to provide a constant modal pattern in the vibrating sensor during angular motion. This condition leads to improved linearity of the scale-factor and reduced bias. The combination of the automatic gain control (agc) loop and the secondary drive loop removes any ' Q ' (quality factor) dependence from the scale-factor. This secondary loop is quite complex as it has two parts, a rate loop and a quadrature loop to null the quadrature motion in the resonator. The control of the so-called quadrature motion effects helps to minimise errors in the measurement of the applied angular motion about the input axis, resulting from frequency differences between the primary and secondary modes.

It is worth noting that in an open-loop configuration the scale-factor would have a Q^2 dependence.

It can be shown that the scale-factor is proportional to

$$\frac{\omega V_{\text{agc}}}{g_{\text{ds}} g'_{\text{p}} B^2}$$

The rate equivalent noise is given by:

$$\frac{K_1 f_n V_n}{g_{\text{d}} Q B^2 V_{\text{pd}}} \sqrt{\frac{4 f_{\text{BW}}^3}{3}}$$

where V_{pd} is the potential applied to the primary, V_n is the pick-off noise, V_{agc} is the potential of the agc loop, f_n is the resonant frequency ($\omega/2\pi$), f_{BW} is the output bandwidth, ω is the angular frequency, Q is the resonance quality factor, B is the magnetic field, K_1 is a term that includes the dimensions of the resonator and the mode shape (Bryan factor), g_{d} is the gain of the current amplifier, g_{ds} is the gain of the secondary amplifier and g'_{p} is the derivative of the gain of the pick-off amplifier.

This theory predicts that a sensor with a 40 Hz bandwidth will have an rms noise figure of the order of $0.2^\circ/\text{s}$, reducing to $0.025^\circ/\text{s}$ for a sensor with a 0–10 Hz bandwidth.

This sensor has been developed for commercial applications by BAE Systems in conjunction with Sumitomo Precision Products Company Limited. The resulting device, fabricated using silicon, lends itself to batch production and is relatively inexpensive to produce. Whilst the sensor is suited to a wide range of

Table 7.2 SiVSG performance figures

Parameter	Value	Comments
Operating range ($^{\circ}/\text{s}$)	± 1000	
Turn-on bias stability ($^{\circ}/\text{s}$)	<0.06	1σ
In-run bias stability ($^{\circ}/\text{s}$)	0.05	1σ (0–30 min)
Scale-factor stability temperature sensitivity (%)	< ± 1	–40 to +85°C
Scale-factor linearity (% of FS)	<1	Input rates $\pm 100^{\circ}/\text{s}$
Noise ($^{\circ}/\text{s rms}$)	<0.5	0–45 Hz

commercial applications, particularly automotive, it has been designed to withstand the environments typical of military and space applications.

A summary of key performance figures is given in Table 7.2.

More recent developments have explored the possibilities for detecting angular rates applied about three mutually orthogonal axes using the Coriolis coupling between in-plane and out-of-plane displacements.

Along with BAE Systems, other corporations, such as Delphi² and QinetiQ are developing ring gyroscope components which stimulate the primary ring mode using electrostatic actuation techniques. Such implementations will offer significant advantages in the future, such as reduced power consumption, less complex assembly and a reduction in component size.

For multi-axis operation, the additional response of the ring owing to angular velocity applied about axes in the plane of the ring has been examined. In a similar manner to the single-axis ring resonator gyroscope, a carrier mode is maintained at constant amplitude. The carrier mode can be either an in-plane or an out-of-plane mode. When subject to a turn rate, Coriolis coupling is induced between the carrier mode and one or more response modes (in-plane or out-of-plane), depending on the axis about which the rate is applied. The magnitude of the induced response is proportional to the applied rate. The implementation of a number of concepts for sensors having two-axis and three-axis rate sensitivity is the subject of continuing research at the present time.

7.4 MEMS accelerometer technology

7.4.1 Introduction

As noted above, the use of silicon to make the precision micro parts of an accelerometer to measure the specific force being applied to an input axis is well established. The current state of development is that the entire sensor may be constructed entirely

² Originally developed in conjunction with Michigan University, USA.

from silicon, with the exception of the hermetically sealed case, which still tends to be metallic. Modern designs and precision micro-machining enable a precision sensor to consist of as few as five parts.

MEMS devices may be divided into two distinct classes, reflecting the manner in which acceleration applied to the case of the device is sensed:

- the displacement of a proof mass supported by a hinge or flexure in the presence of an applied acceleration, that is, a mechanical sensor using silicon components;
- the change in frequency of a vibrating element caused by the change in tension in the element as a result of the mechanical loading that occurs when the element is subjected to acceleration.

These MEMS devices are analogous to the pendulous open-loop and force-feedback accelerometers, and the vibrating-beam sensors described separately in Chapter 6.

As a general guide to the current state of technological development of these two distinct classes of inertial sensor, it may be considered that:

- pendulous types of MEMS accelerometer can provide acceleration measurements to an accuracy compatible with inertial (25 micro-g) or sub-inertial quality (1 milli-g);
- vibrating beam devices or resonant sensors tend to have a potentially higher accuracy capability approaching 1 micro-g.

7.4.2 Pendulous mass MEMS accelerometers

Both out-of-plane (sometimes known as z-plane) and in-plane pendulous devices have been developed and are in quantity production. Figure 7.12 shows a typical out-of-plane MEMS accelerometer in which a hinged pendulous proof mass, suspended by torsional spring flexures over a glass substrate, rotates when subjected to acceleration perpendicular to the plane of the sensor. One of the major attractions of this type of sensor is the versatility of the packaging, which enables planar mounting of the sensor.

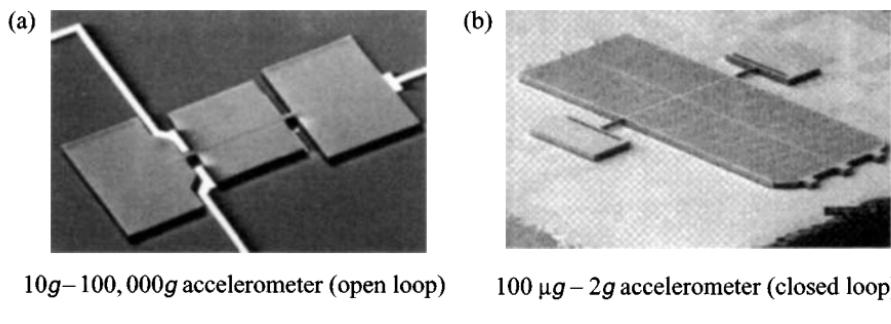


Figure 7.12 MEMS out-of-plane pendulous accelerometer (© The Charles Stark Draper Laboratory, Inc. All rights reserved. Reprinted with permission)

The device shown is a closed loop sensor having a dynamic range of 100 micro-g to 2g. Both open and closed loop devices have been developed, and open-loop devices capable of measuring 100 000g are available.

Motion is detected by the change in the capacitance gap between the proof mass and the substrate using electrodes on an insulator substrate. Under a 1g acceleration, the change in angle of the proof mass is typically 70 µrad resulting in a change in the sense gap of approximately 3×10^{-8} m corresponding to a peak change in capacitance of 15 fF (fF = femtofarad; 1 fF = 10^{-15} F). Typical measurement ranges are from 100 micro-g to 15g. To achieve this dynamic range, it is necessary to resolve motion of 3×10^{-12} m, or about 22.5 electrons charge change on the proof mass per carrier cycle.

Careful characterisation of this type of sensor is required as the scale-factor tends to decrease with increasing temperature [1]. The temperature dependence is systematic, and approximately linear, so it is easy to model and correct through a compensation routine.

Devices of this type, such as the Northrop Grumman SiAc™, have found broad application in a range of military applications, both tactical and inertial grade sensors having been produced. Similar devices have been produced through technical collaboration between The Charles Stark Draper Laboratory, Inc. and Honeywell. These are silicon devices which have been evaluated in extended range guided-munitions applications.

Figure 7.13 shows an in-plane (lateral) accelerometer which uses a comb finger construction as its sensing element. Acceleration is measured by detecting the change in capacitance across the comb fingers. This class of ‘mechanical’ sensor is much more sensitive to acceleration applied in the horizontal plane (as shown – or left to right) than in the orthogonal direction (or top to bottom).

Clearly, if a very small package is required then the use of a combination of these two types of pendulous accelerometer will give a very small and planar package that can sense acceleration along two or three axes. For example, the use of an out-of-plane sensing element and two lateral in-plane sensors, with their input axes orthogonal, will give a triple-axis sensor on a chip.

The analogue devices ADXL150 and ADXL250 are examples of this type of accelerometer. It is noted that the combination of in-plane and out-of-plane devices of the type described here allows systems capable of measuring acceleration along three mutually perpendicular axes to be constructed using a minimal volume package.

7.4.3 Resonant MEMS accelerometers

This class of sensors covers the general category of vibrating beam accelerometers which can be configured to sense accelerations acting in directions either in the plane of the sensor or perpendicular to it. Acceleration is sensed as a result of the change in the resonant frequency of beam oscillators under inertial loading of a proof mass, rather than the measurement of its displacement.

Out-of-plane resonant devices have been made by micro-machining at least one of the flexure members in a structure to form a piezoelectric resonator. The etching

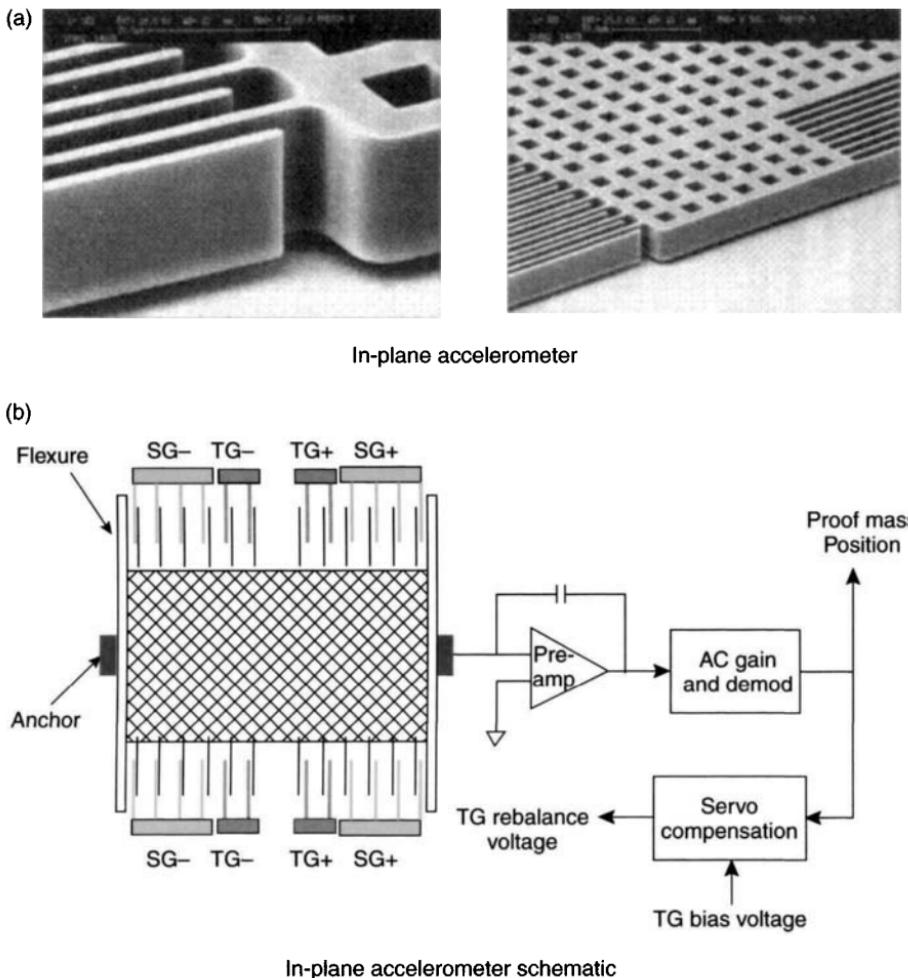


Figure 7.13 MEMS lateral in-plane accelerometer (© The Charles Stark Draper Laboratory, Inc. All rights reserved. Reprinted with permission)

of the structure is undertaken at positions where the resonator is calculated to have a zone, or zones, of high stress in its modal pattern. As the flexure element is distorted under the displacement of the proof mass the resonant frequency of the structure changes.

Both silicon and quartz devices have been fabricated. Figure 7.14 shows the in-plane (lateral) vibrating beam structure of a silicon oscillating accelerometer (SOA) developed by The Charles Stark Draper Laboratory, Inc. In this case the fundamental configuration is a monolithic vibrating tuning fork structure with a large silicon proof mass, which is driven electrostatically. The beams are loaded axially when an acceleration is applied in the plane of the vibratory motion (in the plane of the wafer) and the resonant frequency changes. Oscillator resonance and sensing is accomplished using a silicon comb drive structure as shown in the figure.

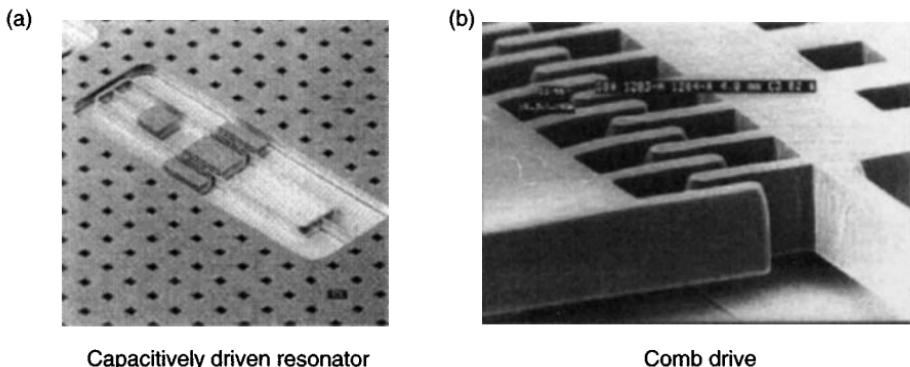


Figure 7.14 Silicon oscillating accelerometer (© The Charles Stark Laboratory, Inc. All rights reserved. Reprinted with permission)

The silicon oscillating accelerometer (SOA) MEMS accelerometer manufacturing process is silicon on glass; the silicon is perfectly elastic allowing very high precision frequency control and stability. The sensor is contained within a ceramic vacuum package, which provides high oscillator *Q* factor (typically $>100\,000$). For a device having a nominal oscillator frequency of 20 kHz and a scale-factor of 100 Hz/g, a frequency stability of 5 parts per billion yields a bias stability of 1 micro-g.

A device developed in France by ONERA³ is of particular interest. This sensor has a mechanical isolating system that thermally isolates the vibrating beam from the mounting base and protects the sensing element from thermal stress caused by the differences in the thermal expansion coefficients of quartz and the case material. This class of accelerometer is capable of inertial-grade performance: accuracy of up to 1 micro-g bias stability has been reported.

7.4.4 Tunnelling MEMS accelerometers

This class of MEMS accelerometers is a recent development that offers significant enhancements over the devices described above that use read-out methods based on measurement of changes in capacitance. In this type of accelerometer the read-out has very high sensitivity and consequently offers better resolution, higher bandwidth and reduced size.

The control electrode deflects the cantilevered beam using an electrostatic force into a position known as the tunnelling position, a deflection of less than 1 μm . The device has a servo system that holds the beam in a position that maintains the gap between the tunnelling tip and the surface of the beam, and thereby maintains the tunnelling current, typically of the order of 1 nA. When an acceleration force is applied it attempts to move the beam and the servo system changes the applied potential at the electrode, and this change is a measure of the applied acceleration. A schematic view of this type of sensor is shown in Figure 7.15. The construction uses

³ Office National d'Etudes et de Recherches Aérospatiales.

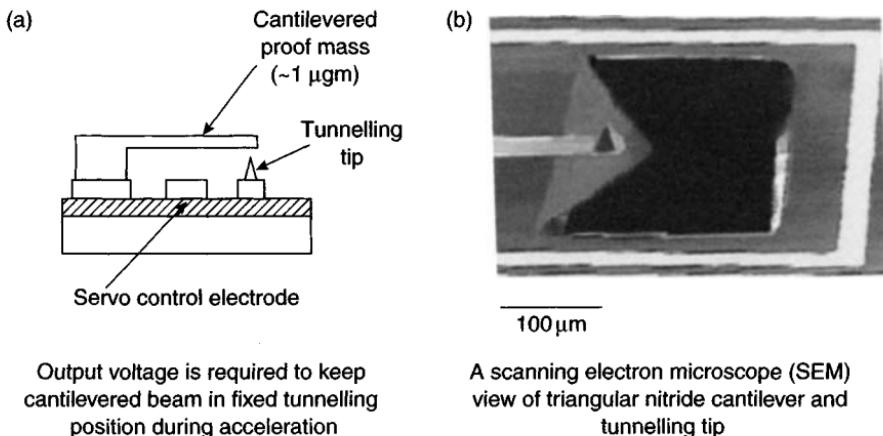


Figure 7.15 Schematic of a MEMS tunnelling accelerometer (© The Charles Stark Draper Laboratory, Inc. All rights reserved. Reprinted with permission)

a structure with low resonant frequency proof mass cantilever beams with a mass of the order of $1 \mu\text{g}$ and read-out circuits with sub-Angstrom resolution.

This class of sensors is designed to have a resolution in the nano-*g* regime. However, the dynamic range is relatively modest at approximately 10^6 , so the maximum input acceleration is in the 1 milli-*g* range.

7.4.5 Electrostatically levitated MEMS accelerometers

There have been many diverse studies and developments aimed at removing the constraints imposed by the elastic restraint and non-linear response of supporting mechanisms. The dynamically tuned gyroscope (Section 4.2.6) and the mass unbalance sensor (Section 6.5.4) are two particularly successful devices. However, the sensing element is only free over a very small angle. Novel gyroscopes such as the nuclear magnetic resonance device (Section 4.5.1) and the electrostatically suspended sensor (Section 4.6) are other examples aimed at very specialised applications.

A new approach involving MEMS technology and electrostatic levitation aims to produce a high performance accelerometer with high sensitivity, very accurate resolution and an easily adaptable response bandwidth, without modification to the structure. Additionally, the use of levitation of a small proof mass could ease some of the demanding fabrication tolerances. As with all of these novel approaches, there is a shift in the design difficulty and in this case it is to the control loop which becomes complex. However, with modern computational techniques and electronic systems this should not be an insurmountable problem.

This new approach has a small sphere 1 mm diameter with a mass of 1.2 mg suspended in an electrostatic field. The position of this sphere is sensed by changes in capacitance and a closed loop servo system maintains the position of the sphere by

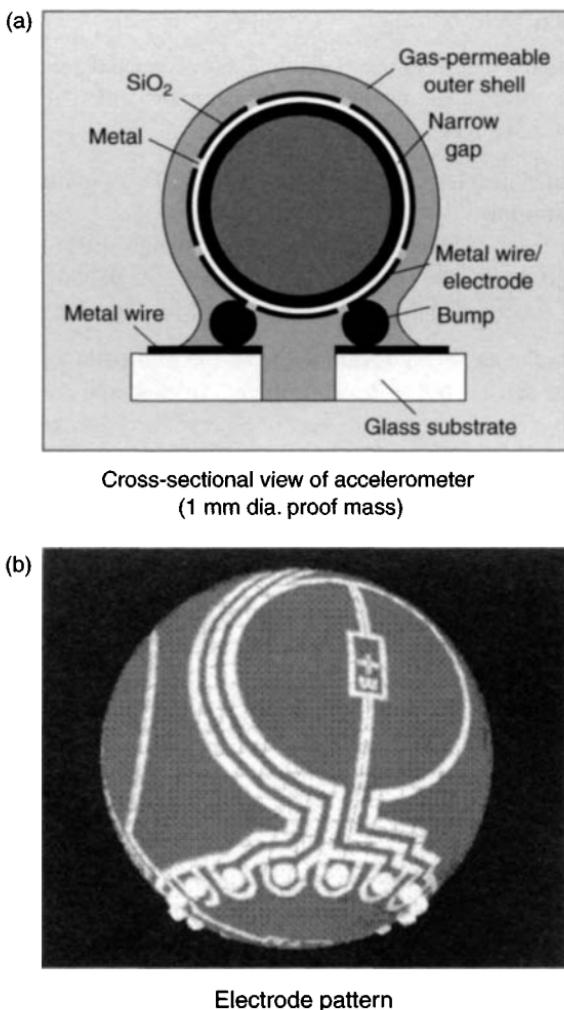


Figure 7.16 Levitation accelerometer (© The Charles Stark Draper Laboratory, Inc. All rights reserved. Reprinted with permission)

controlling the electrostatic force on the sensing element. A schematic representation of the sensor is shown in Figure 7.16.

The gap between the sphere and the outer shell is formed by removal of a sacrificial layer of polycrystalline silicon, which is etched through the outer shell structure.

An alternative approach using a levitated disc is under consideration at the University of Southampton in the United Kingdom [11]. This class of technology with a levitated spinning mass sensing element could lead to a very capable inertial sensor.

This type of sensor is aimed at the measurement of micro-gravity measurements in space and is expected to have a noise value in the region of $40 \text{ micro-}g/\sqrt{\text{Hz}}$.

7.4.6 Dithered accelerometers

There has long been a desire to use a single class of inertial sensor to provide all of the inertial measurements for an inertial measurement unit (IMU). Some examples of approaches that have had some success include:

- use of multiple accelerometers, but this is difficult owing to the real-time processing required;
- use of mass unbalanced tuned rotor gyroscopes, which also requires quite complex processing and the sensor package is quite large; see Section 6.5.4;
- use of multi-sensors, which again are not small sensors; see Section 6.5.

A novel technique has been developed from the multi-sensor approach to enable angular rate to be sensed by an accelerometer. In this sensor, the three opposing pairs of monolithic resonating beam accelerometers are dithered about an axis on a vibrating structure. Acceleration applied about the input axes is sensed from the change in resonant frequency of the vibrating beams in the device. The angular motion is sensed from synchronous demodulation of the Coriolis force acting on the accelerometers.

This device has been developed in the form of the micro-machined silicon Coriolis Inertial Rate and Acceleration Sensor (μ SCIRAS) [13].

7.5 MOEMS

A new approach to micro-machined sensors is micro-opto electromechanical systems (MOEMS). This class of technology offers a true solid-state sensor with an optical readout, so the limitations in performance of MEMS devices using capacitative techniques for measuring small displacements are eliminated.

Various optical pick-off techniques are current topics of research. These are either interferometric approaches, which offer low noise and high resolution, or attenuation methods, involving the interruption of a light beam from a diode. These approaches, or others, may be adapted as the characteristics and origin of the noise are understood, and fully characterised. Another important aspect of an optical read-out is the installation and alignment/harmonisation of the optical source and its detector; particular considerations are low-cost installation and maintainability.

7.6 Multi-axis/rotating structures

The approach of mounting in-plane sensors on a rotating ceramic mounting block and using demodulation techniques to extract the multi-axis inertial data from a reduced number of sensors, has enabled a considerable size reduction in multi-axis systems to be achieved. Off-axis sensors may be used for measuring orthogonality.

The use of the more modern in-plane and out-of-plane sensors is likely to lead to a further reduction in IMU size and make the need for a rotating structure redundant for many applications.

7.7 MEMS based inertial measurement units

A number of projects are underway to produce multi-axis sensors on a chip, which will provide estimates of angular rate as well as linear acceleration. Examples include the combination of two in-plane and one out-of-plane MEMS sensors on a single chip. Such approaches offer many advantages including:

- ease of manufacture;
- a vast reduction in volume;
- negligible power consumption;
- ability to carry out a complete characterisation of the unit in a single operation.

7.7.1 Silicon IMU

The 'Draper' laboratory has reported one chip that has two tuning fork gyroscopes and one out-of plane gyroscope. A complementary chip with three accelerometers has been produced; this integrated sensor also has two in-plane accelerometer sensors and a single out-of-plane pendulous device. The resulting inertial measurement unit has a volume of about 3.3 cm^3 [1]. However, further work is required before a high performance IMU is available.

BAE Systems has developed a silicon-sensor based IMU for a number of applications [12]. The IMU has three sensing axes, which are in a right-handed orthogonal set. Each sensor axis has an associated set of devices to sense angular rate about that axis, acceleration along that axis and temperature of the sensors. An exploded view showing the layout of the components is shown in Figure 7.17.

This IMU has been designed to operate in an environment that is subjected to high angular rates. The inertial sensors are arranged on the edges of a cube to form

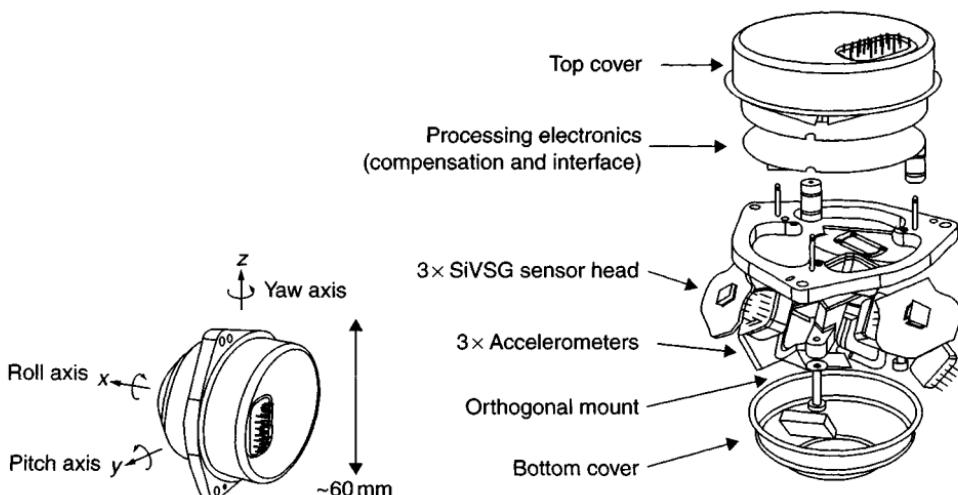


Figure 7.17 SiIMU® (Courtesy of BAE Systems)

a right-handed orthogonal set of axes. This axis system is rotated away from the vertical; however, the relative orientations of the sensors with respect to each other are not skewed. This configuration has the input axes of all sensors at 54.73° to the vertical and is equally spaced at 120° to each other in the horizontal plane.

A direction cosine matrix is used to convert the sensed motion by each of the sensors into a body reference frame. This approach is necessary as no one sensor is aligned with the body axis reference frame in this configuration of sensor axes.

This skew or rotated axis configuration enables

- the x -axis (of the vehicle) to sense up to $\sqrt{3}$ times the maximum input angular rate of the gyroscope, when the input rates about the other two orthogonal body axes are zero; similarly for the accelerometer;
- common mode errors, such as sensor bias, to be magnified by $\sqrt{3}$ for the x -axis, but reduced for the orthogonal axes.

The downside of this configuration is that if the vehicle containing the IMU is subjected to the same maximum rates simultaneously, then this configuration will require sensors with a higher maximum rate capability than would be needed for the non-rotated cluster.

One of the aspects for achieving high navigational data capability from these small sensors is to undertake characterisation and modelling of the sensors. This type of sensor requires equipment that allows application of maximum temperature, acceleration and rates to at least two independent axes. Hence, a single-axis rate table is inadequate, as scale-factor errors and misalignments need to be observed independently; therefore dual-axis rate tables are used in the characterisation.

This IMU compensates for the following systematic errors:

- sensor bias;
- sensor scale-factor errors;
- ‘ g -dependent’ bias of the silicon gyroscopes;
- misalignment of axes from harmonisation errors and non-orthogonality;
- size effect.

The compensation routine for compensating the above errors runs in real time. This leads to a unit with a run-to-run bias of $200^\circ/\text{h}$ (1σ) and ‘ g ’-dependent bias of $7^\circ/\text{h}/g$ (1σ). The corresponding performance for the measurement of acceleration is 20 milli- g (1σ).

7.7.2 Quartz IMU

The quartz rate sensor (QRS), described in Section 7.3.2.2, together with a vibrating quartz accelerometer have been built into an IMU; the Systron Donner Digital Quartz IMU (DQI). This is a solid-state, six-degree of freedom inertial measurement system that provides measurements of angular rate and linear acceleration about three orthogonal axes [4]. The integrated unit combines the inertial sensor assembly, the inertial sensor assembly electronics and a processor in which digital filtering and compensation algorithms are implemented.



Figure 7.18 BEI C-MIGITS™ III miniature INS/GPS integrated system (Courtesy of BEI Systron Donner Inertial Division)

The inertial sensor assembly is machined out of a single aluminium block to form a cube, and is designed to eliminate structural resonances that would coincide with any of the sensor drive modes. To further eliminate coupling between the individual sensors, each sensor operates at a unique drive frequency. One of the advantages of using quartz sensors is the stability of this material, which allows each gyroscope to be produced to a known frequency which can be accurately maintained.

MIGITS™ is a family of products for applications requiring guidance, navigation and control. The C-MIGITS™ III system, shown in Figure 7.18, is a compact and lightweight system that contains the Systron Donner DQI and a commercial GPS receiver. The inertial system and GPS measurements are combined using a tightly coupled integration architecture; discussed separately in Section 13.7.

7.8 System integration

An array of MEMS sensors may be integrated into a single chip to provide multiple independent measurements of inertial motion. A direct advantage of this class of technology is that the sensors may be integrated directly to the electronic control circuits, in an applications specific integrated circuit (ASIC) configuration, in a single

hermetically sealed package. The logic control may be achieved by the use of field-programmable gate arrays.

The characteristics of the MEMS devices tend to be of a form that is readily corrected leading to a sub-system that can be hard mounted in a vehicle. These devices have been used in cannon-launched projectiles that experience many thousands of 'g' during the launch of the weapon. The power requirements are modest and compatible with normal electronic circuitry.

MEMS devices are ideal for integration with other navigation systems, both in terms of complementary error characteristics and small size with rugged operational characteristics.

7.9 Summary

MEMS sensors have developed substantially over the last 15 years in many respects. A number of new approaches have been very successful and this has led to high performance devices. Advances in the micro-machining techniques have provided a substantial enhancement to the technology, leading to substantial cost reduction. As the technology has advanced there has been a greater understanding of the cause and effects of the error mechanisms and the need for a close integration between the sensing element and the electronic control circuits.

The performance of MEMS devices has increased by many orders of magnitude over a decade or so, owing to successful development projects. Consequently, MEMS accelerometers and gyroscopes are capable of providing inertial-grade measurements of acceleration and angular motion for long-range navigation systems.

Techniques have been demonstrated that enable the sensing devices of an inertial measurement unit to be fabricated on a single chip for lower-grade requirements. Methods for characterisation and compensation of systematic errors are proven and have been applied successfully to a range of systems. However, there is still scope for further improvement!

There is optimism that further enhancements are possible as a greater understanding is reached of the effects that geometry, size, packaging and interference from electronic circuits have on the performance of these miniature devices. It would be advantageous to improve the turn-on to turn-on repeatability, and reduce the initial transient response particularly for rapid reaction applications. Other performance areas where there is scope for improvement include: reduction in noise, improved fabrication precision, improved electronic control and reduced sensitivity to packaging.

The development of MOEMS devices is likely to lead to a further enhancement in performance owing to the true solid-state nature of these devices.

References

- 1 BARBOUR, N.: 'Inertial navigation sensors', NATO RTO Lecture Series-232, 'Advances in Navigation Sensors and Integration Technology', Oct. 2003

- 2 BARBOUR, N., ANDERSON, R., CONNELLY, J., *et al.*: 'Inertial MEMS system applications', NATO RTO Lecture Series-232, 'Advances in Navigation Sensors and Integration Technology', Oct. 2003
- 3 GAIßER, A., FRECH, J., SCHUMACHER, A., *et al.*: 'Evaluation of DAVED – microgyros realised with a new 50 µm SOI-based technology'. Presented at DGON symposium, *Gyro Technology*, Stuttgart, 2003
- 4 BAKER, G.: 'Quartz MEMS GPS/INS technology developments'. Presented at DGON symposium, *Gyro Technology*, Stuttgart, 2003
- 5 LÉGER, P.: 'Quapason – a new low cost vibratory gyroscope'. Presented at DGON symposium, *Gyro Technology*, Stuttgart, 1996
- 6 ELEY, R., FOX, C.H.J., and McWILLIAM, S.: 'The dynamics of a vibrating-ring multi-axis rate gyroscope', *Proc. Institution of Mechanical Engineers*, 2000, **214**, Part C
- 7 ROURKE, A.K., FOX, C.H.J., and McWILLIAM, S.: 'Development and testing of novel, multi-channel vibrating structure rate sensor'. Presented at DGON symposium, *Gyro Technology*, Stuttgart, 2003
- 8 FOX, C.H.J., ROURKE, A.K., ELEY, R., FELL, C., and McWILLIAM, S., 'Multi-channel and multi-axis inertial sensor concepts based on vibrating structures'. Presented at SPIE conference, *Smart Structures and Materials*, San Diego, CA, March 2003
- 9 ROYLE, C.M., and FOX, C.H.J.: 'The mechanics of an oscillatory rate gyroscope with piezoelectric film actuation and sensing', *Proc. Institution of Mechanical Engineers*, 2001, **215**, Part C
- 10 FOUNTAIN, J.R.: 'Characteristics and overview of a silicon vibrating structure gyroscope', NATO RTO Lecture Series-232, 'Advances in Navigation Sensors and Integration Technology', Oct. 2003
- 11 HOULIHAN, R.: 'Modelling of an accelerometer based on a levitated proof mass', *Journal of Micromechanics and Microengineering*, 1992, **13** (4), pp. 495–503
- 12 FOUNTAIN, J.R.: 'Silicon IMU for missile and munitions applications', NATO RTO Lecture Series-232, 'Advances in Navigation Sensors and Integration Technology', Oct. 2003
- 13 HULSING, R.: 'MEMS Inertial rate and acceleration sensor,' ION National Technical Meeting, CA, January, 1998

Chapter 8

Testing, calibration and compensation

8.1 Introduction

The evaluation testing of inertial sensors is required to establish their suitability for a given application, that is, to ensure that they satisfy all the performance requirements of that application.

Inertial sensors and systems are designed and manufactured for a very wide range of applications which include the provision of extremely precise navigation, in ships and submarines, for example, to measurements for flight control of short time of flight missiles. The performance required for such diverse applications spans eight or nine orders of magnitude. Similarly, the environment in which the sensors and system are required to operate varies widely, from the potentially more benign maritime applications to the very high dynamic forces experienced by highly agile surface-to-air and air-to-air missiles travelling at supersonic or hypersonic speeds.

The testing and calibration methods need to reflect the type of application and also, but very importantly, the environment in which the sensors and systems are to operate. It is crucial to establish that the sensors not only survive and operate reliably whilst being subjected to the vibrations, shocks and accelerations induced by the host vehicle, but also have sufficient endurance and resistance to survive the testing and calibration procedures.

It is possible to represent the behaviour or performance of an inertial sensor by use of a mathematical expression, as described in Chapters 4–6. One purpose of testing an inertial sensor is to evaluate the coefficients of these equations, the various ‘error’ terms, so that the performance of a sensor can be predicted for particular circumstances. Having established the performance figures or ‘characterised’ the sensor, any systematic errors may be compensated thus enhancing its accuracy. Other purposes of testing are to enable the output signals to be calibrated and to understand the behaviour of the device in various situations and environments.

Although often neglected, the testing and calibration of sensors and systems has great importance with very significant consequences, in terms of cost and performance

of the host vehicle, if the testing is either inadequate or too demanding. As a consequence of the various requirements and demands of a test programme, there are many different approaches that can be followed.

There are typically three distinct categories of testing undertaken on sensors: qualification, acceptance and reliability tests. The qualification tests tend to be the most extensive and stringent tests to which a sensor is subjected. These tests usually precede the production of a sensor by a manufacturer and are intended to show that a particular design will meet the requirement of a customer with adequate margins for production tolerances. These tests are likely to include all the investigations discussed in the sections that follow. Acceptance tests are undertaken on sensors during production in order to check selected parameters and to establish data for the calibration of the sensors. Tests may be undertaken on each sensor produced or on a sample batch selected from a production run, the number tested being determined by some statistically based rules. Reliability testing usually involves a sample of sensors selected at random from a production batch and run under normal operating conditions in order to establish the mean time between failure.

8.2 Testing philosophy

Depending on the sensor or system to be investigated, and the form of evaluation required, use may be made of either static or dynamic test methods. In the case of a static test, the device is kept fixed and the response to some natural effect or phenomenon observed. For example, the specific force due to the Earth's gravity could be observed with an accelerometer in various orientations. When dynamic testing is undertaken, the sensor under test is moved and the response of the device to that disturbance is monitored and compared with the stimulus.

A three-stage process may be followed to characterise the performance of a sensor or system:

1. coarse checking or evaluation using very simple tests, such as a single stationary position test on a bench, to establish that the response is compatible with the designer's or manufacturer's predictions;
2. static testing and/or calibration to derive performance parameters of the device from multi-position tests as defined in Sections 8.5 and 8.6;
3. dynamic testing where the device under test is subjected to motion such as an angular rotation or linear movement with acceleration. This form of testing requires specialised test equipment such as a rate table or vibrating table.

Inertial sensors and systems are subjected to different testing and calibration schemes throughout their development. At the prototype or initial research and development phase, the testing strategy will be designed to estimate the boundaries of performance of the device, establishing what it is 'good for', without breaking it as prototype devices are usually very expensive and in short supply! Depending on the type of sensor, the testing may well be on a 'bench' rather than in a sophisticated test laboratory. Tests will normally be arranged so that only one of the vast

range of environmental stimuli are changed during any one series of tests, to enable the response of the sensor to be understood and characterised. Throughout this and subsequent stages of testing, it is crucial that an accurate log is kept to record details of all tests and the results obtained.

Many projects have a so-called integrated test plan. The idea is to have a structured plan to determine the performance and reliability of a system using the minimum amount of testing, as this is often time consuming and expensive. Consequently, in the structured test plan, the data gathered from the qualification and acceptance evaluation tests can form an input to the integrated test plan, particularly if data are recorded on run times and any variation in performance with time. All these data can help estimate the potential reliability and mean time between failures for this system.

As an inertial sensor or system progresses through its research and development phase, the testing becomes more intensive in order to investigate progressively more of its performance envelope. Generally, the tests are undertaken in a sophisticated test laboratory, specialised test equipment being necessary for the more accurate evaluation of sensors. Some units will be tested to destruction, but this will usually be in a very controlled and deliberate manner. It is important to evaluate as many sensors as possible in order to establish the confidence limits of the parameters measured in the tests.

When the inertial sensors are being produced in significant quantities, the manufacturer's testing is usually to establish that the sensors are conforming to the production specification, and also to enable the sensor to be calibrated. In the case of large production quantities where there may be a possible significant range to the performance of the individual sensors, the testing can be used to grade the sensor performance, and hence direct the sensors to the appropriate application. This technique has been applied with success by many manufacturers.

The philosophy of acceptance testing is usually to establish that a sensor, or class of sensors, but more generally, an inertial system is compatible with the host vehicle. These tests will establish that the device, in whatever form, will operate satisfactorily within the vehicle and not jeopardise its integrity or safety. It will also be established that the device will fit in its specified location and that it will achieve the required accuracy. Generally, the device will undergo very specialised testing to establish the desired compliance with the objectives of the project application.

Various standards exist [1–8] to ensure the use of common terminology for users and manufacturers. Consideration is also given to recommended procedures for testing.

8.3 Test equipment

Over the years, during the development of inertial sensors and systems, many different testing methods and procedures have been perfected [2–7]. Many, if not most, require very specialised, accurate and costly equipment, housed in laboratories that are often built specifically for this special requirement. Special foundations are normally required so that the test equipment can be isolated from shocks, vibrations and other

perturbations induced by the local environment. The temperature of the environment, as a rule, is also carefully controlled, although environmental cabinets are generally used for thermal cycle testing. The application of the digital computer to the field of testing, both for the control of equipment and the analysis of data collected during the testing, has advanced the testing procedures significantly during the last three decades.

It is important not only to match the testing schedule and procedures to the requirements of the application, but also to ensure that the test equipment is of sufficient accuracy and precision to be compatible with the desired test accuracies. This enables a given and known stimulus to be applied to the inertial sensors or systems and their response observed. Generally, the test equipment should be capable of measuring a given quantity to an accuracy significantly in excess of that required by the sensor – a factor of between 5 and 10 is typical. Similarly, the data collection system and the algorithms manipulating those data must be compatible with the anticipated accuracy of either the sensor or its application. Additionally, the test equipment should be calibrated regularly where necessary, and investigated to ensure that it is providing the desired stimulus or disturbance to the device under test.

An example is a vibrator which should provide oscillatory motion along an axis which should be, for example, normal to the horizontal plane. As the equipment degrades with use, it is possible for the axis of motion to move about a cone. Thus any gyroscope under test may detect this angular motion, depending on its accuracy and alignment, and therefore produce an output signal reflecting its detection of this motion. If the investigator was not aware of the defect in the test equipment, then the gyroscope would be reported to have a particular vibration sensitivity, which of course had been induced by a defect in the test equipment. Similarly, this defect may be masking or compensating a real sensitivity in that gyroscope.

8.4 Data-logging equipment

The electrical signals produced by inertial sensors or inertial systems can be in various forms, such as direct current or alternating current, and may be in continuous or pulsed form. The form of the signals is dependent on the type of sensor, their pick-offs and the nature of any re-balance loops used. Depending on the signal being monitored and the accuracy being required, a chart recorder or some form of digital meter may be used. Chart recorders are often used to show trends in sensor performance.

The other technique, most common for accurate work, is to use either a digital voltmeter or a digital ammeter to monitor the signals. The signals are often integrated over a period of time, its length depending on the testing, and then logged in a suitable form, usually through a data bus to a computer. For many years, small computers have played a most important role in the control and conduct of tests, as well as the collection of data, its subsequent manipulation and ultimate presentation.

The ‘rules’ concerning the accuracy of the data monitoring and recording equipment are similar to those for the test equipment as outlined in the previous section. It is vital that the resolution of the whole of the chain of data recording equipment

exceeds that given by the sensors being tested, preferably by about an order of magnitude or more. Similarly, the dynamic range and stability of the monitoring and recording equipment, including any data buses, should exceed those of the instruments to be tested. Care has to be taken to ensure that various transient effects are not masked by inadequate data-logging equipment or during the subsequent data manipulation.

8.5 Gyroscope testing

The gyroscopes to be tested are usually mounted in a test fixture, often a cube with very accurately machined faces to achieve very precise mounting in the test equipment. This enables the sensor to be transferred between the various test equipments used in a test programme and also allows various designs of sensors to be tested on the same equipment. For the more accurate sensors, that is, those with a bias in the region of $0.01^\circ/\text{h}$, mounting accuracies of the order of 10 arc s are required.

Prior to undertaking a series of tests to evaluate the performance of a gyroscope, it is usual to undertake some preliminary investigations. Such checks include measurement of electrical resistance and insulation strength, polarity, time for the rotor to reach its operating speed, time to stop rotating and power consumption.

The IEEE has published a number of documents [2–6] defining procedures for testing various types of gyroscope. The following text concentrates on general test techniques for strapdown gyroscopes, examples of test equipment and methods of analysis enabling the performance of an instrument to be evaluated.

8.5.1 Stability tests – multi-position tests

The purpose of stability testing is normally to evaluate the run-to-run, or switch-on to switch-on, drift and in-run drift of a gyroscope. The gyroscope is placed on a block in a series of fixed orientation with respect to geographic axes and the local gravity vector of the Earth. Alternatively, the gyroscope may be positioned accurately in the required orientation using a Graseby table, a stabilised gimbal system with controlled and instrumented gimbal angles, as shown Figure 8.1. For more accurate work, the table, or gyroscope under test, is mounted on a plinth of granite that has its own foundations separate from, and vibrationally isolated from, the laboratory.

The gyroscope is operated within a fixed temperature range and is positioned in one of a set of eight (possibly up to 12) standard orientations as indicated in Table 8.1. The gyroscope is switched on and after a pre-determined time, to allow thermal transients to settle, the signals from the gyroscope are recorded. This test is repeated a number of times, the number and duration being dependent on the required accuracy of the test, determined by the usual statistical rules for confidence levels [9–11], with a pre-determined time for cool-down between runs.

For sub-inertial quality sensors a set of runs in one test in this series may last for up to one hour. For higher quality sensors, the duration may be many hours or longer.

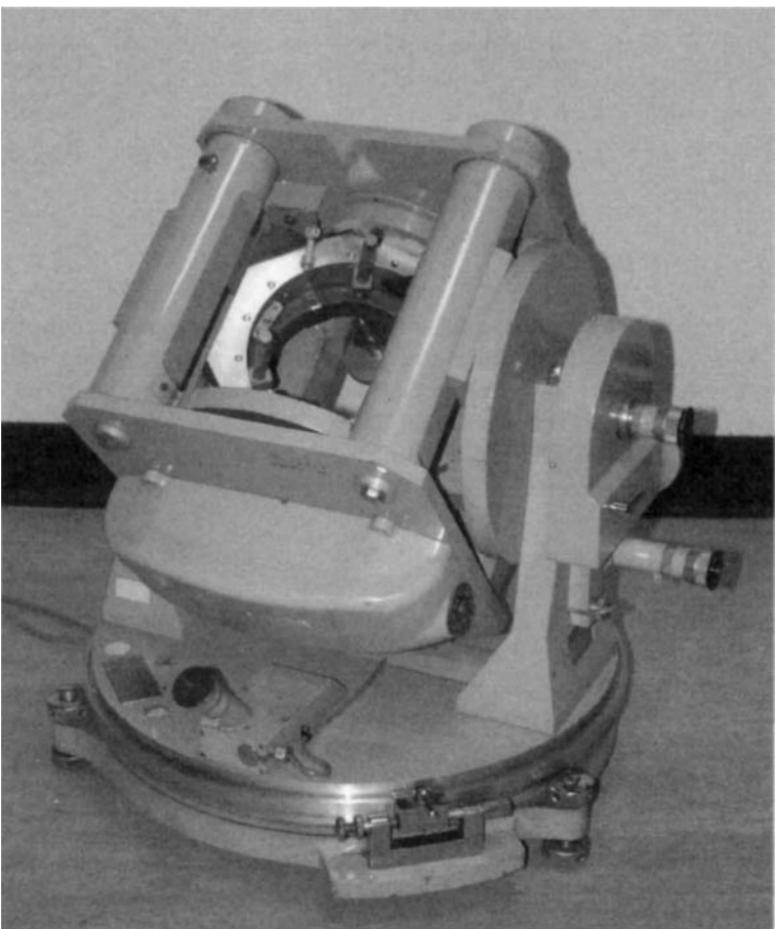


Figure 8.1 *Graseby table (Courtesy of QinetiQ)*

On completion of the series of runs with the gyroscope in a particular axial setting, a new orientation is selected and the test procedure repeated. From the different data sets, the effects of various systematic errors, and the effects of the rotation of the Earth can be removed. The data are analysed to find the mean drift rate and its variance, or scatter, about this value, for each test run.

The run-to-run stability of the gyroscope is evaluated from the scatter in the mean level of gyroscope output for each run recorded during these tests when the gravity vector was not coaxial with the input axis of the gyroscope.

The in-run stability of the gyroscope drift rate is deduced from the average scatter of the measured drift in the output of the gyroscope about its mean value during a run, calculated for each test in a series. This value can be averaged over the whole series of tests, but care must be taken to avoid the inclusion of any data that have any

Table 8.1 Gyroscope axial settings for stability tests

Position	Direction of axis			Acceleration of axis			Component of Earth's rotation along axis	
	Spin	Input 1	Input 2	Spin	Input 1	Input 2	Input 1	Input 2
1	Up	N	W	+1	0	0	$\Omega \cos L$	0
2	Up	W	S	+1	0	0	0	$-\Omega \cos L$
3	Up	S	E	+1	0	0	$-\Omega \cos L$	0
4	Up	E	N	+1	0	0	0	$\Omega \cos L$
5	N	E	Down	0	0	-1	0	$-\Omega \sin L$
6	N	Up	E	0	+1	0	$\Omega \sin L$	0
7	N	W	Up	0	0	+1	0	$\Omega \sin L$
8	N	Down	W	0	-1	0	$-\Omega \sin L$	0

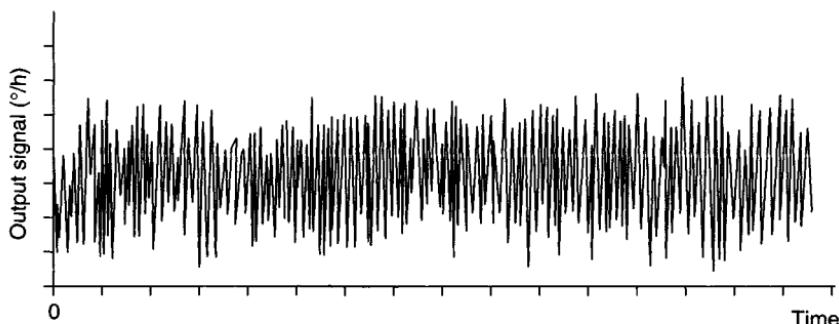


Figure 8.2 Raw data collected during stability testing

anomalies without establishing the reasons for the anomalies. Figure 8.2 shows the form of the raw data collected over a period in a single run.

Table 8.1 illustrates typical axial settings for a dual-axis gyroscope used during stability tests, it also shows the components of gravity along each vector axis and the rotation of the Earth about each axis of the gyroscope. In Table 8.1 N, S, E, W are horizontal directions north, south, east and west, respectively, Ω is the Earth's rate, $15.041^\circ/\text{h}$, L is the local angle of latitude, $\Omega \cos L$ is the horizontal component of Earth's rate and $\Omega \sin L$ is the vertical component of Earth's rate.

The Earth's gravitational force is taken to be an apparent acceleration acting vertically upwards, and this direction is positive in the convention used here. Data from these tests are used to solve a set of simultaneous equations expressing the gyroscopic drift associated with each external stimulus in each position.

For example, from this series of tests, it is possible to establish the acceleration dependence of the gyroscopic drift (g -dependent drift) by comparing the mean levels

of signals for the cases when the sensor is mounted with its input axis orthogonal to and coaxial with the gravity vector. This is illustrated in the following analysis.

Under static conditions, assuming allowance has been made for the Earth's rate component, the output of the gyroscope (ω_0) may be expressed as:

$$\omega_0 = B_f + B_{gx}a_x + B_{gy}a_y + B_{gz}a_z \quad (8.1)$$

where B_f is the g -independent bias and B_{gx} , B_{gy} , B_{gz} are the g -dependent biases induced by accelerations a_x , a_y and a_z acting along the x , y and z axes of the sensor, respectively. For the conventional rate integrating gyroscope, these correspond to the spin, input and output axes of the gyroscope.

If measurements are taken with the gyroscope positioned with its x -axis coincident with the gravity vector pointing up and down, then the measurements obtained, m_1 and m_2 , may be expressed as follows:

$$\begin{aligned} \text{For input axis up: } m_1 &= B_f + B_{gx}g \\ \text{For input axis down: } m_2 &= B_f - B_{gx}g \end{aligned} \quad (8.2)$$

The coefficients B_f and B_{gx} may be calculated from the sum and difference of these two measurements. Similarly, the g -dependent bias coefficients may be determined by taking sets of measurements with the y and z axes of the gyroscope aligned with the gravity vector.

8.5.2 Rate transfer tests

The purpose of these tests is to investigate the various characteristics of the scale-factor of the gyroscope which relates the output signal to the input motion, and the maximum and minimum angular rotation rates that the gyroscope can measure or 'capture'. A schematic representation of this characteristic is shown in Figure 8.3. This diagram also shows how the scale-factor, resolution, dead-band and threshold are defined.

The usual characteristics of the scale-factor that are evaluated in these tests are:

- its mean value and scatter about this value;
- the change in mean value as the angular rotation rate of the rate table changes, that is, its linearity;
- the variation of the mean value of the scale-factor as the ambient temperature changes;
- any hysteresis in the response of gyroscope.

In the case of 'two-axis' gyroscopes, such as the dynamically tuned gyroscope, this form of testing can also assess the orthogonality of the two input or sensitive axes. These tests are undertaken using a precision turntable called a rate table, as shown in Figure 8.4.

The equipment has a flat circular plate on which the gyroscope can be mounted. The plate can be rotated about an axis normal to the surface at very precise angular velocities, which are monitored and controlled very accurately. The angular velocity

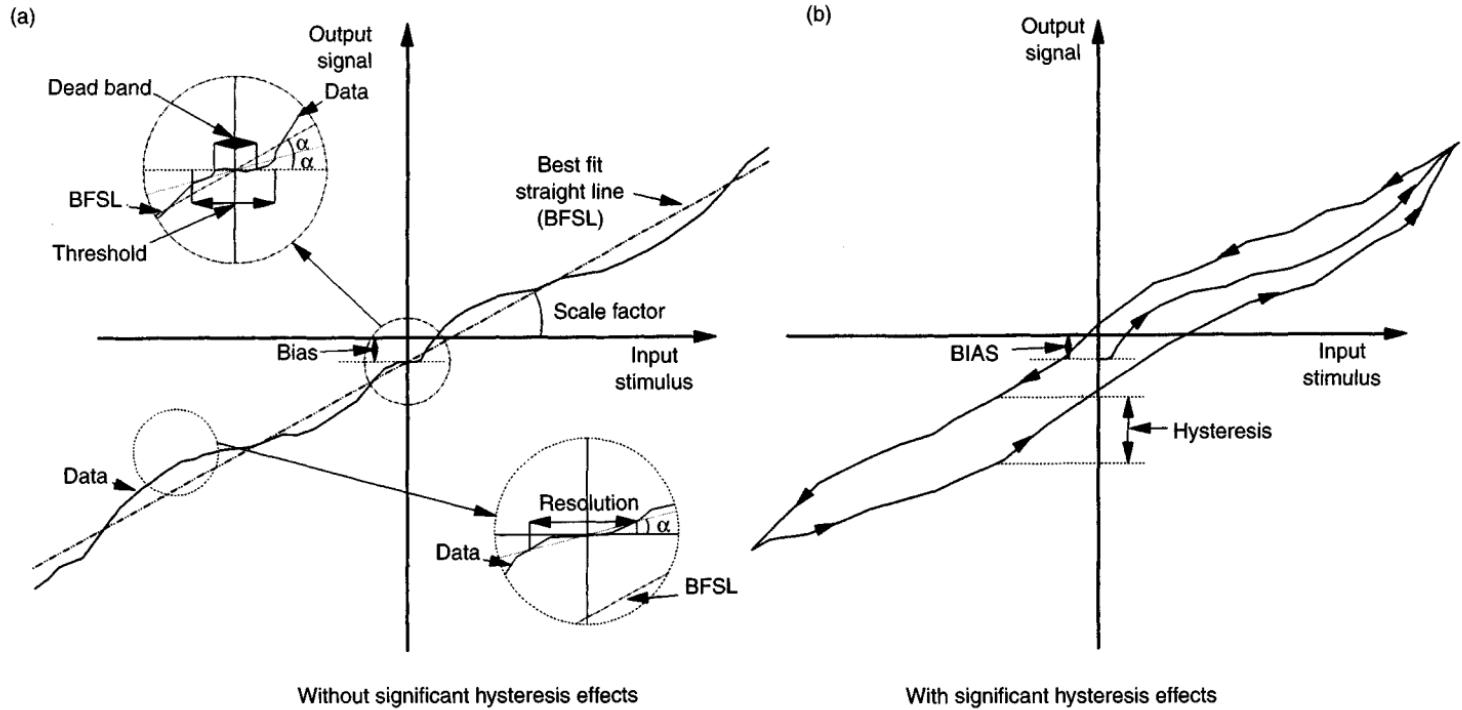


Figure 8.3 Scale-factor characteristic of a gyroscope



Figure 8.4 Rate table with a controlled environmental chamber (Courtesy of QinetiQ)

can often be varied from a fraction of one degree per hour to many hundreds of degrees per second. Manufacturers currently offer rate tables with maximum angular rates of about $3000^{\circ}/\text{s}$. The equipment is usually mounted with its axis of rotation vertical on a granite plinth to provide isolation and stability, and those operating at very high rotation rates are often caged for safety reasons. An environmental chamber may enclose the test table for thermal evaluations, see Section 8.5.3.

During rate transfer tests, the gyroscope is mounted securely on the turntable with its sensitive axis (or one of its sensitive input axes in the case of a twin axis sensor) parallel to, although not necessarily coaxial with, the axis of rotation of the rate table.

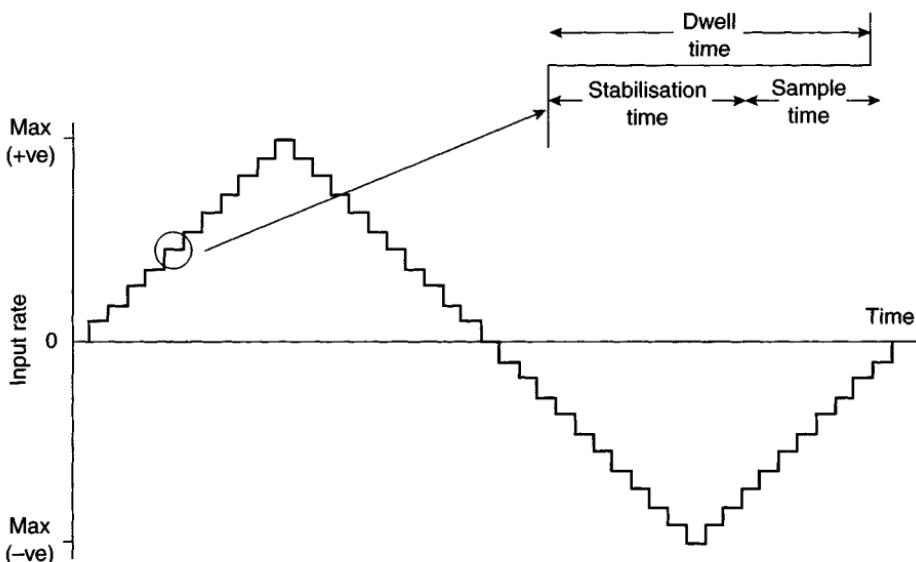


Figure 8.5 Rate table step sequence

The rate table can be used in many ways, but the basic principle is to compare the angular rate or displacement measured by the gyroscope with that given by the rate table. Care is necessary to ensure that the rate table does not 'overshoot' its intended angular rotation rate for any given measurement. Examples of the way in which the rate table may be used are:

1. During a typical test schedule, the rotation rate of the rate table is stepped through a series of angular rates starting at zero, recording data at each stage. The rotation speed is kept constant for a set period at each step and the sensor outputs allowed to stabilise, before recording the output signals. The applied angular rate is varied in incremental steps between the maximum and minimum desired rotation rates as shown in Figure 8.5. At each step, the signals from the gyroscope are recorded when the sensor is in equilibrium.

By cyclical variation of the applied rate over the expected measurement range of the sensor, this form of test allows hysteresis effects in the scale-factor to be observed, sometimes called the bow-tie or butterfly effect owing to the shape of the plot of the error in the indicated rate from the gyroscope against the actual turntable rotation rate, as shown in the example plot in Figure 8.6.

2. This is essentially a variation of method 1 in which the rotation rate is stepped rapidly from rest to the maximum rotation rate in one direction and then slowed down in steps until at rest and stepped to the maximum rotation rate in the opposite direction and then brought to rest in steps. In this case, the 'dwell' at each step is a few hundred milliseconds to allow the turntable to stabilise and allow the output signals from the gyroscope to be recorded. Collecting data at the correct moment is crucial for this form of test, it being co-ordinated by the

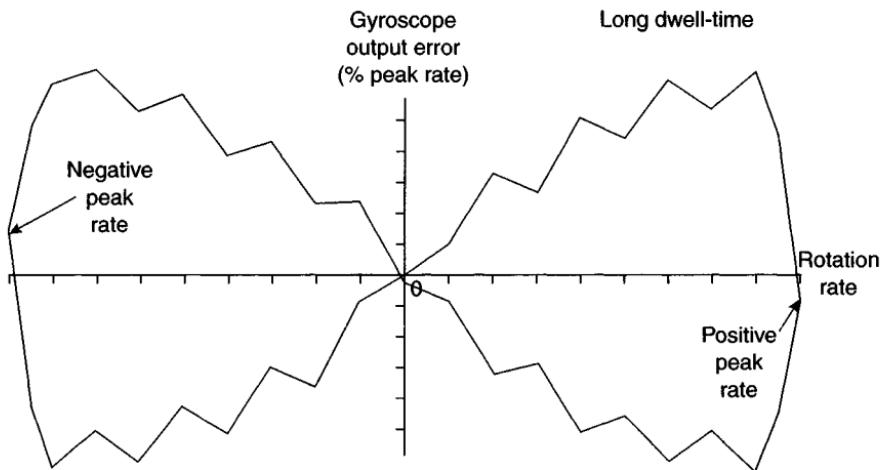


Figure 8.6 Sample data from a rate transfer test

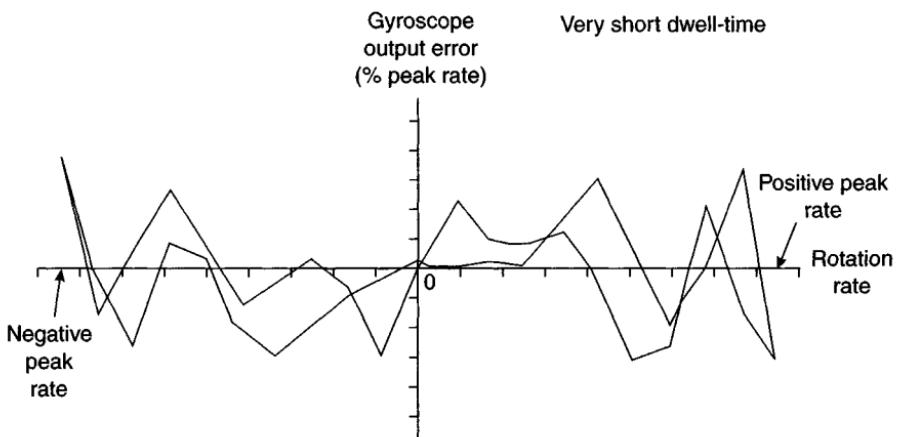


Figure 8.7 Sample data from a rapid rate transfer test

controlling computer. This form of rate transfer test is often used when assessing gyroscopes for tactical missile applications as it can be completed in times that are comparable with the time of flight of the missile. This so-called rapid rate transfer test can be completed, in some circumstances, in as little as 10 s. Typical data from such a test are shown in Figure 8.7.

3. The rate table is set to a constant angular rate and data collected for a given number of complete rotations which are monitored by optical means to give high precision read-out and control. This method is used for precision gyroscopes, such as the ring laser gyroscope, which are often as precise as the normal rate table control systems.

Orthogonality of the axes of a 'two-axis' sensor may be investigated by recording data during the above tests from the second input axis which is nominally at right angles to the axis of apparent rotation. When the data from this axis have been corrected for systematic errors, the residual rate is a result of the input axes not being orthogonal.

Data from the rate transfer tests are normally analysed by comparing the output signal from the gyroscope with the corresponding turn rate of the table, usually measured by a tacho-generator. This process is repeated for all the data collected in a test sequence and a straight line is constructed through the data using a least-squares fit procedure. The gradient of this line is the scale-factor of the gyroscope. In order to take account of any non-linear trends, a curve may be fitted to the data. This is represented mathematically by a polynomial expression, the coefficients of which define the scale-factor non-linearity of the gyroscope. Typical plots of the 'processed' input to output characteristic are shown in Figures 8.6 and 8.7. These curves show deviations from the theoretical linear input/output plot. Although somewhat idealised, these plots illustrate the influence that time has on the recorded output signals from a gyroscope. With a long dwell time, there is a sufficiently long period for various phenomena to reach equilibrium, particularly thermal effects. Further, it is usually possible to observe any hysteresis in the torquer scale-factor with the longer dwell time, whereas this is less evident for the very short measurement period.

Gyroscopes may be calibrated using the data provided by the rate transfer tests. The output signal(s) from the sensor are compared with the accurately known applied rotation rate and the scale-factor defined, for instance, as so many millivolts per degree per second of rotation rate.

8.5.3 Thermal tests

Variations in the performance of a gyroscope with changes in the temperature within the case of the gyroscope can be observed by enclosing the turntable in a climatic chamber, as shown in Figure 8.4. The temperature within the chamber can be varied from 'sub-zero' temperatures, typically down to about -55°C , to temperatures in the region of 75 or 80°C . Cooling is normally achieved by using carbon dioxide, so care must be exercised in venting the used gas. A schematic diagram of a typical experimental arrangement is shown in Figure 8.8.

There are various thermal tests that can be undertaken such as allowing the temperature of the gyroscope to stabilise, that is, a 'soak' test, or allowing a controlled increase or decrease over a given period, that is, a thermal ramp test. The rate tests described above are repeated at the various temperatures and the output signal from the gyroscope recorded. Using this technique, the scale-factor can be evaluated at various temperatures throughout the operating range of the sensor. A typical plot is shown in Figure 8.9. Any correlation with temperature variation can be defined by a mathematical expression which may be stored in a computer and used to provide on-line compensation for temperature variation, provided a thermal sensor is supplied with the gyroscope.

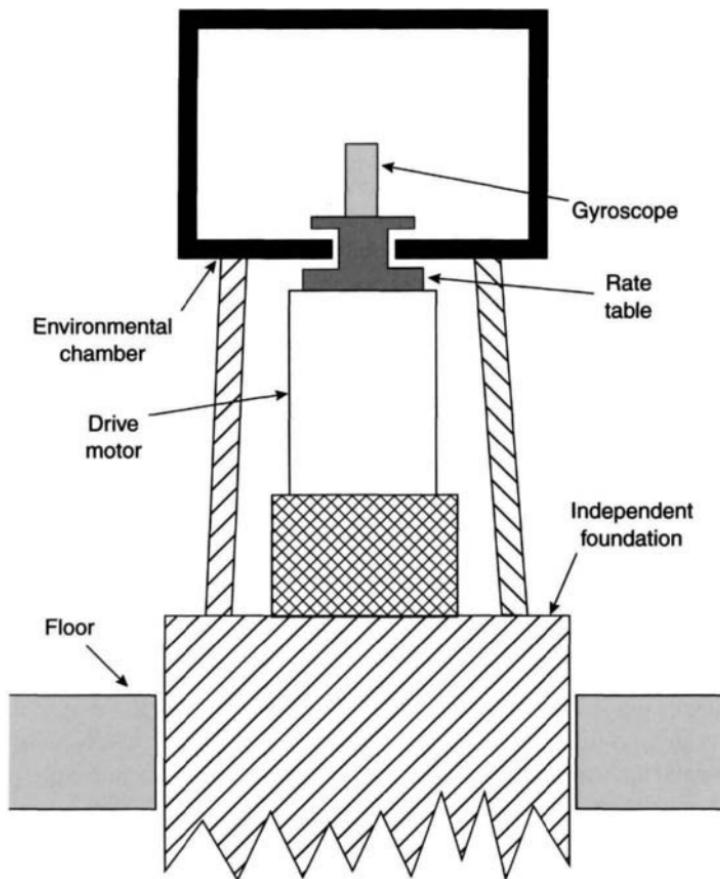


Figure 8.8 Illustration of thermal test equipment

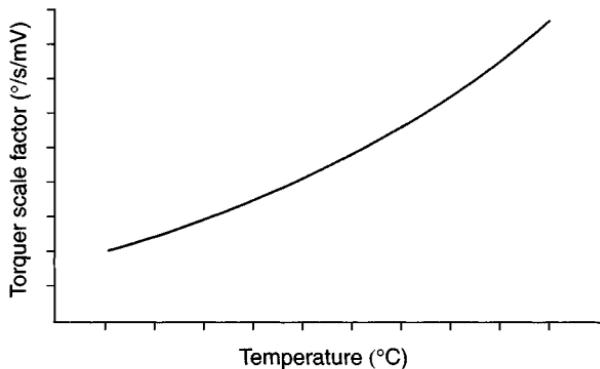


Figure 8.9 Typical variation of torquer scale-factor with temperature

Variations in other performance parameters of the gyroscope with temperature can be evaluated similarly by enclosing the gyroscope in a climatic chamber for a specific test, such as a multi-position test, and recording the response of the gyroscope. Care has to be taken that the imposition of thermal variation does not affect the test equipment itself.

8.5.4 Oscillating rate table tests

The purpose of these tests is to determine the frequency response characteristics of a gyroscope and its associated electronic control circuits to oscillatory rotation applied to the input axes of the sensor. Normally, both the bandwidth and the natural frequency of response of the sensor are evaluated in this test. The test equipment is very similar to the rate table already described for the rate transfer tests. In this case, the turntable, again mounted on a suitable plinth to provide stability, applies an oscillatory angular motion at various pre-set frequencies and so requires low inertia about the rotation axis. A photograph of an oscillating rate table is shown in Figure 8.10.

The gyroscope to be evaluated is mounted and fixed to the turntable with its sensitive axis parallel to the axis of rotation of the turntable. A given maximum rotation rate is selected along with a maximum frequency of oscillation. The frequency of oscillation is increased in pre-determined steps up to the maximum value. The response of the gyroscope is recorded at each frequency step in the series, up to the maximum value, and then, as the frequency is reduced to the starting condition. The tests are generally repeated for various input rate maxima.

Results from these tests are usually plotted as gain and phase graphs by comparing the amplitude and the phase of the signal generated by the gyroscope with the actual amplitude and phase of the disturbance applied by the turntable. A typical response is shown in Figure 8.11 from which the bandwidth and damping factor can be deduced.

Such information is needed to determine the speed of response of the sensor. This is particularly crucial in applications where the sensor is to be used to provide feedback control, as in a missile autopilot for instance.

8.5.5 Magnetic sensitivity tests

The purpose of these tests is to examine and quantify any influence that external magnetic fields may have on the 'drift' characteristics of a sensor. The form of the testing is identical to the multi-position testing described earlier, with the addition of a pair of Helmholtz coils, as shown in Figure 8.12, which may be positioned to apply a magnetic field along each of the principle axes of the sensor under test.

Initially, data from the gyroscope are recorded for a given period, with the sensor aligned to a particular geographic orientation and with the coils unpowered. This test is then repeated with the appropriate current passing through the coils to generate the desired magnetic field strength. Generally, a whole series of measurements are made for different orientations of the magnetic field with respect to the axes of the gyroscope in which the magnetic field strength is incremented in suitable steps up



Figure 8.10 Oscillating rate table (Courtesy of QinetiQ)

to a maximum value. It is usual to keep the gyroscope running during these tests to eliminate any switch-on to switch-on effects, and similarly the gyroscope is allowed to stabilise before the first measurements in the series are recorded.

The recorded data are analysed to establish the mean value of the drift rate for each series of tests. Comparison is made between the mean value in the presence and absence of a magnetic field for each orientation of the magnetic field to establish the magnitude, if any, of the dependence of the drift rate on strength and orientation of an applied magnetic field. Typical sensitivity of a conventional gyroscope is illustrated in Figure 8.13.

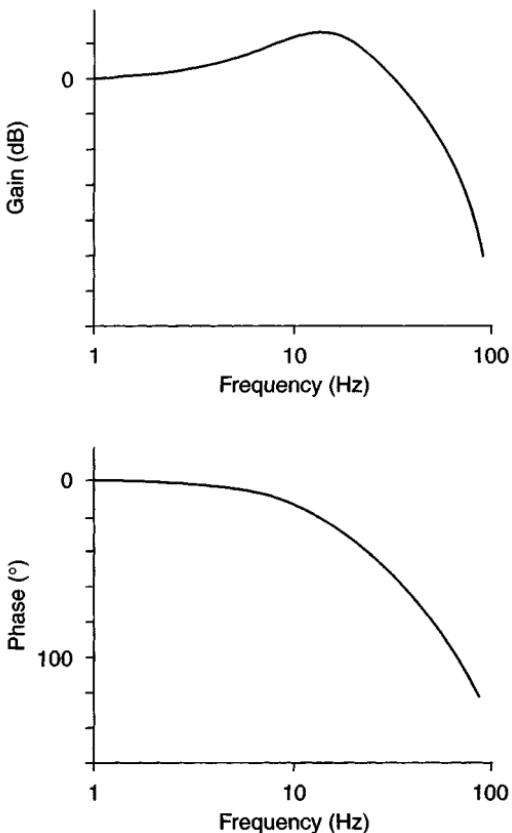


Figure 8.11 Typical gyroscope frequency response

8.5.6 Centrifuge tests

A centrifuge provides a means of applying large steady or fluctuating accelerations to a gyroscope. A photograph of a centrifuge is shown in Figure 8.14. The purpose of the centrifuge testing is to investigate the response of the gyroscope to large accelerations and to establish its ability to withstand large continuous or fluctuating accelerations whilst the sensor is either operational or in a quiescent condition.

These tests are normally one of the last in a series of tests applied to an inertial sensor, as great care has to be taken to prevent permanent damage to the sensor. This is particularly true when the extremes of the gyroscope's performance envelope are being investigated. Care must be taken to align the sensitive axis, or axes, so that the maximum rotation rate of the gyroscope is not exceeded and allowance has to be made for the applied input rate which is sensed by the gyroscope. Alternatively, the sensor can be mounted on a counter-rotating table to null the effects of the rotation of the centrifuge. In this case, the applied acceleration is sinusoidal at the table rotation frequency. The magnitude of the acceleration applied along each axis of the

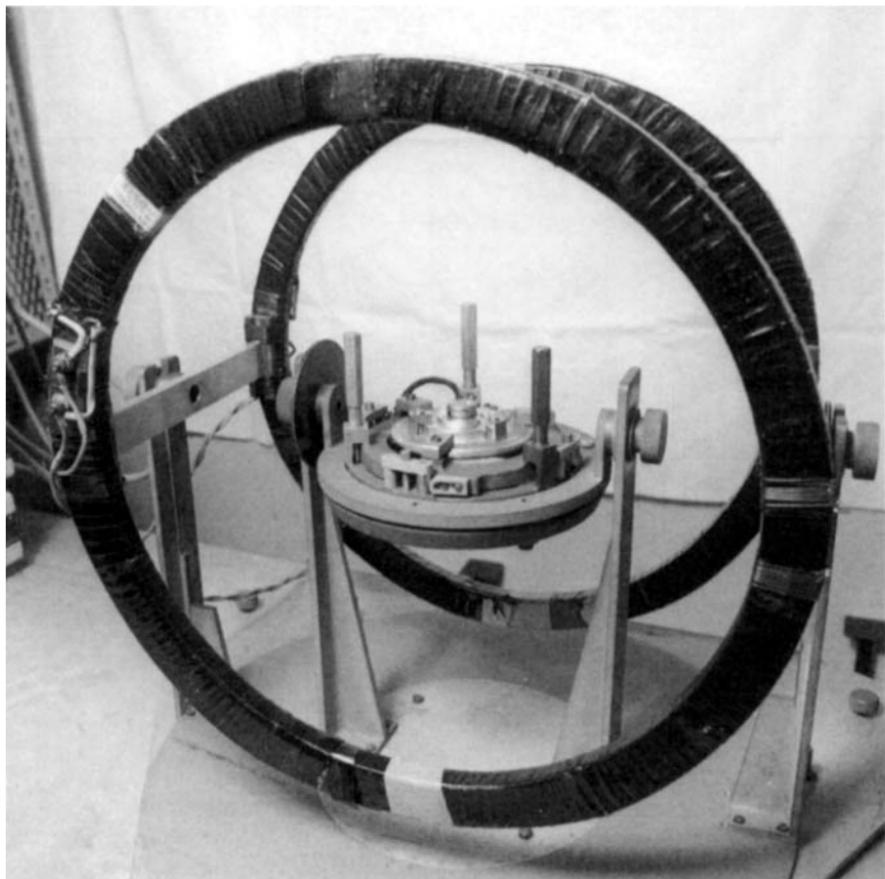


Figure 8.12 Helmholtz coils (Courtesy of QinetiQ)

gyroscope can be either controlled by the rotation speed or the distance of the sensor from the centre of rotation, the radius arm. The addition of an accelerometer adjacent to the gyroscope under test provides an accurate measure of the acceleration acting on that sensor.

To investigate the acceleration sensitivity of a gyroscope, the applied acceleration is increased in steps until a given maximum level is reached. This process is repeated with the gyroscope mounted in different orientations in order to investigate its sensitivity to acceleration along different axes. Comparison is then made between the mean values of the output signals produced by the gyroscope for each acceleration applied to it, and hence the acceleration sensitivity of the sensor can be calculated. This value can be compared with the value deduced from the multi-position tests described earlier, where the acceleration sensitivity is evaluated at a low value of applied acceleration.

In the case of quiescent testing, a particular parameter of the gyroscope is evaluated, or a range of parameters, are evaluated prior to the test. The sensor is then

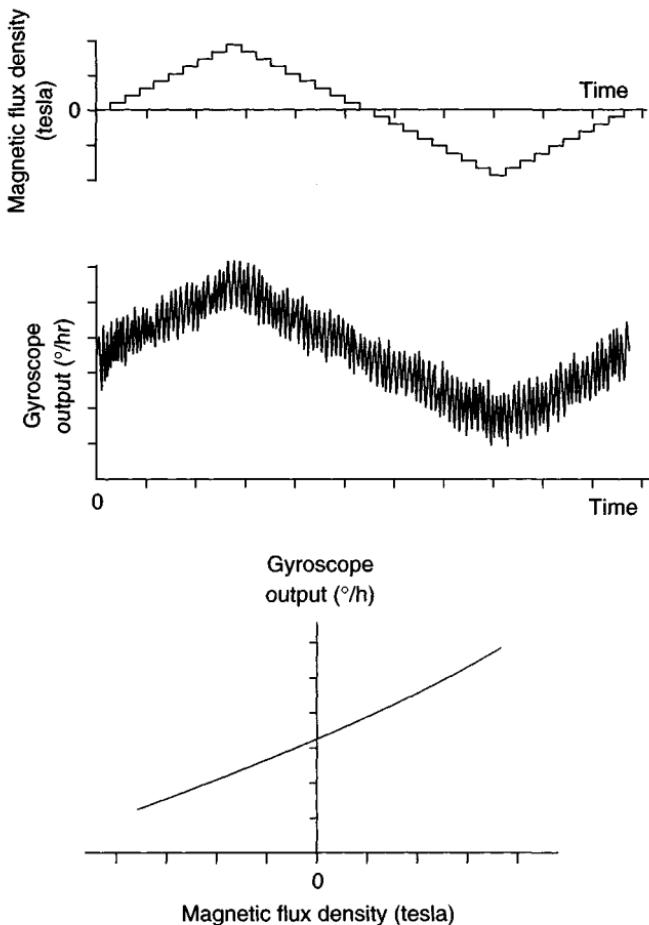


Figure 8.13 Illustration of magnetic sensitivity

subjected to the desired steady state acceleration, or acceleration profile, as described earlier, but with the sensor in an unpowered state. The gyroscope is then re-evaluated to establish if any change in performance has taken place.

8.5.7 Shock tests

The purpose of this form of test is to measure the response of the gyroscope to an applied shock, and to establish the resilience of the sensor to such an applied acceleration over a very short duration, typically of the order of milliseconds. As in the case of the centrifuge tests, the gyroscope may be operational or quiescent during the test.

A shock may be applied to a gyroscope by either using a shock table or a vibrating table. When using a shock table, the sensor is mounted on a heavy metal table and

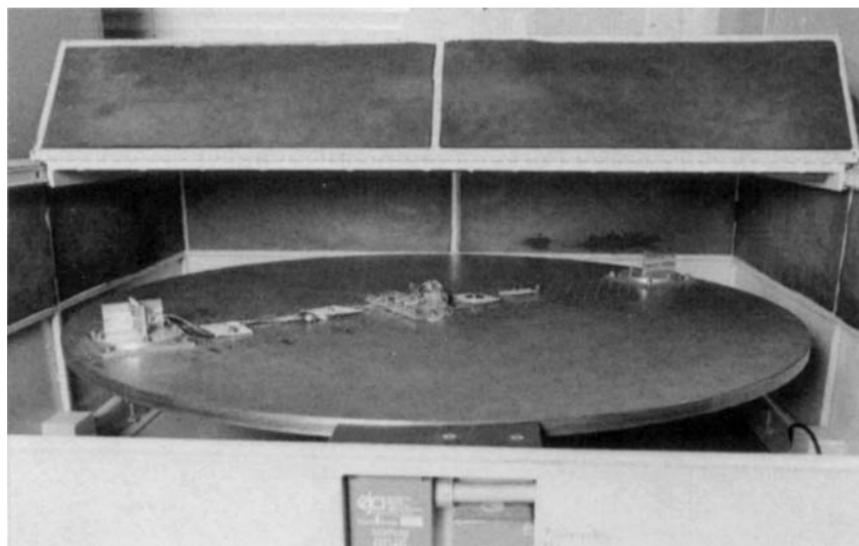


Figure 8.14 Photograph of a centrifuge used for sensor testing (Courtesy of QinetiQ)

this table is dropped over a given distance on to a suitably shaped piece of lead. In the latter case, a short duration single displacement is applied to the vibrating table on which the sensor is rigidly mounted [12].

In order to measure the response of a gyroscope in an operating condition to an applied shock, the sensor is suitably oriented on the table of the test apparatus and very firmly bolted to it. The output signals from the sensor are recorded for a given period before the shock is applied. If possible, the output signals are recorded during the application of the shock and then also for a period after the shock. Comparison with the mean value of the gyroscopic drift before and after the application of the shock will indicate any transient or permanent change in the characteristic of the gyroscope.

When a gyroscope is tested in a quiescent state, it is normal to follow the procedure already described above for the quiescent centrifuge test. Again a comparison is made of suitable characteristics of the tested sensor before and after the application of the shock acceleration.

8.5.8 Vibration tests

This is normally the last series of tests undertaken on a gyroscope owing to a potentially high risk of permanent damage to the sensor. Great care must be taken to ensure that resonances in the mounting fixture do not amplify the applied accelerations. It is strongly recommended that a thorough investigation of the modal behaviour of the structure used to mount the gyroscope on the vibrating table is undertaken before any investigations of random vibration are started.

The purposes of this form of testing are usually fourfold:

1. Investigation of the frequency at which any resonant responses of the sensor occur and their magnitude.
2. Evaluation of the anisoelasticity or acceleration squared (g^2) bias dependence of the gyroscope.
3. Examination of the resilience and survivability of the sensor in a particular vibratory environment. The sensor may be either quiescent or operational depending on the reason for the particular test.
4. Estimation of the change in noise characteristic of the output signals of a sensor experiencing a vibratory environment.

A vibration table provides a means of applying various forms of vibratory motion to a sensor. There are several different forms of such equipment. Usually, the vibrator has a table on which the sensor is mounted. The table is driven along a well-defined axis by an electromagnet fed with a current of the required frequency and wave shape. The table may oscillate vertically or it may be rotated through 90° and be attached to a slip table and vibrated horizontally. A typical vibration table is shown in Figure 8.15.

Two forms of motion may be applied to the test table, either sinusoidal vibrations or random frequency vibrations. In the case of the former type, the displacement of the table is varied sinusoidally over a given frequency band without exceeding a pre-defined acceleration level. In the latter case, random vibration is applied according to a given power spectral density and frequency bandwidth.

Throughout testing with a vibrator, it may be necessary to use de-gaussing coils as some vibratory test equipment generate very large magnetic fields. These fields can alter the drift performance of a gyroscope as indicated by the magnetic sensitivity testing of Section 8.5.5. Thus, care must be taken to ensure that genuine effects are being observed and analysed correctly, not merely artefacts introduced by the test equipment.

It is good practice to carry out an initial series of tests at a low acceleration level with the gyroscope operating in order to establish if any resonances arise within the sensor and the frequencies at which they occur. This enables the investigation of the sensor to be undertaken at frequencies well away from those that will excite any of these identified resonances. The sensor is suitably orientated and firmly clamped to the vibration table. During these tests, a small feedback accelerometer is mounted on the test sensor or its mounting fixture to provide a measure of the accelerations applied to the sensor. The vibrator is used in a feedback mode to ensure that a fairly even acceleration amplitude is applied to the sensor under test. A small peak acceleration in the region of $1g$ is usually chosen and a sinusoidal displacement is applied to the sensor by the vibrator. The vibration frequency is changed slowly from an initial value of a few hertz to an upper limit, normally in the region of 10 kHz. This is often called a 'sine sweep'. During this 'sweep', the output signals from the gyroscope are monitored continuously to ensure that any resonances encountered do not destroy the sensor, and the frequency at which they occur is recorded. This test can be repeated for other orientations of the gyroscope to establish resonant free frequency zones.

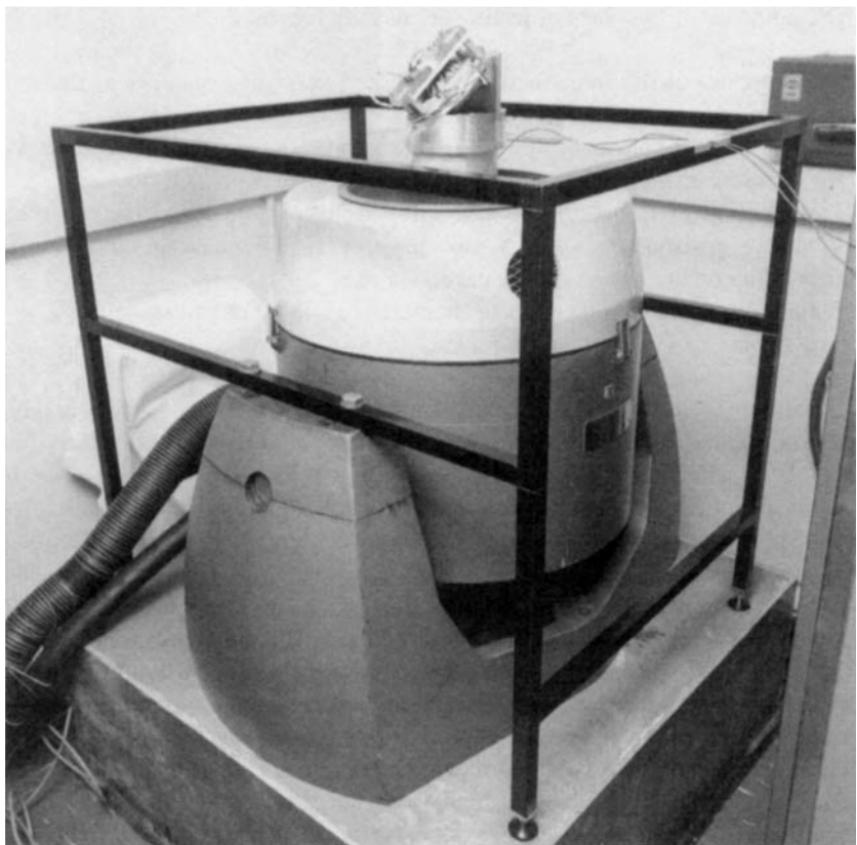


Figure 8.15 Photograph of a vibration table used for sensor testing (Courtesy of QinetiQ)

When testing mechanical gyroscopes, it is usual to mount the sensor with two of its axes at 45° to the axis of table movement in order to maximise the effect of any anisoelastic biases and to allow them to be identified. In this case, the third axis of the gyroscope is perpendicular to this axis of motion and aligned with one of the cardinal geographic axes. The gyroscope is tested to establish the mean and standard deviation of its drift in this selected orientation over a suitable period. Sinusoidal motion is applied to the sensor at a frequency of oscillation selected well away from any resonances detected during the preliminary 'sine sweep' tests. A pre-selected maximum acceleration is applied and the output signals of the gyroscope are collected over the period of oscillation, typically a duration of many minutes. Such tests are repeated with different peak accelerations at a number of spot frequencies. The statistics of the gyroscope data collected during these tests are evaluated and a comparison made of the mean bias in the absence of vibration to allow the acceleration squared dependency of the sensor to be evaluated. The anisoelastic bias coefficient may be

calculated by finding the averaged increase in the output bias caused by the vibratory motion at each frequency and dividing it by the appropriate square of the acceleration. The various values of this drift can then be averaged to give the estimate of the acceleration squared (g^2) drift rate.

Quadrature effects are deduced by repeating the above tests, but with the gyroscope orientated with the input axis perpendicular to the axis of motion. Noise introduced into the output signals of the gyroscope can be analysed by examination of the statistics calculated from the signals produced by the sensor during its vibratory tests and comparing them with those deduced when it was stationary, i.e. before and after testing.

To test the endurance, survivability and resilience of a sensor, it is usual to use random motion of the vibration table. The tests are undertaken as described above but the frequency and amplitude of the motion are continuously varied randomly across the spectrum defined by the frequency bandwidth of the power spectral density used to specify this form of vibration. This power spectral density defines the maximum acceleration that the sensor mounted on the vibrator will experience at any frequency. This is the so-called random vibration testing.

Depending on the aspect of application of the sensor that is being investigated, the sensor may or may not be operational. If, for example, its susceptibility to damage during transportation is being investigated, the sensor may be vibrated for weeks or months whilst in a quiescent state. In this case the sensor would be characterised before and after the tests, to establish any changes in performance, as described above, for quiescent centrifuge or shock testing. If a particular regime of flight of a tactical missile is being examined when, for example, the dynamic forces are high, the sensor will be operational, but the duration of application of the appropriate vibration spectrum may be short, possibly only 10 s or less. During this form of testing, the output signals from the gyroscope would be recorded before and after the vibratory tests to establish the significance of any observed changes in the response of the sensor during the vibratory tests.

Care must be exercised when evaluating mechanically dithered ring laser gyroscopes in order to prevent any unrepresentative interaction between the vibratory motion and the sensor leading to spurious bias errors. It is important that the sensor is attached to a rigid base plate and also that these sensors are tested individually.

8.5.9 Combination tests

Having analysed the response of a sensor to a single disturbance such as rotation or acceleration, it may be appropriate to undertake some limited evaluation of the sensor by combining, for example:

1. rotation and acceleration by mounting a rugged rate table on a large centrifuge so that the gyroscope can be rotated about its input axis whilst an acceleration is applied along that axis;
2. vibration and acceleration by mounting a small vibrator on a large centrifuge enabling an acceleration to be applied along an axis whilst vibrating the sensor about an orthogonal axis or axes;

3. rotation and vibration by mounting a rugged rate table on a vibrator and either applying a vibratory motion along one axis whilst simultaneously applying a rotational disturbance, or applying those disturbances about orthogonal axes.

Various other combinations are possible including the application of shock accelerations or thermal cycling the sensor during these combination tests.

This form of testing is usually valuable for assessing the response of the gyroscope when experiencing the total environment to which it is likely to be subjected in the host vehicle. It is useful to compare the response of the sensor resulting from the combination of stimuli with that expected from the sensor when experiencing the individual stimuli, such as the linear acceleration or rotation used in the combination test. Any anomalies can then be investigated in the laboratory and the actual basis of the behaviour established before the ultimate combination test series takes place, such as actual flight trials. Care must be taken to ensure that combinations of disturbances applied in the laboratory are realistic and not beyond the capability of the host vehicle, and consequently beyond the specified capability of the sensor.

Flight trials allow a qualitative assessment rather than a definitive analysis, owing to the difficulty of defining all the input stimuli very precisely. It is not normally possible to define the stimuli with the same precision as for the normal single stimulus testing such as rotation or multi-position testing.

8.5.10 Ageing and storage tests

Many project applications require a gyroscope to have a 'shelf life' of many years, 15 is not untypical, and still provide performance within specification at the end of that period. It is not easy to provide realistic accelerated ageing or storage tests other than through multiple thermal cycling and enhanced vibration testing to simulate transportation. One method of evaluating ageing is by analogy or reference to the behaviour of similar known components in other devices, such as permanent magnets and establishing the change in magnetic flux expected over the life of the sensor. This would then enable the designer to allow a suitable margin for the anticipated degradation in performance.

A further method for evaluating ageing is to characterise a large number of gyroscopes using say, multi-position tests and then storing them in a typical environment. Periodic withdrawal of a few at say 1- or 2-year intervals for re-evaluation should provide a guide to the form of ageing or change in performance of the sensor and thus allowing a comparison with the predicted changes. It is vital that the identical test procedure, along with identical calibrated test equipment, is followed during the various series of follow-up tests.

8.6 Accelerometer testing

The performance of an accelerometer is usually investigated using a series of static and dynamic test procedures similar to those already described for a gyroscope. However, a reduced scale of testing is required to characterise the performance of this device [13]. For example, rate table testing is generally unnecessary and the

3. rotation and vibration by mounting a rugged rate table on a vibrator and either applying a vibratory motion along one axis whilst simultaneously applying a rotational disturbance, or applying those disturbances about orthogonal axes.

Various other combinations are possible including the application of shock accelerations or thermal cycling the sensor during these combination tests.

This form of testing is usually valuable for assessing the response of the gyroscope when experiencing the total environment to which it is likely to be subjected in the host vehicle. It is useful to compare the response of the sensor resulting from the combination of stimuli with that expected from the sensor when experiencing the individual stimuli, such as the linear acceleration or rotation used in the combination test. Any anomalies can then be investigated in the laboratory and the actual basis of the behaviour established before the ultimate combination test series takes place, such as actual flight trials. Care must be taken to ensure that combinations of disturbances applied in the laboratory are realistic and not beyond the capability of the host vehicle, and consequently beyond the specified capability of the sensor.

Flight trials allow a qualitative assessment rather than a definitive analysis, owing to the difficulty of defining all the input stimuli very precisely. It is not normally possible to define the stimuli with the same precision as for the normal single stimulus testing such as rotation or multi-position testing.

8.5.10 Ageing and storage tests

Many project applications require a gyroscope to have a 'shelf life' of many years, 15 is not untypical, and still provide performance within specification at the end of that period. It is not easy to provide realistic accelerated ageing or storage tests other than through multiple thermal cycling and enhanced vibration testing to simulate transportation. One method of evaluating ageing is by analogy or reference to the behaviour of similar known components in other devices, such as permanent magnets and establishing the change in magnetic flux expected over the life of the sensor. This would then enable the designer to allow a suitable margin for the anticipated degradation in performance.

A further method for evaluating ageing is to characterise a large number of gyroscopes using say, multi-position tests and then storing them in a typical environment. Periodic withdrawal of a few at say 1- or 2-year intervals for re-evaluation should provide a guide to the form of ageing or change in performance of the sensor and thus allowing a comparison with the predicted changes. It is vital that the identical test procedure, along with identical calibrated test equipment, is followed during the various series of follow-up tests.

8.6 Accelerometer testing

The performance of an accelerometer is usually investigated using a series of static and dynamic test procedures similar to those already described for a gyroscope. However, a reduced scale of testing is required to characterise the performance of this device [13]. For example, rate table testing is generally unnecessary and the

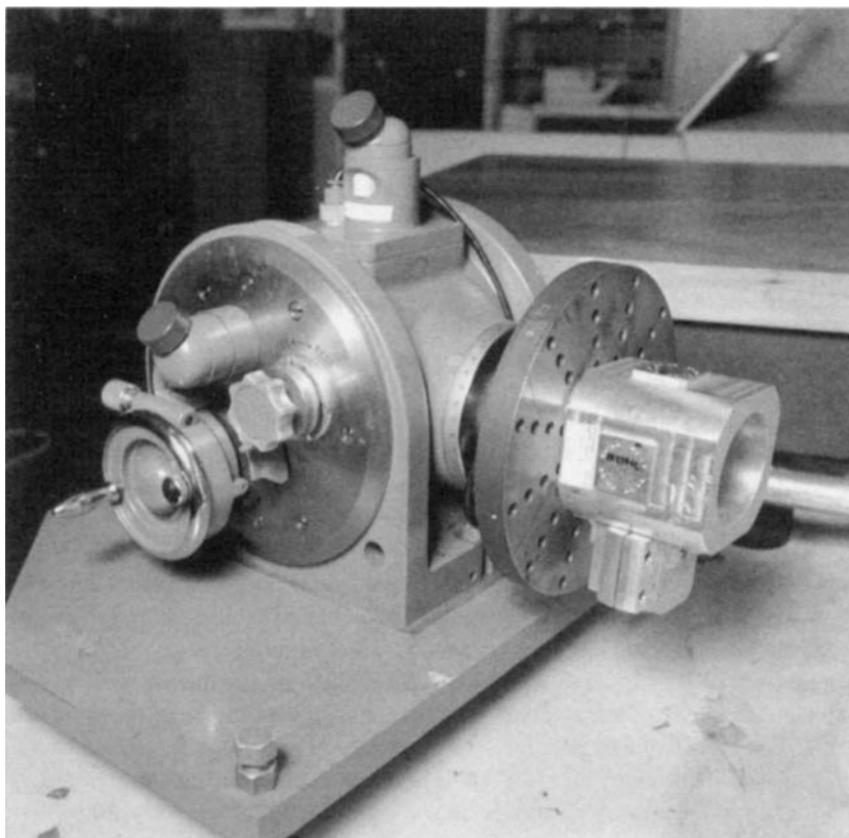


Figure 8.16 Photograph of a precision dividing head (Courtesy of QinetiQ)

multi-position tests are undertaken using a precision dividing head. This piece of equipment, which usually has a setting accuracy of about one second of arc, enables the sensitive (input) axis of an accelerometer to be rotated with respect to the gravity vector. Hence, the component of gravity acting along the input axis of the sensor may be varied very precisely. A photograph of such equipment is shown in Figure 8.16.

When a pendulous accelerometer is under test, care should be used when mounting the sensor to ensure that whenever possible the hinge axis is not vertical. This avoids the effects of frictional forces in the hinge influencing the output signals. Before conducting a series of tests to evaluate the performance of an accelerometer, preliminary tests are usually undertaken to ensure that the accelerometer is functioning as designed by the manufacturer. Typical tests include observation of the output for a short period (10–20 min) after switch-on to check the warm-up trends and the determination of the threshold acceleration level which produces an output signal. Small accelerations may be applied along the input axis of the sensor using the precision dividing head described above.

As in the case of gyroscope testing, the IEEE has issued a number of test procedure documents [7, 8, 14] for the testing of accelerometers. The following sections

concentrate on general test techniques for accelerometers, examples of test technique and methods of analysis, enabling the performance of an instrument to be evaluated.

8.6.1 Multi-position tests

The purpose of these tests is to determine the following parameters of an accelerometer:

1. scale-factor;
2. scale-factor linearity;
3. null bias error;
4. axis alignment error;
5. switch-on to switch-on repeatability.

The device to be tested is fastened firmly to a precision dividing head and generally, the output signals from the accelerometer are recorded for four different attitudes of the sensitive axis corresponding to $0g$, $1g$, $0g$, $-1g$ acting along this axis. At each setting, a series of data points are recorded and the whole test may be repeated a number of times, from which the above performance parameters may be calculated. Typical data showing the residual error between the calculated and measured value of the gravity vector for a fixed temperature are shown in Figure 8.17.

In the case of switch-on to switch-on repeatability evaluation, the tests in a given position are repeated at least twelve times, the whole procedure being repeated usually in each of the normal fixed positions for the tests. Sample data for a single switch-on stability test are shown in Figure 8.18. It is usual to monitor the temperature of the sensor during the warm-up period as shown in the figure. These tests are sometimes known as ‘tombstone’ tests.

8.6.2 Long-term stability

For these tests, the output signals from the accelerometer are recorded with the sensor fixed in a particular orientation. The duration of the test may be hours, weeks or even

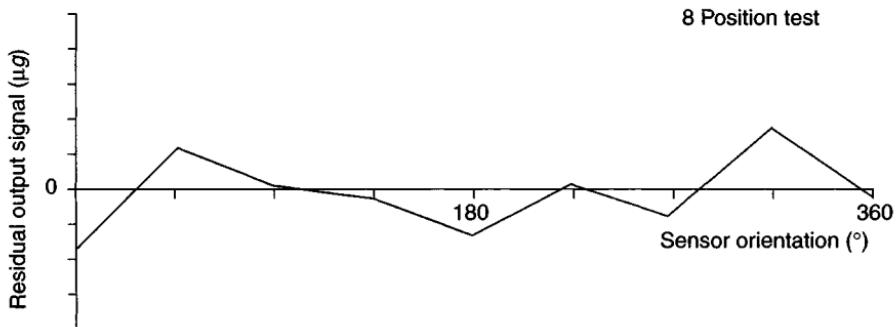


Figure 8.17 Sample data from multi-position testing

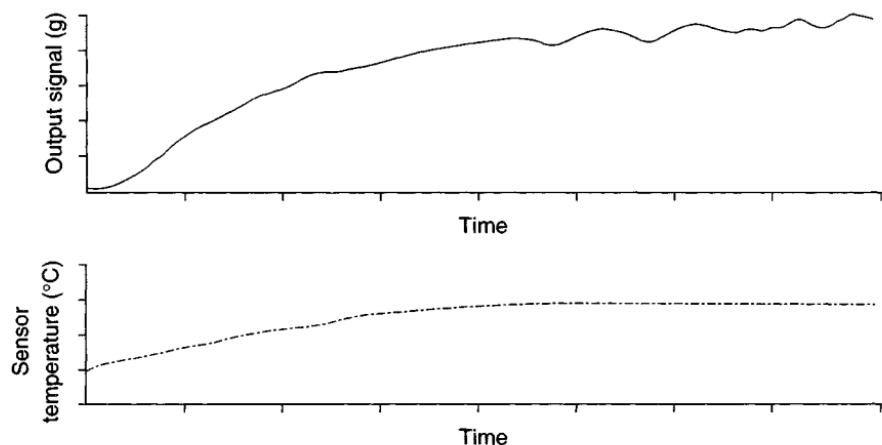


Figure 8.18 Sample data from switch-on stability testing

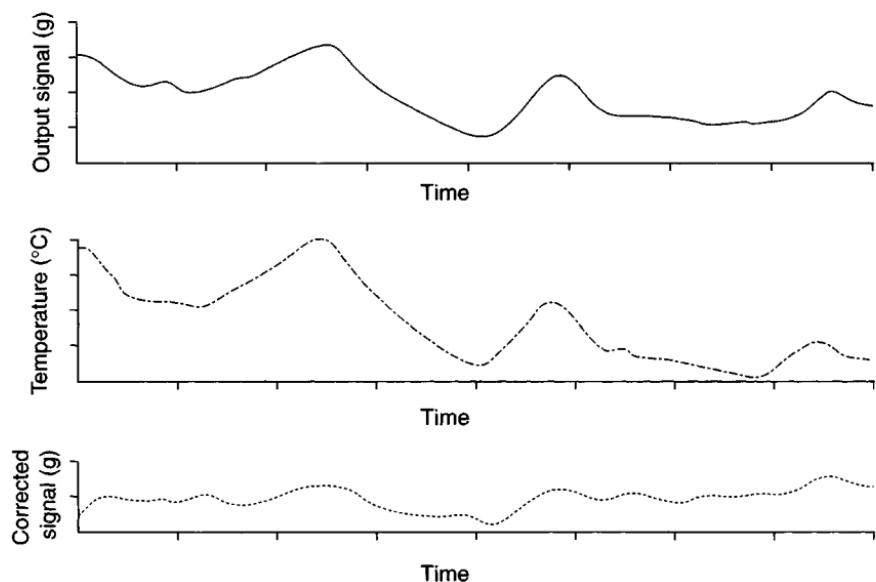


Figure 8.19 Sample data from long-term stability testing

longer. As in the case of gyroscope testing, the sensor is switched off for a period between tests before the procedure is repeated with the accelerometer in a different orientation, or series of different orientations. At each position, the ambient temperature is monitored and recorded for the duration of the test, enabling the accelerometer output signal to be corrected as shown in Figure 8.19.

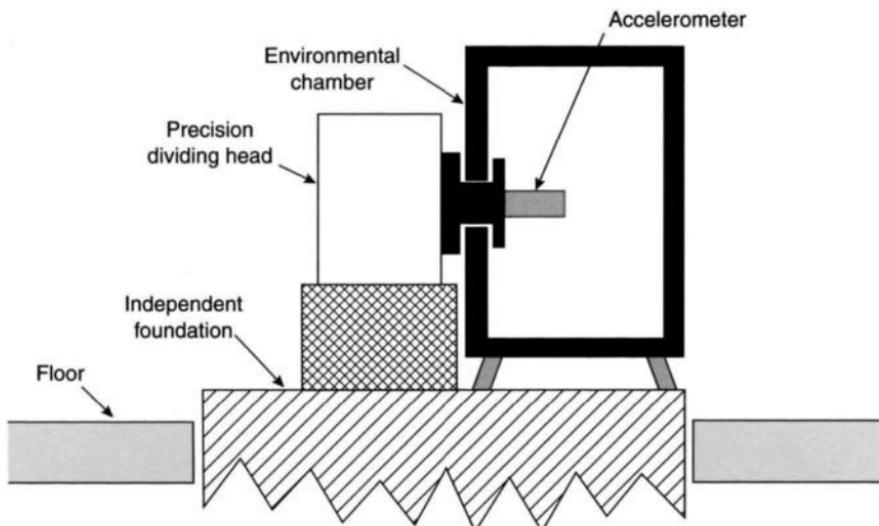


Figure 8.20 Schematic diagram of thermal test equipment

8.6.3 Thermal tests

The purpose of the thermal tests are to establish the variation in the basic parameters used to describe the performance of an accelerometer with temperature; either when the sensor is at a uniform elevated or depressed temperature, known as soaking, or with a temperature gradient across the sensor. The sensor under test on a precision dividing head is usually enclosed in a climatic chamber enabling tests to be carried out at sub-zero temperatures, typically down to -55°C , and elevated to high temperatures, usually up to 75 or 80°C . Figure 8.20 shows a typical experimental arrangement. Various tests can be undertaken to monitor the behaviour of the accelerometer when 'soaked' at elevated and depressed temperatures as the output signals are recorded at the four principal attitudes of the accelerometer. This enables variations in the parameters of performance to be estimated and where possible correlated with the temperature of the sensor during the test. Sample data are shown in Figure 8.21.

Additionally, the response of the accelerometer can be monitored during different rates of temperature decrease or increase. In this case, the accelerometer is set to a given orientation and the ambient temperature is varied linearly over a given range during a given period of time. During this period, the temperature is recorded and the tests are then repeated for a different orientation of the accelerometer.

8.6.4 Magnetic sensitivity tests

The purpose of these tests is to establish the effect that external magnetic fields have on the performance of an accelerometer. The tests take a similar form to those described for a gyroscope in Section 8.5.5. The sensor is set to the desired orientation with respect to the gravity vector and a pair of Helmholtz coils are arranged to enable

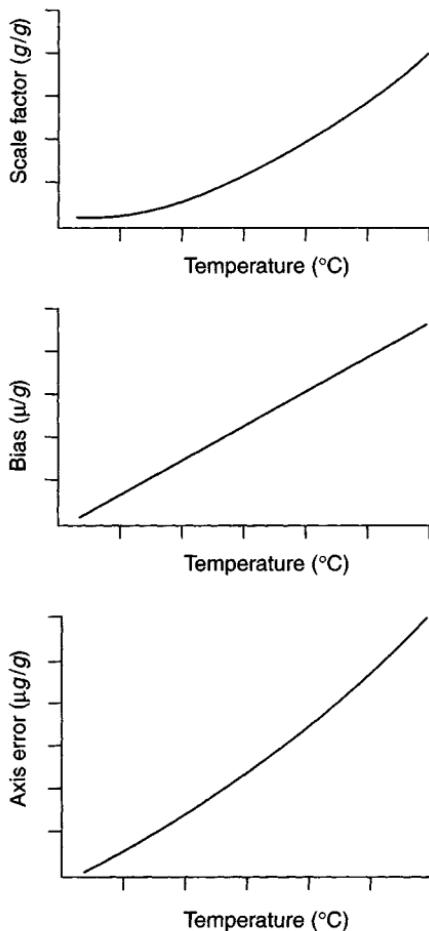


Figure 8.21 Temperature sensitivity of accelerometer parameters

a magnetic field to be applied along a chosen axis of the accelerometer. The response of the accelerometer is recorded before, during and after the application of the magnetic field. This series of tests is usually repeated for different field strengths and different orientations of the accelerometer with respect to the gravity vector and the magnetic field. From these tests, the behaviour of the accelerometer in the presence of magnetic fields, and the change in characteristics it may exhibit, can be assessed from any correlation between changes in performance and the applied magnetic field vector. An example of magnetic sensitivity is given in Figure 8.22.

8.6.5 Centrifuge tests

As noted in the case of the centrifuge testing of gyroscopes, this type of testing must be undertaken with care. It is normal for these tests to be taken towards the end of the

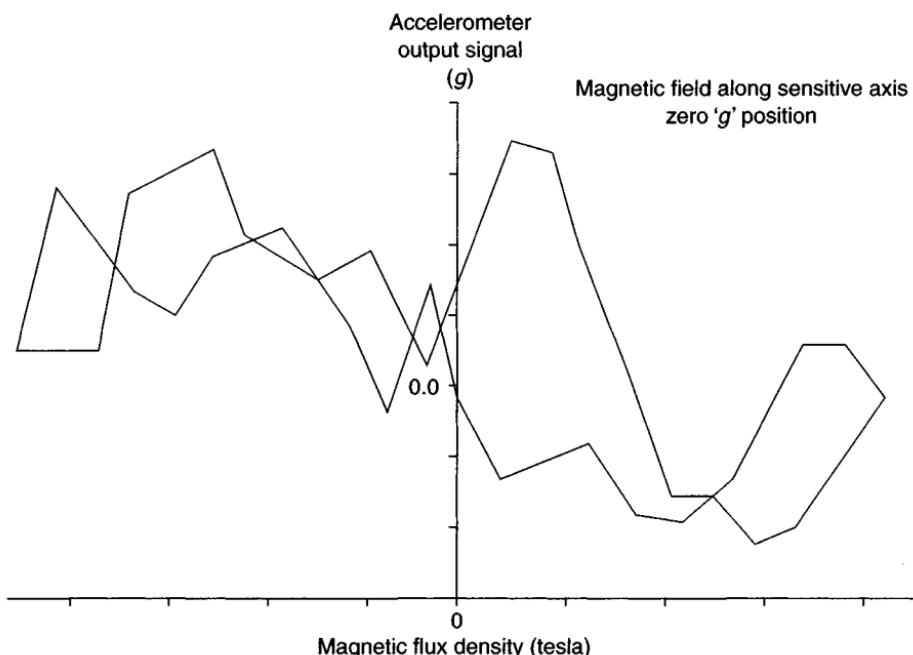


Figure 8.22 Magnetic sensitivity of an accelerometer

test schedule. The purpose of these tests is to record the response of the accelerometer to high values of acceleration compared with gravitational acceleration, and evaluate the linearity of the scale-factor at maximum values of the input acceleration [14]. This equipment can also be used to examine the resilience and tolerance of the sensor to applied accelerations in excess of the recommended or designed maximum acceleration that the sensor is designed to measure along its input axis. This form of testing is commonly referred to as over-range testing. Great care must be taken with these latter tests to avoid permanent damage to the sensor under test.

The equipment used and the form of testing is very similar to that already described for testing a gyroscope on a centrifuge in Section 8.5.6 where the sensor is either operational or in a quiescent state. Again it is crucial that the sensor under test is secured firmly and rigidly to the test platform. Testing can be undertaken with the sensitive axis parallel to the applied acceleration for the scale-factor linearity response tests. Sample data showing the residual error, the difference between the measured and actual acceleration, are shown in Figure 8.23.

When the sensitive axis is set orthogonal to the applied acceleration, then the cross talk or orthogonality of the axes can be examined. In these cases, the accelerometer will be operating in an energised state and the acceleration applied to the sensor is increased in small steps up to the maximum value it should measure and then reduced to zero step by step. At each step in this cycle, the response of the accelerometer is recorded for a given period and the response compared with that predicted by the

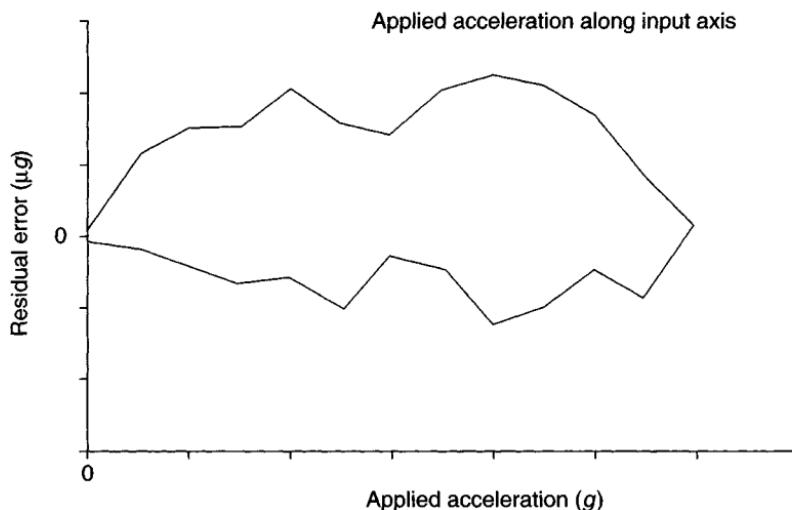


Figure 8.23 Accelerometer data from a centrifuge test

multi-position tests with the sensor on the dividing head. This cycle is repeated as many times as is necessary to generate confidence in the data, and where appropriate, the sensor is rotated through 180° to record the response of the accelerometer to accelerations in the opposite direction with respect to the sensitive axis.

The sensor will also be operational for the over-range testing where the applied acceleration is increased in small steps above the normal permitted maximum, the response of the accelerometer being recorded at each incremental step. Usually, an acceleration in the region of 10–20 per cent above the maximum value is used in these tests unless the objective is to test the sensor to destruction. After an over-range test, it is usual to repeat at least part of the multi-position testing to establish that the accelerometer has not been damaged permanently by the over-range testing. If the objective was to damage the sensor, then multi-position tests can establish the new characteristics of the sensor after damage.

The sensor can be mounted on the centrifuge at an angle of 45° to the applied acceleration vector to establish the response of the sensor when a large acceleration is applied simultaneously along two orthogonal axes. As with the conventional arrangement of the axes, the response of the sensor is recorded at each step as the applied acceleration is incremented to a maximum value, and then reduced in steps to zero. This form of testing allows cross-axis sensitivity to be estimated at high values of acceleration through comparison of recorded data with those predicted from the earlier tests.

For some types of resilience testing, the sensor will not be operational. In this case, the sensor is characterised by, for example, multi-position testing and then tested in the centrifuge in a quiescent state. It is then re-characterised to establish the form of any change in the performance characteristics of the accelerometer.

Another form of dynamic test is to mount the accelerometer on a wedge, usually with a 30° or 60° angle, with the normal to the slope pointing away from the axis of rotation. The orientation of the input axis of the sensor is then aligned so that it senses a particular or desired component of the gravity vector when the centrifuge is at rest. When the accelerometer is rotated, this component of the gravity vector can be nulled by the applied acceleration, and hence, the acceleration can be taken through the zero position dynamically.

8.6.6 Shock tests

The purpose of these tests is to examine the response of the accelerometer to an applied shock, or to evaluate the resilience of the device to such an applied acceleration of very short duration, typically of the order of milliseconds [12]. As in the case of the centrifuge testing described above, the accelerometer may be operating or quiescent during these tests.

The form of the testing and the equipment used is very similar to that already described for evaluating the response of a gyroscope to applied shocks in Section 8.5.7 and will not be repeated here.

8.6.7 Vibration tests

As in the case of gyroscope testing, this is normally the last series of tests undertaken on the sensor owing to the possibility of permanent and irreversible damage to the accelerometer. The type of equipment used is identical to that discussed in Section 8.5.8.

The purpose of these tests is normally five-fold:

1. investigation of any resonant responses of the sensor;
2. evaluation of the vibro-pendulous error of the accelerometer;
3. examination of the frequency response of the instrument;
4. examination of the resilience of the device in a particular vibratory environment, depending on the reason for the test the accelerometer may be quiescent or operational;
5. estimation of the change in noise characteristics of the signals generated by an accelerometer experiencing a vibratory environment.

The testing is undertaken using equipment identical to that already described in Section 8.5.8 for the testing of gyroscopes in a vibratory environment. Additionally, as already described in that section, the motion imparted to the sensor can be random or sinusoidal.

An initial series of tests is usually undertaken with the accelerometer operational in order to establish the frequency and magnitude of any resonances that can be excited in the sensor. This enables the investigation of the sensor to be undertaken at frequencies well away from those that will excite resonances in the instrument. This so-called resonance search is undertaken in an identical manner to that described already for

gyroscopes, in Section 8.5.8, using the so-called sine sweep. The warnings given in Section 8.5.8 regarding bolting the interface or mounting block firmly to the test table, and also undertaking a thorough investigation of this mounting block to detect any resonances, should be observed.

Vibro-pendulous rectification is a bias produced by the forced motion of the pendulum in a pendulous accelerometer when it is subjected to linear vibratory motion. This bias appears when the motion is applied along an axis at 45° to the input axis and lying in a plane containing both the input and pendulum axes. The magnitude of this error is evaluated by aligning the accelerometer to be tested so that the applied vibratory motion is along an axis, as described above, enabling the bias to be observed. Careful alignment is necessary, and it is usual to mount the accelerometer on a precision machined interface block that positions the sensor in the requisite orientation.

The accuracy of the mounting can be checked with the sensor operational by rotating the vibrating table through exactly 45° to bring the sensitive, or input, axis to the horizontal and ensuring it registers zero g output. The position is adjusted to give a precise zero g output. The vibration table should then be rotated back through exactly 45°. A control accelerometer is also used to monitor the applied motion and other sensors may be applied to the test block, to monitor the motion about the orthogonal axes.

Initially, the bias of the accelerometer is observed by collecting data over a short period. The vibratory motion is then applied to the sensor, at a given frequency and acceleration value. The output signals from the accelerometer under test are recorded over a period of typically one minute, having been filtered if necessary. The change in bias is evaluated and divided by the peak acceleration squared to give the value of the vibro-pendulous error. The test is repeated a number of times to give confidence in the estimate of this error, and repeated at different frequencies and peak acceleration values. Each frequency is chosen to be well removed from any resonant frequencies. Finally, an average value can be estimated from the matrix of values calculated.

The frequency response of an accelerometer is established by mounting the accelerometer firmly on a vibration table and applying sinusoidal vibratory motion in the direction of its sensitive axis. During any one test, the peak acceleration is held constant and the frequency is varied across a given bandwidth, typically from 25 Hz to 2 kHz. The output signal from the accelerometer is recorded and compared with a reference signal from a piezoelectric crystal using a transfer function analyser. This enables gain and phase responses to be derived as shown in Figure 8.24.

The resilience of an accelerometer to a vibratory environment may be evaluated using techniques identical to those already described for a gyroscope. This is discussed in detail in Section 8.5.8. Observations of the recorded data, along with calculated statistical data, enable estimates to be made of changes in the noise characteristics of a sensor resulting from the vibratory motion experienced by that device.

8.6.8 Combination tests

This type of testing is sometimes appropriate to estimate the performance of an accelerometer in a realistic environment where there are combinations of forces or disturbances acting on the sensor. The basic philosophy for this form of testing has

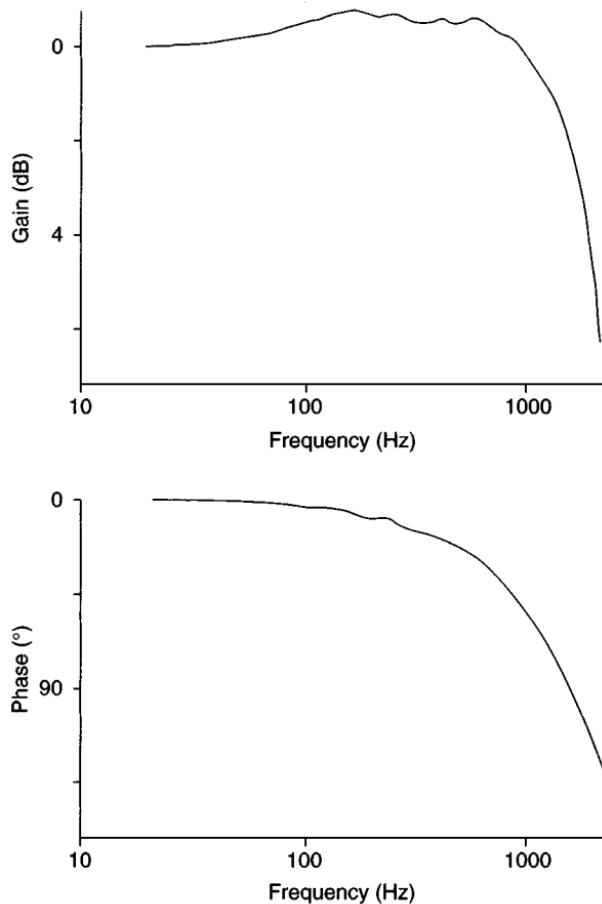


Figure 8.24 Accelerometer dynamic response

already been described in Section 8.5.9 for the evaluation of gyroscopes and will not be repeated here. The warnings given in Section 8.5.9 regarding damage to the sensor through over stressing are equally important here.

8.6.9 Ageing and storage tests

The problems, basic philosophy and constraints described for the evaluation of gyroscopes in such tests in Section 8.5.10, apply to the testing of accelerometers. It is possible that an agency will undertake these studies on behalf of a project and evaluate simultaneously both the accelerometers and the gyroscopes used in a particular vehicle. Generally, the storage and ageing problems associated with pendulous accelerometers centre around deterioration of the permanent magnets, so some prediction in the change in scale-factor can be made from knowledge of the magnetic material used.

8.7 Calibration and error compensation

8.7.1 Introduction

The sensors are calibrated by comparing the analogue or digital signals produced by the sensor, with the known input motion. Thus, from the rate transfer tests, the output signals from a gyroscope can be compared with the accurately known rotation rate and the scale-factor defined as so many millivolts per degree per second of rotation rate, for instance. Similarly, using the gravity vector as an accurate standard, the scale-factor of an accelerometer can be defined.

There are several different levels at which error compensation can be applied. However, the fundamental idea is the same for all; to correct the effects of a predictable systematic error, or errors, on the accuracy of a sensor. Additionally, a basic requirement is that an error process can be represented by an equation and hence modelled mathematically, and that a signal corresponding to the disturbing effect, such as temperature or acceleration, is available and can be measured to the required accuracy.

Predictable error components can be estimated from observations of performance, and used in the opposite sense to correct, or compensate, for the imperfections in the sensor performance. Often this technique relies on the use of a constant coefficient in the error representation, but for more demanding applications, or complex error behaviour, it is common to use polynomial representations.

The scope of possible error compensations can range from the correction of a single error parameter using a constant applied to a whole batch of instruments, to multi-parameter compensation of individual instruments using complex time varying polynomials. In the case of the former type of compensation, averaged data are used from a 'production line' evaluation for example, and it is essentially compensating a systematic trend of error displayed by a particular class or design of sensor. The other extreme requires the characterisation of each sensor through a series of laboratory tests and then correcting the performance of each sensor according to the test results. In this case, the error coefficients for the observed errors on each sensor are evaluated quantitatively and thus the observed measurements are corrected for the known systematic errors. This approach has become more feasible with modern computer and data storage technology and has been used with MEMS sensor technology to give high performance devices.

The former type of error compensation is usually very easy to implement at the system level, particularly as each unit is treated identically. In the latter case, the compensation is more complex and difficult to handle as the degree of compensation can vary from system to system. Of course, each sensor will have its own set of error coefficients which have to be read into the system processor. The coefficients for correction may be stored electronically in a 'chip' located in each sensor. However, this technique does offer considerable scope for a significant improvement in the performance of each sensor and, consequently, for the system in which they are used. The choice of compensation technique is usually a trade-off between various factors, but particularly the balance between the ease of implementation and the

benefits gained from improved sensor accuracy. Another consideration of course, is a trade-off in the extra cost of the detailed characterisation and the price to be paid in using a more accurate instrument, or instruments, with inherently better performance.

The common types of errors which occur in gyroscopes and accelerometers were described in Chapters 4–6. Some examples are given below to illustrate how such errors can be modelled and thus compensated in systems using different types of sensor.

8.7.2 Gyroscope error compensation

As described in Chapter 4, the measurement of turn rate ($\tilde{\omega}_x$) provided by a conventional gyroscope may be expressed in terms of the applied rate about its input axis (ω_x) as:

$$\tilde{\omega}_x = (1 + S_x)\omega_x + M_y\omega_y + M_z\omega_z + B_f + B_{gx}a_x + B_{gz}a_z + B_{axz}a_ya_z + n_x \quad (8.3)$$

where a_x and a_z are the accelerations of the gyroscope along its input and spin axes respectively. B_f is the g -insensitive bias, B_{gx} , B_{gz} are the g -sensitive bias coefficients, B_{axz} is the anisoelastic bias coefficient, n_x is the zero-mean random bias, M_y , M_z are the cross-coupling coefficients, and S_x is the scale-factor error which may be expressed as a polynomial in ω_x to represent scale-factor non-linearities.

As discussed in Chapter 4, each of the error coefficients has repeatable and non-repeatable components. The repeatable components of S_x , M_y , M_z , B_f , B_{gx} , B_{gz} and B_{axz} are measurable and their effects can therefore be compensated. It is not practical to attempt to compensate the in-run random error and such effects can only be controlled by careful sensor design and manufacture.

Similar forms of the above expression may be used to model the bias and scale-factor errors arising in two axis gyroscopes such as the dynamically tuned gyroscope. For optical gyroscopes, such as the ring laser gyroscope, the acceleration sensitive errors are negligible and measured rate may be modelled for many applications as:

$$\tilde{\omega}_x = (1 + S_x)\omega_x + M_y\omega_y + M_z\omega_z + B_f + n_x \quad (8.4)$$

As with mechanical sensors, fixed bias, cross-coupling and fixed scale-factor errors can generally be measured to sufficient accuracy to allow some effective compensation to take place. Other errors are less predictable and therefore cannot be controlled using on-line correction methods of the type considered here.

8.7.3 Accelerometer error compensation

As described in Chapter 6, a measurement provided by an accelerometer (\tilde{a}_x) may be expressed in terms of an applied acceleration and the sensor error coefficients as follows:

$$\tilde{a}_x = (1 + S_x)a_x + M_ya_y + M_za_z + B_f + B_v a_x a_y + n_x \quad (8.5)$$

where a_x represents the acceleration applied in the direction of the sensitive axis and a_y and a_z are the accelerations applied perpendicular to the sensitive axis, S_x is the scale-factor error, usually expressed in polynomial form to include non-linear effects, M_y , M_z are the cross axis coupling factors, B_f is the measurement bias, B_v is the vibro-pendulous error coefficient and n_x is the random bias.

In general, the fixed bias, cross-coupling and the scale-factor error coefficients can be measured and corrections can therefore be applied to offset the repeatable components of these errors. The error contributed by the vibro-pendulous effect in the presence of slowly varying accelerations may also be compensated to a large extent. However, random biases and vibration-dependent errors cannot be compensated accurately.

8.7.4 Further comments on error compensation

Compensation techniques are designed to remove the predictable error terms in the inertial sensors by measuring the appropriate error coefficients and using these values to apply corrections to the measurements. These are usually implemented in the system software. In addition to the compensation of sensor errors, it is also common practice to compensate for any system errors such as sensor mounting misalignments which can be measured and are thus predictable in their effects on system performance.

It will be remembered from the discussions in Chapters 4–6 that the measured coefficients vary with time, temperature, vibration, applied motion and from switch-on to switch-on, and it is these variations which ultimately determine system performance. Whilst it is not possible to compensate for most of these effects, temperature compensations are often essential to achieve a given performance goal.

Thermal effects on bias and scale-factor errors can be very significant and are often difficult to model accurately. This is because in some sensors, especially mechanical gyroscopes, temperature gradients within the sensor can alter the performance of many of its components. Therefore, it is usual for a very accurate inertial system to control the temperature of its sensors very precisely. Consequently, such a system may have a significant warm-up time. However, this does of course alleviate the need for complex and difficult thermal compensation.

The accuracy that may be achieved from the application of compensation techniques is dependent on precisely how the coefficients in the ‘error’ equation represent the actual sensor errors. This representation can often vary as a function of time, the environment in which the sensor is used and how often it is used. For the more demanding applications, it may be necessary to re-calibrate the sensor regularly, to ensure that the compensation routines are as effective as required by the particular application.

8.8 Testing of inertial navigation systems

Depending on the form and nature of the inertial navigation system, it may be appropriate to test either the complete inertial navigation system, or just the inertial

measurement unit in the laboratory. Usually, when the development of a system has reached the point for laboratory testing, the characteristics of the component sensors used will be very well known and the purpose of the testing will be somewhat different from the objectives of the ‘component tests’. Often similar tests are undertaken, but with the aim of checking the performance of the system, for example, that it behaves as predicted from the knowledge of the performance of its component inertial sensors. Sometimes, there can be adverse interaction between these components. Examples include cross-talk between vibratory sensors or changes in the lock-in characteristics of mechanically dithered ring laser gyroscopes mounted on the same structure. A manufacturer will wish to check that the units built on a production line will meet the design specification and the customer will also wish to confirm that the inertial system will fulfil the requirements of his or her particular application.

Typically, such tests involve mounting an inertial measurement unit, or a full navigation system, on a multi-axis table or on a rig. The unit may then be rotated through a series of accurately known angles and positioned in different orientations with respect to the local gravity vector, as shown in Figure 8.25. The dominant sensor errors may then be determined from static measurements of acceleration and turn rate taken in each orientation of the unit. An example of this type of approach is described by Joos and Krogmann [15] in relation to a system which uses conventional sensors. The unit under test is placed on a precision three axis table, and a series of constant rate tests and multi-position tests are then used to allow the major sources of error to be identified.

Alternatively, estimates of the system errors may be obtained by testing the inertial measurement unit as a component of a full strapdown inertial navigation system. In this case, the estimates made by the navigation system of angle turned through and/or linear acceleration in the navigation reference frame may be used to deduce the various system errors as discussed below. An example of the latter approach is described by Brown *et al.* [16] for the testing of an inertial system containing ring laser gyroscopes. In the scheme described, the unit is made to rotate through a sequence of turns using a two degree of freedom table. Immediately prior to each turn, the system performs a self-alignment with respect to the navigation frame. The system itself then keeps track of the table rotations throughout each turn. Components of acceleration in the navigation reference frame are computed immediately on completion of each turn, by resolving the measured accelerations into the navigation frame using the computed attitude information. Errors in the computed accelerations may then be attributed to a combination of acceleration and angular rate measurement errors. By rotating the unit through a carefully chosen set of turns, it is possible to obtain estimates of the dominant sensor errors. A possible advantage of this type of approach is the avoidance of any need to use a highly accurate and expensive test table, since the positioning of the unit with respect to the local gravity vector is less critical.

A scheme of tests is suggested below which allows a detailed investigation of an inertial measurement unit or a full inertial navigation system.

Static acceleration tests. By mounting the navigation system on a level table with each sensitive axis pointing alternately up and down (six-position test), it is possible to

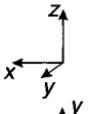
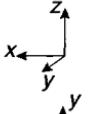
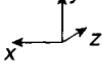
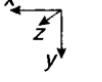
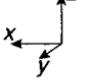
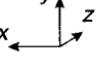
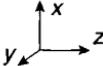
Rotation number	Start position	Turn	Test position
1		90° about y	
2		90° about x	
3		-180° about x	
4		90° about x	
5		-180° about y	
6		90° about y	
7		90° about x	
8		90° about z	
9		-180° about z	
10		90° about z	
11		-90° about x	

Figure 8.25 Test rotations for a strapdown inertial measurement unit (IMU)

extract estimates of the accelerometer biases, scale-factor errors and the sensitive axis misalignments with respect to a set of datum mounting faces. These estimates can be computed by summing and differencing various combinations of accelerometer measurements.

Static rate tests. By monitoring the angular rate measurements provided by the system for a pre-defined period of time, for a number of different orientations of the unit, it is possible to extract estimates of the gyroscope fixed biases and g -dependent biases. As in the previous tests, sum and differencing techniques may be used to separate the various error contributions.

Angle tests. Using a precision multi-position test table, the inertial measurement unit may be rotated through very accurately known angles. By comparing these known rotations with estimates of these turn angles derived by integrating the rate outputs provided by the gyroscopes, it is possible to derive estimates of the various errors in the gyroscopic measurements. For instance, if the table is rotated clockwise and counter-clockwise through the same angle, estimates of gyroscope to case mounting error may be derived along with gyroscope biases and scale-factor errors.

Inertial navigation system multi-position tests. This method makes use of the propagation of sensor imperfections as errors in the components of acceleration derived in the navigation reference frame. Provided a sufficient number of test rotations take place, this method may be used to extract estimates of most of the dominant sensor errors associated with a strapdown system containing conventional gyroscopes.

In selecting a suitable set of rotations for the testing of a particular unit, it is often useful to include separate 90° and 180° turns about each axis. For the testing of systems containing conventional gyroscopes, compliance with this requirement allows many different linear combinations of the gyroscope's fixed biases, g -dependent errors and mounting misalignments to appear in the measurement equations, and hence to become observable. Further, it is necessary to ensure that each axis of the unit is aligned with the local vertical in separate test orientations, both the up and the down positions, to allow the accelerometer biases, scale-factor errors and mounting misalignments to be identified. A sample set of rotations which satisfies these requirements, and which may be implemented using a two degree of freedom test table, is shown in Figure 8.25.

In practice, the series of tests outlined above may be carried out in a recursive manner, using the error estimates from one test to update or correct the error model used in subsequent tests. Various signal processing techniques are used to extract estimates of the system errors. For laboratory testing of the type described here, it is often sufficient to use least-squares methods to compute the magnitude of the errors. Kalman filtering techniques are often used as an alternative to the least-squares method.

On completion of these test procedures, a sequence of separate rate, centrifuge and vibration tests would normally be used in order to assess system performance over a range of dynamic conditions commensurate with those expected during operational use.

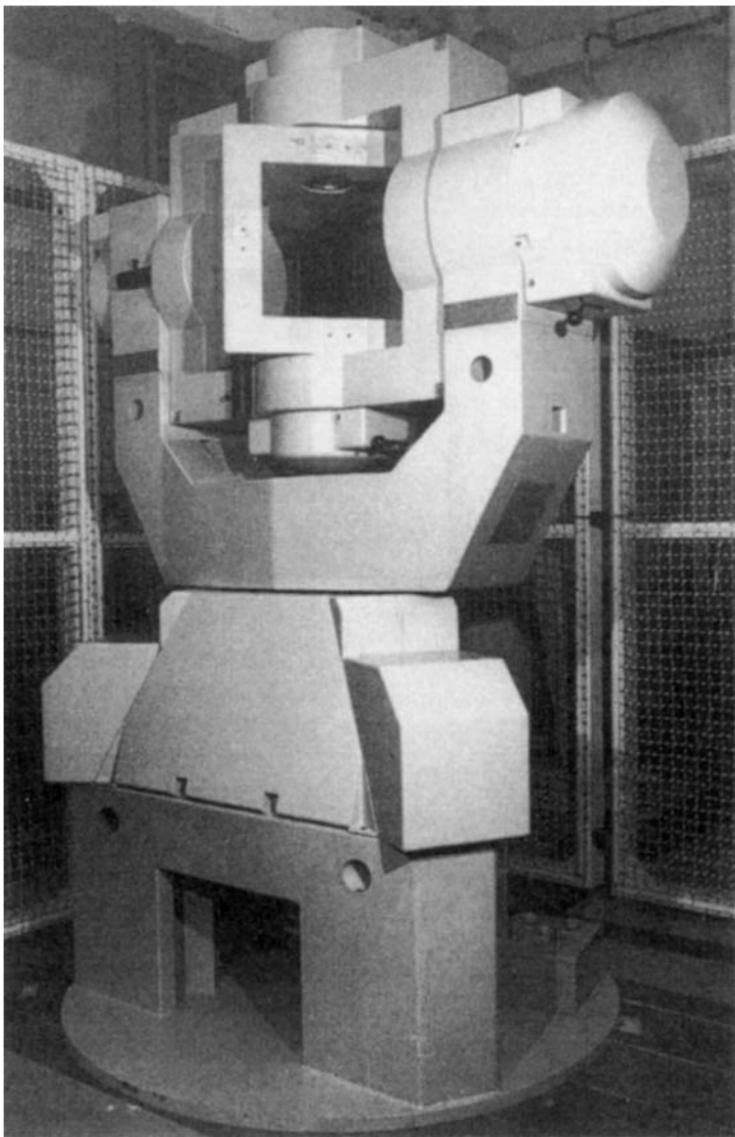


Figure 8.26 Photograph of a three-axis table used for navigation system testing
(Courtesy of CPE)

8.9 Hardware in the loop tests

One form of evaluation that is currently very popular with missile designers and engineers, which is a consequence of the enormous advances in computer technology, is called hardware in the loop testing. This form of testing is often used by a project during its research and development phase in order to establish an accurate estimate of

the performance of a missile, for example. Sometimes an inertial sensor, but usually a complete inertial system, is mounted on a test rig. This is usually a multi-axis test table, as shown in Figure 8.26, that can reproduce accurately the angular motion that the system would experience during its operational life, such as a flight in a missile. Normally, the test rig is limited to angular rotations. The output signals from the device under test are connected, through a suitable interface, to a computer simulating the motion and performance of the vehicle. This mathematical model in the computer is also generating the signals that control the test rig, hence a simulation operating in real time, using actual hardware, can be configured to enable realistic performance assessments to be established for complex systems operating in various flight regimes.

References

- 1 ANDREWS, T.R.: 'Standard gyro and accelerometer terminology' (Compiled by the Gyro and Accelerometer Panel, Aerospace and Electronic Systems Society, IEEE, 1975)
- 2 'IEEE test procedure for single-degree-of-freedom spring restrained rate gyros'. ANSI IEEE Standard 293, 1969
- 3 'IEEE standard specification format guide and test procedure for single-degree-of-freedom rate-integrating gyros'. IEEE Standard 517, 1974
- 4 'Supplement for strapdown applications to IEEE standard specification format guide and test procedure for single-degree-of-freedom rate-integrating gyros'. IEEE Standard 529, 1980
- 5 'IEEE specification format guide and test procedure for two-degree-of-freedom dynamically tuned gyros'. ANSI IEEE Standard 813, 1988
- 6 'IEEE specification format guide and test procedure for single axis laser gyros'. ANSI IEEE Standard 647, 1981
- 7 'IEEE standard specification format guide and test procedure for linear single axis, pendulous, analogue, torque balance accelerometer'. IEEE Standard 337, 1972
- 8 'IEEE standard specification format guide and test procedure for linear, single-axis, digital torque-balance accelerometers'. IEEE Standard 530, 1978
- 9 KIRKPATRICK, E.G.: 'Introductory statistics and probability for engineering science and technology' (Prentice-Hall, 1974)
- 10 TOPPING, J.: 'Errors of observation and their treatment' (Chapman and Hall, 1975)
- 11 BOX, G.E.P., HUNTER, W.G., and HUNTER, J.S.: 'Statistics for experimenters' (Wiley, 1978)
- 12 HARRIS, C.M., and CREDE, C.E. (Eds.): 'Shock and vibration handbook. vol. 1: Basic theory and measurements, vol. 2: Data analysis, testing and methods of control, vol. 3: Engineering design and environmental conditions' (McGraw-Hill, 1961)

- 13 SMITHSON, T.G.: 'A review of the mechanical design and development of a high performance accelerometer'. *Proceedings Institute of Mechanical Engineers. Mechanical Technology of Inertial Devices*, 1987, **C49/87**
- 14 'IEEE recommended practice for precision centrifuge testing of linear accelerometers'. IEEE Standard 836, 1991
- 15 JOOS, D.K., and KROGMANN, U.K.: 'Estimation of strapdown sensor parameters for inertial system error compensation'. AGARD Symposium, *Precision positioning and inertial guidance sensors. Technology and operational aspects*, October 1980
- 16 BROWN, A., EBNER, R., and MARK, J.: 'A calibration technique for a laser gyro strapdown inertial navigation system'. DGON Proceedings, *Gyro Technology Symposium*, Stuttgart, 1982

Chapter 9

Strapdown system technology

9.1 Introduction

The preceding chapters have described the fundamental principles of strapdown navigation systems and the sensors required to provide the necessary measurements of angular rate and specific force acceleration. In this chapter, aspects of strapdown system technology are discussed.

9.2 The components of a strapdown navigation system

As indicated in the earlier discussion, a strapdown inertial navigation system is basically formed from a set of inertial instruments and a computer. However, for reasons which will shortly become clear, such a system may be sub-divided further into the following component parts:

- instrument cluster;
- instrument electronics;
- attitude computer;
- navigation computer.

These components, which form the basic building blocks of a full strapdown navigation system, are shown schematically in Figure 9.1. These units will be mounted in a case, together with the necessary electrical power supplies and interface electronics, which may then be installed in a vehicle requiring an on-board navigation capability. Whilst it is often assumed that the unit will be fixed rigidly in the vehicle, it is usually necessary for it to be installed on anti-vibration (AV) mounts to provide isolation from vehicle motion at frequencies to which the unit is particularly sensitive.

Whilst we are primarily concerned here with the implementation of a full inertial navigation system, applications arise in which the full navigation function is not required. For example, in some short range missile applications, inertial

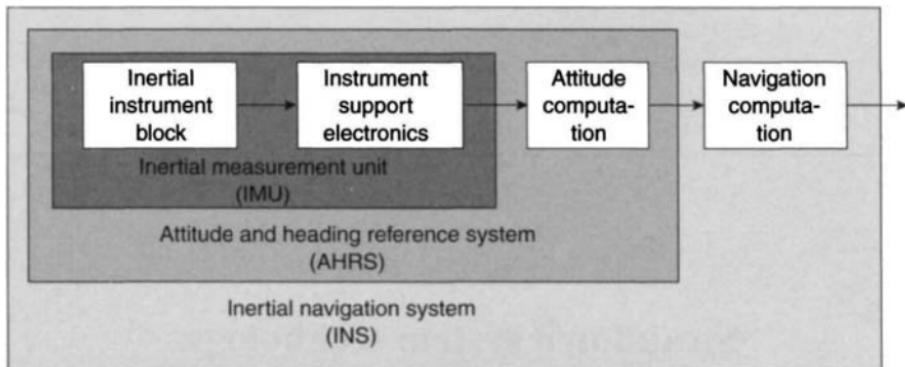


Figure 9.1 Strapdown inertial navigation system building blocks

measurements, typically of angular rate and specific force, are required purely for flight control purposes. In such cases, the instrument cluster and instrument electronics blocks alone are used to form what is known as an inertial measurement unit or IMU. For other applications requiring attitude and heading information alone, the IMU is combined with a processor in which the attitude equations are solved. The resulting system is known as an attitude and heading reference system or AHRS, the processor being referred to here as the attitude computer. An AHRS is sometimes used in combination with a Doppler radar to form a navigation system. Finally, the addition of a further computer in which the navigation equations are solved provides a full inertial navigation capability.

In the following sections, the components of the strapdown inertial navigation system defined above are described separately in more detail. This includes some discussion of the requirements for internal power supplies and AV mounts.

9.3 The instrument cluster

9.3.1 Orthogonal sensor configurations

The instrument cluster usually includes a number of gyroscopes and accelerometers which provide measurements of angular rate and specific force, respectively. The unit may contain either three single-axis gyroscopes or two dual-axis gyroscopes, as well as three single-axis accelerometers, all attached to a rigid block which can be mounted in the body of the host vehicle, either directly or on AV mounts. The sensitive axes of the instruments are most commonly mutually orthogonal in a Cartesian reference frame, as illustrated in Figure 9.2.

This arrangement of the instruments allows the components of angular rate and specific force in three mutually orthogonal directions to be measured directly, thus providing the information required to implement the strapdown computing tasks.

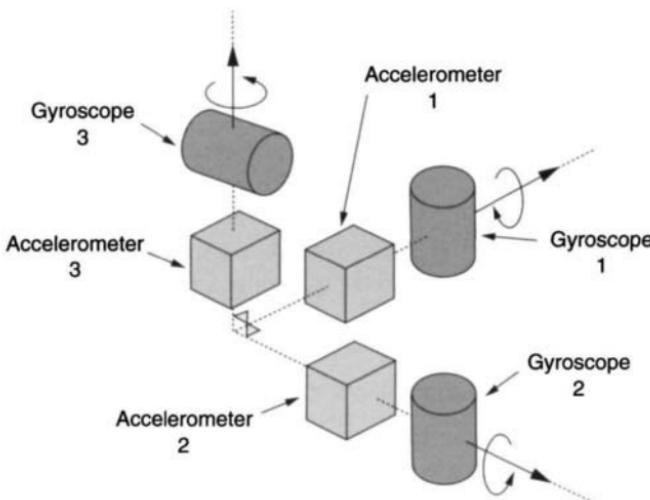


Figure 9.2 Orthogonal instrument cluster arrangement

As indicated earlier, systems using dual-axis sensors such as the dynamically tuned gyroscope as an alternative to the single-axis rate-integrating gyroscope require one less sensor. A dual-axis sensor configuration also provides an additional rate measurement. Through careful choice of the relative orientation of the two dual-axis sensors, the redundant measurement provided by one gyroscope may be used to monitor the performance of the other as part of a built-in test facility. In practice of course, many other factors will influence the choice between these two gyroscopes and their different characteristics are discussed in Chapter 4.

Other instrument arrangements are possible using modern sensing techniques which offer various novel approaches. For instance, a pair of multi-sensors mounted orthogonally or a single laser triad with suitable accelerometers may be used to form an instrument cluster. Such sensors are described in Chapters 5 and 6.

9.3.2 Skewed sensor configurations

Theoretically, it is possible to mount the instruments in orientations other than the orthogonal arrangement illustrated in Figure 9.2. This type of configuration will function provided the measurements may be expressed as independent linear combinations of the orthogonal components of angular rate and specific force. The orthogonal components may then be extracted from the measurements as part of the strapdown processing task. Such instrument configurations are referred to as skewed sensor configurations and may be used to advantage in certain applications. A practical implementation is the silicon drive discussed in Section 7.7.1.

Skewed sensor arrangements are used primarily in applications requiring on-line failure detection and fail-safe operation as discussed in Section 9.3.4.

However, they may also be used in situations where the turn rate about a single-axis of a vehicle may exceed the normal operating range of a gyroscope having performance characteristics suitable for the particular application. By mounting the gyroscopes so that their sensitive axes form an angle with the high rate axis of the vehicle, it is possible to ensure that the resolved component of the turn rate falls within the maximum range of the sensor. Given knowledge of the skew angle, it is possible to calculate the turn rate about the vehicle axis using the measurements provided by the skewed sensors. An example of this technique based on a system using two dual-axis gyroscopes is discussed in the following section.

9.3.3 A skewed sensor configuration using dual-axis gyroscopes

Two dual-axis gyroscopes may be configured in the symmetrical form illustrated in Figure 9.3. A body axis frame $Ox_b y_b z_b$ is indicated along with gyroscope axes $Ox_1 y_1 z_1$ and $Ox_2 y_2 z_2$. The gyroscope spin axes lie in the $Ox_b y_b$ plane in the directions Ox_1 and Ox_2 , inclined at angle Φ to Ox_b .

The turn rates about the respective body axes are denoted ω_x , ω_y and ω_z . This particular configuration may be used where the rotation rate (ω_x) about the axis x_b exceeds the normal operating range of the gyroscopes. For the arrangement shown in the figure, the turn rates ω_A , ω_B , ω_C and ω_D , sensed by the gyroscopes may be

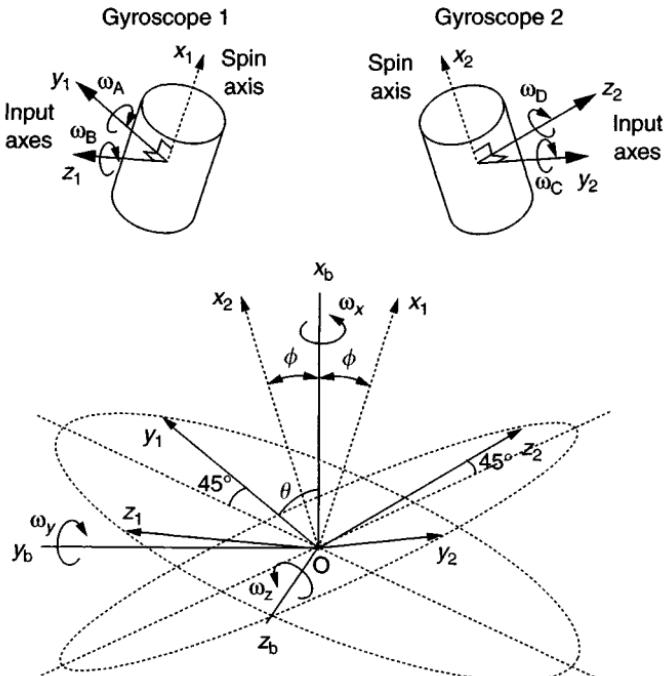


Figure 9.3 Dual-axis gyroscope skewed configuration

expressed in terms of the body rates as follows:

$$\begin{aligned}
 \text{Gyroscope 1:} \\
 \omega_A &= \frac{(\omega_x \sin \Phi + \omega_y \cos \Phi)}{\sqrt{2}} - \frac{\omega_z}{\sqrt{2}} \\
 \omega_B &= \frac{(\omega_x \sin \Phi + \omega_y \cos \Phi)}{\sqrt{2}} + \frac{\omega_z}{\sqrt{2}} \\
 \omega_C &= \frac{(\omega_x \sin \Phi - \omega_y \cos \Phi)}{\sqrt{2}} + \frac{\omega_z}{\sqrt{2}} \\
 \text{Gyroscope 2:} \\
 \omega_D &= \frac{(\omega_x \sin \Phi - \omega_y \cos \Phi)}{\sqrt{2}} - \frac{\omega_z}{\sqrt{2}}
 \end{aligned} \tag{9.1}$$

where Φ is the absolute value of the angular displacement between the spin axis of each gyroscope and the body axis x_b . Estimates of the body rates may be derived by summing and differencing the measurements of the turn rates provided by the gyroscopes as shown below, where the $\hat{\omega}$ notation is used to denote an estimated quantity.

$$\begin{aligned}
 \hat{\omega}_x &= \frac{(\omega_A + \omega_B + \omega_C + \omega_D)}{2\sqrt{2} \sin \Phi} \\
 \hat{\omega}_y &= \frac{(\omega_A + \omega_B - \omega_C - \omega_D)}{2\sqrt{2} \cos \Phi} \\
 \hat{\omega}_z &= \frac{(-\omega_A + \omega_B + \omega_C - \omega_D)}{2\sqrt{2}}
 \end{aligned} \tag{9.2}$$

This equation is in fact a least-squares solution to the measurement eqn. (9.1).

The component of ω_x sensed by each gyroscope is equal to ω_x times the cosine of the direction angle, θ , between the body axis x_b and the gyroscope's input axes. For the instrument configuration considered here,

$$\theta = \cos^{-1} \left\{ \frac{\sin \Phi}{\sqrt{2}} \right\} \tag{9.3}$$

If the maximum body rate about the axis x_b is $1200^\circ/\text{s}$ and the maximum rate which can be measured by the gyroscopes is $600^\circ/\text{s}$, then in the absence of any motion about the other axes, the angle θ must be greater than 60° , that is, the angular displacement of the spin axis (Φ) as shown in the Figure 9.3 should not exceed 45° . In general, the value of Φ will need to be less than this figure to cope with turn rates about the other axes of the body.

In order to satisfy a particular set of performance objectives using a skewed sensor configuration of this type, it will be necessary to use higher quality gyroscopes or to compensate the sensors more precisely than would be required for a conventional strapdown arrangement. It can be shown that biases on the measurements of turn rate provided by the sensors and the accuracy of mounting alignment become more critical in inertial systems which use skewed sensor arrangements.

From eqn. (9.2), it can be shown that biases in the four rate measurements, denoted $\delta\omega_A$, $\delta\omega_B$, $\delta\omega_C$ and $\delta\omega_D$, and an error in the skew angle Φ will give rise to biases in

the estimates of the rates about the body axes, $\delta\omega_x$, $\delta\omega_y$ and $\delta\omega_z$, given by

$$\begin{aligned}\delta\omega_x &= \frac{(\delta\omega_A + \delta\omega_B + \delta\omega_C + \delta\omega_D)}{2\sqrt{2}\sin\Phi} - (\omega_A + \omega_B + \omega_C + \omega_D) \frac{\cos\Phi}{2\sqrt{2}\sin^2\Phi} \delta\Phi \\ \delta\omega_y &= \frac{(\delta\omega_A + \delta\omega_B - \delta\omega_C - \delta\omega_D)}{2\sqrt{2}\cos\Phi} + (\omega_A + \omega_B - \omega_C + \omega_D) \frac{\sin\Phi}{2\sqrt{2}\cos^2\Phi} \delta\Phi \\ \delta\omega_z &= \frac{(-\delta\omega_A + \delta\omega_B + \delta\omega_C - \delta\omega_D)}{2\sqrt{2}}\end{aligned}\tag{9.4}$$

with the result that the effects of the measurement biases on the estimates of turn rate in the x or y direction can be magnified through the use of the skewed sensor configuration. The system is also particularly susceptible to misalignment of the sensor mounts. Hence, accurate knowledge of the skew angles is required in order to obtain precise estimates of the body rates. This is particularly true in the situation illustrated in the Figure 9.3, where the sensitive axes of the gyroscopes are displaced by large angles with respect to a potentially high rate axis.

In general, skewed systems based on conventional angular momentum gyroscopes are expected to be applicable in situations where body rates exceed the sensor maximum angular rate capability only transiently. In many applications where high turn rates are likely to be sustained, it is considered that optical rate sensors, such as the ring laser gyroscope or the fibre gyroscope, now offer the best solution because of the high rotation rate capability and excellent scale-factor linearity offered by these types of sensors. The ring laser gyroscope in particular, offers superior scale factor performance.

9.3.4 Redundant sensor configurations

For reasons of safety and reliability, many applications require navigation systems with on-line failure detection and fail-safe operation [1–3]. To satisfy this objective using a strapdown system, additional sensors are required to provide a level of redundancy in the measurements. This may be achieved using an orthogonal sensor arrangement by adding additional sensors to detect the turn rates and accelerations in each vehicle axis. Alternatively, and more commonly, a skewed sensor configuration is often proposed. For instance, a skewed sensor system employing four dual-axis gyroscopes and eight accelerometers may be used to provide quadruplex redundancy for an aircraft's flight control and avionics sensor unit. The gyroscopic input axes are equally distributed on a cone, the axis of which is aligned with the pitch axis of the aircraft for the purposes of this example. The accelerometers may be oriented in a similar manner.

The four separate sources of rate information provided by such a system are indicated in Table 9.1 where ω_{ix} and ω_{iy} are the rates measured about the x and y input axes of the i th gyroscope, and K_1 and K_2 are geometrical constants. For the instrument configuration considered here, $K_1 = 1/\sqrt{2}$ and $K_2 = 1/2$. Similar equations may be written for the acceleration in aircraft body axes. This arrangement

Table 9.1 Sources of rate information provided by a skewed sensor system employing four dual-axis gyroscopes

Sources of information		
Pitch rate	Roll rate	Yaw rate
$K_1(\omega_{1x} + \omega_{1y})$	$K_1(\omega_{1x} - \omega_{1y})$	$K_1(\omega_{3x} - \omega_{3y})$
$K_1(\omega_{2x} + \omega_{2y})$	$(\omega_{2x} - \omega_{2y}) - K_1(\omega_{3x} - \omega_{3y})$	$(\omega_{2x} - \omega_{2y}) - K_1(\omega_{1x} - \omega_{1y})$
$K_1(\omega_{3x} + \omega_{3y})$	$K_2(\omega_{2x} - \omega_{2y}) - K_2(\omega_{4x} - \omega_{4y})$	$(\omega_{4x} - \omega_{4y}) + K_1(\omega_{1x} - \omega_{1y})$
$K_1(\omega_{4x} + \omega_{4y})$	$(\omega_{4x} - \omega_{4y}) + K_1(\omega_{3x} - \omega_{3y})$	$K_2(\omega_{2x} - \omega_{2y}) + K_2(\omega_{4x} - \omega_{4y})$

of sensors, which is discussed in detail in Reference 1, offers high reliability with a ‘fail-operational, fail-safe’ level of fault tolerance. Fail-operational means that a fault must be detected, localised and the system must be dynamically reconfigured. Fail-safe refers to the capability to detect a fault which must not affect system safety. To achieve the same level of redundancy with an orthogonal sensor arrangement would require up to eight two-axis gyroscopes and twelve accelerometers.

The reader interested in redundant strapdown sensor configurations is referred to the many excellent papers on the subject which include References 2 and 3 given at the end of this chapter. Redundant sensor configurations are discussed further in Section 15.4.5 in the context of the Segway machine.

9.4 Instrument electronics

The instrument electronics unit contains the dedicated electronics needed to operate the inertial sensors. Typically, this includes instrument power supplies, read-out electronics to provide signals in the form needed by a navigation processor and possibly a computer. The precise requirements vary in accordance with the types of instruments used and the level of performance which is needed.

For the vast majority of applications, the electronic signals provided by the inertial sensors are required in a digital format for input directly to a computer. Whilst many sensors naturally provide output in digital form, this is not always the case. Where analogue output is provided, this will need to be converted to digital form. The analogue-to-digital conversion process forms part of the instrument electronics.

The output signals from the inertial sensors are often provided in incremental form, that is, as measurements of incremental angle and incremental velocity corresponding to the integral of the measured angular rate and linear acceleration respectively over a short period of time, τ . The incremental angle output ($\delta\theta$) which may be provided by a gyroscope may be expressed mathematically as:

$$\delta\theta = \int_t^{t+\tau} \omega \, dt \quad (9.5)$$

where ω is the measured turn rate. Similarly, the incremental velocity output (δv) from an accelerometer may be written as:

$$\delta v = \int_t^{t+\tau} f \, dt \quad (9.6)$$

where f is the measured acceleration.

This form of sensor output is very convenient since it eases the tasks of updating attitude and velocity. Many contemporary sensors, ring laser gyroscopes for example, naturally provide output signals in this form, whilst for others it is a result of the digitising process carried out within the inertial measurement unit. The way in which the incremental measurements are used is discussed in Chapter 11 together with other aspects of the strapdown processing tasks.

Many conventional sensors such as spinning mass gyroscopes and pendulous accelerometers typically operate in a null seeking or re-balance loop mode in order to achieve a linear and accurate response characteristic. In such cases, instrument re-balance electronics will form part of the instrument electronics block along with gyroscope spin motor and pick-off power supplies.

The use of a computer within the inertial measurement unit (IMU) [4] enables some form of on-line compensation of the instrument outputs to be performed based on instrument characterisation data obtained during laboratory or production testing, as described in Chapter 8. Since such computing tasks are very specific to the type of instrument used, they may well be implemented here rather than as part of the subsequent attitude and navigation processing.

Because instrument characteristics are often temperature dependent, there may also be a need to compensate the instrument outputs for temperature variation in order to achieve satisfactory performance. It follows therefore that instrument temperature monitoring is often required. This is often the subject of some dilemma as to where to monitor the temperatures. In many sensors, the variation in performance with temperature is the result of the temperature sensitivity of magnetic material used in the very core of the sensor. Hence, it may not be adequate to sense temperature outside the instrument cluster or merely close to the case of the instrument.

Finally, it may well be advisable for most applications to carry out some form of on-line testing of the inertial sensors and associated electronics. This may involve checks to confirm that the outputs of the sensors remain within certain known limits appropriate to the application and that they continue to vary in the expected manner whilst operational. For instance, a sensor output remaining at a fixed level for an extended period of time may well suggest a failure has occurred and a warning should be given. Such tasks may also be implemented within the IMU processor which would form part of the built-in test equipment (BITE) within the unit.

It follows from the above remarks that the instrument electronics block could typically comprise the following:

- instrument power supplies;
- re-balance loop electronics;
- temperature monitoring electronics;

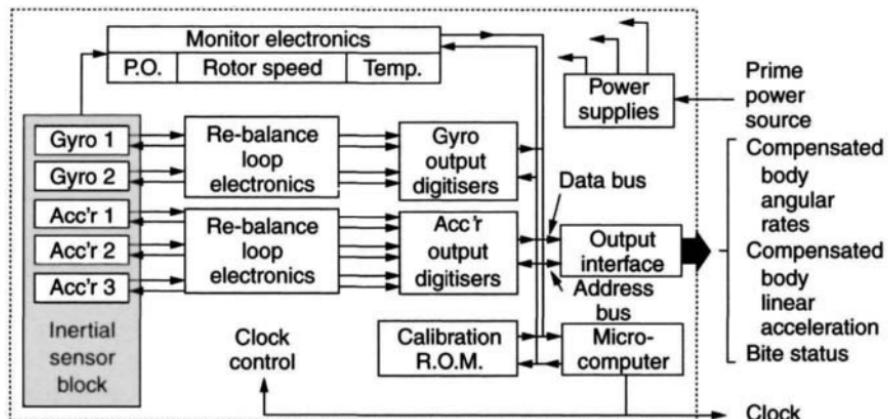


Figure 9.4 Inertial measurement unit functions

- instrument compensation processing;
- analogue-to-digital conversion electronics;
- output interface conditioning;
- built-in test facility.

These components are illustrated schematically in Figure 9.4 which shows an inertial measurement unit containing two dual-axis gyroscopes and three accelerometers.

9.5 The attitude computer

The attitude computer essentially takes the measurements of body rate about three orthogonal axes provided by the inertial measurement unit and uses this information to derive estimates of body attitude by a process of 'integration'. The attitude is usually represented within the computer as a set of direction cosines or quaternion parameters as discussed in Chapter 3, either of which are appropriate for on-line attitude computation. The Euler angle representation described in Chapter 3 is not generally recommended for implementation in strapdown systems. As a result of the preponderance of trigonometric terms in the equations coupled with the presence of singularities for pitch angles of $\pm 90^\circ$, the Euler equations do not lend themselves to real-time solution in an on-board navigation processor. However, it should be borne in mind that there may well be a requirement to extract the Euler angles from the direction cosines or quaternion parameters for control purposes in some applications.

The equations to be solved in the attitude computer are summarised below assuming quaternion parameters are to be used to define the attitude of the vehicle body with respect to the navigation reference frame. The quaternion may be expressed as a four-element vector $[a \ b \ c \ d]^T$, the elements of which are calculated by solving

the following set of differential equations:

$$\begin{aligned}\dot{a} &= -0.5(b\omega_x + c\omega_y + d\omega_z) \\ \dot{b} &= 0.5(a\omega_x - d\omega_y + c\omega_z) \\ \dot{c} &= 0.5(d\omega_x + a\omega_y - b\omega_z) \\ \dot{d} &= -0.5(c\omega_x - b\omega_y - a\omega_z)\end{aligned}\tag{9.7}$$

where ω_x , ω_y and ω_z are estimates of the components of vehicle turn rate with respect to the navigation reference frame. These quantities are computed by differencing the measurements of body rate output by the IMU and estimates of the turn rate of the navigation frame calculated in the navigation computer.

The quaternion parameters may be used to construct the direction cosine matrix which relates the body reference frame to the navigation reference frame (\mathbf{C}_b^n) using:

$$\mathbf{C}_b^n = \begin{pmatrix} (a^2 + b^2 - c^2 - d^2) & 2(bc - ad) & 2(bd + ac) \\ 2(bc + ad) & (a^2 - b^2 + c^2 - d^2) & 2(cd - ab) \\ 2(bd - ac) & 2(cd + ab) & (a^2 - b^2 - c^2 + d^2) \end{pmatrix}\tag{9.8}$$

It is customary in most strapdown attitude computation schemes to carry out self-consistency checks. In the case of the quaternion, the self-consistency check involves confirming that the sum of the squares of the individual quaternion elements remains equal to unity, that is,

$$a^2 + b^2 + c^2 + d^2 = 1\tag{9.9}$$

The attitude computation algorithm used for a given application must be able to keep track of vehicle orientation whilst it is turning at its maximum rate and in the presence of all of the motion of the vehicle, including vibration. Algorithms which may be used to implement the attitude computation function in the presence of such motion are described in Chapter 11.

9.6 The navigation computer

The solution of the navigation equations is carried out in the navigation computer. To implement the navigation function, it is first necessary to transform, or resolve, the specific force measurements provided by the accelerometers, denoted here by the vector \mathbf{f}^b , into the navigation reference frame. This can be accomplished using the attitude information provided by the attitude computer. Using the direction cosine representation of attitude for instance, the required transformation is achieved using:

$$\mathbf{f}^n = \mathbf{C}_b^n \mathbf{f}^b\tag{9.10}$$

where \mathbf{f}^n is the specific force expressed in navigation axes and \mathbf{C}_b^n is the direction cosine matrix described earlier. Both the specific force and the direction cosine matrix are time-varying quantities. Therefore, care must be taken to ensure that all significant movements of the vehicle, including turn rates and vibratory motion, can be accommodated in the computer implementation of this equation.

The resolved specific force components form the inputs to the navigation equations which are used to calculate vehicle velocity and position. The navigation equations are described in Chapter 3 but are repeated here for completeness. For a system which is required to navigate in the vicinity of the Earth to provide estimates of north and east velocity, latitude, longitude and height above the Earth, the equations to be solved may be written as follows:

$$\dot{v}_N = f_N - v_E(2\Omega + \dot{\ell}) \sin L + v_D \dot{L} \quad (9.11)$$

$$\dot{v}_E = f_E + v_N(2\Omega + \dot{\ell}) \sin L + v_D(2\Omega + \dot{\ell}) \cos L \quad (9.12)$$

$$\dot{v}_D = f_D - v_E(2\Omega + \dot{\ell}) \cos L - v_N \dot{L} + g \quad (9.13)$$

$$\dot{L} = \frac{v_N}{R_0 + h} \quad (9.14)$$

$$\dot{\ell} = \frac{v_E \sec L}{R_0 + h} \quad (9.15)$$

$$\dot{h} = -v_D \quad (9.16)$$

where v_N , v_E , v_D , are the north, east and vertical components of vehicle velocity with respect to the Earth, f_N , f_E , f_D , are the components of specific force resolved in the local geographic reference frame, L is the vehicle latitude, ℓ is the vehicle longitude, h is the vehicle height above ground, R_0 is the mean radius of the Earth, Ω is the turn rate of the Earth and g is the acceleration due to gravity.

Refinements to these equations needed to take account of the shape of the Earth and variation in gravitational attraction over the surface of the Earth are given at the end of Chapter 3.

The turn rate of the vehicle with respect to the local geographic navigation frame $\omega_{nb}^b = [\omega_x \ \omega_y \ \omega_z]^T$ which is required to implement the attitude computation process described above is given by:

$$\omega_{nb}^b = \omega_{ib}^b - \mathbf{C}_n^b \omega_{in}^n \quad (9.17)$$

where ω_{ib}^b is the turn rate of the body with respect to inertial frame as measured by the strapdown gyroscopes in the inertial measurement unit and ω_{in}^n is the turn rate of

the navigation frame with respect to the inertial frame, which is computed as follows:

$$\omega_{\text{in}}^n = \begin{pmatrix} \Omega \cos L + \frac{v_E}{R_0 + h} \\ -\frac{v_N}{R_0 + h} \\ -\Omega \sin L - \frac{v_E \tan L}{R_0 + h} \end{pmatrix} \quad (9.18)$$

Algorithms which may be used to implement the navigation function are described in Chapter 11.

9.7 Power conditioning

The raw power supplies available in the host vehicle, whether it is an aircraft, a ship or a land vehicle, will not usually be sufficiently stable or provide the particular voltage levels required by the inertial navigation system. Therefore, it will be necessary to include power conditioning within the unit to generate the supply voltages required which are smoothed sufficiently and controlled to the desired amplitude to ensure satisfactory operation of the navigation system.

9.8 Anti-vibration mounts

A strapdown inertial navigation system will usually be installed on anti-vibration (AV) mounts to provide isolation from vehicle motion at frequencies to which the unit is particularly sensitive. In many applications, the unit may need to be isolated from certain frequencies in the vibration spectrum of the vehicle which may excite resonances within the inertial sensors or give rise to computational errors. The design of suitable AV mounts is frequently a complex task requiring careful matching of the mount design to the characteristics of the inertial sensors within the unit as well as the range and frequency of the perturbing characteristics of the host platform. The effects of vibration are discussed in more detail in Chapter 12 in relation to both instrument errors and overall system performance.

9.9 Concluding remarks

A strapdown inertial navigation system providing navigation in three dimensions will have the following components in one form or another:

- Instrument cluster to sense translational and rotational movements;
- Instrument electronics to provide the control of the sensors and to produce measurement information;
- Attitude computer to compute the attitude of the vehicle for resolution of the specific force measurements;

- Navigation computer to resolve the specific force data and to solve the navigation equations to generate estimates of position and velocity;
- Gravitational model to allow compensation for the effects of gravitational attraction on the translational measurements;
- Power conditioning to provide smoothed and controlled voltage levels needed for satisfactory system operation;
- Input/output interface to communicate with the host vehicle.

These units are mounted in a case which is installed in a vehicle. The case is usually attached to the body of the vehicle via AV mounts.

Reduced configurations are possible to produce an attitude and heading reference unit, or a unit for navigation in a single plane. A photograph of a strapdown navigation system incorporating dynamically tuned gyroscopes is shown in Figure 9.5.

A modern equivalent to this system based on MEMS sensor technology is shown in Figure 9.6.

In a strapdown system, the inertial sensors provide measurements of angular rates and specific force in axes that are usually aligned with the principal body axes of the vehicle. Skewed sensor arrangements may be used in some designs to allow the instruments to cope with very high rates about a single-axis or in systems employing multiple sensors to provide redundancy for fault tolerance purposes. The inertial measurements have to be transformed to the appropriate axis

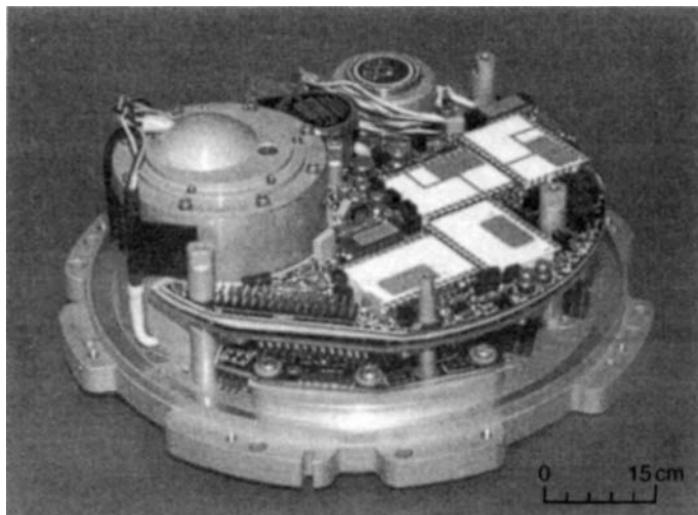


Figure 9.5 Photograph of a strapdown system incorporating dynamically tuned gyroscopes



Figure 9.6 Photograph of a strapdown system incorporating MEMS sensors

set for navigation. A variety of reference frames are used depending on the particular application; typically, a local geographic frame is used to provide estimates of latitude, longitude and height, for navigation in the vicinity of the Earth.

A variety of methods are available for the transformation procedure, direction cosine matrices or quaternion parameters being most commonly used since both are free from singularities at $\pm 90^\circ$ pitch angles. Quaternions are generally preferred because they ensure self-consistency.

The algorithms required for specific force transformation, the correction for the gravitational attraction and the solution of the navigation equations are implemented in the navigation computer. This processor produces the estimates of vehicle velocity and position in whichever axis set the vehicle is using for its navigation.

References

- 1 KROGMANN, U.: 'Optimal integration of inertial sensor functions for flight control and avionics'. AIAA – DASC, San Jose, October 1988
- 2 KROGMANN, U.: 'Design considerations for highly reliable hard- and software fault tolerant inertial reference systems'. DGON proceedings, *Gyro Technology Symposium*, Stuttgart, 1990
- 3 HARRISON, J.V., and GAI, E.G.: 'Evaluating sensor orientations for navigation performance and failure detection', *IEEE Transactions*, 1977, AES-13 (6)
- 4 EDWARDS, C.S., and CHAPLIN, R.J.: 'Strapdown dynamically tuned gyroscopes and the use of microprocessors to simplify their application', DGON proceedings, *Gyro Technology Symposium*, Stuttgart, 1979

Chapter 10

Inertial navigation system alignment

10.1 Introduction

Alignment is the process whereby the orientation of the axes of an inertial navigation system is determined with respect to the reference axis system. The basic concept of aligning an inertial navigation system is quite simple and straight forward. However, there are many complications that make alignment both time consuming and complex. Accurate alignment is crucial, however, if precision navigation is to be achieved over long periods of time without any form of aiding.

In addition to the determination of initial attitude, it is necessary to initialise the velocity and position defined by the navigation system as part of the alignment process. However, since it is the angular alignment which frequently poses the major difficulty, this chapter is devoted largely to this aspect of the alignment process.

In many applications, it is essential to achieve an accurate alignment of an inertial navigation system within a very short period of time. This is particularly true in many military applications, in which a very rapid response time is often a prime requirement in order to achieve a very short, if not zero, reaction time.

There are two fundamental types of alignment process: self-alignment, using gyrocompassing techniques, and the alignment of a slave system with respect to a master reference. There are various systematic and random errors that limit the accuracy to which an inertial navigation system can be aligned, whichever method is used. These include the effects of inertial sensor errors, data latency caused by transmission delays, signal quantisation, vibration effects and other undesirable or unquantifiable motion.

Various techniques have been developed to overcome the effects of the random and systematic errors and enable slave systems in missiles, for example, to be aligned whilst under the wing of an aircraft in-flight, or in the magazine of a ship underway on the ocean. Differing techniques, such as angular rate matching or velocity matching, can be used to align the slave system, the actual circumstances determining the technique which produces the more accurate alignment. In general, a manoeuvre

of the aircraft or ship speeds up the alignment process and increases the accuracy achieved.

The basic principles of alignment on both fixed and moving platforms are described in Section 10.2, whilst the particular problems encountered when aligning on the ground, in the air and at sea are discussed in Sections 10.3, 10.4 and 10.5, respectively.

10.2 Basic principles

The inertial system to be aligned contains an instrument cluster in which the gyroscopes and accelerometers are arranged to provide three axes of angular rate information and three axes of specific force data in three directions, which are usually mutually perpendicular. In a conventional sensor arrangement, the sensitive axes of the gyroscopes are physically aligned with the accelerometer axes. Essentially, the alignment process involves the determination of the orientation of the orthogonal axis set defined by the accelerometer input axes with respect to the designated reference frame.

Ideally, we would like the navigation system to be capable of aligning itself automatically following switch-on, without recourse to any external measurement information. In the situation where the aligning system is mounted in a rigid stationary vehicle, a self-alignment may indeed be carried out based solely on the measurements of specific force and angular rate provided by the inertial system as described in the following section.

10.2.1 *Alignment on a fixed platform*

Consider the situation where it is required to align an inertial navigation system to the local geographic co-ordinate frame defined by the directions of true north and the local vertical. For the purposes of this analysis, it is assumed that the navigation system is stationary with respect to the Earth. In this situation, the accelerometers measure three orthogonal components of the specific force needed to overcome gravity whilst the gyroscopes measure the components of the Earth's turn rate in the same directions.

It is instructive to consider first the alignment of a stabilised platform system in which the instrument cluster can be rotated physically into alignment with the local geographic reference frame. In this situation, it is usual to refer to the accelerometers whose sensitive axes are to be aligned with the north, east and vertical axes of the reference frame as the north, east and vertical accelerometers respectively. Similarly, north, east and vertical gyroscopes may be defined.

In a platform mechanisation, alignment is achieved by adjusting the orientation of the platform until the measured components of specific force and Earth's rate become equal to the expected values. The horizontal components of gravity acting in the north and east directions are nominally zero. The instrument cluster is therefore rotated until the outputs of the north and east accelerometers reach a null, thus levelling the platform. Since the east component of Earth's rate is also known to be

zero, the platform is then rotated about the vertical until the east gyroscope output is nulled, thus achieving an alignment in azimuth. This type of process is referred to as gyrocompassing and is described extensively in the literature [1]. An equivalent alignment process, sometimes referred to as analytic gyrocompassing, can be used to align a strapdown inertial navigation system as described next.

In a strapdown system, attitude information may be stored either as a direction cosine matrix or as a set of quaternion parameters, as described in Chapter 3. The objective of the angular alignment process is to determine the direction cosine matrix or the quaternion parameters which define the relationship between the inertial sensor axes and the local geographic frame. The measurements provided by the inertial sensors in body axes may be resolved into the local geographic frame using the current best estimate of the body attitude with respect to this frame. The resolved sensor measurements are then compared with the expected turn rates and accelerations to enable the direction cosines or quaternion parameters to be calculated correctly. The principles of the method are illustrated below with the aid of single plane examples to show how the attitude of the strapdown inertial sensors with respect to the local geographic reference frame may be extracted from the inertial measurements.

Since the true components of gravity in the north and east directions are nominally zero, any departure from zero in the accelerometer measurements resolved in these directions may be interpreted as an error in the stored attitude data, and in particular as an error in the knowledge of the direction of the local vertical. A single plane illustration is given in Figure 10.1.

The accelerometers provide measurements of the true acceleration in body axes, $-g \sin \theta$ and $-g \cos \theta$ respectively. These measurements are resolved through an angle θ' which is an estimate of the true body angle θ , or the angle that the body makes with the estimated reference frame shown in the figure. It can be seen from the figure that the resolved component in the estimated horizontal plane, denoted g_x , is given by:

$$g_x = -g \sin(\theta - \theta') \quad (10.1)$$

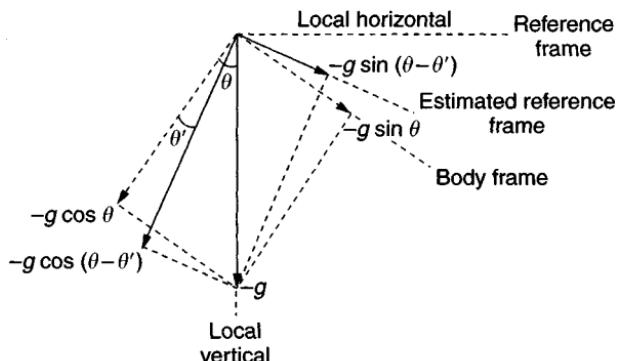


Figure 10.1 Alignment to the gravity vector in a single plane

θ' may be adjusted until g_x becomes zero, at which time $\theta' = \theta$, that is, the estimated body angle becomes equal to the true body angle and the estimated reference frame becomes coincident with the true reference frame.

Given accurate measurements of the specific force acceleration, this process allows the orientation of the axis set defined by the accelerometers with respect to the local vertical to be defined accurately, and is analogous to the process of levelling the stable element in a platform inertial navigation system.

Having defined the local horizontal plane, and so effectively achieved a 'level' in the alignment process, it is then necessary to determine the heading or azimuthal orientation of the inertial instrument frame in the horizontal plane, that is, to determine direction with respect to true north. This is achieved from knowledge of the true components of Earth's rate in the local geographic frame. Assuming that the gyroscopes are of sufficient precision to detect Earth's rate accurately, the stored attitude information is now adjusted until the resolved component of the measured rate in the east direction reduces to zero. A diagram illustrating the alignment in azimuth is shown in Figure 10.2.

In this case, ψ is the true orientation of the x -axis of the instrument frame with respect to true north whilst ψ' is the estimate of that quantity. The components of Earth's rate (Ω) detected by the x - and y -axis gyroscopes shown in the figure are $\Omega \cos L \cos \psi$ and $\Omega \cos L \sin \psi$, respectively, where L is the latitude of the aligning system. The east component of Earth's rate as determined by the navigation system, denoted ω_E , may be expressed as follows:

$$\omega_E = \Omega \cos L \sin(\psi - \psi') \quad (10.2)$$

ψ' is adjusted until ω_E becomes zero, in which case $\psi' = \psi$.

10.2.2 Alignment on a moving platform

In order to align a strapdown inertial navigation system in a moving vehicle, a technique which is similar in principle to that described above may be used.

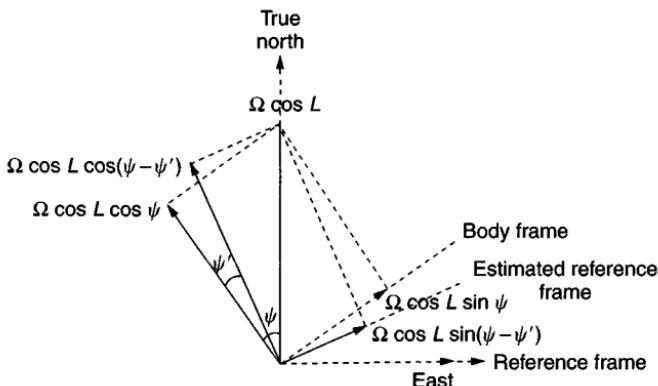


Figure 10.2 Alignment in azimuth

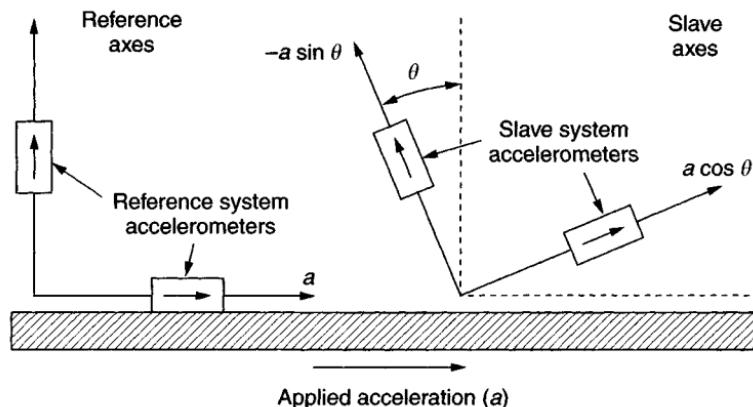


Figure 10.3 Measurement matching alignment in a single plane

However, when aligning in a moving vehicle, the accelerations and turn rates to which the system is subjected are no longer well defined in the way that they are when the system is stationary. It therefore becomes necessary to provide some independent measure of these quantities against which the measurements generated by the aligning system may be compared.

Consider the situation depicted in Figure 10.3 in which the axes defined by the strapdown sensors are shown rotated through an angle θ in a single plane with respect to the navigation reference frame.

If the acceleration of the vehicle in the reference x -direction is a , then the accelerations sensed by the strapdown system accelerometers will be as follows:

$$\begin{aligned} a_x &= a \cos \theta \\ a_y &= -a \sin \theta \end{aligned} \tag{10.3}$$

In the absence of any instrument measurement inaccuracies, alignment of the strapdown system may be achieved by resolving the accelerometer measurements through an angle θ' and adjusting its magnitude using a feedback process so as to null the difference between the resolved components of the slave system measurements and the accelerations measured by the reference system.

Mathematically, θ' is adjusted to allow the following relationships to be satisfied:

$$\begin{aligned} a_x \cos \theta' - a_y \sin \theta' &= a \\ a_x \sin \theta' + a_y \cos \theta' &= 0 \end{aligned} \tag{10.4}$$

Substituting for a_x and a_y from eqn. (10.3) yields:

$$\begin{aligned} a \cos(\theta - \theta') &= a \\ a \sin(\theta - \theta') &= 0 \end{aligned} \tag{10.5}$$

It can be seen that these relationships will be satisfied when $\theta' = \theta$.

Therefore, it is possible to determine the orientation of the strapdown sensors by comparing the accelerometer measurements resolved into the reference frame with independent measurements of these same quantities. An estimate of θ can also be derived in a similar manner by comparing angular rate measurements. Whichever method is adopted, it will be noted that alignment about a given axis is dependent on the measurement of an acceleration or turn rate taking place along or about an axis which is orthogonal to the axis in which the misalignment exists.

As an alternative to the type of procedure described above, alignment may be achieved by comparing estimates of velocity or position generated by the strapdown system with similar estimates provided by an external source over a period of time. Velocity and position errors will propagate with time as a result of the angular alignment errors. Therefore, any difference in the velocity and position estimates generated between the aligning system and the external source over this time will be partially the result of an alignment error. Such methods are discussed in more detail below in the context of in-flight and shipboard alignment.

With aircraft and shipboard systems, the independent measurement information may be provided by a separate inertial navigation system on-board the same vehicle. By comparing the two sets of inertial measurements it is possible to deduce the relative orientation of the two frames on a 'continuous' basis. The precise measurements available will be dependent on the reference system mechanisation on-board the ship or aircraft. As a rule, a stable platform navigation system will only output estimates of position, velocity, attitude and heading. A strapdown reference system offers greater flexibility, potentially providing linear acceleration and angular rate information in addition to the usual navigation outputs listed above. Alternatively, position fixes may be derived on-board the vehicle from signals transmitted by a radio beacon or from satellites.

10.3 Alignment on the ground

10.3.1 Introduction

Attention is now turned to the alignment of an inertial navigation system in a ground based vehicle. Clearly, the scope for carrying out manoeuvres or applying motion to aid the process of alignment is very limited in such applications. Attention is focused here on a requirement which often arises in practice, that of determining the orientation of a set of sensor axes with respect to the local geographic frame. For convenience, the local geographic axis set is often chosen to be the reference frame.

In the past, a site survey would be carried out to establish a north line. Heading information would then be transferred to the aligning navigation system using theodolites and a prism attached to the aligning system. Although high accuracy can be obtained using this approach, it is both time consuming and labour intensive. The methods discussed in the following sections are usually more convenient to implement and avoid such problems.

10.3.2 Ground alignment methods

In principle, the techniques outlined in Section 10.2 for the self-alignment of a strapdown inertial system on a stationary platform can be used. We now look in more detail at the computation required to implement that alignment process. As described above, the objective of the angular alignment process is to determine the direction cosine matrix, \mathbf{C}_n^b , or its quaternion equivalent, which relates the body and geographic reference frames. The body mounted sensors will measure components of the specific force needed to overcome gravity and components of Earth's rate, denoted by the vector quantities \mathbf{g}^b and $\boldsymbol{\omega}_{ie}^b$, respectively. These vectors are related to the gravity and Earth's rate vectors specified in the local geographic frame, \mathbf{g}^n and $\boldsymbol{\omega}_{ie}^n$, respectively, in accordance with the following equations:

$$\mathbf{g}^b = \mathbf{C}_n^b \mathbf{g}^n \quad (10.6)$$

$$\boldsymbol{\omega}_{ie}^b = \mathbf{C}_n^b \boldsymbol{\omega}_{ie}^n \quad (10.7)$$

where $\mathbf{g}^n = [0 \ 0 \ -g]^T$ and $\boldsymbol{\omega}_{ie}^n = [\Omega \cos L \ 0 \ -\Omega \sin L]^T$ in which Ω and L denote Earth's rate and latitude, respectively. Given knowledge of these quantities, estimates of the elements of the direction cosine matrix may be computed directly from the measurements of $\mathbf{g}^b = [g_x \ g_y \ g_z]^T$ and $\boldsymbol{\omega}_{ie}^b = [\omega_x \ \omega_y \ \omega_z]^T$ as follows:

$$\begin{aligned} c_{31} &= -\frac{g_x}{g} & c_{11} &= \frac{\omega_x}{\Omega \cos L} - \frac{g_x \tan L}{g} \\ c_{32} &= -\frac{g_y}{g} & c_{12} &= \frac{\omega_y}{\Omega \cos L} - \frac{g_y \tan L}{g} \\ c_{33} &= -\frac{g_z}{g} & c_{13} &= \frac{\omega_z}{\Omega \cos L} - \frac{g_z \tan L}{g} \end{aligned} \quad (10.8)$$

where $c_{11}, c_{12}, \dots, c_{33}$ are elements of the direction cosine matrix \mathbf{C}_n^b . The remaining direction cosine elements (c_{21}, c_{22} and c_{23}) may be determined by making use of the orthogonality properties of the direction cosine matrix which yield:

$$\begin{aligned} c_{21} &= -c_{12}c_{33} + c_{13}c_{32} \\ c_{22} &= c_{11}c_{33} - c_{31}c_{13} \\ c_{23} &= -c_{11}c_{32} + c_{31}c_{12} \end{aligned} \quad (10.9)$$

It can be seen from the above equations that the direction cosine matrix is uniquely defined provided that L is not equal to $\pm 90^\circ$, that is, there is a unique value so long as the aligning system is not located at either the north or south poles of the Earth. This clearly would lead to singularities in the equations for some of the direction cosine elements which therefore become indeterminate. However, over much of the Earth's surface, a single set of inertial measurements can provide all of the information needed to compute the direction cosine matrix, and so achieve a strapdown system alignment.

The accuracy with which such an alignment can be accomplished is largely determined by the precision of the available measurements and the resolution of the

instrument outputs. As a result of instrument biases, the above procedure will yield an estimate of the direction cosine matrix $\tilde{\mathbf{C}}_b^n$ which will be in error. As described in Chapter 11, $\tilde{\mathbf{C}}_b^n$ may be expressed as the product of the true matrix \mathbf{C}_b^n and a matrix \mathbf{B} which represents the misalignment between the actual and computed geographic frames:

$$\tilde{\mathbf{C}}_b^n = \mathbf{B} \mathbf{C}_b^n \quad (10.10)$$

For small angular misalignments, this can be written in skew symmetric form as:

$$\mathbf{B} = \mathbf{I} - \boldsymbol{\Psi} \quad (10.11)$$

where \mathbf{I} is a 3×3 identity matrix and

$$\boldsymbol{\Psi} = \begin{pmatrix} 1 & -\delta\gamma & \delta\beta \\ \delta\gamma & 1 & -\delta\alpha \\ -\delta\beta & \delta\alpha & 1 \end{pmatrix} \quad (10.12)$$

$\delta\alpha$, $\delta\beta$ and $\delta\gamma$ are the misalignments about the north, east and vertical axes of the geographic frame, respectively, and are equivalent to the physical misalignments of the instrument cluster in a stable platform navigation system. The ‘tilt’ errors ($\delta\alpha$ and $\delta\beta$) which result, are predominantly determined by the accelerometer biases while the azimuth or heading error ($\delta\gamma$) is a function of gyroscopic bias as described in the following section.

The direction cosine matrix, $\tilde{\mathbf{C}}_b^n$, is adjusted through the alignment process until the residual north and east components of accelerometer bias are off-set by components of g in each of these directions, effectively nulling the estimates of acceleration in these directions. The resulting attitude errors correspond to the ‘tilt’ errors which arise when aligning a stable platform system. In azimuth, the platform rotates about the vertical to a position where a component of the Earth’s horizontal rate ($\Omega \cos L$) appears about the east axis to null the east gyroscopic bias. An equivalent process takes place in a strapdown system, again through appropriate adjustment of the direction cosine matrix.

The resulting attitude and heading errors may be expressed as follows for the particular situation in which the body frame is nominally aligned with the geographic frame, that is, where $\mathbf{C}_b^n = \mathbf{I}$, it can be shown that:

$$\begin{aligned} \delta\alpha &= \frac{B_y}{g} \\ \delta\beta &= -\frac{B_x}{g} \\ \delta\gamma &= \frac{D_y}{\Omega \cos L} + \frac{B_y \tan L}{g} \end{aligned} \quad (10.13)$$

More generally, where the system is not aligned with the geographic frame, the sensor biases arising in each of the above equations will be made up of a linear combination of the biases in all three gyroscopes or all three accelerometers.

10.3.2.1 Derivation of azimuth error, $\delta\gamma$

The angular rates sensed about the x -, y - and z -axes may be expressed in vector form, as the sum of the Earth's rate components in each axis and the residual gyroscope biases, as follows:

$$\begin{bmatrix} \omega_x \\ \omega_y \\ \omega_z \end{bmatrix} = \begin{bmatrix} 1 & -\delta\gamma & \delta\beta \\ \delta\gamma & 1 & -\delta\alpha \\ -\delta\beta & \delta\alpha & 1 \end{bmatrix} \begin{bmatrix} \Omega \cos L \\ 0 \\ -\Omega \sin L \end{bmatrix} + \begin{bmatrix} D_x \\ D_y \\ D_z \end{bmatrix}$$

The process of gyrocompassing acts to null the east component of measured angular rate; the ω_y term:

$$\omega_y = \delta\gamma\Omega \cos L + \delta\alpha\Omega \sin L + D_y = 0$$

Substituting for $\delta\alpha$ from eqn. (10.13) and rearranging yields,

$$\delta\gamma = \frac{D_y}{\Omega \cos L} + \frac{B_y \tan L}{g}$$

as given above.

The azimuth misalignment term ($\delta\gamma$) contains two components; the first being the result of a residual gyroscopic bias acting in the east direction, the second term being the result of a level or tilt error about the north axis causing a component of vertical Earth's rate ($\delta\alpha\Omega \sin L$) to appear as a further bias about the east axis.

It can be shown using eqn. (10.13) that a 1 milli- g accelerometer bias will give rise to a level error of 1 mrad (~ 3.4 arc min) whilst a gyroscopic drift of $0.01^\circ/\text{h}$ will result in an azimuthal alignment error of 1 mrad at a latitude of 45° . The relationship between gyroscope bias and azimuthal error is illustrated graphically in Figure 10.4. It is clear that good quality gyroscopes are needed to achieve an accurate alignment in azimuth. It is noted that for some inertial system applications, it is the alignment requirements which can dictate the specification of the inertial sensors rather than the way in which the sensor errors propagate during navigation.

The alignment method as described here, using a single set of instrument measurements, would allow only a coarse alignment to take place. To achieve a more accurate estimate of the direction cosine matrix, sequential measurements would be used to carry out a self-alignment over a period of time. Some Kalman filtering of the measurement data would normally be applied under these circumstances.

In addition to the alignment error mechanisms described above, errors in azimuth also arise as a result of gyroscopic random noise (n) and accelerometer bias instability (b). Noise on the output of the gyroscopes (random walk in angle), which is of particular concern in systems using mechanically dithered ring laser gyroscopes, gives rise to a root mean square azimuth alignment error which is inversely proportional to the square root of the alignment time (t_a), viz. $\delta\gamma = n/\Omega \cos L \sqrt{t_a}$. Therefore, given a random walk error of $0.005^\circ/\sqrt{\text{h}}$, an alignment accuracy of 1 mrad can be achieved at a latitude of 45° in a period of 15 min. The effect of this noise can be reduced by extending the alignment period, that is, extending the time over which the noise is filtered. Small changes in the north component of accelerometer bias (b) with time are equivalent to an east gyroscope drift. Therefore such errors can also

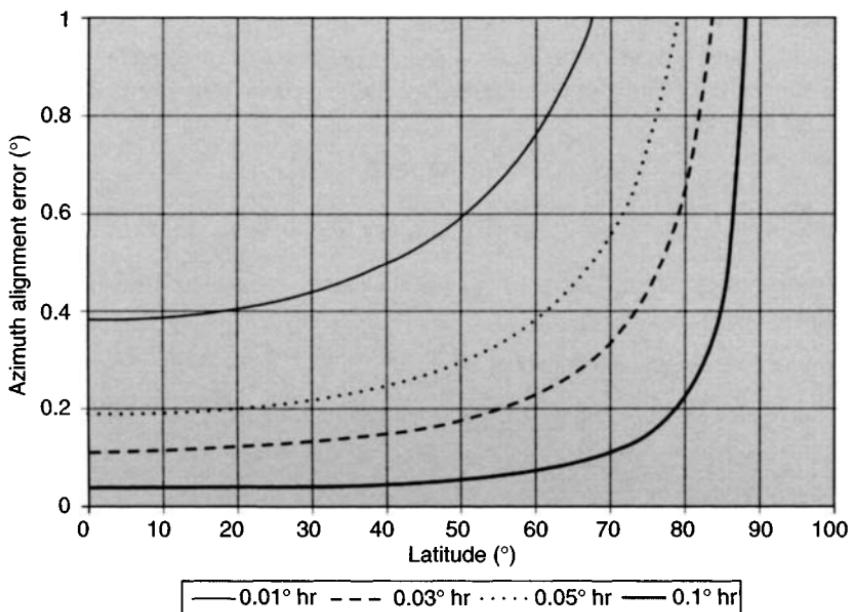


Figure 10.4 Azimuth alignment error versus latitude as a function of residual east gyroscope bias

introduce an azimuth alignment error which may be expressed as $\delta\gamma = b/g\Omega \cos L$. A bias drift of 1 micro-g/s will result in an alignment error of 20 mrad at a latitude of 45°. The minimisation of bias shifts with temperature as well as switch-on transients is vital for applications where this effect becomes significant.

10.3.2.2 Vehicle perturbations

A process very similar to that described above may be adopted to align an inertial navigation system mounted in a vehicle which is not perfectly stationary, but subjected to disturbances. For instance, it may be required to align a navigation system in an aircraft on a runway preparing for take-off which is being buffeted by the wind and perturbed by engine vibration. In such a situation, the mean attitude of the aligning system with respect to the local geographic frame is fixed, and the specific force and turn rates to which the aligning system is subjected are nominally fixed. In this situation, some form of base motion isolation is needed to allow the alignment errors to be deduced from the measurements of turn rate and specific force provided by the sensors [1].

A self-alignment may be carried out in the presence of the small perturbations using a Kalman filter incorporating a model of the base motion disturbance. Failure to take account of any filter measurement differences caused by the disturbances will result in an incorrect alignment, since the measurements of the disturbance will be interpreted incorrectly as resulting from alignment errors. The application of Kalman filtering techniques for the alignment of strapdown inertial navigation systems is

discussed more fully in Sections 10.4 and 10.5 in relation to the alignment of such systems in-flight and at sea.

10.3.3 Northfinding techniques

In view of the limitations of both of the aforementioned techniques, various designs for special purpose equipment, which would allow the directions of the local vertical and true north to be defined within a land-based vehicle, have been produced. Such devices, often referred to as northfinders, are designed with a view to establishing the direction of true north within a short period of time using relatively inexpensive inertial sensors.

One possible mechanisation uses measurements of two orthogonal components of Earth's rate to establish a bearing angle of a pre-defined case reference axis with respect to north. The sensing element is a two-degrees-of-freedom gyroscope such as a dynamically tuned gyroscope (DTG) with its spin axis vertical. The DTG assembly is suspended by a wire to provide automatic levelling of the two input axes which are at right angles to one another. Hence, the input axes are maintained in the horizontal plane. The input axes are held in a torque re-balance loop to provide measurements of the rate of turn about each axis. The pendulous assembly is enclosed within a container which is filled with a fluid to provide damping.

In this configuration, the gyroscope measures two horizontal components of the Earth's rotation rate as indicated in Figure 10.5.

The angular rates (ω_x and ω_y) measured about the two input axes of the gyroscope may be expressed as follows:

$$\begin{aligned}\omega_x &= \Omega \cos L \cos \psi \\ \omega_y &= \Omega \cos L \sin \psi\end{aligned}\tag{10.14}$$

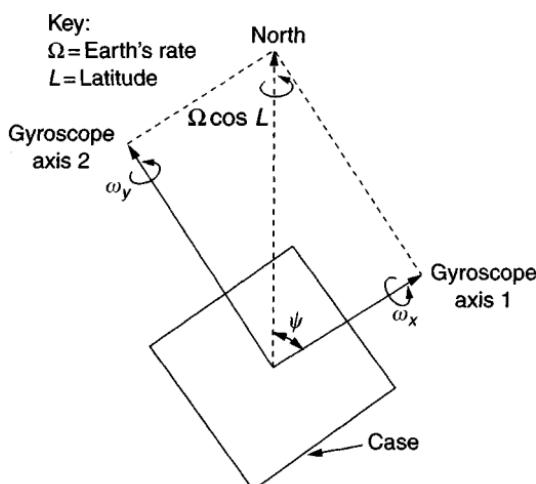


Figure 10.5 A northfinder

where Ω is the Earth's rate, L is the latitude and ψ is the heading of gyroscope axis with respect to true north.

By taking the ratio of the two independent gyroscopic measurements, the latitude dependent terms cancel, allowing the gyroscope heading angle, ψ , to be computed.

$$\frac{\omega_y}{\omega_x} = \frac{\Omega \cos L \sin \psi}{\Omega \cos L \cos \psi} = \tan \psi$$

$$\psi = \arctan \left(\frac{\omega_y}{\omega_x} \right) \quad (10.15)$$

Heading can be calculated in this way provided $\omega_x \neq 0$. In the event that ω_x is close to zero, the following equation may be used:

$$\psi = 90 - \arctan \left(\frac{\omega_x}{\omega_y} \right) \quad (10.16)$$

It can be seen that the northfinder does not require knowledge of latitude, or prior orientation in any particular direction, to enable a measure of heading to be obtained.

In order to achieve useful accuracy from a device of this type, gyroscope measurement accuracy of $0.005^\circ/\text{h}$ or better may be required. However, the need for a highly accurate gyroscope may be avoided by rotating the entire sensor assembly through 180° about the vertical, without switching off, and then taking a second pair of measurements in this new orientation. The measurements obtained in each position are then differenced, allowing any biases on the measurements to be largely eliminated. The heading angle is then computed from the ratio of the measurement differences. This process is identical to the 'indexing technique' used in inertial systems to enhance accuracy.

The rotation of the sensor may be accomplished using a small d.c. motor to drive the assembly from one mechanical stop to another which are nominally 180° apart. The stops are positioned so that the gyroscope input axes are aligned with the case reference axis, or at right angles to it, when the measurements are taken. Over the short period of time required to rotate the sensor (typically 5 s) and to take these measurements, all but the gyroscope in-run random measurement errors can be removed. This technique also helps to reduce any errors arising through the sensitive axes of the gyroscope not being perfectly horizontal.

There are a number of variations of this method, one of which involves positioning the gyroscope with one of its input axes vertical and the spin axis in the horizontal plane. Two measurements of the horizontal component of Earth's rate are taken with the gyroscope in two separate orientations 90° apart. An estimate of heading can then be obtained from the ratio of these two measurements in the manner described above. This scheme allows the Northfinder to be used as a directional gyroscope after the heading angle has been determined. Other variations incorporate accelerometers to allow the inclination to the vertical to be determined as well as heading.

10.4 In-flight alignment

10.4.1 Introduction

The requirement frequently arises to align an inertial navigation system in an air-launched missile prior to missile release from an aircraft platform. A convenient reference for this purpose may be provided by the aircraft's own inertial navigation system. Such an alignment of the missile system may therefore be achieved by the transfer of data from the aircraft's navigation system to the missile by a process known as transfer alignment. This may be achieved quite simply by the direct copying of data from the aircraft to the missile navigation system, or more precisely by using some form of inertial measurement matching process of the type outlined in Section 10.2.2. Alternatively, the missile inertial navigation system may be aligned in-flight using position fixes provided by satellite or airborne radar systems. All such methods are discussed below, but with particular emphasis on the use of transfer alignment.

It is noted that it is sometimes neither desirable nor possible to have the inertial system in a guided missile 'run-up' and aligned waiting for the launch command. In this situation, it is required to align the missile's inertial navigation system very rapidly, immediately prior to launch of the missile.

10.4.2 Sources of error

As a result of physical misalignments between different mounting locations on an aircraft, the accuracy with which inertial data can be transferred from one location to another on-board the aircraft will be restricted. Such errors may be categorised in terms of static and dynamic components as follows:

Static errors will exist as a result of manufacturing tolerances and imprecise installation of equipment leading to mounting misalignments between different items of equipment on the aircraft.

Dynamic errors will exist because the airframe will not be perfectly rigid and will bend in response to the aerodynamic loading on the wings and launch rails to which a missile is attached. Such effects become particularly significant in the presence of aircraft manoeuvres. Significant error contributions can also be expected to arise as a result of vibration.

Methods of alleviating such problems are discussed in the following section.

10.4.3 In-flight alignment methods

Attention is focused here on the alignment of an inertial navigation system contained in an air-launched missile which may be attached to a fuselage or wing pylon beneath a 'carrier' aircraft.

10.4.3.1 'One-shot' transfer alignment

One of the simplest alignment techniques which may be adopted in this situation is to copy position, velocity and attitude data from the aircraft's own navigation system

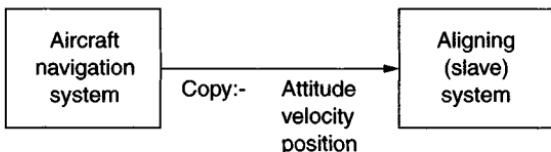


Figure 10.6 ‘One-shot’ transfer alignment

directly to the missile system. This is sometimes referred to as a ‘one-shot’ alignment process and is depicted in Figure 10.6.

Clearly, any angular displacement between the aircraft and missile systems which exists at the instant when the data are transferred will appear as an alignment error in the missile’s navigation system. Therefore, the success of such a scheme is reliant on the two systems being physically harmonised to high accuracy, or on accurate knowledge of their relative orientation being available when the alignment takes place. In the latter situation, the data from the aircraft’s navigation system may be resolved accurately in missile axes before being passed to the missile navigation system.

In general, the precise harmonisation of one system with respect to the other will not be known, for the reasons outlined in the previous section. Furthermore, the aircraft navigation system will be positioned some distance from the aligning system in the missile and there will be relative motion between them should the aircraft turn or manoeuvre; the so-called lever-arm motion. In this situation, the velocity information passed to the missile will be in error. As a result, the accuracy of alignment which can be achieved using a ‘one-shot’ alignment procedure will be extremely limited and more precise methods are usually sought.

10.4.3.2 Airborne inertial measurement matching

An alternative method of transfer alignment, which has received much attention in recent years [2–5], is that of inertial measurement matching. This technique relies on the comparison of measurements of applied motion obtained from the two systems to compute the relative orientation of their reference axes, as introduced in the discussion of basic principles in Section 10.2 and depicted in Figure 10.7. An initial coarse alignment may be achieved by the ‘one-shot’ process, discussed earlier, before initiating the measurement matching process which is described below.

In theory, a transfer alignment between two inertial navigation systems on an aircraft can be achieved most rapidly by comparing measurements generated by the aircraft system and the missile system of the fundamental navigation quantities of specific force acceleration and angular rate, resolved into a common co-ordinate frame. In the absence of measurement errors, and assuming the two systems are mounted side by side on a perfectly rigid platform, the measurement differences arise purely as a result of alignment errors. Under such conditions, it is possible to identify accurately the misalignments between the two systems.

In practice, this approach is often impractical for a number of reasons. The reference system may use ‘platform’ technology, in which linear acceleration and turn

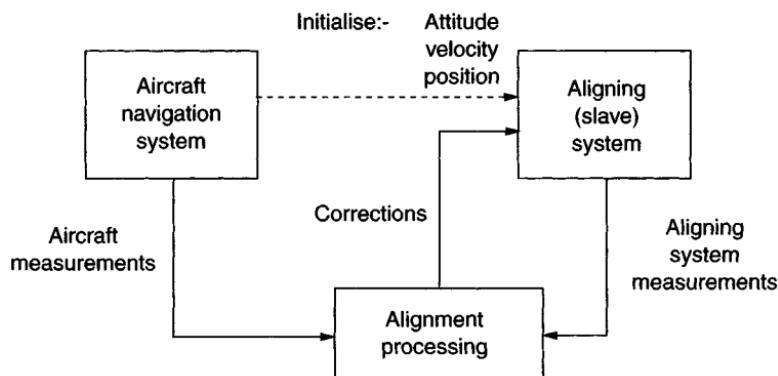


Figure 10.7 Inertial measurement matching alignment scheme

rate data are not standard outputs. This is particularly true in the case of many older military aircraft, although the situation has changed with the wider use of strapdown technology in modern combat aircraft inertial navigation systems.

There are also technical reasons which may preclude the use of linear acceleration and angular rate matching procedures as a viable option for airborne transfer alignment. This is particularly true where the physical separation between the reference and aligning system is large, and where significant flexure motion is present. The turn rates and linear accelerations sensed by the reference and aligning systems will differ as a result of the flexure motion which is present. These differences will then be interpreted incorrectly as errors in the stored attitude data, and so degrade the accuracy of alignment which can be achieved. Acceleration matching and angular rate matching are particularly sensitive to the effects of flexure. Whilst it is possible theoretically to model the flexural motion, and thus separate the components of the measurement differences caused by flexure from those attributable to alignment errors, adequate models of such motion are rarely available in practice.

Even when attempting to carry out an alignment on a perfectly rigid airframe, the translational motion sensed at the reference and the aligning system locations will differ, as the aircraft rotates, as a result of lever-arm motion. The measurement differences which arise as a result of lever-arm motion as the aircraft manoeuvres will also be interpreted incorrectly as alignment inaccuracies and therefore inhibit the alignment process. These additional measurement differences are functions of aircraft turn rate, angular acceleration and the physical separation between the two systems. Whilst it is theoretically possible to correct one set of measurements before comparison with the other, such corrections are dependent on the availability of sufficiently precise estimates of these quantities. Although it is reasonable to assume that distance would be known to sufficient accuracy and the actual turn rates may be provided directly by a strapdown system, angular acceleration measurements are not usually available and without the use of angular accelerometers are not easy to estimate.

For the reasons outlined above, acceleration and rate matching are not generally recommended for alignment of inertial systems on-board aircraft, even when both the reference and aligning systems are configured in a strapdown form. An alternative approach is the use of velocity matching described in Section 10.4.3.3. Velocity errors propagate in an inertial navigation system as a result of alignment inaccuracy, as well as through inertial instrument imperfections. By comparing the velocity estimates provided by the reference and aligning systems, it may therefore be possible to obtain estimates of the alignment errors and, under some circumstances, estimates of the sensor biases. Hence, it is possible to achieve a measure of sensor calibration as part of the same process.

Because of the smoothing effect of the integration process which takes place between the raw measurements from the instruments and the velocity estimates within an inertial navigation system, the effects of flexure and sensor noise on the process of alignment is much less severe than experienced with acceleration matching. Further, it has the advantage of allowing lever-arm corrections to be implemented more easily, such corrections at the 'velocity' level being purely functions of turn rate and separation distance.

10.4.3.3 Velocity matching alignment

As suggested in the preceding section, an in-flight alignment may be achieved by comparing estimates of velocity generated by the aligning system with estimates of the same quantities provided by the aircraft's own navigation system. The nature of the alignment problem, which involves the identification of a number of interrelated and time varying error sources using measurements which are corrupted with noise, is well suited to statistical modelling techniques. These techniques include Kalman filtering, the principles of which are discussed in Appendix A.

This section outlines the system and measurement equations required to construct a Kalman filter which may be used to process the velocity information and so obtain estimates of the alignment errors. For the purposes of this Kalman filter illustration, a number of simplifying assumptions have been made in the formulation given here and these are described below.

The system equations

It is required to determine accurately the attitude and velocity of the aligning system with respect to a designated reference frame. Typically, this may be a body fixed axis set within the aircraft or the local geographic navigation frame. The aligning system and reference frames are denoted here by the superscripts and subscripts b and n, respectively. Following the notation used in Chapter 3, the propagation of the direction cosine matrix (C_b^n) which relates the sensor axes of the aligning system to the reference frame is governed by the following differential equation:

$$\dot{C}_b^n = C_b^n \Omega_{nb}^b \quad (10.17)$$

where Ω_{nb}^b is a skew symmetric matrix formed from the turn rates of the aligning system with respect to the reference frame. This turn rate is obtained by differencing

the angular rates sensed by the aligning system (ω_{ib}^b) and the turn rate of the reference frame (ω_{in}^n). An estimate of the direction cosine matrix, denoted \hat{C}_b^n , is calculated using measurements of the turn rate to which the aligning system is subjected ($\hat{\omega}_{ib}^b$) and an estimate of the reference frame rate ($\hat{\omega}_{in}^n$) to determine $\hat{\Omega}_{nb}^b$, updating from some initial estimate using:

$$\dot{\hat{C}}_b^n = \hat{C}_b^n \hat{\Omega}_{nb}^b \quad (10.18)$$

As described in Section 10.3.2, for small angle misalignments, the true and estimated direction cosine matrices may be related by the equation:

$$\hat{C}_b^n = [\mathbf{I} - \Psi] C_b^n \quad (10.19)$$

where \mathbf{I} is the identity matrix and Ψ is a skew symmetric matrix which may be written as:

$$\Psi = \begin{pmatrix} 0 & -\delta\gamma & \delta\beta \\ \delta\gamma & 0 & -\delta\alpha \\ -\delta\beta & \delta\alpha & 0 \end{pmatrix}$$

in which the off-diagonal elements $\delta\alpha$, $\delta\beta$ and $\delta\gamma$ represent the attitude errors in the aligning system.

It can be shown that the attitude errors propagate according to:

$$\dot{\Psi} = -\omega_{in}^n \times \Psi - C_b^n \delta\omega_{ib}^b + \delta\omega_{in}^n \quad (10.20)$$

where $\Psi = [\delta\alpha \ \delta\beta \ \delta\gamma]^T$, is the alignment error vector; $\delta\omega_{ib}^b = (\hat{\omega}_{ib}^b - \omega_{ib}^b)$ is the gyroscopic measurement error in the aligning system; $\delta\omega_{in}^n = (\hat{\omega}_{in}^n - \omega_{in}^n)$ is the error in the reference frame rate estimates and \times denotes the cross product of two vector quantities.

For the purposes of this example Kalman filter formulation, the gyroscopic errors are modelled in the filter as additive Gaussian white noise and the reference rate errors are assumed to be zero. The derivation of this equation is given in Chapter 12 where the propagation of errors in strapdown inertial navigation systems is discussed in greater detail.

The velocity equations may be expressed approximately as:

$$\dot{\mathbf{v}}^n = C_b^n \mathbf{f}^b - \mathbf{g} \quad (10.21)$$

where \mathbf{v}^n is the velocity of the aircraft, \mathbf{f}^b is the specific force sensed by the accelerometers in the aligning system in body axes and \mathbf{g} is the local gravity vector. The propagation of the errors in the estimates of velocity computed by the aligning system ($\delta\mathbf{v}^n$) may be expressed as:

$$\dot{\delta\mathbf{v}}^n = \mathbf{f}^n \times \Psi + C_b^n \delta\mathbf{f}^b \quad (10.22)$$

where \mathbf{f}^n is the specific force measured by the aligning system resolved in reference axes and $\delta\mathbf{f}^b$ represents the errors in the accelerometer measurements. This is modelled in the Kalman filter as additive Gaussian white noise.

Equations (10.20) and (10.22) may be combined and expressed in state space form as:

$$\delta\dot{\mathbf{x}} = \mathbf{F}\delta\mathbf{x} + \mathbf{Gw} \quad (10.23)$$

where $\delta\mathbf{x}$ is the error state vector, \mathbf{F} is the system error matrix, \mathbf{G} is the noise input matrix and \mathbf{w} is the system noise which represents the instrument noise together with any unmodelled biases. The error state vector may be expressed in component form as:

$$\delta\mathbf{x} = [\delta\alpha \quad \delta\beta \quad \delta\gamma \quad \delta v_N \quad \delta v_E]^T \quad (10.24)$$

where $\delta\alpha$, $\delta\beta$, $\delta\gamma$ are the components of the vector Ψ , the attitude errors; and δv_N , δv_E are the north and east velocity errors, respectively.

The error equation may be expressed in full as follows:

$$\begin{aligned} \begin{pmatrix} \dot{\delta\alpha} \\ \dot{\delta\beta} \\ \dot{\delta\gamma} \\ \dot{\delta v_N} \\ \dot{\delta v_E} \end{pmatrix} &= \begin{pmatrix} 0 & \omega_D & -\omega_E & 0 & 0 \\ -\omega_D & 0 & \omega_N & 0 & 0 \\ \omega_E & -\omega_N & 0 & 0 & 0 \\ 0 & -f_D & f_E & 0 & 0 \\ f_D & 0 & -f_N & 0 & 0 \end{pmatrix} \begin{pmatrix} \delta\alpha \\ \delta\beta \\ \delta\gamma \\ \delta v_N \\ \delta v_E \end{pmatrix} \\ &+ \begin{pmatrix} -c_{11} & -c_{12} & -c_{13} & 0 & 0 & 0 \\ -c_{21} & -c_{22} & -c_{23} & 0 & 0 & 0 \\ -c_{31} & -c_{32} & -c_{33} & 0 & 0 & 0 \\ 0 & 0 & 0 & c_{11} & c_{12} & c_{13} \\ 0 & 0 & 0 & c_{21} & c_{22} & c_{23} \end{pmatrix} \begin{pmatrix} w_{gx} \\ w_{gy} \\ w_{gz} \\ w_{ax} \\ w_{ay} \\ w_{az} \end{pmatrix} \end{aligned} \quad (10.25)$$

where

$$\omega_N = \Omega \cos L + v_E / (R_0 + h)$$

$$\omega_E = -v_N / (R_0 + h)$$

$$\omega_D = -\Omega \sin L - v_E \tan L / (R_0 + h)$$

Ω = Earth's rate

L = latitude

R_0 = radius of the Earth

h = aircraft altitude

f_N , f_E , f_D = north, east and vertical components of vehicle acceleration, respectively

c_{11} , c_{12} , ... = direction cosine elements of the matrix C_b^n

w_{gx} , w_{gy} , w_{gz} = gyroscope noise components

w_{ax} , w_{ay} , w_{az} = accelerometer noise components.

It can be seen from the system error eqn. (10.22) that an acceleration of the aircraft in the north or east direction is required to cause the azimuthal misalignment ($\delta\gamma$) to propagate as a velocity error.

The error model may be augmented by modelling the gyroscope and accelerometer errors explicitly. For example, additional states may be included to represent the fixed biases in the sensor measurements.

To enable the Kalman filter to be mechanised in discrete form, the system error model is converted to a difference equation by integrating between successive measurement instants to give:

$$\delta \mathbf{x}_{k+1} = \Phi_k \delta \mathbf{x}_k + \mathbf{w}_k \quad (10.26)$$

where $\Phi_k = \exp[\mathbf{F}_k(t_{k+1} - t_k)]$, the system transition matrix between time t_k and t_{k+1} and \mathbf{w}_k is a zero mean white noise sequence.

The measurement equations

The measurements of north and east velocity provided by the aircraft's navigation system constitute the Kalman filter measurements ($\tilde{\mathbf{z}}$):

$$\tilde{\mathbf{z}} = \begin{pmatrix} \tilde{v}_N \\ \tilde{v}_E \end{pmatrix} \quad (10.27)$$

Estimates of these measurements ($\hat{\mathbf{z}}$) are obtained from the aligning system:

$$\hat{\mathbf{z}} = \begin{pmatrix} \hat{v}_N \\ \hat{v}_E \end{pmatrix} \quad (10.28)$$

Where the reference and aligning systems are installed some distance apart on the aircraft, it will be necessary to compensate for the rotation-induced velocity components, \mathbf{v}_r , the lever-arm motion. Such corrections are calculated using measurements of the aircraft's turn rate (ω_a) and knowledge of the physical separation between the two systems (\mathbf{r}) using $\mathbf{v}_r = \omega_a \times \mathbf{r}$ resolved in the reference frame. Measurements of ω_a may be provided either by the aircraft's navigation system or by the aligning system with sufficient accuracy.

The velocity measurements are compared at each measurement update to generate the filter measurement differences or innovations, denoted as $\delta \mathbf{z}$, where:

$$\delta \mathbf{z} = \begin{pmatrix} \tilde{v}_N & -\hat{v}_N \\ \tilde{v}_E & -\hat{v}_E \end{pmatrix} = \begin{pmatrix} -\delta v_N \\ -\delta v_E \end{pmatrix} \quad (10.29)$$

The measurement differences at time t_k ($\delta \mathbf{z}_k$) may be expressed in terms of the error states ($\delta \mathbf{x}_k$) as follows:

$$\delta \mathbf{z}_k = \mathbf{H}_k \delta \mathbf{x}_k + \mathbf{v}_k \quad (10.30)$$

where \mathbf{H}_k is the Kalman filter measurement matrix which takes the following form:

$$\mathbf{H}_k = \begin{pmatrix} 0 & 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 0 & -1 \end{pmatrix} \quad (10.31)$$

and \mathbf{v}_k is the measurement noise vector. This represents the noise on the reference measurements and model-mismatch introduced through aircraft flexure and lever-arm motion.

The Kalman filter

In eqns. (10.23) and (10.30), we have the necessary system and measurement equations with which to construct a Kalman filter. The form of the filter equations are given in Appendix A.

The filter provides estimates of the attitude errors and the north and east velocity errors. These estimates are used to correct the aligning system estimates of attitude and velocity after each measurement update. Where instrument bias states are included in the error model, the bias estimates so generated may be used to correct the sensor outputs as part of the alignment process. A block diagram representation of the alignment scheme is given in Figure 10.8.

Whilst it is often recommended that the aircraft should perform a well-defined manoeuvre to aid the alignment process, such as the weave trajectory illustrated in Figure 10.9, analysis of the problem has shown that an alignment can often be achieved in the presence of relatively small perturbations, as would be experienced normally during flight.

Example results

Some simulation results which illustrate the alignment that may be achieved using velocity matching are given in Figure 10.10. The results show the reduction in the alignment error of an airborne navigation system, over a period of 100 s, as the aircraft executes a weave manoeuvre, and have been obtained using a filter formulation similar to that described above, but with the addition of instrument bias states. These results were obtained using a typical aircraft quality system, capable of navigating to an accuracy of 1 nautical mile per hour, to provide the reference measurements. The aligning system was of sub-inertial quality incorporating gyroscopes and accelerometers with 1σ biases of $10^\circ/\text{h}$ and 2 milli-g, respectively.

The figure shows the reduction in the standard deviation of the yaw error as a function of time. The roll and pitch errors, which are not shown here, converge very

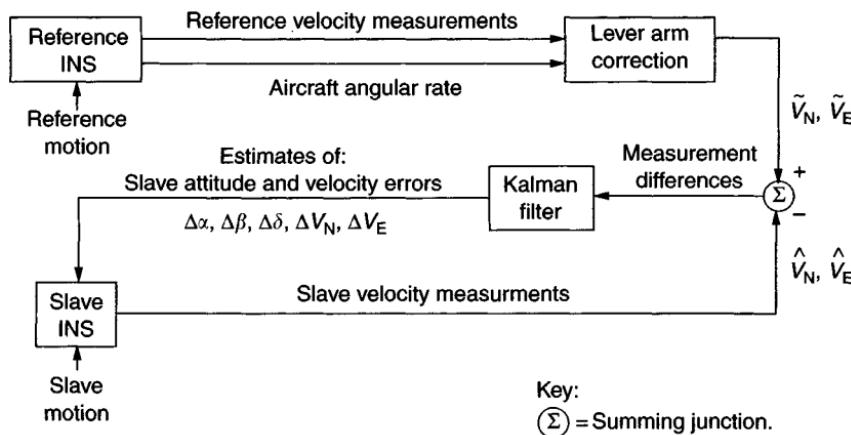


Figure 10.8 Velocity matching alignment scheme

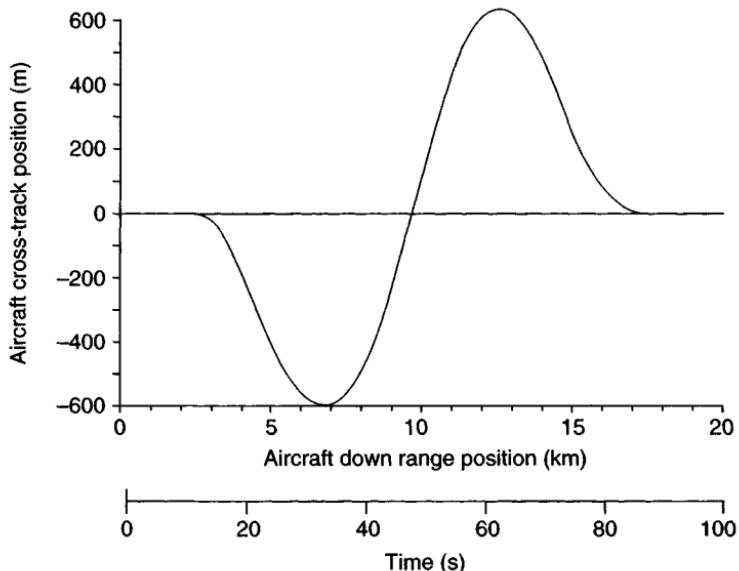


Figure 10.9 Aircraft alignment/calibration manoeuvre

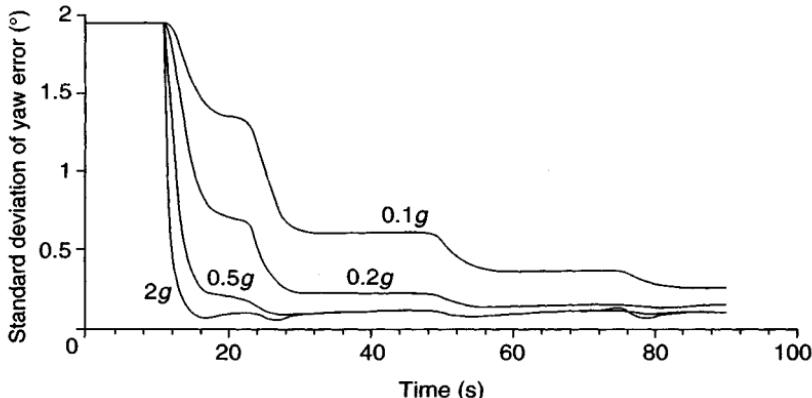


Figure 10.10 Alignment by velocity matching in the presence of an aircraft weave manoeuvre

rapidly as the system effectively aligns itself to the local gravity vector. The accuracy of alignment in level (tilt error) is limited by any residual bias in the accelerometer measurements. In the case shown here, the accelerometer bias is 2 milli-g, which results in tilt errors of approximately 0.1° . The yaw alignment error does not begin to converge until the aircraft commences its manoeuvre, since it only propagates as a velocity error and therefore only becomes observable when the aircraft manoeuvres.

The effects of the manoeuvres are clearly shown in the figure. It can be seen that the yaw alignment error falls each time the aircraft starts to change direction.

In the presence of more severe manoeuvres, mean errors also arise which are correlated with the motion of the aircraft. These errors must be summed with the standard deviations shown in the figure to give the full alignment error. The bias terms are principally the result of geometric effects induced as the aircraft banks to turn. Alignment information can only be deduced about axes which are perpendicular to the direction of the applied acceleration, with the result that some redistribution of the alignment errors tends to take place as the aircraft manoeuvres.

10.4.3.4 Position update alignment

An aircraft may be equipped with various sensors or systems capable of providing position fix information which may be used to align an on-board inertial navigation system during flight. Suitable data may be provided by satellite updates [6] or generated through the use of a ground-based tracking radar or a terrain referenced navigation system of the type discussed later in Chapter 13.

As described earlier, position errors will propagate in an inertial navigation system as a result of alignment inaccuracies. By comparing the external position fixes with the estimates of position generated by the aligning navigation system, estimates of the position errors are obtained. Based on a model of the errors in the aligning system it is possible to deduce the alignment errors from these differences in position. A block diagram of such a scheme is given in Figure 10.11.

This method of alignment is precisely equivalent to the inertial aiding process described in Chapter 13. In the context of integrated navigation systems, or aided inertial navigation systems, the external measurements are assumed to be available throughout all or much of the period for which the navigation system is required to navigate. In the context of pre-flight alignment, it refers to the use of the external measurement data purely to carry out an alignment prior to a period of navigation.

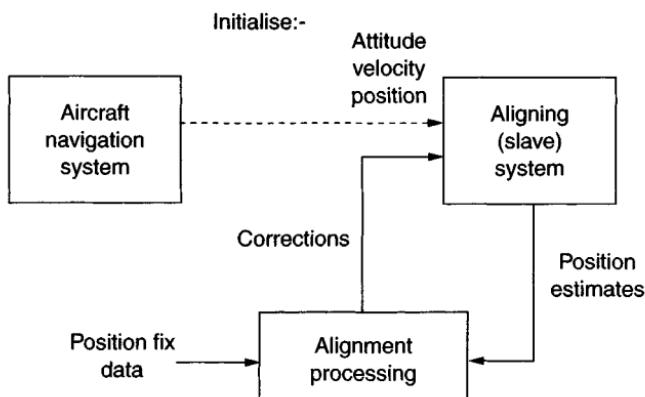


Figure 10.11 Position update alignment scheme

Since the principles of the method are as described in Chapter 13, no further discussion of this topic appears in this chapter.

10.4.3.5 Attitude matching

Recent work has shown that the use of attitude matching, as well as velocity matching, increases the observability of the INS attitude errors, enabling a more accurate alignment to be obtained, or the same accuracy to be obtained with a shorter alignment time or using less manoeuvring of the aircraft. Most importantly, attitude and velocity matching enables an alignment to take place in the presence of a wing rock manoeuvre alone. This is in contrast to velocity matching only which generally requires some heading change manoeuvre, and therefore imposes tactical constraints on the pilot. The attitude difference between the aligning and reference INS is the sum of the attitude error of the aligning INS and the physical relative orientation of the two INS. To separate the two, the Kalman filter must also estimate relative orientation. Adding attitude matching was first proposed by Kain and Cloutier [7]. Flight trials of this technique on a fast jet have been conducted by Graham *et al.* [8] and at QinetiQ, Farnborough [9].

Attitude matching was originally proposed for helicopters where the lever-arm between the reference and aligning INS is relatively rigid. For aircraft where the weapon is mounted on a wing pylon, the flexure environment is more severe. Lever-arm vibration effects can be averaged out by selecting suitably low gains in the Kalman filter. However a more serious problem is presented by the flexure of the wings and pylons in response to aircraft manoeuvre. This can seriously disrupt the performance of transfer alignment using attitude matching. The solution is to introduce additional Kalman filter states that model the variation of the relative orientation with the forces on the wing and to increase the assumed measurement noise in the Kalman filter as a function of the departure of the forces on the wing from their steady state values.

Transfer alignment performance is enhanced by estimating inertial instrument errors as well as velocity and attitude. Estimating accelerometer and gyroscope biases has a huge effect on performance. Further improvements can be attained for some types of IMU by separating the biases into static and dynamic (Markov) states and by estimating scale-factor and cross-coupling errors for both accelerometers and gyros.

The best navigation performance that a transfer aligned INS can attain is that of the reference. Thus, if the aircraft contains an integrated INS–GPS navigation system, this will generally provide a more accurate reference than a pure INS. However, when GPS signals are suddenly re-acquired after a period of jamming (e.g. if the jammer is destroyed) the transient in the aircraft velocity solution as GPS corrects the inertial drift can disrupt the transfer alignment process. The crude solution is to use pure INS as the main transfer alignment reference and just use the integrated solution to correct the weapon position at launch. However, this discards the GPS calibration of the aircraft INS velocity and attitude. Thus, it is better to use the INS–GPS solution as the reference and add a transient handling algorithm.

The best approach to transient handling is to detect transients directly, either by comparing the integrated and pure INS solutions or by taking correction information

from the aircraft navigation filter. In this case, the transient is applied to the missile velocity solution outside the transfer alignment Kalman filter to keep it in step. Where this cannot be done, the transfer alignment algorithm must monitor the measurement residuals for the effects of transients and, if it finds one, selectively increase the error covariance, to make the velocity error estimates more receptive to the corrected aircraft solution.

10.5 Alignment at sea

10.5.1 Introduction

A modern warship contains a wide variety of sensors and weapon systems. In order that the ship can deploy the forces at its disposal and use them in an effective manner, all such equipment must operate in harmony. For example, information about an attacking missile or aircraft derived from a sensor at one location must be in a form that can be used to direct or control a weapon system at a different remote location.

10.5.2 Sources of error

It is common practice to set up a series of datum levels and training marks at strategic locations around the ship to which all equipment is referenced or harmonised when it is installed on the ship. In this way, it is hoped to ensure that all equipment will operate in a common frame of reference. It has long been suspected that whilst the accuracy to which equipment is harmonised during the construction of the ship is very high, the accuracy of this harmonisation degrades when the ship goes to sea. This view has been reinforced by observations of ships at sea and the results of ship trials which have attempted to measure the amount by which ships bend or flex in different sea conditions. Such errors may be categorised as follows:

Long-term deformations occurring through the action of ageing and the effects of solar heating. A gradual movement of the structure takes place as the ship ages and as the load state changes. It has also been observed that significant bending of the ship structure can occur under the action of solar heating. Angular variations of the order of 1° are believed to take place over the period of a day as the sun moves around the vessel.

Ship flexure can occur in heavy seas as the ship moves in response to the motion of the waves, the magnitude of the angular displacement between any two locations becoming larger as the separation increases. Attempts to measure the amount by which ships flex when at sea have revealed significant angular displacements at typical ship motion frequencies of 0.1–0.3 Hz, the dominant flexure motion being the twisting of the hull about the roll axis of the vessel. The magnitude of ship flexure is a function of sea state and the direction in which the waves are approaching the vessel. Further transient distortion may occur as the ship manoeuvres, or through the action of the stabilisers.

Other abrupt changes which are expected to arise from underwater shock, induced for instance by a depth charge, and as a result of slamming in heavy seas, where the bows leave the water and impact on re-entry.

In addition, battle damage will introduce potentially very large distortions of a ship's structure, probably rendering some weapon systems ineffective unless a static reharmonisation takes place.

10.5.3 Shipboard alignment methods

To overcome the problems outlined in the previous section, it is necessary to devise means by which the harmonisation of the various shipboard systems can be maintained under all operational conditions. Whilst an accurate reference is provided on naval ships by the ship's attitude and heading reference system (AHRS) or even more precisely by a ship's inertial navigation system (SINS), the accuracy with which that reference may be transferred about the ship is limited by bending and flexure of the ship. For this reason, other means are sought for the alignment of equipment on-board ships.

10.5.3.1 Shipboard transfer alignment methods

Assuming a master reference can be maintained accurately, slave systems may be aligned to that reference. There are various methods which may be adopted to achieve this end. The simplest technique is to transfer data – attitude, velocity and position – directly from the master system to the slave using the one-shot alignment scheme described above for airborne alignment. However, as with airborne alignment, any physical misalignments resulting from ship flexure, for example, will contribute directly to the errors in the aligning system if this approach is adopted.

One possible method of overcoming this limitation on-board a ship is to use an optical harmonisation scheme to determine the relative orientation of the master reference of the launch platform and a missile system directly. An auto-collimator, fixed in one co-ordinate reference frame, may be used to determine the rotation of a reflector which is attached to the second reference frame. Although such techniques have been used in some applications, they are not generally feasible because of the difficulty of maintaining line-of-sight contact between the two locations which could be some considerable distance apart. For example, a missile silo in a ship may be installed 50 m, or more, away from the ship's inertial reference system.

Alternatively, alignment may be achieved on board a ship by comparing inertial measurements generated by the aligning system with similar measurements provided by a reference unit [10, 11]. The velocity matching scheme described in Section 10.4 for in-flight alignment is of limited use for shipboard applications since it is dependent on a manoeuvre of the vehicle, particularly if an alignment is to take place within a short period of time. In many circumstances this may be totally impractical. Studies of shipboard alignment methods have suggested that the use of velocity and pitch rate matching offers a possible solution [11]. Such a scheme is discussed in more detail in the following section.

10.5.3.2 Shipboard inertial measurement matching

In this section, the scope for achieving an alignment at sea using velocity and angular rate matching is discussed. The application of velocity matching alone is of limited use for shipboard alignment because ships are clearly unable to manoeuvre in the way

that aircraft can to aid the alignment process. However, velocity matching may be used to achieve a level alignment, since errors in the knowledge of the local vertical will cause the measurements of specific force needed to overcome gravity to be resolved incorrectly and to propagate as apparent components of north and east velocity.

On-board a ship, an alignment in azimuth may be achieved within a relatively short period of time by comparing angular rate measurements, provided the ship exhibits some motion in pitch or roll. The measurements may be processed using a Kalman filter based on an error model of the aligning system, as described in the context of in-flight alignment in Section 10.4. The form of the measurement equation is described below.

The measurements of turn rate provided by the reference and aligning systems are assumed to be generated in local co-ordinate frames denoted a and b, respectively. The rates sensed by a triad of strapdown gyroscopes mounted at each location with their sensitive axes aligned with these reference frames may be expressed as ω_{ia}^a and ω_{ib}^b in line with the nomenclature used in Chapter 3. The measurements provided by the gyroscopes in the reference and aligning systems are resolved into a common reference frame, the a-frame, for instance, before comparison takes place.

Hence, the reference measurements may be expressed as:

$$\hat{\mathbf{z}} = \omega_{ia}^a \quad (10.32)$$

assuming errors in the measurements to be negligible. The estimates of these measurements generated by the aligning system are denoted by the \wedge notation.

$$\hat{\mathbf{z}} = \hat{\mathbf{C}}_b^a \hat{\omega}_{ib}^b \quad (10.33)$$

The gyroscope outputs ($\hat{\omega}_{ib}^b$) may be written as the sum of the true rate (ω_{ib}^b) and the error in the measurement ($\delta\omega_{ib}^b$) whilst the estimated direction cosine matrix may be expressed as the product of a skew symmetric error matrix, [$\mathbf{I} - \Psi$], and the true matrix \mathbf{C}_b^a to give:

$$\hat{\mathbf{z}} = [\mathbf{I} - \Psi] \mathbf{C}_b^a [\omega_{ib}^b + \delta\omega_{ib}^b]$$

Expanding the right-hand side of this equation, writing $\Psi = [\Psi \times]$ and ignoring error product terms gives:

$$\begin{aligned} \hat{\mathbf{z}} &= \mathbf{C}_b^a \omega_{ib}^b - \Psi \times \mathbf{C}_b^a \omega_{ib}^b + \mathbf{C}_b^a \delta\omega_{ib}^b \\ &= \omega_{ib}^a + \omega_{ib}^a \times \Psi + \mathbf{C}_b^a \delta\omega_{ib}^b \end{aligned} \quad (10.34)$$

The turn rate of the aligning system may be expressed as the sum of the turn rate sensed by the reference system and any ship flexure which may be present (ω_f). Hence, eqn. (10.34) may be rewritten as follows:

$$\hat{\mathbf{z}} = \omega_{ia}^a + \omega_f + \omega_{ib}^a \times \Psi + \mathbf{C}_b^a \delta\omega_{ib}^b \quad (10.35)$$

The measurement differences may then be written as:

$$\begin{aligned}\delta \mathbf{z} &= \tilde{\mathbf{z}} - \hat{\mathbf{z}} \\ &= -\omega_{ib}^a \times \Psi - \mathbf{C}_b^a \delta \omega_{ib}^b - \omega_f\end{aligned}\quad (10.36)$$

The measurement differences (δz_k) at time t_k may be expressed in terms of the error states (δx_k) as follows:

$$\delta \mathbf{z}_k = \mathbf{H}_k \delta \mathbf{x}_k + \mathbf{v}_k \quad (10.37)$$

where \mathbf{H}_k is the Kalman filter measurement matrix which takes the following form:

$$\mathbf{H}_k = \begin{pmatrix} 0 & \omega_Z & -\omega_Y & 0 & 0 \\ -\omega_Z & 0 & \omega_X & 0 & 0 \\ \omega_Y & -\omega_X & 0 & 0 & 0 \end{pmatrix} \quad (10.38)$$

where ω_X , ω_Y and ω_Z are the components of the vector ω_{ib}^a and \mathbf{v}_k is the measurement noise vector. This represents the noise on the measurements and model-mismatch introduced through ship flexure.

A Kalman filter may now be constructed using the measurement eqn. (10.37) and a system equation of the form described earlier, Section 10.4.3.3, eqn. (10.23). A block diagram of the resulting alignment scheme is given in Figure 10.12.

Example result

The simulation result shown in Figure 10.13 illustrates the accuracy of alignment which may be achieved using a combination of velocity and angular rate matching. The results show the convergence of the azimuthal alignment error in calm, moderate and rough sea conditions where the waves are approaching the ship from the side.

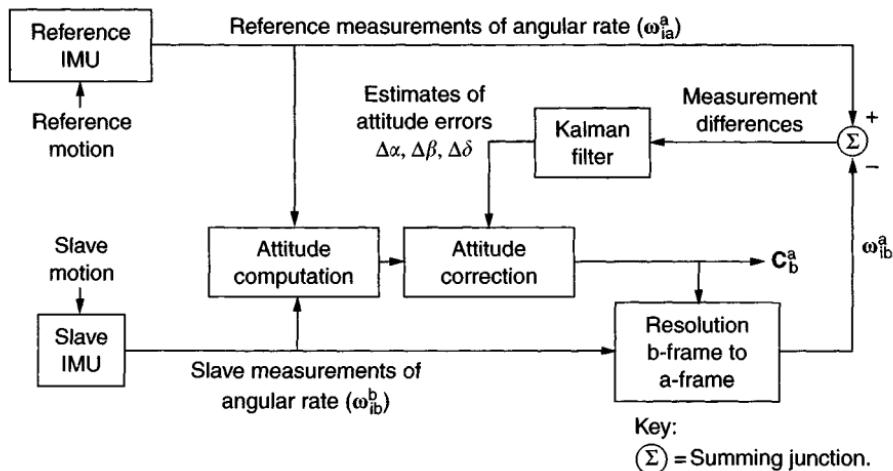


Figure 10.12 Angular rate matching alignment scheme

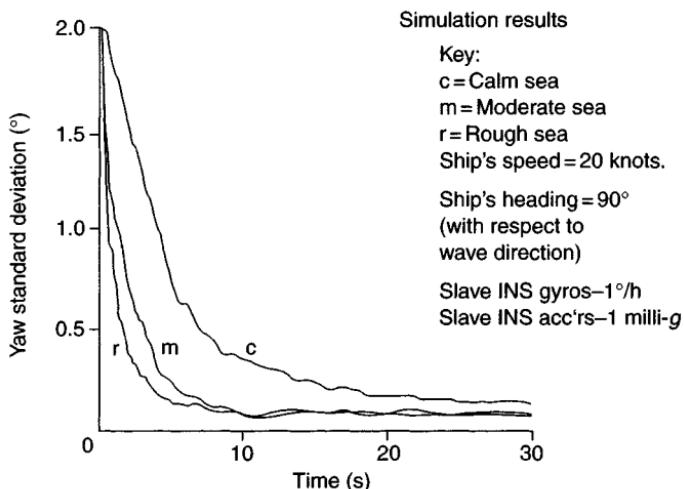


Figure 10.13 Illustration of measurement matching at sea

These results were obtained assuming no knowledge of the ship's flexure characteristics. However, the measurements of velocity were compensated for relative motion of the reference and aligning systems caused by the rotation of the ship. The aligning system contained medium grade inertial sensors with accelerometer biases of 1 milli-g and gyroscope biases of 1°/h; a higher quality reference system was used. The Kalman filter used here was found to be robust in that it is able to cope with initial alignment errors of 10° or more.

The effects of ship flexure

Whilst it is possible in theory to model the ship's flexure explicitly in the Kalman filter and so derive estimates of the flexure rates, a sufficiently precise model is unlikely to be available in practice. Besides, this will result in a 'highly tuned' filter which will be very sensitive to parametric variations. For these reasons, a sub-optimal Kalman filter may be used in which the flexure is represented as a noise process, as described above. The way in which ship flexure limits the accuracy of alignment which can be achieved when using a filter of this type is demonstrated by the simplified analysis which follows.

Consider the two axis sets shown in Figure 10.14 which correspond to the orientations of the reference and aligning systems at two locations remote from each other on a ship. The reference frame is taken to be aligned perfectly with the roll, pitch and yaw axes of the ship whilst the aligning system, denoted here as the slave system, is misaligned in yaw by an angle $\delta\psi$.

In Figure 10.14, O_aXY denotes reference axes at reference system origin; O_bXY denotes a parallel reference axes at the slave system origin and O_{bxy} denotes the slave system axes to be brought into alignment with O_bXY .

The angular rates p and q sensed about the reference axes are the roll and pitch rates of the vessel, respectively. The slave system senses rates $p + \delta p$ and $q + \delta q$

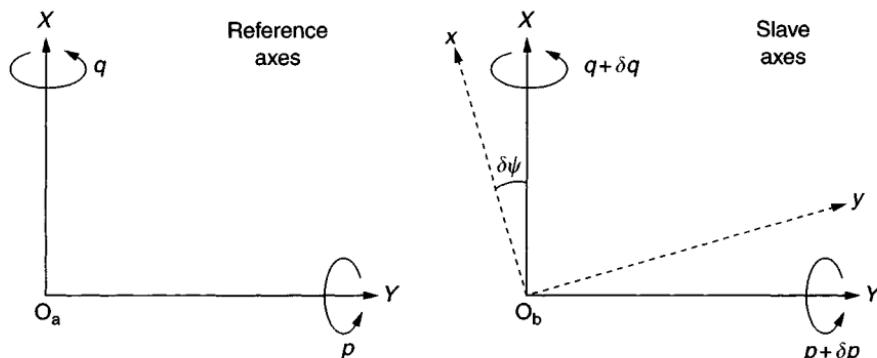


Figure 10.14 Illustration of the effects of ship flexure on axis alignment

resolved in slave system axes, where δp and δq represent the relative angular rates between the two systems, the rates at which the ship is bending or flexing.

Consider first the mechanism by which alignment occurs in the absence of flexure. Using pitch rate matching, the rate measured by the reference system, q , is compared with the slave system rate, $q \cos \delta\psi - p \sin \delta\psi$, to yield a measurement difference δz , where:

$$\delta z = q(1 - \cos \delta\psi) + p \sin \delta\psi \quad (10.39)$$

It can be seen from the above equation that δz becomes zero when the misalignment is zero. Hence, by adjusting $\delta\psi$ in order to null this measurement difference, it is possible to align the slave system perfectly in the absence of ship flexure.

In the presence of ship flexure, additional turn rates δp and δq are present at the slave system and the rate sensed about the nominal pitch axis of the slave system becomes $(q + \delta q)\cos \psi - (p + \delta p)\sin \psi$. The measurement difference is now:

$$\delta z = q(1 - \cos \delta\psi) + p \sin \delta\psi - \delta q \cos \delta\psi + \delta p \sin \delta\psi \quad (10.40)$$

which may be expressed to first order in $\delta\psi$ as:

$$\delta z = (p + \delta p)\delta\psi - \delta q \quad (10.41)$$

In this case, the measurement difference settles to zero when:

$$\delta\psi = \frac{\delta q}{(p + \delta p)} \quad (10.42)$$

It is clear from this result that the magnitude of the residual yaw misalignment will reduce as the roll rate of the ship becomes larger, or as the flexure about the measurement axis, pitch in this case, becomes smaller. By a similar argument, it can be shown that the accuracy of the estimate of yaw error obtained using roll rate matching will be limited by the relative magnitude of the roll flexure and the pitch rate of the vessel. Since flexure about the roll axis is believed to be larger than the pitch rate flexure in

general, and ships tend to roll more rapidly than they pitch, pitch rate matching is the preferred option.

Figure 10.15 shows the azimuthal alignment accuracy achieved as the ratio of pitch rate flexure to roll rate is varied. In line with theoretical expectations discussed here, the accuracy of alignment is shown to improve as this ratio becomes smaller.

10.5.3.3 Shipboard alignment using position fixes

Accurate harmonisation between different items of equipment on-board a ship may be achieved using inertial navigation systems installed alongside each item, or system, to maintain a common reference frame at each location. Such a scheme is shown in Figure 10.16.

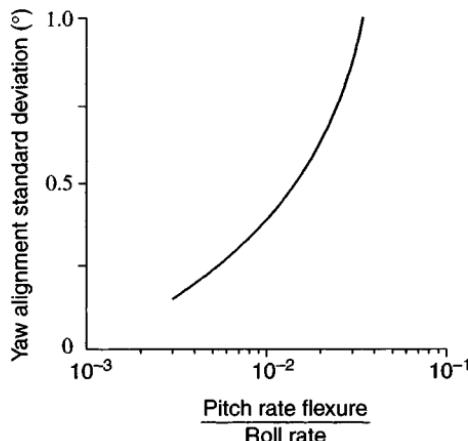


Figure 10.15 Azimuth alignment accuracy as a function of the ratio pitch rate flexure to roll rate

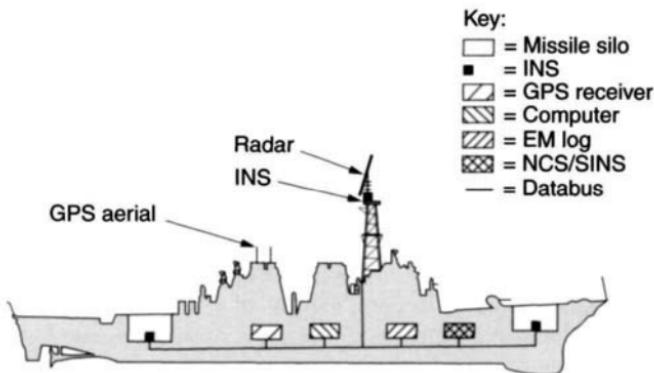


Figure 10.16 Shipboard harmonisation scheme

The reference may be maintained at each location by using accurate position fixes, provided by satellite updates for instance. It is envisaged that each system could be equipped with a GPS satellite receiver and antenna to facilitate its alignment to the local geographic frame. Alternatively, with appropriate filtering and lever-arm corrections, a single GPS receiver could feed all of the inertial systems on the ship with positional data. It is noted that the GPS receiver gives the location of the phase centre of the antenna which is likely to be located at the top of a mast.

The alignment of each inertial system may be accomplished independently of ship motion, although the speed of convergence is greatly increased in the presence of the ship's manoeuvres. This technique would enable the accurate alignment of each system to be achieved, largely irrespective of any relative motion between the different locations resulting from bending of the ship's structure. Clearly, this approach is dependent on the continuing availability of satellite signals. In the event of loss of such transmissions, the period of time for which alignment can subsequently be maintained is dependent on the quality and characteristics of the sensors in each inertial unit.

References

- 1 BRITTING, K.R.: 'Inertial navigation system analysis' (John Wiley and Sons, 1971)
- 2 DEYST, J.J., and SUTHERLAND, A.A.: 'Strapdown inertial system alignment using statistical filters: a simplified formulation', *AIAA Journal*, 1973, **11** (4)
- 3 HARRIS, R.A., and WAKEFIELD, C.D.: 'Co-ordinate alignment for elastic bodies', *NAECON* 1977
- 4 SCHULTZ, R.L., and KEYS, C.L.: 'Airborne IRP alignment using acceleration and angular rate matching'. Proceedings *Joint automatic control conference*, June 1973
- 5 BAR-ITZHACK, I.Y., and PORAT, B.: 'Azimuth observability enhancement during inertial navigation system in-flight alignment', *Journal of Guidance and Control*, 1980, **3** (4)
- 6 TAFEL, R.W., and KRASNJANSKI, D.: 'Rapid alignment of aircraft strapdown inertial navigation systems using Navstar GPS'. AGARD conference proceedings on *Precision positioning and guidance systems*, 1980
- 7 KAIN, J.E., and CLOUTIER, J.R.: 'Rapid transfer alignment for tactical weapon applications'. Proceedings of *AIAA Guidance, Navigation and Control Conference*, 1989
- 8 GRAHAM, W.R., SHORTELLE, K.J., and RABOURN, C.: 'Rapid alignment prototype (RAP) flight test demonstration'. Proceedings of the *Institute of Navigation National Technical Meeting*, 1998
- 9 GROVES, P.D.: 'Optimising the transfer alignment of weapon INS', *Journal of Navigation*, 2003, **56**

- 10 BROWNE, B.H., and LACKOWSKI, D.H.: 'Estimation of dynamic alignment error in shipboard fire control systems'. Proceedings of IEE conference on *Decision and control*, 1976
- 11 TITTERTON, D.H., and WESTON, J.L.: 'Dynamic shipboard alignment techniques'. DGON proceedings, *Gyro Technology Symposium*, Stuttgart, 1987

Chapter 11

Strapdown navigation system computation

11.1 Introduction

The analytical equations which must be solved in order to extract attitude, velocity and position information from the inertial measurements provided by the gyroscopes and accelerometers in a strapdown system have been described in Chapter 3. This chapter is concerned with the real time implementation of these equations in a computer. The main computing tasks, those of attitude determination, specific force resolution and solution of the navigation equation, are indicated in the block diagram shown in Figure 11.1.

The most demanding of the processing tasks, in terms of computer loading, are the attitude computation and the specific force vector resolution. In the presence of high frequency motion, the implementation of these tasks in real time creates a substantial computing burden for the strapdown system computer, even with modern processors. The navigation processing task, which is common to all types of inertial navigation system, both strapdown and platform mechanisations, is less demanding computationally. In addition to these tasks, the determination of attitude in terms of Euler rotations is required in some applications.

Early attempts to produce a strapdown inertial navigation system were limited, in part, by the computer technologies available at the time. Apart from the physical size of early computers which delayed the development of strapdown systems, particularly for airborne applications, the lack of computing speed, or throughput, was a major obstacle to the achievement of fast and accurate attitude computation. As a result, the performance which could be achieved in early strapdown systems was limited, particularly under high frequency motion conditions.

Such difficulties prompted much effort to be directed towards the development of efficient computing algorithms, and in particular, the splitting of the strapdown computing processes into low- and high-speed segments. The low-speed calculations are designed to take account of low frequency, large amplitude, body motions arising from vehicle manoeuvres, whilst the high-speed section involves a relatively simple

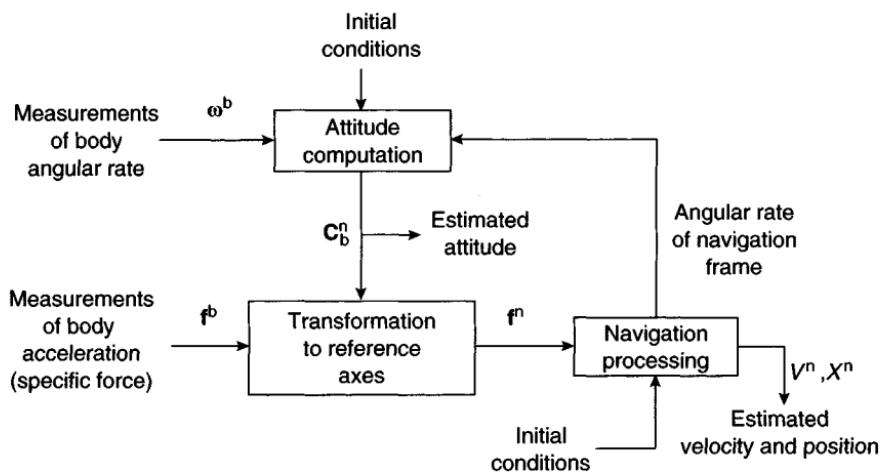


Figure 11.1 Strapdown inertial navigation computing tasks

algorithm which is designed to keep track of the high frequency, low amplitude, motions of the vehicle. Contemporary algorithms often adopt this approach for both the attitude computation and the specific force resolution.

11.2 Attitude computation

As indicated previously, it is the computation of attitude which is particularly critical in a strapdown system. It should therefore come as no surprise to find that it is this topic which has been the subject of much study [1–5]. In many applications, the dynamic range of the angular motions to be taken account of can be very large, varying from a few degrees per hour to 2000° per second or more. In addition, the system may be subjected to high frequency dynamic motion in some applications. For example, a strapdown system in a guided missile may be subjected to such motion as a result of body bending and rocket motor induced vibration. The ability of the strapdown algorithm to keep track of body attitude accurately in a severe vibratory environment may well be the critical factor in determining its performance, if accurate navigation is to be achieved.

The conventional approach to attitude determination is to compute the direction cosine matrix, relating the vehicle body reference frame to the reference co-ordinate system, or its quaternion equivalent, using a numerical integration scheme. It is theoretically possible to calculate body attitude sufficiently accurately, even in the presence of high frequency angular motion, provided that the computational frequency is sufficiently high. However, in practice this may impose an intolerable burden on the processor.

An alternative formulation advocated by Bortz [1] involves the representation of changes in attitude as a rotation vector. As described below, this approach allows

the attitude computation to be split conveniently into low- and high-speed sections, denoted here as the k -cycle and j -cycle rates. The lower-speed part of the calculation is designed to take account of the relatively low frequency, large amplitude, body motion arising as a result of vehicle manoeuvres. The high-speed section involves a simple algorithm, which is designed to track the high frequency, low amplitude, motions of the vehicle.

Using this approach, coning motion at frequencies near to the lower computational rate may be accounted for, without the need to increase the speed at which the bulk of the computation is implemented.

11.2.1 Direction cosine algorithms

In order to update the direction cosine matrix, \mathbf{C} , defined in Chapter 3, it is required to solve a matrix differential equation of the form:

$$\dot{\mathbf{C}} = \mathbf{C}\boldsymbol{\Omega} \quad (11.1)$$

where $\boldsymbol{\Omega}$ is a skew symmetric matrix formed from the elements of the turn rate vector $\boldsymbol{\omega}$. For clarity, the subscripts and superscripts used earlier have been omitted in the following development.

Over a single computer cycle, from time t_k to t_{k+1} , the solution of the equation may be written as follows:

$$\mathbf{C}_{k+1} = \mathbf{C}_k \exp \int_{t_k}^{t_{k+1}} \boldsymbol{\Omega} dt \quad (11.2)$$

Provided that the orientation of the turn rate vector, $\boldsymbol{\omega}$, remains fixed in space over the update interval, we may write:

$$\int_{t_k}^{t_{k+1}} \boldsymbol{\Omega} dt = [\boldsymbol{\sigma} \times] \quad (11.3)$$

Hence, eqn. (11.2) becomes:

$$\begin{aligned} \mathbf{C}_{k+1} &= \mathbf{C}_k \exp[\boldsymbol{\sigma} \times] \\ &= \mathbf{C}_k \mathbf{A}_k \end{aligned} \quad (11.4)$$

where \mathbf{C}_k represents the direction cosine matrix which relates body to reference axes at the k th computer cycle, and \mathbf{A}_k the direction cosine matrix which transforms a vector from body coordinates at the $k + 1$ th computer cycle to body coordinates at the k th computer cycle. The variable $\boldsymbol{\sigma}$ is an angle vector with direction and magnitude such that a rotation of the body frame about $\boldsymbol{\sigma}$, through an angle equal to the magnitude of $\boldsymbol{\sigma}$, will rotate the body frame from its orientation at computer cycle k to its position at computer cycle $k + 1$. The components of $\boldsymbol{\sigma}$ are denoted by σ_x , σ_y and σ_z and its magnitude is given by:

$$\sigma = \sqrt{(\sigma_x^2 + \sigma_y^2 + \sigma_z^2)} \quad (11.5)$$

and

$$\boldsymbol{\sigma} \times = \begin{bmatrix} 0 & -\sigma_z & \sigma_y \\ \sigma_z & 0 & -\sigma_x \\ -\sigma_y & \sigma_x & 0 \end{bmatrix} \quad (11.6)$$

Expanding the exponential term in eqn. (11.4) gives:

$$\mathbf{A}_k = \mathbf{I} + [\boldsymbol{\sigma} \times] + \frac{[\boldsymbol{\sigma} \times]^2}{2!} + \frac{[\boldsymbol{\sigma} \times]^3}{3!} + \frac{[\boldsymbol{\sigma} \times]^4}{4!} + \dots \quad (11.7)$$

and using eqn. (11.6) it can be shown that:

$$[\boldsymbol{\sigma} \times]^2 = \begin{bmatrix} -(\sigma_y^2 + \sigma_z^2) & \sigma_x \sigma_y & \sigma_x \sigma_z \\ \sigma_x \sigma_y & -(\sigma_x^2 + \sigma_z^2) & \sigma_y \sigma_z \\ \sigma_x \sigma_z & \sigma_y \sigma_z & -(\sigma_x^2 + \sigma_y^2) \end{bmatrix}$$

$$[\boldsymbol{\sigma} \times]^3 = -(\sigma_x^2 + \sigma_y^2 + \sigma_z^2)[\boldsymbol{\sigma} \times]$$

$$[\boldsymbol{\sigma} \times]^4 = -(\sigma_x^2 + \sigma_y^2 + \sigma_z^2)[\boldsymbol{\sigma} \times]^2$$

$$\vdots \quad (11.8)$$

Thus, we may write:

$$\begin{aligned} \mathbf{A}_k &= \mathbf{I} + [\boldsymbol{\sigma} \times] + \frac{[\boldsymbol{\sigma} \times]^2}{2!} - \frac{\sigma^2 [\boldsymbol{\sigma} \times]}{3!} - \frac{\sigma^2 [\boldsymbol{\sigma} \times]^2}{4!} + \dots \\ &= \mathbf{I} + \left[1 - \frac{\sigma^2}{3!} + \frac{\sigma^4}{5!} - \dots \right] [\boldsymbol{\sigma} \times] + \left[\frac{1}{2!} - \frac{\sigma^2}{4!} + \frac{\sigma^4}{6!} - \dots \right] [\boldsymbol{\sigma} \times]^2 \end{aligned} \quad (11.9)$$

which may be written as follows:

$$\mathbf{A}_k = \mathbf{I} + \frac{\sin \sigma}{\sigma} [\boldsymbol{\sigma} \times] + \frac{(1 - \cos \sigma)}{\sigma^2} [\boldsymbol{\sigma} \times]^2 \quad (11.10)$$

Provided that σ is the angle vector as defined above, eqn. (11.10) provides an exact representation of the attitude matrix which relates body attitude at times t_{k+1} and t_k . If it was possible to implement this equation perfectly, it would yield an orthogonal matrix which need only be evaluated when transformation of the measured specific force vector is required. In practice of course, it is necessary to truncate the mathematical functions in eqn. (11.10), in order to produce an algorithm which can be implemented in real time. Following eqn. (11.9), \mathbf{A}_k may be calculated using:

$$\mathbf{A}_k = \mathbf{I} + a_1 [\boldsymbol{\sigma} \times] + a_2 [\boldsymbol{\sigma} \times]^2 \quad (11.11)$$

where

$$a_1 = 1 - \frac{\sigma^2}{3!} + \frac{\sigma^4}{5!} - \dots$$

and

$$a_2 = \frac{1}{2!} - \frac{\sigma^2}{4!} + \frac{\sigma^4}{6!} - \dots$$

The direction cosine matrix may therefore be updated for body motion, as sensed by the strapdown gyroscopes, using a recursive algorithm based on eqns. 11.4 and 11.9. The order of the algorithm will be determined by the number of terms included in eqn. (11.9). For example, both of the infinite series would be truncated at the σ^2 level in a fourth order algorithm. The computation rate should be selected to ensure that the magnitude of σ remains small at the maximum turn rate, thus avoiding the need to include a large number of terms in the expression for \mathbf{A}_k .

11.2.1.1 The definition of attitude computation errors

The computed attitude matrix, written here as $\hat{\mathbf{A}}$, may be expressed in terms of the true attitude matrix, \mathbf{A} , and an error matrix \mathbf{E} as follows:

$$\hat{\mathbf{A}} = \mathbf{A}[\mathbf{I} + \mathbf{E}] \quad (11.12)$$

Rearranging, we have:

$$\mathbf{E} = \mathbf{A}^T \hat{\mathbf{A}} - \mathbf{I} \quad (11.13)$$

Substituting for \mathbf{A} and $\hat{\mathbf{A}}$ following eqns. 11.10 and 11.11, we may write:

$$\begin{aligned} \mathbf{E} &= \left[\mathbf{I} - \frac{\sin \sigma}{\sigma} [\sigma \times] + \frac{(1 - \cos \sigma)}{\sigma^2} [\sigma \times]^2 \right] \left[\mathbf{I} + a_1 [\sigma \times] + a_2 [\sigma \times]^2 \right] - \mathbf{I} \\ &= (\sigma a_1 \cos \sigma - \sin \sigma + \sigma^2 a_2 \sin \sigma) \frac{[\sigma \times]}{\sigma} \\ &\quad + (1 - \cos \sigma - \sigma a_1 \sin \sigma + \sigma^2 a_2 \cos \sigma) \frac{[\sigma \times]^2}{\sigma^2} \end{aligned} \quad (11.14)$$

The first term in eqn. (11.14) represents a skew symmetric matrix, following the form of $[\sigma \times]$, and is denoted below by the symbol \mathbf{U} . The second term is symmetric, following $[\sigma \times]^2$, and may be represented by a symmetric matrix \mathbf{S} . Hence, we may write:

$$\mathbf{E} = \mathbf{U} + \mathbf{S} \quad (11.15)$$

as shown in Reference 5.

\mathbf{A} is an orthogonal matrix, if the equation $\mathbf{A}^T \mathbf{A} = \mathbf{I}$ is satisfied. For the computed matrix, we may write:

$$\hat{\mathbf{A}}^T \hat{\mathbf{A}} = [\mathbf{I} + \mathbf{E}]^T [\mathbf{I} + \mathbf{E}] \quad (11.16)$$

Ignoring second- and higher-order terms gives:

$$\hat{\mathbf{A}}^T \hat{\mathbf{A}} = \mathbf{I} + \mathbf{E} + \mathbf{E}^T \quad (11.17)$$

Substituting for \mathbf{E} and \mathbf{E}^T in terms of their symmetric and skew symmetric components, and noting that $\mathbf{S}^T = \mathbf{S}$ and $\mathbf{U}^T = -\mathbf{U}$, gives:

$$\hat{\mathbf{A}}^T \hat{\mathbf{A}} = \mathbf{I} - 2\mathbf{S} \quad (11.18)$$

We can also write:

$$\mathbf{A}^T \hat{\mathbf{A}} = \mathbf{I} + \mathbf{S} + \mathbf{U} \quad (11.19)$$

From the last two equations, the following conclusions can be drawn:

\mathbf{S} represents the deviation of the matrix $\hat{\mathbf{A}}$ from the orthogonal form. The diagonal elements of \mathbf{S} are termed the scale errors, whilst the off-diagonal terms represent the skew errors [5].

If \mathbf{S} is zero, $\hat{\mathbf{A}}$ becomes an orthogonal matrix representing a co-ordinate rotation, which is different from that defined by \mathbf{A} . \mathbf{U} provides a measure of the difference between the two rotations.

A single parameter, D_{dc} , the root sum square of the upper or lower off-diagonal elements of \mathbf{U} , divided by the computer update interval δt , may be used as a measure of the drift in the computed attitude matrix. The parameter D_{dc} may be used to assess the accuracy of the various orders of attitude algorithm considered.

The matrix \mathbf{U} is defined as:

$$\mathbf{U} = (\sigma a_1 \cos \sigma - \sin \sigma + \sigma^2 a_2 \sin \sigma) \frac{[\sigma \times]}{\sigma} \quad (11.20)$$

In the case of a single x -axis rotation, where $\sigma = [\sigma \ 0 \ 0]^T$, we have:

$$D_{dc} = \frac{1}{\delta t} (\sigma a_1 \cos \sigma - \sin \sigma + \sigma^2 a_2 \sin \sigma) \quad (11.21)$$

where

$a_1 = 1, a_2 = 0$ is a first-order algorithm,

$a_1 = 1, a_2 = 0.5$ is a second-order algorithm,

$a_1 = 1 - (\sigma^2/6), a_2 = 0.5$ is a third-order algorithm.

11.2.1.2 Example

Consider cases in which the maximum size of the angular increment (σ_{max}) is 0.1 and 0.05 rad. If the maximum angular rate of the body is 10 rad/s, these figures correspond to update intervals of 0.01 and 0.005 s.

The drift errors in the computed attitude are shown in Table 11.1 for these two cases using different orders of algorithm.

The substantial improvement in accuracy which may be achieved by reducing the update interval is clearly seen. The large reduction in the drift rate obtained through the inclusion of third-order terms, occurs as a result of the cancellation of both the third- and fourth-order terms in the expression for drift error with this level of truncation. In many applications, high turn rates will not normally be sustained for long periods, the mean rates expected being substantially lower than the figure of

Table 11.1 Drift errors in the computed attitude using different orders of direction cosine algorithm

Order of algorithm	Attitude drift error ($^{\circ}/\text{h}$)	
	$\sigma_{\max} = 0.1 \text{ rad}$	$\sigma_{\max} = 0.05 \text{ rad}$
1	6870	1720
2	3430	860
3	7	0.4
4	1.7	0.1

10 rad/s considered in the above analysis. In such cases, the mean drift errors resulting from imprecise attitude computation will be considerably smaller than the figures quoted in the table. It may therefore be feasible to use first- or second-order update algorithms for some applications.

It is assumed here that the major part of the attitude computation described above, along with some other aspects of the navigation processing to be described later, will be implemented at the lower k -cycle data rate. However, some parts of the computation may need to be performed at higher rates, whilst others can be carried out less frequently, as discussed below.

11.2.2 Rotation angle computation

A further limitation on the accuracy of the direction cosine matrix updates is the accuracy with which the rotation angle, σ , can be determined. Consider first the case where the direction of the angular rate vector, ω , remains fixed in space over an update interval. In this case, σ is determined quite simply as the integral of ω over the computer cycle, k :

$$\sigma = \int_{t_k}^{t_{k+1}} \omega \, dt \quad (11.22)$$

that is, σ is the sum of the incremental measurements provided directly by some gyroscopes over the time interval t_k to t_{k+1} . The relationship between σ and ω can also be expressed as $d\sigma/dt = \omega$ in this situation.

In general, however, it is not possible to determine σ precisely by simply summing the measurements of incremental angle. If the direction of ω does not remain fixed in space but is rotating, as occurs in the presence of coning motion, for example, then following Bortz [1], we may write:

$$\dot{\sigma} = \omega + \dot{\epsilon} \quad (11.23)$$

in which ω represents the inertially measurable angular motion and $\dot{\epsilon}$ is a component of $\dot{\sigma}$ owing to the non-inertially measurable motion, known as the non-commutativity rate vector.

An expression for $\dot{\sigma}$ under general motion conditions, that is, where the motion is not restricted to a single-axis, may be derived by differentiating eqn. (11.10), writing $d\mathbf{A}/dt = \mathbf{A}[\boldsymbol{\omega} \times]$ and manipulating vectors formed from the resulting expressions as described in Reference 1 to give:

$$\dot{\sigma} = \boldsymbol{\omega} + \frac{1}{2}\boldsymbol{\sigma} \times \boldsymbol{\omega} + \frac{1}{\sigma^2} \left[1 - \frac{\sigma \sin \sigma}{2(1 - \cos \sigma)} \right] \boldsymbol{\sigma} \times \boldsymbol{\sigma} \times \boldsymbol{\omega} \quad (11.24)$$

As a result of non-commutativity effects, the final orientation after a series of rotations is dependent on both the individual rotations and the order in which they have occurred. The above equation indicates how the history of previous rotations ($\boldsymbol{\sigma}$) and the current angular rate ($\boldsymbol{\omega}$) affect the build up of the non-commutativity term ($\dot{\epsilon}$).

A practical implementation would require the right-hand side of eqn. (11.24) to be truncated. For instance, writing the sine and cosine terms as series expansions and ignoring terms higher than third order in σ , the above equation may be written as

$$\dot{\sigma} = \boldsymbol{\omega} + \frac{1}{2}\boldsymbol{\sigma} \times \boldsymbol{\omega} + \frac{1}{12}\boldsymbol{\sigma} \times \boldsymbol{\sigma} \times \boldsymbol{\omega} \quad (11.25)$$

In Reference 3, the following algorithm is proposed:

$$\delta\alpha_{k+1} = \int_{t_k}^{t_{k+1}} \boldsymbol{\alpha} \times \boldsymbol{\omega} dt$$

where

$$\begin{aligned} \boldsymbol{\alpha} &= \int_{t_k}^t \boldsymbol{\omega} dt \\ \boldsymbol{\sigma} &= \boldsymbol{\alpha}_{k+1} + \delta\alpha_{k+1} \end{aligned} \quad (11.26)$$

and $\boldsymbol{\alpha}_{k+1}$ represents the sum of the incremental angle outputs provided by the gyroscopes from t_k to t_{k+1} . In general, where significant levels of angular vibration are present, it will be necessary to solve eqn. (11.26) at a higher rate, denoted the j -cycle update rate, than that at which the direction cosine matrix is updated. It will be necessary to select a j -cycle update rate which is sufficiently fast to ensure that the value of $\boldsymbol{\sigma}$ obtained using eqn. (11.26) agrees well with the true value, given by eqn. (11.24), in the presence of the maximum body rates and vibratory motion.

11.2.3 Rotation vector compensation

In this section, an expression is derived for the drift error in the computed attitude ($\delta\alpha$) arising in the presence of coning motion. Coning refers to the motion which arises when a single-axis of a body describes a cone, or some approximation to a cone, in space. Such motion results from the simultaneous application of angular oscillations about two orthogonal axes of the system where the oscillations are out of phase.

It is assumed here that the body is oscillating at a frequency f about the x and y axes. The amplitudes of the x and y motions are θ_x and θ_y , respectively. Additionally, a phase difference of ϕ is assumed to exist between the two channels. Thus, we may write:

$$\boldsymbol{\omega} = 2\pi f [\theta_x \cos 2\pi ft \quad \theta_y \cos(2\pi ft + \phi) \quad 0]^T$$

and

$$\boldsymbol{\alpha} = [\theta_x \{ \sin 2\pi ft - \sin 2\pi ft_k \} \quad \theta_y \{ \sin(2\pi ft + \phi) - \sin(2\pi ft_k + \phi) \} \quad 0]^T$$

Substituting for $\boldsymbol{\alpha}$ and $\boldsymbol{\omega}$ in the $\delta\boldsymbol{\alpha}$ equation gives:

$$\begin{aligned} \delta\boldsymbol{\alpha} &= \pi f \int_{t_k}^{t_{k+1}} \begin{bmatrix} \theta_x \{ \sin 2\pi ft - \sin 2\pi ft_k \} \\ \theta_y \{ \sin(2\pi ft + \phi) - \sin(2\pi ft_k + \phi) \} \\ 0 \end{bmatrix} \\ &\quad \times \begin{bmatrix} \theta_x \cos 2\pi ft \\ \theta_y \cos(2\pi ft + \phi) \\ 0 \end{bmatrix} dt \end{aligned}$$

which yields a z -component,

$$\delta\alpha_z = \pi f \theta_x \theta_y \sin \phi \int_{t_k}^{t_{k+1}} \{ 1 - \cos 2\pi f(t - t_k) \} dt$$

Integrating between the appropriate limits, we have:

$$\delta\alpha_z = \pi f \theta_x \theta_y \sin \phi \left[t_{k+1} - t_k - \frac{\sin 2\pi f(t_{k+1} - t_k)}{2\pi f} \right]$$

Writing $t_{k+1} - t_k = \delta t$ gives:

$$\delta\alpha_z = \pi f \theta_x \theta_y \delta t \sin \phi \left[1 - \frac{\sin 2\pi f \delta t}{2\pi f \delta t} \right] \quad (11.27)$$

Thus, although the rate $\boldsymbol{\omega}$ is cyclic about the x and y axes, a z -component of $\delta\boldsymbol{\alpha}$ arises, which is a constant proportional to the sine of the phase angle and the amplitude of the motion. It can be seen, from the above equation, that $\delta\boldsymbol{\alpha}$ is maximised when $\phi = \pi/2$. Under such conditions, the motion of the body is referred to as coning, owing to the motion of the z -axis which describes a cone as space.

Over the time interval $k\delta t$, the drift error in the computed attitude, which arises if the above correction term is not applied, may be expressed as:

$$\delta\dot{\alpha}_z = \pi f \theta_x \theta_y \sin \phi \left[1 - \frac{\sin 2\pi f \delta t}{2\pi f \delta t} \right] \quad (11.28)$$

If $\delta\dot{\alpha}_z$ is small compared with the overall system performance requirement, the need to implement the correction terms described is avoided.

11.2.3.1 Example

Consider a situation where the body exhibits coning motion at a frequency, f , of 50 Hz. The angular amplitudes of the motion in x and y is taken to be 0.1° . If the attitude update frequency is 100 Hz, that is, $\delta t = 0.01$ s, the resulting drift is $100^\circ/\text{h}$. By increasing the computational frequency to 500 Hz, the drift figure falls to $\sim 6^\circ/\text{h}$.

A more general development of the above equation allows an RMS value for $\delta\alpha_z$ to be calculated in the presence of a given vibration spectrum [3].

11.2.4 Body and navigation frame rotations

Returning now to the continuous form of the attitude eqn. (11.1), the vector $\boldsymbol{\omega}$ represents the turn rate of the body with respect to the navigation reference frame. When navigating with respect to the local geographic frame, this equation takes the following form, as discussed in Chapter 3:

$$\dot{\mathbf{C}}_b^n = \mathbf{C}_b^n \boldsymbol{\Omega}_{ib}^b - \boldsymbol{\Omega}_{in}^n \mathbf{C}_b^n \quad (11.29)$$

The first term, $\mathbf{C}_b^n \boldsymbol{\Omega}_{ib}^b$, is a function of the body rates, as sensed by the strapdown gyroscopes, whilst the second term, $-\boldsymbol{\Omega}_{in}^n \mathbf{C}_b^n$, is a function of the lower navigation frame rates. The updating of the direction cosine matrix to take account of the body motion, that is, the solution of the equation $d\mathbf{C}_b^n/dt = \mathbf{C}_b^n \boldsymbol{\Omega}_{ib}^b$ may be accomplished using eqns. 11.4 and 11.11, as described above.

A similar algorithm may be used to take account of navigation frame rotations. In order to update the direction cosine matrix for navigation frame rotations, an equation similar to eqn. (11.4) may be used, in which \mathbf{A} is replaced with a navigation frame rotation direction cosine matrix, \mathbf{B} , as follows:

$$\mathbf{C}_{l+1} = \mathbf{B}_l \mathbf{C}_l \quad (11.30)$$

where \mathbf{B}_l represents the direction cosine matrix relating navigation axes at time t_{l+1} to navigation axes at time t_l . \mathbf{B}_l may be expressed in terms of a rotation vector $\boldsymbol{\theta}$ as follows:

$$\mathbf{B}_l = \mathbf{I} + \frac{\sin \theta}{\theta} [\boldsymbol{\theta} \times] + \frac{(1 - \cos \theta)}{\theta^2} [\boldsymbol{\theta} \times]^2 \quad (11.31)$$

where $\boldsymbol{\theta}$ is a rotation vector with magnitude and direction such that a rotation of the navigation frame about $\boldsymbol{\theta}$, through an angle equal to the magnitude of $\boldsymbol{\theta}$, will rotate the navigation frame from its orientation at time t_l to its position at time t_{l+1} . The angle θ may be written as:

$$\boldsymbol{\theta} = \int_{t_l}^{t_{l+1}} \boldsymbol{\omega}_{in}^n dt \quad (11.32)$$

In view of the fact that the navigation frame turn rates will generally be much slower than the body rates, direction cosine updates for rotations of the navigation frame may be implemented at a much slower rate, denoted the l -cycle update rate. Additionally, the mathematical functions on which the algorithm is based can be truncated at a lower level.

11.2.4.1 Example

Considering first-order terms only in eqn. (11.31), and applying navigation frame updates at 1 s intervals to take account of rotation of the Earth, the net drift error can be maintained at a negligibly low level. In some short-range missile applications, in which angular rate measurement accuracies of several hundred degrees per hour may be acceptable, the need to take account of navigation frame rotations, for example, Earth's rate at 15°/h, becomes superfluous.

11.2.5 Quaternion algorithms

Using the quaternion attitude representation, it is required to solve the equation:

$$\dot{\mathbf{q}} = \frac{1}{2} \mathbf{q} \cdot \mathbf{p} \quad (11.33)$$

where $\mathbf{p} = [0, \boldsymbol{\omega}^T]$ as defined in Section 3.4.3.2. This equation may be expressed in matrix form as:

$$\dot{\mathbf{q}} = \frac{1}{2} \mathbf{W} \mathbf{q} \quad (11.34)$$

where

$$\mathbf{W} = \begin{bmatrix} 0 & -\omega_x & -\omega_y & -\omega_z \\ \omega_x & 0 & \omega_z & -\omega_y \\ \omega_y & -\omega_z & 0 & \omega_x \\ \omega_z & \omega_y & -\omega_x & 0 \end{bmatrix} \quad (11.35)$$

and ω_x , ω_y , and ω_z are the components of $\boldsymbol{\omega}$.

For the situation in which the orientation of the rate vector, $\boldsymbol{\omega}$, remains fixed over a computer update interval, the solution to the above equation may be written as:

$$\mathbf{q}_{k+1} = \left[\exp \frac{1}{2} \int_{t_k}^{t_{k+1}} \mathbf{W} dt \right] \mathbf{q}_k \quad (11.36)$$

$$\int_{t_k}^{t_{k+1}} \mathbf{W} dt = \Sigma = \begin{bmatrix} 0 & -\sigma_x & -\sigma_y & -\sigma_z \\ \sigma_x & 0 & \sigma_z & -\sigma_y \\ \sigma_y & -\sigma_z & 0 & \sigma_x \\ \sigma_z & \sigma_y & -\sigma_x & 0 \end{bmatrix} \quad (11.37)$$

to give

$$\mathbf{q}_{k+1} = \exp \left(\frac{\Sigma}{2} \right) \mathbf{q}_k \quad (11.38)$$

By expanding the exponential term and following a development similar to that used to obtain the direction cosine solution above, it can be shown that the exponential term may be written in quaternion form as:

$$\mathbf{q}_{k+1} = \mathbf{q}_k \cdot \mathbf{r}_k \quad (11.39)$$

where

$$\mathbf{r}_k = \begin{bmatrix} a_c \\ a_s \sigma_x \\ a_s \sigma_y \\ a_s \sigma_z \end{bmatrix} \quad (11.40)$$

$$a_c = \cos\left(\frac{\sigma}{2}\right) = 1 - \frac{(0.5\sigma)^2}{2!} + \frac{(0.5\sigma)^4}{4!} - \dots \quad (11.41)$$

$$a_s = \frac{\sin(\sigma/2)}{\sigma} = 0.5 \left(1 - \frac{(0.5\sigma)^2}{3!} + \frac{(0.5\sigma)^4}{5!} - \dots \right) \quad (11.42)$$

and

$$(0.5\sigma)^2 = 0.25(\sigma_x^2 + \sigma_y^2 + \sigma_z^2)$$

By comparison with the quaternion definition given in Chapter 3, eqn. (3.32), it can be seen that \mathbf{r}_k is a quaternion representing a rotation of magnitude σ , about a vector σ . This is the quaternion which transforms from body axes at time t_{k+1} to body axes at time t_k , whilst \mathbf{q}_k represents the quaternion relating body to navigation axes at time t_k . Therefore, the quaternion \mathbf{q} may be updated for body motion, as sensed by the strapdown gyroscopes, using eqns. (11.39)–(11.42) recursively. The parameter σ is determined as described in Section 11.2.1.

As for the direction cosine algorithm, the update interval is selected to ensure that σ remains small at the maximum body rate, thus avoiding the need to retain a large number of terms in the expressions for a_c and a_s . The order of the quaternion updating algorithm will be determined by the truncation point selected in eqns. 11.41 and 11.42.

11.2.5.1 Definition of attitude errors

In order to quantify the performance of the quaternion update algorithm, a drift parameter D_q is defined following a development parallel to that used in Section 11.2.1 to determine the drift error in the direction cosine update algorithm. The error in the computed quaternion $\delta\mathbf{r}$ may be expressed in terms of the true and computed quaternions, denoted \mathbf{r} and $\hat{\mathbf{r}}$, respectively, as follows:

$$\delta\mathbf{r} = \mathbf{r}^* \cdot \hat{\mathbf{r}} \quad (11.43)$$

For a single x -axis rotation, $\mathbf{r} = [\cos(0.5\sigma) \ \sin(0.5\sigma) \ 0 \ 0]$, $\hat{\mathbf{r}} = [a_c \ \sigma a_s \ 0 \ 0]$ and

$$\delta\mathbf{r} = \begin{bmatrix} a_c \cos(0.5\sigma) + \sigma a_s \sin(0.5\sigma) \\ \sigma a_s \cos(0.5\sigma) - a_c \sin(0.5\sigma) \\ 0 \\ 0 \end{bmatrix} \quad (11.44)$$

This may be expressed as a direction cosine error matrix in accordance with eqn. (3.59). Following the procedure used to define the direction cosine drift in Section 11.2.1, an expression for the quaternion drift may be defined in terms of the off-diagonal elements of the error matrix, viz:

$$\begin{aligned} D_q &= \frac{2}{\delta t} \{a_c \cos(0.5\sigma) + \sigma a_s \sin(0.5\sigma)\} \{\sigma a_s \cos(0.5\sigma) - a_c \sin(0.5\sigma)\} \\ &= \frac{1}{\delta t} \{2\sigma a_c a_s \cos \sigma - a_c^2 \sin \sigma + \sigma^2 a_s^2 \sin \sigma\} \end{aligned} \quad (11.45)$$

where a_c and a_s are selected according to the order of the algorithm needed as follows:

$$a_c = 1, a_s = 0.5 \quad \text{is a first-order algorithm,}$$

$$a_c = 1 - \frac{(0.5\sigma)^2}{2}, a_s = 0.5 \quad \text{is a second-order algorithm,}$$

$$a_c = 1 - \frac{(0.5\sigma)^2}{2}, a_s = 0.5 \left(1 - \frac{(0.5\sigma)^2}{6}\right) \quad \text{is a third-order algorithm.}$$

11.2.5.2 Example

The drift in the attitude computed using different orders of the quaternion algorithm is summarised in Table 11.2, using the same conditions considered in Section 11.2.1 to evaluate the performance of the direction cosine algorithm.

Comparing these results with those tabulated in Section 11.2.1, it can be seen that the quaternion drift figures are smaller than those obtained using direction cosines. This arises principally because the quaternion equations involve the expansion of terms in $\sin(\sigma/2)$ and $\cos(\sigma/2)$, whereas the direction cosine equations contain similar terms in σ . This also accounts for the correspondence between the quaternion drift figures which arise when σ is 0.1 rad and the direction cosine drifts when σ is 0.05 rad. It is therefore apparent that the quaternion representation yields the more accurate attitude solution, for a given level of truncation, in the presence of a single-axis rotation.

In order to update the quaternion for navigation frame rotations, an equation similar to eqn. (11.39) may be used, in which \mathbf{r} is replaced with a navigation frame

Table 11.2 Drift in the computed attitude using different orders of the quaternion algorithm

Order of algorithm	Attitude drift error ($^{\circ}/\text{h}$)	
	$\sigma_{\max} = 0.1 \text{ rad}$	$\sigma_{\max} = 0.05 \text{ rad}$
1	1720	430
2	860	215
3	0.4	0.06
4	0.2	0.06

rotation quaternion, \mathbf{p} , as follows:

$$\mathbf{q}_{l+1} = \mathbf{p}_l \cdot \mathbf{q}_l \quad (11.46)$$

where \mathbf{q}_l represents the quaternion relating body to navigation axes at computer cycle l , and \mathbf{p}_l is the quaternion which transforms from body axes at time t_{l+1} to body axes at time t_l . The quaternion \mathbf{p}_l may be expressed in terms of the rotation vector $\boldsymbol{\theta}$, as follows:

$$\mathbf{p}_l = \begin{bmatrix} b_c \\ b_s\theta_x \\ b_s\theta_y \\ b_s\theta_z \end{bmatrix} \quad (11.47)$$

where

$$b_c = \cos\left(\frac{\theta}{2}\right) = 1 - \frac{(0.5\theta)^2}{2!} + \frac{(0.5\theta)^4}{4!} - \dots \quad (11.48)$$

$$b_s = \frac{\sin(\theta/2)}{\theta} = 0.5 \left(1 - \frac{(0.5\theta)^2}{3!} + \frac{(0.5\theta)^4}{5!} - \dots\right) \quad (11.49)$$

As in the case of the direction cosines, it is assumed that these equations may be implemented at the reduced rate, the l -cycle, compared with the body motion updates. Additionally, a low order algorithm will usually be sufficient to provide accurate navigation frame updates.

In the situation where the quaternion attitude representation is selected, the following equation may be used to compute the direction cosine matrix \mathbf{C}_b^n for use in the acceleration vector transformation algorithm, which is discussed in Section 11.3.

$$\mathbf{C}_b^n = \begin{pmatrix} 1 - 2(c^2 + d^2) & 2(bc - ad) & 2(bd + ac) \\ 2(bc + ad) & 1 - 2(b^2 + d^2) & 2(cd - ab) \\ 2(bd - ac) & 2(cd + ab) & 1 - 2(b^2 + c^2) \end{pmatrix} \quad (11.50)$$

11.2.6 Orthogonalisation and normalisation algorithms

It is common practice, in the implementation of strapdown attitude algorithms, to apply self-consistency checks in an attempt to enhance the accuracy of the computed direction cosines or quaternion parameters. The rows of the direction cosine matrix represent the projection of unit vectors which lie along each axis of the orthogonal reference co-ordinate frame in the body frame. It follows therefore, that the rows of the direction cosine matrix should always be orthogonal to one another and that the sum of the squares of the elements in each row should equal unity. In the case of the quaternion representation, the self-consistency test is to check that the sum of the squares of the four parameters is unity.

Self-consistency checks may be applied as part of the attitude algorithm to ensure that the above conditions are satisfied. If required, it is usually sufficient to carry out these checks at a relatively low rate. This part of the computation may be carried out at the l -cycle frequency.

11.2.6.1 Direction cosine checking

The condition for orthogonality of the i th and j th rows of the direction cosine matrix, denoted \mathbf{C}_i and \mathbf{C}_j , is that their dot product should equal zero, that is, $\mathbf{C}_i \mathbf{C}_j^T = 0$. In practice, this is not necessarily the case, and we define:

$$\Delta_{ij} = \mathbf{C}_i \mathbf{C}_j^T \quad (11.51)$$

where Δ_{ij} is an angle error defined about an axis perpendicular to \mathbf{C}_i and \mathbf{C}_j , the orthogonality error between the two rows.

Since either row, \mathbf{C}_i or \mathbf{C}_j , is equally likely to be in error, the correction is apportioned equally between them using:

$$\hat{\mathbf{C}}_i = \mathbf{C}_i - \frac{1}{2} \Delta_{ij} \mathbf{C}_j \quad (11.52)$$

$$\hat{\mathbf{C}}_j = \mathbf{C}_j - \frac{1}{2} \Delta_{ij} \mathbf{C}_i \quad (11.53)$$

where the \wedge notation denotes the corrected quantity.

Normalisation errors may be identified by comparing the sum of the squares of the elements in a row with unity, that is,

$$\Delta_{ii} = 1 - \mathbf{C}_i \mathbf{C}_i^T \quad (11.54)$$

and corrections applied using:

$$\hat{\mathbf{C}}_i = \mathbf{C}_i - \frac{1}{2} \Delta_{ii} \mathbf{C}_i \quad (11.55)$$

An alternative approach is to operate on the columns of the direction cosine matrix, in a similar manner to that described above for the rows, as described in Reference 3.

11.2.6.2 Quaternion normalisation

The quaternion can be normalised by comparing the sum of the squares of its elements with unity. The normalisation error is given by:

$$\Delta q = 1 - \mathbf{q} \cdot \mathbf{q}^* \quad (11.56)$$

The quaternion may be normalised by dividing each element by $\sqrt{(\mathbf{q} \cdot \mathbf{q}^*)}$. Thus, we may write:

$$\begin{aligned}\mathbf{q} &= \frac{\mathbf{q}}{\sqrt{(\mathbf{q} \cdot \mathbf{q}^*)}} \\ &= \{1 - \Delta q\}^{-0.5} \mathbf{q} \\ &\approx \left\{1 + \frac{1}{2}\Delta q\right\} \mathbf{q}\end{aligned}\tag{11.57}$$

It should be noted that the orthogonalisation and normalisation cannot correct for errors that have occurred in the previous computation. For example, an error arising in a single element of the quaternion will be ‘spread’ amongst all of the elements, as a result of a normalisation correction. In the opinion of the authors, such techniques should be used with caution. Emphasis should be placed on the design of the basic attitude update algorithm, rather than placing any reliance on the normalisation process described here, which may serve only to compound fundamental errors in the update algorithm.

11.2.7 The choice of attitude representation

The relative benefits of using direction cosines or quaternion parameters to represent attitude has received considerable attention in the published literature [3–5]. The results have largely been found to be inconclusive, although the quaternion method does offer some advantage because, inherently, it gives rise to an orthogonal attitude matrix. In addition, the analysis, given in the preceding sections, has shown the accuracy of attitude computation obtained using quaternions to be superior to the accuracy which may be achieved using the direction cosine representation. These factors account, in part at least, for the popularity of the quaternion method in recent years.

However, the final selection of attitude algorithm is unlikely to be made solely on the grounds of computational accuracy. It is the computing burden and the memory requirements which will be the major factors in determining the method to be adopted for a particular application. In this context, there is still a tendency to opt for the quaternion approach, primarily because there are fewer parameters to update, four quaternions as opposed to nine direction cosines. However, when account is taken of the overall strapdown computing task, including the resolution of the measured specific force vector, the relative benefits of the two methods become less clear cut.

11.3 Acceleration vector transformation algorithm

This section is concerned with the computational algorithm required to resolve the measured specific force acceleration components into the navigation reference frame. Care must be taken to allow for changes in body rotation occurring over a computer update interval and a two-speed algorithm may be required for applications where

the system is called upon to operate in a highly vibratory environment [6]. For many applications, a relatively low speed algorithm may be sufficient to resolve accelerations associated with vehicle manoeuvres, whilst a high speed correction term may be included to take account of vibration.

11.3.1 Acceleration vector transformation using direction cosines

The measured specific force vector, \mathbf{f}^b , is expressed in the navigation co-ordinate frame using:

$$\mathbf{f}^n = \mathbf{C}_b^n \mathbf{f}^b \quad (11.58)$$

where \mathbf{C}_b^n is the direction cosine matrix which transforms from body to reference axes, as defined earlier.

An algorithm to implement this function, which accepts incremental velocity measurements, may be developed by first integrating eqn. (11.58) to give:

$$\mathbf{u}^n = \int_{t_k}^{t_{k+1}} \mathbf{f}^n dt \quad (11.59)$$

where \mathbf{u}^n represents the change in velocity, expressed in the navigation frame, over the computer cycle t_k to t_{k+1} . The velocity vector, \mathbf{v}^n , may be determined by summing the values of \mathbf{u}^n computed at each cycle, and correcting for Coriolis accelerations and the effects of gravity as described later, in Section 11.4. The matrix \mathbf{C}_b^n varies continuously with time over the update interval and may be written in terms of the matrix \mathbf{C}_k , the value of \mathbf{C}_b^n at time t_k , and \mathbf{A} , a matrix representing the transformation from body axes at time t to body axes at the start of the update interval, t_k , that is,

$$\mathbf{C}_b^n = \mathbf{C}_k \mathbf{A} \quad (11.60)$$

Substituting for \mathbf{C}_b^n in eqn. (11.58) gives:

$$\mathbf{u}^n = \mathbf{C}_k \int_{t_k}^{t_{k+1}} \mathbf{A} \mathbf{f}^b dt \quad (11.61)$$

Following eqn. (11.11), \mathbf{A} may be expressed in the form:

$$\mathbf{A} = \mathbf{I} + [\boldsymbol{\alpha} \times] + 0.5[\boldsymbol{\alpha} \times]^2 - \dots$$

where

$$\boldsymbol{\alpha} = \int_{t_k}^t \boldsymbol{\omega}^b dt$$

Substituting for \mathbf{A} in eqn. (11.61) gives:

$$\mathbf{u}^n = \mathbf{C}_k \int_{t_k}^{t_{k+1}} [\mathbf{f}^b + \boldsymbol{\alpha} \times \mathbf{f}^b + 0.5[\boldsymbol{\alpha} \times]^2 \mathbf{f}^b - \dots] dt \quad (11.62)$$

If second- and higher-order terms are ignored, we may write:

$$\mathbf{u}^n = \mathbf{C}_k \left[\int_{t_k}^{t_{k+1}} \mathbf{f}^b dt + \int_{t_k}^{t_{k+1}} \boldsymbol{\alpha} \times \mathbf{f}^b dt \right] \quad (11.63)$$

If we now write:

$$\mathbf{v} = \int_{t_k}^t \mathbf{f}^b dt$$

and integrate the cross product term in eqn. (11.63) by parts, it can be shown that

$$\mathbf{u}^n = \mathbf{C}_k \left(\mathbf{v}_{k+1} + \frac{1}{2} \boldsymbol{\alpha}_{k+1} \times \mathbf{v}_{k+1} + \frac{1}{2} \int_{t_k}^{t_{k+1}} (\boldsymbol{\alpha} \times \mathbf{f}^b - \boldsymbol{\omega}^b \times \mathbf{v}) dt \right) \quad (11.64)$$

where $\boldsymbol{\alpha}_{k+1} = \boldsymbol{\alpha}$ evaluated over the time interval t_k to t_{k+1} , by summing the incremental angle measurements over this period. Similarly, $\mathbf{v}_{k+1} = \mathbf{v}$ evaluated over the same time interval, by summing the incremental velocity measurements provided by the inertial measurement unit.

In eqn. (11.64), we have an expression for \mathbf{u}^n containing three terms:

- the sum of the incremental velocity measurements produced by the inertial measurement unit;
- a cross product of the incremental angle, accumulated from t_k to t_{k+1} , and the incremental velocity change over the same period. In the literature, this term is referred to as the rotation correction;
- a dynamic integral term.

In the situation where \mathbf{f}^b and $\boldsymbol{\omega}^b$ remain constant over the update interval, we have $\boldsymbol{\alpha} = \boldsymbol{\omega}^b t$ and $\mathbf{v} = \mathbf{f}^b t$. Substituting for $\boldsymbol{\alpha}$ and \mathbf{v} in eqn. (11.48), it can be shown easily that the integral term becomes identically zero under such conditions. However, if \mathbf{f}^b and $\boldsymbol{\omega}^b$ vary significantly over the update interval, it may be necessary to evaluate the integral in order to provide compensation for the dynamic motion. In this case, the rate at which the integral is evaluated will need to be well above the frequency of the dynamic motion, that is, at the j -cycle update rate referred to in the previous section. An example to illustrate this effect is given in the following section.

11.3.2 Rotation correction

In order to illustrate the need for the rotation correction term in a missile application using strapdown technology, consider a situation in which such a vehicle is manoeuvring in a lateral plane such that, $\mathbf{f}^b = [0 \ f \ 0]^T$ and $\boldsymbol{\omega}^b = [0 \ 0 \ \omega]^T$, that is, simultaneously accelerating and rotating in the yaw plane.

If the acceleration transformation update interval is δt , then, following eqn. (11.61), the true velocity change over the time interval from $t = 0$ to $t = \delta t$ is

given by:

$$\begin{aligned}\mathbf{u}^n &= \int_0^{\delta t} \begin{bmatrix} \cos \omega t & \sin \omega t & 0 \\ -\sin \omega t & \cos \omega t & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 0 \\ f \\ 0 \end{bmatrix} dt \\ &= \int_0^{\delta t} \begin{bmatrix} f \sin \omega t \\ f \cos \omega t \\ 0 \end{bmatrix} dt \\ &= \left[\frac{f}{\omega} (1 - \cos \omega \delta t) \quad \frac{f}{\omega} \sin \omega \delta t \quad 0 \right]^T\end{aligned}$$

If no allowance is made for the vehicle's body rotation during the update interval, the computed velocity change simply becomes:

$$\mathbf{u}^{n'} = [0 \quad f \delta t \quad 0]^T$$

Differencing the expressions for $\mathbf{u}^{n'}$ and \mathbf{u}^n gives:

$$\begin{aligned}\delta \mathbf{u}^n &= \mathbf{u}^{n'} - \mathbf{u}^n \\ &= \left[\frac{f}{\omega} (1 - \cos \omega \delta t) \quad -f \delta t \frac{(1 - \sin \omega \delta t)}{\omega \delta t} \quad 0 \right]^T \\ &= \left[\frac{f \omega \delta t^2}{2} - \dots \quad \frac{f \omega^2 \delta t^3}{6} - \dots \quad 0 \right]^T\end{aligned}\tag{11.65}$$

An equivalent acceleration error may be written as:

$$\frac{\delta \mathbf{u}^n}{\delta t} = \left[\frac{f \omega \delta t}{2} - \dots \quad \frac{f \omega^2 \delta t^2}{6} - \dots \quad 0 \right]^T\tag{11.66}$$

11.3.2.1 Example

Take the case of a missile which is accelerating laterally at $20g$, whilst travelling at 800 m/s . The associated turn rate is $\sim 0.25 \text{ rad/s}$. Substituting for f and ω in the above expression, and assuming an update interval of 0.01 s , gives an x -component of acceleration bias of $25 \text{ milli-}g$. In the presence of a sustained missile manoeuvre, a bias of this magnitude would give rise to significant velocity and position errors, particularly in the case of medium to long-range missile systems. The figure derived above may be put into context by comparing it with the magnitudes of acceleration measurement error which may be acceptable for such applications, typically $\sim 5\text{--}10 \text{ milli-}g$, sometimes less.

It can be shown that the simple rotation correction term, given in eqn. (11.64), is able to compensate for the second-order velocity error which appears in the above example. Higher-order corrections may be applied if considered necessary, the relevant expressions being derived by including extra terms in the expansion of the transformation matrix \mathbf{A} , by substituting in eqn. (11.61).

11.3.3 Dynamic correction

In calculating the change in velocity over a computer cycle, expressed in the navigation co-ordinate frame, the following correction term may need to be applied:

$$\delta\mathbf{u}^n = \frac{1}{2} \int_{t_k}^{t_{k+1}} (\boldsymbol{\alpha} \times \mathbf{a}^b - \boldsymbol{\omega}^b \times \mathbf{v}) dt \quad (11.67)$$

A good test of the performance of the vector transformation algorithm is its ability to cope with sculling motion. Sculling is characterised by the simultaneous application of in-phase components of angular and linear oscillatory motion with respect to two orthogonal axes. Such motion can be particularly detrimental to system performance if the computational frequency is too low, or if sculling corrections are not applied.

An expression for $\delta\mathbf{u}^n$ is now derived under conditions where the body performs an angular oscillation about the x -axis, whilst simultaneously oscillating linearly in the y direction. Writing

$$\boldsymbol{\alpha} = \int_{t_k}^t \boldsymbol{\omega}^b dt$$

where

$$\boldsymbol{\omega}^b = [2\pi f \theta_x \cos 2\pi ft \quad 0 \quad 0]^T$$

and

$$\mathbf{v} = \int_{t_k}^t \mathbf{a}^b dt$$

where

$$\mathbf{a}^b = [0 \quad A_y \sin(2\pi ft + \phi) \quad 0]^T$$

in which A_y is the amplitude of the cyclic acceleration. Substituting in the equation for $\delta\mathbf{u}^n$ yields a z -component

$$\begin{aligned} \delta\mathbf{u}_z^n &= \frac{1}{2} \theta_x A_y \cos \phi \int_{t_k}^{t_{k+1}} \{1 - \cos 2\pi f(t - t_k)\} dt \\ &= \frac{1}{2} \theta_x A_y \cos \phi \left(t_{k+1} - t_k - \frac{\sin 2\pi f(t_{k+1} - t_k)}{2\pi f} \right) \end{aligned} \quad (11.68)$$

Writing $t_{k+1} - t_k = \delta t$ gives

$$\delta\mathbf{u}_z^n = \frac{1}{2} \theta_x A_y \delta t \cos \phi \left(1 - \frac{\sin 2\pi f \delta t}{2\pi f \delta t} \right) \quad (11.69)$$

It can be seen that $\delta\mathbf{u}_z^n$ is maximised when $\phi = 0$. Under such conditions, the motion of the body is referred to as sculling. Over the time interval δt , an effective acceleration error arises if the above correction term is not applied. This error may be expressed as

$$\delta\dot{\mathbf{u}}_z^n = \frac{1}{2} \theta_x A_y \cos \phi \left(1 - \frac{\sin 2\pi f \delta t}{2\pi f \delta t} \right) \quad (11.70)$$

If $\delta\dot{\mathbf{u}}_z^n$ is small compared with the overall system performance requirement, the need to implement the correction term is avoided.

11.3.3.1 Example

Consider a situation in which the body exhibits sculling motion at a frequency, f , of 50 Hz. The angular amplitude, θ_x , of the motion is taken to be 0.1° , whilst the cyclic acceleration, with respect to an orthogonal axis, is taken to vary between $\pm A_y$ where $A_y = 10g$.

If the attitude update frequency is 100 Hz, that is, $\delta t = 0.01$ s, the resulting acceleration error is 8.7 milli-g. By increasing the computational frequency to 500 Hz, the magnitude of the error is reduced to ~ 0.5 milli-g, the required precision being determined by the actual application, and possibly, its particular phase of flight.

11.3.4 Acceleration vector transformation using quaternions

In cases where the attitude is computed in quaternion form, the acceleration vector transformation may be effected using the quaternion parameters directly, using the following equation in place of eqn. (11.64):

$$\mathbf{u}^n = \mathbf{q} \cdot \left(\mathbf{v}_{k+1} + \frac{1}{2} \boldsymbol{\alpha}_{k+1} \times \mathbf{v}_{k+1} + \frac{1}{2} \int_{t_k}^{t_{k+1}} (\boldsymbol{\alpha} \times \mathbf{f}^b - \boldsymbol{\omega}^b \times \mathbf{v}) dt \right) \cdot \mathbf{q}^* \quad (11.71)$$

Alternatively, the direction cosine matrix, \mathbf{C}_b^n , may be calculated from the quaternion parameters, using eqn. (11.50), and the acceleration transformation performed as described in Section 11.3.1.

11.4 Navigation algorithm

The computational processes required to determine vehicle velocity and position are not unique to strapdown systems and are described in many of the standard inertial navigation texts. More recently, Bar-Itzhack [7] has considered the benefits of using different computational rates for different parts of the navigation computation. For example, terms involving Earth's rate do not need to be evaluated as often as terms which are functions of the body rate. Such considerations are also taken account of in the analysis which follows.

The velocity and position equations, given in Chapter 3, may be expressed in integral form as follows:

$$\mathbf{v}^n = \int_0^t \mathbf{f}^n dt - \int_0^t [2\boldsymbol{\omega}_{ie} + \boldsymbol{\omega}_{en}] \times \mathbf{v}^n dt + \int_0^t \mathbf{g} dt \quad (11.72)$$

$$\mathbf{x}^n = \int_0^t \mathbf{v}^n dt \quad (11.73)$$

It is required to evaluate the integral terms within the navigation processor, in order to determine vehicle velocity and position. The vectors in the above equations may be expressed in component form, as follows:

$$\mathbf{v}^n = [v_N \quad v_E \quad v_D]^T$$

$$\mathbf{x}^n = [x_N \quad x_E \quad -h]^T$$

$$\boldsymbol{\omega}_{ie} = [\Omega \cos L \quad 0 \quad -\Omega \sin L]^T$$

$$\boldsymbol{\omega}_{en} = \left[\frac{v_E}{R_0 + h} \quad \frac{-v_N}{R_0 + h} \quad \frac{-v_E \tan L}{R_0 + h} \right]^T$$

$$\mathbf{g} = [0 \quad 0 \quad g]^T$$

The above expressions apply for navigation in the vicinity of the Earth, in the local vertical geographic frame. The first integral term in eqn. (11.72) represents the sum of the velocity changes over each update cycle, \mathbf{u}^n :

$$\mathbf{u}^n = \int_{t_k}^{t_{k+1}} \mathbf{f}^n dt$$

\mathbf{u}^n may be determined using the eqn. (11.48), developed in the previous section. Since this term is a function of vehicle body attitude, $\mathbf{f}^n = \mathbf{C}_b^n \mathbf{f}^b$, it must be calculated at a rate which is high enough to take account of vehicle dynamic motion. Including the gravity contribution, the velocity vector may be updated over the time interval t_k to t_{k+1} using:

$$\mathbf{v}_{k+1}^n = \mathbf{v}_k^n + \mathbf{u}^n + \mathbf{g}\delta t$$

The second integral term in eqn. (11.50) includes the Coriolis correction. In general, the contribution to \mathbf{v}^n is small compared with the other terms in the equation. Since the rate of change of the magnitude of the Coriolis term will be relatively low, it is sufficient to apply this correction at the relatively low l -update rate, referred to in Section 11.3, using a fairly simple algorithm, viz:

$$\mathbf{v}_{l+1}^n = [\mathbf{I} - 2\boldsymbol{\Omega}_{ie}\delta t_l - \boldsymbol{\Omega}_{en}\delta t_l] \mathbf{v}_l^n \quad (11.74)$$

where

$$\boldsymbol{\Omega}_{ie} = [\boldsymbol{\omega}_{ie} \times],$$

$$\boldsymbol{\Omega}_{en} = [\boldsymbol{\omega}_{en} \times],$$

\mathbf{v}_l^n = velocity vector at time t_l ,

$$\delta t_l = t_{l+1} - t_l.$$

Finally, position may be derived by integrating the velocity vector, as shown in eqn. (11.73).

The choice of integration scheme will, of course, be dependent on the application. For relatively short range, low accuracy, applications, a low order scheme such as

rectangular or trapezoidal integration is likely to be adequate. Equations for updating position over the time interval t_k to t_{k+1} are shown below.

Rectangular integration:

$$\mathbf{x}_{k+1}^n = \mathbf{x}_k^n + \mathbf{v}_k^n \delta t \quad (11.75)$$

Trapezoidal integration:

$$\mathbf{x}_{k+1}^n = \mathbf{x}_k^n + \left(\frac{\mathbf{v}_k^n + \mathbf{v}_{k+1}^n}{2} \right) \delta t \quad (11.76)$$

For aircraft and ship-borne inertial system applications, in which the performance requirements are more demanding, a higher order integration scheme such as Simpson's rule or fourth-order Runge-Kutta integration [8] may be needed.

Simpson's rule:

$$\mathbf{x}_{k+1}^n = \mathbf{x}_k^n + \left(\frac{\mathbf{v}_{k-1}^n + 4\mathbf{v}_k^n + \mathbf{v}_{k+1}^n}{3} \right) \delta t \quad (11.77)$$

It is often required to determine position relative to the Earth in terms of height above the Earth and the angular orientation of the current local vertical navigation reference frame with respect to the Earth frame; commonly expressed as latitude and longitude. To avoid mathematical singularities, the angular position parameters may be expressed in the form of a direction cosine matrix relating the navigation (n) and Earth (e) frames. The position direction cosine matrix and height propagate in accordance with the following differential equations:

$$\dot{h} = -\mathbf{v}_D \quad (11.78)$$

$$\dot{\mathbf{C}}_n^e = \mathbf{C}_n^e \boldsymbol{\Omega}_{en}^n \quad (11.79)$$

where $\boldsymbol{\Omega}_{en}^n = [\omega_{en}^n \times]$ and $\omega_{en}^n = [\dot{\ell} \cos L \ -\dot{L} \ -\dot{\ell} \sin L]^T$ as described in Chapter 3.

An algorithm for implementation in the navigation software can be formulated based on the integral of the above equation.

$$h_l = h_{l-1} + \Delta h_l \quad (11.80)$$

$$\mathbf{C}_{nl+1}^e = \mathbf{C}_{nl}^e \Delta \mathbf{C}_{l+1}^l \quad (11.81)$$

where $\Delta \mathbf{C}_{l+1}^l$ represents the direction cosine matrix relating navigation axes at time t_{l+1} to navigation axes at time t_l . $\Delta \mathbf{C}_{l+1}^l$ may be expressed in terms of a rotation vector $\boldsymbol{\theta}$ as follows:

$$\Delta \mathbf{C}_{l+1}^l = \mathbf{I} + \frac{\sin \zeta}{\zeta} [\zeta \times] + \frac{(1 - \cos \zeta)}{\zeta^2} [\zeta \times]^2 \quad (11.82)$$

where ζ is a rotation vector with magnitude and direction such that a rotation of the navigation frame about ζ , through an angle equal to the magnitude of ζ , will rotate the navigation frame from its orientation at time t_l to its position at time t_{l+1} .

$$\zeta \approx \int_{t_l}^{t_{l+1}} \boldsymbol{\omega}_{en}^n dt \quad (11.83)$$

To achieve precision updating of the rotation angle vector ζ in the presence of rotations of the position vector from computer cycle l to $l + 1$, an algorithm of the form used for velocity updating may be used; as described by Savage [9]. The resulting algorithm includes terms to compensate for the effects of position vector rotation and the combined dynamic effect of angular rate and specific force which gives rise to rectification errors. The term 'scrolling' has been used to refer to the dynamic error effect [11, 12], which is analogous to the sculling effect which affects the updating of the velocity vector under dynamic conditions.

11.5 Summary

Strapdown navigation computation involves the determination of vehicle attitude, velocity and position from measurements of angular rate and specific force, obtained from a set of inertial instruments rigidly mounted in the vehicle. Measurements made by the inertial sensors are used in various equations to provide the desired navigation information. In the foregoing discussion, three distinct areas of strapdown computation have been described, namely, attitude computation, transformation of the specific force vector and navigation computation. Techniques have been outlined which may be used to implement each of these functions, in real time, together with some analysis of their application.

In this chapter, the equations to be implemented in an inertial navigation system processor have been described. Algorithms may be developed, based on the equations given, which will accept and process inertial measurements in incremental form directly. In order to achieve the desired real time solution of the various algorithms, it is proposed that the strapdown computation should be split into low-, medium- and high-frequency sections as follows:

Low-frequency computation (l-cycle). Certain parts of the strapdown navigation equations involve terms which vary very slowly with time, 'Earth's rate' terms for instance. It is therefore sufficient to implement these sections of the algorithms at relatively low rates. In particular, the updating of the computed attitude for navigation frame rotations, and the application of Coriolis corrections to the navigation calculations, may be carried out at the low rate. In addition, algorithms for attitude orthogonalisation and normalisation, if required, may be implemented at the low rate.

Medium-frequency computation (k-cycle). It is intended that the bulk of the computation should be implemented at the medium rate which will need to be selected to cope with the large amplitude dynamic motion, arising as a result of vehicle manoeuvres. This includes the updating of the quaternions or direction cosines, the transformation of the acceleration vector and the solution of the major part of the navigation equations.

High-frequency computation (j-cycle). In order to cope with vibratory motion, coning and sculling, for instance, it may be necessary to carry out some relatively simple calculations at high speed. This allows compensation for variations in the applied

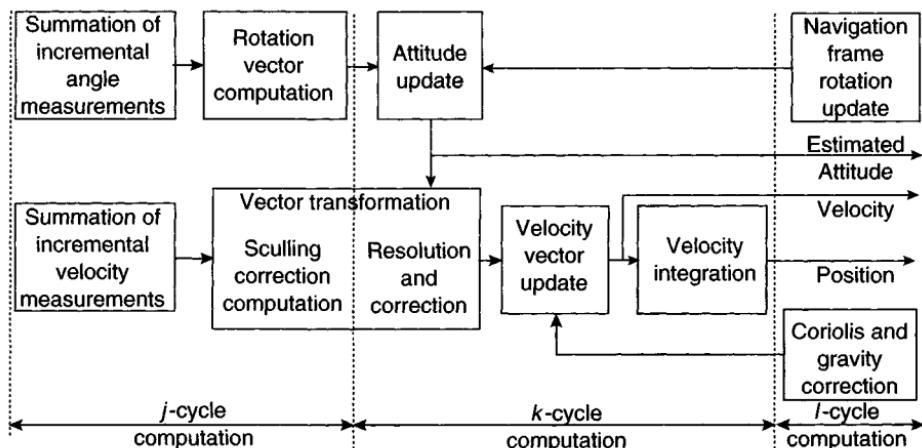


Figure 11.2 Summary of strapdown computational tasks

angular rate and linear acceleration occurring between the updates which take place at a lower frequency.

The various computational functions to be implemented are summarised in Figure 11.2. It is assumed that the inertial measurement unit will be capable of providing outputs at intervals, consistent with the highest computational frequency which is required.

In modern strapdown systems, emphasis is placed on the use of more precise algorithms based, wherever possible, on closed-form solutions of the analytically exact integral solutions to the attitude and navigation differential equations described in this chapter. Resulting improvements in computational precision allow the strapdown algorithms to be validated more easily, since test results can be expected to be in good agreement with analytical solutions under defined motion conditions. The general use of more precise algorithms has become possible through the application of modern computer technology, in which fast processing speed and long floating point word lengths combine to allow such algorithms to be implemented. The reader wishing to gain greater insight into the mathematical derivation of modern strapdown algorithms and processing techniques is referred to the various publications by Paul Savage who has worked and written extensively on this subject [9–12].

References

- 1 BORTZ, J.E.: 'A new mathematical formulation for strapdown inertial navigation', *IEEE Transactions on Aerospace and Electronic Systems*, 1971, **AES-7** (1)
- 2 MILLER, R.B.: 'A new strapdown attitude algorithm', *Journal of Guidance, Control and Dynamics*, 1983, **6**

- 3 SAVAGE, P.: 'Strapdown system algorithms', AGARD Lecture Series No. 133, 'Advances in strapdown inertial systems', May 1984
- 4 NASA: 'A study of the critical computational problems associated with strapdown inertial navigation systems', NASA CR-968
- 5 MORTENSEN, R.E.: 'Strapdown guidance error analysis', *IEEE Transactions on Aerospace and Electronic Systems*, 1974 **AES-10** (4)
- 6 LEVINSON, E.: 'Laser gyro strapdown inertial system applications'. AGARD Lecture Series No. 95, 'Strapdown inertial systems', June 1978
- 7 BAR-ITZHACK, I.Y.: 'Navigation computation in terrestrial strapdown inertial navigation systems', *IEEE Transactions on Aerospace and Electronic Systems*, 1977, **AES-13** (6)
- 8 JEFFREY, A.: 'Mathematics for engineers and scientists' (Nelson, 1979)
- 9 SAVAGE, P.G.: 'Strapdown analytics', Strapdown Associates, Inc.
- 10 SAVAGE, P.G.: 'Strapdown inertial navigation integration algorithm design, Part 1: attitude algorithms', *AIAA Journal of Guidance, Control and Dynamics*, 1998, **21** (1), pp. 19–28
- 11 SAVAGE, P.G.: 'Strapdown inertial navigation integration algorithm design, Part 2: velocity and position algorithms', *AIAA Journal of Guidance, Control and Dynamics*, 1998, **21** (2), pp. 208–221
- 12 SAVAGE, P.G.: 'Strapdown system computational elements'. NATO RTO Lecture Series No. 232, October 2003

Chapter 12

Generalised system performance analysis

12.1 Introduction

In a practical implementation, the accuracy to which a strapdown inertial navigation system is able to operate is limited as a result of errors in the data which are passed to it prior to the commencement of navigation, as well as imperfections in the various components which combine to make up the system. The sources of error may be categorised as follows:

- initial alignment errors;
- inertial sensor errors;
- computational errors.

Many of the contributions to the errors in these different categories have been described in Chapters 4–7, 10 and 11.

Any lack of precision in a measurement used in a dead reckoning system such as an inertial navigation system is passed from one estimate to the next with the overall uncertainty in the precision of the calculated quantity varying or drifting with time. In general, inertial navigation system performance is characterised by a growth in the navigation error from the position co-ordinate values which are initially assigned to it. It is common practice to refer to an inertial navigation system in terms of its mean drift performance; a one nautical mile per hour system is a typical performance class of a system. This would be typical of an inertial navigation system used in a commercial aircraft.

During the early stages of design and system specification, it is necessary to estimate navigation system performance under the conditions in which that system will be called upon to operate. A combination of analysis and simulation techniques are commonly used to predict system performance. In this chapter, equations are derived relating system performance to sources of error. These equations are

used to illustrate the propagation of the various types of error with time. Later in the chapter, the limitations of the analysis techniques are highlighted and an outline of simulation methods which may be used to assess system performance is presented.

12.2 Propagation of errors in a two-dimensional strapdown navigation system

12.2.1 Navigation in a non-rotating reference frame

The manner in which errors propagate in a strapdown inertial navigation system is discussed first in the context of the simple, two-dimensional, navigator considered at the start of Chapter 3, and illustrated in Figure 3.1. An error block diagram of this system is given in Figure 12.1.

The figure shows errors in the initial values of position, velocity and attitude as well as biases on the measurements of specific force and angular rate provided by the inertial sensors. For the purposes of simplifying this initial analysis, imperfections in the representation of the gravitational field have been ignored.

These errors propagate throughout the system giving rise to position errors which increase with time. The propagation of the errors can be represented in mathematical form as a set of differential equations which are derived from the system equations given in Chapter 3 by taking partial derivatives. The differential equations for the two-dimensional navigation system, correct to first order in the various system errors, are shown below.

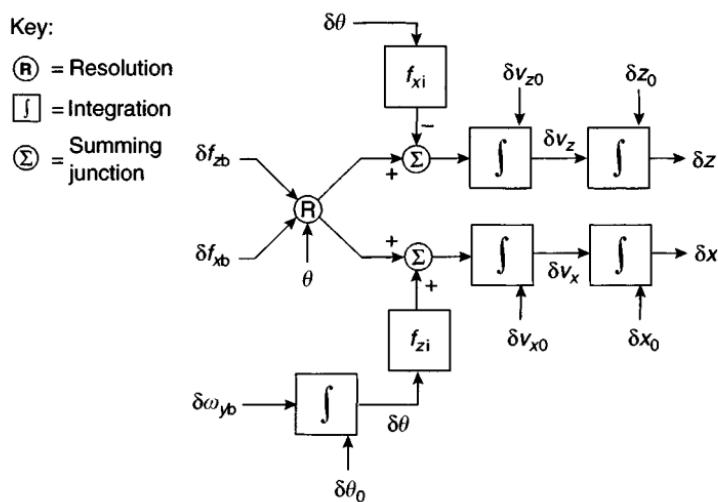


Figure 12.1 Error block diagram of two-dimensional inertial navigation system

Navigation error equations

$$\begin{aligned}
 \dot{\delta\theta} &= \delta\omega_{yb} \\
 \delta f_{xi} &= (-f_{xb} \sin \theta + f_{zb} \cos \theta) \delta\theta + \delta f_{xb} \cos \theta + \delta f_{zb} \sin \theta \\
 &= f_{zi} \delta\theta + \delta f_{xb} \cos \theta + \delta f_{zb} \sin \theta \\
 \delta f_{zi} &= -(f_{xb} \cos \theta + f_{zb} \sin \theta) \delta\theta - \delta f_{xb} \sin \theta + \delta f_{zb} \cos \theta \\
 &= -f_{xi} \delta\theta - \delta f_{xb} \sin \theta + \delta f_{zb} \cos \theta \\
 \delta \dot{v}_{xi} &= \delta f_{xi} \quad \delta \dot{v}_{zi} = \delta f_{zi} \quad \delta \dot{x}_i = \delta v_{xi} \quad \delta \dot{z}_i = \delta v_{zi}
 \end{aligned} \tag{12.1}$$

Consider now the position errors resulting from the various error sources. An initial error in the estimation of position simply contributes a constant off-set to the estimated position which remains fixed during the period the system is navigating, whilst initial velocity errors are integrated to induce position errors which increase linearly with time. The effects of attitude errors and instrument biases on system performance are a little more complex, since individual errors will, in general, affect both channels of the navigation system.

For example, a bias on the output of the x -axis accelerometer, δf_{xb} , introduces an acceleration error which forms error components $\delta f_{xb} \cos \theta$ and $-\delta f_{xb} \sin \theta$ in the x and z channels of the reference frame, respectively. These errors then propagate as position errors which increase as the square of time, $\delta f_{xb} \cos \theta t^2/2$ and $-\delta f_{xb} \sin \theta t^2/2$, the result of the double integration process required to compute position estimates. Initial attitude errors propagate in a similar manner, whilst a bias on the output of the gyroscope causes a position error which increases with time cubed owing to the additional integration process required to determine body attitude.

The form of the position errors caused by the various error contributions shown in Figure 12.1 are tabulated in Table 12.1.

The outline analysis given above has drawn attention to the way in which different sources of error propagate in an inertial navigation system. It is evident that there is inherent coupling of errors between the navigation channels caused by the process of resolving the specific force measurements into the designated reference frame. Hence, a simple rigorous calculation of errors is not usually practical.

12.2.2 Navigation in a rotating reference frame

Consideration is now given to the particular situation in which it is required to navigate in the vicinity of the Earth. Navigation is assumed to take place in the local geographic reference frame as considered in Chapter 3. The revised two-dimensional system mechanisation is as described in Figures 3.3 and 3.4.

In this system, the x and z reference axes are coincident with the local horizontal and the local vertical respectively, and the navigation system provides estimates of velocity in each of these directions. The estimated horizontal velocity divided by the radius of the Earth, which constitutes the transport rate term referred to in Chapter 3, is fed back and subtracted from the measured body rates for the purpose of calculating

Table 12.1 Propagation of errors in two-dimensional strapdown inertial navigation system

Error source	Position error	
	<i>x</i> -axis	<i>z</i> -axis
Initial position errors	δx_0	δx_0
	δz_0	—
Initial velocity errors	δv_{x0}	δv_{0t}
	δv_{z0}	—
Initial attitude error	$\delta\theta_0$	$\delta\theta_0 f_{zi} \frac{t^2}{2}$
Accelerometer biases	δf_{xb}	$\delta f_{xb} \cos \theta \frac{t^2}{2}$
	δf_{zb}	$\delta f_{zb} \sin \theta \frac{t^2}{2}$
Gyroscope bias	$\delta\omega_{yb}$	$\delta\omega_{yb} f_{zi} \frac{t^3}{6}$
		$-\delta\omega_{yb} f_{xi} \frac{t^3}{6}$

body attitude with respect to the local geographic frame. The effect of this feedback is to revise the system error dynamics as described in the following text.

The error dynamics of the local geographic system is analysed here for the condition where true body attitude is zero, that is, $\theta = 0$. In this case, the coupling between the channels is nominally zero, allowing each channel to be analysed separately. Despite the simplification of the analysis which this approach affords, the propagation of errors in the vertical and horizontal channels of the system can still be illustrated without loss of generality in the form of the results. In addition, it is assumed that the navigation system is mounted in a vehicle which is at rest on the Earth, or one which is travelling at a constant velocity with respect to the Earth. Under such conditions, the only force acting on the vehicle is the specific force needed to overcome the gravitational attraction of the Earth. In this situation, $f_{xg} = 0$ and $f_{zg} = g$.

The error equations, correct to first order, for the vertical and horizontal channels can now be written as shown in Table 12.2.

Errors in the vertical channel propagate with time in a similar manner to those of the inertial frame indicated earlier. However, in the horizontal channel, there is a closed loop as shown by the block diagram representation in Figure 12.2. This loop is oscillatory owing to the presence of two integrators in the closed path.

The single-axis navigator shown here is an electronic analogue of a hypothetical simple pendulum of length equal to the radius of the Earth. This is referred to as the Schuler pendulum, the properties of which are described in the following section.

Table 12.2 Error equations for the vertical and horizontal channels

Horizontal channel equations	Vertical channel equations
$\dot{\delta\theta} = \delta\omega_{yb} - \delta v_x/R_0$	$\dot{\delta v_z} = \delta f_{zb}$
$\dot{\delta v_x} = g\delta\theta + \delta f_{xb}$	$\dot{\delta z} = \delta v_z$
$\dot{\delta x} = \delta v_x$	

Key:

\int = Integration
 Σ = Summing junction

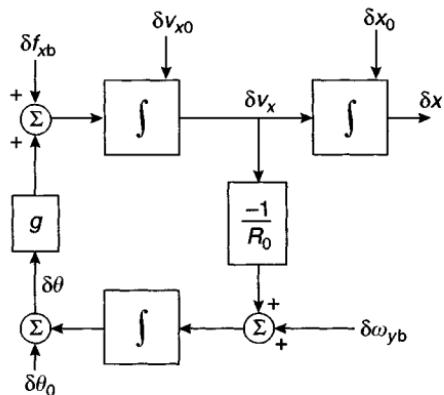


Figure 12.2 Inertial navigation system horizontal channel Schuler loop

12.2.3 The Schuler pendulum

The direction of the local vertical on the surface of the Earth can be determined using a simple pendulum consisting of a mass suspended from a fixed support point by a string. However, if the support point is moved from rest with an acceleration a , the string which supports the pendulum mass will be deflected with respect to the vertical by an angle θ equal to $\arctan(a/g)$ and therefore no longer defines the direction of the local vertical. Hypothetically, if the length of the support string is increased to equal the radius of the Earth, it will always remain vertical irrespective of the acceleration of the support point relative to the centre of the Earth. This is referred to as the Schuler pendulum after its formulation by Professor Max Schuler. A single-axis navigation system configured in the manner described above, as shown in Figure 12.2, is referred to as a Schuler tuned system since it behaves like a Schuler pendulum, as will now be demonstrated.

The measured specific force is resolved into the reference frame stored within the inertial navigation system. The resolved component of specific force is integrated once to give vehicle velocity and then again to give position. The transport rate is

calculated by dividing the indicated velocity by the radius of the Earth (R_0). This signal is used to modify the stored attitude reference as the inertial system moves around the Earth. In the event that the stored attitude reference is in error by an angle θ , the direction of the horizontal indicated by the system will be tilted with respect to the true horizontal by this angle and the resolved accelerometer measurement will include a component of gravity equal to $g\theta$. The closed loop that results is referred to as the Schuler loop. The loop is unstable as a result of the two integrators in the closed path and its dynamic behaviour is governed by the characteristic equation:

$$1 + \frac{g}{s^2 R_0} = 0 \quad (12.2)$$

where s is the Laplace operator.

Equation (12.2) may be rewritten as:

$$s^2 + \frac{g}{R_0} = 0$$

or

$$s^2 + \omega_s^2 = 0 \quad (12.3)$$

which is the equation for simple harmonic motion with natural frequency $\omega_s = \sqrt{g/R_0} = 0.00124$ rad/s for the Schuler pendulum. This is the Schuler frequency. The period of the Schuler oscillation is given by:

$$T_s = \frac{2\pi}{\omega_s} = 2\pi \sqrt{\frac{R_0}{g}} = 84.4 \text{ min} \quad (12.4)$$

This is of the same form as the equation for the period of a simple pendulum of length l , viz:

$$T = 2\pi \sqrt{\frac{l}{g}} \quad (12.5)$$

Therefore, the Schuler oscillation can be considered as the motion of a hypothetical pendulum of length equal to the radius of the Earth, R_0 . A pendulum tuned to the 'Schuler frequency' will always indicate the vertical on a moving vehicle provided it has been initially aligned to it. It is for this reason that Schuler tuned systems are most commonly employed for inertial navigation in the vicinity of the surface of the Earth.

12.2.4 Propagation of errors in a Schuler tuned system

In the single-axis navigation system, oscillations at the Schuler frequency will be excited in the presence of system errors. The block diagram given in Figure 12.2 shows errors in the initial estimates of attitude, velocity and position, $\delta\theta_0$, δv_0 and δx_0 , respectively, and fixed biases in the gyroscopic and accelerometer measurements, $\delta\omega_{yb}$

Table 12.3 Single-axis error propagation

Error source	Position error
Initial position error (δx_0)	δx_0
Initial velocity error (δv_0)	$\delta v_0 \left(\frac{\sin \omega_s t}{\omega_s} \right)$
Initial attitude error ($\delta \theta_0$)	$\delta \theta_0 R_0 (1 - \cos \omega_s t)$
Fixed acceleration bias (δf_{xb})	$\delta f_{xb} \left(\frac{1 - \cos \omega_s t}{\omega_s^2} \right)$
Fixed angular rate bias ($\delta \omega_{yb}$)	$\delta \omega_{yb} R_0 \left(t - \frac{\sin \omega_s t}{\omega_s} \right)$

and δf_{xb} . The propagation of these error terms with time may be derived from the differential equations (Table 12.2).

12.2.5 Discussion of results

It is apparent from these results that over long periods of time, several Schuler periods or more, the errors in the simple navigation system are bounded as a result of the Schuler tuning. This is true for all sources of error with the exception of the bias of the gyroscope which gives rise to a position error which increases linearly with time, $\delta \omega R_0 t$, in addition to an oscillatory component. It is clear, therefore, that the performance of the gyroscope is critical in the achievement of long term system accuracy. This is one of the reasons why so much effort has been expended over the years in perfecting the performance of gyroscopes.

It follows that an approximate indication of inertial navigation system performance can be deduced solely from knowledge of gyroscopic measurement accuracy. For example, a system incorporating 0.01°/h gyroscopes should be capable of navigating to an accuracy of ~ 1 km/h. This relationship is often used to provide a 'rule of thumb' guide to navigation system performance. The physical significance of this effect will be appreciated when it is remembered that the gyroscopes are used to store an attitude reference within the navigation system, and that the stored reference changes at the drift rate of the gyroscope. On the surface of the Earth, at the equator, 1° of longitude corresponds to 111 km (approximately 60 nautical miles). Hence, 1 min of arc is equivalent to a displacement of approximately 1 nautical mile.

The analyses presented thus far relate to a simplified navigation system operating in a single plane. As will now be shown, the complexity of the error model is increased greatly in a full three-dimensional inertial navigator, particularly in the presence of vehicle manoeuvres and as a result of coupling between the respective channels of the system. The following section describes a generalised set of error equations which may be used to predict inertial navigation system performance.

12.3 General error equations

In this section, the growth of errors in a full three-dimensional navigation system is examined. The equations given here relate to a terrestrial system operating close to the Earth in a local geographic reference frame.

12.3.1 Derivation of error equations

12.3.1.1 Attitude errors

The orientation of the instrument cluster in a strapdown system with respect to the navigation reference frame may be expressed in terms of the direction cosine matrix, $\tilde{\mathbf{C}}_b^n$. The estimated attitude, denoted by $\tilde{\mathbf{C}}_b^n$ may be written in terms of the true direction cosine matrix, \mathbf{C}_b^n , as follows:

$$\tilde{\mathbf{C}}_b^n = \mathbf{B} \mathbf{C}_b^n \quad (12.6)$$

where \mathbf{B} represents the transformation from true reference axes to estimated reference axes, the misalignment of the reference frame stored in the inertial navigation system computer. For small angles of misalignment, the matrix \mathbf{B} may be approximated as a skew symmetric matrix which may be written as follows:

$$\mathbf{B} = [\mathbf{I} - \Psi] \quad (12.7)$$

where \mathbf{I} is a 3×3 identity matrix and Ψ is given by:

$$\Psi = \begin{pmatrix} 0 & -\delta\gamma & \delta\beta \\ \delta\gamma & 0 & -\delta\alpha \\ -\delta\beta & \delta\alpha & 0 \end{pmatrix} \quad (12.8)$$

The elements, $\delta\alpha$ and $\delta\beta$, correspond to the attitude errors with respect to the vertical, the level or tilt errors, whilst $\delta\gamma$ represents the error about vertical, the heading or azimuth error. These terms are analogous to the physical misalignments of the instrument cluster in a stable platform navigation system and may be equated approximately to the roll, pitch and yaw Euler errors for small angle misalignments.

The estimated direction cosine matrix may now be written as follows:

$$\tilde{\mathbf{C}}_b^n = [\mathbf{I} - \Psi] \mathbf{C}_b^n \quad (12.9)$$

which may be rearranged to give:

$$\Psi = \mathbf{I} - \tilde{\mathbf{C}}_b^n \mathbf{C}_b^{nT} \quad (12.10)$$

Differentiating this equation yields:

$$\dot{\Psi} = -\dot{\tilde{\mathbf{C}}}_b^n \mathbf{C}_b^{nT} - \tilde{\mathbf{C}}_b^n \dot{\mathbf{C}}_b^{nT} \quad (12.11)$$

As shown in Chapter 3, the direction cosine matrix, \mathbf{C}_b^n , propagates as a function of the absolute body rate (Ω_{ib}^b) and the navigation frame rate (Ω_{in}^n) in accordance with the following equation:

$$\dot{\mathbf{C}}_b^n = \mathbf{C}_b^n \Omega_{ib}^b - \Omega_{in}^n \mathbf{C}_b^n \quad (12.12)$$

Similarly, the time differential of the estimated matrix $\tilde{\mathbf{C}}_b^n$ is given by:

$$\dot{\tilde{\mathbf{C}}}_b^n = \tilde{\mathbf{C}}_b^n \tilde{\boldsymbol{\Omega}}_{ib}^b - \tilde{\boldsymbol{\Omega}}_{in}^n \tilde{\mathbf{C}}_b^n \quad (12.13)$$

where $\tilde{\boldsymbol{\Omega}}_{ib}$ and $\tilde{\boldsymbol{\Omega}}_{in}$ represent the measured body rate and the estimated turn rate of the navigation reference frame respectively.

Substituting for $\dot{\tilde{\mathbf{C}}}_b^n$ and $\dot{\tilde{\mathbf{C}}}_b^n$ in eqn. (12.11) gives:

$$\begin{aligned}\dot{\Psi} &= -\tilde{\mathbf{C}}_b^n \tilde{\boldsymbol{\Omega}}_{ib}^b \mathbf{C}_b^{nT} + \tilde{\boldsymbol{\Omega}}_{in}^n \tilde{\mathbf{C}}_b^n \mathbf{C}_b^{nT} + \tilde{\mathbf{C}}_b^n \tilde{\boldsymbol{\Omega}}_{ib}^b \mathbf{C}_b^{nT} - \tilde{\mathbf{C}}_b^n \mathbf{C}_b^{nT} \tilde{\boldsymbol{\Omega}}_{in}^n \\ &= -\tilde{\mathbf{C}}_b^n [\tilde{\boldsymbol{\Omega}}_{ib}^b - \boldsymbol{\Omega}_{ib}^b] \mathbf{C}_b^{nT} + \tilde{\boldsymbol{\Omega}}_{in}^n \tilde{\mathbf{C}}_b^n \mathbf{C}_b^{nT} - \tilde{\mathbf{C}}_b^n \mathbf{C}_b^{nT} \tilde{\boldsymbol{\Omega}}_{in}^n\end{aligned} \quad (12.14)$$

Substituting for $\tilde{\mathbf{C}}_b^n$ from eqn. (12.9) gives:

$$\dot{\Psi} = -[\mathbf{I} - \Psi] \mathbf{C}_b^n [\tilde{\boldsymbol{\Omega}}_{ib}^b - \boldsymbol{\Omega}_{ib}^b] \mathbf{C}_b^{nT} + \tilde{\boldsymbol{\Omega}}_{in}^n [\mathbf{I} - \Psi] \mathbf{C}_b^n \mathbf{C}_b^{nT} - [\mathbf{I} - \Psi] \mathbf{C}_b^n \mathbf{C}_b^{nT} \tilde{\boldsymbol{\Omega}}_{in}^n \quad (12.15)$$

writing $\delta\boldsymbol{\Omega}_{in} = \tilde{\boldsymbol{\Omega}}_{in} - \boldsymbol{\Omega}_{in}$ and $\delta\boldsymbol{\Omega}_{ib} = \tilde{\boldsymbol{\Omega}}_{ib} - \boldsymbol{\Omega}_{ib}$ and ignoring error product terms, we have:

$$\dot{\Psi} \approx \Psi \tilde{\boldsymbol{\Omega}}_{in}^n - \boldsymbol{\Omega}_{in}^n \Psi + \delta\boldsymbol{\Omega}_{in}^n - \mathbf{C}_b^n \delta\boldsymbol{\Omega}_{ib}^b \mathbf{C}_b^{nT} \quad (12.16)$$

It can be shown from an element by element comparison that the above equation may be expressed in vector form as:

$$\dot{\Psi} \approx -\omega_{in}^n \times \Psi + \delta\omega_{in}^n - \mathbf{C}_b^n \delta\omega_{ib}^b \quad (12.17)$$

where $\Psi = [\delta\alpha \ \delta\beta \ \delta\gamma]^T$, the misalignment vector and

$$\Psi \times = \Psi \quad \omega_{in}^n \times = \boldsymbol{\Omega}_{in}^n \quad \delta\omega_{in}^n \times = \delta\boldsymbol{\Omega}_{in}^n \quad \delta\omega_{ib}^b \times = \delta\boldsymbol{\Omega}_{ib}^b$$

12.3.1.2 Velocity and position errors

The velocity equation may be expressed as:

$$\dot{\mathbf{v}} = \mathbf{C}_b^n \mathbf{f}^b - (2\omega_{ie}^n + \omega_{en}^n) \times \mathbf{v} + \mathbf{g}_l \quad (12.18)$$

where \mathbf{f}^b represents the specific force in body axes.

Similarly, the estimated velocity may be assumed to propagate in accordance with the following equation in which estimated quantities are again denoted by a tilde:

$$\dot{\tilde{\mathbf{v}}} = \tilde{\mathbf{C}}_b^n \tilde{\mathbf{f}}^b - (2\tilde{\omega}_{ie}^n + \tilde{\omega}_{en}^n) \times \tilde{\mathbf{v}} + \tilde{\mathbf{g}}_l \quad (12.19)$$

Differencing these two equations, we have:

$$\begin{aligned}\delta\dot{\mathbf{v}} &= \dot{\tilde{\mathbf{v}}} - \dot{\mathbf{v}} \\ &= \tilde{\mathbf{C}}_b^n \tilde{\mathbf{f}}^b - \mathbf{C}_b^n \mathbf{f}^b - (2\tilde{\omega}_{ie}^n + \tilde{\omega}_{en}^n) \times \tilde{\mathbf{v}} + (2\omega_{ie}^n + \omega_{en}^n) \times \mathbf{v} + \tilde{\mathbf{g}}_l - \mathbf{g}_l\end{aligned} \quad (12.20)$$

Substituting for $\tilde{\mathbf{C}}_b^n = [\mathbf{I} - \Psi] \mathbf{C}_b^n$ and writing $\tilde{\mathbf{f}}^b - \mathbf{f}^b = \delta\mathbf{f}^b$, $\tilde{\mathbf{v}} - \mathbf{v} = \delta\mathbf{v}$, $\tilde{\omega}_{ie}^n - \omega_{ie}^n = \delta\omega_{ie}^n$ and $\tilde{\omega}_{en}^n - \omega_{en}^n = \delta\omega_{en}^n$, and expanding, ignoring error product

terms, gives:

$$\delta\dot{\mathbf{v}} = -\Psi \mathbf{C}_b^n \mathbf{f}^b + \mathbf{C}_b^n \delta\mathbf{f}^b - (2\omega_{ie}^n + \omega_{en}^n) \times \delta\mathbf{v} - (2\delta\omega_{ie}^n + \delta\omega_{en}^n) \times \mathbf{v} - \delta\mathbf{g}$$

writing $\mathbf{C}_b^n \mathbf{f}^b = \mathbf{f}^n$ and rearranging, gives:

$$\delta\dot{\mathbf{v}} = [\mathbf{f}^n \times] \Psi + \mathbf{C}_b^n \delta\mathbf{f}^b - (2\omega_{ie}^n + \omega_{en}^n) \times \delta\mathbf{v} - (2\delta\omega_{ie}^n + \delta\omega_{en}^n) \times \mathbf{v} - \delta\mathbf{g} \quad (12.21)$$

Ignoring errors in the Coriolis terms and in knowledge of the gravity vector, this equation reduces to:

$$\delta\dot{\mathbf{v}} = [\mathbf{f}^n \times] \Psi + \mathbf{C}_b^n \delta\mathbf{f}^b \quad (12.22)$$

Finally, the position errors, $\delta\mathbf{p}$, may be expressed as follows:

$$\delta\dot{\mathbf{p}} = \delta\mathbf{v} \quad (12.23)$$

The velocity and position errors are predominantly functions of the specific force to which the inertial navigation system is subjected, \mathbf{f}^n , the attitude errors, Ψ , and inaccuracies in the measurements of specific force provided by the accelerometers, $\delta\mathbf{f}^b$. In addition, errors arise in a local vertical terrestrial navigator through errors in the Coriolis terms, imperfect knowledge of the local gravity vector and incorrect assumptions regarding the shape of the Earth.

Equations of the above form may be used to describe the propagation of errors in each of the strapdown mechanisations discussed in Section 3.3. For example, for a system operating in local geographic axes, ω_{in} represents the sum of the ‘Earth’s rate’ and ‘transport rate’ terms, whilst it becomes zero for a system operating in space-fixed coordinates.

12.3.1.3 State space form

Equations (12.17), (12.21) and (12.23) may be combined to form a single matrix error equation as follows:

$$\dot{\delta\mathbf{x}} = \mathbf{F}\delta\mathbf{x} + \mathbf{Gu} \quad (12.24)$$

where

$$\delta\mathbf{x} = [\delta\alpha \ \delta\beta \ \delta\gamma \ \delta v_N \ \delta v_E \ \delta v_D \ \delta L \ \delta \ell \ \delta h]^T \quad (12.25)$$

$$\mathbf{u} = [\delta\omega_x \ \delta\omega_y \ \delta\omega_z \ \delta f_x \ \delta f_y \ \delta f_z]^T \quad (12.26)$$

$$\mathbf{G} = \begin{pmatrix} -c_{11} & -c_{12} & -c_{13} & 0 & 0 & 0 \\ -c_{21} & -c_{22} & -c_{23} & 0 & 0 & 0 \\ -c_{31} & -c_{32} & -c_{33} & 0 & 0 & 0 \\ 0 & 0 & 0 & c_{11} & c_{12} & c_{13} \\ 0 & 0 & 0 & c_{21} & c_{22} & c_{23} \\ 0 & 0 & 0 & c_{31} & c_{32} & c_{33} \end{pmatrix} \quad (12.27)$$

$$\mathbf{F} = \begin{pmatrix}
 0 & -\left(\Omega \sin L + \frac{v_E}{R} \tan L\right) & \frac{v_N}{R} & 0 & \frac{1}{R} & 0 & -\Omega \sin L & 0 & -\frac{v_E}{R^2} \\
 \left(\Omega \sin L + \frac{v_E}{R} \tan L\right) & 0 & \Omega \cos L + \frac{v_E}{R} & -\frac{1}{R} & 0 & 0 & 0 & 0 & \frac{v_N}{R^2} \\
 -\frac{v_N}{R} & -\Omega \cos L - \frac{v_E}{R} & 0 & 0 & -\frac{\tan L}{R} & 0 & -\Omega \cos L - \frac{v_E}{R \cos^2 L} & 0 & \frac{v_E \tan L}{R^2} \\
 0 & -f_D & f_E & \frac{v_D}{R} & -2\left(\Omega \sin L + \frac{v_E}{R} \tan L\right) & \frac{v_N}{R} & -v_E \left(2\Omega \cos L + \frac{v_E}{R \cos^2 L}\right) & 0 & \frac{1}{R^2} (v_E^2 \tan L - v_N v_D) \\
 f_D & 0 & -f_N & \left(2\Omega \sin L + \frac{v_E}{R} \tan L\right) & \frac{1}{R} (v_N \tan L + v_D) & 2\Omega \cos L + \frac{v_E}{R} & \left(2\Omega (v_N \cos L - v_D \sin L) + \frac{v_N v_E}{R \cos^2 L}\right) & 0 & -\frac{v_E}{R^2} (v_N \tan L + v_D) \\
 -f_E & f_N & 0 & -\frac{2v_N}{R} & -2\left(\Omega \cos L + \frac{v_E}{R}\right) & 0 & 2\Omega v_E \sin L & 0 & \frac{1}{R^2} (v_N^2 + v_E^2) \\
 0 & 0 & 0 & \frac{1}{R} & 0 & 0 & 0 & 0 & -\frac{v_N}{R^2} \\
 0 & 0 & 0 & 0 & \frac{1}{R \cos L} & 0 & \frac{v_E \tan L}{R \cos L} & 0 & -\frac{v_E}{R^2 \cos L} \\
 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0
 \end{pmatrix} \quad (12.28)$$

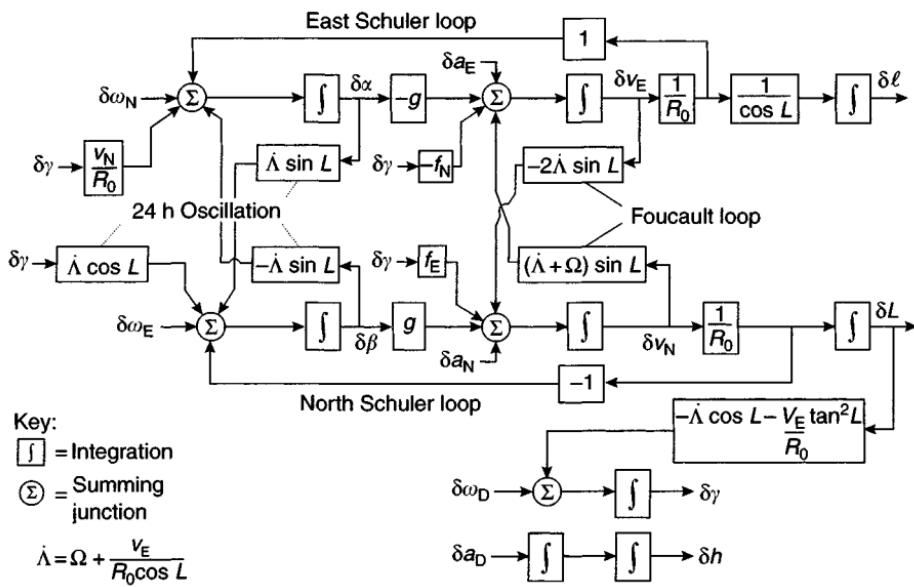


Figure 12.3 Error block diagram

12.3.2 Discussion

The set of coupled differential equations given in the last section define the propagation of errors in a local geographic inertial navigation system. A simplified block diagram representation of the error model is given in Figure 12.3. This model is applicable to a navigation system installed in a vehicle which is moving over the surface of a spherical earth. The diagram shows the Schuler loops and a number of the cross-coupling terms which give rise to longer-term oscillations described below.

The errors which propagate in an inertial navigation system over long periods of time are characterised by three distinct frequencies:

The Schuler oscillation, $\omega_s = \sqrt{g/R_0}$ which manifests itself as an oscillation in each horizontal channel. The period of this oscillation is approximately 84.4 min as described in the Section 12.2.3.

The Foucault oscillation, $\omega_f = \Omega \sin L$. This maintains itself as a modulation of the Schuler oscillation, the modulation in the two horizontal channels being 90° apart in phase. The Foucault oscillation has a period of $2\pi/\Omega \sin L$ where Ω is the angular frequency of the Earth's rotation and L is the latitude of the system. The period of this oscillation is about 30 h for moderate latitudes.

A 24 h oscillation, ω_e , which is directly related to the period of rotation of the Earth, showing itself mainly as a latitude/azimuthal oscillation.

As described earlier, the dynamics of the horizontal channels of an inertial navigation system are analogous to the motion of a simple pendulum having a length

equal to the Earth's radius. The Schuler motion of a freely swinging pendulum of length R_0 suspended above a rotating Earth will be modulated at the Foucault frequency, which corresponds to the vertical component of the Earth's rate. The Foucault oscillation is named after the French physicist who used a freely swinging pendulum to demonstrate the rotation of the Earth. In a moving system, this frequency is modified by motion of the navigation system about the Earth. These effects are illustrated with the aid of the error propagation examples given below.

12.3.2.1 Examples

In Figure 12.4, plots are given illustrating the propagation of navigation errors with time over a 36 h period. The system is assumed to be stationary and located on the surface of the Earth at a latitude of 45° . The following errors sources are included.

Initial alignment errors with respect to the vertical ($\delta\alpha_0, \delta\beta_0$)	= 0.1 mrad
Initial heading error ($\delta\gamma_0$)	= 1.0 mrad
Gyroscopic bias (δB_g)	= $0.01^\circ/\text{h}$
Accelerometer bias (δB_a)	= 0.1 milli-g

The distribution of the errors is assumed to be Gaussian and the earlier figures represent 1σ values, as described in Appendix B. The resulting attitude and position errors given in the figure are also Gaussian 1σ values, that is, there is a 68 per cent probability of each error lying within the limits indicated.

The presence of the Schuler, Foucault and Earth's rate oscillations are clearly apparent in the figure. At a latitude of 45° considered here, the period of the Foucault oscillation is approximately 34 h. The Schuler components of the attitude errors are shown to be modulated at this frequency, whilst the affect on the propagation of position errors is second order.

In general, the full error model described in the preceding section is only required to assess the performance of inertial navigation systems operating for long periods of time, several days. For many applications, including aircraft and missile systems, flight times are typically of the order of hours or minutes, rather than days. In such cases, some simplifications can be made in the error models used to assess system performance, since the terms which give rise to the Foucault and 24 h oscillations can often be disregarded. This is illustrated in Figure 12.5 where the growth of navigation errors over a 4 h period is shown. It can be seen that the Schuler frequency components are dominant in this situation.

For navigation over a few hours or less, much of the coupling between the north, east and vertical channels of the inertial navigation system can be ignored allowing each channel to be treated largely in isolation. The analysis of such systems becomes more tractable, as illustrated in the following section.

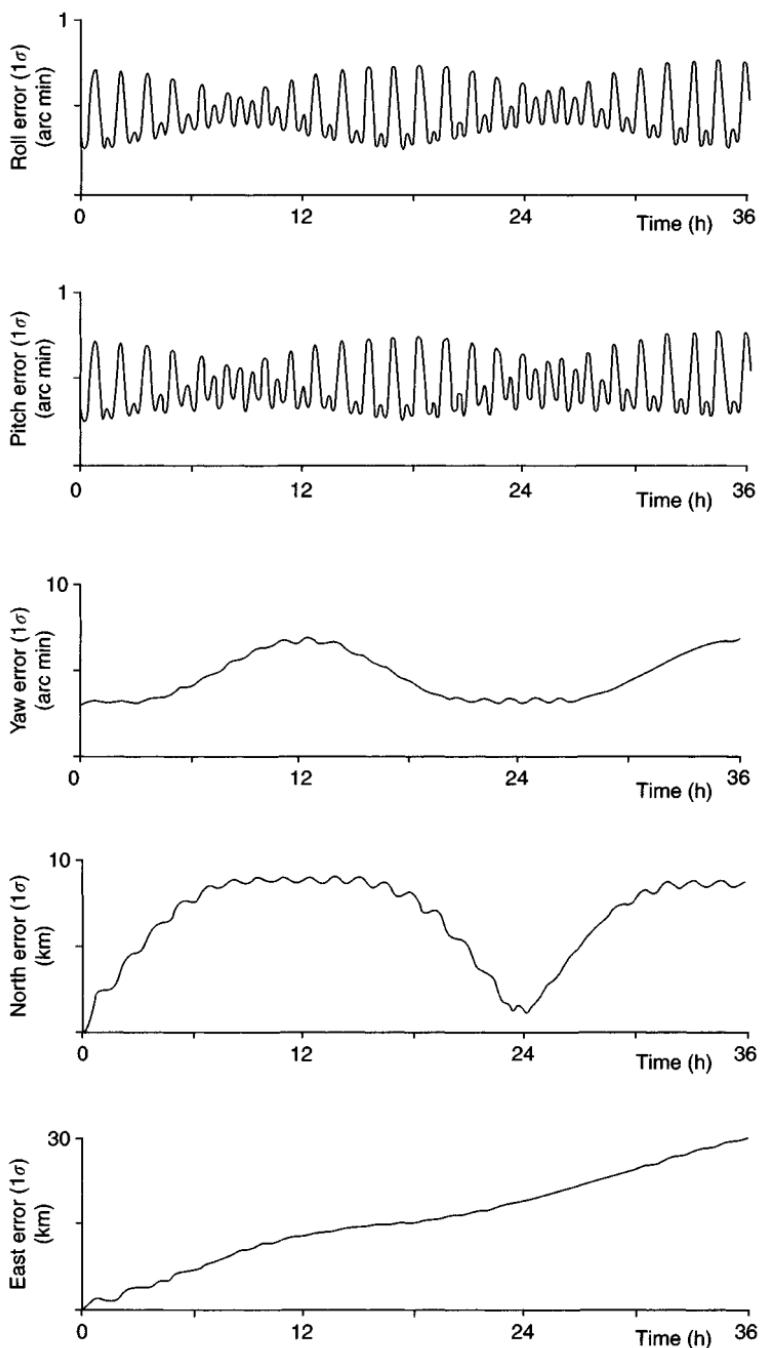


Figure 12.4 Simulated navigational accuracy (36 h period)

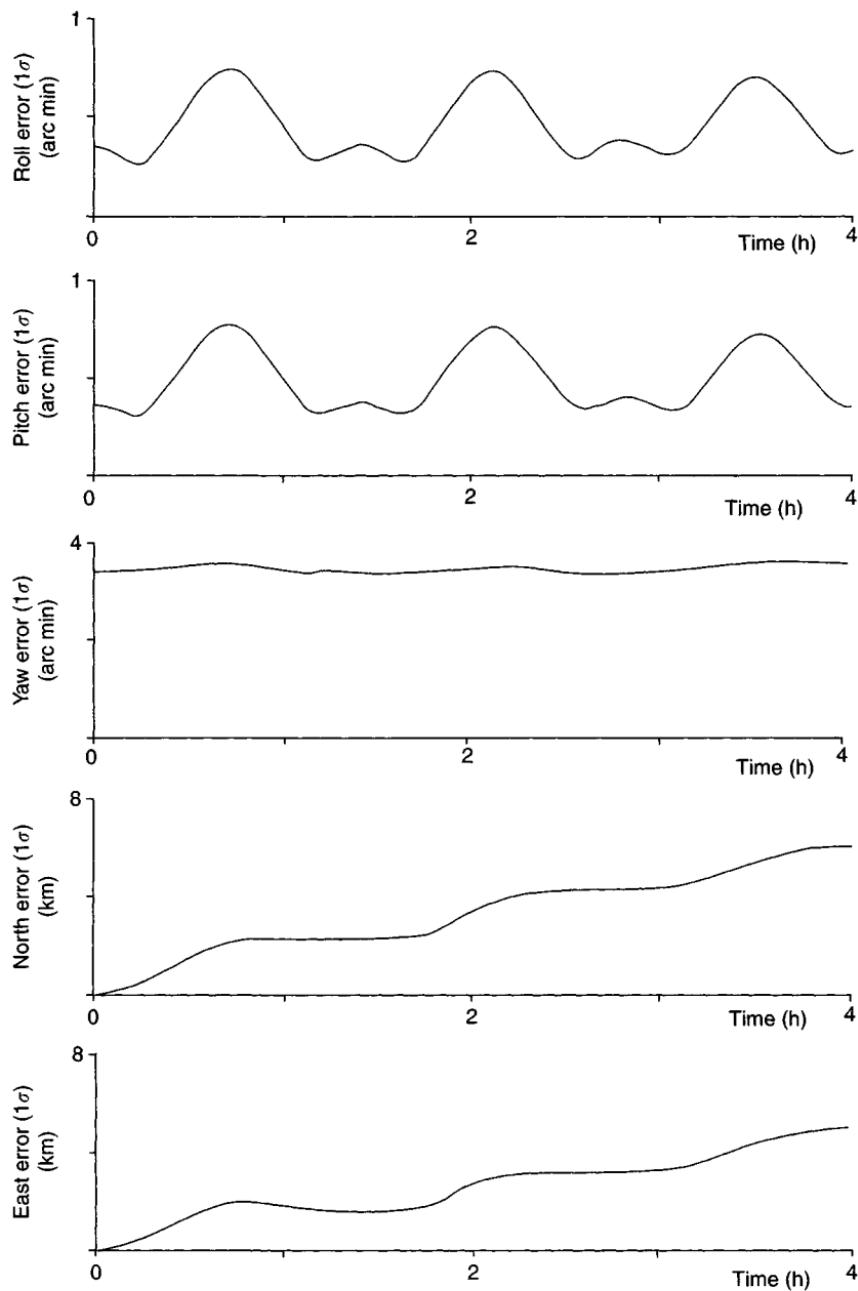


Figure 12.5 Simulated navigational accuracy (4 h period)

12.4 Analytical assessment

A full analytical solution of the error eqn. (12.24) given in Section 12.3 is extremely onerous mathematically and it is common practice to solve it using a computer. Methods by which this may be accomplished using a computer model are the subject of Section 12.5. However, for periods of navigation up to a few hours, the effects of the Foucault and 24 h oscillations may safely be ignored for many applications and the propagation of errors in the north, east and vertical channels can be considered separately. Under such conditions, a simplified analysis similar to that described in Section 12.2 may be undertaken. To illustrate the analytical methods which may be applied, the propagation of navigation errors in the north channel alone is examined.

12.4.1 Single channel error model

For a strapdown inertial navigation system mounted in a vehicle travelling at constant speed and at constant height above the Earth, the error dynamics for the north channel may be expressed in terms of the following set of coupled differential equations, in accordance with the error equations given in the preceding section:

$$\begin{aligned}\delta\dot{\beta} &= \left(\Omega \cos L + \frac{v_E}{R}\right)\delta\gamma - \frac{\delta v_N}{R_0} - \delta B_{gE} \\ \delta\dot{\gamma} &= -\delta B_{gD} \\ \dot{\delta v_N} &= g\delta\beta + \delta B_{aN} \\ \dot{\delta x_N} &= \delta v_N\end{aligned}\tag{12.29}$$

where δB_{gE} and δB_{gD} represent the effective gyroscopic biases acting about the east and vertical axes respectively, and δB_{aN} is the net accelerometer bias acting in the north direction. These terms may be expressed in terms of the gyroscopic measurement errors (δB_{gx} , δB_{gy} , δB_{gz}) and the accelerometer errors (δB_{ax} , δB_{ay} , δB_{az}) as follows:

$$\begin{aligned}\delta B_{gE} &= c_{21}\delta B_{gx} + c_{22}\delta B_{gy} + c_{23}\delta B_{gz} \\ \delta B_{gD} &= c_{31}\delta B_{gx} + c_{32}\delta B_{gy} + c_{33}\delta B_{gz} \\ \delta B_{aN} &= c_{11}\delta B_{ax} + c_{12}\delta B_{ay} + c_{13}\delta B_{az}\end{aligned}\tag{12.30}$$

If, as assumed here, the gyroscopic and accelerometer errors may be represented as fixed biases, the instrument bias dynamics may be represented by the following set of trivial differential equations:

$$\delta\dot{B}_{gE} = 0 \quad \delta\dot{B}_{gD} = 0 \quad \delta\dot{B}_{aN} = 0\tag{12.31}$$

These equations may be expressed in matrix form as follows:

$$\delta \dot{\mathbf{x}} = \mathbf{F} \delta \mathbf{x} \quad (12.32)$$

where

$$\delta \mathbf{x} = [\delta \beta \ \delta \gamma \ \delta v_N \ \delta x_N \ \delta B_{gE} \ \delta B_{gD} \ \delta B_{aN}]^T \quad (12.33)$$

and

$$\mathbf{F} = \begin{pmatrix} 0 & \dot{\Lambda} \cos L & -\frac{1}{R} & 0 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & -1 & 0 \\ g & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix} \quad (12.34)$$

in which $\dot{\Lambda} = \Omega + v_E/R \cos L$.

It is noted that position is given here in terms of distance (x_N) rather than latitude (L). A block diagram representation showing the instrument errors and the initial condition errors is shown in Figure 12.6.

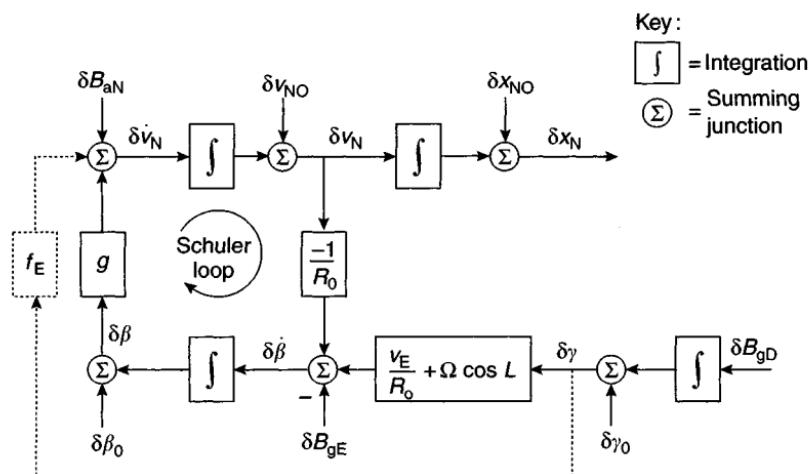


Figure 12.6 Simplified block diagram of the north channel of an inertial navigation system – medium term navigation

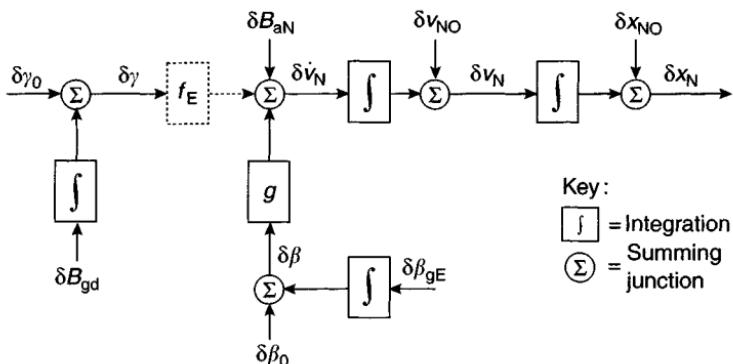


Figure 12.7 Simplified block diagram of the north channel of an inertial navigation system – short term navigation

The diagram shows the various error sources and the Schuler loop. It is noted that in the presence of a vehicle acceleration, the azimuth alignment error is coupled directly into the horizontal accelerometer, as indicated by the additional signal path shown dotted in the block diagram in Figure 12.6. It will be noticed that this representation is equivalent to the error model used for the simplified inertial navigation system described earlier.

For very short term navigation, that is, when the navigation time is a small fraction of a Schuler period, the Schuler feedback has relatively little effect on the growth of errors and the single channel error model can be reduced further to the form given in Figure 12.7.

Expressions for the north position errors which are applicable for medium and short term applications may be derived using state transition matrix methods, as described in the following section.

12.4.2 Derivation of single channel error propagation equations

The solution to eqn. (12.32) may be expressed in terms of the state transition matrix, $\Phi(t) = e^{Ft}$ as:

$$\delta x(t) = \Phi(t - t_0)\delta x(t_0) \quad (12.35)$$

where $\Phi(0) = I$ and $x(t_0)$ defines the initial states of the system.

The transition matrix is obtained using:

$$\Phi(t) = L^{-1}(sI - F)^{-1} \quad (12.36)$$

in which s is the Laplace operator and L^{-1} denotes the inverse Laplace transform.

The state transition matrix may be written as:

$$\Phi = \begin{pmatrix} \cos \omega_s t & \dot{\Lambda} \cos L \left(\frac{\sin \omega_s t}{\omega_s} \right) & -\frac{\sin \omega_s t}{\omega_s R_0} & 0 & -\frac{\sin \omega_s t}{\omega_s} & \dot{\Lambda} \cos L \left(\frac{1 - \cos \omega_s t}{\omega_s^2} \right) & -\left(\frac{1 - \cos \omega_s t}{g} \right) \\ 0 & 1 & 0 & 0 & 0 & -t & 0 \\ g \frac{\sin \omega_s t}{\omega_s} & \dot{\Lambda} \cos L g \left(\frac{1 - \cos \omega_s t}{\omega_s^2} \right) & \cos \omega_s t & 0 & -g \left(\frac{1 - \cos \omega_s t}{\omega_s^2} \right) & -\dot{\Lambda} \cos L R_0 & \frac{\sin \omega_s t}{\omega_s} \\ R_0(1 - \cos \omega_s t) & \dot{\Lambda} \cos L R_0 \left(t - \frac{\sin \omega_s t}{\omega_s} \right) & \frac{\sin \omega_s t}{\omega_s} & 1 & -R_0 \left(t - \frac{\sin \omega_s t}{\omega_s} \right) & -\dot{\Lambda} \cos L R_0 & \left(\frac{1 - \cos \omega_s t}{\omega_s^2} \right) \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix} \quad (12.37)$$

where $\omega_s = \sqrt{g/R_0}$, the frequency of the Schuler oscillation. The terms in a particular row of the transition matrix describe the dynamic effect of each error term on a particular error state. For example, the first term in row four indicates that a tilt error ($\delta\beta_0$) will give rise to a position error which propagates with time as $\delta\beta_0 R_0(1 - \cos \omega_s t)$. Similarly, it may be inferred that:

- a constant velocity error (δv_0) gives rise to a position error of:

$$\delta v_0 \frac{\sin \omega_s t}{\omega_s}$$

- an effective acceleration bias acting in the north channel (δB_{aN}) gives rise to a position error of:

$$\delta B_{aN} \left(\frac{1 - \cos \omega_s t}{\omega_s^2} \right)$$

- an effective angular rate bias acting about the east axis (δB_{gE}) gives rise to a position error of:

$$\delta B_{gE} R_0 \left(t - \frac{\sin \omega_s t}{\omega_s} \right)$$

- similarly, a heading error ($\delta\gamma_0$) gives rise to a position error of:

$$\delta\gamma_0 \dot{\Lambda} \cos L R_0 \left(t - \frac{\sin \omega_s t}{\omega_s} \right)$$

- an effective angular rate bias acting about the vertical axis (δB_{gD}) gives rise to a position error of:

$$\delta B_{gD} \dot{\Lambda} \cos L R_0 \left\{ \frac{t^2}{2} - \left(\frac{1 - \cos \omega_s t}{\omega_s^2} \right) \right\}$$

Over very short periods of navigation, that is, navigation over a small fraction of a Schuler period, further simplifications may be made to these expressions. The position error contributions in the medium and short term are summarised in Table 12.4.

These equations can be used to assess system performance or to specify an inertial navigation (IN) system to fulfil a particular application. The propagation of each error type with time is illustrated in Figures 12.8–12.12. In each of these graphs, the vertical scale is defined (as a function of the relevant error) by the figures given in the text box inset in each plot. Figure 12.8 shows a plot of the north position error (δx_N) resulting from the horizontal component of gyroscope bias (δB_{gE}).

$$\delta x_N = R_0 \left[t - \frac{\sin \omega_s t}{\omega_s} \right] \delta B_{gE}$$

It can be seen that the resulting position error comprises a ramp error with superimposed Schuler oscillation. Figures are given for the mean build-up of position error with time and the associated mean velocity error. It can be seen that given a residual gyro bias of $0.01^\circ/\text{h}$, the position error will grow at the rate of 0.6 nm/h . This corresponds to a mean velocity error of approximately 0.3 m/s (1 ft/s).

Table 12.4 Growth of position errors in the medium and short term

Error source	Position errors	
	Medium term	Short term
Initial attitude error ($\delta\beta_0$)	$R_0(1 - \cos \omega_s t)\delta\beta_0$	$g\delta\beta_0 \frac{t^2}{2}$
Initial attitude error ($\delta\gamma_0$)	$R_0 \dot{\Lambda} \cos L \left(t - \frac{\sin \omega_s t}{\omega_s} \right) \delta\gamma_0$	$\dot{\Lambda} \cos L g \delta\gamma_0 \frac{t^3}{6}$
Initial velocity error (δv_{N0})	$\left(\frac{\sin \omega_s t}{\omega_s} \right) \delta v_{N0}$	$\delta v_{N0} t$
Initial position error (δx_{N0})	δx_{N0}	δx_{N0}
Gyroscope bias (δB_{gE})	$R_0 \left(t - \frac{\sin \omega_s t}{\omega_s} \right) \delta B_{gE}$	$g\delta B_{gE} \frac{t^3}{6}$
Gyroscope bias (δB_{gD})	$-R_0 \dot{\Lambda} \cos L$ $\times \left\{ \frac{t^2}{2} - \frac{1 - \cos \omega_s t}{\omega_s^2} \right\} \delta B_{gD}$	$\dot{\Lambda} \cos L \delta B_{gD} g \frac{t^4}{24}$
Accelerometer bias (δB_{aN})	$\left(\frac{1 - \cos \omega_s t}{\omega_s^2} \right) \delta B_{aN}$	$\delta B_{aN} \frac{t^2}{2}$

Note: $\dot{\Lambda} = \Omega + v_E/R_0 \cos L$, where Ω = Earth's rate, R_0 = Earth's radius, v_E = east velocity and L = latitude.

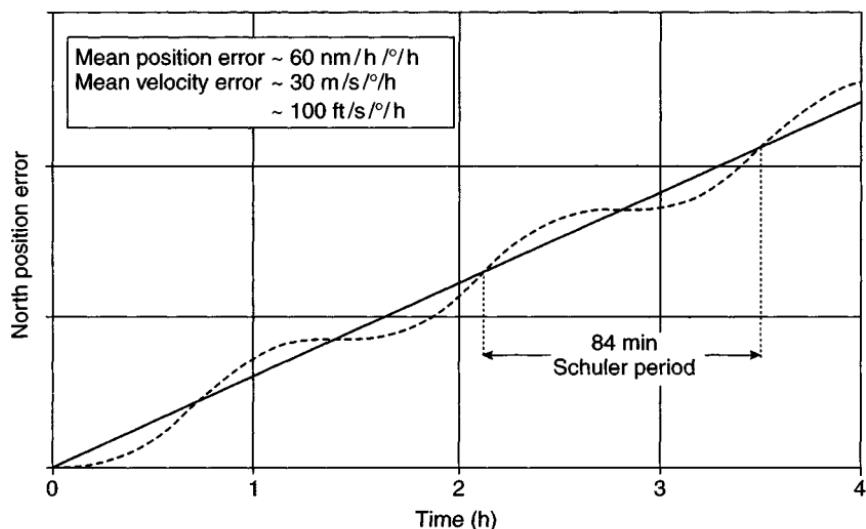


Figure 12.8 North position error versus time caused by gyroscope bias

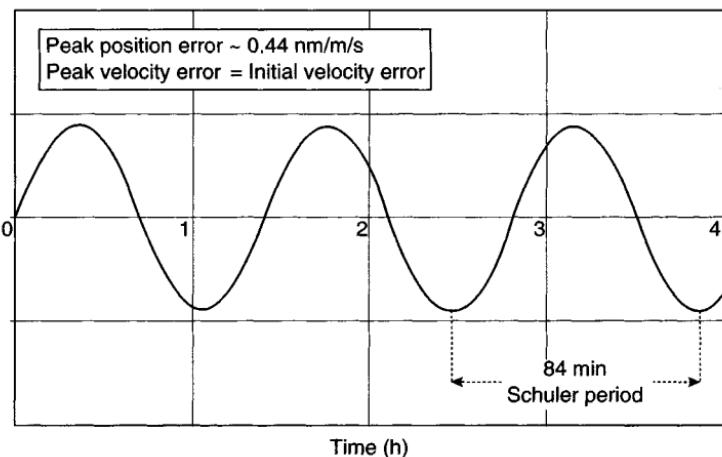


Figure 12.9 North position error versus time caused by initial velocity error

Figure 12.9 shows a plot of the north position error resulting from an initial north velocity error (δv_{N0}).

$$\delta x_N = \frac{\sin \omega_s t}{\omega_s} \delta v_{N0}$$

This error is bounded by the effect of the Schuler loop, and a figure is given for the maximum position error caused by a 1 m/s initial velocity error.

Figure 12.10 shows a plot of the north position error resulting from the horizontal component of accelerometer bias (δB_{aN}):

$$\delta x_N = \left(\frac{1 - \cos \omega_s t}{\omega_s^2} \right) \delta B_{aN}$$

or an initial attitude (tilt) error ($\delta \beta_0$):

$$\delta x_N = R_0 (1 - \cos \omega_s t) \delta \beta_0$$

The propagation of this type of error is also bounded by the effect of the Schuler loop. A residual accelerometer bias of 0.1 milli-g (corresponding to an initial tilt error of 0.1 mrad) will result in peak position and velocity errors of 0.7 nm and 0.8 m/s (2.6 ft/s), respectively.

Figure 12.11 shows a plot of the north position error resulting from an initial error in azimuth alignment ($\delta \gamma_0$).

$$\delta x_N = R_0 \dot{\Lambda} \cos L \left[t - \frac{\sin \omega_s t}{\omega_s^2} \right] \delta \gamma_0$$

The resulting position error comprises a Schuler oscillation superimposed on a ramp function as shown in the figure, the magnitude of the error varying with system latitude and speed. Some examples of the resulting error magnitudes are given for an azimuth misalignment of 1 mrad.

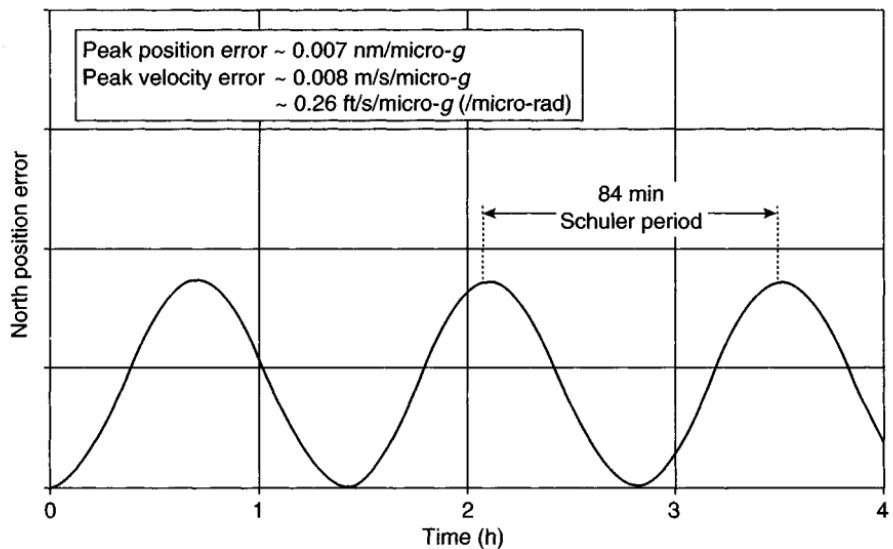


Figure 12.10 North position error versus time caused by accelerometer bias

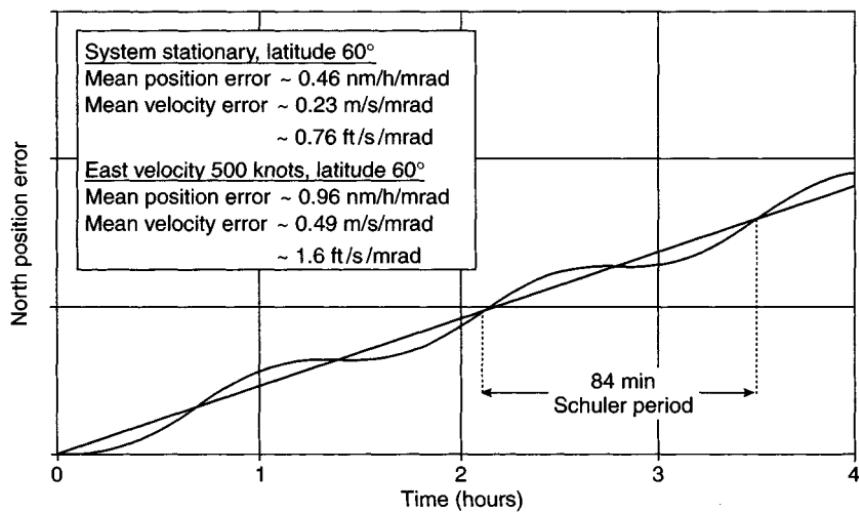


Figure 12.11 North position error versus time caused by initial azimuth misalignment

Figure 12.12 shows a plot of north position error resulting from a vertical gyroscope bias component (δB_{gD}).

$$\delta x_N = -R_0 \dot{\Lambda} \cos L \left[\frac{t^2}{2} - \left(\frac{1 - \cos \omega_s t}{\omega_s^2} \right) \right] \delta B_{gD}$$

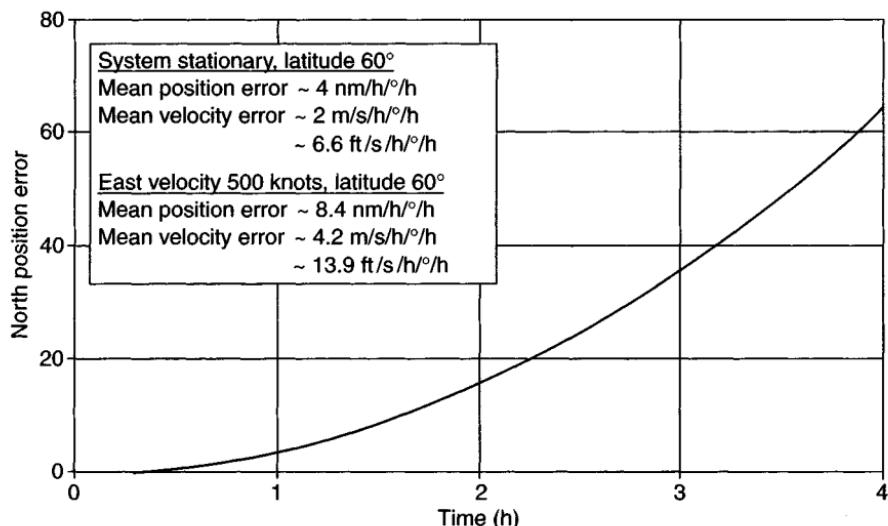


Figure 12.12 North position error versus time caused by vertical gyroscope bias component

The position error comprises a quadratic term with a superimposed Schuler oscillation; the relative magnitude of the oscillatory component is small and hence is not readily apparent in the figure with the scaling adopted here. The corresponding velocity error comprises a ramp with a Schuler oscillation superimposed. The propagation of this error also varies with latitude and example figures are given in the figure for a $1^{\circ}/\text{h}$ gyroscope bias.

Expressions similar to those given in Table 12.4 can be derived for the east channel of the inertial navigation system. As with the north channel, the growth of many of the errors is bounded by the effects of the Schuler tuning. However, this is not the case in the vertical channel where the errors increase rapidly with time. For example, a vertical accelerometer bias, B_{az} , will give rise to a position error of $B_{az}t^2/2$. It is for this reason that aircraft navigation systems commonly operate in conjunction with a barometric or radar altimeter in order to restrict the growth of vertical channel errors. The scope for aiding inertial systems in this way is discussed more fully in Chapter 13, particularly in relation to integrated navigation systems. As shown in that chapter, aided systems rely on error models of the form discussed here to predict the growth of inertial navigation system errors with time.

12.4.3 Single-channel error propagation examples

Some example calculations are given here to illustrate how navigation errors may be determined using the single-axis error models described in the previous section. Examples calculations for both aircraft and missile applications are presented.

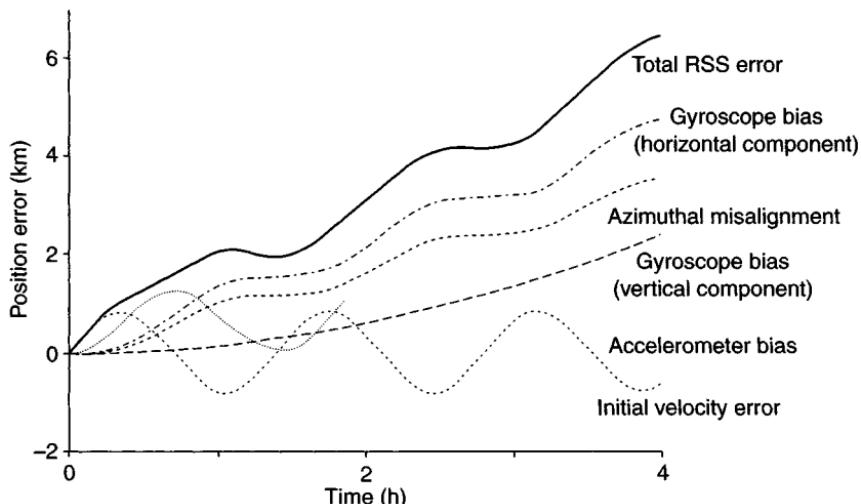


Figure 12.13 Single-channel error propagation

12.4.3.1 Aircraft INS error propagation

The growth of navigation errors under benign flight conditions may be assessed using the medium-term error equations given in Table 12.4. Sample plots are given in Figure 12.13 which show the growth of north position error with time in an aircraft navigation system over a 4 h period. For the purposes of this simple example, the aircraft is assumed to be cruising at a constant speed. Alignment accuracies of 0.05 mrad (1σ) in level and 1 mrad (1σ) in azimuth have been assumed, whilst the instrument performance is typical of a high grade airborne inertial navigation system. The gyroscope and accelerometer biases have been set to constant values of $0.01^\circ/\text{h}$ and 50 micro-g (1σ), respectively.

The figure illustrates the form of the position errors resulting from the various error sources whilst the upper curve represents the combined effect of the individual errors. The upper curve has been obtained by summing the individual error components quadratically to give the total navigation error. It will be seen that the gyroscope bias and azimuthal misalignment contributions grow with time whilst the other terms are bounded as a result of Schuler tuning. It should be noted that these are simplified results for a single channel. In a full model, the Foucault effect is noticeable even at the first Schuler period.

The analysis of inertial navigation system performance for airborne applications rapidly becomes complex when account is taken of realistic vehicle trajectories and manoeuvres, in which case the analyst will usually turn to simulation to aid the design process. However, under many circumstances, some useful analysis can still be carried out to obtain an initial indication of system performance.

In addition to the usual effects of alignment errors and sensor biases illustrated in the above error plots, a number of error contributions arise as a result of the

acceleration experienced by an aircraft during take-off. Both alignment errors and mounting misalignments of the accelerometers will give rise to cross-track velocity errors as the aircraft accelerates during take-off. For example, an azimuthal misalignment of $\delta\gamma_0$ will give rise to a velocity error of $V\delta\gamma_0$, where V is the cruise speed of the aircraft. The velocity error which has built up during take-off then propagates during the subsequent cruise phase of flight in the same way as an initial velocity error. Similarly, gyroscopic mass unbalance introduces a tilt error during take-off which propagates in the same manner as an accelerometer bias during the cruise phase of flight. Sensor scale-factor inaccuracy and acceleration dependent biases will give rise to additional navigation errors in the event of an aircraft manoeuvre. The effects of such manoeuvres are most conveniently assessed through simulation. Errors which are dependent on the motion of the host vehicle are discussed separately later in the chapter. As will be shown, many of these errors are of particular concern in strapdown navigation systems.

12.4.3.2 Tactical missile INS error propagation

During short periods of flight, navigation errors propagate as simple functions of time. For example, in the absence of a missile manoeuvre, a gyroscopic bias (B_G) propagates as $gB_Gt^3/6$, as shown in Table 12.4. However, in the presence of missile accelerations and turn rates, a number of other errors exert a considerable influence on the navigation performance, as illustrated in the following example.

Consider a missile which accelerates from rest at 200 m/s^2 (approximately $20g$) for a period of 5 s by which time it reaches a speed of 1000 m/s. Thereafter, the vehicle is assumed to maintain this speed for a further 10 s. Hence, the total flight duration is 15 s and the overall distance travelled is 12.5 km. The on-board inertial system is assumed to contain gyroscopes and accelerometers having 1σ measurement biases of $50^\circ/\text{h}$ and 10 milli- g , respectively. Table 12.5 indicates the dominant contributions to cross-track position error for the flight path described, together with approximate mathematical expressions for the error propagation. The values used for the instrument errors and misalignments are typical for this type of application.

It is clear from the above results that the largest contributions to position error are caused by the initial angular misalignment in yaw and gyroscope mass unbalance, the g -dependent bias.

12.5 Assessment by simulation

12.5.1 Introductory remarks

Whilst the analytical techniques described earlier provide a broad indication of inertial navigation system performance accuracy for various applications as a function of instrument quality and alignment accuracy, such methods are limited for the following reasons:

- they take only limited account of coupling between channels of the inertial navigation system;

Table 12.5 Tactical missile INS error analysis

Error source	1σ magnitude	Cross-track position error	Cross-track position error at $t = 15$ s (m)
<i>Initial misalignment</i>			
Position (δx_0)	1 m	δx_0	1
Velocity (δv_0)	1 m/s	$\delta v_0 t$	15
Attitude pitch ($\delta\theta_0$)	0.5°	$g\delta\theta_0 t^2/2$	10
Yaw ($\delta\psi_0$)	0.5°	$\int \delta\psi_0 a(t) dt dt$	109
<i>Accelerometer errors</i>			
Fixed bias (B_A)	10 milli-g	$0.5B_A t^2$	11
Mounting misalignment/ cross-coupling (M_A)	0.25%	$\int M_A a(t) dt dt$	31
<i>Gyroscope errors</i>			
Fixed bias (B_G)	$50^\circ/h$	$\int B_G a(t) dt dt$	7
g -Dependent bias (B_g)	$25^\circ/h/g$	$\int \int B_g a(t)^2 dt dt dt$	72
Anisoelastic bias (B_a)	$1^\circ/h/g^2$	$\int \int \int B_a a(t)^2 g dt dt dt$	3
1σ root sum square position error			136 m

Note: $a(t)$ is the longitudinal acceleration of the missile.

- it is difficult to take account of realistic vehicle manoeuvres without the solution to the error equations becoming mathematically intractable;
- it is necessary to make simplifying assumptions about the instrument errors in each channel. In general, the effective angular rate and specific force measurement errors in each of the north, east and vertical channels are functions of the errors in all three gyroscopes and accelerometers.

A more detailed investigation of inertial system errors and their interactions can be carried out using simulation.

12.5.2 Error modelling

The model of the inertial system that is to be assessed must include all sources of error which are believed to be significant. A full simulation must therefore incorporate alignment errors, representative models of the inertial sensors, including their errors, and any imperfections in the computational processes which are to be implemented.

12.5.2.1 Alignment errors

Unless the alignment process is to be modelled in detail, typical values for alignment errors are summed with the true attitude, velocity and position to define the navigation system estimates of these quantities at the start of navigation. Note that the alignment

process itself can result in correlation between initial errors and the sensor errors as described in Section 12.6.1.1.

12.5.2.2 Sensor errors

Generalised sensor error models suitable for simulation purposes are given here based upon the gyroscope and accelerometer error models discussed in Chapters 4–6. The errors in the measurements of angular rate provided by a set of gyroscopes ($\delta\omega_x$ $\delta\omega_y$ $\delta\omega_z$), whose sensitive axes are orthogonal, may be expressed mathematically as shown below:

$$\begin{pmatrix} \delta\omega_x \\ \delta\omega_y \\ \delta\omega_z \end{pmatrix} = \mathbf{B}_G + \mathbf{B}_g \begin{pmatrix} a_x \\ a_y \\ a_z \end{pmatrix} + \mathbf{B}_{ae} \begin{pmatrix} a_y a_z \\ a_z a_x \\ a_x a_y \end{pmatrix} + \mathbf{B}_{ai} \begin{pmatrix} \omega_y \omega_z \\ \omega_z \omega_x \\ \omega_x \omega_y \end{pmatrix} + \mathbf{S}_G \begin{pmatrix} \omega_x \\ \omega_y \\ \omega_z \end{pmatrix} + \mathbf{M}_G \begin{pmatrix} \omega_x \\ \omega_y \\ \omega_z \end{pmatrix} + \mathbf{w}_G \quad (12.38)$$

where a_x , a_y , a_z are the accelerations acting along the principle axes of the host vehicle, and ω_x , ω_y , ω_z are the applied angular rates acting about these same axes, as defined by the reference model. The measurements of angular rate generated by the navigation system gyroscopes are generated by summing errors with the true rates.

For example, the measured rate about the x -axis, $\tilde{\omega}_x$, may be expressed as:

$$\tilde{\omega}_x = \omega_x + \delta\omega_x$$

In eqn. (12.38):

\mathbf{B}_G is a three element vector representing the residual fixed biases which are present;
 \mathbf{B}_g is a 3×3 matrix representing the g -dependent bias coefficients;
 \mathbf{B}_{ae} is a 3×3 matrix representing the anisoelastic coefficients;
 \mathbf{B}_{ai} is a 3×3 matrix representing the anisoinertia coefficients;
 \mathbf{S}_G is a diagonal matrix representing the scale-factor errors;
 \mathbf{M}_G is a 3×3 skew symmetric matrix representing the mounting misalignments and cross-coupling terms;
 \mathbf{w}_G is a three element vector representing the in-run random bias errors.

All of the gyroscopic errors listed above are present to a greater or lesser extent in conventional gyroscopes, rate sensors and vibratory devices. However, as described in Chapter 5, the acceleration dependent biases are usually insignificant in optical sensors such as ring laser and fibre optic gyroscopes.

The modelling of in-run random errors has been discussed at some length by King [1]. Of particular concern is the effect of random walk errors which arise in optical gyroscopes and propagate as an angular error which is a function of the square root of time. The propagation of this type of error has been described by Flynn [2].

The errors in the measurements of specific force provided by an accelerometer triad may be expressed as shown below. It is assumed that the sensors are mounted

with their sensitive axes nominally aligned with the principal axes of the host vehicle:

$$\begin{pmatrix} \delta f_x \\ \delta f_y \\ \delta f_z \end{pmatrix} = \mathbf{B}_A + \mathbf{B}_V \begin{pmatrix} a_y a_z \\ a_z a_x \\ a_x a_y \end{pmatrix} + \mathbf{S}_A \begin{pmatrix} a_x \\ a_y \\ a_z \end{pmatrix} + \mathbf{M}_A \begin{pmatrix} a_x \\ a_y \\ a_z \end{pmatrix} + \mathbf{w}_A \quad (12.39)$$

In this equation,

\mathbf{B}_A is a three-element vector representing the fixed biases;

\mathbf{B}_V is a 3×3 matrix representing the vibro-pendulous error coefficients;

\mathbf{S}_A is a diagonal matrix representing the scale-factor errors;

\mathbf{M}_A is a 3×3 skew symmetric matrix representing the mounting misalignments and cross-coupling terms;

\mathbf{w}_A is a three-element vector representing the in-run random bias errors.

The accelerometer errors described are particularly relevant for the pendulous force-feedback accelerometer which is most commonly used at the present time for many different strapdown system applications.

12.5.2.3 Computational errors

As described in Chapter 11, inaccuracies will arise in a strapdown navigation system computer as a result of:

- bandwidth limitations, that is, restricted computational frequency;
- truncation of the mathematical functions used in the strapdown attitude and navigation algorithms;
- limitations on the order of numerical integration schemes selected.

The effects of computational imperfections are most effectively assessed in a forward time stepping simulation by implementing the attitude and navigation algorithms in full. Comparison is then made between the navigation estimates generated in this way with those obtained using more precise calculations carried out within the reference model.

The inertial system designer will usually endeavour to ensure that the navigation errors arising from computational inadequacies are small compared with the alignment and instrument error contributions. Where this is so, attention may well be concentrated on the alignment and sensor error contributions, certainly during the early stages of a project. However, any computational imperfections which are expected to give rise to a sustained error should be taken into account at this stage of the design process. For example, a net linear acceleration or angular rate bias may be expected to arise in a system designed to operate in a particular vibratory environment. Such an effect may be modelled as a bias of appropriate magnitude and summed with an instrument bias in an attempt to model the effect approximately.

12.5.3 Simulation techniques

Alternative techniques which may be used for computer assessment of inertial systems are indicated in the following sections.

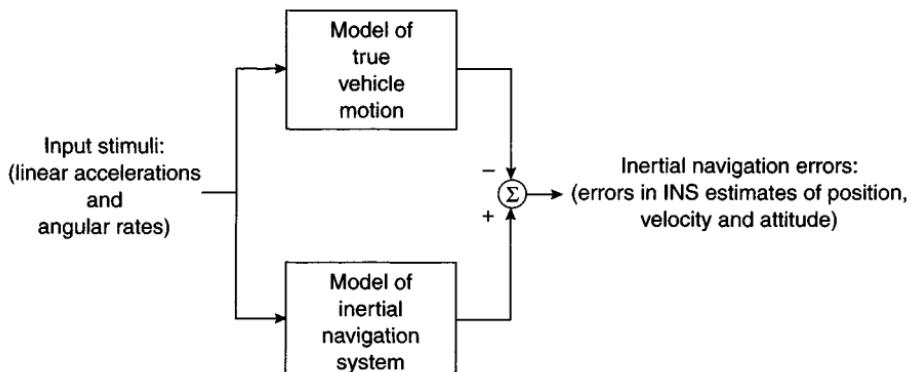


Figure 12.14 Block diagram of strapdown system simulation

12.5.3.1 Time-stepping simulation

A forward time-stepping simulation is commonly adopted for the detailed assessment of an inertial navigation system.

When using a simulation of this type, it is required to define some standard or reference against which the performance of the simulated inertial system can be judged, as depicted in Figure 12.14. The reference defines the actual, or 'true', motion of the vehicle in which the inertial system is required to operate. In defining this reference system, all aspects of the computation must be implemented as precisely as possible in order to ensure that it represents, as closely as possible, the 'true' motion of the vehicle in response to the stimuli which are causing it to move. Thus, it may be considered to be a 'perfect' inertial system in which all alignment, sensor and computational errors are set identically to zero.

The inertial system senses the specific force acceleration and turn rates to which it is subjected, and these same stimuli are used to 'drive' the reference model. An assessment of inertial system performance can then be made by comparing the outputs of the inertial system model with those of the reference system.

Using this approach, a so-called Monte Carlo simulation may be carried out in which each error source in the inertial navigation system is modelled as a random process in the error model. A large number of simulation 'runs' are then undertaken to generate estimates of the variances of the position, velocity and attitude estimates provided by the inertial system.

12.5.3.2 Covariance simulation

Using this approach, the error model equations given in Section 12.4 are transformed into covariance equations. These covariance equations are solved directly to determine the variances of the outputs as functions of time [3].

This technique avoids the need to perform very large numbers of individual runs to carry out a statistical assessment of system performance. Unlike the time-stepping

simulation method described in the previous section, a single covariance simulation yields the standard deviations of the position, velocity and attitude errors caused by initial misalignments and instrument errors.

12.5.3.3 Adjoint simulation

The adjoint technique is a computer-efficient method of determining the effect of initial condition errors, deterministic errors and random inputs on the values attained by a number of parameters of a system at a particular time [4–6]. For example, given an inertial navigation system with alignment and sensor errors, an adjoint simulation may be used to determine the contribution of each error source to the north or east position error at a given time.

The contribution to the north, or the east, position error, made by each of the separate error components is indicated by a single simulation run. Using a conventional forward time-stepping simulation, separate runs would be required for each error source to extract the same information.

The adjoint method can also be used to assess the performance of an inertial system when operating in a manoeuvring vehicle.

Where the sensitivity of a full strapdown system is to be examined in this way, it will be appreciated that the computational savings over a forward time-stepping simulation can become very significant indeed. However, it should be noted that this technique provides no information about the transient behaviour of a system and is not valid for non-linear systems. To examine such effects, a conventional forward time-stepping simulation will usually need to be used.

Whichever method is used, it will be vital to establish confidence in the simulation through verification and validation before proceeding with any detailed analysis. Hence, comparisons of the results of the simulation against theoretical results are recommended wherever possible to verify that the simulation is operating correctly. In some circumstances, it may be helpful to use a combination of methods in order to validate the complete model.

12.6 Motion dependence of strapdown system performance

The error propagation equations given in Section 12.4 are broadly applicable to all types of inertial navigation systems. In this section, attention is focused on system imperfections which are dependent on vehicle motion, many of which are of particular concern in strapdown inertial navigation systems.

The performance of a strapdown inertial navigation system is very dependent on the motion of the host vehicle. Strapdown inertial sensors are subjected to the full range of heading and attitude changes and turn rates which the vehicle experiences along its flight path. This is in marked contrast to the inertial sensors in a stable platform navigation system which remain nominally fixed in the chosen reference frame and are not subjected to the rotational motion dynamics of the vehicle. For example, aircraft turn rates sensed by strapdown gyroscopes are typically four or five

orders of magnitude greater than the inertial rates of the local geographic frame to which instruments in a platform system are subjected.

The need to operate in a relatively harsh dynamic environment whilst being able to measure large changes in vehicle attitude with sufficient accuracy has a major effect on the choice of inertial sensors. For example, gyroscope scale-factor accuracy and cross-coupling must be specified more precisely in a strapdown system than would be necessary for a platform system of comparable performance. In addition, a number of error effects need to be taken into account which do not have a major impact on the performance of platform systems and hence are not addressed in many earlier texts on the subject of inertial navigation. A number of motion dependent errors are discussed below, including various manoeuvre dependent terms and inaccuracies introduced through cyclic or vibratory motion of the host vehicle.

12.6.1 Manoeuvre-dependent error terms

The turn rates and accelerations which act on a vehicle as it manoeuvres excite a number of error sources within an on-board strapdown inertial navigation system. These include gyroscope and accelerometer scale-factor errors, cross-coupling effects and sensitivity to non-orthogonality of the sensors' input axes. In addition, for systems which use conventional spinning mass gyroscopes, various acceleration-dependent errors are induced as a result of mass unbalance and anisoelasticity, as described in Chapter 4. Therefore, the accuracy of the angular rate and specific force measurements generated by the on-board sensors during a manoeuvre can degrade substantially compared with that achieved under more benign conditions. The resulting measurement errors propagate giving rise to additional inaccuracies in the navigation system estimates of vehicle attitude, velocity and position.

Such effects are of particular significance in airborne inertial navigation systems used for combat aircraft and agile missile applications. In missile applications for example, the on-board sensors are often subjected to high levels of acceleration and rapid rates of turn, which give rise to substantial navigation errors. An example is given in Figure 12.15 showing the growth of navigation error in a short-range tactical missile navigation system. The missile contains a medium grade strapdown inertial navigation system with gyroscope biases of $10^\circ/\text{h}$ and accelerometer biases of 10 milli-*g*. The growth of attitude and position errors is largely determined by *g*-dependent gyroscopic biases, sensor scale-factor inaccuracy and initial misalignments of the on-board system. Also shown in the figure, for comparison purposes, are the errors which arise in a missile flying straight and level for a similar period of time. The importance of taking account of vehicle manoeuvres when assessing inertial system performance is clear.

Similarly, a combat aircraft may need to perform a variety of manoeuvres during the course of a mission. Examples include 'jink' or 'S' manoeuvres in which the aircraft performs a series of co-ordinated turns for purposes of low-level terrain avoidance or tactical evasion and 'pop-up' manoeuvres for ground attack sorties. The effect of these manoeuvres on the overall navigation accuracy achieved during flight is often highly dependent on the precise order and timing of the aircraft manoeuvres [7].

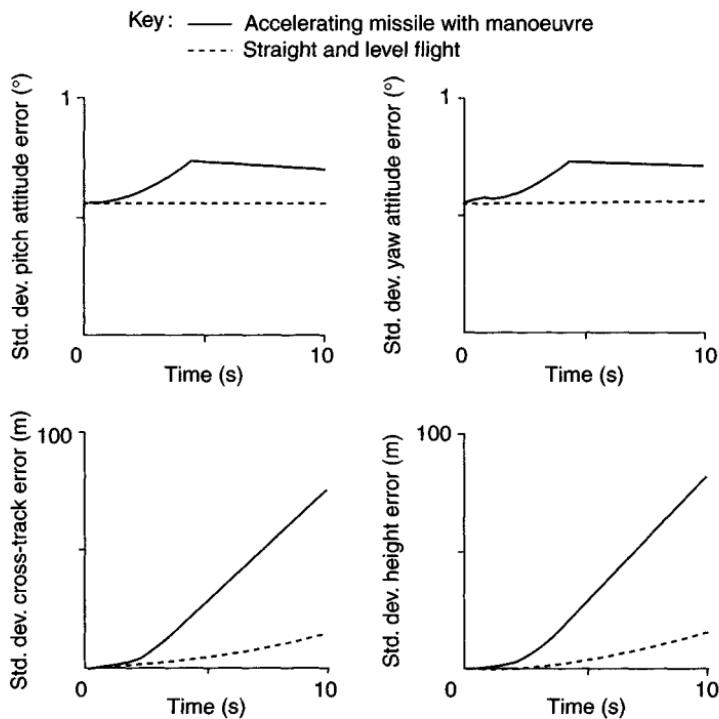


Figure 12.15 Illustration of manoeuvre-dependent error propagation

12.6.1.1 De-correlation of error terms

When an inertial system is aligned using gyrocompassing techniques, as described in Chapter 10, residual tilt and heading errors remain which are caused by gyroscope and accelerometer biases. For example, alignment to the local vertical may be achieved by adjusting the stored direction cosine matrix until the resolved measurements of specific force in the horizontal plane become zero. On achieving this condition, biases in the accelerometer measurements are off-set or cancelled by the 'tilt' errors. Consequently, the 'tilt' errors and biases are then said to be correlated.

Following this alignment process, provided the orientation of the inertial sensors with respect to the navigation reference frame in which the 'tilt' errors are defined remains fixed, as occurs normally in a platform system, neither the biases nor the 'tilt' errors propagate as navigation errors. However, in a strapdown system, the orientation of the sensors with respect to the navigation reference frame is unlikely to be maintained with the result that these errors will not remain correlated for long. As soon as the vehicle rotates in the reference frame, the instrument biases are no longer cancelled by the 'tilt' errors. In fact, a rotation of 180° about the vertical will result in a reinforcement rather than a cancellation of the errors, in which case both the 'tilt' error and the bias propagate separately giving rise to errors in the navigation system estimates of velocity and position. It is precisely this effect which gives rise to the 'Schuler pumping' effect discussed in the following section.

12.6.1.2 Schuler pumping

As shown in Section 12.4, for flight times up to a few hours, error propagation in the horizontal channels of an inertial navigation system is governed by the behaviour of the so-called Schuler loop. The Schuler loop may be represented as an undamped oscillator with a natural period of 84.4 min. It is to be expected that by stimulating this loop with a particular error signal at or near its natural frequency, a large and increasing error would result. For example, errors of this type may arise in an airborne navigation system if the aircraft executes a series of 180° turns at intervals of 42 min. This effect is referred to as ‘Schuler pumping’ [7, 8]. Although not unique to strapdown inertial navigation systems, it is more likely to occur in strapdown systems than other types of inertial system mechanisation. This is because the correlation which exists between certain of the error sources, attitude errors and accelerometer biases, for example, is maintained in a platform system but is lost in a strapdown system when the host vehicle changes course.

12.6.2 Vibration dependent error terms

This section is concerned specifically with the effects of vibratory and oscillatory motion on the performance of a strapdown navigation system. The various errors considered below are categorised as follows:

Instrument rectification errors: as the name implies, such errors arise through the rectification of the applied oscillatory motion by the sensor, and manifests itself as a bias giving rise to an erroneous measurement of the vehicle motion.

System errors: these errors refer to bandwidth limitations and imperfections in the strapdown computation which inhibit the capability of the system to follow both angular and translational oscillatory motion correctly. Coning and sculling motion are of particular significance in this context.

Pseudo-motion errors: these errors are caused by false instrument outputs which the navigation system interprets incorrectly as true vehicle motion; pseudo-coning motion is a typical example.

Examples of each are described briefly below.

12.6.2.1 Instrument rectification errors

Many of the inertial sensor errors described in Section 12.5 are functions of products of the applied angular rates or linear accelerations – anisoelasticity, scale-factor linearity and vibro-pendulous errors for example. Vibratory motion will be rectified by such effects resulting in additional biases on the inertial sensor outputs.

As an example, consider the effect of anisoelasticity in a spinning mass gyroscope which will give rise to a measurement bias ($\delta\omega$) which is a function of the linear acceleration acting simultaneously along orthogonal axes, viz:

$$\delta\omega = B_{ae}a_xa_y \quad (12.40)$$

where a_x and a_y represent components of applied acceleration acting along orthogonal axes x and y , respectively, and B_{ae} is the anisoelastic coefficient. In the presence of

sustained oscillatory motion of frequency ω and phase difference φ between the two axes of motion, $a_x = A \sin \omega t$ and $a_y = A \sin(\omega t + \varphi)$, the bias becomes:

$$\delta\omega = B_{ae} A^2 \sin \omega t \sin(\omega t + \varphi) = 0.5 B_{ae} A^2 \{\cos \varphi - \cos(2\omega t + \varphi)\} \quad (12.41)$$

The mean value of this expression, $0.5 B_{ae} A^2 \cos \varphi$, represents a constant bias which will be maximised when the accelerations in the two channels are exactly in phase. Consider a single-axis spinning mass gyroscope which is subjected to a sustained sinusoidal oscillation of amplitude $10g$ in a direction at 45° to its spin and input axes, and normal to its output axis. If the magnitude of its anisoelastic coefficient is $0.5^\circ/h/g^2$, an angular rate bias of $25^\circ/h$ would result.

12.6.2.2 System errors

Oscillatory motion can give rise to navigation system errors owing to limited sensor bandwidth, dynamic mismatch between sensors and insufficient computational speed which prevent the system from interpreting such motion correctly. The effects of cyclic angular motion by the vehicle, known as coning motion, or combinations of angular and translational motion, known as sculling, can be particularly detrimental to system performance. If the navigation system fails to detect such motion or to process accurately the inertial measurements obtained in the presence of such motion, significant navigation errors can arise.

12.6.2.3 Coning errors

Coning is the conical (or near conical) motion in inertial space of one of the gyroscope input axes, as illustrated in Figure 12.16. Such motion results from the simultaneous application of angular oscillations about two orthogonal axes of the system, where the oscillations are out of phase.

Taking the situation where a single gyroscope is subjected to motion such that its input axis follows a closed conical path, it can be shown [9, 10] that σ , the attitude of the gyroscope with respect to its initial position, after a full cycle of the

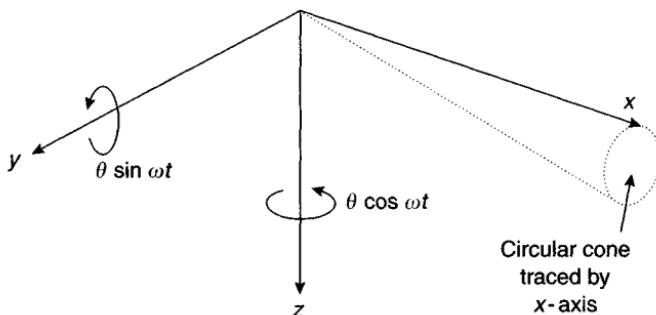


Figure 12.16 Coning motion

motion is given by:

$$\sigma = \int_0^T \omega dt + \varepsilon \quad (12.42)$$

where ω is the angular rate of the gyroscope about the line of the cone axis (as if no coning motions were taking place), T is the time taken to complete one coning cycle and ε is an additional rotation caused by the movement of the input axis around the closed path. The error term ε is identically equal to the area traced out by the input axis on the surface of a unit sphere centred on the origin of the gyroscope axis. The 'error' ε , is a real effect, correctly measured by the gyroscope. It results from the fact that the input axis of the gyroscope is slightly displaced from its nominal x -direction, as illustrated in Figure 12.16. The consequence is that when a small rotation is caused by $\theta \sin \omega t$ about the y -axis, the gyroscope senses a small amount of the $\theta \cos \omega t$ rotation about the z -axis and vice versa. These small rotations keep changing size and sign. If, as shown in Figure 12.16, the motions about y and z are at the same frequency and not in phase, then there is a net angle sum.

Thus, coning is purely a geometric effect resulting from the real motion of the gyroscope. If the attitude is computed solely on the basis of this one measurement, in the absence of knowledge of the cyclic rotations which have taken place about the other axes, the value of σ will be in error by ε . It therefore appears that the measurement of angular rate is in error owing to the presence of a bias of ε/T . This bias is referred to as the coning error and would be present even if a perfect gyroscope without any errors was used.

In a strapdown inertial navigation system which contains three single-axis gyroscopes, or an equivalent configuration, mounted such that their input axes are mutually orthogonal, the cyclic motions can be measured accurately and taken account of in the full attitude computation process. However, a coning error will still arise if the bandwidths of these gyroscopes are insufficient to measure or observe the angular motion, or if the attitude computation process is not performed at a sufficiently high rate.

Consider now the case of classical coning motion in which sinusoidal motions which are 90° out of phase are applied about two orthogonal axes. In addition, a constant rate is applied about the third axis in order to ensure that the body returns to its original position at the end of each coning cycle. For small rotations, the instantaneous attitude of the body may be expressed as:

$$\sigma = [-\theta \cos \beta t \quad \theta \sin \beta t \quad 0]^T \quad (12.43)$$

where θ is the amplitude of the coning motion and β is the frequency. The body may be returned to its original position at any time by rotating through an angle equal to the magnitude of σ , θ in this case. The associated angular rate is given by:

$$\omega = \left[\begin{array}{ccc} \beta \theta \sin \beta t & \beta \theta \cos \beta t & \frac{\beta \theta^2}{2} \end{array} \right]^T \quad (12.44)$$

Failure of the inertial navigation system to keep track of the oscillatory components of ω means that the measured motion, denoted ω' , and the computed attitude, σ' ,

will be as follows:

$$\omega' = \begin{bmatrix} 0 & 0 & \frac{\beta\theta^2}{2} \end{bmatrix}^T \quad (12.45)$$

$$\sigma' = \begin{bmatrix} 0 & 0 & \frac{\beta\theta^2 t}{2} \end{bmatrix}^T \quad (12.46)$$

that is, the computed attitude drifts at a rate $\beta\theta^2/2$ about the coning axis.

Hence, coning motion of 0.1° at a frequency of 50 Hz (~ 300 rad/s), for instance, can result in a drift in the computed attitude of almost $100^\circ/\text{h}$. Clearly, this can be a very significant error. The effects of coning motion on navigation system performance are considered in detail in References 9 and 11.

A more general development shows that the coning error which arises when the phase shift between the two orthogonal rotations is γ , and their respective amplitudes are θ_x and θ_y , can be expressed as follows:

$$\text{Coning error} = \frac{1}{2}\beta\theta_x\theta_y \sin \gamma \quad (12.47)$$

Clearly, the resulting drift error is maximised when γ is 90° and falls to zero when γ is zero.

12.6.2.4 Sculling errors

Sculling is made up of a combination of linear and angular oscillatory motions of equal frequency in orthogonal axes. In the presence of such motion, errors can arise in the strapdown computing task which is concerned with the resolution of the measured specific force vector into the chosen navigation reference frame. An acceleration bias can arise through failure to take account of the rapid changes of attitude occurring between successive specific force vector resolutions.

For example, if a vehicle rotates sinusoidally about its y -axis such that $\theta_y = \theta \sin(\omega t + \varphi)$, whilst oscillating linearly along its z -axis such that $a_z = A \sin \omega t$, a component of the acceleration ($A_z \sin \theta_y$) will appear in the x -direction if these rotations are not correctly sensed and resolved into the navigation reference frame. For small angle perturbations, the x -component of acceleration can be approximated as:

$$a_B = a_z \theta_y \quad (12.48)$$

Substituting for a_z and θ_y gives:

$$a_B = 0.5A\theta \{ \cos \varphi - \cos(2\omega t + \varphi) \} \quad (12.49)$$

Therefore, a steady acceleration component of $0.5A\theta \cos \varphi$ occurs in the x -direction. It is stressed that this error term can arise even when using perfect accelerometers, being purely a function of the inaccuracy in the resolution process. If, for example, $A = 10 \text{ g}$, $\theta = 0.1^\circ$ and the phase difference is zero, the resulting acceleration bias is $\sim 9 \text{ milli-g}$.

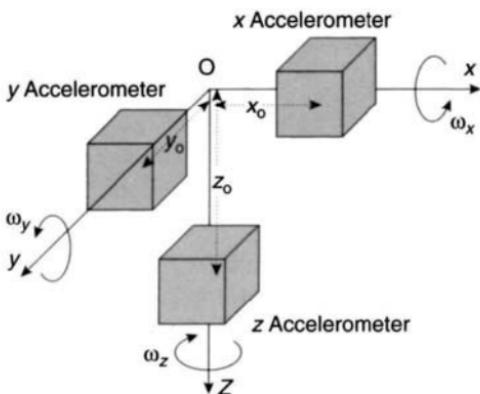


Figure 12.17 Illustration of size effect

12.6.2.5 Size effect errors

The specific force acting on a vehicle is detected by sensing motion along three orthogonal axes, often using a triad of linear accelerometers. In order to navigate, it is required to sense the linear accelerations acting at a particular point in the vehicle, at its centre of gravity for example. Whether the inertial navigation system is mounted precisely at the vehicle centre of gravity or, as is more usual, at some off-set location, it provides a measure of the motion of that point within the vehicle. This assumes that the inertial system is able to sense all motion accurately, including any centripetal and tangential forces induced by vehicle rotation, and to process accurately the inertial measurements which are generated.

Ideally, it is required that all three accelerometers should be mounted precisely at the same location in the vehicle. This is clearly impossible because the sensors are of finite size and because of design constraints on the positioning of hardware. The centripetal and tangential forces sensed by the accelerometers because of their physical displacements with respect to the desired position are referred to as the 'size' effect.

Consider the situation depicted in Figure 12.17 where the sensitive axes of the accelerometers intersect at the point O. The x-axis accelerometer is mounted with its sensitive element displaced a distance x_0 from O and its sensitive axis pointing along that axis. In the presence of angular rates, ω_y and ω_z about the y and z axes, respectively, the x accelerometer will be subject to a centrifugal acceleration:

$$a_x = -(\omega_y^2 + \omega_z^2)x_0 \quad (12.50)$$

Similarly, the y and z sensors will sense accelerations:

$$a_y = -(\omega_x^2 + \omega_z^2)y_0 \quad (12.51)$$

$$a_z = -(\omega_x^2 + \omega_y^2)z_0 \quad (12.52)$$

In the presence of continuous rotations, as occur in a freely rolling missile for example, the effect of these additional accelerations will integrate to zero over a few cycles provided that the resolution of the measurements into the navigation reference frame is implemented accurately. Size effect errors can arise in this situation as a result of imperfections in the strapdown computational algorithms.

Of particular concern here is the effect of oscillatory motions which will be rectified to give steady acceleration errors. For instance, if $\omega_y = \omega\theta_y \sin \omega t$ and $\omega_z = \omega\theta_z \sin(\omega t + \varphi)$, then:

$$\begin{aligned} a_x &= -\{\omega^2\theta_y^2 \sin^2 \omega t + \omega^2\theta_z^2 \sin^2(\omega t + \varphi)\}x_0 \\ &= -\frac{1}{2}\omega^2(\theta_y^2 + \theta_z^2)x_0 + \frac{1}{2}\omega^2\{\theta_y^2 \cos 2\omega t + \theta_z^2 \cos(2\omega t + 2\varphi)\}x_0 \end{aligned} \quad (12.53)$$

Thus, a steady bias acceleration of magnitude $0.5\omega^2(\theta_y^2 + \theta_z^2)x_0$ is introduced as a result of size effect. In the presence of cyclic angular motion of amplitude 0.1° , and frequency 50 Hz, for example, a bias of ~ 1.5 milli-*g* arises for a 10 cm displacement. This is not related in any way to imperfections in the accelerometer and will arise even when perfect sensors without any errors are used in this configuration.

12.6.2.6 Pseudo-motion errors

The inertial sensors themselves may produce false signals which are correlated with each other. The navigation system will then interpret these signals as indicating the presence of coning or sculling motion for instance, where no such motion is actually present. The apparent motions are sometimes referred to as pseudo-coning or pseudo-sculling. If these motions are treated as true by the navigation system computer, the performance of the system will be degraded.

For example, pseudo-coning can arise in systems which use spinning mass gyroscopes as a result of the angular acceleration sensitivity of such sensors. In the presence of a cyclic angular rate about a single-axis, $\omega\theta \sin \omega t$, an apparent rate may be detected about an orthogonal axis proportional to the applied angular acceleration, $\omega^2\theta \cos \omega t$. The vehicle will therefore appear, according to the measured rates, to exhibit coning motion as characterised by the two cyclic oscillations which are 90° out of phase. For a system using rate-integrating gyroscopes, it can be shown [12] that the resulting bias, ω_b , is given by:

$$\omega_b = \frac{I}{2H}\omega^2\theta^2 \quad (12.54)$$

where H is the angular momentum of the gyroscope and I is the moment of inertia of the float assembly, as described in Chapter 4.

Pseudo-coning can also arise in the absence of any applied motion, that is, purely as a result of gyroscope imperfections. In a dual-axis sensor, such as the dynamically tuned gyroscope, outputs will arise if there is a misalignment between the rotor and the pick-offs which sense angular displacements about the two input axes of the sensor. These outputs will approximate to sinusoidal functions of time. Since the pick-offs are displaced by 90° about the spin axis, the two outputs will be 90° out of phase thus giving an erroneous indication of coning motion at the gyroscope spin frequency.

If the spin frequency is 1000 rad/s, a misalignment of 1 arc minute will result in a coning error of 1°/h.

12.7 Summary

The performance accuracy of an inertial navigation system can be expressed in terms of a series of equations. Inaccuracies arise in such a system because of initial alignment errors, imperfections in the performance of the inertial instruments and limitations in the computational process. These errors can be quantified enabling a designer to estimate the performance accuracy of a proposed navigation system. The analysis can be simplified in some circumstances, such as for very short duration flight. In other cases, particularly where there is coupling between channels, a deterministic solution is not possible, and simulation is necessary to provide accurate information on performance.

Inertial sensors are sensitive to various external stimuli. Fortunately these sources of error are frequently well behaved and consequently can be expressed as a deterministic equation, the coefficients of each term representing the various sensitivities to a given stimulus. Care must be taken when processing the various sensor signals in the presence of angular motion, particularly in the presence of coning and sculling motion. In these cases, the bandwidths of the sensors and the speed of the computation must be high enough to sense and record the actual motion, otherwise significant errors can arise, even if 'perfect' sensors were to be available.

References

- 1 KING, A.D.: 'Characterisation of gyro in-run drift'. DGON proceedings, *Gyro Technology Symposium*, Stuttgart, 1984
- 2 FLYNN, D.J.: 'The effect of gyro random walk on the navigation performance of a strapdown inertial navigator'. DGON proceedings, *Gyro Technology Symposium*, Stuttgart, 1982
- 3 LIN, C.-F.: 'Modern navigation, guidance and control processing' (Prentice-Hall, 1991)
- 4 LANING, J.H., and BATTIN, R.H.: 'Random processes in automatic control' (McGraw-Hill, New York, 1956)
- 5 NESLINE, F.W., and ZARCHAN, P.: 'Miss distance dynamics in homing missiles'. AIAA Paper 84-1845, August 1984
- 6 ALPERT, J.: 'Miss distance analysis for command guidance missiles', *Journal of Guidance*, 1988, 11 (6)
- 7 WATSON, N.F., and CAMPBELL, E.A.F.: 'Cost-effective strapdown INS design and the need for standard flight profiles'. Proceedings DGON, *Gyro Technology Symposium*, Stuttgart, 1987
- 8 FENNER, P.J.: 'Requirements, applications and results of strapdown inertial technology to commercial airplanes'. AGARD Lecture Series No. 133, 'Advances in strapdown inertial systems', May 1984

- 9 GOODMAN, L.E., and ROBINSON, A.R.: 'Effect of finite rotations on gyroscopic sensing devices', *Journal of Applied Mechanics*, 1958
- 10 FLYNN, D.J.: 'A discussion of coning errors exhibited by inertial navigation systems'. Royal Aircraft Establishment Technical Memorandum, Rad-Nav 243, 1984
- 11 CRAIG, R.J.G.: 'DTGs in strapdown systems'. AGARD conference proceedings, No. 116, AD-758 127, February 1973
- 12 NASA: 'A study of the critical computational problems associated with strapdown inertial navigation systems'. NASA CR-96, 1968

Chapter 13

Integrated navigation systems

13.1 Introduction

For many vehicles requiring a navigation capability, there are two basic but conflicting requirements to be considered by the designer, namely those of achieving high accuracy and low cost. This chapter examines the scope for satisfying these demanding requirements by using integrated navigation systems, in which strapdown inertial navigation systems are used in conjunction with other navigation aids. The variety of modern navigation aids now available is extensive and, coupled with advances in estimation processing techniques and high-speed computer processors, have resulted in greater application of integrated navigation systems in recent years.

As discussed in Chapter 12, the performance of an inertial navigation system is characterised by a time-dependent drift in the accuracy of the position estimates it provides. The rate at which navigation errors grow over long periods of time is governed predominantly by the accuracy of the initial alignment, imperfections in the inertial sensors that the system uses and the dynamics of the trajectory followed by the host vehicle. Whilst improved accuracy can be achieved through the use of more accurate sensors, there are limits to the performance that can reasonably be achieved before the cost of the inertial system becomes prohibitively large. Very expensive inertial navigation systems are incompatible with most applications except for those special strategic systems where there is no easy alternative, such as submarine navigation or other strategic platforms and missiles. However, as discussed in Chapter 15, alternative mechanisation techniques may be used to ease the demands on sensors' performance, whilst maintaining system performance.

An alternative approach that has received much attention in recent years, and is suitable for many applications, is known as integrated navigation. This technique employs some additional source of navigation information, external from the inertial system, to improve the accuracy of the inertial navigation system. Careful selection of fundamental characteristics leads to low cost, but potentially very accurate and reliable navigation.

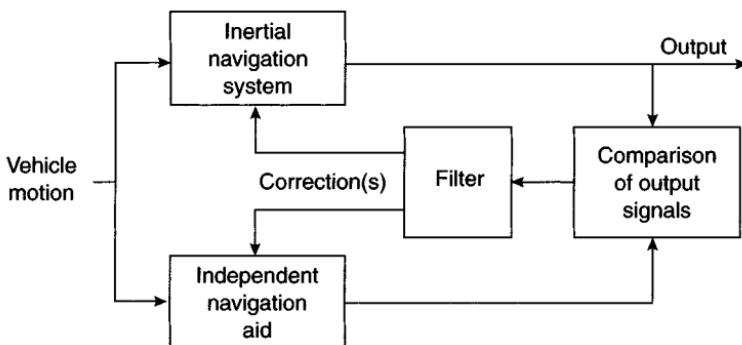


Figure 13.1 Basic principle of an integrated navigation system

13.2 Basic principles

In an aided inertial system, one or more of the inertial navigation system output signals are compared with independent measurements of identical quantities derived from an external source. This is illustrated in Figure 13.1. Corrections to the inertial navigation system are then derived as functions of these measurement differences. By judicious combination of this information, it is possible to achieve more accurate navigation than would be achieved using the inertial system in isolation.

As a simple example, take the case of an aircraft that is able to detect a radio signal when it over-flies a beacon on the ground. Provided the aircraft has precise knowledge of the position of the beacon, an accurate position fix is provided at the instant it over-flies the beacon and this fix may be used to update an on-board inertial navigation system. In the event of a number of such position fixes being available, at discrete intervals of time, it is possible to update other quantities within the inertial system which are not directly measurable. For example, it may be possible to update and improve the inertial system estimates of velocity and heading.

Integrated systems of this type usually make use of two independent sources of information with complementary characteristics; it is common to use one source providing data with good short term accuracy, whilst the second source provides good long term stability. For example, a radio beacon can provide accurate position fixes at discrete intervals of time and so bound the long term drift of an inertial navigation system. Meanwhile, the inertial system provides low noise continuous navigation data between the fixes which are accurate in the short term and not subject to external interference.

In broad terms, the various types of navigation aid that are available may be categorised under the following headings:

External measurements: Measurements obtained by receiving signals or by viewing objects outside the vehicle requiring navigation. Such observations may be provided by radio navigation aids, satellites, star trackers or a ground-based radar tracker, for example. In some cases, data may be transmitted to the vehicle during

its journey, whilst in others, there will be a 'receiver' or 'viewer' to accept the observations. Navigation aids of this type usually provide a position fix, which may be expressed either in terms of vehicle latitude and longitude or as co-ordinates with respect to a local reference frame.

On-board measurements: Measurements derived using additional sensors carried on-board the vehicle requiring navigation. Navigation aiding of this type may be provided by altimeters, Doppler radar, airspeed indicators, magnetic sensors and radar or electro-optical imaging systems. Such sensors may be used to provide attitude, velocity or position updates, any of which may be used to improve the performance quality of the inertial navigation system.

Navigation aids that fall into these two categories are described in the two sections which follow. The later sections are concerned with methods of mixing measurement data provided by different navigation sensors or systems to form an integrated navigation system.

13.3 External navigation aids

13.3.1 Radio navigation aids

Radio navigation, based on ground-based transmitting stations, is perhaps the oldest of the modern navigation aids. The development of radio direction finders for both ships and aircraft allowed bearings to any radio transmission station at a known location to be determined and used for navigation purposes. Given measurements of bearing to two or more ground stations at known locations, the position of the vehicle may be calculated by the process of triangulation.

Many communications and broadcasting stations use low and medium electromagnetic wave frequencies to obtain large areas of coverage and these stations can be used at long ranges, well beyond the visual line of sight. However, radio propagation at these frequencies is affected by atmospheric conditions and care has to be taken when using them at night. More accurate measurements can be obtained at higher frequencies, although their range is more restricted.

To overcome some of the problems that arise when using simple direction finding equipment, a number of systems were developed based on the use of modulated radio beams. In such systems, the modulation received at the aircraft is dependent on the position of the vehicle in the beam, hence providing navigational information. A widely adopted scheme is very high frequency omni-directional radio range (VOR).

13.3.1.1 Very high frequency omni-directional radio range

This is a short-range navigation aid, primarily for aircraft use. The ground station has an omni-directional aerial, which transmits a very high frequency (VHF) carrier amplitude modulated by a reference signal. A series of other aerials are situated around the reference aerial. These transmit a constant carrier frequency, which is switched between them to simulate a cardiac-shaped beam rotating once per cycle of

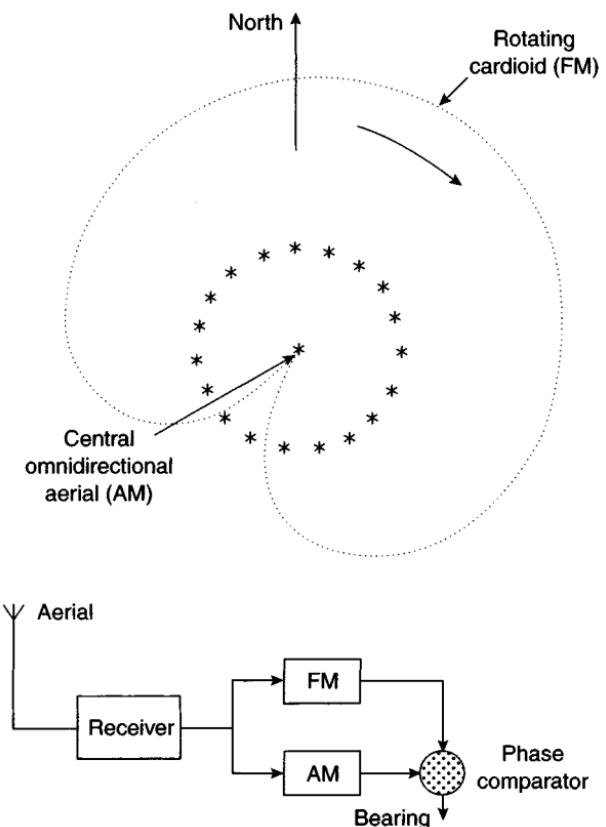


Figure 13.2 Very high frequency omni-directional radio range system

the reference signal, as indicated in Figure 13.2. At a receiver, this gives a frequency modulated carrier modulated at the rotation frequency.

The timing of the rotating beam is adjusted so that for a receiver situated due north of the VOR station, the frequency modulation is in phase with the amplitude modulation. The modulation phase at any other location is then equal to the angle from north along which the receiver is situated. Thus, by measuring this phase angle, the aircraft is able to determine the bearing from the station to itself. The reference direction at the VOR station is magnetic north.

The operation of the VOR navigational aid depends on maintaining line of sight contact between the aircraft and the ground stations. For this reason, operational range varies with aircraft altitude. At 300 m altitude, for example, the effective range is about 75 km, whilst at 6 km altitude this increases to 350 km. Typical bearing measurement accuracies are about 2° . The error in a position fix derived using VOR increases with the range from the ground stations.

A major improvement in navigational accuracy is possible if the distance to the radio stations can be determined. This can be achieved by transmitting signals at

known times and measuring the time of arrival at the receiver. Since the propagation velocity is known, distance can be determined from the measured time delay. This is the basis of virtually all modern radio navigation systems, including satellite-based approaches such as GPS, which is described later, the differences being in the means of timing and the transmission frequencies.

To measure the time delay between transmission and reception of a radio signal, the transmitter and receiver must have clocks that are synchronised to a common time. This is not particularly easy since a $3\text{ }\mu\text{s}$ timing error corresponds to a range error approaching 1 km. Assuming that clock resynchronisation can be performed only once per hour, this corresponds to a drift in the measurement of time of one part in 10^9 . Such accuracies were not achievable before the invention of atomic clocks, and other methods of measuring range were sought to overcome this difficulty.

A number of ranging systems are described below.

Distance measuring equipment (DME)

Many VOR stations are equipped with a microwave transponder, which provides range information in response to a signal emanating from the aircraft. The aircraft transmits pairs of pulses with a unique separation and repetition rate. A ground station receives the signals and retransmits them after a fixed time delay. The aircraft receives the retransmitted signals and measures the time delay between transmission and reception, deducts the fixed delay in the ground station, and so determines the two-way range, i.e. to and from the ground station. Position can then be determined by measuring the range to two or more DME stations. DME observations are accurate typically to within 300 m. The principle of the method is shown in Figure 13.3.

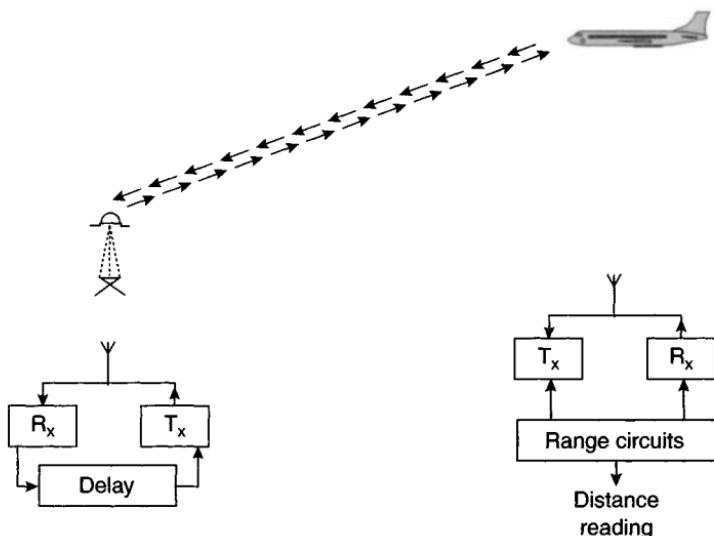


Figure 13.3 Distance measuring equipment

Tactical air navigation system (TACAN)

The tactical air navigation system provides the same type of measurements as those obtained using VOR and DME, as described above, but with increased precision through the use of ultra-high frequency (UHF) transmissions in the 1 GHz frequency band. Typical bearing accuracy is in the region of $\pm 0.5^\circ$ with ranging errors usually better than $\pm 1\%$ of the distance between the aircraft and the beacon. The maximum range is altitude dependent because of the characteristics of the propagation of UHF radio waves.

Hyperbolic systems

The need for accurate absolute time in the receiver is also eliminated if signals are transmitted in synchronism from two or more ground stations and the time interval between their arrivals at the receiver is measured. In systems based on this principle, a master station transmits a signal that is received in the aircraft or ship and also at slave transmitter stations on the ground. The slave stations lock their clocks to the master signal, with allowance for the propagation time for the fixed distance between master and slave. The corrected clocks are then used to generate the slave station transmissions.

At the receiver, its clock can be locked to the received master signal and used to measure the time interval to the receipt of the slave signals. The long-term stability is now governed by the stability of the master station clock, and the short-term stability of crystal clocks is perfectly adequate for the time difference measurements. The time interval between signals received from two stations gives a measure of the difference in range to the stations. A given reading indicates a position for the receiver, which lies on a hyperbola with the two stations at its foci, as indicated in Figure 13.4. By measuring the time intervals obtained between three stations, two hyperbolae and hence a fix is obtained.

Two methods of measuring range differences are used by terrestrial hyperbolic navigation radio aids, namely:

- phase measurements from continuous waves;
- direct time measurements from pulse transmissions.

Decca navigation system

This is a typical hyperbolic system; it uses continuous waves and phase measurements to determine hyperbolic position lines. Transmissions are in the low frequency band from 70 to 130 kHz giving a usable range of around 250 km. Four stations, a master and three slaves, form a chain of transmitting stations. The slave transmissions are phase locked and harmonically related to the transmissions from the master station.

A complete phase cycle of 360° at the comparison frequencies represents a distance of between 500 and 800 m on the baseline between the stations. Hence measuring phase to an accuracy of 10° gives a resolution of around 5 m. However, the resulting position is ambiguous since the same phase measurement repeats every cycle or 500 m. Special facilities, which essentially involve making phase comparisons at synthetically produced lower frequencies, have to be introduced to resolve the ambiguity.

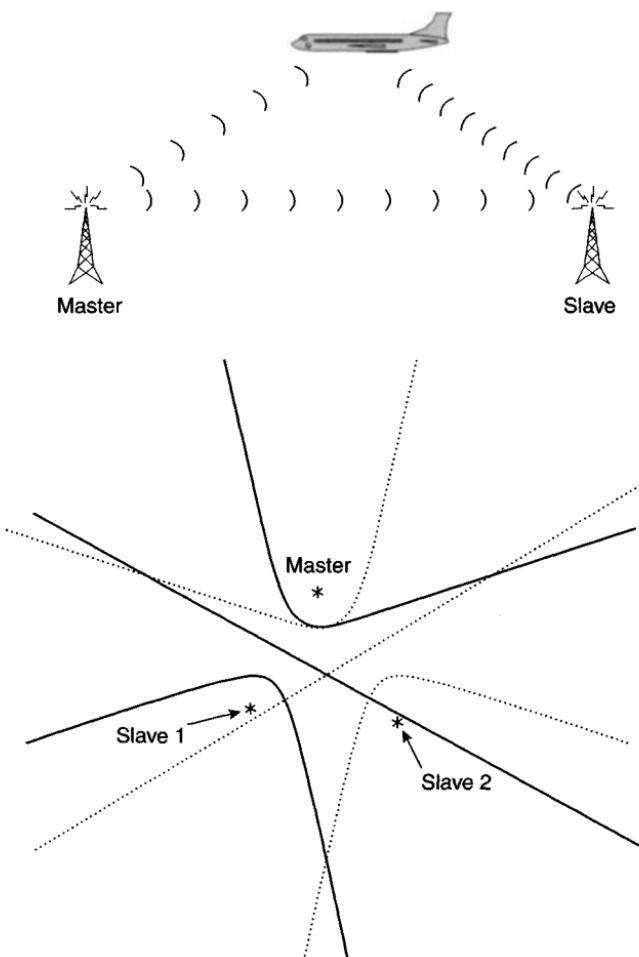


Figure 13.4 Hyperbolic navigation

When originally introduced during World War II, the system required maps with special overlays printed on them to obtain latitude and longitude information. This limitation has been removed through the use of computers to produce systems that provide latitude and longitude information automatically.

Omega

This is a long-range hyperbolic navigation aid operating at very low frequency. The system is based on eight ground transmitting stations distributed around the Earth, each having a nominal range of 8000 nautical miles. Thus, an aircraft or ship located anywhere around the Earth can expect to receive signals from at least three stations, and is able to deduce its position from the phase of the received signals, in the manner outlined above. Typically, a position fix can be obtained to an accuracy of a few nautical miles.

Although the Omega system was developed originally for maritime use, it has been extensively used as an airborne navigation aid, in particular on the trans-oceanic routes.

Long-range navigation (LORAN)

This is a low frequency electronic position fixing system using pulsed radio waves with a frequency of 100 kHz. It is a long-range aid (1500 km or beyond) obtained by using pulse transmissions rather than continuous waves. It is more precise than Omega but does not have the worldwide coverage of that system.

The Loran C system has a master transmitting station and two or more slave stations forming a chain. There are many chains located in the northern hemisphere. The system operates by measuring the difference in time of arrival of pulses from the master and its slave stations. The ground wave from the transmitters travel distances of up to 2000 km, the exact range depending on the power of the transmitter, receiver sensitivity and atmospheric attenuation. Position accuracies are dependent on the distance of the observer from the chain and vary from a few tens of metres at short ranges, typically less than 500 km, to 100 m, or more, at longer distances of around 2000 km. The transmissions also produce sky waves, which can interfere with the ground waves causing distortion of the received signals and errors in the estimates of position provided by the system.

As with the Omega system, LORAN was developed originally for ships but has found wide use as an airborne navigation aid, and more recently, as an aid for land vehicle location.

Following the advent of satellite navigation, some terrestrial radio navigation systems, Omega and Decca, have become obsolete. However, while satellite systems do not meet all the integrity and reliability requirements imposed by some applications, including civil aviation, terrestrial radio navigation systems continue to provide essential back-up. However, as satellite based augmentation systems become established (see Section 13.3.2.1), the need for VOR/DME and Loran-C is likely to diminish.

13.3.2 Satellite navigation aids

Radio positioning, similar to that described in the previous section, can be achieved using satellite transmissions. The first satellite navigation system was developed for the US Navy and became operational in January 1964 [1]. It was known as TRANSIT and provided:

- 24 hour operation;
- all weather operation;
- two-dimensional positioning accuracy.

However, position updates were not continuous owing to the number of observations required to make a navigation estimate, and because there was up to an average of 100 min between successive satellite passes. Moreover, each position fix required observation of the satellite signals over a 10–16 min period causing

accuracy to be degraded by ship's motion. Therefore this system was not suitable for other platforms, such as aircraft that require virtually continuous updates.

Operational support for this pioneering satellite navigation system ceased in 1996. TRANSIT has long been superseded by the global positioning system (GPS) which is described in the following section.

13.3.2.1 Global positioning system

The global positioning system, or GPS, also known as Navstar¹, is a radio positioning system which has now reached full operational status providing a worldwide navigation capability [2–5].

GPS is a satellite navigation system designed to provide highly accurate, three-dimensional position and velocity data to users anywhere on or near to the Earth. The system is available to an unlimited number of users, each equipped with an antenna and a receiver. GPS comprises a constellation of 24 satellites² in near-circular orbit around the Earth, as shown in Figure 13.5, and a ground control system. The satellites orbit the Earth in about 12 h and are arranged in six orbital planes which are inclined at 55° to equatorial plane, at an altitude of 20 180 km. The spacing of the satellites is arranged so that, generally, at least six satellites will be in view to a user at any instant of time.

Each satellite transmits two carrier frequencies in L band, known as L₁ (1575.42 MHz) and L₂ (1227.6 MHz), each signal being derived from an atomic frequency standard. Each of these signals is modulated by either or both of the precise positioning service (PPS or P) signal (10.23 MHz) and the standard positioning service (SPS), also called the coarse/acquisition (C/A) signal (1.023 MHz). The binary signals are created by a P-code or C/A-code, which is modulo 2 added to 50 bps data. The P-code and C/A-code are added to L₁ in phase quadrature. Note that only the P-code is present on the L₂ signal. These techniques are in the process of being revised.

Position is calculated by taking measurements of distance from each of the satellites 'in view'. The distance measurements are made by measuring how long it takes for a radio signal to travel from each satellite to the GPS receiver. It is assumed that the satellites and the GPS receiver are generating the same coded signal at exactly the same time. Distance is determined by comparing the arrival time of the satellite signal with that expected by the receiver. The timing signals used by GPS are pseudo-random sequences which enable each satellite to be identified unambiguously and also allow access to the system to be controlled. Two codes are transmitted by each satellite; the SPS code and the PPS code as described above. As its name suggests, the PPS code yields the full navigational precision of the system, but is only available to selected users.

¹ Navstar – Navigation by satellite time and range.

² The original constellation had 21 operational satellites and three fully functional spare satellites. Reliability analysis shows that this arrangement provides a 98% probability of system availability. In 2002, this was revised to 28 satellites to meet FAA requirements.

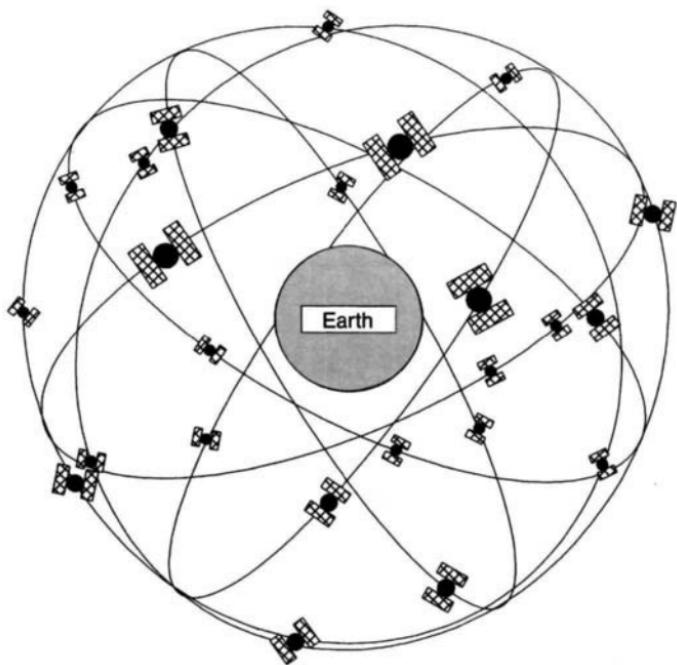


Figure 13.5 GPS satellite constellation

In order to make precise distance measurements, accurate timing of the satellite and receiver signals is clearly essential. The satellite signals are accurate because they have atomic clocks on board. Less accurate clocks are used in the receivers and their timing errors corrected by taking measurements of range to four satellites, four ranges being required to determine the four unknowns; three spatial coordinates and time.

It is also necessary to know the precise location in space of each satellite that is being monitored if position is to be determined accurately. The orbits of the satellites are very predictable. However, minor variations that do occur are monitored regularly by ground-based tracking stations. These data are passed to the satellites enabling them to broadcast information about their exact orbital location in addition to the timing signals discussed earlier.

By using the techniques outlined above, it is possible to obtain very accurate distance measurements. By measuring the Doppler shift of the radio frequency carrier, the range rate to each satellite can be calculated in the receiver. Using this information, the vehicle velocity can be determined since that of the satellite is known.

Measurement errors arise from a variety of causes. The Earth's ionosphere and atmosphere cause delays in the GPS signals³ that can give rise to errors in the

³ The time delay is caused by the free electron population density in the signal path. The delay is different for different frequencies and that is the purpose of having the two different L-band frequencies; a correction for the delay can be made when both frequencies are received.

measurement of position, although their effects can be compensated to some extent by modelling. Other sources of error are satellite and receiver clock imperfections and multi-path reception. Further, at certain times, the geometrical arrangement of the satellites being monitored can magnify the errors in the system by a process referred to as the geometric dilution of precision (GDOP). However, GPS can enable a vehicle to establish its position anywhere in the world and at any time with an accuracy of a few metres and its velocity to better than 0.1 m/s, although this degree of precision is not available to all users. This is described further in Appendix D.

The GPS position measurement is noisy, 2 m (1σ) owing to a variety of reasons:

- low signal strength;
- length of the pseudo-random code, about 300 m;
- resolution of the code-tracking loop.

Additionally, multi-path is a source of correlated noise that is particularly prevalent in applications involving moving vehicles. Velocity estimates are also noisy, owing to varying signal strength, changing multipath conditions and receiver-clock instability.

Differential GPS

A technique has been devised to give a substantial increase in the accuracy of the estimates of position compared to those which can be obtained from the so-called standard positioning service of the GPS satellite navigation system. This technique is known as differential GPS and requires the use of a receiver at a surveyed location, at least one other receiver and an accurate high-speed data link between them. The receiver unit at the surveyed position will be able to compare its GPS estimate of position with the known position from the survey, and thus compensate for range errors produced by the GPS. A correction signal can then be transmitted to other receivers in the immediate vicinity allowing errors in their measurements to be reduced dramatically.

The principle of differential GPS techniques is to take advantage of the fact that a substantial component of the navigation error in a GPS measurement arises from slowly changing biases. Moreover, these biases are correlated in both distance and time between an array of receivers. Therefore if two receivers, or more, are operating simultaneously at different locations and the position of one of them is known, then corrections can be generated in real time to the measurements from one receiver and applied to the other receiver measurements. Over the past year, the GPS control segment has improved its performance, particularly in Ephemeris and clock predictions, so these errors are only 2 m (1σ); a differential system will remove half this error.

It is also possible to undertake the reverse process and undertake precision tracking of a co-operative target, particularly one that is required to carry the minimum amount of equipment. An example is a test target on a firing range. The principle of the reverse technique is to compare the navigation solution achieved at a surveyed point with the known position from an accurate survey to determine the correction for the normal satellite navigational errors, which is normal 'differential GPS'. However, with this reverse technique, the correction to the navigation data recorded at a remote point is applied at a tracking station.

In this case a vehicle, which may be moving under commands on a range, carries a simple satellite navigation system receiver and a transmitter to send the raw received navigation data to the tracking station. The tracking system in the control station can then apply the calculated correction to the received data from the vehicle.

This technique may be extended to multiple remote vehicles, however, it is important that the transmitters do not interfere with each other and confuse the tracking station. The use of suitable spread spectrum transmission techniques, such as frequency hopping over a narrow frequency band may solve this problem.

The errors in this technique are very similar to those encountered with the normal differential GPS navigation technique. The correction is good when the baseline between the receiver at the surveyed position and the 'remote receiver' is small. This correction becomes less valid with increasing range, particularly if the satellites being viewed are not identical.

Relative GPS

Relative GPS refers to techniques that provide high relative positional accuracy between two GPS receivers, even though the absolute position of each receiver is not known precisely. This technique contrasts with absolute GPS where a single GPS receiver is used to determine the navigation estimates, and differential GPS where corrections are applied to absolute measurements made by a single receiver linked to a receiver whose position is well known. The differential technique can be used to correct an array of GPS receivers connected to the surveyed position via a data link, as described above.

The relative GPS technique can be used to remove the large and highly correlated common errors between two GPS receivers at a specific time, enabling relative accuracy estimates of less than a metre. In order to achieve this level of relative navigation accuracy it is important that all of the GPS receivers observe the same set of satellites. This is because the correlated and slowly varying errors cancel out when applying the relative navigation technique.

Clearly, the relative GPS technique applied between two or more GPS receivers aims to cancel out the bias-like correlated errors common to the two receivers. However, not all error sources are perfectly correlated and the correlation between those errors considered to be correlated decreases as separation between the receivers increases. Additionally, the so-called correlated errors also change slowly with time, consequently if there is a time delay between the calculation of the correction and their application there will be an additional error, owing to erroneous compensation.

High-precision navigation services

High-precision navigation techniques are required for some applications such as precision approach and landing systems. These techniques are based on the ability of a GPS receiver to make accurate measurements of the accumulated carrier phase, or integrated Doppler shift to the satellites being viewed. The accumulated phase measurements are calculated with respect to similar measurements made by a reference receiver. The result is high-precision knowledge of the relative positions between the various receivers. Consequently, if the reference receiver is at a surveyed position,

as discussed for the differential GPS method, then absolute differential position data are derived for the user receivers.

Two distinct types of algorithms may be used for the calculation of position from the received satellite data:

- use of relative accumulated carrier phase measurements to smooth relative code-phase measurements, which leads to positional accuracy of the order of a metre or better;
- ambiguity resolution of the relative accumulated carrier phase measurements, which leads to positional accuracy errors in the centimetre regime. Most of the ambiguity resolution algorithms require an accurate initial estimate of position, which is often calculated from the relative code-phase position technique [6].

The GPS space-based augmentation system (SBAS) is being implemented for applications that have a ‘safety-critical’ element, such as civil aviation. The function of SBAS is to warn users of a satellite problem. This system includes the European Geostationary Navigation Overlay System and the Wide Area Augmentation System in North America, which combine differential GPS with additional ranging and integrity monitoring. The SBAS system is designed with the capability to transmit a ‘health warning’ within 6 s of a satellite malfunction.

The GPS system is currently undergoing a period of modernisation in order to meet the changing requirements of both civil and military users [7, 8]. Additional signal codes are to be made available to provide improved correction for propagation delays in the ionosphere, improvements in the accuracy of code phase measurements and reduction of multi-path. Ultimately, these enhancements will be made available through the addition of a third GPS frequency band, centred on 1176.45 MHz. In addition, there is to be a new military code, which will allow precision and standard positioning signals to be separated, yielding enhanced security for military users.

A phased approach to the introduction of these system updates is being adopted over a 15-year period with M-code becoming fully operational in 2010. Full operation of the new frequency band is expected in 2014 and the provision of a spot beam facility, whereby precision positioning signals will be transmitted at higher power in selected regions, in 2018.

GPS is not the only satellite navigation system in operation. The GLONASS system was developed by the former Soviet Union at the same time as the GPS development was taking place. The European Union is also developing a system known as Galileo, principally for civil users. Outline descriptions of these systems are given in the following sections.

13.3.2.2 Global Navigation Satellite System

A system equivalent to GPS has been developed by the former Soviet Union, known as the Global Navigation Satellite System (GLONASS). The GLONASS system is designed to operate with a constellation of 24 satellites, with eight in each of three orbital planes that are 120° apart in longitude. The orbital planes are arranged with

an inclination of 64.8° to the equator [9]. However, the constellation has yet to be completed. The satellites operate at an orbital radius of 25 600 km.

Like GPS, GLONASS transmits on two carrier frequencies, L₁ and L₂, and provides both a military and a civil service. The coarse/acquisition (C/A)-code, available to all users, is modulated at 511 kHz whilst the precision (P)-code is modulated at 5.11 MHz. In contrast to GPS, GLONASS satellites transmit the same ranging codes, but on 21 pairs of frequencies. The L₁ frequencies are 1598.0625–1608.75 MHz at 562.5 kHz spacing. The L₂ frequencies are 1242.9375–1251.25 MHz at 437.5 kHz spacing. The higher frequency assignments in each band are being phased out owing to interference problems, so from 2005, only the lowest twelve pairs of frequencies will be used. As there are fewer frequency pairs than satellites, satellites on opposite sides of the Earth will share the same frequency.

All GLONASS satellites are equipped with a caesium-based frequency standard; the RMS accuracy of the mutual synchronisation is 20 ns. GLONASS accuracy and coverage is a little poorer than that of GPS as there are fewer operational satellites and there has been less investment in system improvements in recent years.

13.3.2.3 *Galileo*

This satellite navigation system is currently under development within Europe, funded by the European Union, The European Space Agency, national governments and the private sector. Unlike GPS and GLONASS, Galileo is intended primarily for civil users and will be under civil control. It is planned to launch the first satellite in 2005 with full service becoming available in 2012 at the earliest.

In addition to a planned basic service available to all users worldwide, with horizontal position accuracy of ~2 m (1σ), it is proposed to provide various commercial subscription services with higher levels of accuracy. These include a regulated system giving very high integrity, availability, continuity and resistance to signal interference for safety critical applications. The precise range of services and areas of availability have yet to be finalised, including the service available to the general public.

The space segment will comprise 30 satellites uniformly spaced in three orbital planes, 120° apart in longitude with an inclination of 54° to the equator. The orbital radius will be 29 600 km.

Details of the Galileo system regarding operating frequencies have recently been defined. An agreement with the US, secured in June 2004, defined a baseline L₁ signal as BOC (Binary Offset Carrier) (1, 1) with a BOC (15, 2.5) as a public regulated service with time and co-ordinate standards.

13.3.2.4 *Multi-system global navigation satellite systems*

With two constellations of satellites (GPS and GLONASS) currently orbiting the Earth, the user can have access to up to double the number of satellites needed to obtain a navigation solution at any time. This provides a number of advantages to the user equipped to receive signals from both systems, viz.

- the capability to monitor the integrity of the navigation solution, which is vital for safety-critical applications such as civil aviation;

- improved accuracy of the navigation solution through the reduced likelihood of having to compute position at a time when the geometrical arrangement of the satellites in view is unfavourable, that is, improved geometric dilution of precision (GDOP);
- reduced susceptibility to interference since the two systems operate at different frequencies.

Care must be taken when combining GPS and GLONASS data, since the two systems are not entirely compatible in terms of the timescales used and the geodetic reference/Earth model adopted by each system. However, such problems are not expected to arise in the future when integrating the GPS and Galileo systems, the Galileo system having been designed to be compatible with GPS.

Appendix D compares and contrasts the characteristics of the two currently operational satellite-based navigation systems. Additionally, some further consideration is given to the issues concerning the integration of data from the two constellations to provide an integrated system with greater integrity.

13.3.3 Star trackers

The stars may be regarded as fixed points, which can be used as references for the purposes of celestial navigation. The geographical position of an observer on or close to the Earth can be determined at any time given knowledge of:

- the positions of two or more celestial objects in relation to the observer;
- the exact time of the observation.

The basis of celestial navigation is that if the altitude (the angle between the line of sight and the horizontal) of a celestial object is measured, then the observer's position must lie on a specific circle on the Earth. This circle is centred on the point on the Earth that lies directly below the object. If the time of the observation is known, then this point can be found from pre-computed astronomical tables. Given sightings of two objects, then two circles of position are defined and the observer must be located at their intersection, as indicated in Figure 13.6a.

To enable the technique of position estimation from star sightings, or celestial observation, to be used in aircraft, automatic star trackers have been developed. A star tracker is basically a telescopic device having a detector and a scanning mechanism. Sightings of stars may be achieved using a star tracker to provide measurements of the azimuth and elevation angles of a star with respect to a known reference frame within the vehicle. Typically, this would be a space-stabilised reference frame, defined by a stable platform on which the star tracker is mounted. For navigation purposes, knowledge of the direction of the local vertical is needed in order to relate the measurements to an Earth-based reference frame.

Alternatively, measurements may be made with respect to a body-fixed frame and used to update a strapdown inertial navigation system in the manner outlined below. Such observations can be compared with stored knowledge of the observed star's position to derive a position fix, or an estimate of vehicle heading.

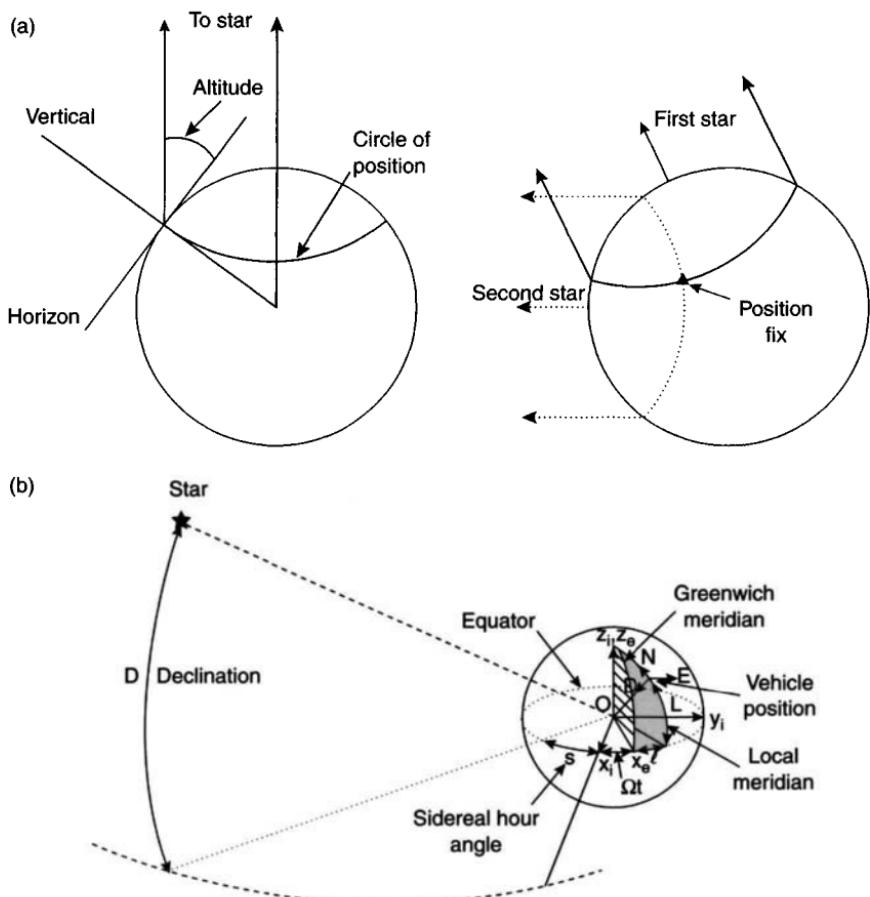


Figure 13.6 (a) Star tracker position fixing; (b) star tracker geometry

For navigation purposes, stars may be considered to be positioned on the inner surface of a geocentric sphere of infinite radius, often referred to as the celestial sphere. The projection of the Earth's lines of latitude and longitude on to this sphere establishes a grid in which the position of a celestial object may be defined. The position north or south of the equator is referred to as the declination of a star, whilst its longitudinal position is expressed in terms of a sidereal hour angle, as indicated in Figure 13.6b. Hence, the direction of a star with respect to an inertial frame, which has its origin at the centre of the Earth, may be expressed in terms of its declination and sidereal hour angle.

A star tracker fixed in the body of a vehicle would provide measurements of the star's bearing and elevation with respect to the body frame. These measurements may then be compared with predictions of these same quantities derived from knowledge of the declination and sidereal hour angle of a star. These quantities may also be expressed in body axes given knowledge of the vehicle's latitude and longitude and its orientation with respect to the local geographic frame. The resulting measurement

differences may then be used to update the on-board inertial navigation system in a manner similar to that discussed in Section 13.6.

Star trackers combined with an inertial navigation system are believed to be capable of measurement accuracies of a few arc seconds in a space environment. When a star tracker is used close to the surface of the Earth, corrections to the measurements have to be made for the refraction of the Earth's atmosphere. Typically, accuracies of better than 10 arc s may be achieved, which corresponds to a position error of around 300 m on the surface of the Earth. However, it is a passive system and its errors are potentially independent of elapsed time.

13.3.4 Surface radar trackers

A ground- or surface-based radar station may be used to provide line-of-sight observations of an aircraft or missile during flight. These observations usually take the form of measurements of a vehicle's range, elevation and bearing as indicated in Figure 13.7. The measurements are derived with respect to a local reference frame, usually the local vertical geographic frame at the location of the radar tracker. The measurement data may be transmitted to the vehicle for in-flight aiding of an on-board inertial navigation system.

The measurements of a vehicle's range (R), azimuth (ψ) and elevation (θ) may be expressed in terms of the Cartesian position coordinates of the aircraft (x, y, z) as follows:

$$R = \sqrt{x^2 + y^2 + z^2}$$

$$\psi = \tan^{-1} \left(\frac{y}{x} \right)$$

$$\theta = \tan^{-1} \left\{ \frac{z}{\sqrt{x^2 + y^2}} \right\}$$

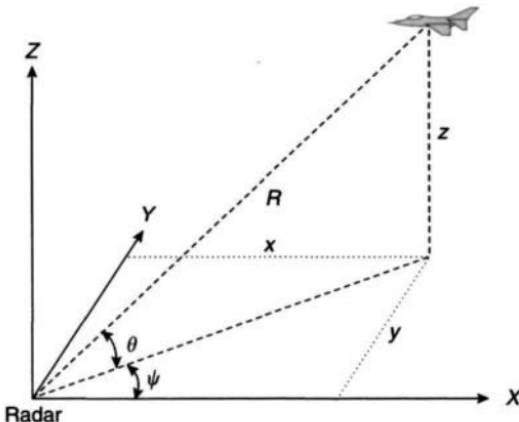


Figure 13.7 Ground radar measurements

These measurements may be used to update the on-board inertial navigation system by comparing them with predictions of the same quantities obtained from information provided by the on-board navigation system. A design example, based on such a system, is presented in Section 13.6.2.

13.4 On-board measurements

13.4.1 Doppler radar

Doppler radar, which provides a means of measuring a vehicle's velocity, is often used to provide navigation aiding for airborne systems and in some cases, in conjunction with an attitude and heading reference system, as the primary source of navigation data. A Doppler radar operates by transmitting a narrow beam of microwave energy to the ground and measures the frequency shift that occurs in the reflected signal as a result of the relative motion between the aircraft and the ground. In the situation where the aircraft velocity is V and the radar beam slants down towards the ground at an angle θ , the frequency shift is:

$$\frac{2V}{\lambda} \cos \theta$$

where λ is the wavelength of the transmission. For a typical system operating in the frequency range 13.25–13.4 GHz ($\lambda \approx 2.2$ cm), the frequency shift is approximately 47 Hz per knot. Given the knowledge of the wavelength of the transmission and the slant angle, an estimate of the velocity of the aircraft can be determined from the measured frequency shift. Because the aircraft is able to move in three dimensions, the minimum number of beams needed to establish aircraft velocity is three. The beams are often directed forwards and to the rear of the aircraft as illustrated in Figure 13.8.

Modern Doppler systems generate the beams using a planar array, the aerial being attached rigidly to the body of the aircraft. The reflected signals from each of the beams are processed separately, enabling estimates of aircraft velocity components to be calculated in a computer processor. Such estimates are derived in a co-ordinate

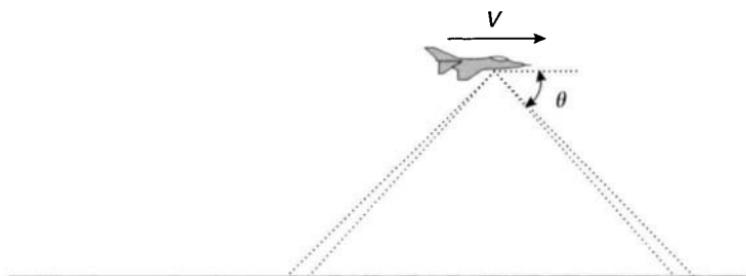


Figure 13.8 Doppler radar beam geometry

frame, which is fixed with respect to the aerial. In order to carry out the navigation function, it is necessary to resolve these velocity estimates into the chosen navigation reference frame. An on-board attitude and heading reference system is required for this purpose. Alternatively, the Doppler sensor may be integrated with an on-board inertial navigation system. In this case, the Doppler velocity measurements would be compared with estimates of those same quantities generated by the inertial system.

Typically, a Doppler sensor operating over land is able to provide measurements of velocity to an accuracy of about 0.25 per cent, or better. Performance is degraded during flights over water owing to poor reflectivity, scattering of the reflected signal giving rise to a bias in the measurement, wave motion, tidal motion and water currents. However, this navigation aid offers good long-term stability and a chance to bound the position and velocity estimates provided by an inertial navigation system.

13.4.2 Magnetic measurements

The Earth has a magnetic field similar to that of a bar magnet with poles located close to its axis of rotation. This means that the direction of the horizontal component of the Earth's magnetic field lies close to true north, and the magnetic north, as determined by a magnetic field sensor, or compass, can be used as a working reference. Unfortunately, the angle between true north and magnetic north is not constant. It varies with the position of the observation on the Earth and slowly with time, although it is possible to compensate for both of these effects. The direction of the Earth's magnetic field at any point on the Earth is defined in terms of its orientation with respect to true north, known as the angle of 'magnetic declination' (sometimes referred to as 'magnetic variation'), and its angle with respect to the horizontal, the

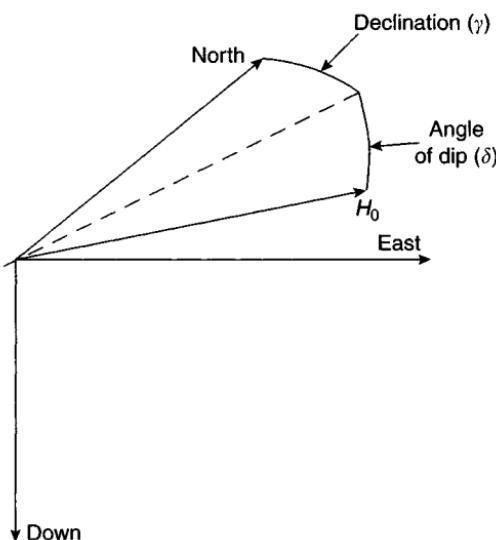


Figure 13.9 Components of Earth's magnetic field

angle of 'dip', as indicated in Figure 13.9. The vehicle in which the magnetic sensor is mounted will almost always have a magnetic field that cannot be distinguished from that of the Earth and consequently it is necessary to compensate for this effect as well as the others mentioned earlier.

The magnetic compass is one of the oldest navigation aids known to man, having been used for hundreds of years to provide a directional reference for steering and dead-reckoning navigation. A practical device, which may be used to sense the Earth's magnetic field and to provide a measure of the attitude of a moving vehicle, is the fluxgate magnetometer, as described in Section 4.7.2. In the absence of local magnetic disturbances, this device senses the components of the Earth's magnetic field (H_0) acting along its sensitive axis. A three-axis device may be mounted in a vehicle to sense the components of the Earth's field along its principal body axes, (H_x H_y H_z). Expressing body attitude, with respect to the local geographic frame, as a direction cosine matrix, C_n^b , the relationship between the magnetic measurements and body attitude may be written as follows:

$$\begin{bmatrix} H_x \\ H_y \\ H_z \end{bmatrix} = C_n^b \begin{bmatrix} H_0 \cos \delta \cos \gamma \\ H_0 \cos \delta \sin \gamma \\ H_0 \sin \delta \end{bmatrix}$$

where δ is the angle of dip and γ is the angle of declination. Given knowledge of the angle of dip, and the variation between true and magnetic north, estimates of vehicle attitude can be deduced from the measurements provided by the magnetometer. Rotations about the local magnetic vector cannot be detected. For this reason, such a device must operate in conjunction with a vertical reference system to determine body attitude in full.

Other possibilities for navigation aiding have been suggested that involve using measurements of the Earth's magnetic field in different ways [10]. For example, it would be possible in principle to obtain position fixes either by comparing field measurements in the local geographic frame with stored maps of magnetic variation and dip angle, or by attempting to match magnetic anomalies. The former method would require precise knowledge of the directions of true north and the local vertical to provide accurate fixes. The latter scheme, which is analogous to the terrain matching technique discussed in Section 13.4.4, would clearly be reliant on the availability of sufficiently detailed magnetic anomaly maps and on the stability of these anomalies. In regions with a large number of significant and stable anomalies, this system has the potential for good positional fix accuracy.

13.4.3 Altimeters

Barometric altimeters are invariably used for height measurement in aircraft to satisfy height accuracy standards in controlled air space. As supplementary navigation aids, they are widely used for restricting the growth of errors in the vertical channel of an inertial navigation system. In a Schuler-tuned inertial navigation system, whilst the propagation of errors in the horizontal channels is bounded, the velocity and position errors in the vertical channel are unbounded. Consequently, these errors can become very large within a relatively short period of time, unless there is an independent means

of checking the growth of such errors. For example, the effect of a net acceleration bias (B) acting in the vertical direction gives rise to a positional error $Bt^2/2$. Hence, a bias of only 10 micro-g would result in a height error in excess of 2.5 km over a 2-h period.

A barometric altimeter, relying on atmospheric pressure readings, provides an indirect measure of height above a nominal sea level, typically to an accuracy of much less than 0.01 per cent. Most airborne inertial systems requiring a three-dimensional navigation capability operate with barometric aiding in order to bound the growth of vertical-channel errors.

A radar altimeter provides a direct measure of height above ground, which is equally important for many applications. Such measurements may be used in conjunction with a stored map of the terrain over which an aircraft is flying to provide position updates for an inertial navigation system. The subject of terrain referenced navigation is addressed separately in the following section.

13.4.4 Terrain referenced navigation

The development of terrain referenced navigation techniques, such as terrain contour matching began in the 1970s, and a number of systems have become commercially available.

The most established form of terrain referenced navigation (Figure 13.10) uses a radio altimeter, an on-board baro-inertial navigation system and a stored contour map of the area over which the aircraft or missile is flying. The radio altimeter measurements of height above ground, in combination with the estimates of height above sea level, provided by the inertial navigation system, allow a reconstruction of the ground profile beneath the flight path to take place within a computer on-board the vehicle. The resulting terrain profile is then compared with the stored map data to achieve a fit, from which the position of the vehicle may be identified.

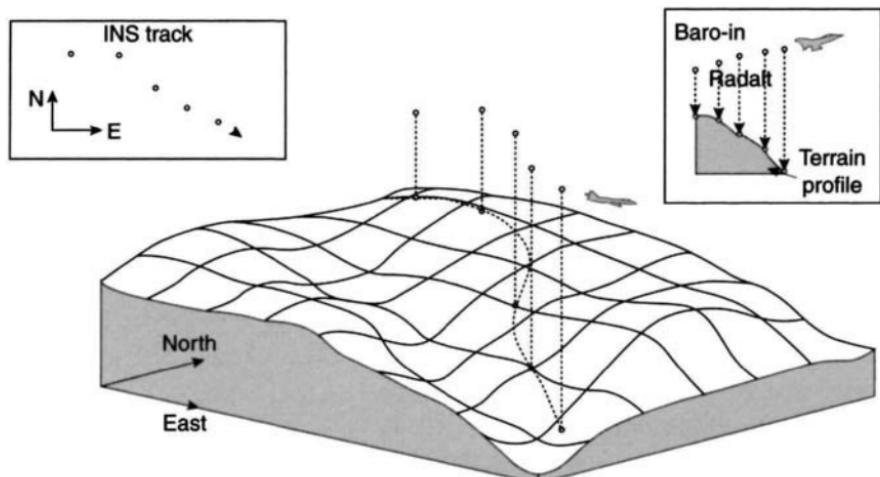


Figure 13.10 Terrain referenced navigation

A position fix is possible because of the random nature of the Earth's surface, which tends to give each section of terrain a shape that is unique to its location. The accuracy of the position fix that can be obtained is generally a function of the roughness of the terrain, a more precise fix being achievable when the contour variation is between 20% and 40%.

Various schemes for extracting vehicle position during flight have been devised. The scheme outlined above relies on taking a succession of height measurements, which may be used retrospectively, to determine the location of the vehicle. An alternative method involves the comparison of estimates of terrain height, derived by differencing the inertial system indicated height with the radar altimeter measurement, with estimates of terrain height extracted from the stored contour map. The map heights are derived at the co-ordinate location of the vehicle indicated by its on-board inertial navigation system. Updates of vehicle position may be derived given knowledge of the terrain slope beneath the vehicle. In this way, each altimeter measurement is processed separately, and then used to update the on-board navigation system on an almost continuous basis.

Operation of this scheme is clearly dependent on the availability of a good quality terrain database for the area over which the vehicle is to be flown. Typically, radial position accuracies of a few tens of metres can be achieved, the precise accuracy varying with the roughness of the terrain beneath the aircraft. Clearly, navigation accuracy degrades when over-flying large flat and featureless areas of terrain, and particularly over water, where the accuracy of navigation becomes solely dependent on the performance of the unaided inertial system. Further degradations in performance can result from tree foliage and snow cover, which can affect the radio altimeter accuracy.

The accuracy of terrain contour matching navigation may be enhanced by using a more precise height sensor. Research activities in recent years have focused on the use of laser line scanners, which provide measurements of range from the air vehicle to the ground to an accuracy of a few centimetres. Such devices are capable of scanning the terrain beneath the vehicle; the effect of the scanning, in combination with the forward motion of the aircraft, allows an elevation contour to be constructed over a two-dimensional surface, rather than the one-dimensional elevation strip generated using a radar altimeter. The correlation of a scanned area of terrain with the stored terrain database, coupled with the improved accuracy of the basic range measurement potentially allows greater positional accuracy and reduces the probability of a false fix.

Alternative techniques that allow terrain-based navigation over areas of flat terrain have also been developed. The increased flexibility this affords in terms of route planning adds to the attraction of such options. Methods adopted include scene matching area correlation and, more recently, continuous visual navigation, which are described separately in the following sections.

13.4.5 Scene matching

Scene matching area correlation (SMAC) techniques may be used to provide highly accurate position fixes based on images of the ground beneath an aircraft

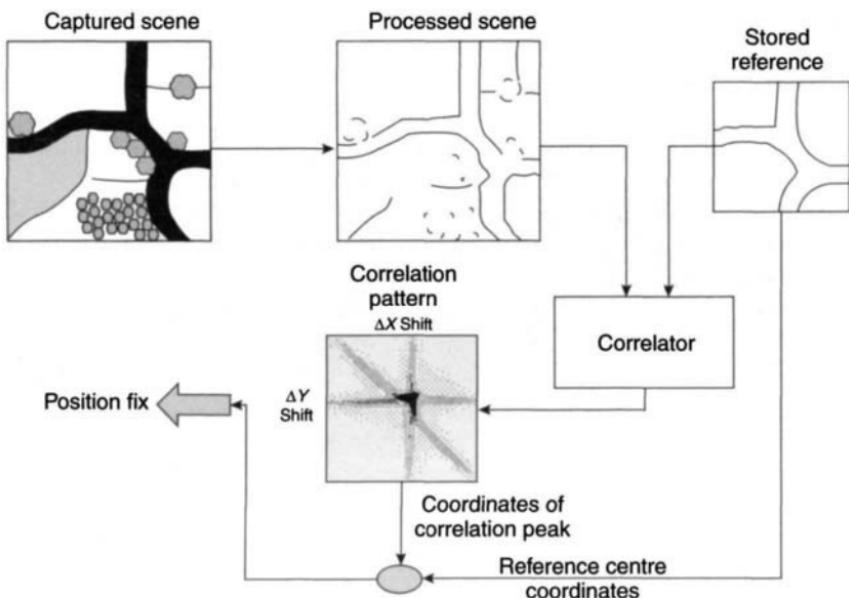


Figure 13.11 The principle of scene matching.

or missile. The operation of this system is equivalent to the technique used by a human navigator, navigating from recognition of remembered landmarks or ground features.

The fundamental principle is illustrated in Figure 13.11. An imaging system, such as an infrared line-scan device, builds up a picture of the terrain beneath the aircraft as it moves forward. When a fix is required, a portion of the infrared line-scan image is stored to form a 'scene' of the terrain beneath the aircraft. By this process, the image is converted into an array of 'pixels', each pixel having a numerical value indicating its brightness of that part of the image. The 'captured' scene is processed in order to remove noise and to enhance those features that are likely to provide navigation information, road junctions and railway lines for example. In the next stage of analysis, a correlation algorithm is used to search for recognisable patterns, which appear in a pre-stored database of ground features. Having found a match between a feature in the scene and one in the database, geometrical calculations based on the aircraft's attitude and height above ground enable its position to be calculated at the time the scene was captured. The various stages in the scene matching process are illustrated in Figure 13.11.

13.4.6 Continuous visual navigation

It is vital to provide precise position, velocity and attitude data for medium and long range missile systems and on military aircraft to ensure mission success. Navigation

data with the required accuracy can be provided using systems incorporating INS integrated with GPS using the tightly coupled integration strategies discussed in Section 13.7. However, the possibility of jamming of satellite receivers, and the consequent risk to mission success, remains a concern in the minds of military users. This concern has led to the use of alternative, or additional, navigation aids for some military applications, in addition to those indicated earlier.

Terrain referenced navigation, described in Section 13.4.4, provides an obvious option in such applications, and has been developed for precisely this type of mission. However, such systems do not yield the precise navigation data sought by modern military planners, the resulting positional data being less accurate than that derived using GPS. The reduced accuracy of terrain contour matching systems when flying over very flat terrain also limits the usefulness of this approach. This factor introduces constraints on mission planning, owing to the reduction in guidance accuracy that could be achieved if routes over flat terrain were to be chosen. In view of the concerns outlined here, further methods of complementing the navigation strategies described above have been sought.

Recent research effort by the Navigation Research Group at QinetiQ, Farnborough, UK (originally the Navigation Department, RAE, Farnborough) has focused on an alternative method known as continuous visual navigation (CVN), which appears to offer a viable and robust approach, whilst addressing many of the concerns raised about other methods. Conceived originally as a reversionary navigation mode for integrated INS–GPS systems, CVN is a linear feature-matching navigation technique [11–13]. It was developed originally as an evolution of scene matching (see Section 13.4.5); CVN operates by matching observed linear features with an array of features stored in a database, and is capable of providing civil GPS positional accuracy over flat terrain. The main components of such a system are:

- an INS to provide continuous position, velocity and attitude data;
- a radar altimeter to provide continuous measurements of vehicle height above ground;
- a device, such as a video in infrared camera, to provide digital images of the terrain beneath the vehicle;
- an on-board database to store linear features for the area over which the vehicle is required to fly during the execution of its mission;
- a processor in which the integrated navigation algorithms are implemented.

The CVN system has been designed to match lines in the stored terrain feature map, rather than points, and to produce a measurement update to the navigation filter following the match of a single linear feature. A diagrammatic representation of the CVN primary algorithm is shown in Figure 13.12. Efforts have been made to maximise the robustness of the system through the implementation of at least two distinct position fixing algorithm methods, and by maintaining multiple INS error hypotheses in the form of several Kalman filters. These issues are discussed further in Section 13.8.

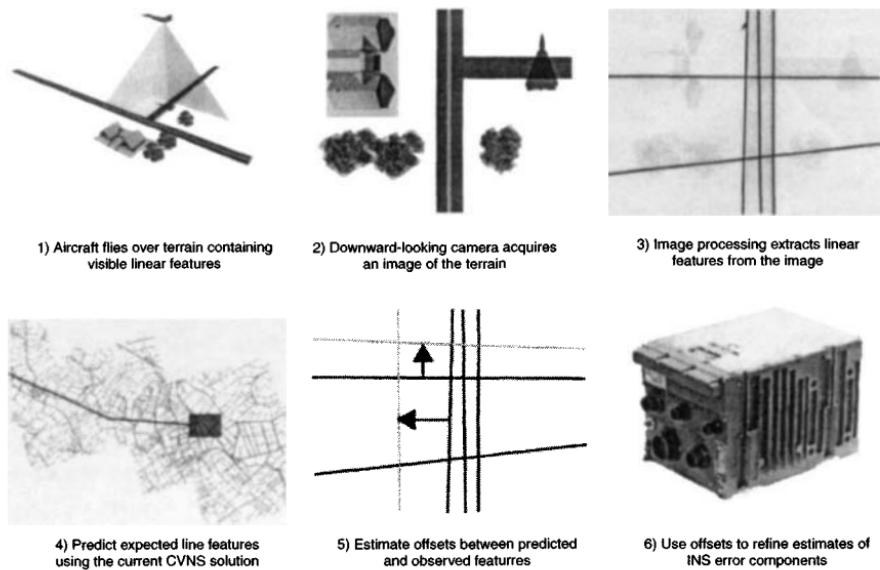


Figure 13.12 Diagrammatic representation of continuous visual navigation primary algorithm

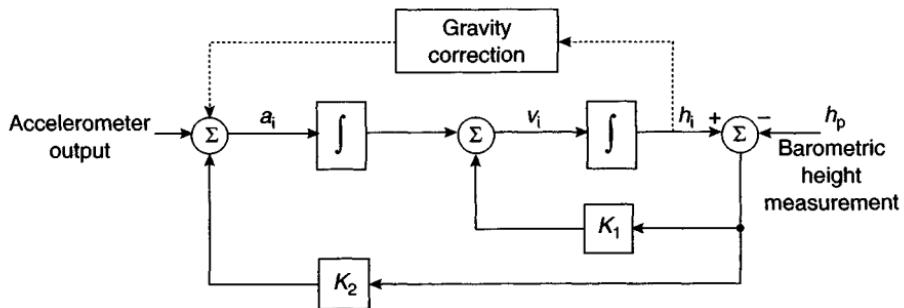


Figure 13.13 Baro-inertial height measurement

13.5 System integration

The remainder of this chapter is concerned with techniques that may be used to combine inertial measurement data with information provided by one or more of the navigation aids discussed in the preceding section. In general, the available measurements are corrupted with noise. Consequently, some form of on-line filtering technique is required to achieve good navigation performance.

Early systems used complementary filtering techniques of the form described below for a baro-inertial system and illustrated in Figure 13.13.

The difference between the inertial system estimate of height and the pure barometric height measurement is fed back via the gains K_1 and K_2 , shown in

the figure, to correct the velocity and height estimates. Values for the gains are selected to allow the baro-inertial system to follow the long-term variations in the barometric measurements, whilst filtering out any higher frequency fluctuations. Typically, $K_1 = 2/T$ and $K_2 = 1/T^2$, where the value of T may be 30 s [14]. In the integrated system, a bias (B_z) on the inertial estimate of vertical acceleration no longer propagates as a height error with time squared, but settles to a steady state value of $T^2 B_z$. Hence, a bias of 100 micro-g gives rise to a height error of approximately 1 m. The major limitation of this technique is that any longer-term errors in the barometric altimeter, resulting from weather conditions and the position of the device, persist in the integrated system.

As with all filtering techniques, the objective is to make use of the available knowledge about the long-term behaviour of a signal, contaminated with noise, in order to derive a better estimate of the signal than could be obtained using a single measurement. A practical form of filter, which is applicable for on-line estimation, relies on generating a mathematical model of the process that is producing the signal, and adjusting the parameters of the model to minimise the mean square deviation between the signal and the output of the model. A best estimate of the signal is derived based on knowledge of the expected errors in the model and the measured signal using a Kalman filter. Kalman filtering has become a well-established technique for the mixing of navigation data in integrated systems. It is particularly suitable for on-line estimation, being a recursive technique which lends itself to implementation in a computer.

The principles of Kalman filtering are described in Appendix A and its application is illustrated in the following section.

13.6 Application of Kalman filtering to aided inertial navigation systems

13.6.1 Introduction

As described in Appendix A, Kalman filtering involves the combination of two estimates of a variable to form a weighted mean, the weighting factors being chosen to yield the most probable estimate. One estimate is provided by updating a previous best estimate, in accordance with the known equations of motion, whilst the other is obtained from a measurement. In an integrated navigation system, the first estimate is provided directly by the inertial navigation system, that is, in filtering terms, the inertial system constitutes the model of the physical process that produces the measurement. The second estimate, the measurement, is provided by the navigation aid. The same technique can be applied irrespective of the source of measurement information.

A generalised block diagram representation is shown in Figure 13.14, and a design example is described in the following section.

However, the measurements provided by navigation aids are often non-linear combinations of the inertial navigation system estimates. Additionally, the inertial system equations themselves are non-linear, which means that a modified approach

the figure, to correct the velocity and height estimates. Values for the gains are selected to allow the baro-inertial system to follow the long-term variations in the barometric measurements, whilst filtering out any higher frequency fluctuations. Typically, $K_1 = 2/T$ and $K_2 = 1/T^2$, where the value of T may be 30 s [14]. In the integrated system, a bias (B_z) on the inertial estimate of vertical acceleration no longer propagates as a height error with time squared, but settles to a steady state value of $T^2 B_z$. Hence, a bias of 100 micro-g gives rise to a height error of approximately 1 m. The major limitation of this technique is that any longer-term errors in the barometric altimeter, resulting from weather conditions and the position of the device, persist in the integrated system.

As with all filtering techniques, the objective is to make use of the available knowledge about the long-term behaviour of a signal, contaminated with noise, in order to derive a better estimate of the signal than could be obtained using a single measurement. A practical form of filter, which is applicable for on-line estimation, relies on generating a mathematical model of the process that is producing the signal, and adjusting the parameters of the model to minimise the mean square deviation between the signal and the output of the model. A best estimate of the signal is derived based on knowledge of the expected errors in the model and the measured signal using a Kalman filter. Kalman filtering has become a well-established technique for the mixing of navigation data in integrated systems. It is particularly suitable for on-line estimation, being a recursive technique which lends itself to implementation in a computer.

The principles of Kalman filtering are described in Appendix A and its application is illustrated in the following section.

13.6 Application of Kalman filtering to aided inertial navigation systems

13.6.1 Introduction

As described in Appendix A, Kalman filtering involves the combination of two estimates of a variable to form a weighted mean, the weighting factors being chosen to yield the most probable estimate. One estimate is provided by updating a previous best estimate, in accordance with the known equations of motion, whilst the other is obtained from a measurement. In an integrated navigation system, the first estimate is provided directly by the inertial navigation system, that is, in filtering terms, the inertial system constitutes the model of the physical process that produces the measurement. The second estimate, the measurement, is provided by the navigation aid. The same technique can be applied irrespective of the source of measurement information.

A generalised block diagram representation is shown in Figure 13.14, and a design example is described in the following section.

However, the measurements provided by navigation aids are often non-linear combinations of the inertial navigation system estimates. Additionally, the inertial system equations themselves are non-linear, which means that a modified approach

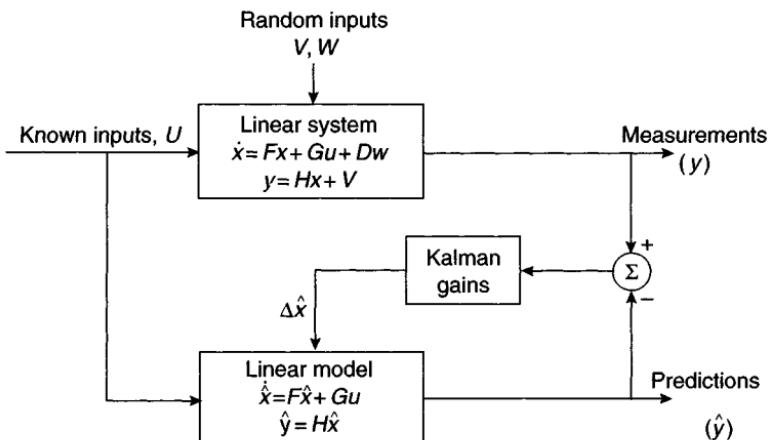


Figure 13.14 Kalman filter for linear systems

is needed. Consequently, it is customary to use an extended Kalman filter formulation for an aided inertial navigation system (see Appendix A).

13.6.2 Design example of aiding

In this section, a scheme for aiding a hypothetical missile on-board inertial navigation system is described, which relies upon tracking the missile during flight using a sensor on the launch platform. Suitable measurements may be provided by a multi-function radar or an infrared tracker in combination with a laser range-finder. In either case, it is assumed that measurements of a missile's range, elevation and bearing with respect to the chosen navigation frame will be provided. These measurements may be passed via an uplink transmitter to the missile and used to aid the on-board navigation system.

The transmitted measurements are combined, using a Kalman filter, with the measurements provided by the missile's inertial navigation system. This not only allows improved estimates of the missile's position to be derived from noisy measurement data, but also provides a means of estimating errors in the states of the navigation system which are not directly measurable; the velocity and attitude estimates for example. The form of the filter is as shown in Figure 13.15.

In order to provide estimates of the three attitude errors, the three velocity errors and the three position errors in the on-board inertial navigation system, a nine state Kalman filter is required. The associated system and measurement equations are described in the sections which follow.

13.6.2.1 The system equations

To formulate an extended Kalman filter to update the on-board navigation system, it is necessary to develop a linear dynamic model of the errors that are to be estimated. For the purposes of this design example, a simplified version of the error model given

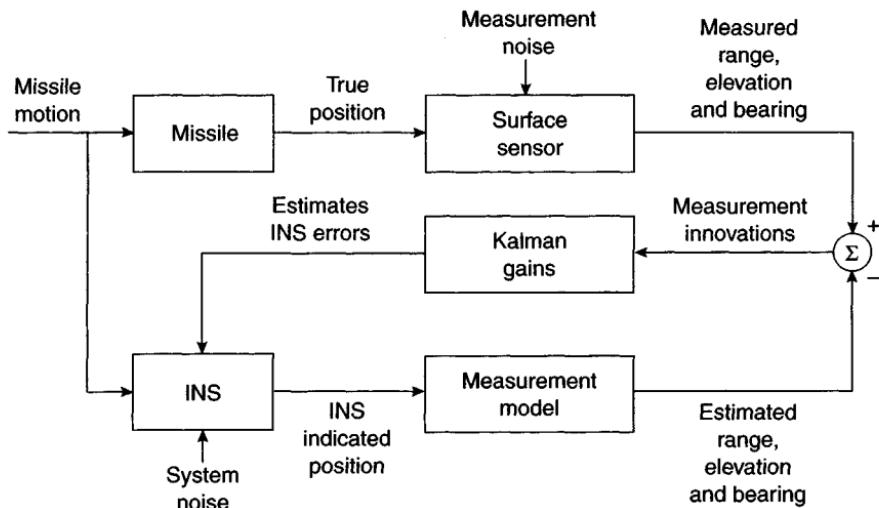


Figure 13.15 INS aiding using ground tracker measurements

in Chapter 12 may be used. The error model may be expressed in matrix form as:

$$\delta\dot{\mathbf{x}} = \mathbf{F}\delta\mathbf{x} + \mathbf{Gw} \quad (13.1)$$

The vector $\delta\mathbf{x}$ represents the error state of the system. For the purposes of this illustration, $\delta\mathbf{x}$ consists of the three attitude errors ($\delta\alpha \delta\beta \delta\gamma$), three velocity errors, ($\delta v_x \delta v_y \delta v_z$) and three position errors ($\delta x \delta y \delta z$).

In order to allow a discrete Kalman filter to be constructed, it is necessary to express the system error eqn. (13.1) in discrete form. If $\delta\mathbf{x}_k$ represents the inertial navigation system error states at time t_k , and $\delta\mathbf{x}_{k+1}$ the error states at time t_{k+1} , we may write:

$$\delta\mathbf{x}_{k+1} = \Phi_k \delta\mathbf{x}_k + \mathbf{w}_k \quad (13.2)$$

where Φ_k is the system transition matrix at time t_k , which may be expressed in terms of the system matrix \mathbf{F} as follows:

$$\Phi_k = \exp[\mathbf{F}(t_{k+1} - t_k)] \quad (13.3)$$

13.6.2.2 The measurement equations

Measurements of the missile's position with respect to the radar are assumed to be available at discrete intervals of time throughout flight. The radar provides measurements in polar coordinates, that is, measurements of range (R), elevation (θ) and bearing (ψ). The polar quantities may be expressed in terms of the

Cartesian coordinates (x, y, z) as follows:

$$\begin{aligned} R^2 &= x^2 + y^2 + z^2 \\ \theta &= \arctan\left\{\frac{z}{\sqrt{x^2 + y^2}}\right\} \\ \psi &= \arctan\left(\frac{y}{x}\right) \end{aligned} \quad (13.4)$$

Writing $z = [R \ \theta \ \psi]^T$, the radar measurements, denoted by \tilde{z} , may be expressed as:

$$\tilde{z} = z + n \quad (13.5)$$

where n represents the error in the measurements. n is assumed to be a zero mean, Gaussian white-noise process.

Estimates of the radar measurements, \hat{z} , may be obtained from the inertial navigation system estimates of position (x, y, z) as follows:

$$\hat{z} = \begin{pmatrix} \hat{R} \\ \hat{\theta} \\ \hat{\psi} \end{pmatrix} = \begin{pmatrix} \sqrt{\hat{x}^2 + \hat{y}^2 + \hat{z}^2} \\ \arctan\left\{\frac{\hat{z}}{\sqrt{\hat{x}^2 + \hat{y}^2}}\right\} \\ \arctan\left\{\frac{\hat{y}}{\hat{x}}\right\} \end{pmatrix} \quad (13.6)$$

The difference between the radar measurements (\tilde{z}) and the estimates (\hat{z}) of those quantities is referred to as the filter measurement innovation (δz) and is generated as follows:

$$\delta z = \tilde{z} - \hat{z} = H \delta x \quad (13.7)$$

where

$$H = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & \frac{x}{R} & \frac{y}{R} & \frac{z}{R} \\ 0 & 0 & 0 & 0 & 0 & 0 & \frac{-xz}{R^2\sqrt{x^2+y^2}} & \frac{-yz}{R^2\sqrt{x^2+y^2}} & \frac{\sqrt{x^2+y^2}}{R^2} \\ 0 & 0 & 0 & 0 & 0 & 0 & \frac{-y}{x^2+y^2} & \frac{x}{x^2+y^2} & 0 \end{bmatrix} \quad (13.8)$$

13.6.2.3 The Kalman filter

Equations (13.2) and (13.7) are the system and measurement equations needed to construct a Kalman filter. The equations for the Kalman filter, given in Appendix A, take the following form for the radar-aided inertial system considered here.

Filter prediction step

Following each measurement update, the inertial navigation system is corrected using the current best estimates of the errors in position, velocity and attitude. Therefore, after an update, the best estimate of each of the inertial system errors becomes identically zero and the state prediction equation reduces to:

$$\delta \mathbf{x}_{k+1/k} = 0 \quad (13.9)$$

The covariance matrix is predicted forward in time using the expression:

$$\mathbf{P}_{k+1/k} = \Phi_k \mathbf{P}_{k/k} \Phi_k^T + \Delta \mathbf{Q}' \Delta^T \quad (13.10)$$

where Φ_k is the transition matrix given by eqn. (13.3). $\mathbf{P}_{k+1/k}$ denotes the expected value of the covariance matrix at time t_{k+1} predicted at time t_k . It is set up initially as a diagonal matrix, the individual elements being chosen according to the expected variances of the errors in the initial attitude, velocity and position passed to the missile navigation inertial system prior to launch. \mathbf{Q}' , the system noise matrix, is set up according to the expected level of noise on the inertial measurements of linear acceleration and angular rate.

Filter update

The estimates of the errors in the inertial navigation system states are derived using:

$$\delta \mathbf{x}_{k+1/k+1} = \mathbf{K}_{k+1} \delta \mathbf{z}_{k+1} \quad (13.11)$$

and the covariance matrix is updated according to:

$$\mathbf{P}_{k+1/k+1} = [\mathbf{I} - \mathbf{K}_{k+1} \mathbf{H}_{k+1}] \mathbf{P}_{k+1/k} \quad (13.12)$$

where

$$\mathbf{K}_{k+1} = \mathbf{P}_{k+1/k+1} \mathbf{H}_{k+1}^T [\mathbf{H}_{k+1} \mathbf{P}_{k+1/k} \mathbf{H}_{k+1}^T + \mathbf{R}']^{-1} \quad (13.13)$$

\mathbf{H} is defined by eqn. (13.8) and \mathbf{R}' , the measurement noise is a 3×3 diagonal matrix, the elements of which are selected in accordance with the anticipated level of radar measurement noise.

Inertial navigation system correction

The inertial navigation states, $\hat{\mathbf{x}}$, are corrected immediately after each measurement update using the current best estimates of the errors. The correction equations are given below.

Velocity and position correction. Velocity and position may be corrected by simply subtracting the estimate error from the inertial system estimates of these quantities using:

$$\mathbf{x}_c = \hat{\mathbf{x}} - \delta \mathbf{x} \quad (13.14)$$

where \mathbf{x}_c is the corrected state.

Attitude correction. As described earlier (Chapter 11), the computed direction cosine matrix may be expressed in terms of the true matrix using:

$$\hat{\mathbf{C}} = [\mathbf{I} - \Psi] \mathbf{C}$$

The corrected direction cosine matrix, \mathbf{C}_c , may therefore be expressed as follows:

$$\mathbf{C}_c = [\mathbf{I} + \Psi] \hat{\mathbf{C}} \quad (13.15)$$

where $\Psi = \Psi \times$ and $\Psi = [\delta\alpha \quad \delta\beta \quad \delta\gamma]^T$.

By writing \mathbf{C}_c and \mathbf{C} in component form as functions of the corrected and estimated quaternion parameters, denoted $[a_c \quad b_c \quad c_c \quad d_c]$ and $[\hat{a} \quad \hat{b} \quad \hat{c} \quad \hat{d}]$, respectively, and equating terms, it can be shown that the estimated quaternion parameters may be corrected directly using:

$$\begin{aligned} a_c &= \hat{a} + 0.5(\delta\alpha\hat{b} + \delta\beta\hat{c} + \delta\gamma\hat{d}) \\ b_c &= \hat{b} + 0.5(-\delta\alpha\hat{a} + \delta\beta\hat{d} - \delta\gamma\hat{c}) \\ c_c &= \hat{c} + 0.5(-\delta\alpha\hat{d} - \delta\beta\hat{a} + \delta\gamma\hat{b}) \\ d_c &= \hat{d} + 0.5(\delta\alpha\hat{c} - \delta\beta\hat{b} - \delta\gamma\hat{a}) \end{aligned} \quad (13.16)$$

13.6.2.4 Results

The results presented in Figures 13.16 and 13.17 illustrate the effectiveness of the measurements provided by a surface sensor in improving the performance of a missile's on-board inertial navigation system when using the nine state Kalman filter, described in the previous section. Results are presented in graphical form showing the standard deviations of attitude and position errors, with and without aiding, over a typical short-range missile flight of 10 s duration. Over this period of time, the missile follows a boost/coast trajectory, accelerating at 20g for the first 4 s of flight and slowing under the influence of aerodynamic drag thereafter.

Figure 13.16 shows navigation performance when a high-grade inertial navigation system is used. The fixed biases for the gyroscopes and accelerometers used here were 0.01°/h (1σ) and 100 micro-g (1σ), respectively. For the purposes of this analysis, the standard deviations of the initial condition errors in the missile system were chosen to be as follows:

- initial attitude errors: 10 mrad
- initial velocity errors: 1 m/s
- initial position errors: 1 m

Sensor measurement accuracies were set to values of 3 mrad (1σ) in elevation and bearing and 10 m (1σ) in range, and a data update rate of 1 Hz was assumed. The dramatic improvement in navigation performance with aiding is illustrated clearly in Figure 13.16. The position errors are corrected very rapidly and remain within 20 m for the remainder of the flight. The attitude errors settle to less than 0.2° after three measurements have been received. Because the instrument errors are small in this case, the system error model on which the Kalman filter is based provides

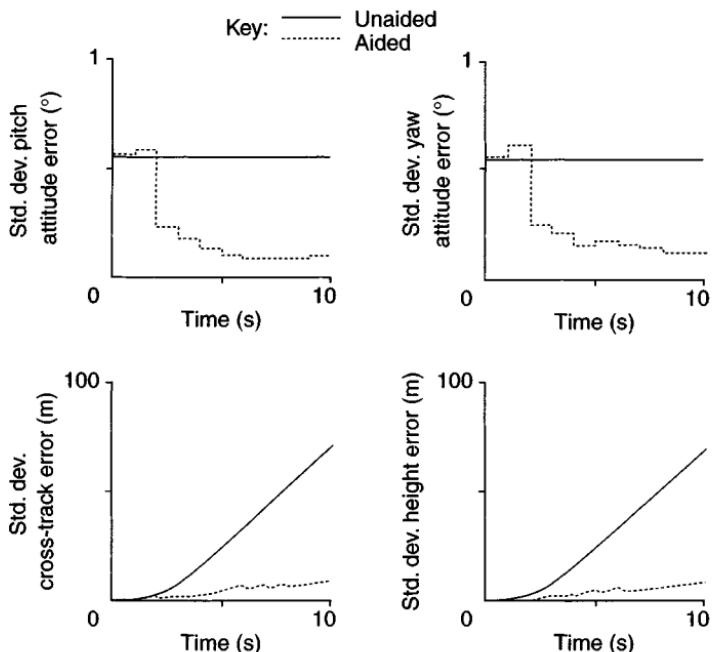


Figure 13.16 Simulation results from aiding – high-grade INS

a representative model of the actual system. This allows filter convergence to take place and accurate estimates of the inertial system errors to be derived.

Where the instrument errors become large, failure to model them correctly in the Kalman filter means the error model does not provide an accurate representation of what is happening in the actual system. Under such conditions, there is said to be a mis-match between the filter error model and the actual system. The sensor errors now introduce additional contributions to the measurement differences, which will be interpreted incorrectly as alignment errors. As a result, the Kalman filter estimates of attitude, velocity and position will be in error.

This effect is illustrated in Figure 13.17. The simulation described above was repeated under identical conditions with the exception that the high-grade inertial navigation system has been replaced by a system of more modest performance. The sensors have been replaced by gyroscopes and accelerometers having biases of $30^\circ/\text{h}$ and 10 milli- g , respectively, typical of the grade of sub-inertial quality instruments often specified for use in tactical missile systems.

Whilst a substantial improvement in navigation performance is still achieved, when compared with the unaided system, the resulting accuracy is reduced and the rate of convergence increased, in comparison with the previous set of results shown in Figure 13.16. The build up of attitude errors that occurs between the measurement updates, is indicative of the mis-match which now exists in the Kalman filter. A major source of error in this particular case is from gyroscopic mass unbalance, which

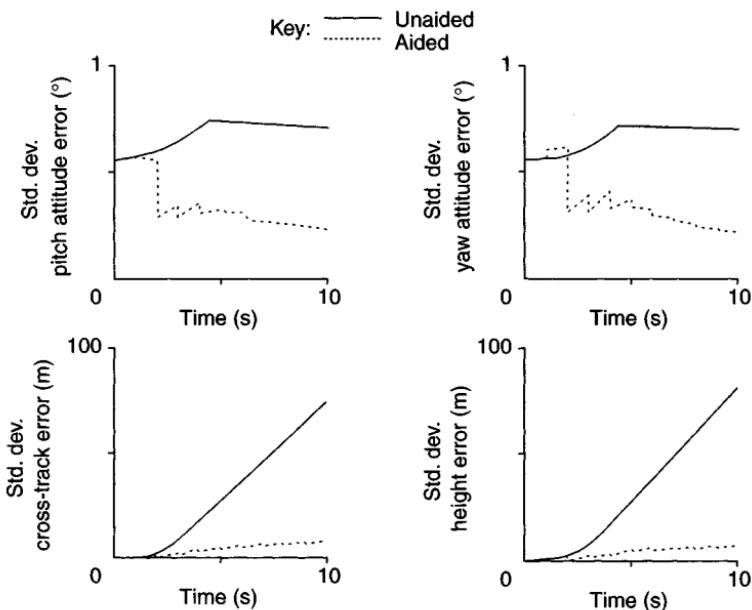


Figure 13.17 Simulation results from aiding – low-grade INS

introduces a rate bias that varies with the applied acceleration. This accounts largely for the shape of the attitude error curves shown in Figure 13.17.

In theory, there may be scope to improve the performance of the aided system by including additional filter states to model explicitly the dominant sensor errors [15]. By adopting this approach, the Kalman filter can be used to achieve some degree of in-flight sensor calibration, so leading to even greater enhancements of overall navigation performance. However, such a system will usually require time for the effects of modelling the sensor errors to become apparent.

In conclusion, the Kalman filter described earlier allows the effect of initial alignment errors to be reduced dramatically, and so provides sufficiently accurate inertial data for many short range missile applications, without recourse to much higher quality and more expensive inertial sensors. It is noted that for some applications, the in-flight aiding scheme described above may well allow some relaxation in the accuracy of the pre-flight alignment required. Pre-flight alignment methods are discussed in Chapter 10.

13.7 INS–GPS integration

As discussed earlier in this chapter, integrated navigation systems attempt to take advantage of the complementary attributes of two or more navigation systems to yield a system that provides greater precision than either of the component systems operating in isolation. Nowhere is this more true than for systems that combine inertial

navigation system (INS) measurements with satellite navigation data, the latter being provided by GPS, GLONASS or the Galileo system. A substantial level of effort has been focused on this topic in recent years, and work is continuing in an effort to produce integration schemes that are resistant to the effects of jamming of the satellite signals.

An INS exhibits relatively low noise, but tends to drift over time. For example, the position errors arising in a typical aircraft system will grow at between 1 and 10 nautical miles per hour of operation. In contrast, satellite navigation system estimates of position are relatively noisy, but exhibit no long-term drift.

Inertial and satellite navigation system measurements are complementary for two reasons:

- their error characteristics are radically different;
- they measure different quantities.

Satellite systems provide measurements of position and velocity, or more specifically pseudo-range and pseudo-range rate, whilst inertial systems measure specific force acceleration, which must be compensated for gravity and resolved into a known co-ordinate reference frame before being integrated twice to yield estimates of position.

Satellite position measurement accuracy is limited as a result of low signal strength, the length of the pseudo-random code and errors in the code tracking loop. Further errors arise as a result of multi-path, variations in the satellite geometry, changes in propagation conditions and user clock instability. Satellite velocity measurements are also noisy, again owing to variations in signal strength, the effects of changing multi-path and user clock instability.

Moreover, there is the possibility of jamming a GPS receiver with modest power jammers. Hence, any system that is dependent on a GPS-based navigation approach is vulnerable, and consequently, its availability may be compromised. The latter fact being a major concern for many users, particularly the armed forces.

The main features of inertial and satellite systems in terms of their respective advantages and disadvantages are summarised in Table 13.1.

Operating the two systems together yields benefits over operating either system alone. By making use of this basic synergy between inertial and satellite systems, it is possible to produce an integrated system that yields low noise and low drift estimates of vehicle position. There are also other relative attributes, which add to the benefits of integrating the two system approaches as discussed next.

The wide availability of GPS satellite navigation updates, coupled with the low cost of GPS receivers, has provided much of the impetus for the continuing development of techniques for the integration of INS and GPS. Given the availability of uninterrupted GPS access, there is considerable scope to combine low-accuracy inertial systems with GPS in order to produce low-cost precision navigation systems capable of operating under a wide range of conditions. The grade of inertial sensors required for such systems is determined to a large extent by the duration of GPS interruptions expected; such interruptions may be expected to occur in military applications as a result of jamming of the satellite signals and more generally as a result of 'signal shading' when attempting to operate in an urban environment. The accuracy

Table 13.1 Comparison of features of inertial and satellite navigation systems

	Advantages	Disadvantages
Inertial navigation systems	High data rate Provides both translational and rotational data Autonomous – not susceptible to jamming	Unbounded errors Knowledge of gravity required
Satellite navigation systems	Errors are bounded	Low data rate No attitude information Susceptible to jamming – both intentional and unintentional

of the inertial sensors is also a factor in applications calling for precision estimates of velocity and attitude, in addition to positional data. For systems in which the threat of interference is minimal, future inertial systems incorporating GPS updates are expected to provide 1 m (CEP)⁴ navigation accuracy.

A number of different integration architectures have been developed to allow INS and GPS to be combined; the level of integration depending in part on whether one is dealing with the creation of a new system, or the addition of GPS updates as a retro-fit to an existing system. A number of INS–GPS integration schemes that are in use or under development for the future are described below. Four main classes of integration architecture may be defined [16, 17], viz.

Uncoupled systems in which GPS estimated position is used simply to reset the INS indicated position at regular intervals of time;

Loosely coupled systems in which the INS and GPS estimates of position and velocity are compared, the resulting differences forming the measurement inputs to a Kalman filter;

Tightly coupled systems in which the GPS measurements of pseudo-range and pseudo range rate are compared with estimates of these quantities generated by the inertial system;

Deep or ultra-tightly coupled systems which combine the GPS signal tracking function and the INS/GPS integration into a single algorithm.

13.7.1 Uncoupled systems

This is the simplest method of deriving the respective benefits of GPS and INS, whilst maintaining the two systems operating independently and providing system redundancy. By using GPS position and velocity estimates to reset the INS, the growth

⁴ Circular error probable (50%).

of errors in the INS estimates of position and velocity is bounded. Whilst this approach involves minimal changes to either system, it does not provide the opportunities for performance enhancement and jamming avoidance that are possible with the coupled systems described in the following section.

13.7.2 Loosely coupled integration

This approach allows the GPS to function autonomously, whilst simultaneously providing measurement updates to the inertial system. The two systems are effectively operated in cascade with the position and/or velocity estimates provided by the GPS navigation calculation forming measurement inputs to an INS–GPS integration Kalman filter. A simplified representation of a loosely coupled INS–GPS integration architecture is given in Figure 13.18.

The two main advantages of loosely coupled integration are simplicity and redundancy. This approach can be used with any INS and any GPS receiver, and is therefore well suited to retro-fit applications. In a loosely coupled configuration, it is usual to provide a stand-alone GPS navigation solution, in addition to the integrated solution. The redundant navigation solution can be used to monitor the integrity of the integrated solution and to facilitate a filter failure recovery process should the need arise.

In the scheme described here, the integration filter provides estimates of the INS errors, which may be used to correct the inertial system following each measurement update. GPS uses the INS purely to aid the satellite signal acquisition process with this level of system integration. As in all of the integration architectures described here, aiding the satellite receiver's code-tracking loops with inertial-sensor information allows the effective bandwidth of these loops to be reduced. This feature allows an

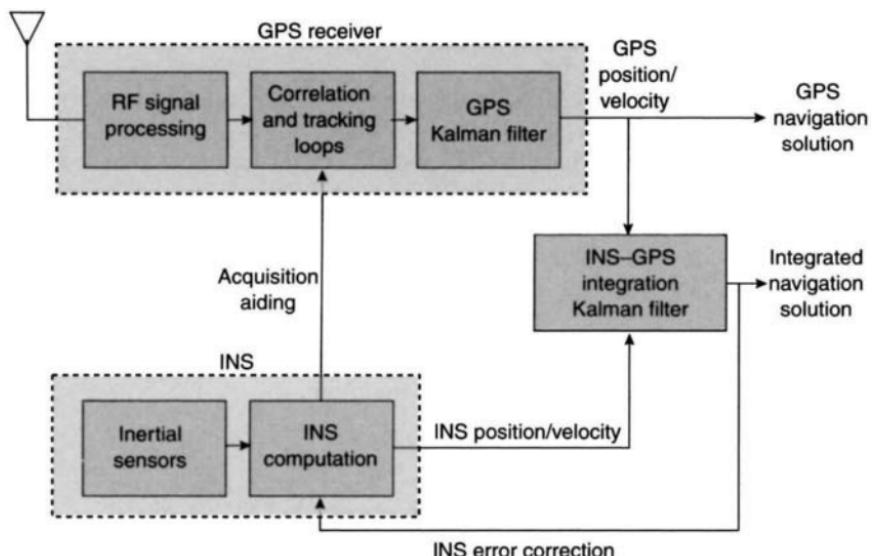


Figure 13.18 Loosely coupled INS–GPS integration architecture

improvement in the ability of the receiver to track signals in a noisy environment, as will exist in the presence of signal jamming.

Whilst GPS position updates alone may be used to aid the inertial system, it is more usual to use both position and velocity measurements for a more robust solution. Because there are fewer integration steps between attitude errors and sensor biases, these errors propagate more rapidly as velocity errors. Therefore, velocity measurements allow more immediate estimates of sensor biases and attitude errors to be obtained. However, the use of velocity measurements alone reduces the observability of position errors in the INS. For these reasons, it is customary to use both GPS position and velocity updates to aid the inertial system in most integration algorithms of this type.

The main problems with loosely coupled INS-GPS integration stem from the use of cascaded Kalman filters; the fact that the output of the GPS Kalman filter is used as a measurement input to the integration filter. In formulating a Kalman filter, the implicit assumption is made that the measurement errors are uncorrelated, that is, that the measurement noise is 'white'. For the system configuration considered here, such an assumption is not necessarily true. For example, situations can arise where the integration algorithm samples the GPS data faster than the tracking loops can supply independent measurements causing the Kalman filter measurement errors to be time correlated. Further time correlation can arise through multi-path effects, the process whereby several delayed copies of a signal reach the antenna following reflection from nearby surfaces.

The correlation time of the GPS Kalman filter outputs varies with the tracking-loop bandwidths, and can be up to 100 s on position and 20 s on velocity and in a dynamic receiver can be 10 s on position and 0.1–1.0 s on velocity. This is too short for the correlated errors to be estimated, but long enough to slow down the process of estimating the INS errors within the integration filter. As a result of these issues, the selection of the Kalman filter measurement update interval becomes critical.

To overcome this problem, the measurement update interval may be increased until the measurement errors are no longer correlated. Alternatively, estimates of the correlated errors can be derived by modelling them as Markov processes and including the correlation time as an additional state in the Kalman filter system model.

Other factors that should be borne in mind when attempting to use loosely coupled integration are as follows. Signals from at least four separate satellites are required to form and maintain a GPS navigation solution, although degraded navigation can be maintained for short periods using only three satellites. Therefore, in situations where fewer satellites are 'in view', the GPS cannot be used to aid the INS. In addition, the integration filter requires knowledge of the covariance of the GPS filter outputs. This varies with satellite geometry and availability, and for many GPS receivers, covariance data are unreliable or not available at all.

13.7.3 *Tightly coupled integration*

Figure 13.19 shows a simplified representation of a tightly coupled INS-GPS scheme; also referred to as closely coupled, centralised or direct integration architecture. In

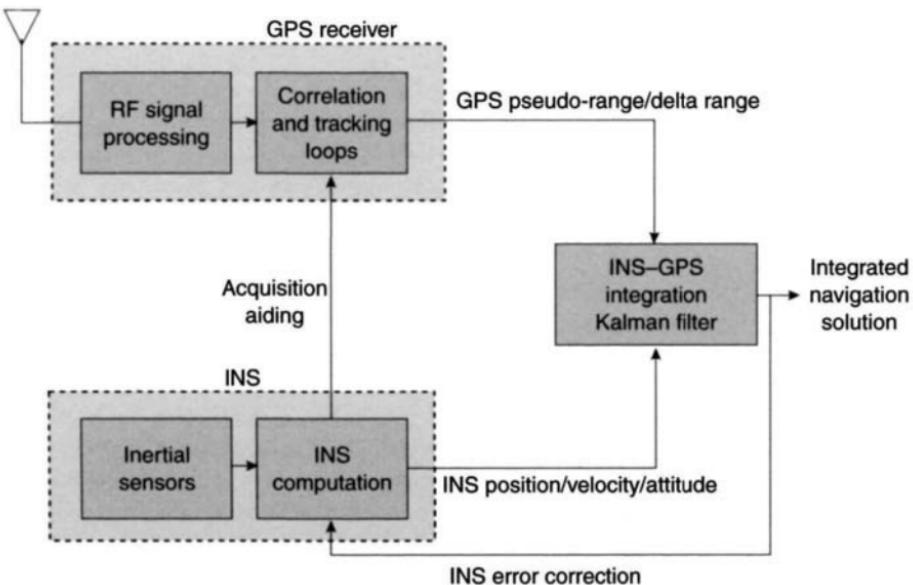


Figure 13.19 Tightly coupled INS–GPS integration architecture

this approach, the GPS filter becomes an integral part of the integration filter, which accepts measurements of pseudo-range and pseudo-range rate provided by the GPS tracking loops. These measurements are used to generate estimates of the errors in the INS. The corrected INS navigation solution forms the integrated navigation solution, and either the corrected or the raw INS data may be used to aid the GPS tracking loops. Signal timing is critical if this system is to operate successfully.

Whilst either pseudo-range or pseudo-range rate measurements may be used, it is common practice to use both. The two measurements are complimentary in that pseudo-range comes from the GPS code-tracking loop, whilst the pseudo-range rate is derived mainly from the more accurate, but less robust carrier-tracking loop.

The benefits of adopting the tightly coupled approach stem mainly from combining the two Kalman filters used in the loosely coupled system as summarised below:

- the statistical problems arising through using the output of one Kalman filter as the measurement input to the second filter are eliminated;
- the handover of GPS position and velocity covariance is done implicitly;
- the system does not require a full GPS solution to aid the INS, GPS data being input to the filter even if only a single satellite signal is being tracked, but accuracy degrades rapidly.

There is no inherent stand-alone navigation solution available from the tightly coupled system as described. However, a GPS only solution may be generated

in parallel with the integrated solution. This solution may be used, when required, for integrity monitoring and failure recovery.

The tightly coupled approach is preferable to the loosely coupled system, giving the better performance in terms of both accuracy and system robustness.

13.7.4 Deep integration

Deep integration, also known as ultra-tightly coupled integration, combines GPS signal tracking and INS–GPS integration into a single Kalman filter, as illustrated in Figure 13.20. Deep integration methods are presently under development. Although many authors have published theory and simulation results [17], a fully working hardware implementation has yet to be published in the open literature.

By tracking the GPS signals together, instead of using independent tracking loops, the tracking of each signal is aided by the others and by the inertial data, bringing three main benefits:

- as fewer independent quantities are tracked using the same data, the effective signal to noise ratio is improved, the more satellites tracked, the greater the improvement;
- multi-path resistance is improved;
- the reacquisition of a signal following a brief interruption as a result of signal obstruction or jamming can be much faster.

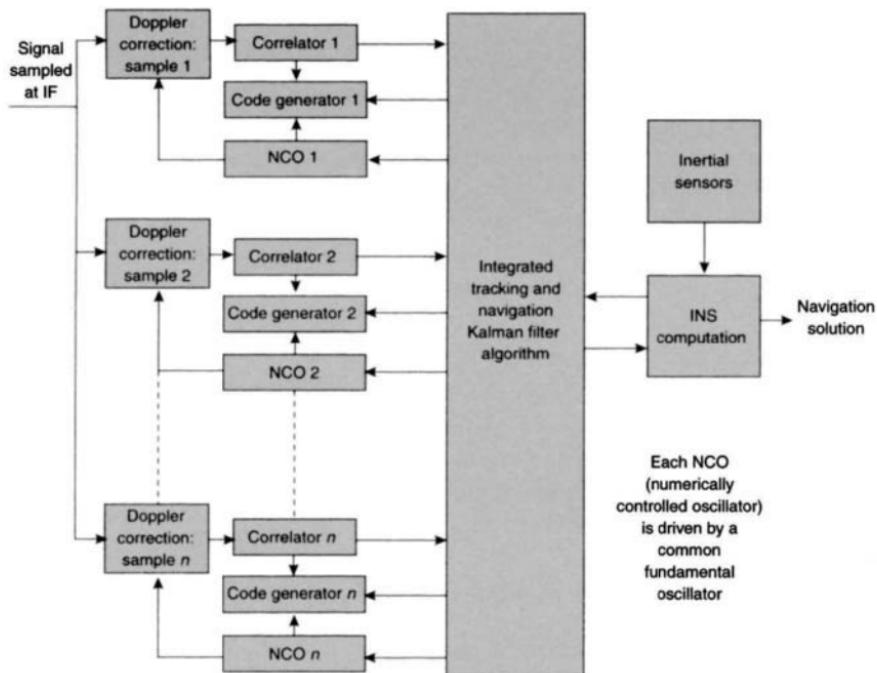


Figure 13.20 Simplified diagram of deep INS–GPS integration architecture

The potential benefits described here may be achieved at the expense of greatly increased complexity; increased computational load and tight time synchronisation requirements for some configurations, as well as high performance INS data to maintain tracking loop lock. A major difference between deep integration and other integration architectures is the need for a very fast update rate, typically 200 to 1 kHz, in order to keep the signal tracking functions in lock.

13.7.5 Concluding remarks

It is not possible to provide hard and fast rules about the performance of one technique over another, because of the range of countermeasures or other techniques that a system may encounter. As a consequence, the resulting performance of any technique is usually critically dependent on the actual scenarios where the platforms are expected to perform, for example where the jammers are positioned or where the obscurations may occur.

Some generalised comments are possible [18]:

- GPS measurement data dominate navigation estimates in an integrated system when the GPS data are available. The steady state navigation error will be reduced through the inertial aiding by effectively averaging out the noise in the GPS measurements.
- Enhancement to navigation accuracy is not critically dependent on the quality of the inertial system, so using a high-grade system has limited benefits.
- Navigation systems with higher-grade sensors benefit more from the in-flight calibration that is possible with an integrated system, owing to the superior compensation that is possible with low-noise sensors, thus the high-grade system provides superior performance during a sustained loss of GPS to the integrated system.
- Tight coupling appears superior for maintaining lock-on to satellites during operation in a jamming environment when compared with loose coupling techniques, however, the benefit is difficult to quantify except in terms of jam to signal ratio.
- The different types of coupling architectures can have an impact on navigation accuracy after short-term losses of the GPS data, owing to the way the data are used for in-flight calibration of the sensors and alignment of the system, however, for long-term loss of GPS it is the sensor quality that will dominate system accuracy.

It is therefore recommended that simulation and hardware evaluation studies are undertaken, paying particular attention to the scenario in which the system is required to operate.

13.7.6 INS aiding of GPS signal tracking

Aiding of GPS carrier tracking by the corrected inertial navigation solution is difficult. This is because very tight time synchronisation is required to keep the carrier tracking locked under high dynamics and to follow the receiver oscillator noise. The development of receivers with software correlators is making this easier as the incoming

GPS signals can be stored and retrieved to match the processing lags in the rest of the system. The tolerances for inertial aiding of GPS code tracking are much wider, so this is an established technique.

Selection of the GPS tracking loop bandwidths is a trade-off between noise resistance and dynamical response. Narrower bandwidths are more tolerant of interference; wider bandwidths respond better to the effects of dynamic motion. By aiding the GPS tracking loops with the corrected INS velocity, the aiding information handles the vehicular motion dynamics, enabling narrow bandwidths to be used to resist noise.

For carrier tracking, the minimum bandwidth is limited by the need to track the oscillator noise. Thus, INS–GPS systems can track GPS code at lower signal to noise levels under dynamical motion than a stand-alone GPS receiver. However, lower-grade inertial systems require constant calibration from the GPS receiver and narrower tracking-loop bandwidths reduce the rate at which independent measurements are provided by the GPS receiver, and hence the INS calibration accuracy. To prevent positive feedback, the gains in the INS–GPS integration algorithm must be matched to the GPS tracking bandwidths. Whilst deep integration does this implicitly [18, 19, 20], this may also be achieved using the tightly coupled integration architecture; a technique known as adaptive tightly coupled (ATC) integration [21]. Recent simulation studies have shown that both deep and ATC integration techniques enable GPS signals to be tracked under noise levels at least 10 dB higher than obtainable with a fixed bandwidth INS–GPS system, when lower-grade inertial sensors are used.

13.8 Multi-sensor integrated navigation

As indicated at the start of Section 13.5, the measurements provided by two or more complementary navigation systems may be combined to yield a navigation solution for a given application; a best estimate of a vehicle's position, velocity and attitude. Prime examples of applications in which it is often desirable, if not essential, to incorporate multiple sources of navigation data are modern military aircraft, long-range missile systems and, more recently, precision guided munitions. Whilst such systems may rely upon an integrated INS–GPS system as their primary source of navigation data, use is frequently made of terrain-reference navigation systems to obtain a more robust navigation solution over the full range of operational conditions anticipated. Terrain-based navigation systems that are used in such applications are terrain contour matching, scene matching area correlation (SMAC) and, more recently, continuous visual navigation (CVN) as described earlier in this chapter.

Issues to be considered when designing such systems are:

- dealing with ambiguous measurement updates;
- choice of integration architectures.

Terrain referenced navigation systems are liable to produce false position fixes on occasions. CVN systems are designed to generate multiple hypothesis position fixes where there is not a unique match between the measurements and the database. Each candidate fix will be accompanied by an associated covariance and an estimated

probability. Ambiguities between these fixes can be treated in different ways. The system may be designed either to accept the position fix that has the highest probability, to compute a weighted fix based upon the estimated probabilities of each, or to reject a fix altogether in the event that its probability is too low.

The multiple hypothesis technique maintains several hypotheses, as the term suggests; each hypothesis is contained in the form of an alternative position-fix Kalman filter, and providing a different set of navigation estimates and an associated probability. Subsequent fixes are used to clarify which of the hypotheses are unlikely to be true, and these are ignored thereafter. Hence, a list of hypotheses is held in order of likelihood, the list being constantly updated as the flight proceeds. At all times, the system has an absolute favourite hypothesis, which is taken to provide the best estimate of position at that time. This method has been proposed as a method for dealing with terrain referenced navigation measurements in an optimal manner [21], although the processing requirement can be very large, but is feasible with modern technology.

Various integration architectures may be applied to combine the data generated using multiple sensor navigation. One possibility is to combine the INS data with the data from each measurement source using a separate Kalman filter. It is postulated that a navigation solution that combines all of the available measurement sources may subsequently be derived by taking a weighted sum of the solutions from each filter based on the covariances of the individual solutions. This approach ignores any unmodelled correlations that may exist between the solutions generated by the individual filters and can result in a false navigation solution.

An additional master Kalman filter may be used to combine the outputs of the individual filters. However, the presence of correlated noise on the outputs of the individual filters can present problems causing the master filter to become unstable. An alternative is required to give a robust solution over a broad range of inputs.

A preferred approach to the integration methods outlined above, is to adopt a centralised architecture [22] in which, as the name suggests, all of the measurement sources are processed by a single Kalman filter. The effectiveness of this approach is, of course, reliant on the availability of representative error models for the individual measurement sources.

To minimise the possibility of a false fix corrupting the integrated navigation solution, integrity monitoring techniques are frequently recommended for sensitive applications. This may be achieved by implementing parallel filters and by monitoring the residual measurement errors throughout the filtering process. This so-called federated scheme was proposed some years ago, but has yet to be implemented in a system.

13.9 Summary

There are many sources of navigation data, which may be used to correct inertial system estimates to provide enhanced navigation performance. These include external measurements derived from equipment outside the vehicle, such as radio navigation aids and satellites, and measurements derived from additional sensors on-board the

vehicle such as various types of altimeters and Doppler radar. The various navigation aids often provide attitude, velocity or position updates, any of which may be used to bound the drift errors arising in an inertial navigation system, and so improve its performance.

The resulting integrated navigation systems often permit substantial improvements in navigation accuracy compared with the performance that may be achieved purely from using inertial systems, even when the inertial navigation system uses very accurate inertial sensors. Whilst very accurate navigation performance may be achieved through the use of higher quality inertial sensors and more precise alignment techniques, the use of integrated systems, such as those described here, often provide a more cost-effective solution.

Techniques for mixing inertial and other measurement data have been described, culminating in the description of an algorithm specifically for aided inertial systems. A Kalman filter algorithm may be used for the integration of different measurement data with inertial measurements. The design example described has shown how external measurement data can be combined with the inertial system information to bound the growth of navigation errors. As a result, it may be possible to allow some minor relaxation in pre-flight alignment accuracy and in the precision of the inertial sensors. Such techniques can be extended to achieve a measure of sensor calibration as part of the aiding process.

The increased accuracy and availability of satellite navigation systems (GPS, GLONASS) have encouraged designers of modern combined systems to become more reliant for good performance on the satellite component. So, now military concerns about vulnerability to interference, reliability and receiver jamming are being investigated by a variety of methods including the use of controlled radiation pattern antenna and integration algorithm developments. For particularly sensitive applications, reversionary modes of operation incorporating additional navigation aids are frequently adopted.

References

- 1 PARKINSON, B.: 'Origins. Evolution and future of satellite navigation', *Journal of Guidance, Control and Dynamics*, 1997, **20** (1), pp. 11–25
- 2 van DRIEL, N.: 'A review of the status and capabilities of Navstar-GPS', NLR TP 92042 U, National Aerospace Laboratory, The Netherlands, October 1992
- 3 FITZSIMONS, B.: 'Satellite navigation', 'Aerospace', *Royal Aeronautical Society Journal*, June 1994
- 4 NAVSTAR-GPS User Equipment – Introduction (Public release version) Navstar-GPS Joint Program Office, Los Angeles Air Force Base
- 5 KAPLAN, E.D. (Ed.): 'Understanding GPS principles and applications' (Artech House, Boston, 1996)
- 6 van GRAAS, F.: 'GNSS augmentation for high precision navigation services'. Proceedings AGARD MSP Lecture Series LS 207, on *System implications and innovative applications of satellite navigation*, Paper 8, 1997

- 7 LOVERRO, D.: 'Global position system (GPS) modernisation'. Proceedings of the *National Technical Meeting of the Institute of Navigation*, 2001, pp. 46–57
- 8 McDONALD, K.: 'Performance improvements to GPS in the decade 2000–2010'. Proceedings of the *Annual Meeting of the Institute of Navigation*, 1999, pp. 1–15
- 9 KAYSER, D., and SCHANZER, G.: 'Effects of the specific military aspects of satellite navigation on the civil use of GPS/GLONASS'. AGARD Conference Proceedings No. 556, *Dual usage in military and commercial technology in guidance and control*, March 1995
- 10 HIN, A.: 'Magnetic compasses and magnetometers' (Adam Hilger Ltd., London, 1968)
- 11 HANDLEY, R.J., ABBOT, J.P., and SURAWY, C.R.: 'Continuous visual navigation—an evolution of scene matching'. Institute of Navigation: Proceedings of the *National technical meeting*, 1998, pp. 17–224
- 12 HANDLEY, R.J., DACK, L., and McNEIL, P.: 'Flight testing of the continuous visual navigation (CV) system', *Journal of Defence Science*, 2002, 7 (1)
- 13 HANDLEY, R.J., GROVES, P.D., McNEIL, P., and DACK, D.: 'Future terrain referenced navigation techniques exploiting sensor synergy'. The European Navigation Conference, *GNSS 2003*, Graz, Austria, April 2003
- 14 STEILER, B., and WINTER, H.: 'AGARD flight test instrumentation series, vol. 15 on Gyroscopic instruments and their application to flight testing'. AGARDograph No. 160, September 1982
- 15 TITTERTON, D.H., and WESTON, J.L.: 'In-flight alignment and calibration of a tactical missile INS'. DGON proceedings, *Gyro Technology Symposium*, Stuttgart, 1990
- 16 GROVES, P.D.: 'Principles of integrated navigation'. Issue 1.1 of a training course prepared by the Telematic Solutions Group of QinetiQ Ltd, Farnborough, October 2003
- 17 SCHMIDT, G.T., and PHILLIPS, R.E.: 'INS/GPS integration architectures'. NATO RTO Lecture Series LS 232, October 2003
- 18 PHILLIPS, R.E., and SCHMIDT, G.T.: 'GPS/INS integration'. Proceedings of AGARD MSP Lecture series LS207, *Systems implications and innovative applications of satellite navigation*, Paper 9, 1997
- 19 GUSTAFSON, D., DOWDLE, J., and FLUECKIGER, K.: 'A high anti-jam GPS-based navigator'. Proceedings of *ION National Technical Meeting*, 2000, pp. 438–446
- 20 SENNOTT, J., and SENFFNER, D.: 'A GPS carrier phase processor for real-time high dynamics tracking'. Proceedings of *ION 53rd Annual Meeting*, 1997, pp. 299–308
- 21 OWEN, J., GROVES, P., and HANDLEY, R.: 'The development of aircraft and missile navigation systems precision attack', *Journal of Defence Science*, 2002, 7 (1)
- 22 GROVES, P.D., and LONG, D.C.: 'Adaptive tightly-coupled, a low cost alternative anti-jam INS/GPS integration technique'. *ION National Technical Meeting*, January 2003

Chapter 14

Design example

14.1 Introduction

The objective of this chapter is to apply the various aspects of strapdown inertial navigation, discussed in the preceding chapters, to a design example. Before any design of an inertial navigation system can be undertaken, it is necessary to make a careful study of the requirements to be placed upon the navigation or stabilisation system in order that the resulting design may be capable of meeting that requirement in a cost effective manner.

The definition of the performance requirement of a navigation or stabilisation system will usually define the accuracy to which it must perform. For example, in the case of a navigation system, it will state the accuracy in position, velocity and attitude to be achieved for a given vehicle and the period of time for which that navigation must take place. It is also vitally important to define the conditions existing at the start of navigation, and the total environment in which the system must operate. These factors, taken together, will influence each of the following aspects of the design solution:

- the system configuration required;
- the navigation system error budget;
- the methods by which the system can be aligned immediately prior to the start of navigation;
- the inertial sensors which are most appropriate for the task;
- the design of the system computational algorithms;
- system testing, calibration and compensation requirements.

These aspects are discussed in the context of a generic requirement for a strapdown inertial navigation system which is needed to achieve mid-course guidance of a hypothetical surface-launched tactical missile. It is assumed that preliminary studies have been conducted at the weapon system definition stage resulting in a set of missile system requirements and the formulation of a navigation system performance

specification. The navigation system requirement is first presented followed by a series of analytical steps leading to the definition of a navigation system capable of satisfying the requirement.

14.2 Background to the requirement

A navigation system is required to be designed for installation in a guided missile that will be launched from a ship. This missile is part of a weapon system that has been defined to provide naval vessels with protection against a variety of airborne threat weapons that could attack from any direction or elevation of approach. In the case of a multiple attack, it has been assumed that the attack has been co-ordinated so that the threat weapons could arrive at the ship simultaneously either from the same direction or from different directions.

The proposed guided missile is to be launched vertically, directly from a storage silo (magazine) on the ship, and then undertake a turn manoeuvre in the pitch plane during the first few seconds of its flight. This vertical launch and turnover manoeuvre from a silo gives the weapon system complete flexibility of direction in which it can fly and a high rate of fire enabling a large number of attacking weapons (targets) to be intercepted. Additionally, the missile will be required to be as autonomous as possible after launch in order to achieve this high rate of fire.

The weapon system has a target tracking device, such as a multi-functional radar, that is able to track targets as they approach the ship. A tracking algorithm predicts the future path of each approaching target. This information is used to predict where the missile launched from the ship would intercept an approaching enemy weapon. Hence, during the turnover manoeuvre, the ship-launched missile is turned so that its flight path is directed towards this interception point.

It is proposed in the missile design that the missile should use inertial guidance during the launch, turnover and post-turnover or mid-course phases of flight. However, when the missile is within a given range of the target, a target seeker on-board the missile will be activated to acquire and track the target. At this stage, the missile will use the information provided by the seeker to steer itself to intercept the target.

During the phases of flight that use inertial guidance techniques, the missile is to use knowledge of its position, velocity and attitude provided by the on-board inertial navigation system, together with target position and velocity provided by a target tracking device on the ship and transmitted to the missile during flight using a radio data link. In order that the homing head may acquire the target to allow terminal guidance to be initiated, the homing head must be pointed in the right direction. This may be accomplished during the mid-course phase of flight using position and attitude information provided by the on-board navigation system in combination with the measurements of target position provided by the shipboard sensor.

In addition, throughout the flight of the missile, the navigation system is required to supply measurements of the components of missile linear acceleration and turn rate in body axes for missile control purposes. These quantities are required to provide the missile autopilot with suitable feedback signals.

14.3 The navigation system requirement

14.3.1 Navigation data required

The requirement is for a navigation system to be installed in a ship-launched guided missile to provide position, velocity and attitude data in a suitable reference frame together with measurements of linear acceleration and angular rate in body axes. Ideally, the navigation system should be capable of operating autonomously throughout flight with a view to minimising the quantity of data which need to be passed to the missile when in flight.

14.3.2 Operating and storage environment

Each missile is to be launched perpendicular to the deck of a ship operating in various sea states up to and including rough sea conditions. Typically, the flight profile of the missile will involve a boost phase, during which the missile executes its turn manoeuvre, followed by a period of sustained motor thrust to maintain near constant speed to interception. The missile is assumed to accelerate from rest at up to 500 m/s^2 (approximately $50g$) for a period of 3 s by which time it reaches a speed of 1000 m/s. Thereafter, the vehicle is assumed to maintain this speed for a further 12 s. Example trajectory and speed profiles are shown in Figures 14.1 and 14.2, respectively.

Over the total flight duration of 15 s, the distance travelled is approximately 12.5 km. The precise details of the flight path which the missile may be required to follow will vary depending on the target trajectory. However, it is assumed that the system may be subjected to pitch and yaw turn rates of up to 5 rad/s during flight. The missile is to be roll stabilised throughout flight, but may experience transient roll rates as high as 5 rad/s.

During flight, the on-board navigation system may be subjected to linear vibration having a power spectral density of $0.05 \text{ g}^2/\text{Hz}$ over a frequency band from 10 Hz to

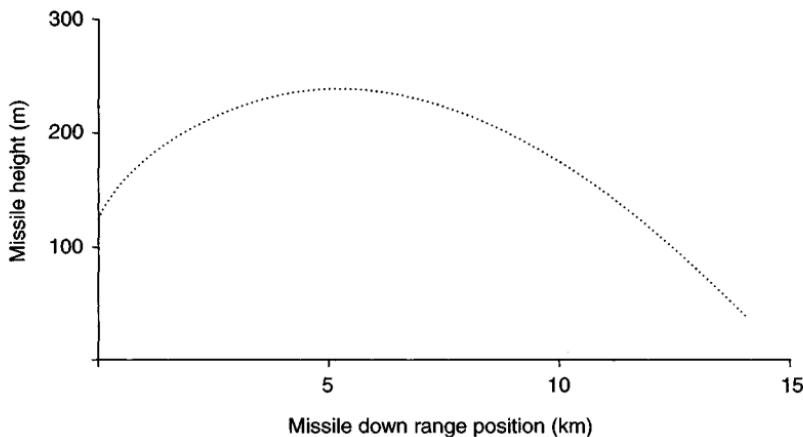


Figure 14.1 Missle trajectory profile

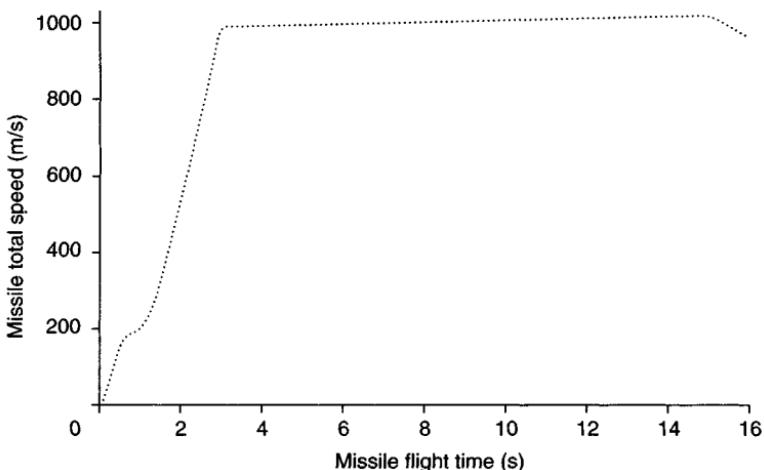


Figure 14.2 Missile speed profile

1 kHz, such motion arising as a result of motor-induced vibration and missile body bending induced by changes in fin commands. In addition, the system must be capable of withstanding shock accelerations of 100g of 0.5 ms duration, and it is required to operate in temperatures ranging from -20 to $+50^{\circ}\text{C}$.

In addition to the requirements given above, it is customary to specify the environmental conditions which the unit must be able to withstand during storage. For military systems, the storage requirements are often more extreme than those specified for the operational system. However, for this design example, the emphasis will be on the in-flight conditions.

14.3.3 Performance

Estimates of missile position, velocity and attitude are required to an accuracy which will enable the missile to be guided from its launch point to a point at which there is a high probability of its homing head acquiring the designated target. However, previous analysis of the missile system reveals that the critical parameters are those which determine the accuracy with which the homing head can be pointed in the direction of the target, namely cross-track position, height, pitch attitude and yaw attitude. In order that a successful transition or handover from inertial mid-course to terminal homing guidance may take place, the navigation system must be capable of providing cross-track position and height information with respect to the chosen reference frame to an accuracy of 50 m (1σ) together with pitch and yaw attitude accurate to 1° (1σ).

Similar accuracies are required in along-track position and roll, and estimates of the velocity components in all three channels are required to an accuracy of 10 m/s (1σ). These accuracies or better must be maintained over a flight time of up to 15 s.

The measurements of body acceleration and angular rate are required for missile control purposes to an accuracy of 0.1g and $1^{\circ}/\text{s}$, respectively. The system must be able to detect variations in these quantities over a bandwidth of 50 Hz.

14.3.4 System reaction time

It is required that the missile is able to be launched a short time after the decision to fire has been taken, say within 2 s. It is therefore important that the missile sensors are able to provide meaningful outputs within a second of switch-on and that the system can be initiated, or aligned, within a similar period.

14.3.5 Physical characteristics

In addition to the performance and environmental requirements, physical constraints are usually imposed on the size and weight of the complete unit. In this application, it is required that the volume of the navigation system should not exceed 3 l litres and that the weight should be less than 3 kg.

Table 14.1 summarises the performance requirements of the on-board navigation system. This table also shows the environment in which it is expected to operate, physical constraints and its required reaction time.

Table 14.1 Summary of navigation system specification

<i>1σ navigational accuracy requirements:</i> ¹	
Position – all axes	50 m
Velocity – all axes	10 m/s
Attitude – yaw, pitch and roll	1°
Linear acceleration – all axes ²	0.1g
Angular rate – all axes ²	1°/s
<i>Environmental requirements:</i>	
Maximum lateral acceleration – pitch and yaw	50g
Maximum longitudinal acceleration	50g
Maximum turn rate – all axes	5 rad/s
Vibration power spectral density	0.05g ² /Hz in bandwidth 10 Hz to 1 kHz
Shock	100g, 0.5 ms
Operating temperature range	-20 to +50°C
Rate of change of temperature	5°C/min
Maximum altitude	15 km
<i>Physical characteristics:</i>	
Mass	<3 kg
Size	System must be capable of being accommodated within a cylinder of length = 10 cm; diameter = 20 cm (~ 3 l in volume)
<i>System reaction time</i>	<2 s

¹ Many of these are independent and may each contribute to the design.

² Autopilot requirements.

14.4 Why choose strapdown inertial navigation?

In order to satisfy the requirement for an autonomous on-board missile navigation capability, an inertial system is considered to be the most suitable for the application considered here. Whilst both strapdown or platform inertial systems may be designed to satisfy the navigation performance requirements, a strapdown system is ideally suited to this type of application in view of the physical constraints of size and weight imposed by the system specification.

In addition, there is a requirement for body mounted gyroscopes and accelerometers to be used to provide the measurements of linear acceleration and angular rate in missile body axes needed for control purposes. By using a strapdown system, the same sensors may be used to provide the measurements needed for control purposes and to implement the navigation function. If a platform navigation system were to be chosen, separate strapdown gyroscopes and accelerometers would be needed for missile control purposes. Additionally, a four-gimbal platform [1] would be required for the system to operate satisfactorily over the full range of dynamic conditions demanded by the system specification. Because of the extra cost incurred with this approach, the strapdown system option is definitely preferred.

The relative merits of strapdown and platform technology for the application considered here are summarised in Table 14.2, the ticks and crosses indicating conformance or non-conformance respectively with the requirement.

14.5 Navigation system design and analysis process

14.5.1 Introduction

In this section, a systematic approach to the design of a strapdown navigation system is outlined. In general, a sequence of design stages of the type described below will need to be followed for any strapdown system application, although the specific

Table 14.2 Comparison of strapdown and platform options

	Strapdown	Platform
Size	✓	✗
Weight	✓	✗
Performance	✓	✓
Environment	✓	✓
Outputs	✓	✗ ¹

¹ Separate body mounted gyroscopes and accelerometers required.

approach adopted and the areas of design emphasis will often vary substantially from one application to another.

For the system considered here, which is to be installed in a tactical missile, particular emphasis must be placed on assessing and analysing the performance of the inertial system during the dynamic conditions arising during flight. The navigation system will be subject to in-flight manoeuvres, high levels of acceleration and an exacting vibratory environment, all of which will greatly influence the choice of inertial sensors and the accuracy of the measurements which they can be expected to provide during flight. In addition, the requirement for the missile to be launched from a moving platform will affect the accuracy to which the navigation can be initialised or aligned prior to missile launch. Throughout the design process, it is important that a reasonable balance is struck between all potential sources of total system error. For example, there is nothing to be gained by relaxing the requirements placed on the inertial sensors at the expense of an alignment specification which is unrealisable or difficult to meet.

As part of the design process, consideration must first be given to the most appropriate system mechanisation for this application. This is followed by an error budget analysis to assess the magnitudes of alignment, instrument and computational errors which can be accepted. Assessments of error budget requirements may be conducted at various levels from relatively simple 'hand calculations', using the single plane error models discussed in Chapter 12, to more rigorous analysis using simulation which allows the effects of dynamic motion to be assessed more precisely. For the type of application being considered here, in which dynamic effects are expected to be of major significance, simplified calculations are likely to be of limited use and the designer will rapidly need to resort to very complex calculations or, more usually, simulation for the assignment of error budgets. In general, several iterations of the error budget analysis process will need to be carried out before a set of error parameter values which are both practical and realisable can be defined.

Having defined instrument performance characteristics and alignment accuracies, and also identified any potential computational difficulties, the tasks of defining suitable sensors and computational algorithms can begin. Throughout this process, further iterations of the error budget calculations may need to be undertaken in the light of the types of sensors considered before converging on a satisfactory design.

14.5.2 Choice of system mechanisation

During the inertial mid-course phase of missile flight, guidance commands are to be generated by combining the estimates of missile position and velocity provided by the on-board inertial navigation system with estimates of target position and velocity generated by a shipboard tracking device. In order that this may be accomplished satisfactorily, both the target and missile estimates must be generated in the same reference frame.

It may be assumed that the ship is fitted with an attitude and heading reference system, or possibly a full ship's inertial navigation system, which defines a co-ordinate reference frame which is nominally aligned with the directions of true north and the local vertical. Further, it is assumed that all shipboard equipment is harmonised with

this reference system. Hence, the tracking device can provide estimates of target position and velocity in this reference frame. Similarly, the missile navigation system may be nominally aligned with this same frame before the missile is launched. Navigation will then take place in this reference frame, the origin of which will be the location of ship's attitude and heading reference system (or navigation system) at the instant when missile navigation begins.

Therefore, the missile navigation system will provide estimates of its position, velocity and attitude with respect to an Earth-fixed reference frame defined at the time when the missile system starts to navigate, at or shortly before the missile is launched. In order that the target position and velocity estimates may be specified in this same frame, it will be necessary to correct the measurements provided by the target tracker, which are generated in a reference frame which translates with the motion of the ship, to take account of any ship motion which takes place during missile flight. This approach is considered preferable to navigating the missile in a reference frame which moves with the ship to avoid the need to transmit ship motion data to the missile during flight. An Earth frame strapdown navigation system mechanism of the type which may be adopted for this application is described in Chapter 3.

An orthogonal inertial sensor arrangement is considered appropriate for this application, there being no requirement for sensor redundancy or the measurement of particularly high turn rates about any axis which could lead to a skewed sensor configuration being considered. The sensitive axes of the respective inertial sensors should be mounted coincident with the principal body axes of the missile, hence providing direct measures of the lateral accelerations and the turn rates required for autopilot feedback purposes.

14.5.3 Error budget calculations

Having defined the inertial system in broad terms, it is now possible to specify the alignment accuracy required and the performance of the gyroscopes and accelerometers needed to achieve the desired navigation accuracy at the end of the inertial phase of flight. As described in Chapter 12, the overall performance of the inertial navigation system depends upon the values of a large number of error parameters and on the way in which they propagate during the flight of the missile.

Each error parameter may be characterised as a random variable, with a probability distribution having a zero mean and a known standard deviation, which varies from system to system and from flight to flight. The effect of each error source may be quantified with the aid of simulation over a given flight profile and its contribution to the total error budget determined in the manner described below.

Assuming conventional (mechanical) sensors are to be used, the dominant sensor error contributions in a tactical missile application of the type considered here are:

- fixed bias uncertainty;
- g -dependent biases (gyroscope only);
- anisoelastic biases (gyroscope only);
- scale-factor errors;
- sensor mounting misalignments/cross-coupling.

It is assumed, in the analysis which follows, that variations with temperature in these various error terms will be compensated, so that the overall effect can be ignored.

In addition, for a system of this type which must be aligned on a moving ship, alignment errors are expected to contribute significantly to the overall error budget. For alternative sensor technologies, the sensor error contributions which need to be considered will need to be revised, as discussed later.

Various approaches may be used to allocate acceptable error contributions between the various processes that contribute to the overall error. A simple approach is to divide the total error equally between the contributing processes. Preliminary values may be determined in this way for a typical flight profile. However, this approach is unlikely to yield a practical set of error parameter values which bear close relation to 'state-of-the-art' technology and alternative design procedures are therefore sought.

A common alternative method is to assess the sensitivity of the total error budget to each of the contributing errors sources. From this sensitivity analysis, combined with an assessment of the ease with which a given level of performance can be achieved, it is possible to make an 'informed' allocation of the various component error sources. This approach relies heavily upon the designer's knowledge and experience of what is technically feasible and practical. As a result of this process, the size of many of the error parameters may be reduced, where this may be achieved without incurring an excessive cost penalty. By reducing the size of some error parameters, it may be possible to accommodate some of the larger or significant error sources, which are difficult or costly to reduce, in the error budget.

An illustration of this design process is given in the following section.

14.5.3.1 Preliminary error allocation process

In Table 14.3, a list is given of 30 errors which are expected to contribute significantly to the overall error budget in the missile application considered here. Initially, values are assigned to each of the error parameters assuming each contributes equally, in the root sum squared sense [2], to the overall cross-track position and height errors which propagate over a 15 s period of flight. In this analysis, particular attention is focused on maintaining the cross-track position and height errors within the performance limits specified since this is crucial for the achievement of a successful inertial mid-course guidance phase. Whilst it is also vital to meet the attitude (pitch and yaw) accuracy specification, these requirements are more easily satisfied as illustrated in the error budget tables (Tables 14.3 and 14.4).

Therefore, assuming the standard deviation of the total navigation errors resulting from sensor imperfection and alignment errors can be 50 m, then the contribution from each individual error source is allowed to be $50/\sqrt{30} \approx 9$ m (1σ), as defined in Table 14.3.

The results given in Table 14.3, which are purely for illustrative purposes, have been derived for a missile which manoeuvres in the pitch plane alone. Under these simplified conditions, the cross-track and height channels of the navigation system remain largely decoupled from each another and the majority of the error sources give

Table 14.3 Sensor error budget based on equal position error contributions

Error source	1σ value	Cross-track position error (m)	Height error (m)	Pitch error (degrees)	Yaw error (degrees)
<i>Alignment errors</i>					
Pitch attitude errors	0.04° (0.8 mrad)	0	9	0.04	0
Yaw attitude error	0.04° (0.8 mrad)	9	0	0	0.04
Cross-track velocity error	0.6 m/s	9	0	—	—
Vertical velocity error	0.6 m/s	0	9	—	—
Cross-track position error	9 m	9	0	—	—
Height error	9 m	0	9	—	—
<i>Gyroscope fixed biases</i>					
B_{gx}	437°/h	9	0	0	0.09
B_{gy}	100°/h	0	9	0.41	0
B_{gz}	120°/h	9	0	0	0.48
<i>Gyroscope g-dependent biases</i>					
B_{gxx}	11.2°/h/g	9	0	0	0.05
B_{gxz}	115°/h/g	9	0	0	0.07
B_{gyx}	2.1°/h/g	0	9	0.07	0
B_{gyz}	21.3°/h/g	0	9	0.01	0
B_{gzx}	2.9°/h/g	9	0	0	0.08
B_{gzz}	21.5°/h/g	9	0	0	0.01
<i>Gyroscope anisoelastic biases</i>					
B_{axx}	2.9°/h/g ²	9	0	0	0.06
B_{ayx}	0.57°/h/g ²	0	9	0.05	0
B_{azx}	0.59°/h/g ²	9	0	0	0.05
<i>Gyroscope scale-factor errors</i>					
S_{gy}	0.05%	0	9	0.04	0
<i>Gyroscope cross-coupling</i>					
M_{gxy}	0.06%	9	0	0	0.04
M_{gzy}	0.08%	9	0	0	0.04
<i>Accelerometer fixed biases</i>					
B_{ax}	115 milli-g	0	9	—	—
B_{ay}	8.2 milli-g	9	0	—	—
B_{az}	8.8 milli-g	0	9	—	—
<i>Accelerometer scale factor errors</i>					
S_{Ax}	0.42%	0	9	—	—
S_{Az}	1.06%	0	9	—	—
<i>Accelerometer cross-coupling</i>					
M_{Axz}	3.07%	0	9	—	—
M_{Ayx}	0.06%	9	0	—	—
$M_{Ay়z}$	1.01%	9	0	—	—
M_{Azz}	0.07%	0	9	—	—
Total RSS errors (1σ)		36 m	33.7 m	0.42°	0.51°

Table 14.4 Sensor error budget based on a minimum risk strategy

Error source	1σ value	Cross-track position error (m)	Height error (m)	Pitch error (degrees)	Yaw error (degrees)
<i>Alignment errors</i>					
Pitch attitude errors	0.16° (3 mrad)	0	36	0.16	0
Yaw attitude error	0.16° (3 mrad)	36	0	0	0.16
Cross-track velocity error	0.6 m/s	9	0	—	—
Vertical velocity error	0.6 m/s	0	9	—	—
Cross-track position error	1 m	1	0	—	—
Height error	1 m	0	1	—	—
<i>Gyroscope fixed biases</i>					
B_{gx}	50°/h	1	0	0	0.01
B_{gy}	50°/h	0	4.5	0.21	0
B_{gz}	50°/h	3.75	0	0	0.2
<i>Gyroscope g-dependent biases</i>					
B_{gxx}	5°/h/g	4	0	0	0.02
B_{gxz}	5°/h/g	0.4	0	0	0.003
B_{gyx}	5°/h/g	0	21.4	0.17	0
B_{gyz}	5°/h/g	0	2.1	0.002	0
B_{gzx}	5°/h/g	15.5	0	0	0.14
B_{gzz}	5°/h/g	3.8	0	0	0.004
<i>Gyroscope anisoelastic biases</i>					
B_{axx}	0.5°/h/g ²	1.6	0	0	0.01
B_{ayx}	0.5°/h/g ²	0	7.9	0.05	0
B_{azx}	0.5°/h/g ²	7.7	0	0	0.04
<i>Gyroscope scale-factor errors</i>					
S_{gy}	0.05%	0	9	0.04	0
<i>Gyroscope cross-coupling</i>					
M_{gxy}	0.1%	15	0	0	0.07
M_{gzy}	0.1%	11.25	0	0	0.05
<i>Accelerometer fixed biases</i>					
B_{ax}	10 milli-g	0	0.8	—	—
B_{ay}	10 milli-g	11	0	—	—
B_{az}	10 milli-g	0	10.2	—	—
<i>Accelerometer scale factor errors</i>					
S_{Ax}	0.3%	0	6.4	—	—
S_{Az}	0.3%	0	2.5	—	—
<i>Accelerometer cross-coupling</i>					
M_{Axz}	0.1%	0	0.3	—	—
M_{Ayx}	0.1%	15	0	—	—
$M_{Ay়z}$	0.1%	0.9	0	—	—
M_{Azx}	0.1%	0	12.9	—	—
Total RSS errors (1σ)		49.2 m	48.2 m	0.32°	0.32°

rise to navigation errors in one channel only. As a result, the total errors in position and attitude are shown to be well within the requirements specified. In general, the missile will be called upon to manoeuvre in the pitch and yaw planes simultaneously, in which case each error source will contribute to the navigation error in each channel. Further, some additional g -dependent biases and scale-factor errors may become significant in the more general situation, owing to the manoeuvring of the missile. However, the simplified analysis given here is adequate for illustrating the fundamentals of the technique.

14.5.3.2 Refinement of the error budget allocation

Referring to the error terms given in Table 14.3, the angular alignment accuracies and some of the sensor bias values suggested will be very difficult to achieve in practice. Therefore, the design engineer must now bring his or her experience of system design and sensor technology performance to bear on the selection of suitable error coefficient values, bearing in mind those parameters to which the overall errors are most sensitive.

There is usually some limited scope for relaxation of some contributions to the error budget at the expense of other parameters. For example, in a design based upon conventional gyroscopes, the fixed bias contributions to the error budget may be allowed to increase whilst raising the g -dependent biases to levels which can more easily be achieved in practice. Analysis reveals that system performance in this type of application is particularly sensitive to the values of the g -dependent bias coefficients. It can also be seen that certain of the cross-coupling terms associated with both gyroscopes and accelerometers will need to be small in order to achieve the required performance.

At all times, it is of course essential to ensure that the contribution of any one error term does not exceed the total error budget. It will usually be necessary to carry out several iterations of this parameter selection process before arriving at a reasonable set of values. An example set of error parameter values is given in Table 14.4 along with their respective contributions to the total position and attitude error budgets.

Table 14.4 shows clearly the dominant sources of error to be the angular alignment errors together with certain of the g -dependent gyroscope biases and accelerometer cross-coupling, both of which give rise to large position errors in the presence of missile longitudinal acceleration. In addition, gyroscope cross-coupling contributes substantially to the overall error budget in the presence of the pitch turn manoeuvre which the missile undergoes during its boosted phase of flight.

Using the alignment and sensor errors given in Table 14.4, the total along-track position error, roll error and velocity errors, which are not given in the table, have been calculated to be as follows:

$$\text{RSS along-track position error} = 41 \text{ m};$$

$$\text{RSS roll error} = 0.3^\circ;$$

$$\text{RSS along-track velocity error} = 0.7 \text{ m/s};$$

$$\text{RSS cross-track velocity error} = 3.7 \text{ m/s};$$

$$\text{RSS vertical velocity error} = 3.8 \text{ m/s}.$$

It can be seen that each lies within the limits defined by the specification.

Having generated detailed sensor specifications, it is essential to evaluate the performance of the resulting system over a representative set of missile trajectories. The effect which many of the errors have on overall navigation performance is often highly dependent on the precise motion to which the system is subjected during flight. It may become necessary to refine further some of the error coefficient values at this stage of the design in order to specify the system adequately.

Whilst following the type of procedure described above, the designer may well wish to combine certain errors, especially those which propagate in a similar manner and have similar effects on navigation system performance. Some examples are given below.

- Gyroscope anisoelasticity – in the presence of cyclic motion, biases arise on the output of a conventional gyroscope owing to the effects of unequal bearing compliance.
- Accelerometer vibro-pendulous error – additional biases arise on the outputs of pendulous accelerometers in the presence of vibration.
- Coning and sculling motion – further angular rate and linear acceleration biases can arise if the inertial sensors are subjected to coning and sculling motion respectively.

The gyroscope and accelerometer biases used in the error budget analysis may need to be increased to take account of such effects, all of which are discussed at some length in Chapter 12.

14.5.4 System alignment

A vital factor in the performance of an inertial navigation system is the accuracy to which it can be aligned, or initialised, prior to the start of navigation. If the alignment contribution to the total error budget is to be as suggested in Section 14.5.3, then a 1σ angular alignment accuracy of approximately 0.16° (10 arc min) is required together with initial velocity and position estimates accurate to 0.6 m/s and 1 m (1σ), respectively.

Whilst the achievement of sufficiently small angular misalignments is of particular concern, significant errors can also arise in the velocity estimates used to initialise the navigation system as a result of lever-arm motion between the ship's reference and the missile. For a tactical missile system which is to be launched from a moving platform, the achievement of satisfactory alignment accuracy is particularly difficult. In addition to the alignment accuracy required, a major factor influencing the choice of the alignment scheme is the time available to accomplish such an alignment.

Following from the discussion of shipboard alignment methods in Chapter 10, it is postulated that an alignment to this order of accuracy may best be achieved within the short period of time available (~ 1 s) by the rapid transfer of alignment data from an inertial navigation system mounted on the missile launch silo on the ship using the so-called 'one-shot' alignment method. The alignment of this unit would need to be maintained independently using shipboard measurements provided by the ship's own inertial reference system, or through the use of satellite data updates.

In view of the potential difficulties of achieving an accurate alignment of the missile system on board a moving ship, it may well be advisable to relax the alignment requirement further at the expense of tightening instrument performance. However, it can be seen from the error budget that there is very limited scope for further relaxation of the alignment requirement and for the purposes of this design exercise it is assumed that the system can be aligned to the accuracy suggested.

14.5.5 Choice of inertial instruments¹

Having defined a set of performance characteristics which must be satisfied by the inertial sensors in order to meet the overall system specification, the task of selecting suitable gyroscopes and accelerometers can be undertaken.

With reference to Chapters 4–7 on inertial sensor technology, there is a wide range of gyroscopes and accelerometers capable of satisfying the performance requirements under the flight conditions described above. Previous experience with other strapdown navigation systems would suggest that possible candidates for the body-mounted gyroscopes could be:

- dynamically tuned gyroscope (DTG) or flex gyroscope,
- rate-integrating gyroscope (RIG),
- fibre optic gyroscope (FOG),
- ring laser gyroscope (RLG),
- vibratory gyroscope.

In the case of the accelerometers, the leading candidate is likely to be a form of pendulous force-feedback accelerometer, although other possibilities are as follows:

- silicon accelerometer,
- surface acoustic wave (SAW) accelerometer.

The compliance of these sensors with the various parameters selected from the requirements in the system specification and the error budget analysis are indicated in the Tables 14.5 and 14.6.

Clearly, a number of different sensors could provide the desired angular rate and acceleration data to the required accuracy. Based on the information given in Table 14.5, it is concluded that there are three leading gyroscope candidates; the two mechanical gyroscopes and the fibre optic sensor, whilst Table 14.6 suggests that any of the three accelerometer types considered may be chosen.

It should be noted that the data provided by the ring laser gyroscopes are likely to be far more accurate than required for this application and, therefore, is less likely to be chosen. However, the use of such sensors could possibly enable the performance requirements to be relaxed on other processes which contribute to the navigation system error budget, such as the alignment accuracy or accelerometer performance.

¹ It should be noted that it is not the intention of this exercise to recommend a particular ‘brand’ of inertial sensor. Moreover, as stated previously, it is not the aim of this book to endorse a particular brand or class of inertial sensor.

Table 14.5 Gyroscope options

	Mechanical		Optical		Other Sensors
	RIG	DTG	RLG ¹	FOG	Vibratory
Measurement range	✓	✓	✓	✓	✓
Measurement accuracy	✓	✓	✓	✓	✗ ²
Environmental performance	✓	✓	✓	✓ ³	✓ ⁴
Size and weight	✓	✓	✓	✓	✓
Reaction time	✓ ⁵	✓ ⁵	✓	✓	✓
Risk ⁶ (technical maturity)	Low	Low	Low/ medium	Low	Low

¹ Even a small RLG is likely to have performance well in excess of this requirement.

² This type of sensor could possibly be compensated to give performance close to the desired accuracy.

³ Care would possibly be required in the design and packaging of the system to make this sensor sufficiently insensitive to various aspects of the environment but now feasible.

⁴ Elaborate temperature compensation of this type of sensor is likely to be necessary and may be prohibitively complex. However, it would be very rugged.

⁵ The ability of mechanical gyroscopes to achieve stable operation shortly after switch-on has been examined in the past. It has been demonstrated that sub-inertial sensors of the type required for this type of application can be run-up in about 1 s by applying extra power to the spin motor at switch-on, using so-called 'over-volting' techniques. Some characterisation of sensor error transients may be needed to achieve meaningful outputs within this period of time.

⁶ Risk is a multi-faceted problem, encompassing technical, financial and temporal aspects. In this case, only the technical maturity and pedigree of the gyroscopes are considered.

Table 14.6 Accelerometer options

	Pendulous force-feedback	SAW	Silicon
Measurement range	✓	✓	✓
Measurement accuracy	✓	✓ ¹	✓ ¹
Environmental performance	✓ ²	✓ ²	✓ ²
Size and weight	✓	✓	✓
Reaction time	✓	✓	✓
Risk ³ (technical maturity)	Low	Low/ medium	Medium

¹ Careful compensation is likely to be necessary.

² Careful packaging design is likely to be necessary for some types.

³ Risk is a multi-faceted problem, encompassing technical, financial and temporal aspects. In this case, only the technical maturity and pedigree of the accelerometers are considered, i.e. consistent with the gyroscopic analysis.

Alternatively, the performance requirement of another sub-system within the missile may be relaxed. This would be the subject of a trade-off study for the whole system, and is beyond the intended scope of this design example.

In the event that the designer decides to choose sensors based on the maturity of the technology and minimum risk, then he or she is most likely to select the force-feedback accelerometer and either the dynamically tuned gyroscope or the single-axis rate-integrating gyroscope. A further attraction of the dynamically tuned gyroscope is the fact that only two sensors are required to provide the three axes of angular rate measurements required, potentially offering good value for money.

Alternatively, the more modern sensors may be chosen, that is, either the SAW or the silicon accelerometer and the fibre optic gyroscope, on the grounds that they could offer low total cost in the future through the ease of manufacture of solid-state technology.

The final choice may well be decided by cost, including purchase price and projected total life cycle costs. The purchase price is a complex issue involving availability of components, total order size, production rate and all aspects of calibration and compensation².

14.5.6 Computational requirements

It is important that the data provided by the inertial sensors are processed correctly in order to achieve the required navigation system performance. The on-board computer is required to calculate the missile attitude in the chosen reference frame, resolve the accelerometer measurements into that reference frame and then integrate these quantities to provide estimates of missile velocity and position. In addition estimates of roll angle are to be extracted to provide the feedback signal needed to stabilise the position of the missile about the roll axis.

As discussed in Section 14.5.2, the Earth-frame strapdown system configuration described in Chapter 3 is considered appropriate for this application. However, in this particular case, the Earth's rate terms in the attitude computation process could be ignored. Since angular rate biases of 50°/h can be tolerated, the effects of neglecting Earth's rate (15°/h) will not cause system performance to degrade significantly. However, the Coriolis correction terms will need to be included in the navigation equation owing to the high missile velocity (1000 m/s). It is noted that the removal of the Earth's rate terms will not reduce the overall computational burden greatly.

The crucial aspects of the computation are:

- data rate;
- processing speed;
- truncation levels of mathematical functions used in the computation.

² Sensor technology has advanced (since the original preparation of this design example) to the point where MEMS devices now provide a viable, if not preferred, option for this application. Sensors of this type are now more than capable of meeting the design requirement initiated here, in terms of measurement accuracy, survivability and risk.

Data rate and processing speed are determined by the frequencies of the angular and translational motion to which the navigation system is to be subjected during flight. For the missile application considered here, it is expected that the bulk of the computation, including the attitude update algorithm, the resolution of the accelerometer measurements and the solution of the navigation equation would need to be carried out at a frequency in the region of 100–200 Hz.

It is important to be aware of any cyclic motion which may arise, such as coning or sculling type motions, since they can have an adverse effect on the performance of the system. These types of motion can arise during flight as a result of forces exerted by aerodynamic control surfaces which can excite natural body bending modes of the missile airframe. The degradation in performance introduced by cyclic motion can often be reduced by the use of multiple-speed algorithms of the type described in Chapter 11, in which fast but relatively simple algorithms are used to implement high speed correction algorithms. This approach may need to be adopted in a missile application where significant levels of vibration can arise. For these fast calculations, computational frequencies of around 400 Hz would normally be sufficient. However, these calculations may need to be carried out more rapidly since the specification indicates that vibratory motion at frequencies up 1 kHz may be present. Detailed calculation is necessary in each specific case to ensure that the processing speed exceeds the frequency of the vibratory motion by a factor of at least 2.

Truncation levels of mathematical functions must be selected to allow sufficient accuracy of computation to be achieved without imposing an excessive burden on the system computer. In general, there are trade-off considerations to be made between the speed of computation and the amount of truncation which can be permitted.

In the past, the choice of wordlength has also been a crucial factor in the design of navigation system computers. However, modern systems tend to use floating point processors in which wordlength restrictions do not usually impose a serious limitation. The current status of computer technology is that this is not a technological issue.

There is often a temptation to assume that computational errors can be contained within acceptably small limits and that their contribution to the overall error budget is small compared with the sensor and alignment errors. Whilst this may often be the case, care should be taken in the choice of algorithms, particularly for agile missile applications of the type considered here. As a result of the rapid advances made in computer technology in recent years, the achievement of sufficient computer speed is certainly far less of a problem than it has been in the past.

14.5.7 Electrical and mechanical interfaces

It is essential to check that the electrical interfaces are compatible on either side of any mechanical interface. Additionally, any mechanical interfaces on the inertial system must be mutually compatible with the fixing points on the missile structure. In particular, any mechanical interface with any other unit must not compromise the structural loading and stiffness requirements. For the purposes of this design study,

it will be assumed that these design characteristics can be achieved, as it is a case of applying sound engineering techniques.

14.6 Testing, calibration and compensation requirements

It is most important to evaluate the prototype navigation system to ensure that it fulfils the design requirement. Suitable test plans for the navigation system, and the individual sensors within it, may be devised in accordance with the overall system specification. A series of laboratory tests will be required including static, rate, centrifuge and temperature tests as discussed in Chapter 8.

During the sensor and system evaluation, various systematic errors may become evident. Compensation techniques can be readily applied to such systematic errors. In some cases, compensation enables the derived specification of basic performance of the inertial sensors to be relaxed further.

A typical systematic error that is likely to require correction is temperature dependence of the scale-factors of the inertial sensors. Hence, it may be necessary to incorporate a temperature sensor in the inertial measurement unit. A typical temperature compensation routine is shown below.

Considering a sensor with a scale-factor that varies systematically with temperature. This dependence can be determined from the testing procedure described in Chapter 8, Section 8.5 for gyroscopes and Section 8.6 for accelerometers. The slope of the graph in Figure 8.9 gives the scale-factor temperature coefficient (SFTC) for a gyroscope. A similar parameter can be deduced from Figure 8.21 for an accelerometer.

The scale-factor of a sensor can be corrected for changes induced by variations in the temperature of the sensor, provided that the temperature (T_c) at which the sensor was calibrated is known and the actual temperature (T_a) of the sensor, when the measurement was made, is also known. Then the scale-factor of the sensor at temperature T_a is calculated as follows:

$$SF_a = SF_c + SFTC(T_a - T_c) \times 10^{-6}$$

where SF_a is the scale-factor at temperature T_a , SF_c is the calibrated scale-factor (at temperature T_c), SFTC is the scale-factor temperature coefficient expressed in parts per million per kelvin (or centigrade degree).

Similarly, any systematic variation in the bias of the sensor can also be compensated in an identical fashion.

14.7 Performance enhancement by aiding

If additional navigation measurements could be provided during flight, it may be possible to relax further the sensor performance and the alignment accuracy requirements.

In a tactical missile application of the type considered here, missile position fixes may be provided by tracking the missile during flight and transmitting the data to the missile during flight. Given that such information may be provided at intervals during the flight with sufficient accuracy and passed to the missile, it may be used to correct the on-board inertial navigation system. As a result of this process, there may be scope to relax the specifications of the inertial sensors and to avoid the need for pre-flight alignment to the accuracy proposed for the unaided system.

The reader is referred to Chapter 13 in which a system of this type is described in some detail.

14.8 Concluding remarks

This chapter has provided an outline of the stages needed to produce a design for an inertial navigation system for a generic short range tactical missile application. By analysis of the allocated system error budget in the system specification, it is possible to derive the performance specification for the inertial sensors and to define the computing processes which must be implemented.

In the example chosen, a design has been formulated that can fulfil the technical requirements set out in the specification. Therefore, the first stage of the design is complete. This would have to be confirmed by laboratory and field testing. However, if a suitable design could not be achieved, even with aiding and superior performance sensors, then it would be necessary to negotiate concessions in the technical requirements specification, with possible implications on the overall performance of the system.

References

- 1 BRITTING, K.: 'Inertial navigation system analysis' (Wiley Interscience, New York, 1971)
- 2 TOPPING, J.: 'Errors of observation and their treatment' (Chapman and Hall, 1975)

Chapter 15

Alternative applications of IN sensors and systems

15.1 Introduction

Much of this text book has concentrated on the design and operation of inertial sensors and systems for inertial navigation. However, inertial sensors have been used in many applications that do not require navigation, as well as those requiring fundamental data for inertial navigation or guidance. A relatively common application is rotation of components or projectiles to give stability and to reduce flight path dispersion, but there are many others. An objective of this chapter is to consider some of the applications that have benefited from the use of inertial sensors or techniques.

The range of applications in which inertial sensors and systems are used, where navigation is not the primary function, is quite diverse. In many cases, the primary purpose of the inertial system is vehicle or equipment stabilisation and control. Examples of such applications include:

- active suspension systems in high-performance vehicles;
- autopilots in aircraft and ships;
- geodesy;
- laser radar functions and terrain following;
- passive missile roll control (rollerons);
- personal transport (Segway and Ibot);
- seeker head stabilisation;
- sightline stabilisation;
- fundamental physical studies.

Other applications use non-inertial methods for navigation aids, or for information and entertainment; in moving map displays, for example, in which additional information may be superimposed to aid recognition of particular features.

This chapter seeks to provide the engineer with an outline of the way in which the inertial sensor or system may be applied to the task. The description includes the role of the inertial sensor, or sensor system in each application, together with a discussion of critical issues. Where appropriate, the use of alternative approaches is discussed, emphasising the relative benefits and drawbacks in each case. Where applicable, an indication is provided of the performance of the system or device.

We begin by examining the role of inertial navigation systems for well-bore surveying, where particular problems may be encountered, chiefly in relation to the environment in which such systems must operate. In the design description given here, the issues that must be investigated and evaluated when attempting to design systems to operate at the extremes of temperature, pressure, shock and vibration that can be encountered by well-bore survey systems are discussed. Somewhat briefer descriptions are provided for a selection of other examples.

15.2 Borehole surveying

15.2.1 Introduction

To extract hydrocarbons, such as oil and gas, from beneath the surface of the Earth, well bores are drilled by rotating a drill bit attached to the end of a drilling assembly. Most oil and gas exploration involves drilling highly deviated and substantially horizontal wells to increase production and to obtain additional hydrocarbons from Earth's formations frequently located several kilometres (horizontally) from beneath the drilling platform. This is illustrated in Figure 15.1.

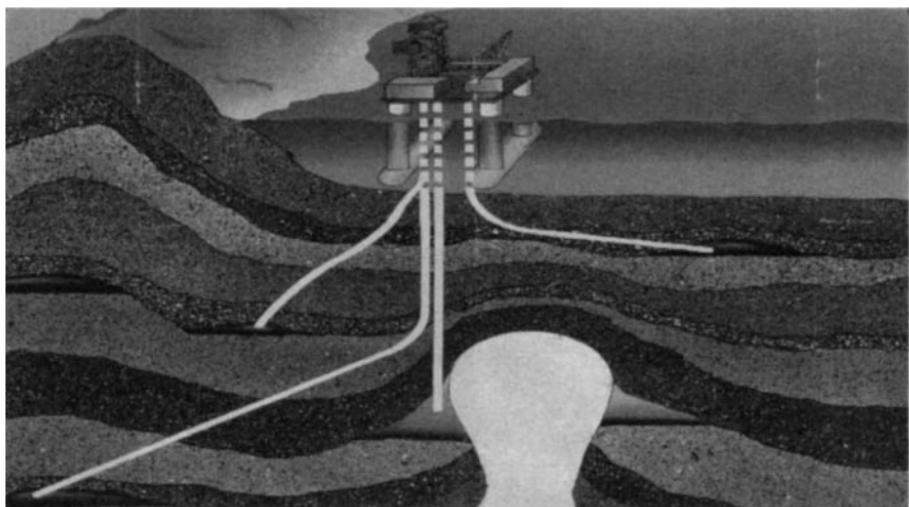


Figure 15.1 An example of a well-bore trajectory

There is a critical requirement for geological survey, mining and, in the drilling of oil and gas wells, to derive accurate representations of the well-bore trajectory. Such information is vital for:

- cost-effective drilling;
- the avoidance of well-bore collisions;
- to allow precise relief wells to be drilled should the need arise.

Directional surveys provide definitive data that depicts the position of a borehole or well path in three-dimensional space. The only point in the path that is truly known and in most cases ever likely to be visible (some civil engineering and mining boreholes excluded) is the start point, generally referred to within the oil industry as the surface position, or the co-ordinates of the well head. The trajectory and position of the well as it progresses under ground is calculated by taking a series of directional surveys, which define the attitude of the well in space at a known measured depth down the well or along the well path. In its simplest form, the well path is represented by a series of vectors calculated using the survey station point angular data obtained from the survey tool and the incremental depth to each successive survey point. The frequency at which surveys are taken and recorded has an impact on the overall accuracy of the calculated well path from its initial starting point. In general, well bores with a continuous and/or variable change in attitude (dogleg) require to be surveyed at a higher frequency than a lined out (tangential or straight) hole.

The development and manufacture of measurement systems capable of operating underground to the levels of accuracy required presents some major engineering design challenges, not least in terms of the extreme environmental conditions under which such systems are required to operate. Well-bore measurement systems may be required to operate at temperatures in excess of 200°C and at pressures up to 25 000 psi (172 MPa). Whilst well-bore surveys may be carried out following the completion of the drilling process, it is economically desirable to generate survey data during the drilling process, in which case the measurement system must survive and operate in a severe shock and vibrational environment.

Additional factors which impact the design are physical size constraints and limitations on the levels of electrical power available under the ground. To allow such systems to penetrate narrow gauge well bores many thousand of metres beneath the surface of the Earth, they are required typically to be contained within a diameter of a few centimetres, although the length of the equipment is less constrained. Well-bore survey tools are therefore installed in long narrow gauge pressure housings.

15.2.2 Historical background

This section provides a brief history of the various techniques used over the years to provide well bore trajectory data, culminating in modern magnetic and gyroscopic systems. There follows a description of a full inertial well-bore survey system in which many of the critical design factors are highlighted.

The original instrumentation used within the industry measured hole inclination only; this technique indicated the degree to which a drilled well path deviated from

the vertical. Subsequently, a magnet was incorporated into the mechanical plumb bob derivative of this instrument with the assembly allowed to float on jewelled bearings to provide both attitude with respect to the vertical and the direction of that attitude; to form a compass inclinometer unit. Many units of this type are still in use today around the world.

Dual-axis semi-solid-state magnetic sensing tools were first introduced in the late 1960s as steering tools, operated on a cable or wire line, primarily to improve the efficient kick-off drilling of deviated directional wells. These wire-line tools provided near continuous data at the surface. Tri-axial/orthogonal magnetic tools were subsequently introduced and developed to provide survey data during the drilling process, or measurement whilst drilling (MWD) systems. Such tools have been refined over the years and, when combined with precise knowledge of the local magnetic field, provide accurate well-bore survey data. The data provided by magnetic survey tools are processed using sophisticated analysis software, which allows down-hole sensor problems to be readily identified, so facilitating the potential for real-time data correction and uninterrupted operations. Magnetic sensor MWD data have become accepted as definitive quality for several years by the oil industry. However, certain well geometries preclude the generation of an accurate magnetic survey, when the measurement system is attempting to operate in the presence of magnetic interference. Under such conditions, gyroscopic survey systems offer an attractive alternative.

It is noted that the use of magnetic sensing tools provides data referenced to magnetic north. Knowledge of magnetic declination at the well site is therefore required and is crucial to allow surveys to be referenced to a geographic/true north reference frame, as is frequently required. Therefore, precise knowledge of the angle of declination at the survey location is vital to allow a survey to be generated with respect to the designated reference frame.

Gyroscope sensor survey tools were developed to maintain an established geographic north reference, set at the surface, when the survey tool is underground. In early gyroscopic survey tools, the instrument attitude data generated were time recorded on photographic film whilst down the hole. The film was read at the surface and each survey data set was matched to the appropriate survey depth with the aid of a synchronised surface time log. The three-dimensional co-ordinate position of the well path was then calculated, as now, by simple trigonometry. This type of gyroscopic tool was used by the oil industry to provide the definitive survey data for most of the directional well survey projects in earlier years.

Significant gyroscope developments resulted in the improvement of gyroscopic data quality. Notably these were the introduction of the first surface recording wire-line gyroscopes¹ followed by the application of rate-sensing gyroscopes, commonly known as north-seeking gyroscopes, so-called as it describes the system referencing technique employed.

¹ Surface recording gyroscopic tools, run on wire line in a cased hole, are capable of producing quality data in well bores inclined at up to 65° to the vertical. Data quality can become erratic at higher angles of inclination, where gravitational forces are no longer sufficient to maintain tool transport along the well path, and it becomes necessary to pump the survey tool down the well.

Whilst magnetic-survey tools are widely used, systems incorporating high-accuracy gyroscopes with a north-seeking capability are used in regions of high magnetic interference. This technique is used to derive directional data (azimuth and inclination) as the survey system is lowered or raised in a well-bore. These data, together with hole-depth measurements, are combined using a survey calculation method, such as minimum curvature, to yield accurate trajectory information in a local geographic reference frame.

Modern developments have led to systems built around a full inertial navigation system capability, which make no assumptions about the shape of the well and facilitate uninterrupted surveys.

The gyroscopic systems currently used to survey underground bore holes are many and varied. In the following, a system incorporating a full inertial navigation system is described, and some of the major factors that influence its design are discussed.

15.2.3 Inertial survey system

A miniature strapdown inertial measurement unit for a well-bore survey system is illustrated in Figure 15.2. This particular unit contains two miniature dual-axis dynamically tuned gyroscopes (visible to the right of the picture) and three force-feedback accelerometers (only those devices that provide measurements with respect to the lateral axes of the probe are visible in the picture).

Systems of the type described here may be used to conduct a well-bore survey following completion of the drilling of a well. In such applications, the survey probe, containing the inertial navigation system, may be lowered and raised in the well on a continuously spooled wire-line. The wire-line also provides the medium through which communication between the surface and the probe is maintained throughout the survey process.

A system of this type comprises surface and down-hole modules as depicted in Figure 15.3. The surface-system includes a PC or notebook computer together with a power supply and control unit. They form the interface between the computer and the wire-line cable on which the probe is lowered and raised in the well. The measuring system that is placed in the well, the down-hole unit, comprises the electronic survey probe containing the inertial navigation (IN) system, support electronics, communications hardware, and associated running gear. Bi-directional communication between

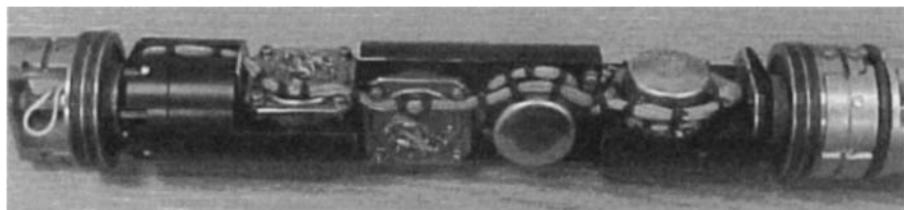


Figure 15.2 Well-bore survey inertial measurement unit

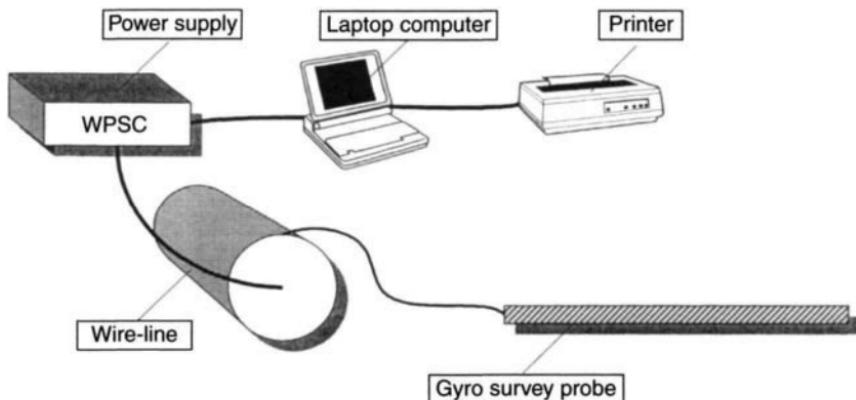


Figure 15.3 Cable-operated well-bore survey system

the surface and the down-hole system is maintained throughout a run allowing depth measurement data to be transmitted to the probe for inertial aiding purposes and survey data to be sent to the surface for storage and display.

More generally, systems of this type are finding application in both bore-hole survey and so-called measurement whilst drilling (MWD) systems for the oil, gas, mining and construction industries. The requirements placed upon such systems are often very demanding both in terms of the measurement accuracy needed, as well as the environmental requirements imposed.

15.2.4 System design requirements

Survey accuracies as small as 1 m per 1000 m of measured depth, or better, are frequently required to allow the well to be positioned to sufficient accuracy with respect to the smaller underground reservoirs from which it is required to extract oil and gas. The achievement of this level of performance calls for azimuth and inclination alignment accuracies better than 0.1° and consequent gyroscope and accelerometer residual bias levels of $\sim 0.01^\circ/\text{h}$ and $\sim 1 \text{ milli-}g$, respectively. In order to maintain the required accuracy throughout a survey, it is customary to combine the IN system data with an independent measurement of well depth in order to bound the growth of inertial system errors and so achieve an integrated system of the required performance.

Achieving the required level of gyroscope performance becomes difficult under the more extreme environmental conditions to which such systems can be exposed. Vibratory motion can induce additional measurement errors; as discussed in Section 12.6 where sensor motion dependence issues are addressed. For example, the effects of rotation of the angular rate vector, which the system is attempting to measure (coning motion), or vibration leading to apparent coning motion (pseudo-coning), need to be examined in systems that are required to meet such exacting performance standards in the presence of cyclic motion. In addition, careful consideration needs to be given to the mounting scheme and configuration of the sensors in order to ensure their survivability under drilling conditions.

Such systems may be required to operate at temperatures of 200°C or greater and at pressures up to 25 000 psi (172 MPa). In addition, the vibration and shock levels that an MWD system must be able to withstand are frequently extreme, whilst the turn rates to which a system may be exposed may be as high as 300 rpm (or 1800°/s). Further design constraints are imposed by the fact that the available power is severely restricted in down-hole systems. Last, but not least, the system must fit into a very small diameter pressure housing, typically less than 3 or 4 cm internal diameter, to facilitate its passage through narrow boreholes and/or tubes.

15.2.5 System design issues

Critical factors that have a major influence on the design of such systems are summarised below:

- survey performance
 - typically 1 m per 1000 m of measured hole depth;
- physical size constraints
 - diameter ~ few centimetres,
 - length constraints – less critical;
- dynamic range of tool motion
 - rotation rate up to 300°/s,
 - dynamic range up to 10^8 ;
- vibration-induced motion whilst running in a bore hole
 - qualification level for MWD tools: 20g RMS (5–500 Hz),
 - substantially less for systems operating on wire-line;
- operating temperature range
 - zero to 200°C;
- limited power available down hole;
- communications limitations between surface and down-hole equipment.

Aspects of the design influenced by these factors are discussed in the following sections.

15.2.5.1 Selection of inertial sensors

The major factors influencing the choice of gyroscopes and accelerometers are the performance characteristics imposed by the system requirement and physical size constraints. Typically, the diameter into which the inertial sensor block is required to fit within the pressure case is less than 4 cm.

In general, a system that is required to measure position to an accuracy of 1 m per 1000 m of along hole depth, must be capable of aligning in azimuth to an accuracy of 1 mrad. A gyroscope having a residual bias of 0.01°/h is capable of

Table 15.1 Gyroscope performance requirement (post-calibration)

Bias repeatability	0.01°/h
g-Dependent bias	0.01°/h/g
Scale-factor error	100 ppm
Random walk	0.006°/ \sqrt{h}
Angular rate limit	$\geq 200^\circ/s$

gyroscope-compassing to this accuracy at Earth latitudes up to 48°. For oil exploration and production at higher latitudes, in the north Sea and in Alaska for example, this would suggest the need to use even higher-grade gyroscopes. However, some relaxation of this performance figure can be accepted in practice, since well paths are not horizontal over their full length.

Typical gyroscopic sensor performance requirements, which are considered adequate for the application described here, are given in Table 15.1. The figures given represent the residual errors that can be accepted following system calibration.

There is a range of modern gyroscopic sensors capable of meeting the levels of performance defined here; ring laser gyroscopes, fibre optic gyroscopes and conventional spinning rotor sensors such as the dynamically tuned gyroscope. However, it is only the latter type which can be accommodated within the limited space envelope. This situation could well change in the not too distant future as a result of developments in newer technologies; advances in MEMS gyroscope performance, for example, considered in Chapter 7.

Accelerometers with bias repeatability in the region of 100 micro-g are needed to meet the system performance objectives defined here. Quartz flexure sensors of the required size and performance are available, and which are sufficiently rugged to survive and operate under the environmental conditions to which they will be subjected in this role.

For well-bore survey applications, account must also be taken of the reliability and cost of the chosen sensor, as well as the complexity of the sensor or sensor system. For example, in some applications of this type, the inertial sensors may be platform mounted² to facilitate rotational motion of the sensors [1]. As a consequence of this motion, a measure of on-line calibration of the inertial sensors can be accomplished as part of the operational procedure.

15.2.5.2 Sensor re-balance loops

For precision survey systems, the sensor re-balance loops are a critical part of the inertial system electronic support unit. Careful design is necessary to meet system performance objectives, particularly to satisfy the requirements for high resolution of

² A single axis platform is frequently used to give rotational freedom in a direction coincident with the longitudinal axis of the survey tool.

the angular measurements, and to achieve accurate scale-factor symmetry in respect of positive and negative rotations of the survey tool.

For systems using angular momentum gyroscopes, the use of analogue re-balance loops helps to minimise the power consumed by the probe during operation. Moreover, analogue systems have the potential opportunity to give greater resolution.

15.2.5.3 Quantiser design

The gyroscope quantiser circuits, along with the gyroscope re-balance loops, are amongst the most sensitive parts of the system electronics in precision well-bore survey systems. The dynamic range requirement is frequently very large $\sim 10^8$; arising because the probe may be subjected to turn rates as high as $300\text{--}400^\circ/\text{s}$ during surface handling, whilst needing to resolve rates as low as $0.01^\circ/\text{h}$ in order to satisfy the most exacting survey performance requirements. The realisation of this capability over the extremes of the expected operating temperature ranges necessitates the use of precision (small tolerance) electrical components in these circuits.

15.2.5.4 Down-hole processing

Ideally, all signal processing associated with the IN system and the inertial aiding process is implemented down-hole within the measurement probe. This requirement arises, in part, from an effort to minimise the quantity of data that need to be transmitted between the down-hole and surface modules of the system. The transmission rates that can be achieved in current down-hole communication systems are fairly modest, although research into methods of overcoming these limitations is underway at the present time.

The down-hole processing tasks include:

- on-line compensation of the inertial sensor outputs using stored calibration data, including temperature compensation;
- INS processing, that is, attitude computation, axis transformation of specific force data, and the solution of the navigation equation to yield velocity and position data;
- implementation of the Kalman filtering process incorporating incremental depth measurements to bound the growth of IN system measurement errors.

In addition to the above processing tasks, it is of vital importance in these types of applications to monitor the 'health' of the tool³ continuously during a run with a view to maintaining confidence that the probe is providing meaningful survey data throughout a survey. This information, in the form of a 'quality' indicator, must be transmitted to the surface at regular intervals for display to the operator. The quality indicator may be based upon the comparison between the actual measurement data and the *a priori* expected standard measurement data.

³ 'Health' of the tool: the quality and consistency of the data generated by the survey tool are monitored throughout its use in order to check, as far as is reasonable, that the tool is operating satisfactorily.

15.2.5.5 Compensation for vibration-induced errors

The inertial sensors will be subject to significant levels of shock and vibration whilst operating in the down-hole environment. In order that the system can be capable of surviving and operating under such conditions, careful attention must be given to the mounting of the sensors and electronics in the down-hole assembly. The design of suitable shock/anti-vibration mounts is a critical part of the design of such systems, and one, which occupies much of the designer's attention when developing systems for this role.

Whilst the vibration environment is clearly most severe when attempting to survey during the drilling process (MWD operations), significant levels of oscillatory motion can also arise in cable operated systems depending on how the sensors are supported in the well. For example, pseudo-coning motion may be present as a result of the gyroscope's sensitivity to cyclic angular motion; see Section 12.6.2.

For a dual-axis gyroscope subject to cyclic motion about a single input axis, a pseudo-coning rectification error is induced having a time-invariant component ($\Delta\omega$) given by the following equation:

$$\Delta\omega = \frac{-A\Omega^2}{2H}$$

where Ω is the amplitude of applied cyclic angular rate, H is the angular momentum of gyroscope rotor and A is the moment of inertia of gyroscope rotor about input axis.

An angular acceleration correction term may be calculated as a function of these gyroscopic parameters enabling a software correction to be applied.

15.2.5.6 Thermal design issues

Critical to the achievement of the required inertial sensor measurement accuracy is the temperature compensation of the output signals from each gyroscope and accelerometer. To this end, it is often necessary to install temperature sensors on, or as close as possible to, each inertial sensor. The resulting measurements, combined with temperature coefficient data stored in the probe, facilitate the on-line temperature compensation of the inertial measurements throughout a survey. The calibration process by which the temperature coefficient data are obtained is outlined later.

In order to achieve the required temperature operating range using inertial sensors with limited temperature capability, it is common practice in such applications to run the survey probe within a heat shield coupled to a heat sink. Depending on the capacity of the heat sink, this approach can allow extended operating times at high temperature.

15.2.5.7 Power requirements and management

In the development of systems for a well-bore surveying application, it is vital to minimise the power consumed by the down-hole system and to avoid excessive transient power demands. This is necessary because of down-hole power limitations, as well as a desire to avoid excessive heat generation in the down-hole equipment. Hence, highly efficient systems and techniques are vital, moreover, to cope with the power

surge needed to run up the gyroscopes; the gyroscopes may be run up sequentially following tool switch-on. As mentioned earlier, analogue re-balance loops may be used to minimise the power needed to operate the gyroscopes in the high-accuracy closed-loop mode.

15.2.5.8 IN system aiding techniques

An IN system provides very good short-term accuracy, but its uncompensated long-term performance does not fulfil the survey precision needs of the oil, gas, and construction industries, where run times varying from many hours to weeks may be required. Therefore, for each application, aiding techniques combined with statistical filtering methods are required to bound, or restrict the growth of, the time-dependent inertial sensor errors.

For systems operating underground, the options for IN system aiding are somewhat limited. One possible method for aiding a probe-mounted IN system involves stopping the probe periodically during its descent/ascent in the borehole. Whilst the probe is stationary, any components of velocity indicated by the IN system are clearly attributable to IN system errors and can be used to update that system. The Kalman filtering process implemented in such applications is often referred to as 'zero velocity updating'. As described in Chapter 13, this technique may be used for correcting the IN system velocity estimates; additionally estimates of attitude errors and instrument biases, which propagate as velocity errors within the IN system, may also be generated. The technique allows the appropriate corrections to be applied as part of the filtering process.

This method of aiding is a natural choice for systems designed to generate survey data as part of the bore-hole drilling process; the MWD systems referred to earlier. During the construction of a well, drilling ceases from time to time to allow additional drill pipe(s) to be added before a further section of well can be constructed. At this time, it can usually be assumed that the survey probe is stationary. Through the same process, it is possible to obtain improved estimates of gyroscopic errors and so update the gyroscope calibration process during system operation.

Under dynamic conditions, a measurement of tool motion along the well bore is required; that is, a measurement of position or velocity along the path of the well bore. Whilst there are various methods of generating this information [2], a commonly used technique is to measure the incremental changes in distance moved along the well. For wire-line or cable operated survey systems, proprietary depth measurement systems are available to provide this information. Precision systems of this type allow corrections for cable temperature and tension to be applied as part of the depth measurement process. Knowledge that the IN system cannot move outside the confines of the well bore is implicit within the Kalman filter measurement processes adopted in such applications.

15.2.6 System calibration and test

Vital to the successful operation of any IN system are the calibration processes undertaken during manufacture, and, subsequently, carried out on-site prior to running the survey tool. An on-site calibration is required in order to check for any unexpected

changes in inertial sensor performance prior to the start of the survey process, and to remove any switch-on to switch-on bias variation. This procedure helps to ensure that the best possible performance is extracted from the inertial sensors during the subsequent survey operation.

Detailed calibration and characterisation is carried out following manufacture, using a calibration and test procedure. This process involves a series of multi-positional and rate-transfer tests in order to characterise the system in terms of fixed biases, *g*-dependent biases, scale-factor linearity, mounting misalignment, and cross-coupling errors. The probe is installed in a temperature chamber whilst on the test stand. This arrangement allows data to be collected as the probe is cycled over the specified operational temperature range. As part of this process, it is often necessary to consider the direction of any temperature change in order to take account of temperature hysteresis effects.

In order to extract the best possible performance from the gyroscopes, a further process of calibration may be implemented on-site prior to each run. The objective of the on-site calibration is to estimate and compensate for switch-on to switch-on bias variations. This procedure takes the form of a multi-positional test with the probe nominally horizontal. Similar techniques are described in Chapter 8, in relation to the laboratory testing of inertial sensors and systems. These tests are a vital step in the process for qualifying the performance of the system to ensure that the sensors and the system are suitable as well as being optimised for this survey role.

All survey tools undergo extensive calibration, laboratory testing and environmental qualification prior to operation in the field. Further testing of development tools may be carried out using van trials and other surface testing (funicular railway tests for example) to emulate, as accurately as possible, the movement of a tool in an underground well.

In-hole testing is carried out in known test wells, which have been surveyed many times in the past, for which the well path is believed to be reasonably well defined. However, as mentioned at the start of this section, the only point in the well path that is truly known is the surface position, or the co-ordinates of the well head. This being the case, well tests involving surveys conducted whilst both descending and ascending in the well (usually referred to as in-run and out-run surveys) are undertaken, any shift in the measured position on returning to the surface, with respect to the true surface location, providing a good indicator of survey system performance. Additional tests are carried out to check for survey repeatability over a number of repeat runs, consistent results giving further confidence that the survey tool is operating satisfactorily.

15.2.7 Concluding remarks

Robust and reliable survey systems based on inertial sensor technology can be produced, and a number of sophisticated techniques have been developed to ensure that accurate and reliable operation can be achieved in a hostile environment.

It is the combination of careful calibration and characterisation of the IN system during manufacture, on-site calibration, and the on-going aiding process

implemented during operations that are key features for the successful operation of a precision down-hole system. This process is designed to yield the best possible system performance, whilst constantly checking for errors throughout survey operations; errors which may render the survey of insufficient quality, or invalid.

Continuing research developments in this field are directed towards the design of gyroscope based survey tools capable of delivering precision survey data whilst drilling (measurement whilst drilling systems). The ultimate aim is to produce directional steering tools, which control the path of a well-bore directly throughout the drilling process.

15.3 Ship's inertial navigation systems (SINS)

One of the most demanding applications of inertial systems for achieving very high accuracy is the navigation of naval ships and submarines, which may be required to operate covertly for long periods of time in any part of the ocean, including the polar regions. Early work in the United Kingdom at the Admiralty Compass Observatory (ACO), Slough (formerly part of the Defence Evaluation and Research Agency), led to the development of very stable platform systems: these devices provided an autonomous navigation capability in the 1 nautical mile per day class, incorporating gyroscopes with residual biases of less than $0.001^\circ/\text{h}$. The single-axis rate-integrating gyros used for these applications were highly sophisticated instruments with gas bearings, which needed to operate in a closely controlled temperature environment to achieve the level of performance required for highly accurate navigation.

A British programme to develop a ship's inertial navigation system at the ACO dates from the mid-1950s. The initial system design incorporated rate-integrating gyroscopes developed by the Charles Stark Draper Laboratory in the United States, where the rotor was supported by high-precision ball bearings. This project led to laboratory tests and initial sea trials in 1961. The system became fully operational in 1968. During this time, ACO played a major part in the development of a new rate-integrating gyroscope for SINS in which the rotor was supported by a gas bearing, which resulted in an order of magnitude improvement in performance. SINS Mk. 1 was a completely analogue design, sensitive to temperature and power supply variations and had significant limitations when operating at high latitudes. Its successor, SINS Mk. 2 featured gyroscope pulse torquing, integrated circuits and digital computation and resulted in production equipment with full polar capability in 1976.

The early systems provided very accurate IN performance, but were both large and expensive to manufacture and maintain, in fact the system was so massive that it required a special access hatch to get it into and out of a ship; clearly this limited the number of ship classes that could accommodate this class of technology. A number of efforts were undertaken to devise a smaller but equally accurate device, but this did not progress significantly until the 1980s. The process of developing a smaller, lighter and less costly replacement system, with potential application in a broader range of ships, began in 1983, initially as the ship's low cost inertial navigation system

(SLINS) development, and subsequently as an international collaborative venture culminating in NATO SINS [3]. This is a strapdown system based upon ring laser gyroscope technology.

15.3.1 NATO SINS

NATO SINS incorporates three aircraft-grade ring laser gyroscopes and three high-precision pendulous force-feedback accelerometers. The inertial instruments are mounted in a sensor block assembly, with individual temperature sensors to facilitate on-line real-time temperature compensation of the gyroscopes and accelerometers. The system accepts position-fix updates (when available) provided by a shipboard GPS receiver or dockside information and/or ship's log velocity data, which are processed via a Kalman filter to aid the system and so bound the growth of errors that would otherwise propagate. The key elements of the design are shown in Figure 15.4.

Apart from the switch to strapdown technology, this system differed from the earlier designs in its use of a proprietary 'indexing' technique. This scheme was originally designed for use in aircraft IN systems, prior to the development of high performance strapdown gyroscopes, to average out drifts in all directions by rotating the complete inertial measurement unit (IMU) through angles of ± 90 or $\pm 180^\circ$ about its azimuth and roll axes every few minutes. This is achieved by mounting the sensor block assembly in two gimbals, an azimuth (inner) gimbal and a roll (outer) gimbal, as indicated in schematic form in Figure 15.5.

Normally the inertial sensors used in an airborne strapdown inertial navigation system would yield navigation system performance in the 1 nautical mile/h class.

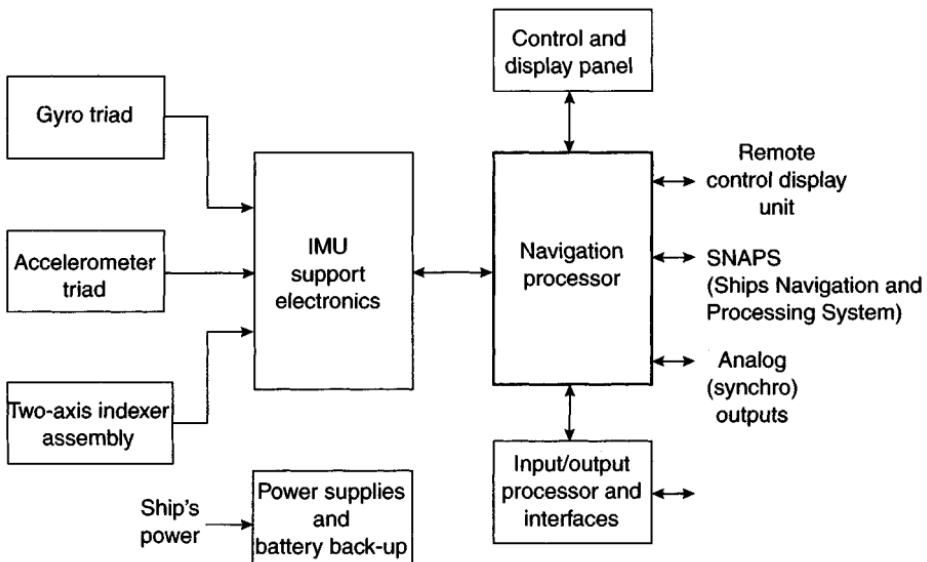


Figure 15.4 SINS functional block diagram

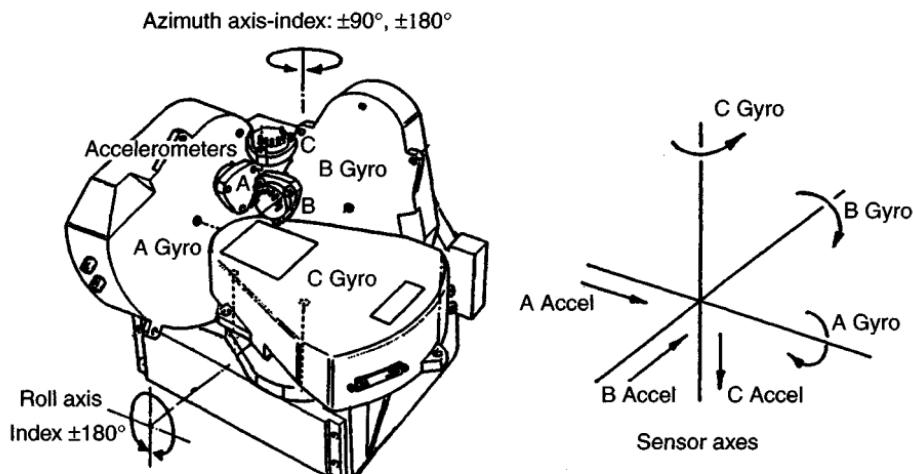


Figure 15.5 Gimbal mounted sensor block

However, NATO SINS may achieve navigation performance in the 1 nautical mile/day class through the use of an indexing mechanism and selecting gyroscopes with a low random bias. System analysis indicates typically random walk error of less than $0.00125^\circ/\sqrt{h}$ is required to achieve the 1 nautical mile/day performance.

The index cycle is designed to average out the effect of gyroscopic drifts in all directions, the full sequence of moves taking of the order of 3 h to complete. For example, a bias in the C gyroscope measurement will cause the computed azimuth to drift to the left when upright, and to the right when the instrument cluster is rolled through 180° .

The indexing mechanism is also used to isolate the system from the effects of ship motion. This allows the effect of a ship's turn, which may be such as to negate the averaging effect of a gimbal index rotation, to be removed. This is achieved by using the computed attitude data (embodied in the direction cosine matrix) to drive the azimuth gimbal so as to maintain the sensor block assembly in a fixed orientation with respect to the local geographic reference frame.

Finally, the indexing procedure is used to facilitate the calibration of the individual sensors as part of a dock-side alignment procedure. During dock-side alignment, the system gyrocompasses to establish direction and uses the roll and azimuth motions available through the indexing mechanism to put the three gyroscopes into controlled test positions. It is then possible to compare the gyroscopic measurements of rotation rate with the known Earth's rate components, which are expected to be sensed by the gyroscopes, and so form estimates of biases, scale-factor errors and misalignments. The Kalman filter, discussed briefly below, is the key to the successful implementation of this process.

Having removed the effects of gyroscopic biases by the indexing procedure, system performance accuracy is largely determined by the random walk error in

the gyroscopes; this is the principal error that remains, and unfortunately cannot be removed through indexing as it is not a systematic error process.

The major functions of the Kalman filter in NATO SINS are as follows:

- to monitor system parameters, gyroscope and accelerometer biases, scale-factor errors and misalignments, measured during dock-side calibration;
- to carry out system resets at sea, based upon navigational updates, the amount by which the system is reset being controlled by the accuracy of the fix;
- to control ship's log damping, by removing the log input when the discrepancy between log and system exceeds a pre-set value in either north or east directions.

The Kalman filter forms estimates of the errors in position, velocity and attitude, as well as inertial sensor biases, scale-factor errors and misalignments. On receipt of a position (latitude/longitude) fix, this information is compared with the latitude and longitude estimates generated by the inertial system, the differences forming inputs to the Kalman filtering process. Tests are applied to assess the 'credibility' of the external fix data, and manual intervention by the operator is allowed to review and assess the data prior to validating the fix before it is applied to the filter. Following acceptance of the fix, the Kalman estimates of system error states are used to 'reset' the system.

Successful implementation of an inertial navigation system that compensates for the inherent errors in inertial sensors has provided ships with precision navigation data anywhere in the world; with positional accuracy $\sim 1 \text{ nm/day}$ error. The system demonstrated that systematic errors could be compensated leading to greater performance, whilst offering major reductions in size, weight and both acquisition and life cycle costs.

15.4 Vehicle stabilisation and control

This section examines some current applications in which inertial sensors and systems are used to maintain control of land, sea and airborne vehicles during normal operations and throughout manoeuvres. These range from autopilots to safety system, as well as fault diagnostic approaches.

15.4.1 Autopilots

Autopilots have a number of functions, as discussed below, but the fundamental objective of an autopilot is to modify the intrinsic natural behaviour of a vehicle to produce the desired response characteristic. This involves making the behaviour of the vehicle more predictable and less influenced by external disturbances and variations in component characteristics. Moreover, the ideal design aim is to make the response of the vehicle constant over its entire operating envelope. This section will consider two common applications; autopilots in guided weapons and in aircraft.

An autopilot is a closed loop system used to stabilise the chosen or demanded flight path of a vehicle. In the case of an aircraft or missile it maintains a response and track that is reasonably resistant to external disturbances, but is still responsive to control demands.

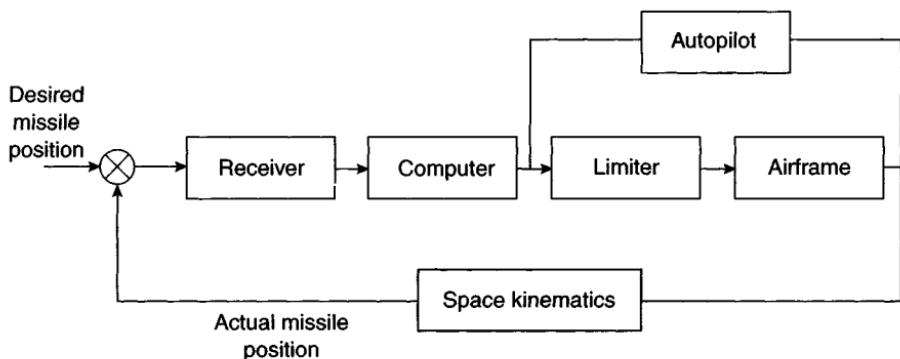


Figure 15.6 Missile guidance loop

15.4.1.1 Conventional missile systems

The fundamental requirement for a guided weapon is for the projectile to hit its target. This is achieved by generating guidance commands, derived from the perceived engagement conditions, usually in the form of lateral acceleration (latax) commands.

At a given speed, a latax command may be thought of as a rate of turn, or curvature of the trajectory. The purpose of the autopilot is mainly to ensure that the guidance command is achieved in the minimum time in the presence of potentially rapidly changing system and environmental conditions.

A missile autopilot translates the input commands of lateral acceleration from the guidance system into control-surface deflections to achieve the desired lateral acceleration or angular motion. When an error is detected between the actual position of a missile and its desired position, as determined by the guidance system, a demand is generated to move the missile. This is a continuous process and is applied until the missile 'is on track'. A simplified diagram of this form of control loop is shown in Figure 15.6.

The generation of aerodynamic lift is the commonest means of generating latax through rotation of the vehicle to create an incidence angle between the airstream and the velocity vector of the vehicle. Changing the roll angle of the vehicle in a bank-to-turn system controls the direction of this lateral force generation.⁴ The lift on aerodynamic control surfaces is largely governed by the same quantities as govern the overall vehicle lift, with the additional variable of the deflection of a control surface. By careful design, the steady-state lateral acceleration can be made a monotonic function of control surface deflection; a given control deflection will result in a particular value of equilibrium incidence being reached. This is the basis of an acceleration-demand control system.

⁴ An alternative steering system is Cartesian control, or skid to turn, by inducing a lateral force in the direction of turn independent of the roll condition of the vehicle.

The aerodynamic design of a missile is a compromise between stability, which determines the response time to an input demand, and the available control power. Control power dictates the maximum steady state latax that may be achieved. However, if both rapid response and high latax are required, they will not be achievable through aerodynamic design alone.

Even when a reasonable aerodynamic compromise between the equilibrium characteristics of static stability and maximum trim incidence is achieved, the resulting dynamic response may not necessarily be satisfactory. Missile configurations typically have poor aerodynamic damping, which unless augmented with rollerons (see Section 15.4.2), results in an undesirable oscillation about the trim condition. Motion about the trim condition is known as the weathercock mode, after the tendency of a weathercock (or weather vane) to align itself with the wind direction. In some missile designs, the guidance command is passed through a notch filter to avoid excitation of this mode, but internal filters do not prevent its excitation by external disturbances, such as wind gusts.

A missile may gain details of a target's motion, with respect to a given reference frame, from the output generated by a seeker, or from a command receiver collecting information from external systems. This information is processed by a computer to generate the lateral force commands required to manoeuvre the missile, and then applied by the autopilot. The use of a limiter is important to ensure that the applied manoeuvre demands do not destroy the missile, either by generating forces beyond its structural capability or by causing it to reach an angle-of-attack (incidence angle) that causes the missile to become unstable. A range of instruments may be used to monitor these effects and, in combination with the autopilot, ensure that missile has a wide operational envelope; this is the form of the system required to achieve the fundamental goal of an autopilot.

A simplified block diagram of a missile autopilot is shown in Figure 15.7. It shows the use of rate gyroscopes and linear accelerometers in the feedback loops of this system. The method of use of these sensors is dependent on the missile application and the way the autopilot is used, that is, the type of autopilot. Autopilots come in several forms [4], as shown in Figure 15.8, but may be considered to be either:

- a lateral autopilot for controlling the pitch or yaw motion of a missile;
- a roll autopilot for motion control about the fore and aft (longitudinal) axis of a weapon.

A lateral autopilot used in a command-guided or homing missile, often uses an accelerometer as the main feedback sensor and a rate gyroscope provides the damping function. If the missile has two planes of symmetry, then two autopilots would be required to provide the control of motion in the orthogonal pitch and yaw planes.

The simplest application of inertial sensors is to compensate for the poor natural damping by the use of rate gyroscopes to provide measurements of a missile's body angular rate, which is used as a feedback signal. Additionally, as the primary function of the autopilot is to control lateral acceleration, it is sensible to measure this state directly, rather than using an indirect derivation. This is the basis of a two-loop autopilot.

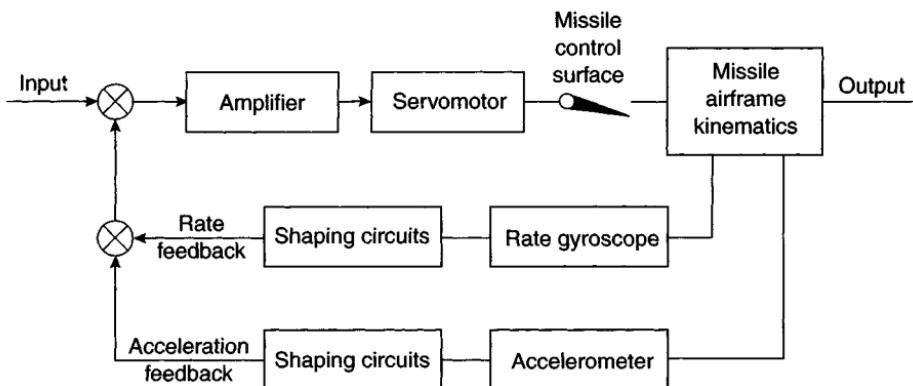


Figure 15.7 Autopilot control loop

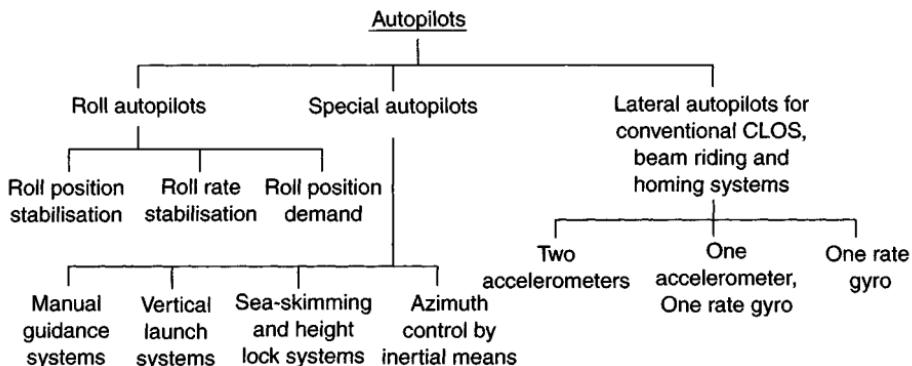


Figure 15.8 Autopilot structures

The requirement for near neutral stability of the airframe over the entire flight envelope is usually very difficult to achieve, particularly when there is a significant shift in the centre of gravity position as fuel is consumed by the rocket motor during flight. Therefore, gains in the control system may be scheduled to vary over the flight envelope to ensure substantially constant bandwidth and damping, and insensitivity to disturbances and system variability. In short, the feedback controller simplifies the missile behaviour, making it more predictable and responsive than the basic airframe.

The accelerometers in this system are usually placed well ahead of the centre of gravity of the vehicle, with their input axes orthogonal to the longitudinal axis of the missile. The recommended distance is in the region of one half to two thirds of the distance of the centre of gravity to the nose. Care is required to ensure that the sensors are not positioned at an anti-node of the principal bending mode of the missile; otherwise there can be detection of significant vibratory motion by the sensors, leading to destruction of the missile from 'false' commands'.

The favoured position for the rate gyroscopes is at an anti-node⁵ where the angular motion, owing to vibratory motion is a minimum, although this is not the optimum position from the overall design point of view. The input axes of the rate sensors are arranged to sense pitch and yaw motion, so they are orthogonal to the missile's longitudinal axis. Consequently, the sensors' output signals are proportional to the pitch and yaw rates. The accuracy of the sensor, such as its drift rate, is not an important requirement for the performance of the gyroscopes, as they are only required to provide a measurement of angular rate, particularly with low-noise characteristics. Similarly, the accelerometers are not required to give measurements to high accuracy with great stability over long periods for this function.

The closed loop, self-correcting behaviour characteristic of the guidance loop reduces the inertial instrument performance requirements compared with those of typical inertial navigation applications. Hence, the inertial sensors in an IMU used in a navigation system may be used to provide the 'feedback' data for the autopilot, provided the inertial sensors are appropriately positioned and have suitable bandwidth responses. This is developed further in the following section.

15.4.1.2 Vertical launch missile systems

During the launch phase, of this class of weapon, guidance information may not be available and inertial measurements must be used to define the missile state vector. In such circumstances, the guidance commands are generated from inertial measurements. For example, during the initial launch manoeuvre from a vertical canister, or silo, the missile state has to be determined from inertial measurements, with a desired end state programmed into the missile's control system during the launch sequence. Initially, the speed is low and less impulse is required to achieve a given heading change than later in flight. Unfortunately, the available aerodynamic lift is also low, so the lateral acceleration autopilot designs described earlier are unlikely to prove capable of exploiting this phenomenon.

During the launch phase, the greatest force acting on a vertically launched missile is typically the boost-motor thrust, and this must be rotated with respect to the velocity vector to generate the required lateral acceleration. Rotating the missile body requires a means of applying moments. As the available aerodynamic moments are also small, provision is typically made to deflect the motor efflux laterally in order to apply controlling moments to the body of the missile and consequently steer the missile. In this phase of flight, maximum lateral acceleration is achieved when the thrust acts perpendicular to the trajectory, that is, with the missile apparently flying sideways!

It is usual to use an attitude-control autopilot during this phase of flight by using a velocity-vector steering technique. The autopilot loop is closed by setting the deflection of the motor's jet to be proportional to the difference between the actual angular body rate, again measured by the angular rate sensors in the IMU, and the demanded body-rate command.

⁵ In a rigid missile, gyroscope and accelerometers are placed at a node and this has very little adverse impact on the performance of the gyroscopes.

If there are no constraints on the apogee height, then a simple attitude autopilot may be adequate. However, in some circumstances the open-loop control of the trajectory inherent in attitude control is unlikely to be effective, so some form of closed-loop control of the trajectory is required, such as velocity-vector control.

As part of its normal operation, the IMU will produce an estimate of the missile velocity vector with respect to the chosen inertial reference frame. The control system will generate a velocity direction demand. The objective of the steering loop is to generate a command signal to be fed into an attitude loop of the form described earlier. In this case the proper function of the attitude loop is to produce a lateral acceleration demand, which is proportional to the velocity error, in order to reduce that error to zero

15.4.1.3 Aircraft

Commercial and military aircraft have made significant advances in their operational characteristics over the last few decades, particularly with respect to operational range and control complexity. Furthermore, the development of IN technology has had a number of significant impacts on aircraft operations, for example the successful development of the IN system has made the traditional role of the navigator redundant. Development of fly-by-wire techniques and advanced avionics systems has made aircraft far more agile. As a consequence there has been an increase in the difficulty in manual control of some combat aircraft, particularly those with so-called negative stability, where the centre-of-gravity is ahead of the centre-of-lift, giving a negative static margin.

Automatic control equipment has been developed to minimise pilot fatigue and reduce the more tedious aspects of flying, so that the pilot can concentrate on the major operational decisions. The so-called automatic pilot's function is to stabilise the flight path by maintaining the aircraft at a constant height and in the commanded direction during cruise phases of flight. During manoeuvres to a new altitude or bearing, the autopilot generates the commands to ensure the transition is smooth and leads to a stable flight path. There are some differences between the missile and aircraft autopilots as aircraft use a combination of roll motion and yaw deflections to turn, in a so-called 'bank-to-turn' manoeuvre. In contrast, most tactical missiles use Cartesian control in a 'skid-to-turn' type of manoeuvre.

Gyroscopes are at the heart of this type of autopilot used in 'cruise control', as shown in Figure 15.9. In this case the various channels of the autopilot are shown. The pitch channel gyroscopic sensors detect motion of the aircraft about its horizontal or pitch axis and provide command signals to the elevator control surfaces in the tail plane to counter the sensed motion. This channel can also respond to signals generated by a pre-set altimeter, such as a barometric device or a radar-based sensor, to ensure the designated altitude is maintained.

The yaw and roll channels respond to motion sensed by the appropriate gyroscopic sensors, in a simplified system the yaw channel motion may be sensed by a gyro-stabilised magnetic compass.

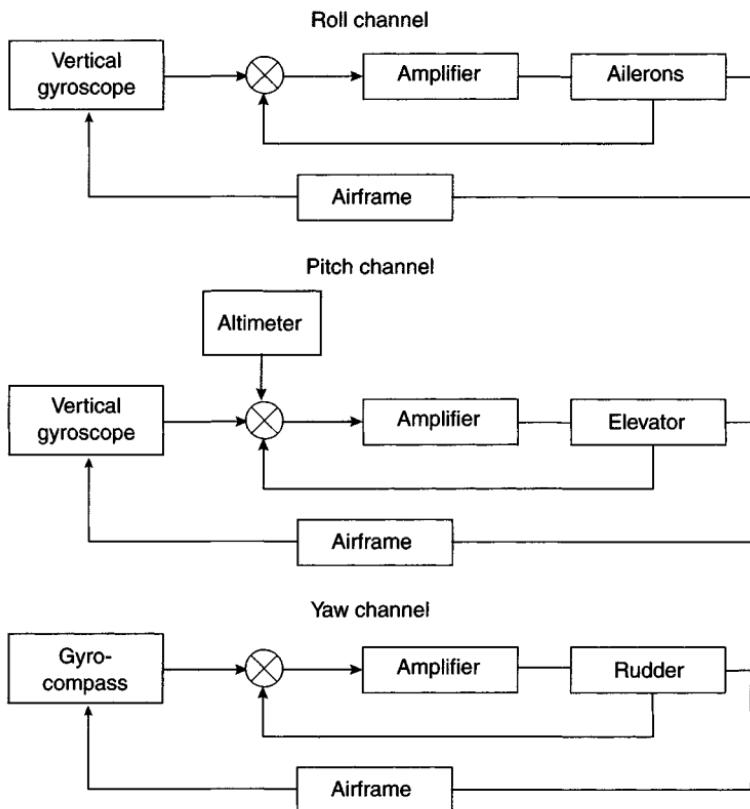


Figure 15.9 Aircraft autopilot

15.4.2 Passive missile roll control (rollerons)

One of the problems frequently encountered in the design of tactical missiles is the provision of adequate roll damping. This problem primarily arises as a consequence of the predominance of low-aspect-ratio 'lifting' surfaces, such as wings and fins, used in most missile configurations. Consequently, the inherent natural damping moment about the longitudinal axis of the missile body is small in comparison with the forces generating the roll motion. Moreover, in some of the simpler missiles there is only a requirement to contain the roll-motion dynamics within certain limits, to allow resolution of autopilot commands, rather than control the roll motion for a roll-position or roll-rate stabilised system.

A number of solutions are possible. One example is a servo-mechanism device, which senses the roll rate via an angular rate sensor and actuates control-surface deflections to provide the necessary damping forces as applied in an autopilot. However, these servo-mechanisms are quite complex and demand space within the missile body and tend to lead to a decrease in the reliability of the overall missile system. This type of solution is not consistent with the low-cost and simple approach of many weapon systems, such as heat-seeking missiles.

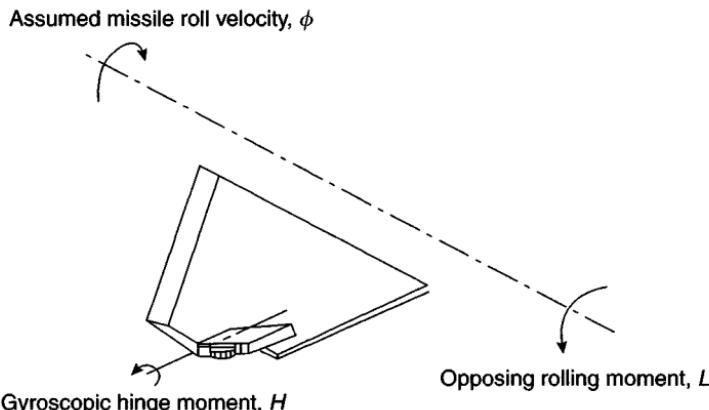


Figure 15.10 Aileron control

A simple technique to provide roll damping may be achieved by the action of gyroscopically actuated, uncoupled wing-tip ailerons. This type of system does not require any components internal to the missile. Roll damping is achieved in this system by the independently acting, wing-tip ailerons with enclosed air stream impelled roll-rate-sensitive gyroscopic rotors. This type of roll damper is referred to as a rolleron with one device per wing.

This system consists of a small aileron control surface positioned close to the wing tip and hinged towards the leading edge of a fixed wing on a missile, as shown in Figure 15.10. Enclosed in the aileron is a gyroscopic rotor with its spin axis perpendicular to the plane of the wing in the undeflected aileron position. This rotor is activated by the movement of the air stream as it flows over the wing during the normal air-carriage of the missile, so the rotor is activated as the missile is conveyed to its launch point. During normal air carriage of the missile, the rolleron is clamped in an in-line or streamlined position, but is unlocked following launch of the missile as the rotor reaches a pre-set spin speed.

When the missile experiences a roll rate during free flight, the aileron (control surface) of this system will be subjected to a gyroscopic hinge moment. This gyroscopic hinge moment causes a deflection of the aileron, which, in turn, creates a rolling moment in a direction opposite to the initial direction of the sensed rolling velocity of the missile; this is indicated in Figure 15.10. As a result, a resistance to the rolling motion is produced from the deflected control surface and roll-motion damping of the missile occurs. This roll-motion damping force is greater than the inherent aerodynamic roll-damping moment generated by the missile configuration alone with its low-aspect-ratio wings and fins. The difference in the damping is the amount determined by the roll effectiveness of the complete set of rollerons.

Care has to be undertaken in the design of this device to avoid simultaneously introducing undesirable effects of the longitudinal motion of the missile and its roll stability. Additionally, care is required in the design of the pivot of the rolleron, in order to ensure it has the appropriate damped motion as it responds to the gyroscopic

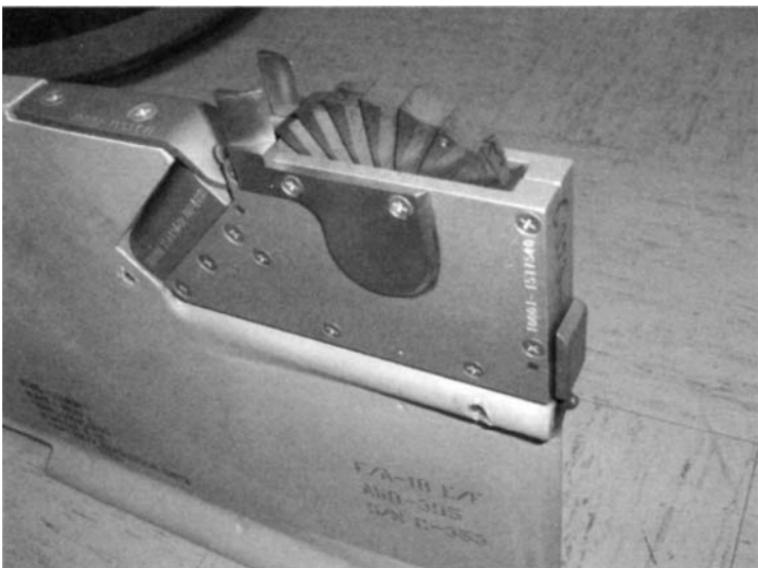


Figure 15.11 Sidewinder missile wing with rollerons

forces, again to avoid introducing undesirable characteristics into the trajectory of the missile during free flight.

This type of gyroscopic device has been demonstrated to operate over a wide range of flight conditions [5], ranging from Mach numbers less than one to about Mach 3 at altitudes up to about 50 000 ft. Typical rotor speeds range from 8000 to 50 000 rpm. A photograph of the system on a Sidewinder air-to-air missile wing is shown in Figure 15.11.

15.4.3 Intelligent transport systems – automotive applications

Miniature inertial sensors are now used in a variety of vehicular applications; systems to monitor motion, and thus provide data for modifying the ride or to provide a logging function are becoming quite common. Whilst many vehicles are fitted with MEMS yaw-rate sensors, there is a move to incorporate a complete six-degrees-of-motion measurement to provide a central control system with vital data from a number of systems. This central control system can provide data to other systems via a central data highway for digital communication throughout a vehicle (known as a CAN-bus data highway). Examples include:

- adaptive cruise control and navigation;
- adaptive suspension;
- headlight steering;
- safety functions, including airbag actuation and control data for intelligent or advanced breaking systems.

Some of these are considered in the following sections.

This class of application demands low-cost sensors, such as MEMS devices. In general, the cost and the reliability are the key aspects of the demanding specification, as relatively low-performance rate sensors are adequate for the vast majority of these applications.

15.4.3.1 Navigation

The navigation systems used in vehicles are based broadly on satellite navigation techniques, such as GPS [6], but use of information from inertial navigation sensors may augment the accuracy of this basic system. The major limitation of the external reference system is the possible loss of signal from one, or more, of the satellites, which can occur for a number of reasons; common examples include:

- obscuration by terrain or foliage;
- obscuration by cultural objects in the landscape, such as tunnels, mountains or buildings.

During these periods of loss of satellite navigation signals, the inertial navigation system fills in the gaps, particularly if the loss of satellite data is intermittent; hence the inertial system acts in a similar fashion to a flywheel in an engine. Provided the database of the area being travelled is accurate, then the road system may be considered as a set of urban canyons (defined by the stored database), so the errors may be bounded and used in Kalman filters to compensate for many of the systematic errors.

In simpler car navigation systems, linking of the output from the odometer (or a speedometer and a clock) to an electronic map is an adequate reversionary mode for intermittent loss of the 'GPS' signal. The true value of a more sophisticated system is to limit the growth of navigation errors during the loss of the 'GPS' navigation data and thus help the receiver re-acquire the 'GPS' signal following a prolonged loss of signal.

It is important to distinguish between position and location: a positioning or navigation system will supply co-ordinates, but a location system requires a map. Car navigation systems usually have the ability to take an address and transform this into a co-ordinate set from a stored map and provide the driver with instructions on how to reach the destination. Map-matching techniques may be used in the simpler systems; in this case it is assumed that the car will be on a road, usually close to the GPS co-ordinates transposed on to a digital map. A good map-matching algorithm relies on maps with a high positional accuracy, usually of the order of 30 m, to avoid incorrect road selection. Dead reckoning techniques, described in Section 2.1, are also used to determine absolute position of the vehicle.

Owing to the short duration navigation requirements and the aiding techniques that are available, the requirements on the inertial sensor are modest. This is an ideal application for a low-cost MEMS device.

15.4.3.2 Suspension and braking systems

Active suspension or adaptive chassis systems [7] use either electrical or hydraulic actuation to stiffen the suspension of the vehicle and thereby control the dynamical

motion of a vehicle. This means better stability as a vehicle turns a corner by controlling the roll motion of a vehicle. The other feature of this type of system concerns the pitching motion of the vehicle, either smoothing out the oscillations as it travels along a bumpy road, or preventing the front of a vehicle from dipping under heavy braking action, thus ensuring the rear tyres maintain maximum adhesion to the road.

An intelligent braking system aims to provide an enhancement to the already widely used anti-skid braking system (ABS). In the case of an intelligent braking system, a single-axis accelerometer and gyroscope are mounted on the vehicle and are required to measure the yaw rate and lateral acceleration of the vehicle. The control system aims to keep an even distribution of the vehicle's mass by adjusting the damping of the shock absorbers in the suspension system.

A more sophisticated system may embody an electronic throttle control, a steering wheel position indicator, wheel speed sensors and a modified braking system. The vehicle's system 'knows' the safe turning radius or speed for a given wheel speed and steering wheel movement from stored data. If the inertial sensors detect motion beyond the designated safe parameters for a given speed, then the control system removes power and applies a suitable braking force independently to each wheel.

These techniques may be combined with an autopilot for navigation and guidance of robotic vehicles. These vehicles may then undertake hazardous activities such as mine clearance, cleaning up contaminated waste products or spraying pesticides.

15.4.3.3 Car monitoring

One research effort with cars involves precision location systems that act in a similar role to an aircraft's 'black box' accident recording system. This system would determine the attitude and position of a vehicle, that is, its trajectory on a road or with respect to any other suitable reference.

In this case, an integrated navigation system with a tuned Kalman filter is used. The inertial navigation system provides the attitude, position and turn rates of the vehicle at a high data rate. A 'GPS' receiver will provide position fixes every second. The data may be combined in a Kalman filter to provide highly accurate information about the speed, position and attitude of a vehicle over a given period immediately before an accident. The current projection is that such an integrated navigation system could provide position to an accuracy of about 5 cm!

Clearly, this is a direct application of a low-cost integrated inertial navigation system.

15.4.3.4 Safety

Other automotive safety projects are concerned with collision avoidance and prevention of vehicles wandering off roads unintentionally. Such object-sensing functions may be implemented using radar or electro-optical systems. Again, the complete safety system would involve an integrated inertial navigation system and a suitable stored-map database.

MEMS accelerometer sensors are used for activation of the explosive train in an airbag passenger protection system.

15.4.4 Intelligent transport systems – trains

Gyroscopes are used on modern high-speed locomotives and carriages to control the tilting as they travel at high speed around tight bends. This enables existing railway lines to be used safely by trains travelling at a higher speed than the designed safe operating speed for a conventional train.

Another use of inertial sensors that is the subject of on-going research involves the use of an inertial measurement unit on a locomotive or carriage to record the motion characteristics of ‘the ride’ throughout its travels. Periodic detailed analysis of the recorded data could be used for a range of analysis such as:

- condition of locomotive and carriage wheels, particularly the occurrence of ‘flat spots’;
- performance of the suspension;
- riding quality of the track.

15.4.5 Personal transport

One of the most common applications of inertial principles in transport is the use of the gyroscopic effect to keep a balance and control of any two-wheeled machine, such as a bicycle or a motor cycle. The angular momentum of the wheels provides the force required to keep the rider vertical when travelling in a straight line. Steering may be accomplished by moving the handlebars so that an orthogonal torque is created to move the rider in the desired direction. Alternatively, the rider may lean to one side to create a force out of the vertical plane, creating a turning moment and consequently disturbing the equilibrium for travel in a straight line.

However, more recently there have been significant developments in ‘two-wheel’ transportation, as described in the following section.

15.4.5.1 Segway and Ibot (human transporter)

A recent development in this aspect of transport is the Segway machine, invented by Dean Kamen, which has a stabilised platform where the rider stands. The application of gyroscopic techniques enables the rider to balance safely on two wheels positioned on either side of the machine, which may be driven backwards or forwards by an electrical motor. A smart wheel chair, known as the Ibot, has also been invented using the same principles of a stabilised platform to allow it to operate or balance on two wheels.

The application of this technology to a wheel chair makes it far more versatile than the conventional device. The self-balancing feature means that the wheel chair may be used safely on two wheels, (wheelie mode), which may be exploited to:

- make the wheel chair stand up on its rear wheels to lift the occupant to the same level as other people in a standing group;
- travel up or down stairs, using the wheels in pairs in ‘wheelie mode’ alternately to move up or down steps. The Ibot descends stairs in a fashion similar to a coiled spring moving from one level or step to another. The reverse process may also be accomplished with this type of wheelie mode. This is shown in Figure 15.12.

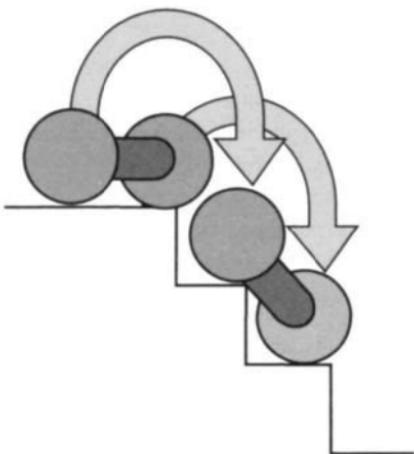


Figure 15.12 Scheme of a wheel chair moving over steps

At the heart of these systems is a so-called balance sensor assembly (BSA): the BSA has five single-axis silicon sensors and two dual-axis tilt sensors that provide augmentation data, as well as its own computer processor. The tilt sensors are effectively liquid potentiometers with four pins arranged around a central fifth pin in an electrolyte. As the stabilised platform tilts the inertial sensors provide angular rate data and the tilt sensors provide signals based on the change in resistance between the various pins.

The tilt sensors replace the function of accelerometers on stabilised platforms, described in Appendix C, as they undertake the 'spirit level' function and thus define the direction of the gravity vector, thereby defining the down direction of the reference system. The sensing characteristics of the two classes of sensors are complementary; the gyroscopes providing the high bandwidth data and the tilt sensors bounding the drift and noise on the output from the gyroscopes. Additionally, the gyroscopes provide a measure of angular rate, not absolute angular position, so the angular motion requires an independent absolute reference (vertical) so that the platform can be kept level. Thus, the stabilisation control-loop can be closed in an optimum fashion with long-term stability using low-cost sensors. Electrical motors provide the torque to null out any motion measured by the BSA and thereby keep the platform horizontal.

The five gyroscopes are arranged on the platform at various angles to provide a high degree of safety through sensor-measurement redundancy and to enable the system to move up and down hills. Only one of the sensors provides a direct measurement of the pitch angle of the platform, the others supply components of pitch, roll and yaw. In this case there are three independent measurements of pitch motion from the gyroscopes and two measurements of roll and yaw motion. A photograph of the BAS unit is shown in Figure 15.13, together with a MEMS rate sensor.

One gyroscopic sensor measures pitch motion directly, as noted above, and acts as the datum reference for the validity of the values of the derived measurements from two of the other sensors that give pitch motion data. As long as the two derived

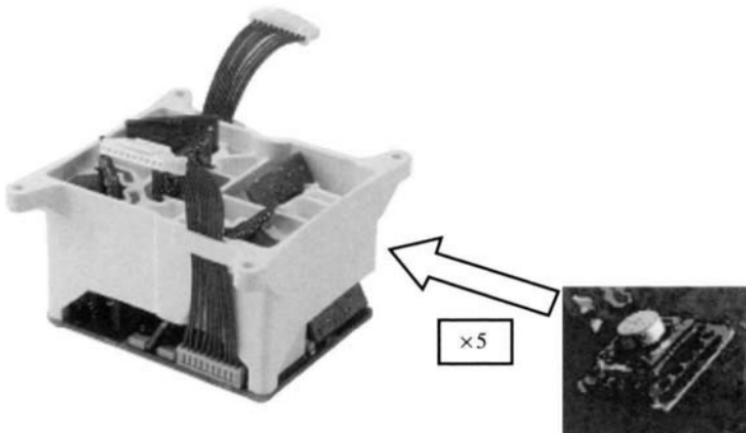


Figure 15.13 Photograph of the balance sensor assembly (Courtesy of BAE Systems)

measurements agree, the reference measurement has no direct role to play; once a discrepancy is detected the system shuts down automatically. Two of the other four gyroscopes are oriented to provide pitch and yaw data in such a format so that summing and differencing the signals will give the two estimates of pitch and yaw motion. The other two sensors provide pitch and roll data. Additional manipulation enables other estimates to be derived to provide redundant estimates of the motion.

Estimates of pitch and yaw are required to enable the ‘human transporter’ to make manoeuvres safely, such as turns whilst travelling up or down a hill, this avoids a cyclic pitch error over confusion in the system of the ‘location’ of the gravity vector. This is overcome by estimating the pitch motion with respect to the gravity vector in all three axes of the reference frame and compensating for the effects of erroneous estimates of pitch motion caused by trigonometric error.

The forward and backward motion of the human transporter is controlled by a handlebar. Moving the handlebar forward causes the device to move forward. An exploded view of the Segway is shown in Figure 15.14.

In this application, the use of integrated sensor techniques has been used to produce a low-cost high-performance stabilisation system. The approach has also enabled the maximum use to be made of the available sensor information. Other approaches to stabilisation are considered in the following section.

15.5 Equipment stabilisation

In this section, the role of inertial sensors and systems is examined in applications where it is necessary to maintain a vehicle-mounted sensor pointing in a given direction in the presence of vehicle motion disturbances. These disturbances may be demanded manoeuvres or random effects such as turbulence-induced motion.

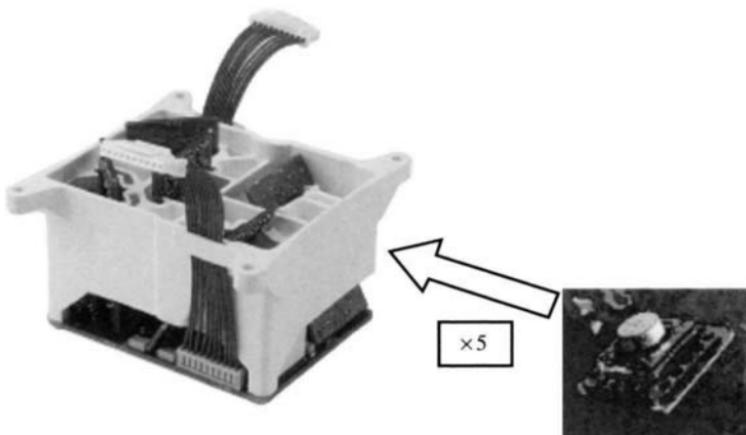


Figure 15.13 Photograph of the balance sensor assembly (Courtesy of BAE Systems)

measurements agree, the reference measurement has no direct role to play; once a discrepancy is detected the system shuts down automatically. Two of the other four gyroscopes are oriented to provide pitch and yaw data in such a format so that summing and differencing the signals will give the two estimates of pitch and yaw motion. The other two sensors provide pitch and roll data. Additional manipulation enables other estimates to be derived to provide redundant estimates of the motion.

Estimates of pitch and yaw are required to enable the ‘human transporter’ to make manoeuvres safely, such as turns whilst travelling up or down a hill, this avoids a cyclic pitch error over confusion in the system of the ‘location’ of the gravity vector. This is overcome by estimating the pitch motion with respect to the gravity vector in all three axes of the reference frame and compensating for the effects of erroneous estimates of pitch motion caused by trigonometric error.

The forward and backward motion of the human transporter is controlled by a handlebar. Moving the handlebar forward causes the device to move forward. An exploded view of the Segway is shown in Figure 15.14.

In this application, the use of integrated sensor techniques has been used to produce a low-cost high-performance stabilisation system. The approach has also enabled the maximum use to be made of the available sensor information. Other approaches to stabilisation are considered in the following section.

15.5 Equipment stabilisation

In this section, the role of inertial sensors and systems is examined in applications where it is necessary to maintain a vehicle-mounted sensor pointing in a given direction in the presence of vehicle motion disturbances. These disturbances may be demanded manoeuvres or random effects such as turbulence-induced motion.

(a)



Figure 15.14 (a) Segway transporter (Courtesy of BAE Systems)

Included in this section is the application of inertial systems for the measurement and control of relative alignment between stations on a flexible vehicle or craft. Inertial measurement techniques for the characterisation of vehicle motion are also considered.

15.5.1 Aero-flexure compensation

Modern combat aircraft are highly sophisticated and capable multi-role systems. They gain much of this capability from the use of distributed systems to provide enhanced functionality, as well as resistance to combat damage and the opportunity to incorporate new function data onto a platform after entering service. Because modern aircraft may well have a service life in excess of 30 years, there is a need to incorporate technology and systems as they develop and mature throughout the total lifetime of a platform. Technology insertion enables enhancement in performance, allowing the full potential of an aircraft to be achieved, as well as optimisation of the investment in the aircraft.

A distributed system of components provides greater flexibility for integration of a new system into a range of host platforms, including retrofit to so-called legacy

(b)

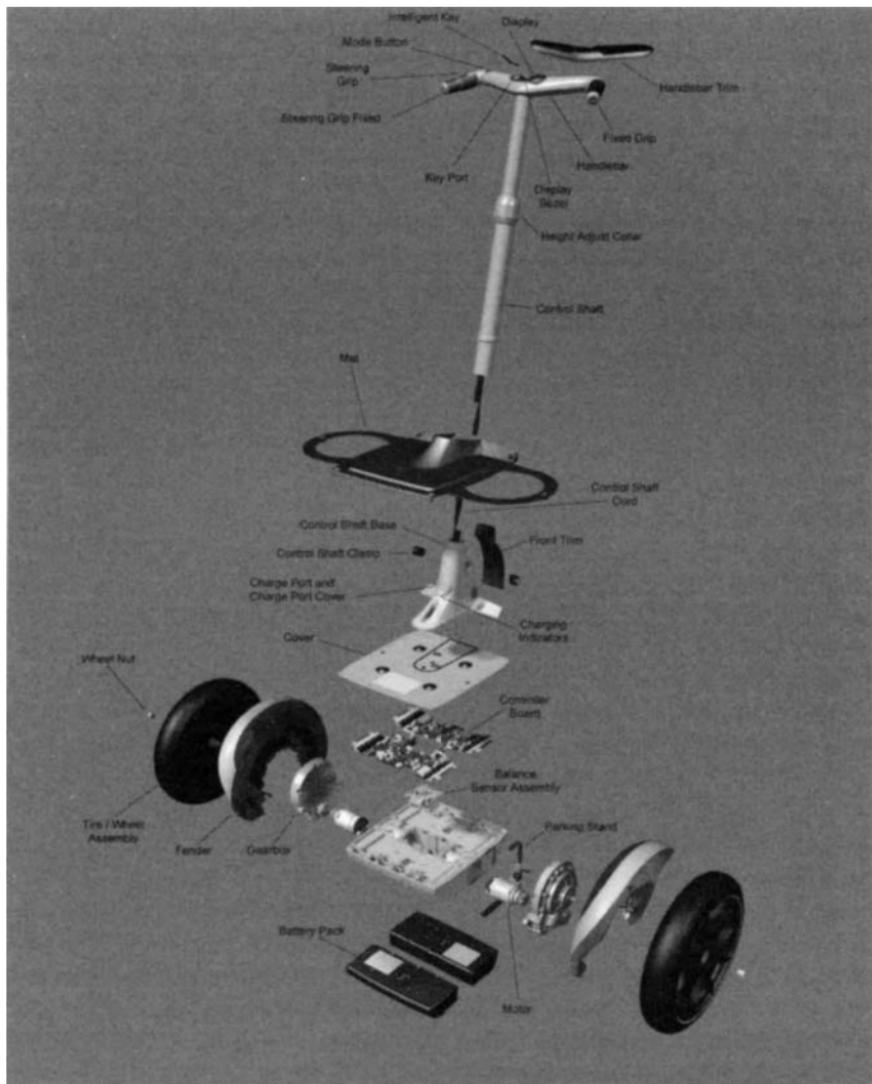


Figure 15.14 Continued. (b) An exploded view of the Segway transporter (Courtesy of BAE Systems)

aircraft. It is considered easier to integrate a number of small sub-system modules and components into an aircraft, rather than attempting to fit a single but larger monolithic structure. This is particularly the case when retro-fitting during technology insertion or through a mid-life update, although this approach has a potential impact on installed performance of the system.

The downside of this approach is the difficulty of achieving total system integrity, particularly if comparison of information measurements is required between various

sensing devices. An example is the need for harmonisation of the reference axes between the various components of a distributed sensor array with respect to the master reference system used within the host platform.

The operation of a defensive-aids suite (DAS) on an aircraft or ship is an example of a distributed system; its purpose is to defend a platform when any type of weapon attacks it. Any aircraft is vulnerable to attack by modern heat-seeking missiles, which may be launched from any bearing. Consequently, the various sensors that form the missile detection system are mounted at a number of locations around the aircraft structure to give complete situational awareness. This form of distributed system, using an array of separate sensors, provides the protection system with the potential for complete 4π steradians of angular coverage. An array of four, or more, wide field-of-view staring sensors are required to provide the DAS or counter-measure system controller with a comprehensive and uninterrupted view of the approach of hostile targets, and thus allow the appropriate countermeasure to be deployed. An idealised example of a distributed system is shown in Figure 15.15.

The electro-optical version of this type of system will be used to indicate some of the issues, particularly in compensating for the continuous or dynamic changes in orientation between the reference axes of the various components of this system. This problem will be considered in terms of a directed infrared counter-measures

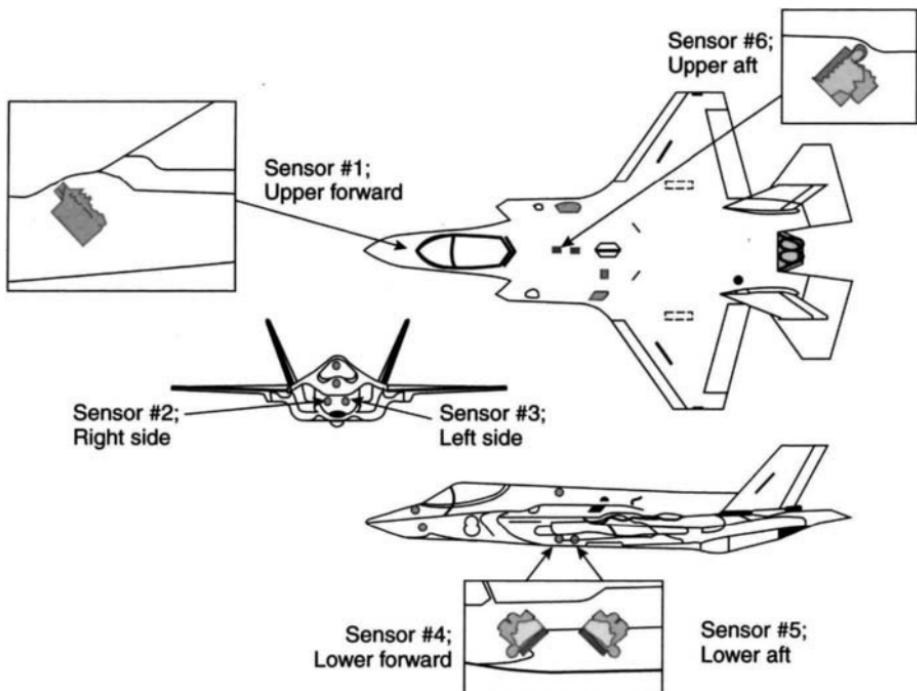


Figure 15.15 DAS distributed system (Courtesy of Northrop Grumman Electronics Systems)

(DIRCM) system on an aircraft, and in particular, achieving transfer of information from an array of situational awareness sensors that form the sensor suite to provide data to the on-board jammer transmitter system.

A photograph of a modern DIRCM system is shown in Figure 15.16 and a schematic diagram of the architecture of a generic DIRCM system is shown in Figure 15.17. This type of system is fitted to new aircraft and retro-fitted to legacy platforms to provide enhanced survivability. Specifically, the purpose of a DIRCM system is to detect and defeat heat-seeking missiles that are fired at an aircraft and clearly, this system needs to have a distributed system in order to fulfil its task.

This system has a processor, an array of situational awareness sensors (or missile approach warning sensors), to detect the launch of a hostile missile, and a beam director with an infrared thermal camera and a mid-wave infrared source of energy.

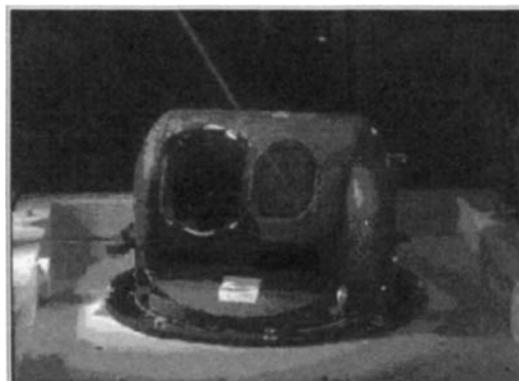


Figure 15.16 DIRCM system (Courtesy Northrop Grumman)

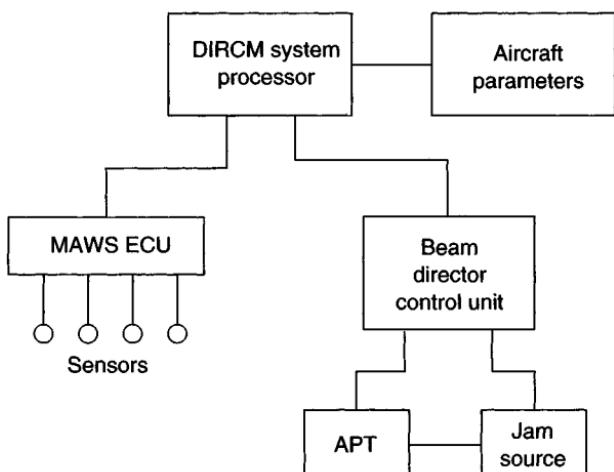


Figure 15.17 DIRCM system architecture

The beam director receives cueing information from the sensor array to locate the threat; it then tracks the target with the thermal camera and provides a series of pulses of infrared energy to confuse the target-tracking system in the approaching missile [8, 9].

The essential aspects of the information flow through the elements of this system are:

- an array of staring electro-optical sensors that detect the direction of approach of a missile in local reference frame axes;
- formation of a track being followed by the detected object, in the computer processor, and determination of its direction of arrival in the chosen reference axes of the aircraft system, and to provide cueing commands to a beam director to slew to the calculated approach direction;
- undertake a search about a narrow field-of-view centred on that bearing (by the tracking sensor in the beam director) to detect and acquire the designated target by that tracking sensor;
- track the designated target and send the infrared energy at the approaching missile.

A critical issue controlling the performance of such a system is error in the accuracy of the bearing information of the approaching missile used to cue the beam director. Errors in this process give rise to so-called handover errors, and lead directly to a reduction in the probability that the beam director, with a narrow field-of-view tracking camera, will find a designated target instantaneously. This is directly analogous to the problem described in Chapter 14 for missile seekers transitioning from inertial navigation guidance to terminal homing. These errors arise from a number of sources, such as measurement inaccuracy of the distributed sensors and uncertainty in the harmonisation of these sensors with respect to the master reference system.

The latter component tends to be dominant owing to the dynamic characteristics of modern military aircraft, particularly transport aircraft, which leads to flexure of the airframe. Hence, there is a potentially significant and random, time-dependent error in the knowledge of the orientation of the local reference frame with respect to the master reference frame. Another significant contributor to this error results from mounting misalignments of the situational awareness sensors on to the structure of the aircraft. However, this is quite easy to compensate for, as it is a systematic error and may be measured by surveying the sensor array with respect to the master reference system in the host vehicle.

The most efficient technique for compensating this undesirable motion within the structure of an aircraft is to use a local inertial measurement unit adjacent to each situational awareness sensor. This provides a data stream with the details of the change in attitude of the remote components of the sensor array, and this data stream may be time tagged and compared with the attitude changes recorded by the master-reference system. This technique is very similar to the transfer alignment processes described in Chapter 10. An alternative, and lower-cost approach, involves the use of a single IMU at a representative location to provide an indication of the flexure of the airframe.

Fortunately, the DIRCM system only requires details of changes in the orientation of the remote sensor array with respect to the master reference system, and therefore

a relatively simple IMU is sufficient for this purpose. Moreover, simple rate sensors are adequate, so a compact IMU can be formed using devices such as the quartz rate sensor, or equivalent devices, to give the three axes of angular motion data to the DIRCM processor. It is important to use sensors with a low-noise characteristic to enable accurate and timely correlation of the data streams. The data rate and sensor bandwidth need to be faster than the fundamental frequency of the flexural motion.

This approach to compensation has a number of benefits to give accurate system performance, and is applicable to a whole range of distributed systems. Additionally, the remote IMU outputs may be used to assist in providing estimates of target sightline rates locally, if the sensor and its processor are capable of providing tracking data, as well as general situational awareness for the platform. This information can then assist with the selection of the most threatening targets.

An alternative technique, using the output information for the aircraft's IN system may also be applied to solve this problem. In this case, the influence of external forces on the distortion of the airframe needs to be known, and to be correlated to the response of the aircraft. Consequently, if there is a good correlation between the motion of the aircraft and its flexure, then the outputs from the IN system may be used for flexure compensation using the real-time outputs from a suitable mathematical model of the phenomenon. Clearly, the success of this approach is dependent on the efficient characterisation of the aero-flexure response characterisation of each class of aircraft. Moreover, there should be very little variation in bending response of the airframe to a given external stimulus from aircraft to aircraft in a particular class of aircraft if this approach is to be viable.

Alternative approaches to account for random and uncorrelated errors in the estimation of the bearing of an approaching missile resulting from flexure of the airframe include:

- Use of a beam director with a wide field-of-view sensor that is used to track the target. This would allow for a larger handover error, but reduce the spatial resolution of the target-tracking system. An optical zoom system may be used to change the focal length of the target-tracking system to regain spatial resolution, but this adds complexity, cost and weight to the tracking system.
- Use of a beam director without an integral target-tracking system, but this requires a very much wider beam that reduces the radiant intensity, and hence the potential effectiveness of the DIRCM system.

However, these alternative techniques and approaches are not particularly effective in comparison with the use of real-time compensation methods.

15.5.2 *Laser beam director*

Laser systems are used for a range of military and civilian applications [10]. Many of the applications require the very small, low-divergence laser beam to be directed to a remote object. This may be accomplished using a range of techniques such as a fibre optical cable, an articulated arm or a beam director. In the case of a beam director this may be a solid-state system using the optical 'flash' from a reflecting object to

provide the tracking cue, and provide the beam direction in a ‘long cavity’ laser. However, by far the most common techniques use a mechanical system involving optical components mounted in a servo-controlled gimballed structure.

Gimballed structures are used where the laser-based system is required to operate over a wide field-of-regard, such as a hemisphere or a hyper hemisphere for scanning a volume of space. Examples of this application include laser radars for remote sensing, such as detecting obstacles and wind shear. Another example is a development of the DIRCM system considered earlier [8].

A typical laser-beam director has a number of tasks to undertake, which require the use of inertial sensors. Examples include:

- measurement of the angular motion of the platform to provide sightline stabilisation of the laser-beam pointing system if the device is mounted on a moving platform;
- feedback of the angular rates about the gimbal axes in the beam director to the control system of the beam director during target tracking of a moving target;
- measurement of the angular dynamics for control of the beam director during the slewing and pointing functions.

This section will consider the aspects of the moving target application with the beam director at a fixed and stable position; the other aspects of the stabilisation are considered in other sections of this chapter. Of course, the complete system for operation on-the-move will require the integration of all of the measurement, control and compensation systems into a consolidated system; moreover, the response characteristics will need to be mutually compatible.

This type of beam directing system has a target-tracking system and a servo control system used to position a beam-directing optical element, such as a mirror. The target-tracking system and the beam-directing element operate in a closed-loop system, with the mirror responding to the commands from the target-tracking device so that the laser beam follows the target.

A modern target-tracking system in a laser beam director uses an electro-optical camera to view a scene around the optical sightline or bore sight of the beam director. This camera may operate in the visible or infrared wave bands of the electromagnetic spectrum, and is dependent on the existence of contrast between the target and the background to enable the target-tracking function to be undertaken. This contrast may be positive or negative. An image-processing algorithm is used to extract the position of the target’s image within the captured scene, and in particular, its displacement from the optical bore sight, or other defined and specific tracking point.

A number of target-tracking algorithms [11] may be applied to the captured data to move the gimbals in the beam director to ensure that the designated target is at the centre of the field-of-view within the so-called tracking box and that it remains there; that is, the target is tracked. If the target appears as a point source then a simple quadrant vector approach may be used, which is analogous to a quadrant detector. If the target has spatial extent, then a more sophisticated approach may be required to determine the target’s centroid, or some other specific point on the image, that will provide the reference point for the target-tracking algorithm.

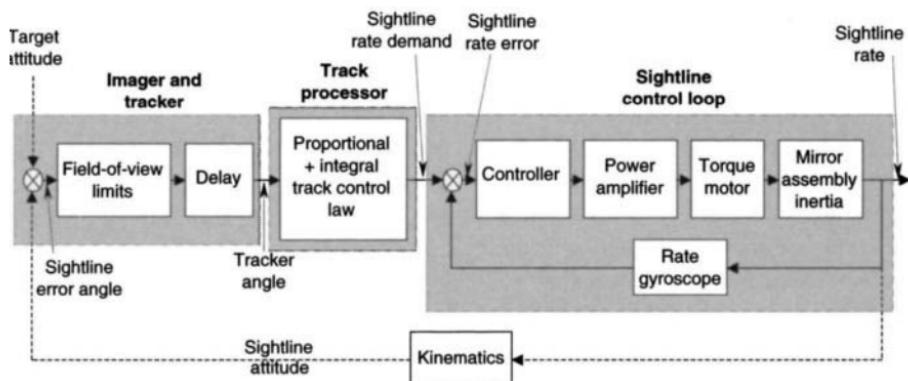


Figure 15.18 Control scheme for laser beam director

As the target moves, the target-tracking system will detect a displacement within the tracking box and this will be converted into a perceived angular error. A control law is used to demand an angular displacement of the gimballed system; a number of control laws can be applied, but a common one is integral plus proportional control. This control law is attractive, as it does not require a displacement of the target from the sightline to the optical axis of the camera to follow a sightline moving at a constant rate. More sophisticated techniques are required to provide very small tracking errors if the sightline is accelerating or decelerating [11].

Rate gyroscopes are used in the sightline-control system to sense the angular motion achieved by the torque motors used to move a gimbal. The sightline control system forms a voltage demand based on the perceived displacement of the target from the sightline. The demand is shaped according to the control law being applied, and this voltage is amplified before being applied to the torque motor. The torque motor, which may be considered to be an electrical lag, a constant and a resistance, provides the force that accelerates the mirror by an amount that is inversely proportional to the inertia of the beam positioning element. The control scheme is shown in Figure 15.18.

There is some choice in the location of the rate gyroscope within a beam steering system. A full strapdown approach may be used, in which case a rate sensor is positioned on the body of the beam director to sense motion of the base of the beam director directly. In this case a resolver, or angular pick-off device, is required to detect the angular motion of the gimbal with respect to the reference frame of the gimbal system. An alternative approach is to place a rate sensor on each of the axes of a multi-axis gimbal. As far as system design is concerned, this is a simpler approach as it gives a direct indication of achieved angular rate, but has the disadvantage of producing a larger and more complex gimbal structure. The system has larger inertia and greater rotational spring restraint, and thus requires larger torquers to give equivalent performance. Generally, structures and systems with large moments of inertia have lower slew rates compared with smaller devices with smaller moments of inertia.

Lower accuracy resolvers are required when the rate sensors are mounted on the gimbal structure, as the resolver signals are not required by the control loop for sightline stabilisation and target tracking. Additionally, the data latency requirements of this sensor are less critical. In this case the role of a resolver is to feedback the position of the gimbal axes during an angular slew command for positioning (and controlling) the optical axis of the beam director within its field of regard. Furthermore, any errors are immediately removed once the target-tracking and sightline-stabilisation functions are active. The angular accuracy requirement is driven by the need for a high probability of transition from slew to tracking; that is, completion of the handover function.

Stabilisation of the gimbal axes in inertial space can be achieved by moving the gimbal in a manner to negate the platform-induced motion detected by the rate sensors. Any rotational motion sensed by the rate sensors in any of the axes can be used for generating an inverse command to the gimbal. This stabilisation of the perceived rate sensor movement directly produces a stabilisation of the ‘gimbal’ line of sight to the target in inertial space. For tracking, a commanded movement of the gimbal mechanism immediately appears as gimbal rotations sensed by the rate sensors and can be used to close the control loop.

Other advantages of making direct measurements of the angular motion:

- There are no structure bending factors or modes to be taken into account.
- The signal from the gyroscopes can have much higher precision than that provided by resolvers. This also means a two-axis sensing capability is possible if the sensors are orientated to measure the two-degrees of angular motion. A three-axis IMU is required in the strapdown case with larger dynamic range owing to the need to sense the full motion undertaken by the host platform.
- An inertially stabilised gimbal, as is likely to be used in a tracking phase, will only need the rate sensors to detect the unstabilised residual motion and can be optimised for small angular deflections.

An important aspect of a successful laser beam director is having knowledge of the direction of the optical axis of the laser beam with respect to the target-tracking camera (so-called cross eye). It is vital that an alignment or harmonisation process is undertaken: this process may be undertaken using a prism to provide a retro-return from a distant position to measure the displacement of the laser beam from the axis of the target-tracking system. This off-set can either be corrected by moving the beam-directing element, or by compensating for the angular offset during the demanded angular displacement of the element.

In this type of application the angular rates achieved by the beam director may be high, of the order of hundreds of degrees per second, possibly exceeding one thousand degrees per second. However, an important, if not crucial consideration, is the noise on the output from the angular sensor. This is because the pointing stability (jitter) required from the beam director may be measured in micro radians.

Very accurate target tracking is often required even though the designated target may be undertaking highly dynamic manoeuvres. In these circumstances sophisticated target-tracking algorithms are required to ensure that the laser spot remains on the

distant target even with a changing sightline rate. In these circumstances, great care is required to ensure that there is no cross-coupling between the input axes of the gyroscopes as this leads to instability in the feedback loops of the control system.

Investigation of the motion of the rotor in a dynamically tuned gyroscope (DTG) during motion of the gimbals in a beam director has shown that the rotor may nutate. This is a consequence of the rotor being 'torqued back' to its null position in the normal operational mode of these high-precision sensors. This motion of the rotor means that the input axes, defined by the rotor position and its pick-off devices, are likely to experience some cross-coupling effects, owing to the nutation dynamics of the rotor. The cross-coupling effects inherent in the DTG nutation dynamics of the free rotor in these sensors may not allow the input axes to be satisfactorily defined in the sensor. As a consequence of this motion of the rotor, its input axes for very accurate beam director systems are not adequately defined and consequently cannot be stabilised with a high-gain control system, such as a type II tracking loop [11].

A solution that may be applied for high-performance beam directors is to de-couple the track loops by replacement of the dual-axis gyroscope with two single-axis angular rate sensors. Independent measurement of azimuth and elevation motion is then possible by using two single-axis sensors within the system.

15.5.3 *Laser radar*

A laser radar system is frequently known as a Lidar or Ladar device, owing to its similarity with conventional radar systems. In this case laser light is used to probe a volume of space to provide information about objects intercepted or encountered by the laser light. Owing to the coherent nature of laser light [10] it is possible to get a great deal of information regarding the characteristics and behaviour of objects that are detected, as well as about the various processes that are occurring in the atmosphere. The principal advantage that a laser-based system has over a radar system is its spatial resolution for a given aperture, but of course, it tends to be more sensitive to environmental propagation conditions.

A laser radar (or ladar) [12–16] has an active laser transmitter, a receiver and a scanning mechanism to search a field-of-regard (FoR). A ladar works by measuring the time delay between the laser-pulse emission and the detection of the returned pulse. This information provides the absolute range to each 'picture element' in the scene, and therefore by scanning the laser beam and its receiver over an FoR, a three-dimensional map of the terrain may be generated.

The ladar-generated image contains spatial and temporal data, which are essentially invariant with environmental changes to a target's surface, such as water or dirt on a surface, or temperature. Additional information may be extracted from the received data such as any frequency shifts induced by motion of an object or vibrations of a target, as well as variations in reflected intensity. The fundamental principles of vibrometry are outlined in the section on calibration and measurement later in this chapter.

Modern computer technology enables the 'picture' formed from the volume of space scanned in the searched sector to be processed rapidly. Image processing

algorithms can filter the data, so for example, the retro-returns from objects that do not have any interest to a particular system, such as clutter or small bushes close to the ground can be rejected. The flexibility of the processing of the captured scene means that these techniques can be applied to a range of applications. Two are considered below, as they require use of inertial data:

- automatic target recognition for target selection during autonomous terminal homing of missiles;
- obstacle avoidance for low-flying aircraft and helicopters.

15.5.3.1 Automatic target recognition

The application of laser-radar techniques to military systems has enabled autonomous weapons to be demonstrated that use automatic target-recognition methods. This involves the search of a volume of space to form a ‘picture of the terrain’ in front of the laser radar and comparing it with stored data. The database may include:

- details of the terrain in the target zone for aiding the navigation system;
- characteristics of the designated targets, which may include spatial and temporal characteristics.

These ‘data points’ are defined with respect to the fixed reference axes of the laser radar system and are recorded in terms of the bearing of the scanning mirror and range from the transmitter. Hence, the Cartesian co-ordinates of measured points may be calculated in terms of the system reference axes. To be able to refer to these co-ordinates, and hence the observations to another reference frame the position and attitude of the laser radar system is required to enable the axis transformation to be undertaken. Moreover, if the host platform is moving, then line-of-sight stabilisation will be required, as discussed in Section 15.5.5. A slave IMU may need to be used if the laser radar system is positioned remotely from the master IN reference system and there is significant motion between the laser radar system and the reference system. Additionally, frequent IN updates are required owing to the motion of the vehicle and it is normal to time tag the measurements to provide suitable synchronisation of the ensemble of data.

The processed returns from the scene may be evaluated with various algorithms to undertake the detection, acquisition and tracking of objects. These may be classified according to a range of criteria, typically according to size and shape. The targets may then be compared with a stored database to provide recognition of military vehicles, buildings, bridges or other man-made structures. A three-dimensional picture may be formed and the target data superimposed on the ‘map’ of the area provided the ladar data are referenced to the stored database co-ordinates. This aspect is considered further in general terms in Section 15.7.1, on moving-map displays.

These techniques are being developed into laser-radar seeker systems for use in tactical weapons to detect, acquire and track surface targets. The seeker contains a scanning system that enables the laser radar to search its field of regard to find a target. A simple single-plane azimuthal scan, with the beam directed downwards at a shallow angle to the flight path, is often adequate for a missile flying

horizontally, as a forward motion will provide the 'second dimension' for the scan mechanism.

The processes undertaken by the ladar seeker are:

- laser beam scans the field of regard of the seeker;
- process the received information to create an image that includes height data;
- filter the captured scene to identify potential targets and eliminate background features and other objects that do not represent targets;
- process the image objects that are potential targets to estimate dimensions (length, width and height), orientation and information about any objects that have been detected on the potential targets;
- classification of the targets in the captured scene and identification of the objects using three-dimensional target matching algorithms;
- generation of guidance commands for target interception.

The seeker requires stabilisation during its various modes of operation, as well as precision control of the scanning mechanism, as the transmitter and receiver are panned across the field of regard. Rate gyroscopes provide the required control function by detecting the disturbance, providing the feedback to the control loops to stabilise the sightline and ensure that a smooth and consistent scan is achieved during the target acquisition process. Sightline stabilisation and beam control are discussed in more detail in Sections 15.5.5 and 15.5.2, respectively.

The inertial measurement unit in a cruise missile may provide the measurement of the angular disturbances, so an accurate angular pick-off device would be required to provide a measure of the precise position of the optical axis of the seeker. Alternatively, rate gyroscopes may be used on the gimbal mechanism to record the angular disturbances and then used, in conjunction with lower-precision angular pick-off devices, to control the scan over the field of regard.

15.5.3.2 Obstacle avoidance

High-performance surveillance systems linked to air-defence units present a significant threat to aircraft attempting to penetrate the defended air space. The effectiveness of these air-defence systems compels interdiction aircraft to fly at very low altitudes, in order to avoid detection by these air-defence systems. The aircraft may have to fly close to the ground at very high speed in poor visibility, and possibly with inadequate information about the local terrain. The possibility of a collision is high, particularly if the aircraft relies on conventional sensors to detect potential obstacles along the flight path. Other roles, such as reconnaissance or search and rescue missions, may also require aircraft to fly close to the ground.

Real-time and accurate situational awareness is, therefore, a crucial safety function for any low-flying aircraft. Obstacle avoidance systems, based on laser-radar technology [15, 16], offer a very valuable aid to aircraft that are required to fly close to the Earth. One of the major hazards encountered during this type of low-level flight over land is the man-made obstacles, which are difficult to detect such as wires between pylons. Laser-radar systems have been developed



Figure 15.19 Examples of laser-radar systems (Courtesy of QinetiQ and FGAN-FOM)

to provide an obstacle avoidance system for helicopters and fixed-wing aircraft. Figure 15.19 shows the system developed in the United Kingdom by QinetiQ and another developed in Germany for use by the German border patrol on their EC 135 helicopters.

In the general situational awareness role the laser-based system scans the terrain and its processor forms a three-dimensional picture of the received reflections produced by the objects in the scene. Advanced three-dimensional algorithms for scene analysis have increased the probability of recognising obstacles and reduced false alarms. The use of angular motion sensors in the system enables the position of the laser beam for each position in its scan to be tagged and the co-ordinates established with respect to a reference, as discussed for target recognition. In this case the data when correlated with the reference frame of the system, allows the detected obstacles to be superimposed on the image, to create a synthetic picture. Image processing may be used to link the various returns superimposed on the scene to indicate positions of wires. Additionally, correlation of the position of the measured and the stored data provides accurate navigation cues for terrain referenced navigation.

Laser-radar systems may also be used to detect wind shear ahead of aircraft, so that the effects of turbulence may be mitigated, if not avoided. In this case, the laser-radar detects the motion of aerosols and other small objects in the atmosphere. Again, the use of inertial sensors means that the position of the detected phenomenon may be given a location in a reference frame and the appropriate avoidance measure applied.

15.5.4 Seeker-head stabilisation

The development of heat-seeking missiles started after World War II with the pioneering work of Dr McLean at the then Naval Ordnance Test Station at China Lake in California. His idea was to produce a simple but effective guided weapon that had a seeker that would home on to the thermal energy emitted by an aircraft.

In recent years there has been great proliferation of the small electro-optical (EO) and infrared (IR) guided missile systems. This proliferation is attributed to the low

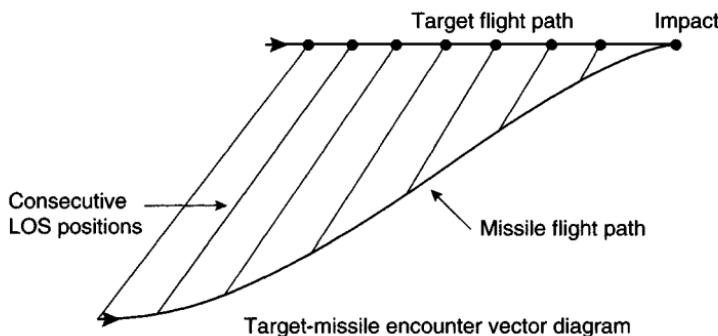


Figure 15.20 Principles of proportional navigation

cost, ease of operation and the excellent lethality record of these devices. These missile systems include air- and ground-launched anti-tank missiles as well as the more common anti-aircraft missiles. One of the novel approaches used in this design concerned the use of inertial sensor principles to stabilise the optical sightline between the seeker and the target. Consequently, the seeker pointing direction was insensitive to missile body motion, enabling the target to be tracked.

Most EO/IR missile systems use some form of proportional navigation guidance during their homing phase. This type of navigation law is effective as it relies on achieving a constant bearing or sightline between the missile and its target, which it achieves by nulling the sightline-angle rate. This form of interception path is very efficient, as the intercepting missile aims for a point ahead of its target's current position, which avoids tail chases. However, this guidance law will only work if the seeker system has access to information about the target-to-seeker sightline rates that are independent of missile motion and other external disturbances. The principles of proportional navigation are illustrated in Figure 15.20.

In order to achieve an inertial line-of-sight angle measurement most EO/IR missiles use gyroscopic motion within their seeker head, so that the gyroscopic inertia (i.e. angular momentum) provides an inertial reference that may be translated into stabilisation. This rotating mass is stabilised independently, i.e. isolated from the missile body motion. This stabilisation is crucial to the seeker because the target would easily be lost from its small field of view if it were subject to the perturbations caused by normal missile body disturbances. A similar situation is described below, in Section 15.5.5.

These EO/IR missile seekers typically use a cassegrain optical telescope to focus the thermal scene on to the detector. This telescope is mounted in a two-axis gimbal structure and the whole of the telescope and detector assembly is rotated about the sightline at a frequency of the order of 100 Hz. In some cases the rotor assembly (telescope and detector) is maintained at the chosen angular speed but in other cases the rotor is allowed to spin down once the rotor has been uncaged, that is, the angular rotation rate is not sustained.

The rotation of the telescope [17] is also useful in the target-tracking task in a reticle-based seeker. In the case of a conically scanning seeker [8, 17], one of the

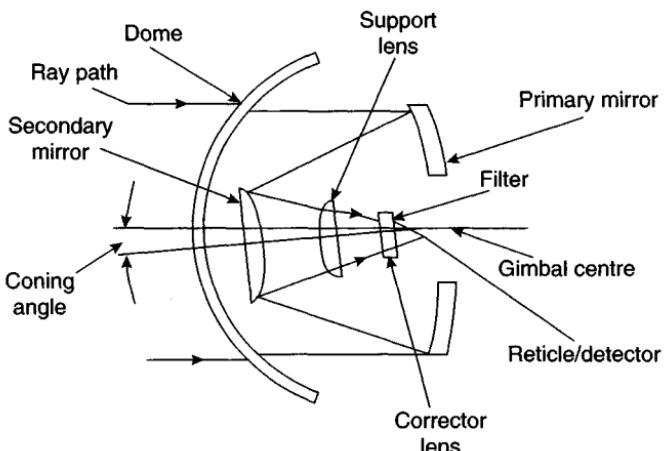


Figure 15.21 Conical scan seeker schematic

mirrors is canted with respect to the primary mirror. This type of seeker traces out a circle of received energy on to the seeker reticle, when it is aimed directly at a point source of thermal radiation in the thermal scene. This is a consequence of the combination of the spin motion of the seeker head and the cant of one of the mirrors forming the telescope. The radius of this nutation circle on the reticle is proportional to the angular misalignment between the mirrors of the cassegrain telescope. However, if the seeker is not pointed directly at the point source in the thermal scene the circle will not be centred on the reticle. The magnitude of the difference between the centre of the reticle and the centre of rotation of the nutation circle is proportional to the angular error between the optical axis of the seeker and the true line-of-sight to the target. Figure 15.21 shows a schematic of a conical scanning optical seeker.

In these systems the gyroscopic body forms the permanent magnet, so that the wire coils embedded in the missile fore body can be used to control the seeker's telescope assembly (i.e. the gyroscopic rotor) and sense its angular position. The magnetic field of the spinning seeker cuts the sensor coils, generating an electric current, which is read with a pick-off device to determine the seeker's angular position. A current is applied to the precession (torquer) coils in the seeker, which induces a magnetic field. This magnetic field interacts with the magnetic field of the rotor body to induce control torques on the seeker telescope.

Figure 15.22 depicts a seeker assembly with a cassegrain optical telescope, and shows the configuration of the sensor and its precession (control) coils. The reference coils are two 'pancake' type coils placed on opposite sides of the missile fore body. As the seeker spins, the magnetic field of the seeker cuts these coils regardless of seeker orientation. The resulting signal induced in the reference coils is a sinusoid at the relative rotor to missile body spin frequency.

The cage-coil sensor coils are wound circumferentially around the missile fore-body like the precession coils shown in Figure 15.22. If the seeker telescope is

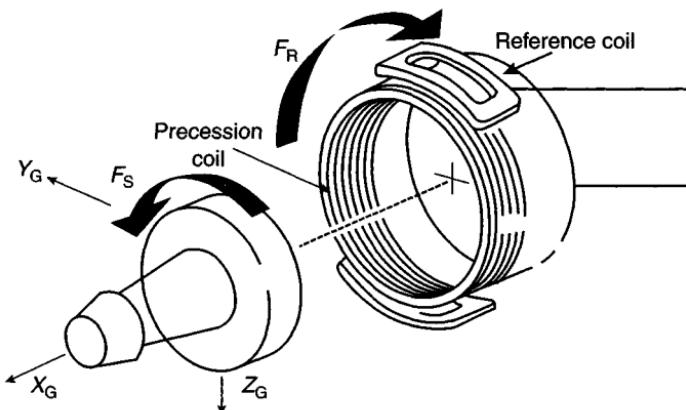


Figure 15.22 Seeker assembly with control coils

angularly aligned with the missile's body axis (bore sighted), then no current is induced in a cage coil. As the seeker telescope processes away from the bore sight axis, the seeker's magnetic field cuts the cage coils and induces a sinusoidal signal in the cage coils. The magnitude of this sinusoid is proportional to the displacement of the rotor's axis from the missile bore sight (usually termed lambda). The direction of the seeker telescope precession can be determined by comparing the relative phase difference between the reference-coil signal and the cage-coil signal. Therefore the missile can determine the angular displacement of the seeker's optical axis with respect to the bore sight, using the reference- and cage-coil signals.

The precession coils are used to control the telescope's angular position by inducing magnetic fields to generate torques on the spinning rotor assembly. A current injected into the precession coils creates a magnetic field, which is aligned with the missile's body longitudinal axis, usually designated as the X_M axis. This magnetic field has one of its poles directed along the longitudinal axis, depending on the direction of the current flowing in the precession coils. This induced magnetic field will generate a torque on the telescope by its interaction with the magnetic flux field of the spinning seeker telescope system.

Assuming that the north-pole of the rotor is aligned with the body-fixed axis of the seeker (Z_G axis), then the precession coil magnetic field will induce a torque on the seeker about a lateral seeker-fixed axis (Y_G). Efficiency of the precession coils decreases as the telescope increases its angle away from the bore sight of the missile. As the seeker processes away from the missile's bore sight axis, the induced precession torque has a component about the seeker's body Y_G axis and a component about the rotor (or seeker assembly) spin-axis (X_G). The X_G component of this torque can change the seeker's spin rate, particularly at large off-bore-sight angles. The axis convention and the layout of the various coils used to control the seeker are shown schematically in Figure 15.23.

As noted before, some missiles spin up the seeker telescope assembly prior to launching the missile and then let the seeker coast during the flight, so its angular

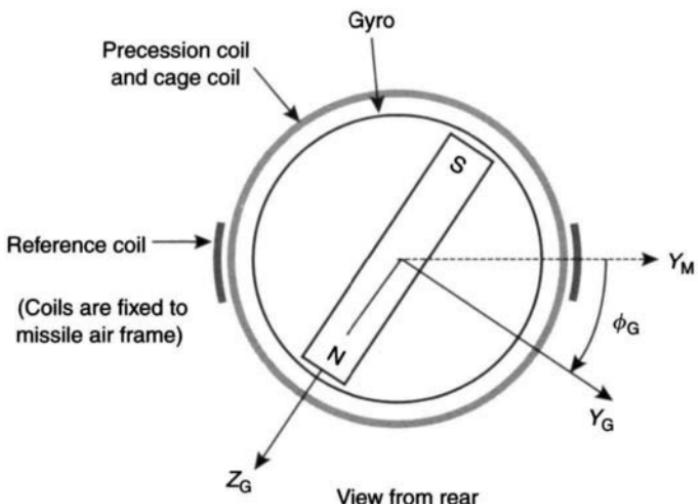


Figure 15.23 Schematic representation of seeker control coils

momentum decreases with time. Others maintain the spin frequency. The former is acceptable for many applications, since the typical flight time of small missiles is relatively short and therefore the seeker's angular momentum will not change significantly during the engagement period. However, some missiles use an on-board motor to maintain the rotor's spin rate, this being important for longer range engagements. The spin motor uses a pair of pancake-type coils, similar to the reference coils described above. A sinusoidal current waveform is used to power the spin coils, which induces a magnetic field, perpendicular to the missile's body axis (X_M). If this current is injected at the designated seeker spin frequency, and provided it is phased correctly, it will create a torque about the seeker body axis (X_G), which will control the spin rate of the seeker assembly. As with the precession coils, the spin coils will also induce a precession torque on the seeker as it moves away from the missile bore sight axis (X_M).

The rotating seeker telescope assembly in this configuration behaves as an inertially free device, owing to its two-axis gimbal suspension and the spinning mass of its components, giving it angular momentum. The seeker telescope assembly behaves as a gyroscopic rotor and remains pointing in the same inertial reference direction, even if its outer gimbals are in motion, that is, it behaves as a free gyroscope. External torques applied to the rotating seeker assembly cause gyroscopic angular motion (precession). The direction of this motion can be found by taking a cross product of the gyroscope's spin vector and the torque vector using the right hand rule. This motion can be interpreted as the gyroscope (i.e. the spinning seeker assembly) trying to align its spin vector with the torque vector, as occurs with any mechanical gyroscopic device.

An alternative approach to the stabilisation of a gimballed seeker assembly involves the use of rate gyroscopes mounted on the gimbal to provide the feedback data to a servo system. This approach is identical in principle to the technique described below for the sightline stabilisation of a laser beam director. However, this approach would be an increase in complexity and contrary to the objectives of these small, cheap and simple weapons.

15.5.5 Sightline stabilisation

The use of EO systems on combat aircraft and other military platforms has become very common for many military systems. Particular examples of airborne applications are laser-guided bombing systems and countermeasure systems to enhance platform survivability. Other examples include the use of electro-optical systems for reconnaissance on the battlefield, using stabilised optical sights. The performance of many optical systems is limited more by the degree of line-of-sight stabilisation that may be achieved than the inherent capability of the optical train, including its detector. Consequently, line-of-sight isolation and motion compensation techniques have to be applied to these systems to allow them to function to their full optical capability. This is a clear example of system optimisation.

These systems have to work on platforms that are highly manoeuvrable and where angular vibrational motion is large in comparison with the spatial resolution of the sensor. Moreover, sensors tend to have a limited field of view in order to provide good spatial resolution by the detection system for identification of targets, so stabilisation is vital for full functional operation of this class of device. However, there is also often a need for very wide-angle viewing of a scene, so the sensor needs to be panned (or scanned) over a field-of-regard. Therefore, there is often a requirement for methods to enable pointing and scanning of an EO system, as well as stabilising the sightline during manoeuvres undertaken by the platform, whilst using the system. Techniques for tracking a target and pointing a laser beam have already been discussed above and may be applied to the problem of scanning a scene and sight-line control when tracking a target.

Line-of-sight control and stabilisation in the presence of motion of the host platform may be achieved in a number of ways. There are two main approaches:

- platform stabilisation;
- and strapdown system stabilisation.

These approaches are both derived directly from inertial navigation system techniques.

In the former case the entire system, be it an EO system or a radar-based system, is mounted on a stabilised platform, similar to the IN system platform approach described in Appendix C. This is usually a two- or three-axis gimballed arrangement to provide the necessary degrees of freedom to isolate the system from the angular motion of the host platform, often termed 'own-ship motion'. The stabilisation principle involved is that the line of sight remains fixed in inertial space, owing to the stabilisation of the base, analogous to the 'fixing' of the inertial element

in an IN system platform system to maintain the accelerometer sensors in a given orientation. In the case of a highly dynamic platform a minimum of three-degrees-of-freedom will be required in order to prevent any singularities occurring in the system, which would lead to gimbal lock. A four-degrees-of-freedom approach may be required with some manoeuvres to ensure that this so-called nadir condition, or singularity, does not arise, where a degree of sightline control is lost.

The major problem with the stabilised platform approach is the size, mass and power consumed by this type of system. This system is inevitably large, as the sensing device being stabilised has to be mounted at the heart of the system, requiring substantial torque motors to provide adequate line-of-sight control and stabilisation.

The use of strapdown-type approaches offers a more compact system, but it is reliant on inertial sensors and other components with the appropriate dynamic range measurement capability and real-time, complex signal processing to provide the requisite dynamic demands to give the line-of-sight control. Moreover, active stabilisation methods require timely information about the line-of-sight-motion, with negligible data latency, if effective sightline isolation is to be achieved. Additionally, the processing also has to provide a direct substitute for the damping given by the inertia of the mechanical elements, such as the gimbals and 'inertial element', in a platform-based stabilisation system.

Strapdown techniques may be considered to be an indirect method of sightline stabilisation and a number of implementations are feasible:

- mounting only the smaller elements, such as a mirror, on a stabilised platform, which offers a light-weight system with low inertia and the potential for high angular-rate capability. The mirror may be used to steer and stabilise a number of sensor sightlines through a single aperture. This minimises the number of apertures required by a device or system;
- a full 'strapdown approach' based on a remote IMU requiring very accurate sensors capable of providing three axes of angular-motion data over the full dynamic range of a platform's motion, which has the potential for a very compact but computationally intensive system.

Inevitably the application of strapdown techniques leads directly to a more complicated sub-system, as the stabilised element has to be incorporated and controlled within the EO device. However, the complete stabilised system is far more compact, generally it has a much smaller mass and consumes far less power. This approach has many similarities with a strapdown IN system; it has traded mechanical complexity and inertia for computational complexity and sophisticated use of algorithms.

A stabilised-mirror system [18] will be used to illustrate the issues involved with implementation of a strapdown stabilised line-of-sight system. The essential elements of such a system are:

- a mirror mounted in a gimballed structure to give the requisite number of degrees of freedom;
- torque motors to compensate for platform motion;

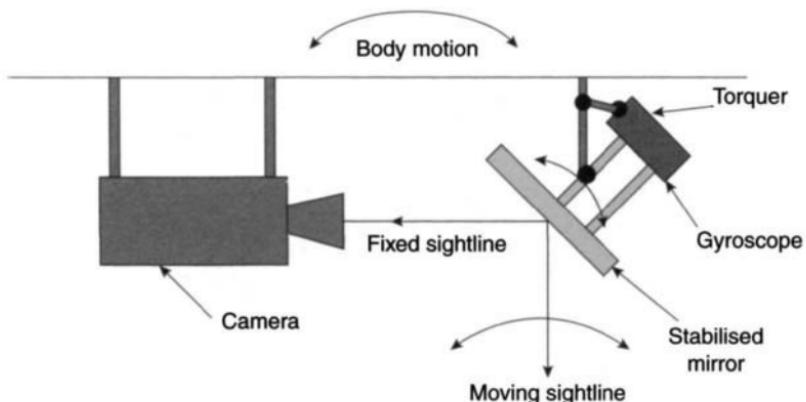


Figure 15.24 Stabilised mirror system

- resolvers/angular transducers to measure angular displacements;
- a rate gyroscope to measure angular rates about each axis for each rotational degree of freedom.

An additional function of the torque motor is to direct the optical sightline to provide the pointing or scanning function of the system within the sensor's field of regard.

The mirror mechanism is a two-axis device, having an inner and outer gimbal. The outer gimbal of this mechanism is aligned with the optical axis of the EO system. On the rotating outer-gimbal structure are two inner-gimbal axis structures, which are parallel to each other and orthogonal to the outer-gimbal axis. One of these inner-gimbal axes carries a 'gyroscopic platform', on which the inertial sensors are mounted. The other inner-gimbal structure carries a mirror. This arrangement is shown in Figure 15.24.

The deflection of a mirror surface through any angle produces a deviation of the reflected ray from that surface through an angle equal to twice the deflection of the mirror. Therefore the beam-steering mirror is required to move through half of the angle detected by the inertial sensors. In order to achieve this gearing, a 2 : 1 ratio link connects the mirror structure to the gyroscopic platform.

This mechanism can be designed to combine high structural rigidity with very low bearing torques about the rotational axes. This may be achieved through the use of pre-loaded duplex pair ball bearings in these gimballed structures.

The inertial (or gyroscopic) platform may have two single-axis rate integrating gyroscopes, which are used to stabilise the inner and outer gimbal axes, using conventional stabilisation techniques. Alternatively, a single two-axis gyroscope can be used, but care is required in the mounting of this device to avoid undesirable cross-coupling effects of the input axes, which will lead to spurious perceived rates, as considered in Section 15.5.2.

A servo-stabilisation loop for a single-axis system is shown in Figure 15.25. A rate-integrating gyroscope is used to measure the angular motion of the platform,

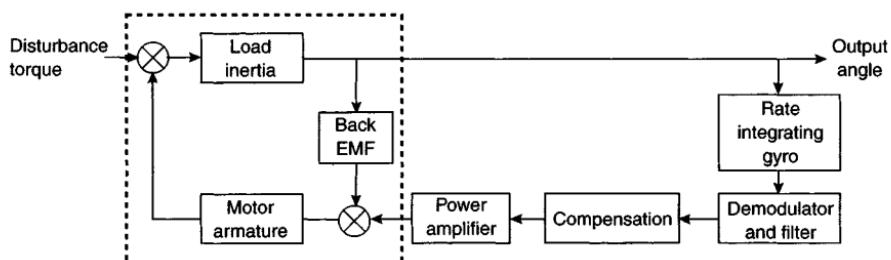


Figure 15.25 Servo-stabilisation loop

in inertial space. The output from the gyroscope's pick-off is modulated and filtered, passed through a compensation stage and applied to a brushless torque motor to correct any angular deflections. This servo loop may be considered to comprise a number of elements, the parameters of which are fixed, and with a compensating element. In this servo loop the fixed elements are the inertial load, the torque motor and the gyroscopic response. A configuration of the compensating element is virtually entirely 'free format' and represents the area where the servo designer can practise his or her skills.

The bandwidth of typical stabilised mirror systems is in the range of 30–70 Hz. Consequently, the servo bandwidth requirements for stabilised mirrors in this type of application are sufficiently high that the natural resonant frequency of a rate gyroscope would be a drawback and preclude its use. The majority of applications require bandwidths in the range of 120–150 Hz with a sharp cut-off beyond this frequency range.

Gyroscopes with low-noise characteristics are desirable to prevent jitter of the sightline and thus blurring the image. Additionally, any slight noise on the output signal can excite structural resonances within the mirror mechanism control system, which may lead to instability in the system response. The effect of resonant frequencies may be reduced by the use of notch filters and some noise may be removed by use of band-pass filters. However, the use of such devices also limits the dynamic performance of this type of system.

The drift characteristics of gyroscopes used in this type of application are particularly important for achieving the high performance required by high-resolution surveillance systems, particularly if operated 'open-loop', that is, without an auto-tracker [11]. In this type of application the field of view and the displayed image may be as small as one or two degrees. For those systems using a joystick for positional control to steer the sightline, the drift rate should be no worse than about a degree per minute. However, for EO systems, which use an auto-tracker to maintain the sightline on the target through a feedback system, a gyroscope with low or small drift characteristics is not required.

Passive methods are possible for reducing vibratory motion using anti-vibration mounts, as described in Chapter 9 for isolation of an IN system. However, these systems are not normally compact and require matching to the precise system being damped.

An alternative approach for line-of-sight stabilisation of equipment on platforms with a low degree of 'own-ship motion' is to use a sensor with a wider field-of-view and use electronic processing to remove the apparent motion of the target. This technique uses a frame-to-frame correlation of the successive images from the EO system. This approach is often used in target-tracking systems, but accurate stabilisation in a highly dynamic environment requires real-time precision processing and high-resolution resolvers to provide the accurate angular-data stream.

In many applications, such as EO systems for use on aircraft or ground-based vehicles, strapdown systems are the only possible feasible solution to the sightline isolation or stabilisation problem.

15.5.6 Relative angular alignment

A common requirement in various industrial processes and military systems is to measure the angle between two surfaces or edges, which are separated from each other, particularly where use of straightforward mechanical or even optical measurement devices is difficult, if not impossible. In some cases the references axes of one system may be moving randomly with respect to another, as occurs with flexure in large aircraft and ships.

This is an application where the use of satellite-based navigation techniques still requires development in order to provide accurate attitude data [19]. Some techniques are being investigated, but currently the standard inertial sensors provide the robust solution.

This section will consider the example of the problem of the measurement of the alignment, or relative misalignment, of the axes of rollers in large and complex machinery for steel or paper production. This problem is analogous to the surveying of the orientation to the relative static orientation of sub-systems, such as distributed sensors, in a platform such as a ship or an aircraft.

There are a number of possible solutions to this problem and three techniques will be considered here.

1. The first method is to place a gyro-compassing inertial system (see Section 10.2) on one surface, enabling its orientation with respect to the Earth's axes to be measured, and then to move the system to another surface, carrying out a fresh gyro-compassing routine. This approach is generally time-consuming, particularly if accurate estimates are required.
2. The second method is to mount a three-axis inertial navigation system on one surface, without gyro-compassing, and then to move the system to the other surface, logging the angular displacements that occur during the movement of the IN system to the new surface to be measured. During the movement from one surface to the other it is crucially important to ensure the angular rate limitations of the instruments in the IN system are not exceeded. In general, this approach is more time-efficient than the gyroscope-compassing method, in terms of operation of the IN system during the relative measurement process.

3. The third approach is the use of dedicated master and slave systems. This is the most rapid approach, although often the least cost-effective for many applications.

15.5.6.1 Gyro-compassing technique

The use of gyro-compassing techniques is well established, but currently, the only reported development of this technique is north-finding systems for military applications. A significant disadvantage of this approach, coupled with the need to use repeated gyro-compassing measurements, is the handling of the system. This type of system needs to be large, and is inevitably expensive, to achieve an angular accuracy of better than around 0.5° . One possible approach, which was investigated in the late 1980s, but never fully developed, is to use smaller, lower-cost instruments on a continuously rotating (or carouselling) cluster. This is similar to the techniques used in a number of inertial navigation stabilised platform systems to give high performance from lower-cost systems (and lower performance sensors) and is considered further in the application to ship's inertial navigation systems (SINS), described in Section 15.3.

In the case of this technique, generally only two 'gyroscope'- and two 'accelerometer'-sensing axes are required, all approximately perpendicular to the rotation axes. The output signals from the gyroscopes and the accelerometers will contain a noisy sinusoidal component, at the rotation frequency of the carousel. The sine-wave output from the gyroscopes passes through its mean value at the instant the gyroscope's sensing axis is perpendicular to the Earth's rotation axis. Additionally, the signal from an accelerometer shows a similar characteristic at the instant that its sensing axis is horizontal.

The success of this technique is critically dependent on the use of appropriate signal processing, such as filtering and curve fitting as required. Moreover, the random walk content of the gyroscope's output has the same constraints as that of any other gyro-compassing system regarding the time taken to achieve any particular orientation estimate accuracy. However, there is no requirement for the absolute drift or bias values to be known or calibrated, nor to be as stable as those required for use in a conventional navigation system.

15.5.6.2 Non-gyro-compassing technique

The use of a non-gyro-compassing approach has been developed by at least one company, Prüftechnik AG of Ismaning, Germany. This company specialises in high-accuracy alignment measurement devices for industry; in particular systems have been developed to measure the alignment of rollers in steel- and paper-making machinery. Their system (named 'Paralign'), uses three inertial-grade Honeywell RLGs, but does not require accelerometers, as basically this system simply measures and logs the angular change from one measurement made at one surface to that made at the next. It achieves a resolution of 4 micro-radians, and accuracy of the order of 16 μrad (about 3 arc s). The basis of this technique is shown in Figure 15.26.

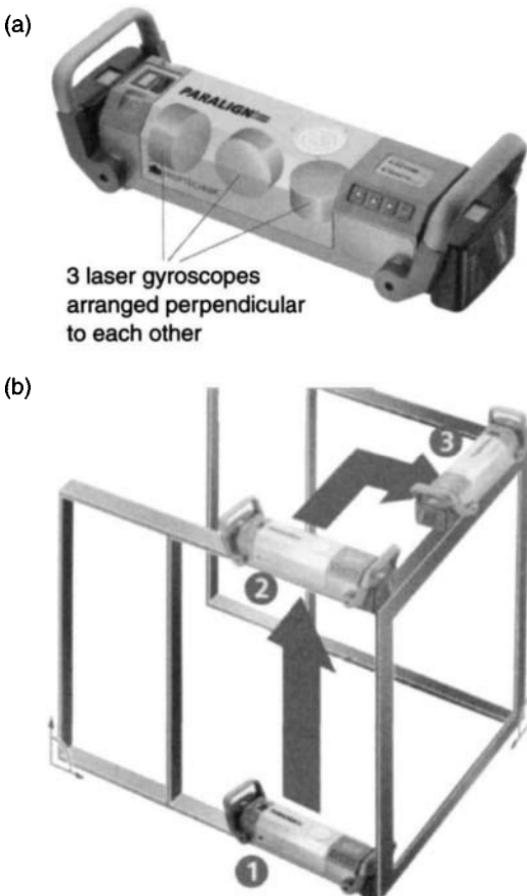


Figure 15.26 Non-gyro-compassing technique for measuring relative alignment (Courtesy of Prüftechnik)

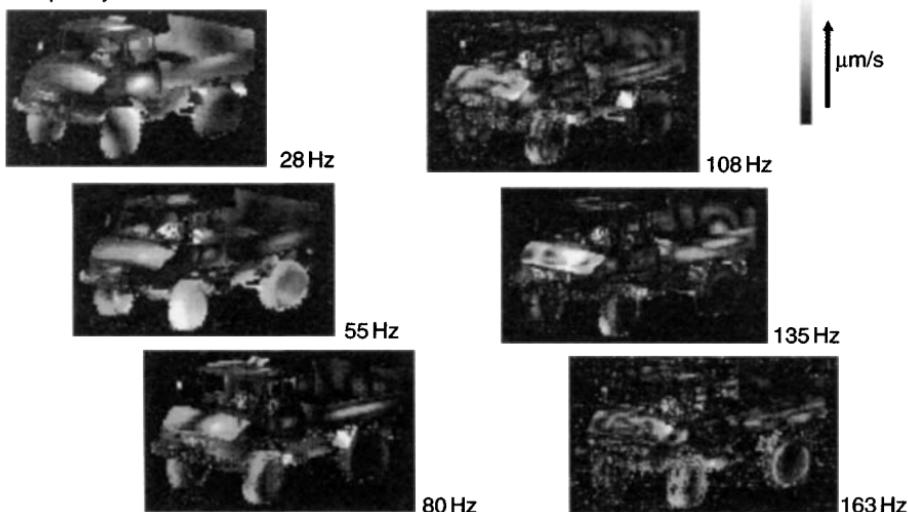
15.5.6.3 Master and slave systems

The fundamental aspects of this approach are covered in Section 15.5.1 on aero-flexure compensation, and in Chapter 10. In particular, this approach is most valuable when there is motion between the components of the complete system.

15.5.7 Calibration and measurement

Vibrometry is the study of the characteristics of vibratory motion and it is being investigated for exploitation as a technique to aid target recognition using laser-radar techniques. The basis of the vibrometry technique is considered here in terms of this military function, as all types of mechanised vehicles vibrate and consequently modify the frequency of the reflected laser light, owing to the physical motion of the reflecting surface. The magnitude of the frequency shift in the reflected light is

Frequency resolved

*Figure 15.27 Vibration spectra of vehicles (Courtesy of FGAN-FOM)*

governed by the vibratory motion and is usually small. This frequency shift in the reflected laser light may be detected by a number of well-established techniques.

The vibratory motion exhibited by a vehicle is a function of many variables, such as its type, style of construction and engine speed, so it is likely that each particular class of vehicle is likely to have a characteristic vibration spectrum, particularly in terms of its power spectral density. Hence, each class of vehicle, such as a 20 tonne truck made by a particular manufacturer, may have a characteristic frequency spectrum, which can be used for identification purposes, especially at very long ranges. Examples of a truck vibration spectra are shown in Figure 15.27.

For the technique to be applied to a recognition system it is necessary to create a database for each potential target. This process requires an extensive series of trials to be undertaken to measure the response of the various types of objects in a range of orientations and engine speeds of the test vehicle to incident probe energy from a given laser system. Moreover, it is essential to instrument the test objects to correlate the responses of the reflected laser energy with the actual motion of the object.

Piezo accelerometers have proved to be ideal sensors for this type of measurement; these sensors are very similar devices to those used to monitor other sensors under test during environmental testing in the laboratory. The input range of these sensors is in the $\pm 50\text{g}$ range, although a smaller input range of less than $\pm 10\text{g}$ would suffice, with sensitivity of the order of 0.1 mV/g . The bandwidth of the devices is usually about 7 kHz . Triple-axis devices are ideal for some aspects of this type of application.

Instrumentation packs can be manufactured to provide the 'truth' measurement. Each pack consists of an array of accelerometers, the conditioning electronics, data



Figure 15.28 Triple-axis piezoelectric accelerometer monitor

conversion and storage module and a power supply. The array of accelerometers is firmly attached to the vehicle and the defined measurement sequence is undertaken. It is necessary to ensure that the attachment of the sensors does not modify the natural response of the vehicle, so it is normal to apply strips of aluminium tape to the object at the various locations being ‘sensed’. The accelerometers can then be fixed into position using an epoxy glue; an example of a triple-axis sensor suitable for use in a measurement campaign, is shown in Figure 15.28.

15.6 Geodetic and geophysical measurements and observation of fundamental physical phenomena

An accurate model of the shape of the Earth is a very important aspect of any terrestrial navigation system, as has already been discussed in Chapter 3. However, fluctuations in the rotational characteristics of the Earth also have an impact on the potential accuracy of terrestrial navigation systems, including those that use precision GPS data. Measurement of other changes to the behaviour of the Earth may be used to predict seismic events, such as earthquakes.

Many factors produce small fluctuations in the characteristics of the Earth’s rotational parameters. The principal factors that induce these fluctuations are;

- continental drift;
- motion of the moon and its phases;
- movement of the subterranean magma in the Earth’s core;
- tides;
- weather.

Chapter 13 and Appendix D considers the operation and use of the popular satellite-based navigation systems, which offer the potential for high-precision navigation for all types of vehicles and platforms. The basis of the technique is the constellation of satellites in geo-stationary orbits locked in a particular known position above the Earth. Consequently, errors in the estimated position of a navigation



Figure 15.28 Triple-axis piezoelectric accelerometer monitor

conversion and storage module and a power supply. The array of accelerometers is firmly attached to the vehicle and the defined measurement sequence is undertaken. It is necessary to ensure that the attachment of the sensors does not modify the natural response of the vehicle, so it is normal to apply strips of aluminium tape to the object at the various locations being ‘sensed’. The accelerometers can then be fixed into position using an epoxy glue; an example of a triple-axis sensor suitable for use in a measurement campaign, is shown in Figure 15.28.

15.6 Geodetic and geophysical measurements and observation of fundamental physical phenomena

An accurate model of the shape of the Earth is a very important aspect of any terrestrial navigation system, as has already been discussed in Chapter 3. However, fluctuations in the rotational characteristics of the Earth also have an impact on the potential accuracy of terrestrial navigation systems, including those that use precision GPS data. Measurement of other changes to the behaviour of the Earth may be used to predict seismic events, such as earthquakes.

Many factors produce small fluctuations in the characteristics of the Earth’s rotational parameters. The principal factors that induce these fluctuations are;

- continental drift;
- motion of the moon and its phases;
- movement of the subterranean magma in the Earth’s core;
- tides;
- weather.

Chapter 13 and Appendix D considers the operation and use of the popular satellite-based navigation systems, which offer the potential for high-precision navigation for all types of vehicles and platforms. The basis of the technique is the constellation of satellites in geo-stationary orbits locked in a particular known position above the Earth. Consequently, errors in the estimated position of a navigation

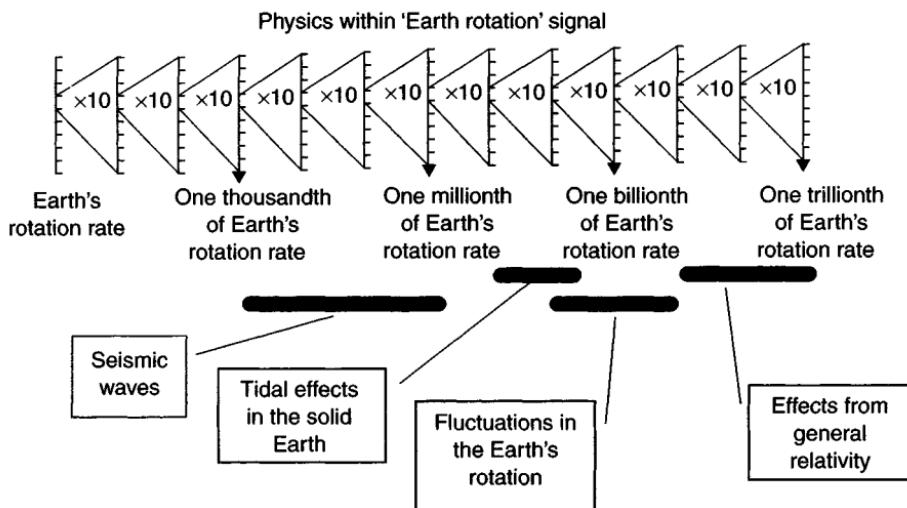


Figure 15.29 Scale of factors perturbing RLG measurements

system's receiver will occur owing to the displacement of the constellation of satellites, resulting from the fluctuations of the Earth's rotational motion. Hence, changes in these characteristics of the motion of the Earth need to be measured and understood, so that an appropriate compensation technique can be applied.

One technique that is being investigated for the measurement and monitoring of the Earth's rotation fluctuations involves the use of a very large ring laser gyroscope. A ring laser gyroscope is capable of providing more information about its experience in the environment than merely the angular motion imparted to it, that is, it is potentially much more than a gyroscope. The main contribution to the Sagnac frequency detected by a stationary ring laser gyroscope is the Earth's rotation rate, provided the input axis is aligned with the rotation axis of the Earth. Very accurate ring laser gyroscopes are capable of measuring other effects that perturb the sensor and thus modify the Sagnac frequency. The scale of some of these effects is illustrated in the schematic diagram of Figure 15.29.

A research group at the University of Canterbury in New Zealand, in collaboration with others, has pioneered studies into these topics and exploited the properties of this type of device through the construction of very large ring laser gyroscopes [20]. Currently there is a project involving collaboration between this group and the German Federal Cartography and Geodesy Office in Frankfurt, Germany.

Measurements are made of the rotation rate of the Earth using very large ring laser gyroscopes, typically having a perimeter of many metres. The enclosed area for the prototype was $\cong 3.5 \text{ m}^2$, but the latest one is $\cong 16 \text{ m}^2$ and is known as the gross ring (G-ring), with an even larger device having an enclosed area of 367 m^2 being commissioned, in this case the structure is made of basalt. The lasers in these sensors use the helium-neon '633 nm transition' and the population inversion for this transition is created by a radio frequency source, which is carefully controlled



Figure 15.30 A large ring laser gyroscope (Courtesy of University of Canterbury, NZ)

to avoid unwanted laser-mode generation. Consequently, unwanted mode-pulling effects in the laser cavity are avoided.

Careful control of the backscatter from the mirrors, with a total optical loss in the part per million category, enables the cavity to be unlocked by use of the Earth's rotation (7.293×10^{-5} rad/s). These lasers combine very high cavity finesse, through low total reflection losses, and large enclosed area to give excellent frequency resolution of better than one part in 3×10^{21} and position accuracy of 300 prad, with the enhanced frequency resolution leading to sub-microhertz resolution of the beat frequency.

The Fourier transform of the output from the laser cavity gives a spectral line associated with the rotation of the Earth often termed 'the Earth line'. Analysis of this spectral line is used to monitor the impact of seismic waves and other events. This ability to make high-precision measurements of rotation rates has led to applications in geodesy.

At the heart of this latest measurement system is a 4.25 m diameter disc of zero-expansion glass that is 25 cm thick and forms the stable ring laser cavity. The rate of expansion is quoted as 60 nm/ $^{\circ}\text{C}$ change in temperature. The 'G-ring' is believed to be the largest and most accurate ring laser in the world [21]. Its enclosed area is 16 m^2 and consequently is about 1000 times more accurate than the RLGs used in commercial aircraft. A photograph of the large RLG is shown in Figure 15.30.

In order for these sensors to achieve the full potential of their performance these ring laser gyroscopes are located in a very carefully controlled environment, such as a cavern; the G-ring RLG is at the Wettzell facility at Kötzing in Germany, others are at Cashmere in New Zealand. The 'G-ring' sensor is mounted on a granite slab that is 60 cm thick and has a mass of 10 tonnes, all embedded in a concrete base in an under-ground laboratory. Moreover, the laser system is encapsulated in a steel tank

with a sophisticated control system to minimise ambient temperature and pressure fluctuations, thus ensuring maximum measurement accuracy from the sensor.

Techniques have been developed to enable the use of these large RLGs for studying a number of geophysical effects; particularly those associated with earthquakes. Of particular interest is the relationship between the ground rotational effects in a seismic event and induced rotational effects in buildings as they are generally particularly vulnerable to this type of motion. There is particular interest in the frequency dependence of responses of buildings in the 0.2–30 Hz band [22]. The effect of seismic events is to induce frequency-modulated side bands, in the 0.2–1 Hz region, around the ‘Earth line’, which indicate the presence of rotational components associated with seismic events.

The ‘G-ring’ will be used for the measurement of short-term fluctuations of the Earth’s rotation rate, with periods in the range of hours to days. Periodic geophysical signals with periods of around one day and half a day have become visible in the time series of ring laser observations. These signals are introduced by variations in the orientation of the ring laser plane and currently limit the uncorrected sensor resolution to 2 parts in 10^8 with respect to the Earth’s rotation rate.

A GEOSENSOR [21] is being established to measure earthquakes. The GEOSENSOR consists of:

- a large single-axis ring laser gyroscope;
- a conventional three-axis broadband seismometer;
- a tiltmeter to measure changes in the orientation of the ring laser gyroscope;
- GPS time receiver to provide the time reference for the data acquisition system;
- auxiliary equipment and instrumentation to monitor the performance of the ring laser gyroscope and other equipment.

Currently, a demonstration system is under construction at the geodetic observatory at the Fundamentalstation Wettzell in Bavaria, Germany.

Alternative approaches to the use of very accurate inertial sensors to measure imperfection in the Earth’s rotation rate include the use of radio telescopes that measure the fluctuation characteristics relative to a very distant inertially stable object such as a quasar. However, this measurement is complex and the measurements are dependent on a network of radio telescopes located around the world.

The large RLG has detected an Earth tide signal at the lunar tidal period of 12 h 25 min [23]. The amplitude of this signal is one-millionth of Earth’s rate $\approx 15 \times 10^{-6}^\circ/\text{h}$. This effect confirms other observations with large ring interferometers. These measurements are made from the changes in the orientation of the sensor resulting directly from lunar tidal effects and atmospheric loading. The strong tidal effects result from ocean-induced loading near the location of the sensor. It is anticipated that the larger ring-laser sensor under construction, this device will enable other effects, such as atmospheric pressure fluctuations at a lower level to be detected. Moreover, the effects of polar motion are expected to be detected by the more sensitive devices.

The large RLGs have been proposed for study of other physical phenomena, such as time-reversal invariance, or both parity and time-parity symmetry. The large ring laser cavity lends itself to this type of study as its beat frequency detects any effects

that contribute differently to the clockwise and counter-clockwise propagating beams in its cavity [24].

Investigation of the Lense-Thirring field, or reference frame drag [25], is an example of using the very large ring laser gyroscopes to detect fundamental physical phenomena. This is a relativistic effect that involves the coupling of an electromagnetic field owing to gravity. Lense and Thirring [26] showed, from general relativity, that the local inertial frames near a rotating mass rotated relative to those at infinity. A very popular term for the Lense-Thirring effect is inertial-frame dragging. This effect is also touted as a particularly characteristic and direct manifestation of the graviomagnetic effects associated uniquely with general relativity.

A number of proposals are being developed to devise experiments that would enable the very large RLGs to be used to measure the Lense-Thirring effect. This involves defining a very large RLG in the reference frame of the stars to very high precision and involves extreme metrology to enable extremely precise definition of the ring laser area normal vector to be defined. This would require an optical stellar interferometer to be positioned very close to the sensor, to provide the requisite attitude of the sensor in inertial reference axes.

15.7 Other applications

A number of additional applications, not covered by the categories of system described in the preceding sections, are discussed in the following section.

15.7.1 Moving-map displays

The principle of a moving-map display is based on terrain referenced navigation techniques and centres on the correlation of inertial positional and attitude data with digitised stored-map data. It is used as an aid to aircrew, being particularly valuable when flying in poor visibility or at night.

At the heart of this type of system is a mass data store, which may comprise the following:

- the flight plan of the proposed route or mission, which would include waypoints, routeing and timing cues;
- hazard data that may include potential constraints or intelligence information for military applications, which could incorporate data about locations of known missile sites, current battle areas, as well information about obstructions, such as pylons, chimneys and masts;
- cultural and environmental data, including diurnal information, which would include location of bridges, railways, rivers and roads;
- elevation profile data.

The flow of information in the database for this system is illustrated in Figure 15.31.

The other vital element of this navigation aid is the IN system, which normally includes aiding of the vertical channel with a barometer, and a precision altimeter.

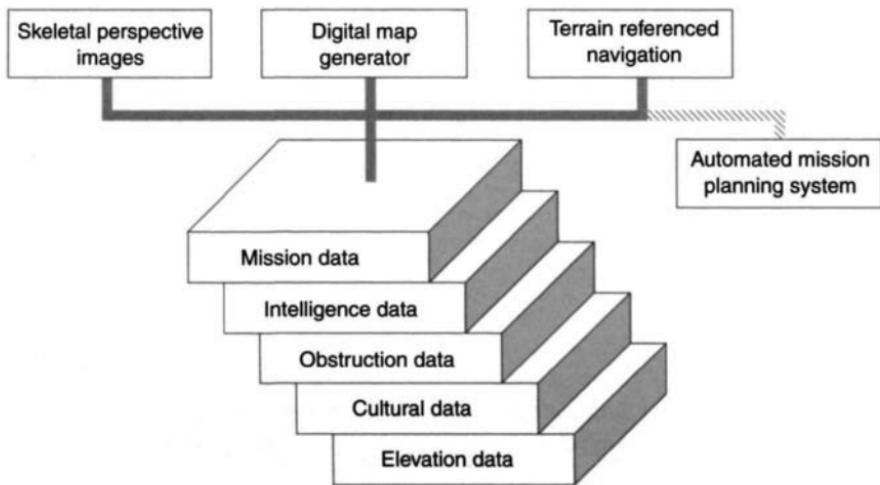


Figure 15.31 Digital database information flow (Courtesy of BAE Systems)

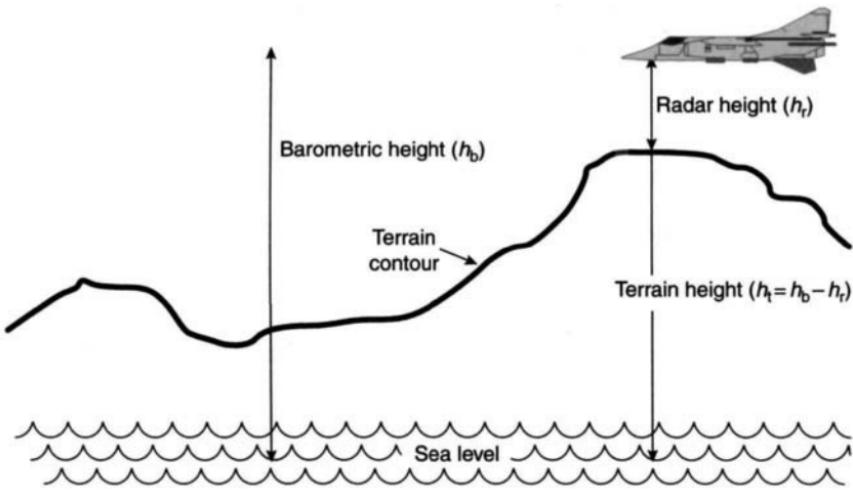


Figure 15.32 Height profiling

This type of system enables a measurement to be made of the aircraft height above the ground, thus creating a profile of the terrain below. The profile is formed by calculating the difference between the IN-indicated height and the height above ground measurement given by the altimeter, as shown in the Figure 15.32.

Comparison of the time-series profile with the terrain-profile information stored in the database enables the exact position of the navigation system to be determined. The position of the aircraft is usually superimposed at its identified position on the map. The map projection may move so that the terrain below the aircraft 'travels' with the aircraft; additionally the position of the aircraft symbol may be selected so

that it is at the correct position with respect to its actual position on the appropriate projected image. This is analogous to moving a window over a scene.

This type of display is very flexible so the scale of the image can be increased to enable additional detailed features to be included, or diminished (i.e. a zoom function) to provide a large-scale view. Modern computer processing and image processing algorithms allow vast flexibility in the variety of information on the projected map, such as natural cultural features. The superimposed cultural information can be based on feature vectors or alternatively from digitised aeronautical maps.

This application is also common in commercial aircraft where the IN system data, from the aircraft's navigation system, are used to provide passengers with an update of progress on their journey, as part of the entertainment system. In this case far less precision is required, so a much simpler system can be implemented without the need for the recourse to precision terrain referenced navigation, to provide the appropriate reference. It is common for the system to add additional features or information such as:

- place names and destination;
- time to go and current time at destination and starting point;
- velocity and altitude of the aircraft;
- distance to go;
- flight path and orientation of the aircraft;
- culture of the land masses;
- shadow indicating night time.

In military applications, this approach may be used in automated mission planning and execution of tasks to reduce the workload of the crew.

Moving-map navigation systems are becoming a common accessory in motor vehicles. These are based on a 'GPS' receiver and stored database in a baseline system. A more accurate system is possible when the output signal from the vehicle's odometer is used to aid the basic system, thus providing an integrated navigation system. Motor car navigation systems are considered in more detail in Section 15.4.3.

Recent developments of moving maps have seen a satellite navigation system (GPS) linked to a stored digital map and a Braille display on a computer to help visually impaired people. A commercially available product enables this class of navigation system to provide a number of types of aid, such as directions or a count down to junctions, a destination or points of interest. However, these types of device are limited to outdoor use, as the 'GPS signals' do not penetrate buildings.

A form of integrated system that combines motion and direction sensors with a satellite-based navigation system is being developed to help blind people navigate inside as well as outside buildings. In this case, the motion sensors are used when satellite-based navigation systems, such as GPS, are not available.

The motion sensor pack contains:

- a digital magnetic compass;
- a barometer;
- a gyroscope.
- an odometer (pedometer)

The motion data are used in conjunction with a pattern recognition algorithm to determine the characteristics of the user's step pattern. Dead-reckoning techniques (see Section 2.1) may then be applied to calculate the user's position. In the case of advanced systems this course may be plotted on a stored plan of a building or a map of any location, urban or rural.

An example of its operation could involve a user marking their position on entry to a building and then using the track mode to plot their movement and provide markers on the plan of the building, within the navigational aid. The stored information may then be used to guide the person to the exit, or any other chosen destination known to the database. Additional navigation and guidance capability would be possible if the integrated system had a stored map of the interior of the building. This would not only enable the user to be guided to particular locations, but also allow regular correction of navigation errors from way marking as discussed in Chapter 13.

15.7.2 Safety and arming units

A safety and arming unit is a device that ensures a missile has been launched successfully and is clear of its launch point before the warhead is armed ready for activation of its detonation train by its fuze. Clearly, the fundamental requirement is for the launch point to be outside of its warhead's lethal radius before the arming of the system is implemented. It is conventional for the safety and arming unit to have received positive responses from two independent channels before it activates 'the switch' to arm the warhead.

Accelerometers are commonly used as motion sensors for this type of application. Very simple sensors are adequate to measure motion along the longitudinal axis of the missile, especially for those weapons launched from a trainable launcher. In this case the sensor can be a simple displacement device, where the seismic mass is displaced axially when the longitudinal acceleration exceeds a given value, normally close to the peak acceleration expected from a nominal launch of the missile as it is boosted from a launch rail. The displacement of the seismic mass provides the actuation of one channel of the safety circuit, so closing a switch. An independent technique, such as elapsed time, may be used for the second channel.

The safety and arming unit for a vertically-launched missile is potentially more complex, as it is crucial to ensure that the weapon has made the appropriate manoeuvre before arming its warhead. In this case the inertial measurement unit can provide the requisite data to indicate a successful turnover manoeuvre has been completed. These data can then be used to identify the position when the missile has moved beyond the safety zone around the launch canister by applying simple navigation algorithms. In this case relatively low accuracy sensors are required. The logic of the double switch system is shown in Figure 15.33.

In the case of command-guided missiles the arming function can be activated once the missile has been 'gathered' by the ground-based tracking system and is being guided towards its target. The ground-based tracker and the longitudinal acceleration measurement may be used to provide the independent channels for arming the lethal

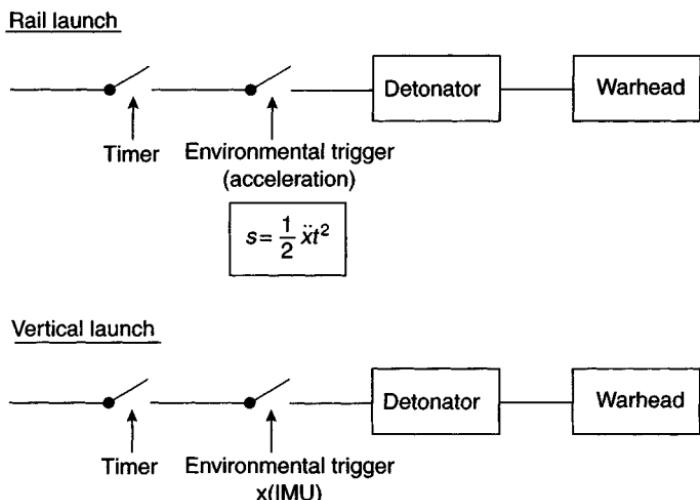


Figure 15.33 Safety and arming unit logic

payload. Figure 15.34 shows the position of a compact safety and arming unit within a nose cone of a shell, where it is an integral part of the system's fuze.

15.7.3 Aircraft ejection seats

The use of ejection seats in fast jet aircraft, and many other military combat aircraft, has saved the lives of many thousands of aircrew since the invention of this safety system more than 50 years ago. As the combat aircraft have become more sophisticated the demands on the ejection seat technology have also increased dramatically, particularly if the ejection process occurs close to the ground and the stricken aircraft is out of control, so that it may be rolling rapidly. Clearly, the use of a simple ballistic ejection technique is not adequate in these circumstances if the crew are going to have a high probability of surviving the ejection from the aircraft.

In the case of any form of rolling motion of the airframe, it is crucially important that the ejection of the seat and its occupant does not occur when the aircraft is inverted and close to the ground. In this case the aircraft's IN system can organise the ejection process so that the aircrew and their seats are directed into the upper hemisphere away from the ground, as the IN system will know 'which direction is up'.

The use of an inertial measurement unit on each ejection seat offers the opportunity to control the direction of the ejection if the ejection motor or the seat assembly has some form of control system to manoeuvre the seat and its occupant. A favoured technique would involve thrust vector control of the ejection-seat motor. This approach could manage the direction of the ejection-motor thrusts, and consequently the direction of the trajectory of the seat, to ensure the aircrew has an optimum chance of survival. This type of control system also offers the opportunity to manage the ejection process and minimise the impact of the ejection-induced shocks on the occupant

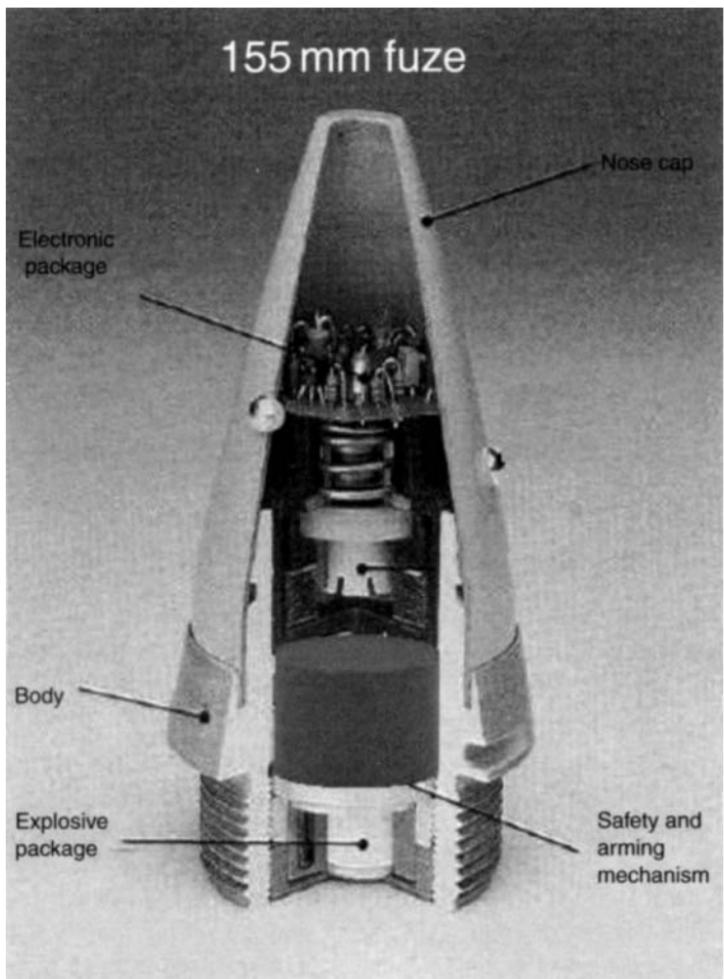


Figure 15.34 Safety and arming unit in a shell

as well as determining when it is best for the occupant to separate from their seat and descend back to Earth on their parachute.

A simple IMU may also be used to direct the trajectory, as well as controlling the thrust dynamics, of each of the ejection seats during a multiple-seat ejection. This would reduce the probability of the seats colliding or interfering with each other during the ejection process from the aircraft.

The requirements on the inertial sensors are not particularly stringent in terms of the performance accuracy; the devices are only required to operate over a relatively short period, consequently low-performance devices will be adequate. The most demanding requirement is for the sensors to withstand the shocks imparted to the seat system during the ejection process.

15.7.4 Agricultural survey

The use of precision navigation techniques has been applied to the farming [27–29] industry in order to optimise the utilisation of the land. A number of institutions are devising techniques based on precision inertial navigation methods, with differential GPS (described in Section 13.3.2), to use the great positional accuracy for guiding the machinery for ploughing and other functions, from planting to harvesting.

Many combine harvesters throughout the world carry a range of sophisticated equipment to analyse the yield of a crop as it is gathered. An inertial navigation system, with GPS aiding and a reference, may provide a data stream with the co-ordinates of the vehicle, so that the yield can be correlated with an area of the field. Processing of the yield and the associated positional data can enable optimised performance to be achieved in the future, for example, by providing information on where additional fertiliser is required, or where an alternative crop may be successful.

A further example is the automatic control of the tractor during the sowing of seeds in very large fields to enable accurate and efficient planting without missing zones or ‘double’ planting. This may be considered a form of advanced cruise control for tractors.

15.7.5 Artillery pointing

An enduring requirement of military forces is to know precisely where they are within the battle space, that is, having the ability to establish a precise geographic reference for each system rapidly. This is particularly true for those units involved in the indirect battle, for example artillery batteries using systems with unguided projectiles.

Up to the 1970s, the deployment of artillery required many hours (or possibly days) of work with theodolites to survey the position of the intended gun battery system, prior to occupation. Once the survey had been completed and the battery centre determined, the guns were then aimed along an azimuth determined by magnetic-heading measurement, or, if cost and time permitted, by alignment/orientation transfer from a gyroscopic theodolite.

By the late 1970s an automated survey system known as ‘PADS’ (position and azimuth determining inertial systems) was developed in the United States and the United Kingdom, which replaced the increasingly outdated method of theodolite survey. PADS allowed a mobile artillery battery to set up, aim, fire its guns in a relatively short period, of the order of 90 min. The introduction of PADS heralded the dawn of mobile artillery operations and gave batteries the unprecedented ability to manoeuvre, set up and fire one accurate salvo, and depart before the enemy could retaliate, a philosophy known as ‘shoot and scoot’.

PADS works using a conventional inertial navigation system, and incorporates an inertial grade floated rate-integrating gyroscope with bias of about $0.005^\circ/\text{h}$. The system is aided by use of a ‘zero-velocity-updating’ algorithm, which allows it to measure and model its error states when it is known to be stationary. With a low-noise



Figure 15.35 *Position and azimuth determining inertial system (Courtesy of BAE Systems)*

gimballed system, errors in positional estimates can thus be reduced to centimetres, and heading orientation can be estimated to fractions of a milliradian. However, a PADS is both heavy and expensive, and generally could only be deployed on a one-per-battery basis, although in a few cases, armies have used them on a one-per-gun basis. A picture of PADS is shown in Figure 15.35.

The development and widespread availability of inexpensive and accurate satellite-based navigation systems has now generally superseded the use of PADS for survey. Indeed GPS and similar technology is sufficiently cheap to enable the fitting of GPS receivers to individual guns. However, there are, as yet, no effective replacements for the use of inertial techniques for rapidly determining azimuthal orientation accurately.

Nowadays, a strapdown system, aided by odometer and/or GPS, with zero-velocity updating, can provide the required azimuth accuracy with a cost and weight of about a quarter of that of earlier-generation PADS. Moreover, the modern system may also be sufficiently light and rugged to be mounted directly on the trunnion, giving accurate elevation measurement as well. A typical example is the BAE Systems FIN 3110 system used on the British Army's new Light Gun, and in several other similar applications. A picture of a gun system is shown in Figure 15.36.

A few other complications associated or encountered with the typical implementation of these systems need to be compensated for. First, the problem of 'track slip' in tracked vehicles where an odometer is used as an aiding source during the navigation phase. Second, the problem of the shock (up to several hundred g's) transmitted at the instant of firing – well beyond the linear range of inertial-grade accelerometers. Algorithms to overcome these problems of track slip are now mature and robust.

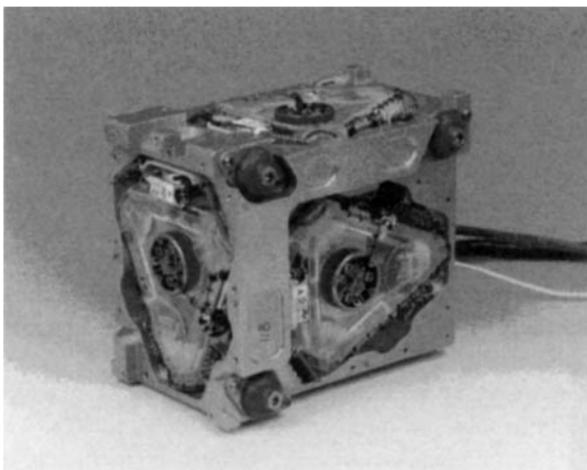


Figure 15.36 Ring laser gyroscope system for artillery use (Courtesy of BAE Systems)

15.7.6 Other unusual applications

The use of inertial sensors and navigation techniques continues to grow rapidly, particularly with the reduction in cost of the sensors and the wide availability of satellite navigation systems. These developments have encouraged the application of these techniques to conservation, recreation, resource management search and rescue and transport. Some examples of these established and proposed applications are considered below:

- survey of forests to locate specific species;
- monitoring of oil slicks and spills;
- survey of remote areas for location of specific features, such as those suitable for archaeological investigation;
- tracking animals during migration or at other times, for example, during release from captivity;
- an aid for hikers in unfamiliar territory;
- an aid for golfers to determine the distance to the next hole;
- provide the proof that a glider pilot has followed a particular course and reached the turning points;
- marking points for further investigation, such as treasure hunting;
- automatic collection of tolls from vehicles fitted with the appropriate systems;
- position marking in remote areas for dispatch of emergency services;
- monitoring the position and operational use of road vehicles in a fleet;
- precision positioning of equipment, such as excavation or drilling equipment in remote or featureless terrain, including on the seabed;
- tunnelling aids for projects under the sea or through mountains;

- stabilisation of hand-held optical devices and equipment, such as video cameras, binoculars and telescopes.

15.8 Concluding remarks

The range of applications which make use of inertial sensor technology is extremely broad, and is expanding rapidly, as illustrated by the examples given in this chapter. In designing systems for these varied roles, it is essential to consider the full context of the application, taking careful account of issues such as dynamic measurement range and the full range of environmental factors that may have a major impact on the design.

In many of these new or novel applications the catalyst for their development has been the availability of low-cost miniature inertial sensors that offer high reliability and require little or no maintenance. In general, these devices are rugged so that they can be used in relatively hostile environments although, quite often, the measurement accuracy is mediocre. However, the quality of the sensor performance has proved to be well matched to the fundamental requirement, that is, only an indication of angular rate was required to fulfil the task.

The development of integrated navigation techniques, particularly with IN sensors and satellite-based navigation systems, has led to devices that provide accurate navigational aids at a low price. This approach has seen the integrated navigation systems and devices displace high-performance IN systems from some traditional applications, a trend that is likely to continue.

References

- 1 GOETZE, D., HOHNER, G., McROBBIE, D., WESTON, J., and GÜNTHER, M.: 'Aided strapdown inertial navigation systems for the oil, gas, coal and construction industry'. Proceedings of symposium *Gyro Technology*, Paper 19, Stuttgart, Germany, 2001
- 2 McROBBIE, D., and WESTON, J.: 'Aided inertial navigation system'. US Patent No. US 6,145,378, November 2000
- 3 MARLAND, P.: 'The NATO ships inertial navigation system (SINS)', *Journal of Naval Engineering*, 1992, **33** (3), pp. 688–700
- 4 GARRELL, P., and EAST, D.J.: 'Guided weapon control systems' (Pergamon Press, 1977)
- 5 NASON, M.L., BROWN, C.A., and ROCK, R.S.: 'An evaluation of a rolleron-roll-rate stabilization system for a canard missile configuration at Mach numbers from 0.8 to 2.3'. NACA report (research memorandum) RM L55C22 September 1955
- 6 KRAKIWSKY, E.J., and BULLOCK, J.B.: 'Smart vehicles: digital road data – putting GPS on the map', *GPS World*, 1995, **5** (5), pp. 43–45
- 7 HONEYWELL, T.: 'Steady progress', *Professional Engineering*, 2003, **16** (4), pp. 47–48

- 8 TRANCHITA, C.J.: 'Active infrared countermeasures', in POLLOCK, D.H. (Ed.): 'Infrared and electro-optical systems handbook vol. 7', Chapter 3 (ERIM and SPIE Optical Engineering Press, 1993)
- 9 FULGHUM, D.A.: 'Laser can foil SAMS and air-to-air missiles', *Aviation Week & Space Technology*, 2001, **154**, pp. 43–45
- 10 WEBB, C.E., and JONES, J. (Eds.): 'Handbook of lasers & applications' (IOP Publications, Bristol, 2003)
- 11 NASBURG, R.E.: 'Tracking and control systems' in DUDZIK, M.C. (Ed.): 'Infrared and electro-optical systems handbook vol. 4', Chapter 5 (ERIM and SPIE Optical Engineering Press, 1993)
- 12 JALALIAN, A.V.: 'Laser radar systems' (Artech House, Norwood, MA, 1992)
- 13 MARTIN, N.J.D., and CHAZELLE, X.: 'Obstacle warning laser radar for fixed and rotary winged aircraft beyond CLARA', private communication, 2004
- 14 MARTIN, N.J.D., and CHAZELLE, X.: 'CLARA – A Franco-British co-operative development of airborne laser radar', private communication, 2004
- 15 SCHULTZ, K.R., and BIERS, K.: 'High reliable laser sensor for military applications'. Private communication, 2004
- 16 BÜCHTEMANN, W., and EBERT, R.: 'Laser based obstacle warning sensor for helicopters'. AGARD Conference Proceedings, 25–27 October, 1994, p. 561, Italy, 1994
- 17 WAGGONER, B.A.: 'A comparison of gyroscope digital models for an electro-optical/infrared guided missile simulation', NavSea Crane Division, Surface Warfare Center, NSWCCR/RDTR-03/12, Master's thesis, February 2003
- 18 MOORE, P.G.: 'Applications of inertial sensors for stabilisation of electro-optic sightlines'. Symposium proceedings, *Technology of inertial sensors and systems* (Institution of Mechanical Engineers, London, UK, 1991)
- 19 LAPACHELLE, G.: 'Attitude determination'. AGARD Lecture Series 207, 'System implications and innovative applications of satellite navigation', Paper 10, June 1996, pp. 10-1–10-11
- 20 ROWE, C.H., et al.: 'Design and operation of a very large ring laser gyroscope', *Applied Optics*, 1999, **38** (12), pp. 2516–2523
- 21 SCHREIBER, U., VELIKOSELTSOV, A., STEDMAN, G.E., HURST, R., and KLÜGEL, T.: 'New applications of very large ring lasers'. Proceedings of Symposium on *Gyro Technology*, Paper 8, Stuttgart, Germany, 2003
- 22 PANCHA, A., et al.: 'Ring laser detection of rotations from teleseismic waves', *Geophysical Research Letters*, 2000, **27** (21), pp. 3553–3556
- 23 SCHREIBER, K.U., KLÜGEL, T., and STEDMAN, G.E.: 'Earth tide and tilt detection by a ring laser gyroscope', *Journal of Geophysical Research*, 2003, **108** (B2), pp. 2132–2137
- 24 STEDMAN, G.E., et al.: 'T violation and microhertz resolution in a ring laser', *Optics Letters*, 1995, **20** (3), pp. 324–326
- 25 STEDMAN, G.E., SCHREIBER, K.U., and BILGER, H.R.: 'On the detectability of the Lense-Thirring field from rotating laboratory masses using the ring laser gyroscope interferometers', *Classical and Quantum Gravity*, 2003, **20**, pp. 2527–2540

- 26 LENSE, J. and THIRRING, H.: *Physikalische Zeitschrift*, 1918, **19**, pp. 156–163
- 27 PALMER, R.J.: ‘Techniques for navigating in a farm field’, *Navigation*, 1989–90, **36** (4), pp. 337–344
- 28 DEDES, G., and GOAD, C.: ‘Real-time cm-level GPS positioning of cutting blade and earth-moving equipment’. Proceedings of the ION, National Technical meeting January 24–26 (San Diego), Institute of Navigation (Alexandria Virginia, USA), 1994, pp. 589–594
- 29 LACHAPELLE, G., *et al.*: ‘GPS system integration and field approaches in precision farming’, *Navigation*, 1994, **41** (3), pp. 323–335

Bibliography

- GARNELL, P., and EAST, D.J.: ‘Guided weapon control systems’ (Pergamon Press, Oxford, 1977)
- LOCKE, A.S.: ‘Guidance’, in GRAYSON, M. (Ed.): ‘Principles of guided missile design’ (Van Nostrand, Princeton, NJ, 1958)
- BLAKELOCK, J.H.: ‘Automatic control of aircraft and missiles’ (Wiley, New York, 1991)
- MacKENZIE, D.: ‘Inventing accuracy, a historical sociology of nuclear missile guidance’ (The MIT Press, Cambridge, MA, 2000)
- STEDMAN, G.E.: ‘Ring-laser tests of fundamental physics and geophysics’, *Reports on Progress in Physics*, 1997, **60**, pp. 615–688
- JALALIAN, A.V.: ‘Laser radar systems’ (Artech House, 1992)
- AGARD Lecture Series 207. ‘System implications and innovative applications of satellite navigation’, June 1996
- <http://people.howstuffworks.com/sidewinder.htm>

Appendix A

Kalman filtering

The combination of independent estimates

It is shown how best to combine two independent estimates of a variable to form a weighted mean value, this operation being central to the process of Kalman filtering [1, 2]. The following development assumes only a knowledge by the reader of elementary statistical principles. A full mathematical derivation of the Kalman filter is beyond the intended scope of this book. Therefore, the reader interested in a mathematical treatise on the subject is referred to the excellent text by Jazwinski [3].

The single-dimension case

Consider the situation in which two independent estimates, x_1 and x_2 , are provided of a quantity x , where σ_1^2 and σ_2^2 are their respective variances. It is required to combine the two estimates to form a weighted mean, corresponding to the ‘best’, or minimum variance, estimate, \hat{x} . In general, the weighted mean may be expressed as:

$$\hat{x} = w_1x_1 + w_2x_2 \quad (\text{A.1})$$

where w_1 and w_2 are the weighting factors and $w_1 + w_2 = 1$. The expected or mean value of \hat{x} , written as $E(\hat{x})$, is given by:

$$E(\hat{x}) = w_1E(x_1) + w_2E(x_2) \quad (\text{A.2})$$

The variance of a quantity x is defined to be $E[(x - E(x))^2]$. Hence the variance of \hat{x} , denoted σ^2 , may be written as:

$$\begin{aligned} \sigma^2 &= E\{(w_1x_1 + w_2x_2 - w_1E(x_1) - w_2E(x_2))^2\} \\ &= E\{w_1^2(x_1 - E(x_1))^2 + w_2^2(x_2 - E(x_2))^2 \\ &\quad - 2w_1w_2(x_1 - E(x_1))(x_2 - E(x_2))\} \end{aligned} \quad (\text{A.3})$$

Since x_1 and x_2 are independent, $(x_1 - E(x_1))$ and $(x_2 - E(x_2))$ are uncorrelated, $E\{(x_1 - E(x_1))(x_2 - E(x_2))\} = 0$. Hence σ^2 can be expressed as follows:

$$\begin{aligned}\sigma^2 &= w_1^2 E\{(x_1 - E(x_1))^2\} + w_2^2 E\{(x_2 - E(x_2))^2\} \\ &= w_1^2 \sigma_1^2 + w_2^2 \sigma_2^2\end{aligned}\quad (\text{A.4})$$

Writing $w_2 = w$ and $w_1 = 1 - w$, the variance σ^2 may be expressed as:

$$\sigma^2 = (1 - w)^2 \sigma_1^2 + w^2 \sigma_2^2 \quad (\text{A.5})$$

The value of w , which minimises σ^2 , is obtained by differentiating the above equation with respect to w . Hence,

$$\frac{d\sigma^2}{dw} = -2(1 - w)\sigma_1^2 + 2w\sigma_2^2 = 0$$

which yields the optimum weighting factor as:

$$w = \frac{\sigma_1^2}{\sigma_1^2 + \sigma_2^2} \quad (\text{A.6})$$

Substitution in eqns. (A.1) and (A.5) gives \hat{x} and its variance σ^2 :

$$\hat{x} = \frac{\sigma_2^2 x_1 + \sigma_1^2 x_2}{\sigma_1^2 + \sigma_2^2} \quad (\text{A.7})$$

$$\sigma^2 = \frac{\sigma_1^2 \sigma_2^2}{\sigma_1^2 + \sigma_2^2} \quad (\text{A.8})$$

By following this process, the two independent estimates, x_1 and x_2 , have been combined to form a weighted mean value, in which the weighting factor has been selected to yield a mean with minimum variance, and hence, maximum probability. In a Kalman filter, one such estimate is usually provided by updating a previous best estimate in accordance with the known equations of motion, whilst the other is obtained from a measurement. If x_2 is taken to be a measurement which is used to improve an updated estimate x_1 , the above equations can be expressed in the following form:

$$\hat{x} = x_1 - w(x_1 - x_2) \quad (\text{A.9})$$

$$\sigma^2 = \sigma_1^2(1 - w) \quad (\text{A.10})$$

which shows how the estimate (x_1) and its variance (σ_1) are improved by the measurement (x_2). This derivation is now generalised to the multi-dimensional form necessary for a full Kalman filter implementation.

The multi-dimensional case

Consider now the situation in which \mathbf{x}_1 and \mathbf{x}_2 are n -element vectors representing two independent estimates of an n -dimensional vector quantity \mathbf{x} . The variances of \mathbf{x}_1 and \mathbf{x}_2 are represented by the two $n \times n$ matrices, \mathbf{P}_1 and \mathbf{P}_2 , respectively.

The weighted mean of \mathbf{x}_1 and \mathbf{x}_2 may be expressed in the same form as presented earlier for the single dimension case, viz:

$$\begin{aligned}\hat{\mathbf{x}} &= (\mathbf{I} - \mathbf{W})\mathbf{x}_1 + \mathbf{W}\mathbf{x}_2 \\ &= \mathbf{x}_1 - \mathbf{W}(\mathbf{x}_1 - \mathbf{x}_2)\end{aligned}\quad (\text{A.11})$$

where \mathbf{W} is an $n \times n$ weighting matrix and \mathbf{I} is a unit matrix. The best estimate of \mathbf{x} , denoted $\hat{\mathbf{x}}$, will be provided by the above equation when \mathbf{W} is selected to minimise the variance of $\hat{\mathbf{x}}$.

In most practical cases, the dimensions of the two estimates are not equal and one of them is often a function of the individual elements of \mathbf{x} . For example, a set of m measurements may be provided, denoted \mathbf{y}_2 , where \mathbf{y}_2 is related only to some of the elements of \mathbf{x} . In this situation, the relationship between \mathbf{y}_2 and \mathbf{x}_2 may be expressed as follows:

$$\mathbf{y}_2 = \mathbf{H}\mathbf{x}_2 \quad (\text{A.12})$$

where \mathbf{H} is an $m \times n$ matrix.

It is, therefore, required to form an optimum estimate of \mathbf{x} from one estimate \mathbf{x}_1 of variance \mathbf{P}_1 and a second estimate $\mathbf{y}_2 (= \mathbf{H}\mathbf{x}_2)$, with variance denoted here by the symbol \mathbf{R} . If we now let the weighting matrix $\mathbf{W} = \mathbf{K}\mathbf{H}$, where \mathbf{K} is another arbitrary weighting matrix, then:

$$\begin{aligned}\hat{\mathbf{x}} &= \mathbf{x}_1 - \mathbf{K}\mathbf{H}(\mathbf{x}_1 - \mathbf{x}_2) \\ &= \mathbf{x}_1 - \mathbf{K}(\mathbf{H}\mathbf{x}_1 - \mathbf{y}_2) \\ &= (\mathbf{I} - \mathbf{K}\mathbf{H})\mathbf{x}_1 + \mathbf{K}\mathbf{y}_2\end{aligned}\quad (\text{A.13})$$

By definition, the variance (\mathbf{P}) of $\hat{\mathbf{x}}$ is given by:

$$\mathbf{P} = E\{[\hat{\mathbf{x}} - E(\hat{\mathbf{x}})][\hat{\mathbf{x}} - E(\hat{\mathbf{x}})]^T\} \quad (\text{A.14})$$

Similar expressions can be written for the variances \mathbf{P}_1 and \mathbf{R} . Substituting for $\hat{\mathbf{x}}$ from eqn. (A.13) yields:

$$\begin{aligned}\mathbf{P} &= E\{[(\mathbf{I} - \mathbf{K}\mathbf{H})\mathbf{x}_1 + \mathbf{K}\mathbf{y}_2 - (\mathbf{I} - \mathbf{K}\mathbf{H})E\{\mathbf{x}_1\} - \mathbf{K}E\{\mathbf{y}_2\}] \\ &\quad - [(\mathbf{I} - \mathbf{K}\mathbf{H})\mathbf{x}_1 + \mathbf{K}\mathbf{y}_2 - (\mathbf{I} - \mathbf{K}\mathbf{H})E\{\mathbf{x}_1\} - \mathbf{K}E\{\mathbf{y}_2\}]^T\}\}$$

but since \mathbf{x}_1 and \mathbf{y}_2 are uncorrelated, this reduces to:

$$\begin{aligned}\mathbf{P} &= (\mathbf{I} - \mathbf{K}\mathbf{H})E\{[\mathbf{x}_1 - E(\mathbf{x}_1)][\mathbf{x}_1 - E(\mathbf{x}_1)]^T\}(\mathbf{I} - \mathbf{K}\mathbf{H})^T \\ &\quad + \mathbf{K}E\{[\mathbf{y}_2 - E(\mathbf{y}_2)][\mathbf{y}_2 - E(\mathbf{y}_2)]^T\}\mathbf{K}^T \\ &= (\mathbf{I} - \mathbf{K}\mathbf{H})\mathbf{P}_1(\mathbf{I} - \mathbf{K}\mathbf{H})^T + \mathbf{K}\mathbf{R}\mathbf{K}^T\end{aligned}\quad (\text{A.15})$$

It is now required to find the value of \mathbf{K} which minimises \mathbf{P} in the sense that the diagonal elements of \mathbf{P} , the variances of \mathbf{x} , are minimised.

It is shown in Reference 1 that this is achieved when

$$\mathbf{K} = \mathbf{P}_1 \mathbf{H}^T [\mathbf{H} \mathbf{P}_1 \mathbf{H}^T + \mathbf{R}]^{-1} \quad (\text{A.16})$$

Under such conditions, the best estimate of \mathbf{x} is given by:

$$\hat{\mathbf{x}} = \mathbf{x}_1 - \mathbf{K}[\mathbf{H}\mathbf{x}_1 - \mathbf{y}_2] \quad (\text{A.17})$$

and its variance is given by:

$$\mathbf{P} = \mathbf{P}_1 - \mathbf{K} \mathbf{H} \mathbf{P}_1 \quad (\text{A.18})$$

where \mathbf{K} takes the value given by eqn. A.16. The weighting process, defined by eqns. (A.16)–(A.18), is implemented in a Kalman filter as discussed in the following section.

The Kalman filter

In this section, we begin by considering the application of Kalman filtering to linear systems, before moving on to show the extensions which are necessary to overcome any system non-linearity. It is noted that the algorithms presented below are applicable to systems which are linear and time varying, not simply constrained to constant parameter systems. The explicit dependence of some of the parameters with time has been omitted to aid the clarity of the development.

Linear systems

The dynamical behaviour of a linear system may be represented by a set of first order differential equations of the form

$$\frac{d\mathbf{x}}{dt} = \mathbf{F}\mathbf{x} + \mathbf{G}\mathbf{u} + \mathbf{D}\mathbf{w} \quad (\text{A.19})$$

where the elements of the n -vector, $\mathbf{x}(t)$, are called the states of the system, $\mathbf{u}(t)$ is a p -vector of deterministic inputs and $\mathbf{w}(t)$ is the system noise. \mathbf{F} is an $n \times n$ matrix, known as the system matrix and \mathbf{G} is an $n \times p$ system input matrix. \mathbf{F} , \mathbf{G} and \mathbf{D} are constant or time-varying matrices. The noise $\mathbf{w}(t)$ has zero mean and is normally distributed (Gaussian), with a power spectral density of \mathbf{Q} .

Let us now assume that there are m measurements of the system, which are a linear combination of the states, but are corrupted with noise. This can be expressed in terms of the system states by the following equation:

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{n} \quad (\text{A.20})$$

Here, the m -vector, $\mathbf{y}(t)$, is called the measurement vector and \mathbf{H} is an $m \times n$ measurement matrix. $\mathbf{n}(t)$ represents the measurement noise, which also has zero mean and is normally distributed, with power spectral density \mathbf{R} .

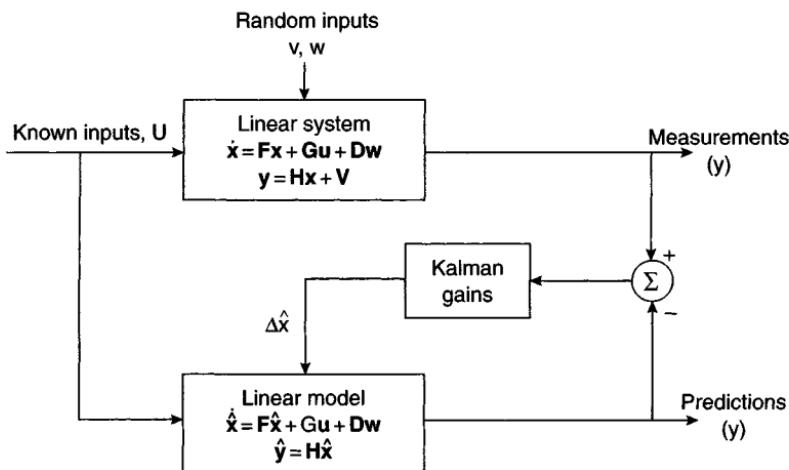


Figure A.1 Block diagram of a Kalman filter

The Kalman filter for the system described here seeks to provide the best estimates of the states, x , using:

- the measurements, y ;
- a model of the system provided by the matrices F , G , H , and D ;
- knowledge of the system and measurement statistics given in the matrices Q and R .

The deterministic or measurable inputs are processed by both the system and the model of the system, as shown in the block diagram representation in Figure A.1.

The measurements of the true system are compared with predictions of those measurements, derived from the latest best estimates of the states provided by the system model. The difference between the true and predicted measurements are fed back through a weighting matrix, the Kalman gain matrix, to correct the estimated states of the model.

The Kalman gains are selected to provide best estimates of the states in a least-squares sense. It can be shown that this is equivalent to the best estimate in the maximum likelihood sense in the linear, Gaussian noise system described above. It should be noted that, it is because there is a feedback of a noisy signal in the Kalman filter that the system must be linear and the noise Gaussian, or normally, distributed. It is only because the distribution of the sum of two normally distributed signals is itself normally distributed, and also because a normally distributed signal remains so after passing through a linear system, that a least-squares or maximum likelihood estimation procedure can be applied repeatedly.

Whilst it is usual for the system to be described mathematically in the continuous differential equation form given above, the measurements are in practice provided at discrete intervals of time. To cope with this, and to provide a computationally efficient filtering algorithm, it is customary to express the continuous equations in the form of

difference equations as shown below:

$$\mathbf{x}_{k+1} = \Phi_k \mathbf{x}_k + \Gamma_k \mathbf{u}_k + \Delta_k \mathbf{w}_k \quad (\text{A.21})$$

with measurements:

$$\mathbf{y}_{k+1} = \mathbf{H}_{k+1} \mathbf{x}_{k+1} + \mathbf{n}_{k+1} \quad (\text{A.22})$$

where \mathbf{x}_k is the state at time t_k , \mathbf{u}_k is the input at time t_k , \mathbf{w}_k is the system noise at time t_k , \mathbf{n}_{k+1} is the measurement noise at time t_{k+1} , Φ_k is the state transition matrix from time t_k to time t_{k+1} , \mathbf{H}_{k+1} is the measurement matrix calculated at time t_{k+1} , and Γ_k and Δ_k are appropriate input matrices.

The noise is zero mean, but now discrete, and will be characterised by the covariance matrices \mathbf{Q}_k and \mathbf{R}_k , respectively.

These equations are used to formulate a recursive filtering algorithm. In such a formulation, it is necessary to consider two distinct sets of equations. The first set is concerned with the prediction of the state of the system based on the previous best estimate, whilst the second involves the updating of the predicted best estimate by combining the prediction with a new measurement.

The prediction process

The best estimate of the state at time t_k is denoted here by $\mathbf{x}_{k/k}$. Since the system noise, \mathbf{w}_k , has zero mean, the best prediction of the state at time t_{k+1} is given by:

$$\mathbf{x}_{k+1/k} = \Phi_k \mathbf{x}_{k/k} \quad (\text{A.23})$$

whilst the expected value of the covariance at time t_{k+1} predicted at time t_k , is given by:

$$\mathbf{P}_{k+1/k} = \Phi_k \mathbf{P}_{k/k} \Phi_k^T + \Delta_k \mathbf{Q}_k \Delta_k^T \quad (\text{A.24})$$

The measurement update

On arrival of a new measurement \mathbf{y}_{k+1} , at time t_{k+1} , it is compared with the prediction of that measurement derived from the system model. The measurement is then used to update the prediction to generate a best estimate, following the procedure outlined in the previous section. Hence, the best estimate of the state at time t_{k+1} is given by:

$$\mathbf{x}_{k+1/k+1} = \mathbf{x}_{k+1/k} - \mathbf{K}_{k+1} [\mathbf{H}_{k+1} \mathbf{x}_{k+1/k} - \mathbf{y}_{k+1}] \quad (\text{A.25})$$

and its covariance by:

$$\mathbf{P}_{k+1/k+1} = \mathbf{P}_{k+1/k} - \mathbf{K}_{k+1} \mathbf{H}_{k+1} \mathbf{P}_{k+1/k} \quad (\text{A.26})$$

where the Kalman gain matrix is given by:

$$\mathbf{K}_{k+1} = \mathbf{P}_{k+1/k} \mathbf{H}_{k+1}^T [\mathbf{H}_{k+1} \mathbf{P}_{k+1/k} \mathbf{H}_{k+1}^T + \mathbf{R}_{k+1}]^{-1} \quad (\text{A.27})$$

where \mathbf{H}^T denotes the transpose of the measurement matrix, \mathbf{H} .

The system states may therefore be updated each time a measurement is received by implementing eqns. (A.25)–(A.27).

Non-linear systems – the extended Kalman filter

So far, we have considered only linear dynamical systems with zero mean, Gaussian noise type disturbances. For such systems, the Kalman filter is optimal in the least-squares sense or maximum likelihood sense. If the system is not linear or if the noise is not Gaussian, the Kalman filter is no longer optimal. In these cases, the only way to regain the ‘optimality’ of the filtering is to design an algorithm specifically for the system under consideration. This, however, is not usually feasible in practice, as the filter would be of infinite dimension. Therefore, it is normal to accept that the performance will be sub-optimal, and use the Kalman filter in such a way as to make the performance as close to optimal as possible. This may involve, for instance, predicting the system and its covariance matrix over relatively short time intervals during which the conditions for linearity hold.

Consider a continuous non-linear dynamical system described by the equations:

$$\frac{dx}{dt} = f(x, t)x + g(x, t)u + d(x, t)w \quad (A.28)$$

with discrete measurements given by:

$$y = h(x, t)x + v \quad (A.29)$$

For simplicity, the explicit dependence on time will be dropped, that is, $f(x, t)$ will be written as $f(x)$. This system may be approximated by linearising it about a nominal set of states commonly referred to as a nominal trajectory.

This approximation will only be valid for short periods of time after which the system will have to be re-linearised. The linearisation technique most often used is the truncation of the Taylor series. Thus, for the function $f(x)$, the Taylor series about a nominal trajectory \tilde{x} is given by:

$$f(x) = f(\tilde{x}) + \left. \frac{df}{dt} \right|_{\tilde{x}} (x - \tilde{x}) + \left. \frac{d^2 f}{dt^2} \right|_{\tilde{x}} \frac{(x - \tilde{x})^2}{2} + \dots \quad (A.30)$$

and similarly for other non-linear functions. Defining the nominal trajectory as:

$$\frac{d\tilde{x}}{dt} = f(\tilde{x})\tilde{x} + g(\tilde{x})u \quad (A.31)$$

$$\tilde{y} = h(\tilde{x})\tilde{x} \quad (A.32)$$

and subtracting from the original equations, we obtain differential equations governing the deviations from the nominal trajectory, viz:

$$\frac{d\delta x}{dt} = \left. \frac{df}{dt} \right|_{\tilde{x}} \delta x + \left. \frac{dg}{dt} \right|_{\tilde{x}} \delta u + \left. \frac{dd}{dt} \right|_{\tilde{x}} w \quad (A.33)$$

$$\frac{dy}{dt} = \left. \frac{dh}{dt} \right|_{\tilde{x}} \delta x + v \quad (A.34)$$

If we define:

$$\mathbf{F} = \left. \frac{d\mathbf{f}}{dt} \right|_{\tilde{\mathbf{x}}} ; \quad \mathbf{G} = \left. \frac{d\mathbf{g}}{dt} \right|_{\tilde{\mathbf{x}}} ; \quad \mathbf{D} = \left. \frac{d\mathbf{d}}{dt} \right|_{\tilde{\mathbf{x}}} ; \quad \mathbf{H} = \left. \frac{d\mathbf{h}}{dt} \right|_{\tilde{\mathbf{x}}} \quad (\text{A.35})$$

then the discrete Kalman filter equations may be used as given in the previous section. At each measurement interval, the following steps must now be taken:

1. Linearise the equations about the nominal trajectory. The nominal trajectory is usually taken to be the latest estimate of the states.
2. Calculate the transition matrix and other matrices of the discrete equivalent to the linearised system.
3. Integrate the state prediction equations. The actual estimated states may be used here, as there is no difference between this and integrating the differential equations for the nominal trajectory and the deviations from it separately and then adding the result.
4. Implement the Kalman filter equations. This will provide best estimates of the deviations from the nominal. The corrections, given by the Kalman gains multiplied by the measurement differences, can again be added directly to the predicted state estimates.
5. Continue to the next time interval, that is, return to step 1.

There are situations in which the measurement update rate may be relatively low, in these cases the prediction stage of the filter will need to be run with a smaller time interval. If this is not done, the non-linearities begin to dominate the deviations of the state estimates from the nominal.

Although this form of the Kalman filter, the so-called extended Kalman filter, is used for systems which are explicitly non-linear, it is also used when there is a requirement to identify certain unknown parameters in the system. In this situation, the unknown parameters are defined as states of the system and the state vector is augmented to include them.

References

- 1 BARHAM, P.M., and HUMPHRIES, D.E.: 'Derivation of the Kalman filtering equations from elementary statistical principles', NATO Agardograph AG139, 'Theory and application of Kalman filters', 1970
- 2 GELB, A. (Ed.): 'Applied optimal estimation' (Massachusetts Institute of Technology, Cambridge, MA, 1974)
- 3 JAZWINSKI, A.H.: 'Stochastic processes and filtering theory' (Academic Press, 1970)

Appendix B

Inertial navigation system error budgets

The accuracy of an inertial navigation system is often expressed as a position uncertainty after a given period of navigation, or on reaching a given destination. Alternatively, it is expressed in terms of the rate at which the navigation error builds up with time, in nautical miles per hour for instance. The actual form of expression used for this overall accuracy figure depends upon the application. For example, for an inter-planetary missions, the accuracy refers to the desired point of closest approach to the 'destination' planet. For navigation in the vicinity of the Earth, it is usually the errors in two dimensions which are of most interest, the along-track and cross-track position errors over the surface of the Earth. These errors are often combined to yield a single number which expresses navigation accuracy after a given navigation time, the circular error probable (CEP) or circular probable error (CPE) as it is sometimes called. Essentially, this defines a circular area within which the navigation system estimates its true position to be, with a certain probability. The 50 per cent CEP is a frequently quoted figure. When the probability value is not stated, it usually means a 50 per cent value should be assumed.

Consider now the composition of this navigation performance figure. In practice, navigation errors propagate owing to a large number of error sources which include alignment errors, a variety of inertial sensor errors and errors attributable to computational inaccuracy. In general, each errors may be regarded as comprising a repeatable or predictable component and a random or unpredictable component. The former category produces predictable effects which can be compensated for, if so desired, that is, electrical signals or software corrections can be applied which should offset the effect of such errors. The remaining errors, which arise as a result of random effects within the system and incomplete compensation of systematic errors, will give rise to navigation inaccuracies.

For the assessment of system performance, random errors are treated statistically to derive a mathematical description of each term and to allow the various contributors to the overall system errors to be combined in what is known as the system error budget. It is common practice to assume that the random errors within the components

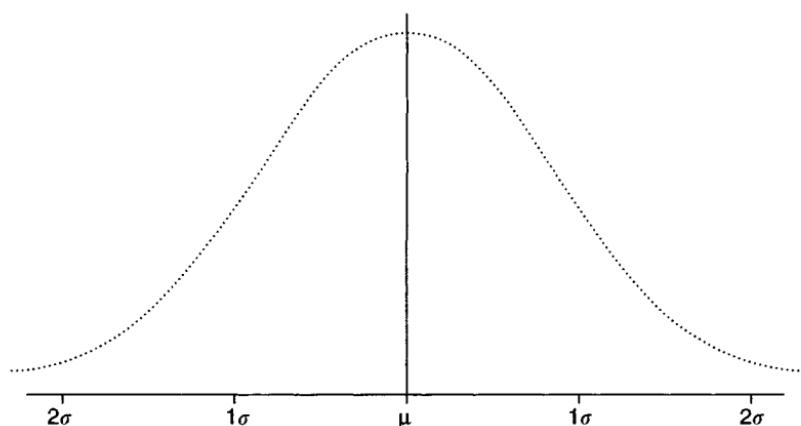


Figure B.1 Gaussian distribution

of an inertial navigation system follow a Gaussian or normal distribution where the probability density function, denoted $p_x(x)$, is expressed mathematically as:

$$p_x(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(\frac{-x^2}{2\sigma^2}\right)$$

A Gaussian distribution is depicted in Figure B.1.

Based on this assumption, the statistical analysis techniques applicable for normal distributions may be conveniently used for the analysis of inertial systems.

If measurements are made of a bias (x_i) on the measurement provided by a gyroscope or an accelerometer, then over a large number of sensor samples, it is assumed that the error distribution approximates to this curve. It is common practice to quote a 1σ error or standard deviation for each sensor error. The standard deviation of a normal distribution may be expressed in terms of the mean error (x_m) and the number of samples (n) as follows:

$$\sigma = \sqrt{\frac{\sum x_i^2 - nx_m^2}{n}}$$

The 1σ value represents 68.26 per cent of the area under the curve. Hence, if a 1σ gyroscope bias is quoted as being $1^\circ/\text{h}$, it is expected that over a large number of gyroscopes of that type, 68.26 per cent would have a bias of between $\pm 1^\circ/\text{h}$ of the mean value. Alternatively, it may be inferred there is a 68.26 per cent probability of the bias uncertainty on a given gyroscope being within $\pm 1^\circ/\text{h}$. The 2σ and 3σ bias values which are obtained by multiplying the 1σ value by 2 and 3, respectively, may also be defined for the sensor, corresponding to 95.44 and 99.73 per cent of the area under the normal distribution curve. Hence, for this example, there is a 95.44 per cent probability of the gyroscope bias being within $\pm 2^\circ/\text{h}$ and a 99.73 per cent probability of the bias being within $\pm 3^\circ/\text{h}$.

In an inertial navigation system, a number of sensors and components are required to operate together, each producing random errors. It is usual to assume that each error component is unrelated to each of the other error components, that is, the error components are said to be independent. In combining the effect of many error sources, simply summing the individual contributions to the error budget arithmetically would give a very pessimistic prediction of system performance. Where a number of independent sources need to be combined, a more correct prediction of overall system performance is obtained by summing the individual 1σ errors quadratically, i.e. by taking the root sum square (RSS) error. Hence, if $\sigma_1, \sigma_2, \dots, \sigma_n$ represent a number of independent errors which combine to produce an overall error, then the total or overall error is obtained using:

$$\sigma_{\text{RSS}} = \sqrt{\sigma_1^2 + \sigma_2^2 + \dots + \sigma_n^2}$$

where σ_{RSS} is the 1σ RSS error. In an inertial navigation system, an RSS position error can be calculated for both the along-track or down-range position error (σ_x) and the cross-track position error (σ_y). Assuming all error contributions to both along- and cross-track errors to be Gaussian, then the behaviour of each is also described by a Gaussian distribution curve. When σ_x and σ_y are combined, a probability ellipse is used to describe the behaviour of the total error. For the special case where $\sigma_x = \sigma_y$, the probability ellipse reduces to a probability circle. In mathematical terms, the radial probability distribution is given by:

$$p(r) = \frac{r}{\sigma^2} \exp\left(\frac{-r^2}{2\sigma^2}\right)$$

where σ is the standard deviation of the error in x and y . This is known as the Rayleigh probability density function (Figure B.2). The probability of r lying

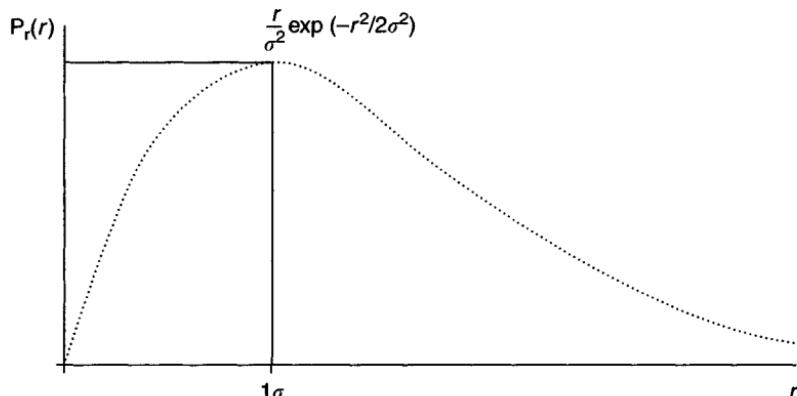


Figure B.2 Rayleigh distribution

between 0 and R , $P_r (< R)$, is given by:

$$\begin{aligned} P_r (< R) &= \int_0^R p(r) dr \\ &= 1 - \exp\left(\frac{-R^2}{2\sigma^2}\right) \end{aligned}$$

$P_r (< R) = 0.5$ defines the 50 per cent probability circle which occurs when $R/\sigma = 1.17741$. The radius of this circle is the 50 per cent CEP referred to earlier.

Hence,

$$50 \text{ per cent CEP} = 1.17741\sigma$$

The reader is referred to any standard text on probability theory, Reference 1 for instance, for a more detailed discussion of the topics outlined in this appendix.

Reference

- 1 LATHI, B.P.: 'An introduction to random signals and communication theory' (International Textbook Company, 1968)

Appendix C

Inertial system configurations

Practical inertial navigation systems may take a variety of forms. These forms generally fall into one of two basic categories,

- stable platform systems;
- strapdown systems.

Although the two types of system are very different, both physically and computationally, it must be stressed that the underlying principles and functionality of the two types of system are identical.

Stable platform systems

- Inertial sensors mounted on stabilised platform
- Inertial sensors isolated from rotational motion of vehicle

Strapdown systems

- Inertial sensors rigidly attached to vehicle
- Inertial sensors subjected to vehicle turn rates

Stable platform systems

The original applications of inertial navigation technology used stable platform techniques. At that time, neither sensors with the necessary dynamic range nor sufficiently powerful computers were available to allow strapdown systems to be produced. At the core of such a system is a structure, called the platform, on which inertial sensors are mounted. This platform is 'isolated' from the rotational motion of the vehicle using a number of gimbals arranged to provide at least three degrees of rotational freedom and so minimise the angular coupling between the vehicle and the platform. Generally, three gimbals are used.

Figure C.1 shows a schematic representation of a three gimbal stabilised platform system. The platform containing the inertial sensors is supported by three gimbals.

The gimbals are mechanical frames, each of which is free to rotate about a single-axis which is nominally perpendicular to the free axis of its neighbouring gimbal(s). Torque motors are used to rotate the gimbals with respect to one another and angular pick-offs provide a measure of their relative orientation. The system is configured so that the three gimbal pick-off angles correspond to the roll, pitch and yaw orientations of the host vehicle with respect to the platform/reference frame. Sometimes, it is necessary to add a fourth gimbal; this is required for very agile vehicle applications to allow the platform to remain isolated from the vehicle irrespective of its orientation. For example, a vertical launch guided weapon test vehicle was produced some years ago which had a four-axis gimbal system to avoid 'gimbal lock' during the dynamic turn-over manoeuvre.

The platform configuration (Figure C.1) illustrated here minimises the amount of computing needed to implement the navigation function of providing position, velocity and attitude of the host vehicle with respect to the designated navigation reference frame. Since the platform, and hence the accelerometer triad, is held in alignment with the reference frame, typically coincident with the local geographic frame (north, east, down), it is simply required to sum the accelerometer outputs with the gravity terms and to integrate the navigation equations to obtain estimates of velocity and position in the reference frame. Any rotational motion of the platform would be detected by a gyroscope, the output of which would be fed back via a torque motor to rotate the appropriate gimbal (and hence the platform) in the opposite sense, so maintaining its initial (and fixed) orientation in space. For systems required to navigate around the Earth, it is required to torque the platform at Earth's rate plus any rate caused by the velocity of the system with respect to the Earth (transport rate) to allow it to remain aligned with the local level frame.

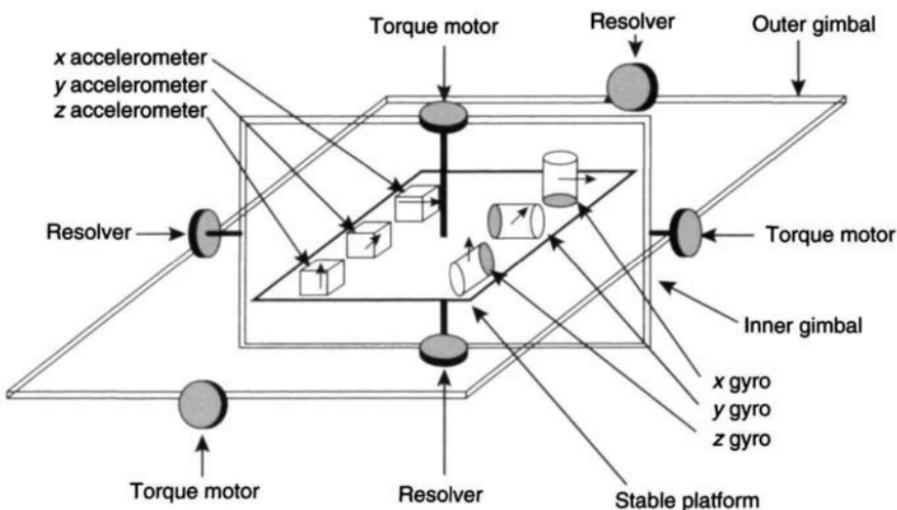


Figure C.1 Three gimbal platform IN system

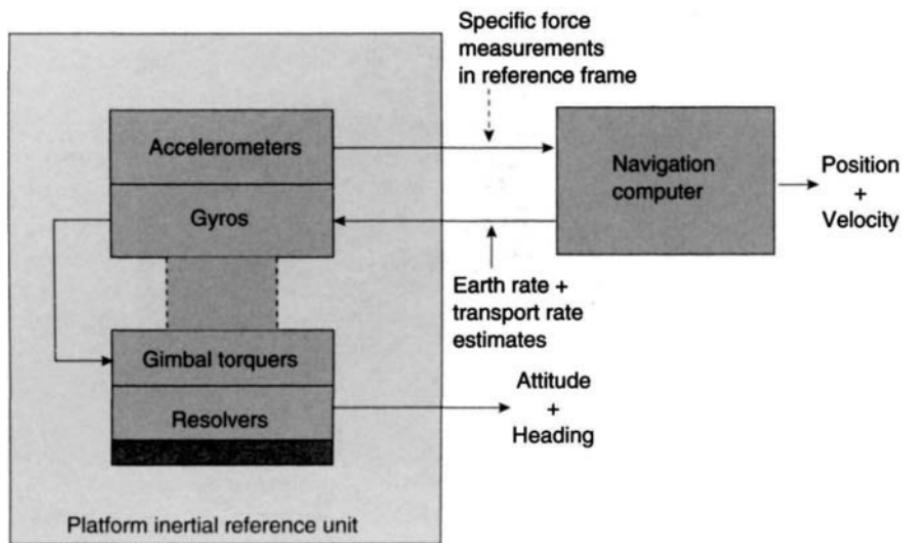


Figure C.2 Stable platform IN system

It should be noted that platform systems are still in common use, particularly for ships and submarines where accurate navigation is required (unaided) over long periods of time.

A functional block diagram representation of a stable platform system is shown in Figure C.2. This diagram shows the functional blocks which combine to form a stable platform INS. Attention is drawn to the inertial reference unit comprising the inertial sensors mounted on the platform and supported by a set of mechanical gimbals with their respective torque motors and angular pick-offs. This may be compared with the equivalent strapdown system configuration in Figure C.4. A photograph of a stable platform is given in Figure C.3.

Strapdown systems

In strapdown systems, the inertial sensors are fastened directly (or via anti-vibration mounts) to the vehicle and hence are not isolated from its angular motion. Consequently, the gyroscopes, as well as the accelerometers, experience the full dynamic motion of the host vehicle. The signals produced by the inertial sensors are resolved mathematically in a computer prior to the calculation of navigation information. This use of a computer to establish and resolve the inertial data reduces the mechanical complexity of the inertial navigation system, thus frequently reducing the cost and size of the system and increasing its reliability. It is stressed that this reduction in mechanical complexity is achieved at the expense of computational complexity.

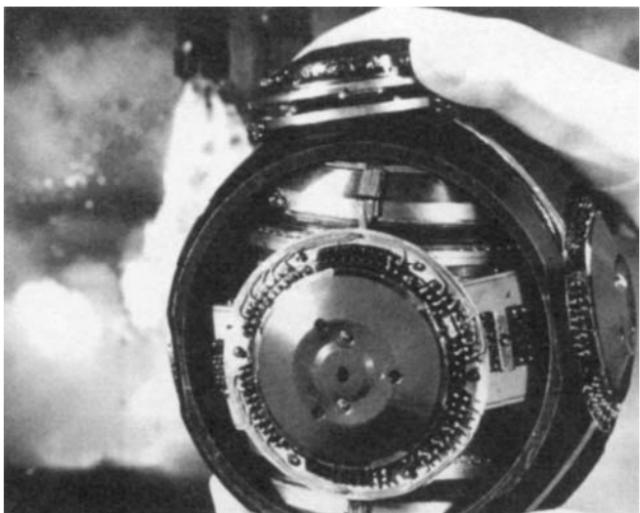


Figure C.3 Marconi stable platform

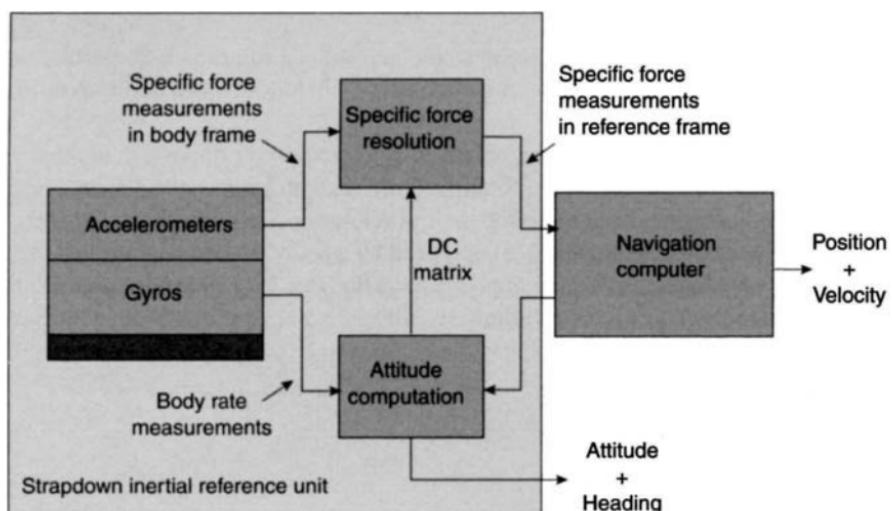


Figure C.4 Strapdown IN system

Figure C.4 shows the main functional blocks of a strapdown system for comparison with the stable platform configuration in Figure C.2. The mechanical complexity of the platform system is seen to be replaced by additional computing tasks in the strapdown inertial reference unit, viz. the attitude computation and specific force resolution tasks. Attitude and heading are now computed rather than being provided directly by electrical pick-off devices. However, outside of the respective inertial

reference units, the navigation computations are the same, that is, both systems are required to solve identical equations.

To summarise, the underlying principles of the stable platform and strapdown systems are common to both and they perform identical functions. The only difference between the two systems, at the functional level discussed here, is the replacement of the mechanical platform with what may be described as an ‘analytic platform’. A strapdown system replaces mechanical complexity with computational complexity.

Whilst this book aims primarily to provide an appreciation of the principles of strapdown navigation and the related areas of technology, it should be remembered that the stable platform system constitutes a precise mechanical analogue of the strapdown system. It is often helpful to refer to the stable platform system to gain a clearer physical insight into many of the processes that take place within the strapdown (analytic) system.

Appendix D

Comparison of GPS and GLONASS satellite navigation systems

There are currently two satellite-based navigation systems deployed, the American controlled system known as GPS or Navstar and the Russian system GLONASS. There are a number of similarities between the two systems, such as the constellations consist of up to 24 satellites, although the configuration within each constellation is somewhat different. There are several other similarities between the two systems, but the characteristics of the two systems are sufficiently different to affect their operation and world coverage.

The similarities and complementary nature of the differences in coverage lead to the possibility of using the systems together in an integrated system to enhance navigation performance.

Comparison of systems

Both systems are space-based approaches to provide the user with precision position, velocity and time data for accurate navigation. The aim of each system is to provide:

- navigation data throughout the day and night;
- navigation data during all types of weather conditions;
- navigation data anywhere over the world's surface, or close to it.

GPS

This system has a space segment with 21 active satellites and three active spares in six 12-h orbits that are at an altitude of 20 180 km; the orbits have an inclination of 55° to one another. The message broadcast continuously from each satellite contains data about its precise position and clock accuracy, and less precise information about its relative position with regard to the other satellites in the constellation, that is, the almanac.

This system has a control segment, consisting of monitor stations and a master control station to provide precise measurement of each satellite's orbital parameters. This information is sent to each satellite for re-transmission to the users.

The user segment comprises the equipment required to track the satellites that are visible to the receiver to enable navigation data to be determined.

The GPS satellites use two carrier frequencies in the L band, known as L₁ (1575.42 MHz) and L₂ (1227.6 MHz). Each of these signals is modulated by either or both of the precise positioning service (PPS or P) signal (10.23 MHz) and the standard positioning service (SPS), also called the coarse/acquisition (C/A) signal (1.023 MHz). The binary signals are created by a P-code or C/A-code, which is modulo 2 added to 50 bps data. The P-code and C/A-code are added to L₁ in phase quadrature. Note that only the P-code is present on the L₂ signal.

The P-code is a pseudo-random sequence with a period of one week. This contrasts with the C/A-code, which has a period of 1 ms and has a Gold code. A receiver duplicates either or both of these codes and the transmission time is derived from the offset that has to be generated at that position to synchronise the locally generated code with the one emitted from a satellite. The C/A-code is a 1 MHz 1.023 bit sequence from a set of Gold codes, it repeats at 1 ms intervals, enabling a receiver to search rapidly through the received signal to obtain lock. The P(r) code is a 10 MHz signal. Receiver technology to acquire the P(v) code is only just becoming available, even so it is expensive and power hungry. The C/A code was designed to enable rapid acquisition of time to allow the position of the P(v) code to be established. However, the P-code provides an accuracy of about a factor of 2 or 3 better than the use of the C/A-code alone.

The American Department of Defence controls the precision of the navigation data available through a technique known as selective availability (SA), where the satellite signals are corrupted. Authorised users are provided with the algorithms required to recover the original satellite signals. The reduction in accuracy caused by SA to the order of 100 m has prompted a number of techniques for recovering some of this lost precision, examples include relative and differential GPS.

The principle of differential GPS techniques is to take advantage of the fact that a substantial component of the navigation error in a GPS measurement arises from slowly changing biases. Moreover, these biases are correlated in both distance and time between an array of receivers. Therefore if two receivers, or more, are operating simultaneously at different locations and the position of one of them is known, then corrections can be generated in real time to the measurements from the positional knowledge of one receiver and applied to the other receiver measurements. Clearly, a data link is required between each receiver to enable this technique to operate.

GLONASS

The space segment has 24 satellites in this constellation at an altitude of 19 100 km, corresponding to an 11 h 15 min period. They are arranged in three orbital planes with a relative inclination of 64.8°. These satellites also broadcast their own precise position as well as less accurate data on the positions of the other satellites in the constellation.

The transmitted data are in the form of Earth-centred, Earth-fixed coordinates and extrapolation terms, rather than the orbital parameters used by the GPS.

The ground-based control segment is similar to GPS and undertakes a similar function. Similarly, the user segment involves the use of a receiver, or set of receivers, to track the satellites.

Each of the satellites in the constellation uses two carrier frequencies in the L band, however, there are different frequencies for each satellite. The L₁ emission ranges from 1602.5625 to 1615.5 MHz in jumps of 0.5625 MHz, whilst L₂ ranges from 1246.4375 to 1256.5 MHz in jumps of 0.4375 MHz. Each of these signals is modulated in a similar fashion to the GPS signals: the P-code modulation is 5.11 MHz whereas the C/A-code is 0.511 MHz. The binary signals are created in an identical fashion. However, the period of the P-code pseudo-random sequence is one second and the corresponding C/A-code period is 1 ms, and added to the L₁ signal in quadrature, again with the P-code present in the L₂ signal.

A single GLONASS code is used for all of the satellites in the constellation, whereas each of the GPS satellites has a unique code. The navigation function is identical to that described above for GPS. At this stage there is no selective availability applied to corrupt the fidelity of the transmitted satellite data, and the accuracy is superior to the standard positioning service of GPS.

The table below compares the characteristics of these two satellite navigation systems.

GPS	GLONASS
24 satellites in six planes with a ~12 h period and 55° inclination	24 satellites in three planes with a ~11 h 15 min period and ~65° inclination
Spread spectrum system	Spread spectrum system
Code division multiplexing	Frequency division multiplexing
C/A-codes and P-codes with selective availability	C/A-codes and P-codes but no selective availability
Broadcast satellite orbital parameters updated every hour	Broadcast satellite E-C, E-F position, velocity and acceleration updated every $\frac{1}{2}$ hour
WGS-84 Earth model (E-C, E-F reference frame ¹)	Soviet geocentric co-ordinate system SGS-90 or PZ-90 Earth model (E-C, E-F reference frame)
GPS time synchronised with UTC (USNO)	GLONASS time synchronised with UTC (Moscow)

¹ Earth Centred, Earth Fixed reference frame.

The differences in the orbital planes of these two constellations of these systems lead to differences in coverage throughout the world. GPS provides very good coverage at mid latitudes, whereas GLONASS provides good coverage at higher latitudes.

Schemes have been devised to enable users to combine information derived from these two systems. The advantages are:

- faster acquisition in the cold-start mode, as more satellites are potentially available;
- better coverage in obstructed environments, as again a larger number of satellites should be available;
- improved accuracy;
- enhanced system integrity as the user is not dependent on a single navigation system and its continued availability;
- integrated operation.

There is a timing synchronisation and definition requirement for dual GPS and GLONASS operation, owing to the different clock time bases used by the two systems. In terms of a GPS-only receiver, or a GLONASS-only receiver, all measurements include a receiver-clock error with respect to either GPS or GLONASS defined time. This error is common to all measurements from a given constellation, therefore it will only affect the time estimate and will not have an impact on the position and velocity estimates. However, in the case of dual satellite navigation system operation using both systems, some measurements will include a GPS-to-receiver clock error, whilst others will include GLONASS-to-receiver clock error. The relationship between the GPS and GLONASS time references, therefore, has to be known or derived for acceptable position and velocity estimates to be formulated during integrated operations.

This factor is normally accounted for in terms of the ‘two receiver’ clock offsets. The state vectors in the Kalman filter could be increased by two to five, to account for the additional unknown parameters. An alternative approach to correcting for this problem is to correlate all of the measurements to the GPS (or GLONASS) time base, which requires knowledge of the relative receiver to clock off-sets prior to the measurements being taken.

GPS time is related to the universal time (UT) standard, but the GLONASS system is related to Moscow time. Analysis has shown that the differences can reach several microseconds, hence this is extremely significant for both position and velocity measurements made by mixing measurements from the two satellite-based systems. The other important issue in dual-mode operation is the difference between the two reference systems used by GPS and GLONASS. As noted above, GPS uses the WGS-84 ellipsoid representation of the Earth, whereas GLONASS uses SGS-90 (also known as PZ-90). Again, this difference between the two systems can be resolved by using a suitable transformation matrix, which allows the data to be referred to a common reference.

The overall accuracy that can be obtained from the use of data transmitted by the two systems, depends on:

- the combined effect of the improved position dilution of precision over that operating for either system, coupled with the difference in accuracy possible from either system alone;
- the residual error between the two time-base references.

The combined use of the two systems offers a significant increase in system integrity as well as improved accuracy. A major issue is the determination of any rogue data that would reduce the accuracy of the navigational estimates. A highly integrated system architecture may be used to make the best use of available data from the two systems. Individual satellite tracking channels can be arbitrarily assigned to any satellite available in either constellation.

The issues concerning the advantages and disadvantages of combined operation, compared with single system use, are given in the table below.

Advantages	Disadvantages
Better accuracy owing to <ul style="list-style-type: none"> • better coverage at all latitudes; • better geometry; • GLONASS performance not impaired by selective availability. 	A more complex receiver and more complex processing.
Enhanced integrity owing to a higher number of satellites in the combined constellations.	GLONASS future availability and reliability?
Enhanced availability owing to a greater selection of satellites in-view at any instant.	GLONASS not under 'Western' control.
Independent of US DoD control	

Clearly, a number of the disadvantages will be resolved with the introduction of the Galileo system. Moreover, this has been designed to be compatible with the GPS.

List of symbols

Scalars

a_x	magnitude of acceleration acting along x -axis of body
a_y	magnitude of acceleration acting along y -axis of body
a_z	magnitude of acceleration acting along z -axis of body
$B_{fx,y,z}$	g -insensitive bias of inertial sensor
$B_{gx,y,z}$	gyroscope g -sensitive bias coefficients
$B_{axz,yx}$	gyroscope anisoelastic bias coefficients
c	speed of light
c_{ij}	i, j element of direction cosine matrix
f	magnitude of specific force, the non-gravitational force per unit mass, acting on a body
f_N	magnitude of north component of specific force
f_E	magnitude of east component of specific force
f_D	magnitude of vertical (down) component of specific force
g	magnitude of acceleration due to Earth's gravity
h	height above ground
H	magnitude of angular momentum
I	moment of inertia
ℓ	longitude
L	latitude
M_x, M_y, M_z	inertial sensor cross-coupling coefficients
n_x, n_y	zero-mean random biases of inertial sensor
p	roll rate
q	pitch rate
r	yaw rate
R_0	mean radius of curvature of the Earth
S_x, S_y	scale-factor error of inertial sensor
t	time
v_N	magnitude of north velocity
v_E	magnitude of east velocity

v_D	magnitude of vertical (down) velocity
δt	time increment
$\delta\phi$	rotation increment in roll
$\delta\theta$	rotation increment in pitch
$\delta\psi$	rotation increment in yaw
ξ	meridian deflection of the local gravity vector
η	deflection of the local gravity vector perpendicular to the meridian
ϕ	roll Euler angle
θ	pitch Euler angle
ψ	yaw Euler angle
π	pi
ω_x	roll rate of body with respect to navigation frame
ω_y	pitch rate of body with respect to navigation frame
ω_z	yaw rate of body with respect to navigation frame
ω_R	gyroscopic nutation frequency
ω_s	spin speed of gyroscope rotor
Ω	turn rate of the Earth

Vectors

\mathbf{a}_i	acceleration with respect to inertial reference frame
\mathbf{f}	specific force vector
\mathbf{g}	mass attraction gravitation vector
\mathbf{g}_l	local gravity vector
\mathbf{H}	angular momentum vector
\mathbf{i}	unit vector in x -direction of co-ordinate frame
\mathbf{j}	unit vector in y -direction of co-ordinate frame
\mathbf{k}	unit vector in z -direction of co-ordinate frame
\mathbf{q}	four-element quaternion vector $[a \; b \; c \; d]$
\mathbf{r}	position vector
\mathbf{T}	torque acting on the body
\mathbf{v}_i	velocity with respect to inertial reference frame
\mathbf{v}_e	velocity with respect to the Earth, the ground speed
$\boldsymbol{\sigma}$	angle vector having components $\sigma_x, \sigma_y, \sigma_z$
$\boldsymbol{\omega}$	gyroscopic precession rate
$\boldsymbol{\omega}_{ie}$	turn rate of the Earth with respect to inertial reference frame
$\boldsymbol{\omega}_{ib}$	turn rate of body with respect to inertial reference frame
$\boldsymbol{\omega}_{eb}$	turn rate of body with respect to the Earth frame
$\boldsymbol{\omega}_{in}$	turn rate of navigation frame with respect to inertial reference frame
$\boldsymbol{\omega}_{en}$	turn rate of navigation frame with respect to Earth frame
$\boldsymbol{\omega}_{ew}$	turn rate of wander azimuth frame with respect to Earth frame

Matrices

B	direction cosine matrix which defines transformation from true reference axes to estimated reference axes
C_b^i	direction cosine matrix relating body frame to inertial reference frame
C_b^e	direction cosine matrix relating body frame to Earth frame
C_b^n	direction cosine matrix relating body frame to navigation frame
C_e^w	direction cosine matrix relating earth frame to wander azimuth frame
Ω_{ib}	skew symmetric matrix defining turn rate of body with respect to inertial reference frame
Ω_{eb}	skew symmetric matrix defining turn rate of body with respect to Earth frame
Ω_{nb}	skew symmetric matrix defining turn rate of body with respect to navigation frame
Ω_{ew}	skew symmetric matrix defining turn rate of wander azimuth frame with respect to Earth frame
Ψ	skew symmetric matrix defining incremental rotation

Other symbols

\times	denotes a vector cross product
\sim	denotes a measured quantity
\wedge	denotes an estimated quantity
\cdot	denotes a quaternion product
$*$	denotes the complex conjugate of a vector

Additional symbols are defined where they are used in the text.

Glossary of principal terms

Inertial navigation systems rely on complex technology, even if the implementation is simple or mundane. Thus it will be inevitable that many technology specific terms and jargon will be in common usage. The more common aspects are covered here together with those that are used in the text of this book. The terms described appear in alphabetical order.

Accelerometers

These devices are used to measure the translational motion of the vehicle in which they are located. In its most basic form, this sensor consists of a proof mass attached, via a spring, to the case of the instrument. The device operates by detecting a displacement of the mass in the presence of an acceleration of the vehicle. An accelerometer actually provides a measure of specific force, the non-gravitational force per unit mass to which it is subjected.

Aiding

This technique refers to the use of external measurement data to enhance the performance of an inertial navigation system. The source of the aiding may be external to the vehicle or derived from an additional on-board sensor. An example of the former type is the use of position information provided by a tracking radar to update periodically an inertial navigation system in a short range tactical missile. Alternatively, an on-board sensor such as a Doppler radar or a barometric altimeter may provide information to aid the inertial system. As with all such techniques which combine measurement data with different characteristics, sophisticated Kalman filtering techniques are often used, and are necessary to produce the best results.

Note: data from the inertial navigation system may be used to aid another sensor, such as a satellite navigation system.

Algorithms

An algorithm is a mechanistic procedure for solving a problem in a finite number of steps. Such a procedure may take the form of a series of mathematical instructions which may be programmed in a computer.

Alignment

This is the process of determining the initial orientation of the measurement axes of the inertial navigation system to the chosen reference frame, that is, the process of determining the angles between the measurement axes and the axes of the reference frame. An alignment must be carried out prior to the start of navigation. The accuracy with which this can be achieved is of fundamental importance as it can severely influence the performance of an inertial navigation system, or indeed any system which operates by the so-called dead reckoning technique. Following the initial alignment process the IN system keeps calculating the orientation of the sensors during the navigation.

Attitude and heading reference system

Such a system will provide, as its name suggests, attitude and heading data for the vehicle in which it is installed. This system is very similar to an inertial navigation system, but does not provide velocity and position of the host vehicle. Generally the sensors are of lower accuracy and consequently less expensive. Such a system is often combined with another sensor such as a Doppler radar to form a full navigation system. The Doppler radar provides measurements of velocity over the ground which can be resolved into the navigation frame defined by the attitude and heading reference system. Integration of these velocity components yields the position of the vehicle.

Bias

Bias refers to the offset in the measurement provided by an inertial sensor. For example, a gyroscope provides a measurement of turn rate about a given axis. The output of the gyroscope may take the form of a voltage or current proportional to the applied turn rate plus a bias term caused by the various imperfections within the sensor.

Calibration

This is the process of establishing the precise value (and meaning) of the electrical signals produced by the inertial sensors. For example, a change of one millivolt in the electrical signal generated may indicate a rotation of a given rate or a specific force acceleration of a given value.

Compensation

All inertial sensors which are subject to systematic errors can have these errors removed by a correction technique called compensation. For example, if a gyroscope has a known tendency to drift when its temperature changes, the effects of this drift can be corrected provided:

- it is a systematic error;
- the temperature at which the measurement being made is known;
- the temperature of the calibration is known;
- the contribution that this change in temperature has to the measured electrical signal is known.

Given this information, the signal produced by the gyroscope can be corrected to give a measure of the true rotation or rotation rate. Similar logic applies to other forms of inertial sensor, and other systematic errors.

Dead reckoning systems

These types of navigation system rely on the continuous updating of the position data derived from inputs of velocity components or speed and heading generated from a known start position. A simple example of this technique is a system which uses a compass heading in combination with a device, such as an odometer, to measure the distance travelled over the ground. Inertial navigation systems also fall into this category. The navigation accuracy that can be achieved using a dead reckoning system is largely influenced by the accuracy to which the start position, the velocity and the heading are known.

Direction cosine

A direction cosine is the cosine of the angle between two vectors. The orientation of a vector with respect to a given co-ordinate frame may be expressed in terms of three direction cosines defining the projection of the vector on to the axes of the frame. The attitude of a body frame with respect to a given reference frame may be expressed as a direction cosine matrix. This is a 3×3 matrix, the columns of which represent unit vectors in body axes projected along the reference axes.

Drift

Drift refers to the rate at which the error in a sensor or system accumulates with time. For example, a mechanical displacement gyroscope will provide a measure of the attitude of the body in which it is installed with respect to the spin axis of the gyroscope. The orientation of the spin axis will change with time as a result of unwanted torques acting on the rotor. As a result, a drift will be present on the measurement of attitude provided by the sensor which may be expressed in units of degrees per hour.

In an inertial navigation system, the estimates of position it provides will drift with time as a result of various errors within the system. Hence, it is customary to quantify navigation system performance in terms of a drift figure, often expressed in nautical miles per hour.

Euler angles

The attitude of a body with respect to a given reference frame may be specified in terms of three successive rotations about different axes. Hence, a transformation from the reference frame to the body frame can be carried out as follows:

- rotate through angle ψ about reference z -axis;
- rotate through angle θ about new y -axis;
- rotate through angle ϕ about new x -axis;

where ψ , θ and ϕ are referred to as the Euler rotation angles. It is of course possible to use the same angles in reverse order to define the transformation from body axes to reference axes.

The three angles correspond to the angles which would be measured between a set of mechanical gimbals, which is supporting a stable element or platform. The axes of the stable element represent the reference frame, whilst the body of the host vehicle is attached via a bearing to the outer gimbal. The order of the rotations is important and corresponds to the gimbal order.

Force-feedback

This is a technique used to return a sensitive element such as an accelerometer's proof mass or a gyroscope's rotor to its null position with respect to the case of the instrument. Null positions can be located and measured far more accurately than measurement of displacements. In some sensors, an electro-magnet is used to move the sensitive element to this zero position, the current flowing in the coils of the electro-magnet usually being proportional to the quantity being measured.

Gravitational model

This is the mathematical representation of the gravitational attraction to which a navigation system is subjected. Such a representation is required so that the specific force measurements provided by the accelerometers can be converted to true acceleration data. This is important for the navigation of vehicles in close proximity to the Earth as well as vehicles moving through interplanetary space. For navigation in the vicinity of the Earth, the Earth's gravitation can be defined to various levels of detail and complexity depending on the accuracy required from the navigation system. For very accurate navigation, it is required to take account of local variations or anomalies in the gravitational vector from the generalised mathematical model.

Guidance

This process refers to the direct control of the vehicle to constrain it to follow a pre-defined trajectory. For instance, consider an aircraft flying at constant speed which is required to follow a straight track between two way-points. It can reasonably be assumed that the aircraft has knowledge of where it should be at any time during the flight. Guidance commands can therefore be generated by differencing the desired and measured aircraft position. These commands are fed to the aerodynamic control surfaces causing the aircraft to move so as to null the detected path difference, thus ensuring that it flies along the desired track. This form of guidance is fairly common in strategic ballistic missiles and some long range tactical guided weapons.

Gyroscopes

These are the sensors which are usually used to measure rotational motion in an inertial system. These sensors commonly take the form of mechanical devices which

rely on the inertial properties of a spinning mass for their operation. For example, in a free (two-axis) gyroscope, the spin axis tends to remain fixed in space allowing rotational motion to be sensed about two orthogonal axes (which are nominally at right angles to the spin axes and to each other). Such devices produce measurements of turn angle with respect to inertial space. Single-axis mechanical gyroscopes are usually designed to provide measurements of turn rate with respect to inertial space.

There are a number of other devices that perform a similar function which rely on different physical phenomena for their operation. For instance, optical gyroscopes which provide a measure of angle or of angular rate by detecting some physical difference between two counter-propagating beams of light, using the Sagnac effect, and vibratory sensors which measure the Coriolis forces acting on masses which undergo linear vibration whilst being rotated.

Harmonisation

This term refers to the relative orientation of equipment case axes.

Inertial guidance

Guidance using knowledge of position, velocity and attitude derived from an inertial navigation system. The distinction is drawn between this process and inertial navigation that concerns a definition of position and velocity vectors, and their orientation, in a reference frame.

Inertial measurement unit

This term refers to a sub-set of the full inertial navigation system, the structure that contains the inertial sensors together with their support electronics and power supplies. The inertial measurement unit often contains the electronics required to operate the sensors and a micro-processor to compensate for some of the biases in the measurement signals generated by the sensors.

Inertial navigation

This is the process of establishing the position, velocity, attitude and heading of a vehicle using information derived from inertial sensors. Such systems are widely used for the navigation of aircraft, guided missiles, space vehicles, ships, submarines and land vehicles.

Integrated navigation

This is the process of combining data or information from two or more navigation systems, usually with complementary error characteristics, in order to produce a system with performance characteristics that surpass those of the component systems acting in isolation. Typical examples of such synergy are the use of position fixes derived from ground-based transmitters or satellite information, terrain reference systems and scene matching techniques to augment information derived from the inertial navigation system.

Integrated systems

This term refers to the combination of inertial data from different independent sources or techniques. For example, loosely coupled systems are essentially independent and data are only combined at the output level, whereas a deeply integrated system involves combination of inertial data at the earliest possible stage for superior performance, but offers very little redundancy.

MEMS

This term refers to micro machined electro-mechanical system technology using Coriolis force principle to detect angular motion and conventional displacement of a proof mass to measure acceleration. Such devices tend to be made from silicon or quartz crystals and are fabricated using well established semi-conductor processing techniques.

Performance

The accuracy with which an inertial navigation system can navigate a vehicle during a journey is governed by the accuracy of the data supplied to it at the commencement of navigation (i.e. the alignment accuracy), the quality of the inertial sensors it uses and the precision with which the navigation computation task is carried out. The errors grow with time and it is therefore customary to characterise inertial navigation system performance in terms of a drift in its navigational accuracy with respect to time. The drift is often specified in terms of nautical miles per hour, although the error growth is not a linear function of time in general.

'Full' inertial performance usually refers to sensors that give navigational accuracy of less than 1 nautical mile per hour. Such systems typically require gyroscopes with drift rates of $0.01^\circ/\text{h}$ or better and accelerometers which can provide measurements to an accuracy of 100 micro-g. The other term sometimes used when describing system performance is 'sub-inertial' which relates to tactical applications. Such systems are used for short duration navigation, or possibly to provide an attitude reference. Typically, sub-inertial systems use gyroscopes and accelerometers with measurement biases of the order of $1^\circ/\text{h}$ and 1 milli-g respectively, which if used in a full inertial navigation system would produce an error growth of ~ 1 nautical mile per minute.

Position fixing systems

Such systems operate by determining their position with respect to a known location, and rely on the observation of an object or effect at that location or a transmitted signal emanating from it. This form of system can determine its position when activated at any time during a journey provided the object or effect is observable. The earliest forms of navigation used this technique of navigation by observation and recognition. There are many techniques, both active and passive, that can be used by the observer to fix his position. Examples are signals from fixed radio beacons such as DECCA, OMEGA or LORAN and signals transmitted from satellites orbiting the Earth as in the case of GPS or GLONASS. Terrain referenced systems also fall

into this category. The received signals are used to derive position updates that are independent of the previous estimate of position. As a rule, such schemes only provide navigation data at discrete time intervals and not on a 'continuous' basis as in dead reckoning systems, although given a knowledge of prior events, some extrapolation is possible.

Quaternion

A four-parameter attitude representation. The quaternion attitude representation allows a transformation from one co-ordinate frame to another to be effected by a single rotation about a vector defined in the reference frame. The four elements of the quaternion are functions of the orientation of this vector and the magnitude of the rotation about it.

Reference frame

This refers to the set of axes to which the measurements and estimated quantities generated within an inertial system are referenced. The reference frame may be either a co-ordinate frame defined by three mutually orthogonal axes or a polar co-ordinate system. Various frames of reference are considered in the text of this book. An inertial reference frame is one which is fixed in space, that is, fixed in relation to the 'fixed' stars. For the purposes of navigation in the vicinity of the Earth, use is made of a geographic axis set which takes the polar form to give latitude, longitude and height above the Earth. In addition a local vertical geographic frame is often used. This is a Cartesian frame most commonly defined by the mutually orthogonal directions of true north, east and the local vertical (down).

Resolution

This is the mathematical process of calculating the various components of a vector quantity in a given co-ordinate reference frame when the vector is given in components referred to another frame. In a strapdown inertial navigation system, it is necessary to resolve the measurements of specific force, provided by the accelerometers in a body-axis reference frame, into the chosen navigation reference frame. The resolved specific force measurements can then be integrated to enable velocity and position with respect to the reference frame to be determined. The attitude of the body may be defined using various mathematical representations including Euler angles, direction cosines and the four parameter quaternion form.

The term resolution is also used in the context of inertial sensors, and refers to the smallest change in a measurement quantity which can be detected and distinguished by such a sensor. For instance, a particular type of gyroscope may be able to detect or resolve angular movements of 1 arc s.

Satellite navigation

Many modern navigation applications rely on radio data transmitted from a constellation of satellites surrounding the Earth. Simple trigonometry involving the timing of the arrival of these radio signals enable the position of the receiver to

be established. Examples of this are the American system known as GPS, the Russian system GLONASS and the future European system Galileo.

Various techniques have been devised to enhance the accuracy that can be achieved from GPS, as the fundamental accuracy is deliberately degraded by the controller of this system. Relative GPS refers to techniques that provide high relative positional accuracy between two GPS receivers, even though the absolute position of each receiver is not known precisely. This technique contrasts with absolute GPS where a single GPS receiver is used to determine the navigation estimates, and differential GPS where corrections are applied to absolute measurements made by a single receiver linked to another receiver whose position is well known. The differential technique can be used to correct an array of GPS receivers connected to the surveyed position via a data link, as described earlier.

Scale-factor

The scale-factor of an inertial sensor is the relationship between the output signal and the quantity which it is measuring. For instance, the scale-factor of a gyroscope which provides an analogue output may be expressed as so many millivolts per degree/second.

Stabilised platform

The original applications of the inertial navigation technology used stable platform techniques which are still in common use, particularly for ships and submarines when accurate navigation is required, unaided, over long time intervals. At the core of this system is a structure, often called the platform, on which inertial sensors are mounted. This platform is ‘isolated’ from the rotational motion of the vehicle using a number of gimbals arranged to provide at least three degrees of rotational freedom and so minimise the coupling between the vehicle and the platform.

The movement of these gimbals is controlled by torque motors which are activated by information provided by the gyroscopes. Additional gimbals may be used for some specific applications, particularly for very agile vehicles to prevent gimbal lock occurring.

The first practical realisation, called the geometric system, had at least five gimbals to provide the basic navigation data of latitude, longitude, roll, pitch and yaw in direct analogue form. These angles being available directly from the angular displacement of the various gimbals. This mechanisation instrumented two reference frames, an inertial non-rotating co-ordinate system and a local navigational co-ordinate system.

The geometric system was superseded by the so-called semi-analytic system which established one of the two reference frames instrumented by the geometric system. Generally, three gimbals are used in this mechanisation, to minimise coupling of the vehicle motion into the stable platform, with the coordinates of latitude and longitude being calculated in a computer. As in the geometric system, roll, pitch and yaw are deduced from the relative positions of the gimbals. A variety of different mechanisations of this type have been used. Some are used to establish an inertially

non-rotating reference frame, as in space stabilised systems. An alternative system commonly used for navigation in the vicinity of the Earth establishes a local level frame so that the input axes of two of the gyroscopes and two of the accelerometers are constrained to remain with their input axes in the horizontal plane, thus avoiding the explicit calculation of the gravity field vector.

Strapdown systems

In strapdown systems, the inertial sensors are fastened directly to the vehicle and hence are not isolated from its angular motion. Such systems are sometimes referred to as analytic systems. The signals produced by the inertial sensors are resolved mathematically in a computer prior to the usual calculation of navigation information. This use of a computer to establish and resolve the inertial data reduces the mechanical complexity of the inertial navigation system, thus frequently reducing the cost and size of the system and increasing its reliability.

Sub-inertial systems

See Performance.

Index

Index terms

Links

A

acceleration							
dependent bias	72	360					
insensitive bias	76	80	272				
matching	291						
sensitive bias	76	79					
squared sensitive bias	239	251					
accelerometer							
bias	30	258	284	297	304	338	
	347	350	355	366	433	456	
errors	295	350	361				
force-feedback	4	160	434	445	454		
silicon	13	97	165	434	436		
fibre optic	161	168	171				
Mach-Zehnder interferometric	168						
pendulous force-feedback	4	363	434	454			
photo-elastic fibre optic	171						
solid-state	161						
surface acoustic wave	163	178					
vibrating fibre optic	169						
aero-flexure compensation	470	493					
aerosols	482						
ageing and storage testing	242						
aided navigation	298	402					
aileron	462						

Index terms**Links**

alignment						
on a fixed platform	278					
on a moving platform	278	280	427	433		
on the ground	278	282				
in-flight	5	277	282	287	289	
shipboard	282	301	306	427	433	
alignment errors	244	282				
algorithms						
direction cosine	311					
quaternion	319					
orthogonalisation	322					
normalisation	322					
acceleration vector transformation	329					
rotation correction	326					
navigation	329					
altimeters	396					
analogue re-balance loop	70	449	451			
analog-to-digital conversion	269	271				
analytic gyro-compassing	279					
angle of dip	396					
angular acceleration sensitivity	73	373				
angular accelerometer						
gas rotor angular accelerometer	184					
liquid rotor angular accelerometer	183					
angular momentum	60	62	68	100		
vector	60	64	67			
gyroscopes	4	62	70	100	179	268
	373	449	483			
angular pick-off	40	66	477	481		

Index terms**Links**

angular rate	19	67	73	76	86	97
	132	172	182	194	204	264
	269	336	370	426		
angular rate matching	277	291	303			
anisoelasticity	84	239	366	368	433	
anisoelastic bias	72	76	80	240	254	
anisoinertia	72					
anisoinertia errors	72	162				
anodic bonding	192	202				
Aronowitz	117	118	120	125		
artificial horizon		12				
artillery pointing		505				
aspect ratio		462				
atomic clocks	381	386				
attitude						
computation		310				
reference	111	340	341			
representations		36				
attitude and heading reference system (AHRS)		264				
attitude control autopilot		460				
automatic target recognition		480				
autopilot	12	59	456	466		
axis rotations, definition of		37				

B

backscattering	120	141
balanced oscillator (MEMS)	193	
ball race	70	
band pass filters	490	

Index terms**Links**

bandwidth	141	177	203	209	233	251
	363	368	494			
bank-to-turn	457	461				
barometric altimeter	397	402				
baro-inertial system	401					
basalt	496					
bearings	70					
ball bearings	70	453	489			
gas bearings	71	453				
beat frequency	119	120	497			
bellows	70					
beryllium	68	104	160			
beryllium – copper	160					
bias						
accelerometer	284	297	350	356	358	360
gyroscope	285	354	357			
birefringence	129	134	136	139		
body axis set	9					
body frame	22	25	39	42	284	311
	322	392				
Bohneberger	12					
borehole surveying	442					
boron carbide	71					
Bortz	310	315				
bow tie – butterfly (hysteresis)	229					
Boykov	12					
Bragg cell acousto-optic modulator	132					
Bragg grating	171					
Brown and Perry	12					
Bryan factor	204					

Index terms**Links**

Bryan, G.H.	12	88				
built-in test equipment (BITE)	270					
C						
capacitive pick-off	91	94	166	196		
calibration	95	148	160	177	219	253
	416	421	438	448	455	493
cantilever	165	174	178	210		
capture rate	84	226				
Cardan joint	78					
Carousel	492					
Cartesian control	461					
Cartesian reference frame	7	264				
cavity	84	106	119	121	140	144
	497	499				
cavity path length control	119	125				
celestial navigation	391					
centripetal acceleration	27	33	48	64		
centrifuge	241	248				
centrifuge testing	235	247				
Cervit/Zerodur	122	166	497			
Charles Stark Draper	13	60				
chronometer	2	12				
Chu, S	143					
circular error probable (CEP)	411	519				
combination tests	241					
compass	11	395	444	461		
compass inclinometer	444					
compensation						
instrument	271					

Index terms**Links**

complementary filtering	401					
computer						
implementation	271	273	402			
update interval	314	319				
computation						
attitude	310					
frequency	332					
strapdown navigation	309					
conical scan	484					
coning						
motion	311	315	368	373	433	437
errors	369	374				
control law	477					
co-ordinate reference frame	37	301	410	427		
Coriolis						
acceleration	27	32	64	89	93	178
	193					
correction	30	34	325	330	332	436
equation	26	344				
forces	30	193	195	202	212	
couple (forces)	63					
cross axis sensitivity	249					
cross-coupling	76	97	174	182	254	346
	363	366	428	432	479	
cross-coupling coefficient	76	80	254			
cross eye	478					
cruise control	461	464	505			
crystal clocks	382					

Index terms**D**

		<u>Links</u>				
data highway	464					
data-logging equipment	222					
dead band	226					
dead reckoning	2	9	335	396	465	502
Decca navigation system	382					
declination	392	395	444			
defensive-aids suite (DAS)	472					
degree-of-freedom	74					
design example	403					
differential GPS	387					
differential laser gyroscope (DILAG)	122					
diffusion doping	168					
digital re-balance loop	70					
dip, angle of	396					
directed infrared counter-measures (DIRCM)	472					
direction cosine						
representation	272	324				
matrix	39					
propagation	39					
computation	283					
relationship with Euler angles, quaternions	41					
directional gyroscopes	12					
distance measuring equipment (DME)	381					
dither (ring laser gyroscope)	121	123	126	131		
dither product	121					
dither motion	121					
dither spillover	121					
Doppler radar	394					

Index terms**Links**

Doppler shift	386	388					
Draper Laboratories	96	195					
drift rate (gyroscope)	71	224	241	341	460	490	
dry tuned gyroscope		77					
dual-axis gyroscope	266						
dual-axis rate transducer (DART)	84						
dynamic range	13	100	116	155	158	182	
	207	210	223	310	447	449	
	478	488					
dynamically tuned gyroscope		77					

E

Earth, the shape of	49						
Earth-fixed reference frame	21	30	428				
Earth frame	28						
Earth line	497						
Earth's magnetic field	107	111	395	396			
Earthquakes	495	498					
Earth's rate	52	150					
Earth tide	498						
ejection seats	503						
elastic restraint	78	210					
electro magnetic radiation	115	168					
electromagnetic torque generator	67	69					
electrostatically suspended gyroscope	103						
elevator	461						
error(s)							
attitude	284	293	296	299	320	337	
	342	344	347	365	368	403	
	407	413	451				

Index terms**Links**

error(s) (*Continued*)

analysis	361						
budget	428						
computation	313						
coning	369						
de-correlation	367						
derivation of general equations	342						
manoeuvre dependence	366						
modelling	361						
position	282	298	327	336	343	347	
	352	359	366	396	403	407	
	410	413	432				
propagation	336						
pseudo motion	368	373					
direction cosine	321						
single-axis	341	358					
state space form	344						
velocity	292	294	296	300	327	337	
	354	356	358	360	403	432	
	451	461					
vibration dependence	368						
estimates, of vehicle velocity and position	19	26	276				
Euler angles	40						
Euler rotations	41	309					
external measurement	5	278	298	378			

F

Faraday rotation	122						
fault tolerance	269	275					
ferromagnetic materials	110	173					
ferroelectric materials	175						

Index terms**Links**

fibre optic accelerometer	168					
fibre optic gyroscope	126					
fibre optic rate sensors	13					
fibre optic ring resonator gyroscope	140					
field of regard	476	478	487	489		
finesse	140	497				
fixed bias (<i>g</i> -independent)	72	156				
fixed stars	1	21	61			
flex gyroscope	81					
flexible pivot	82					
flexure (tuned rotor gyroscope)	77					
flexure						
aircraft	295					
hinges	77	160				
ship	300					
float (gyroscope)	70	76	180			
flotation fluid	69	70	75			
fluidic (fleuric) sensors	105					
flywheel	68	465				
force-feedback	13	155	156			
force-feedback pendulous accelerometer	157					
former	12	70	134	169	239	253
	389	396	486	519		
Foucault	12	88	359			
Foucault oscillation	346	350				
four parameter attitude representation	42					
free gyroscope	71	486				
frequency difference	119	142	204			
frequency nulling	132					
frit sealing	123					

Index terms**Links****G**

gain control	95	119	204			
Galileo satellite navigation system	390					
gas rotor angular accelerometer	184					
Gaussian distribution	520					
Gaussian white noise	293	405				
Geodesy	51	441	496			
Geodetic models	51					
geographic coordinates	19	31	51	53	278	
geographic frame	19	47	56	276	279	280
	282	286	307	318	330	338
	366	392	396			
geographic reference frame	19	47	273	279	283	337
	445	455				
Geoid	51					
G-ring (RLG)	496	498				
gimbal	42	47	61	65	70	76
	96	426	454	477	489	
gimbal lock	47	61	68	488		
Global positioning system (GPS)	143	385				
Global navigation satellite system (GLONASS)			389			
Graseby table	223					
graviomagnetic effects	489					
gravity model	19	36				
gravity gradiometry	56					
gravitation vector	23					
gravitational field	10	12	23	32	48	51
	54	336				
gravitational acceleration	21	27	147	154	248	

Index terms**Links**

gravitational anomalies	48	56				
gravitational attraction	55	145	155	273	275	338
Greenwich meridian	1	21				
ground-based tracking station	386					
ground speed	24	26	28	32	36	
gyrocompassing	277	279	285	367		
Gyromagnetic ratio	100					
gyroscope						
angular momentum	179	268				
conventional	60	84	99	104	234	254
dynamically tuned	258	362	432			
fibre optic						
flex	77					
laser	126					
mechanical	81					
NMR	116					
rate integrating	68					
transmission line	59	100				
vibratory	73					
gyroscopic	112					
ball	88					
inertia	104					
measurements	60	84	483			
precession	293	341	350			
theodolite	62					
gyrotron	505					

H

hardware in the loop testing	259
health of tool	449

Index terms**Links**

heat seeking missiles	462	472	482				
helium-neon gas laser	119	496					
Helmholtz coil	233	236	246				
Hooke's joint	77	82					
hyperbolic navigation	382						
hysteresis	158	160	226	229	231	452	
hysteresis motor	70						

I

Ibot	441	467					
inclinometer	185						
incremental angle	269	315	326				
incremental measurements	270	315					
incremental velocity	269	325					
inertial element	61	68	70	81	84	100	
	104	487					
inertial frame	26						
inertial guidance	13	34	422				
inertial measurement matching	289	301					
inertial measurement unit (IMU)	97	134	174	181	189	212	
	216	256	258	264	270	454	
	467	474	481	502			
inertial navigation							
error budgets		5					
system		14					
system performance	5	335	341	359	371	421	
	433	436	454				
inertial reference frame	1	21	25	61	461		
inertial space	9	19	25	55	62	64	
	78	104	117	154	369	478	
	487	490					

Index terms**Links**

in-flight alignment	289					
initialisation	5					
in-plane device	194	206				
input axis	66	69	73	80	86	91
	94	125	132	156	178	181
	184	195	201	204	224	226
	231	241	243	248	251	254
	369	450	496			
input axis misalignment	282					
input range	95	159	163	167	177	179
	494					
instantaneous speed	7					
instrument						
cluster	264					
compensation	253					
electronics	269					
power supplies	111	269				
re-balance loops	270					
rectification errors	368					
intelligent braking system	466					
intelligent transport systems	464	467				
integrated navigation system	377					
integrated optical components	127					
integration	2	7	8	21	23	70
	172	271	330	391	413	418
	470	476				
interferometer	115	119	128	130	144	168
in-run bias variation	73					
in-run random bias	73	362				

Index terms**Links****J**

jewelled pivot hinges	160
jitter	478 490
Johnson	12

K

Kalman filtering	258 285 292 402 449 451
	456
Kerr effect	121 142

L

Ladar	479
Lambert conical projection	54
Laplace operator	340 352
Larmor frequencies	101
Larmor precession	100 102
laser	
beam director	
gyroscope	4 13 15 60 103 112
	115 118 137 142 144 150
	230 241 254 256 268 270
	285 434 448 454 496
radar	479
radar seeker	480
lasing threshold	142 159 163 167 174 177
	179 184 186 226 243
Latax	457
lateral autopilot	458

Index terms**Links**

latitude	1 225 379	31 273 383	36 276 392	47 288 448	50 346 453	111 358 456
			532			
Law of gyroscopes		65				
least-squares		72	156	258	267	515
Lemarle		12				
Lense-Thirring Field		499				
lever-arm motion		290	295			
Lidar		479				
linear accelerometer		155	185	372	458	
line-of-sight isolation		487				
line-of-sight stabilisation		480	487	491		
lines of position		31	53	392		
liquid potentiometers		468				
local geographic frame		19 286 396	21 307 524	47 318 338	56 366	276 366 392
local gravity vector		27 293	48 297	51 344	56	223 256
local vertical		19 302 396	30 337 427	104 339 545	258 367	278 391
local vertical reference frame		20				
lock-in		120				
lock-in rate		121	141	164	256	
lodestone		11				
Long-range navigation (LORAN)		216	384			
long-term stability		244	382	395		
longitude		1 276	12 341	31 379	36 383	51 392
						273 456

Index terms**Links**

lubricants (mechanical gyroscope)	70	99
M		
Mach-Zehnder accelerometer	144	168
magnetic anomalies	110	396
magnetic declination	395	444
magnetic measurements	395	
magnetic mirror	121	
magnetic mirror biasing	121	
magnetic sensitivity	247	
magnetic sensitivity tests	233	239 246
magnetic shielding	82	86 136
magnetic tuning	83	
magnetism	11	
magnetometer	107	
magnetohydrodynamic gyroscope	86	
map matching algorithm	465	
mass attraction	34	56
mass unbalance	72	76 79 104 174 180
	210	212 366
mass unbalanced gyroscope	179	
measurement while drilling (MWD)	444	446 450 453
mechanical dither	121	123 126
mechanical gimbals	38	47 525 542
mechanical gyroscope	68	
mechanical sensors	155	
mechanical stops	68	
Meissner effect	102	
MEMS technology	95	192 195 210

Index terms**Links**

meridian						
plane		56				
Michelson interferometer		115				
Michelson, A. H.		115				
mid-course guidance commands		30				
missile approach warning sensors		473				
missile guidance	14	457				
moment of inertia	62	67	373	450		
motion compensation		487				
moving coil		69				
moving mop displays		499				
multi-sensor	178	187	212	265	417	
multi-sensor technology		153				
multi-functional sensors		174				
rotating devices		174				
vibratory devices		178				
multi-positional tests		452				
mutual inductance		157				
MWD systems	444	446	451			

N

Nadir	47	488				
navigation	1					
aids	379					
external measurements	298	378	418			
on-board measurements	394					
computation	329	332				
equation	47					
frame	31					
reference frame	256	258	271	324	367	395
	524	545				

Index terms**Links**

navigation (<i>Continued</i>)							
system performance		5	335	341	359	371	421
		433	436	454	541		
navigational grid		49					
Navstar-GPS		325					
natural frequency		67	99	233	368		
negative stability		461					
Newton		2	12				
Newton's laws		63					
non-commutative		37					
non-commutativity effects		316					
non-gravitational force		10	154				
non-linearity							
gyroscope		231					
accelerometer		156	161				
normal distribution		520					
normalisation		322					
northfmder		287					
north line		282					
north seeking device		444					
notch filters		490					
nuclear magnetic resonance gyroscope		100					
null		66	69	74	97	156	186
nulling		67	71	76	110	132	156
		168	186	284	483		
nutation		67	105	479	484		
nutation frequency		67					

Index terms**Links****O**

Oblate spheroid	51
obstacle avoidance	480
odometer	465 501 506
off-axis sensors	212
Omega	383
one-shot transfer alignment	289
operating temperature	85 88 99 161 195 447 449
optical fibre accelerometer	161
optical path difference	118
optical rate sensors, gyroscopes	4 60 268
optical cavity	118 122
ordnance datum	57
orthogonal sensor configurations	264
orthogonalisation	322
oscillating rate table	234
oscillating rate table tests	233
outer gimbal axis	68 489
out-of-plane (z-plane)	206
output angle	490
output axis	66 70 74 80 369
over-range testing	248
over-volting	435
own-ship motion	487 491

P

PADS	505
path difference	116 118 128
path length (optical gyros)	116 125 128 132

Index terms**Links**

performance analytical assessment of						
motion dependency	365					
simulation assessment of	360					
pendulosity	12					
pendulous force-feedback accelerometer	4	434	454			
pendulous accelerometer hinge elements						
flexure hinges	160					
jewelled pivot hinges	160					
pendulous axis	157	178	180	207	243	251
	255					
pendulous integrating gyro accelerometer	269					
personal transport	441	467				
phase nulling	132					
Philpot and Mitchell	77					
photo-elastic fibre optic accelerometer	171					
pick-off	61					
angle	40	67	74	77		
capacitive	91	166	196			
electromagnetic	69					
optical	115	212				
variable reluctance	69					
piezoelectric	94	98	122	132	163	169
	174	195	201	251		
piezoelectric crystals	85	89	177			
piezoelectric transducer	167	174				
planar ring structure	203					
platform	3	8	248	290	366	368
	416	487				
platform mechanisation	278	309				
platform, stabilised	13	201	278	467	487	492
pointing stability	478					

Index terms**Links**

polysilicon	168					
position						
fixing	1	5	11	384	400	
update alignment	298					
potentiometer	69					
precession	62					
precession axis	63	67				
precision dividing head	243	246				
proof mass	154	159	161	171	183	186
	196	199	207	210		
proportional navigation	483					
pseudo-coning motion	368	450				

Q

<i>Q</i> -factor	91	97	204	209		
quadruplex redundancy	268					
quantisation	277					
quartz	89	91	95	161	194	201
	208	214				
quartz rate sensor	94	214	475			
quartz resonator	89	91	167			
quaternion						
computation	42	320				
parameters	44	271	279	322	324	
propagation	44					
relationship with Euler angles, direction						
cosines	45					
representation	38	42	320			

Index terms**Links****R**

radial probability distribution	521					
radar altimeter	358	397	400			
radio beacon systems	282	378				
radio navigation aids	379					
random bias	73	76	80	125	158	254
	455					
random drift rate	224	234	241			
random walk	121	125	137	200	285	448
random walk error	125	285	362	455		
rate gyroscope	17	66	118	182	458	460
	477	481	487	490		
rate-integrating gyroscope	13	59	70	73	80	81
	183	265	373	434	436	453
	489	505				
rate measurements	25	80	258	267	282	302
	414	436				
rate of turn indicators	12					
rate sensor	88	94	98	103	201	214
	462	468	475	477		
rate table	220	226	229			
rate transducers	59					
rate transfer tests	226					
Rayleigh distribution	521					
reaction torque (or force)	65					
re-balance loop	69	158	182	222	270	
receiver coil	69					
rectangular integration	331					
rectification error	450					
redundancy	268	275	411			

Index terms**Links**

redundant sensor configurations	268					
reference axis set	179					
reference ellipsoid	49	50	53			
reference frame						
drag	499					
inertial, space-fixed	18	23				
rotating	24	118	337			
refractive index	122	134	136	139	142	
relative angular alignment	491					
relative misalignment	491					
remote sensing	476					
resolution, specific force measurements	25	274				
resonance						
search	250					
resonant frequency	89	95	162	204	207	212
	490					
restrained pendulum accelerometer	157					
restraining spring	66					
retainer	70					
reticle	484					
ring laser gyroscope	116					
ring laser gyroscope – three-axis device	126					
ring resonator gyroscope	142					
roll autopilot	458					
root sum squared (RSS) error	429					
rotation angle	315					
rotational motion	2	3	8	12	60	62
	68	120	155	448	478	524
rotation correction	326					
rotation vector	316					

Index terms

rotor	
angular momentum	62
moment of inertia	67
speed	78
run-up time	289

S

safety and arming units	499
Sagnac effect	140 144
SARDIN	103
satellite navigation aids	384
scale-factor	
asymmetry	72
error	72 76 121 125 158 254
linearity	137 205 244 368 452
scene matching area correlation (SMAC)	398 417
Schuler, M	12 339
Schuler loop	340 352 356 368
Schuler pendulum	339
Schuler pumping	368
Schuler tuning	12 341 358
Scrolling	332
sculling motion	328 368 373 433
sculling errors	371
seeker	422 458 480 483
seeker head stabilisation	482
segway transporter	467
Segway	1
seismic	
events	495 498
waves	496

Links

Index terms**Links**

seismometer	498				
self-alignment	256	278	283	285	
sensor errors	71				
sensor re-balance loop	448				
sextant	11	12			
shelf life	70	99	134	242	
ship flexure	300				
shipboard alignment	301				
shipboard inertial measurement matching	301				
ship's inertial navigation system (SINS)	453				
ship's log velocity	454				
shock	93	105	442	447	
shock testing	237	250			
shoot and scoot	505				
short term navigation	30	352			
Shupe (winding)	134				
side slip sensors	12				
sidereal day	52				
sidereal hour angle	392				
sightline stabilisation	487				
silicon					
accelerometers	13	97			
carbide	160				
dioxide	165				
oil	183				
sensor	96				
silicone fluid	160				
simple oscillator (MEMS)	193				
Simpson's rule	331				

Index terms**Links**

simulation							
adjoint		365					
co variance		364					
time stepping		364					
sine-sweep		239					
single-axis sensors	9	69	479				
singularities	22	36	271	276	283	331	
		488					
situational awareness		472	481				
size effect		214	372				
skew angle		266					
skew symmetric matrix		25	36	292	311	313	342
		362					
skewed sensor configurations		265					
skid-to-turn		461					
slip rings		174	177				
solar day		52					
solid-state accelerometers		161					
solid-state sensors		13	99	153	190		
space-fixed coordinates		344					
space-fixed reference frame		18	23				
specific force		10	24	27	32	44	154
		205	264	273	278	283	286
		293	312	324	332	338	343
		364	367	371			
specific force measurements		11	19	24	28	47	155
		191	272	278	280	302	336
		362	366				
speedometer		465					
Sperry (brothers)		12					
spider		82	83				

Index terms**Links**

spin axis	12	60	71	74	76	80
	85	100	104	175	180	267
	287	463	485			
spin motor	64	70	76	270	486	
spinning mass gyroscopes	60	72	270	366	368	373
SPIRE		3				
spirit level		468				
SQUID magnetometer		102				
stabilised mirror system		488				
stability	91	106	119	125	132	174
	199	209	215	223	228	233
	478					
stability tests						
gyros		223				
accelerometers		244				
stable element	38	280				
stable platform	3	13	284	391	453	
stable platform inertial navigation system	40	282	284	342	365	
stable platform technology		13				
standard mercator projection		53				
star trackers		391				
static margin		461				
stationary reference frame		118				
strapdown configurations						
earth frame mechanisation		28				
inertial frame mechanisation		26				
navigation frame mechanisation		31				
wander azimuth mechanisation		34				
strapdown inertial navigation system		17				
strapdown system components		47				
strapdown technology	3	4	291	326	454	

Index terms**Links**

strapdown, two-dimensional system	17					
sub-inertial quality systems	206	296	408			
superluminescent diode	136					
surface acoustic wave accelerometer	163					
surface radar trackers	393					
suspension	14	72	77	81	104	157
	161	180	441	464	467	486
suspension and braking systems	465					
switch-on to switch-on bias variation	73	157	223	255	452	
synchronous motor	78	86				
system errors	255	258	336	340	358	361
	368	406	408	446	451	
system integration	215					
system performance	335					
systematic error	77	253	438	456	474	

T

tacho-generator	231					
tactical air navigation system (TAGAN)	382					
system tangent plane	30					
technology insertion	470					
temperature	70	72	76	97	102	106
	124	134	171	214	222	231
	244	246	270	438	442	453
	479	497				
temperature-dependent bias	73					
temperature sensitivity	86	91	247	270		
terrain referenced navigation	397					
terrestrial inertial navigation	21					
test equipment	221	239				

Index terms**Links**

testing							
accelerometers		243					
dynamic		220					
environmental		494					
gyroscopes		223					
static		220					
thermal		231					
theodolite		282	505				
threshold		142	159	160	163	167	174
		177	179	184	186	226	243
thrust vector control		503					
tie bars		166					
time reversal invariance		498					
tiltmeter		498					
tilt sensors		468					
tombstone tests		244					
topographical models		51					
torque		62	64	69	75	79	85
		100	202				
torque axis		63	65	67			
torque (force) re-balance accelerometer		182					
torque generator		67	69	76	183		
torque motor, torquer							
electromagnet		69					
permanent magnet		69					
torque re-balance		69	78	180			
TRANSIT		384					
transfer alignment		289	299				
transition matrix		295	352	404	406		
translation motion		153					

Index terms**Links**

transmission line gyroscope	112						
transport rate	19	21	34	47	50	339	344
trapezoidal integration	331						
trials	242	300	453	494			
truncation, of mathematical functions	363	437					
tumbling	61	68					
tuned rotor gyroscope	77	212					
tuned speed	78						
tuning fork sensor	94						
turbulence	469	482					
turn rate							
twin-axis gyroscope	59	228					
two-degrees-of-freedom gyroscope	61	67	287				
two-dimensional strapdown navigation system	17						

U

ultra-high frequency (UHF)	382						
Universal Transverse Mercator (UTM)	54						
unpredictable errors	157						
urban canyons	465						

V

variable reluctance	69						
vehicle stabilisation	456						
velocity							
matching	277	292	296	301			
vector steering	460						
vertical plane	19	467					

Index terms**Links**

very high frequency omni-directional radio	
range (VOR)	379
vibrating disc sensor	93
vibrating fibre optic accelerometer	169
vibrating string accelerometer	98
vibrating wine glass sensor	89
vibrating wire rate sensor	98
vibration-dependent errors	255
vibration sensitivity	126 222
vibration table	239 241 251
vibration tests	238
vibratory gyroscope	88
vibro-pendulous error – accelerometer	250
viscosity	70 157 177
viscous coupling	85

W

wander azimuth	
frame	34
system	36
warm-up time	255
way point	1 542
weathercock model	458
wheel speed sensitivity	466
wire line tools	444
Wrigley,W	12 15

Y

Young's modulus	190
-----------------	-----

Index terms**Z**

zero mean random bias

76 80 254

zero offset

158

Zerodur, Cervit

122

Links