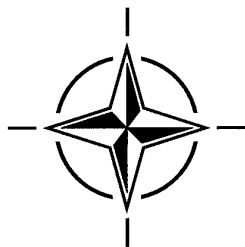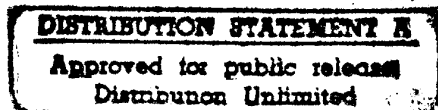# AGARD

ADVISORY GROUP FOR AEROSPACE RESEARCH & DEVELOPMENT

7 RUE ANCELLE, 92200 NEUILLY-SUR-SEINE, FRANCE

AGARDograph 331

Aerospace Navigation Systems
(Les systèmes de navigation aérospatiaux)

DTIC
S ELECTE D
AUG 2 9 1995
F

This AGARDograph has been sponsored by the Guidance and Control Panel of AGARD.

NORTH ATLANTIC TREATY ORGANIZATION

Published June 1995

Distribution and Availability on Back Cover

AGARD-AG-331

AGARD

ADVISORY GROUP FOR AEROSPACE RESEARCH & DEVELOPMENT

7 RUE ANCELLE, 92200 NEUILLY-SUR-SEINE, FRANCE

AGARDograph 331

Aerospace Navigation Systems

(Les systèmes de navigation aérospatiaux)

by

J. Niemela, M. Kumar, B. Stieler, H. Buell, D. Boozer, R. Langley,
D. Liang, S. Levine, D. Moore, S. Lewantowicz, R. Paschall, N. Carlson,
R. Greenspan, D. DiFilippo, S. Ausman, C. Bedoya, J. Miller and C. Hunt

Edited by

John Niemela, Programme Director

Accesion For		
NTIS CRA&I		☑
DTIC TAB		☐
Unannounced		☐
Justification		
By		
Distribution /		
Availability Codes		
Dist	Avail and / or Special	
A-1		

North Atlantic Treaty Organization
Organisation du Traité de l'Atlantique Nord

19950825 111

The Mission of AGARD

According to its Charter, the mission of AGARD is to bring together the leading personalities of the NATO nations in the fields of science and technology relating to aerospace for the following purposes:

— Recommending effective ways for the member nations to use their research and development capabilities for the common benefit of the NATO community;

— Providing scientific and technical advice and assistance to the Military Committee in the field of aerospace research and development (with particular regard to its military application);

— Continuously stimulating advances in the aerospace sciences relevant to strengthening the common defence posture;

— Improving the co-operation among member nations in aerospace research and development;

— Exchange of scientific and technical information;

— Providing assistance to member nations for the purpose of increasing their scientific and technical potential;

— Rendering scientific and technical assistance, as requested, to other NATO bodies and to member nations in connection with research and development problems in the aerospace field.

The highest authority within AGARD is the National Delegates Board consisting of officially appointed senior representatives from each member nation. The mission of AGARD is carried out through the Panels which are composed of experts appointed by the National Delegates, the Consultant and Exchange Programme and the Aerospace Applications Studies Programme. The results of AGARD work are reported to the member nations and the NATO Authorities through the AGARD series of publications of which this is one.

Participation in AGARD activities is by invitation only and is normally limited to citizens of the NATO nations.

Preface

The need for an up to date, comprehensive treatise on aerospace navigation systems has been recognized by the Guidance, Control and Navigation community. Such a document will serve as a lasting reference as did material prepared in the late 60's and early 70's. These earlier documents, though remarkably prescient in their technical forecasts, have been overtaken by very rapidly advancing technology.

This new document should be of interest to those who are involved in the integration of navigation equipment aboard an aerospace vehicle and who may not be knowledgeable about, but would like to have an expert's perspective on the capabilities and limitations of the various navigation sensors, integration techniques and concepts. In addition, it is expected that navigation specialists will have an interest in broadening their understanding of aerospace navigation. Great care has been given to the selection of references to previously published material.

The editor would like to acknowledge Dr. George Schmidt of the Charles Stark Draper Laboratory, who conceived of this AGARDograph and Ir Pieter Ph van den Broek of Delft University of Technology, who provided many constructive suggestions in the course of its preparation.

JOHN NIEMELA
U.S. Army Member
AGARD Guidance and Control Panel

Préface

La communauté Guidage, Pilotage et Navigation reconnaît la nécessité d'un traité moderne et complet sur les systèmes de navigation aérospatiaux. Un tel document servirait de référence permanente, à l'instar des textes établis pendant les années '60 et '70. Ces documents, d'une prescience remarquable dans leurs prévisions techniques, ont toutefois été vite dépassés par des technologies dont l'avancée s'est révélée très rapide.

Ce nouveau document devrait intéresser tous ceux qui sont impliqués dans l'intégration des équipements de navigation dans les véhicules aérospatiaux et qui ne connaîtraient peut-être pas, mais qui aimeraient connaître, le point de vue d'un expert sur les capacités et les limitations des différents senseurs de navigation, des techniques et des concepts d'intégration. En outre, il est prévisible que les spécialistes en navigation voudront élargir le champ de leurs connaissances dans le domaine de la navigation aérospatiale. Un soin particulier a été apporté à la sélection des références qui sont faites aux textes déjà publiés.

Le rédacteur tient à remercier le Dr George Schmidt du Charles Stark Draper Laboratory, qui a conçu cette AGARDographie, ainsi que l'Ir Pieter Ph van den Broek de l'Université de technologie de Delft pour les nombreux conseils positifs qu'il a bien voulu fournir lors de l'élaboration du manuscrit.

JOHN NIEMELA
U.S. Army Member
AGARD Guidance and Control Panel

Guidance and Control Panel

Chairman: Mr. S. Leek
PB 121
British Aerospace Defence Ltd
PO Box 19
Six Hills Way, Stevenage
Herts SGI 2DA
United Kingdom

Deputy Chairman: Mr. J.K. Ramage
Chief,
Flight Control Systems Branch
WL/FIGS, Bldg 146
2210 Eighth St, Ste 11
Wright-Patterson AFB, OH 45433-7521
United States

TECHNICAL PROGRAMME

PANEL EXECUTIVE

ACKNOWLEDGEMENT/REMERCIEMENTS

Contents

SECTION I

INTRODUCTION

DR. JOHN NIEMELA
U.S. ARMY COMMUNICATIONS — ELECTRONICS COMMAND
RESEARCH AND DEVELOPMENT CENTER
FORT MONMOUTH, NJ 07703-5000
USA

1. MOTIVATION AND OVERVIEW

The need for an up to date, comprehensive treatise on aerospace navigation systems has been recognized by the navigation community. Such a document will serve as a lasting reference as did previous material prepared in the late 60's and early 70's. [1] These earlier documents, though remarkably prescient in their technical forecasts, have been overtaken by very rapidly advancing technology.

It is anticipated that the target reader of this document will be an individual who has responsibility for the integration of navigation equipment aboard an aerospace vehicle. He may not have, but is desirous of obtaining, an expert's perspective on the capabilities and limitations of the various navigation sensor and integration techniques. In addition, it is expected that navigation specialists will have an interest in broadening their understanding of aerospace navigation. Great care has been given to the use of references to previously published material.

The AGARDograph is organized into six sections with this section providing the motivation for establishing the requirements to assure that the development of an aerospace navigation system will meet its operational requirement. The relationship between system characteristics and the mission scenario has been often vaguely defined and typically requires much attention prior to development. Included is a discussion of the methods to rigorously specify the technical requirements to meet the operational requirement. The second section of this document reviews navigation coordinate frames with a discussion of inertial, terrestrial and geodetic coordinate systems.

The third section of the document describes the navigation sensor technologies that are employed in modern aerospace navigation systems. Emphasis is placed on the physical principles of operation rather than specific or unique implementations.

The fourth section of this document addresses system analysis and synthesis methods. Examination and trade-off of the technology, for each sensing, processing and control and display element, generates alternative preliminary designs. This section discusses the identification and

partitioning of the navigation system functions and definition of the interface between functional elements and the rest of the aerospace system.

A vital aspect in the development of an integrated navigation system is the processing algorithm that combines the navigation sensor data. Also in this section is a methodology for Kalman filter development that includes error models, covariance and Monte Carlo simulation techniques with means to verify the error models.

The fifth section of the document describes representative state-of-the-art navigation system implementations in fixed wing, rotary wing aircraft and spacecraft.

The sixth and concluding section of the document describes the various test methods employed to verify the performance of aerospace navigation systems. This includes instrumentation techniques with emphasis on standard statistical methods used by NATO countries. This discussion provides closure with the first section on requirements by relating the means to verify that the navigation performance meets the technical and operational requirement.

In general, modern aerospace navigation systems are comprised of more than one and in many instances several independent subsystems. This

considerably complicates the process of translating the operational requirement to a technical requirement. The importance of a concise statement of the technical requirements on which to base the development of a system design is of paramount importance. For this reason, the following discussion outlines the critical steps in development of the technical requirement that is based on an explicit or implicit operational requirement. The methodology discussed draws in part on the experience gained in formulation of the EFA - [2].

2. OPERATIONAL AND TECHNICAL REQUIREMENTS FOR AEROSPACE NAVIGATION SYSTEMS

2.1 OPERATIONAL REQUIREMENTS

The need for a navigation capability aboard aerospace vehicles is dependent on many qualitative and quantitive factors. For the design to fulfill the need requires that the operational requirement, which is often stated in terms of a mission scenario and an aerospace vehicle capability, be faithfully translated into a technical requirement. The technical requirement can then be stated in terms of parameters on which a design can be accomplished.

In virtually all modern aerospace vehicles the navigation system must provide more functionality than the classical service of providing direction and distance to a succession of waypoints. In many instances the navigation

system must provide data to other on-board systems so that they may accomplish their function to include flight control, weapon direction, envelope limiting, approach and landing, to name a few. Similarly, in modern aerospace system designs, the navigation system is often dependent on information and processing assets residing outside the confines of the navigation system. In these highly integrated designs it is difficult, if not impossible, to circumscribe the navigation system as part of the aerospace system. Consequently, the above mentioned navigation functionality must be provided with appropriate levels of performance, to include accuracy and reliability.

The complexity of these modern aerospace system designs makes traceability from the operational requirement to the technical requirement to the navigation system design of paramount importance. For this to be accomplished requires that the operational requirement be described in sufficient detail, particularly scenario and aerospace vehicle characteristics that impact the technical requirement of a navigation system. Below are listed some of the principle considerations and factors.

  ° Detailed description of nominal mission (altitude, velocity, dynamics) by mission segment.

  ° Accuracy for each mission segment

  ° Reliability

  ° Availability

  ° Size

  ° Weight

  ° Power

  ° ECM Environment

2.2 Technical Requirement

Requirements for aerospace platforms often give overall descriptions of the functions needed to carry out the basic missions of the platform but leave the requirements for the navigation system rather vague. In the past it has been adequate to specify that the platform shall carry, for instance, an inertial navigator with a given general performance. However, the introduction of satellite navigation systems whose performance may be substantially improved by integration with dead reckoning systems has changed the way in which navigation system requirements are to be defined.

Where the sensors to be used in a navigation system are likely to be affected by the maneuvers of the vehicle, their location, and activities of hostile parties, are now required to be specified in much more detail than previously. It is to assist in the preparation of the technical requirements that this portion of the paper is offered.

2.2.1 Characteristics of Integrated Navigation Systems

In general, an integrated system will consist of a number of subsystems, linked together by means of interfaces and software that are located in processors which may be separate devices or integral parts of one or more of the sensors. Each of the subsystems may be capable of independent operation under some conditions but will benefit by making use of information derived from other subsystems. In the primary mode of operation all subsystems will be contributing information to the integrating computer and the resulting total system will have greater accuracy than any individual subsystem.

An important aspect of the integration is that many of the deterministic errors of the subsystems may be calibrated during the integration process by cross comparison of the subsystem outputs. This has the effect that when one or more subsystems become non-operative, e.g., due to terrain screening, the other systems may continue at a higher level of accuracy than would otherwise be possible. The accuracy will degrade slowly, depending on the variability of the calibrated error sources, rather than suddenly as would occur with an uncalibrated sensor. When operation of the missing sensor is restored, the system will once again return to full operation.

Thus an integrated navigation system has the property of robust operation with graceful degradation in the event of sensor non-operation and the ability to recover full capability should the non-operating sensors become operational.

2.2.2 Relation to the Operational Requirement

Generally, an operational requirement will give general descriptions of the scenarios of operation and the desired mission capabilities of an aerospace platform. An integrated navigation system will have various modes of operation which will react in various ways to the particular environment in which they operate. It is therefore necessary that the operational requirement state prospective missions in some detail, and to define the requirements for the navigation system for each stage of the mission. The technical requirement must translate the several mission capabilities and environments into technical capabilities and parameters on which a system design can be developed.

For instance, the initial part of a sortie may be carried out under benign conditions during which time the integrated system is able to calibrate the sub-system error sources. Knowledge of the duration of this period is important in determining the accuracy to which the errors may be calibrated. The sortie may then continue with a low level penetration in rugged terrain

during which intermittent acquisition of satellites may occur. The duration and conditions of this phase and the required navigation accuracy need to be specified.

Following the penetration phase, the attack phase may be in less rugged terrain but with enemy jamming in operation. The level of jamming needs to be specified in order to determine the optimum level of rate aiding for the satellite system.

During each mission segment, evasive or attack maneuvers may be carried out. These will have an effect on the INS, particularly if it is a strapped down system, and the likely maneuver levels and allowable overall system accuracy should be specified.

2.2.3 Design Constraints and Degrees of Freedom

The requirements for the system accuracy should be stated and the times and conditions under which these are to be met, defined. However, such technical requirements should not define the details of the integrated system mechanization. With these systems, there is a considerable trade-off of complexity, accuracy, integrity and cost which should be the prerogative of the system designers to determine, within the constraints laid down by the technical requirement. Implicit in the process is the need for close liaison between the drafters of the technical requirement and system designers.

2.2.4 Affordability

An important aspect of all system developments is that of affordability. Most new requirements call for "state-of-the-art" performance which is not readily met and could be very costly. There is a need to specify the allowable cost for a system designed to meet the requirements, and those areas in which performance (in the general sense) may be traded for cost, and any areas in which such trade-offs are not permitted.

2.2.5 Integrity Aspects

The requirements on the navigation subsystem which are necessary to enable the total aircraft system to meet the mission success requirements must be specified. The specification should state any constraints on the methods which are to be used to detect and handle the loss or degradation of data from individual sensors. This may be necessary to meet flight safety or operational requirements.

The specification should distinguish between the methods being used to ensure adequate integrity for the flight safety requirements and those needed to meet the mission success requirements.

2.2.6 Factors Contained in a Technical Requirement

The technical requirement for an integrated navigation system is frequently developed and integrated into technical requirement for an aerospace vehicle. In this form, or as a stand-alone document, a typical technical requirement

for an integrated navigation system should address the following points:

° The functionality and performance * required from the system at each stage of flight.

° The characteristics of each flight segment in terms of:

- The duration of the segment

- The occurrence of the segment in relationship to other segment (e.g. the number of times it is likely to occur, will it always follow other segments, etc.)

- The aircraft linear and angular acceleration and velocity conditions likely to occur in each segment

- The nature of external conditions at each segment (e.g. altitude, terrain, jamming conditions, etc.)

- The integrity requirements at each segment, including the allowable degradation in function and performance

° The system reliability and maintainability requirements.

° The maximum acceptable system size, weight, power and cooling requirements.

° The maximum cost of the system, both in development, acquisition and life cycle cost terms, and areas where cost trade-offs cannot be permitted.

3. SUMMARY

The purpose of an integrated navigation system is twofold - to provide information to the operator and data to other systems of an aerospace vehicle with sufficient accuracy to accomplish the mission. Implicit is the function of positioning and generation of guidance information to a destination defined within a specific coordinate frame. The next section of this document discusses navigation coordinate frames used for computational mechanization and display to the aerospace vehicle crew.

[*]   The means to rigorously specify navigation accuracy has been given much attention by navigation specialists. The results of the NATO technical community deliberation is contained in Ref [3].

[1]   Kayton, M. & Fried, W.R., Avionics Navigation System, John Wiley & Sons, New York 1969.

[2]   Couperthwaite, W.J. & Stokes, r.F., Suggestion concerning the context of the Navigation Sub-System Section of the EFA Weapon system Design and Performance Specification, Royal Aircraft Establishment, Farnborough, Haute, UK, March 1987

[3]   NATO Standardization Agreement 4278, Method of Expressing Navigation Accuracies, 17 August 1993

SECTION II
COORDINATE FRAMES

Muneendra Kumar
Defense Mapping Agency
8613 Lee Highway
Fairfax, VA 22031, USA.

1   INTRODUCTION

Accurate navigation requires an understanding of the three coordinate frames and systems which define positions in space. The first is the inertial frame in which the earth revolves around the sun annually (Section 2). This annual motion is affected by the general precession and astronomic nutation and it takes place in the Earth Centered Inertial (ECI) frame or Conventional Inertial System (CIS).

Second is the reference frame which is defined by the daily rotation of the earth around its polar axis (Section 4). This Instantaneous Terrestrial System (ITS) frame requires knowledge of the sidereal time relationship with the CIS.

Third frame is the geodetic or Earth Centered Earth Fixed (ECEF) coordinate system defining the three-dimensional positions on the earth's surface or in its adjoining space. This frame is also known as Conventional Terrestrial System (CTS) and requires the knowledge of the earth's polar motion, the gravity field, and its size and shape (Section 5).

The ECI (or CIS) transformation to the ECEF (or CTS) frame makes use of the new theories of precession [1,2], astronomic nutation [3,4,5], change to a new J2000.0 Standard Time Epoch [2], the new definition of Universal Time as defined and adopted by the International Astronomical Union (IAU) [6,7], and the latest ECEF frame, viz., World Geodetic System (WGS) 1984 [8].

In the discussion that follows, all the coordinate frames or systems are right-handed and orthogonal, and positive rotation is clockwise when viewed from the origin towards the positive axis. Further, the three rotation matrices R_x, R_y, and R_z represent the positive rotations about the orthogonal axes X, Y, and Z respectively.

2   EARTH CENTERED INERTIAL (ECI) SYSTEM

In the mean ECI system of epoch (or CIS), the Z_1-axis lies along the earth's instantaneous axis of rotation positive towards the mean celestial pole of epoch (P_0), the X_1-axis is positive towards the vernal equinox of epoch (γ_0) and lies in the plane of celestial equator at right angle to the Z_1-axis, and the Y_1-axis is perpendicular to the Z_1 and X_1 axes and contained in the equatorial plane. These axes complete the right-handed orthogonal coordinate frame (Figure 1). The epoch J2000.0 is defined at noon on 1 January 2000 (Section 3).

This inertial system is based on the Fundamental Katlog 5 (FK 5) system [9].

3   TIME AND EPOCHS

The two time systems of interest here, the Sidereal Time (ST) and Universal (UT) or solar time, are both based on earth's diurnal rotation [10]. The Sidereal Time is determined by observing the transits of stars across the observer's meridian. However, since the meridian is involved, one must consider the effect of polar motion on the meridian's position. Mean solar time is associated with a mean or "fictitious" sun that moves along the

celestial equator with a uniform sidereal motion. The motion is approximately equal to the mean rate of the true sun's annual motion along the ecliptic. If the hour angle of the mean sun is referred to the zero meridian ($\lambda = 0^0$) as defined by the International Earth Rotation Service (IERS), the resulting time is known as the Universal Time (UT).

Universal time and sidereal time are both affected by the irregularities of the earth's rotation. These irregularities take the form of polar motion, i.e., the variations in the position of the earth's axis of rotation with respect to earth's crust and the variations in the angular rotation (ω) of the earth about its rotational axis.

Since January 1 1972, the time scale distributed by most of the broadcast time services is Coordinated Universal Time (UTC) [11]. This redefined time differs from the International Atomic Time (TAI) by an integral number of seconds. TAI is the most precise and accurate time scale as determined by the Bureau International de Poids et Mesures (BIPM), Paris, from the analysis of data from the atomic time standards of many countries worldwide.

Universal time, as determined with its irregularities, is known as UT0. The time UT1, which is independent of the position of the observer, is derived by removing the effect of polar motion from the UT0. The leap second or one second time step is introduced periodically to maintain UTC time within 0.9 second of UT1. This correction is normally made at the end of June or December.

3.1 Sidereal Time (ST)

Mean Sidereal Time (MST) is directly obtained from the apparent right ascension of transiting stars [12]. On any given date, a star's computed apparent position is a function not only of its catalogued coordinates and proper motion, but also of the adopted constants and theories of the general precession, astronomic nutation, aberration, etc., [7].

It is to be noted that the change from the old FK4 to the current FK5 star system and adoption of new astronomical constants do produce a complex and subtle change on the values of universal time UT1 which is <u>indirectly</u> derived from the transit time of stars.

3.2 Universal Time (UT)

The UT1, which forms the basis for the worldwide system of time, is related to MST by the equation for Greenwich Mean Sidereal Time (GMST) of 0^h UT1 (Table 1).

The sidereal time equation or matrix [B] in Section 4.4 gives specific values for the angular rotation velocity (ω) of the earth and the ratio of solar time to sidereal time as 0.997269566329084 (or its inverse as 1.0023737909350795) at epoch J2000.0 [7].

UT1 defined as above maintains time continuity in value and rate of ST at the change of epoch to FK5 system, i.e., on 1 January 1984.0.

3.3 Unit of Time (T)

The unit of time (T) in the formulas for astronomical precession and nutation is the Julian Century of 36525 days. The conventional relationship between the Julian Epochs (JE) and Julian Ephemeris Date (JED) is [2]:

$$JE = 2000.0 + \frac{(JED - 2451545.0)}{365.25} \quad (1)$$

It is to be noted here that the Julian Day (JD) begins at noon (12^h).

3.4 Changing Epochs

The relationship between the <u>disconti-nued</u> Besselian time epoch (BE) and the new JE is expressed as [2]:

$$BE(1950.0) = 1900.0 + \frac{(JED - 2415020.3)}{365.24219878} \quad (2)$$

The correspondence between the six different epochs of BEs and JEs and the matrix (M) to transform precessing position and velocity components from BE to JE, as computed through equation (2), are also available in [2].

4 INSTANTANEOUS TERRESTRIAL SYSTEM (ITS)

4.1 General

The transformation of position components from the CIS (or mean ECI system of epoch) to the true ITS (or ECEF) system (Figure 1) requires formation of three matrices, viz., the general precession matrix [D], the astronomic nutation matrix [C], and the sidereal time matrix [B]. The starting epoch J2000.0 is defined at noon time on 1 January 2000 (Section 3.3).

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix}_{ITS} = [B][C][D] \begin{bmatrix} X \\ Y \\ Z \end{bmatrix}_{CIS} \quad (3)$$

For transformation of three velocity components, rate of change of the matrices in equation (3) is to be considered:

$$\begin{bmatrix} \dot{X} \\ \dot{Y} \\ \dot{Z} \end{bmatrix}_{ITS} = [\dot{B}][C][D] \begin{bmatrix} X \\ Y \\ Z \end{bmatrix}_{CIS} + [B][C][D] \begin{bmatrix} \dot{X} \\ \dot{Y} \\ \dot{Z} \end{bmatrix}_{CIS} \quad (4)$$

4.2 General Precession Matrix [D]

The motion of general precession in space can be specified by three angles ζ, z, and Θ (Table 2). The precession matrix [D] transforms the mean CIS coordinate system $X_1(X_1,Y_1,Z_1)$ of epoch J2000.0 (or ECI) of Figure 1 to the orthogonal mean coordinate system $X_2(X_2,Y_2,Z_2)$.

The precession, consisting of three rotation angles ($90-\zeta$, Θ, $90-z$), is given

as the product matrix [D] in Table 3.

After the precession matrix [D] has been applied, the new coordinate frame $X_2(X_2,Y_2,Z_2)$ represents the mean inertial system of date (Figure 2).

As the matrix [D] is orthogonal, the vector equations to transform are:

$$X_2 = [D] X_1 \quad (5)$$

$$X_1 = [D]^T X_2 \quad (6)$$

4.3 Astronomic Nutation Matrix [C]

The astronomic nutation is resolved in two components, nutation in ecliptic longitude ($\Delta\psi$) and nutation in obliquity ($\Delta\epsilon$). It transforms the coordinates from the <u>mean</u> inertial of date to the <u>true</u> inertial system of date (Figure 3).

The $X_2(X_2,Y_2,Z_2)$ axes (Figure 2) are transformed with three nutation angles or rotations ($\epsilon,\Delta\psi,\epsilon$) into the $X_3(X_3,Y_3,Z_3)$ axes (Figure 3). These angles are also depicted symbolically in Figure 4 and the equations and related coefficients for computing them are listed in Tables 4 and 5. The matrix [C], orthogonal like [D], is in Table 6.

The Z_3-axis now points towards the Celestial Ephemeris Pole (CEP) of date and its orthogonal X_3 and Y_3 axes lie in the true equatorial plane of date. In vector form:

$$X_3 = [C] X_2 \quad (7)$$

$$X_2 = [C]^T X_3 \quad (8)$$

4.4 Sidereal Time Matrix [B]

The ST represents a positive rotation about the Z_3-axis by an angle Λ, the longitude of the zero meridian (which used to be the Greenwich Meridian but now as defined by IERS) from the <u>true</u> vernal equinox of date. This sidereal time matrix [B] or its rotational transformation

accounts for both the earth's rotation towards the east and precession of the mean equinox towards the west. Table 1 lists the equation to compute Λ.

The matrices [B] and [B'] are given in Table 7 (Section 4.1).

The matrix [B] transforms the coordinates from the true inertial system of date $X_3(X_3, Y_3, Z_3)$ to the true ECEF system $X_4(X_4, Y_4, Z_4)$ or the Instantaneous Terrestrial System (ITS). Figure 5 shows symbolically only the Z_4-axis of the new transformed orthogonal system. In vector form:

$$X_4 = [B] X_3 \qquad (9)$$

$$X_3 = [B]^T X_4 \qquad (10)$$

The X_4 coordinate frame (ITS) is fixed to the earth and thus rotates with it.

5   CONVENTIONAL TERRESTRIAL SYSTEM (CTS)

5.1 General

The Z_4-axis of the ITS frame is still affected by the polar motion or relative motion of the earth's instantaneous axis of rotation with respect to its crust. When corrected for this polar motion, the true ITS system transforms to a mean geodetic ECEF or CTS system [8].

5.2 Polar Motion

The polar motion affecting the X_4 frame consists of two rotation angles (Figure 6):

- A negative rotation of angle y_p about the X_4-axis.

- A negative rotation of angle x_p about the Y_4-axis.

The angles x_p and y_p are small and this allows the use of approximations for the sine and cosine functions of these angles. The resulting matrix [A] for the polar motion is shown in Table 8. In

vector form:

$$X_5 = [A] X_4 \qquad (11)$$

$$X_4 = [A]^T X_5 \qquad (12)$$

The X_5 coordinate frame (CTS) is also fixed to the earth and rotates with it. The first CTS, defined for global use, was the World Geodetic System (WGS) 1960, which in turn has evolved through WGS 66 and WGS 72 to the latest and more accurate version, viz., WGS 84 [8].

6   WORLD GEODETIC SYSTEM 1984

6.1 General

The WGS 84 coordinate frame (depicted without the suffixes) is defined for the epoch 1984.0 (Figure 7). The CTS implied in the WGS 84 is defined by the Bureau International de l'Heure or BIH which is now known as IERS.

6.2 Realization

The CTS/ECEF frame in the WGS 84 was realized by modifying the Navy Navigation Satellite System (NNSS) or TRANSIT Doppler reference frame in origin and scale, and also rotating it about the Z_5-axis to bring its reference meridian to the BIH defined zero meridian [8].

The WGS 84 coordinate system (Figure 7) represents a standard earth rotating with a constant velocity (ω) around an average astronomical pole (CTP) for the epoch 1984.0.

In addition to defining the basic geodetic reference frame, the WGS 84 defines the earth's geometric figure, models the gravity field and the related geoid, and establishes the transformation constants relating local geodetic datums to a global reference system (Section 8). This provides a global reference for worldwide Mapping, Charting, and Geodetic (MC&G) and navigational products and applications.

6.3 Geometry of the Earth

In MC&G and navigational applications, three different surfaces are involved which are depicted in Figure 8:

- The earth's actual physical topographical surface.

- The geometric or mathematically represented earth's surface, the reference ellipsoid (Section 6.3.1)

- The equipotential surface defined in the earth's gravity field, viz., the geoid (Section 6.3.2; Figure 9).

6.3.1 The WGS 84 Reference Ellipsoid

As the geometric approximation of the earth's shape, WGS 84 reference ellipsoid is based on the Geodetic Reference System (GRS) 1980 [13], as defined by the International Union of Geodesy and Geophysics (IUGG). The GRS 80 definition consists of a geocentric equipotential ellipsoid of revolution defined by the semimajor axis (a), the earth's gravitational constant (GM), the normalized second degree zonal gravitational harmonic coefficient (C_{20}), and the angular rotational velocity (ω) of the earth.

The GRS 80 ellipsoid, tied through the gravity and other data sets [14], was used to best represent the earth's geoid by minimizing the geoidal heights between the ellipsoid and the geoid globally.

This global solution, made possible with the availability of a dense gravity data base worldwide and intercontinental geodetic ties available through satellite observations since the early 60's, should not be confused with the local or regional datums which are in use all over the world. To simplify the mapping complexity of using the different position coordinates defined in the local and regional datums, these datums have been geodetically tied to the WGS 84 through transformation constants (Section 8). However,

due to the non-availability of common control points, which are essential to establish trans- formation constants, there are limited number of local/regional geodetic datums which have been tied to WGS 84 [8].

Table 9 lists the defining or the fundamental constants of the WGS 84 [8]. Some of the important derived geometric and physical constants (Section 6.3.2) and the conversion factors are provided in Table 10. Further, it is important to note that the flattening f of the WGS 84 ellipsoid is a derived constant and is based on the gravitational zonal coefficient C_{20} (or $J_{20}/\sqrt{5}$ of the GRS 80). The eccentricity e is derived or computed from flattening f (Equation 14).

The spatial position in the WGS 84 reference frame is given as rectangular coordinates X,Y,Z (by dropping the CTS suffixes in Figure 5) or as equivalent geodetic coordinates viz., latitude ϕ, longitude λ, and ellipsoidal height h in Figure 10. The relationship between the rectangular and the geodetic coordinates is in Table 11.

a. WGS 84 Ellipsoidal Geometry - Figure 10 shows a meridian section of the WGS 84 ellipsoid with its semimajor axis a and semiminor axis b where:

$$b \quad = \quad a\,(1-f) \tag{13}$$

$$e^2 \quad = \quad (a^2 - b^2)/a^2 \tag{14}$$

$$= \quad 2f - f^2 \tag{15}$$

The geocentric radius R intersects the semimajor axis (or the equatorial plane of the ellipsoid) at angle ψ, the geocentric latitude, where:

$$R^2 \quad = \quad X^2 + Y^2 + Z^2 \tag{16}$$

$$\tan \psi \quad = \quad (1 - e^2) \tan \phi \tag{17}$$

b. Ellipsoidal Radii of Curvature - The two WGS 84 ellipsoidal radii, which

12

are of fundamental importance in dead-reckoning navigation to define the latitudinal and longitudinal velocities $(\dot{\phi}, \dot{\lambda})$, are the radius of curvature R_M in the plane of meridian (or N-S), and radius of curvature R_N in the plane of prime vertical (or E-W):

$$R_M = \frac{a(1 - e^2)}{(1 - e^2 \sin^2 \phi)^{3/2}} \qquad (18)$$

$$R_N = a/(1 - e^2 \sin^2 \phi)^{\frac{1}{2}} \qquad (19)$$

c. Ellipsoidal Height - The distance or height of any point P measured along the ellipsoidal normal or a line perpendicular to the ellipsoid, is known as ellipsoidal height (h) as shown in Figures 8 and 10.

6.3.2 Gravity Model and Geoid

a. Gravity Model - The earth's total gravity potential ($\bar{W}$) is defined as:

$$W = V + \Phi \qquad (20)$$

where V is the gravitational potential or the Earth Gravitational Model (EGM) and Φ is the centrifugal potential [8]. The EGM is modeled as a spherical harmonic expansion, complete through degree (n) and order (m) equal to 180.

The WGS 84 gravity potential (W) then provides the accurate model of the earth's gravity field (g) and its variations on or near its surface. The direction of g is along the plumb line or the astronomical vertical [15].

b. Geoid - The general representation of the geoid (Figure 9) can now be defined in the WGS 84 gravity field as the particular equipotential surface (GEOP) as defined by the equation:

$$W_{84}(X,Y,Z) = W_0 \qquad (21)$$

where the latest value of W_0 is listed in Table 10. The direction of gravity field g is perpendicular to the geoid at every point.

For some practical applications, the geoid, defined as above, is approximated by the mean sea level (msl) at 1 to 2 meter accuracy level along the coast lines and over the ocean areas. Under the continental land masses, the hypothetical extension of msl is sometimes used as an approximation for the geoid.

It may be necessary to clarify here that msl is not an equipotential surface. By definition, msl is the average or mean of hourly sea level surfaces observed at a point (or tidal benchmark) over a period of 18.67 years. However, all msl surfaces so determined in different parts of the world differ in definition and also do not belong to a common zero reference.

In a mathematical sense, the geoid is defined (or realized) as so many meters (m) above (+ N) or below (- N) the reference ellipsoid, the geometric figure of the earth (Figure 8). Figure 9 depicts one version of the WGS 84 geoid as a contour chart with respect to the WGS 84 ellipsoid.

NOTE: More detailed contour charts or denser grid values of the WGS 84 geoid with absolute accuracy range of ± 2 to 6 m are now UNCLASSIFIED and available).

The distance of any point P from the actual physical surface of the earth to the geoid (Figure 8), measured along the direction of gravity or plumb line is called the orthometric height (H) of P. If P is above the geoid, the H is positive and, if below the H is negative. In practice, the height H is also approximated by elevation above or below the msl.

6.4 Relationship Between h, H, and N

The three heights, viz., the geodetic or ellipsoidal height h (Section 6.3.1), the orthometric height or elevation H, and the geoidal height N (Section 6.3.2) can be related to each other (Figure 8) as:

$$ h \approx H + N \tag{22} $$

or its equivalent form:

$$ H \approx h - N \tag{23} $$

Equation (23) is of great practical importance and it illustrates the use of geoid heights (Ns) in the determination of the orthometric heights (Hs) from geodetic heights (hs), which are directly obtainable from satellite point positioning solutions of the GPS satellites.

6.5 Deflection of the Vertical

The angle (Θ) between the ellipsoidal normal and the direction of gravity (or plumb line) is known as the deflection of the vertical (Figure 8).

The angle Θ, with an average of about 10 arc seconds and a maximum of about 1 arc minute, cannot be neglected in the modern day navigation when it extends over time or long distances globally.

7   OTHER COORDINATE FRAMES

7.1 Geocentric Spherical Coordinates

These are defined by the geocentric radius R (Equation 16), the geocentric latitude ψ (Figure 10) which is directly related to the geodetic latitude ϕ through equation 17, and the geodetic longitude λ. From Table 11, the radius R can be related to the geodetic coordinates (ϕ, λ, h).

7.2 Local Geodetic/Topocentric Systems

In Figure 11, the three axes E, N, and U defines a local geodetic or topocentric coordinate system. The system origin is at a point (ϕ, λ, h) with the positive axes E and N pointing geodetic east and north, and the axis U pointing upwards along the ellipsoidal normal or towards positive geodetic height h respectively.

The coordinates (E,N,U) of P in the topocentric local geodetic ENU-system are then obtained by the equation:

$$ \begin{bmatrix} E \\ N \\ U \end{bmatrix} = R_X(90-\phi) \; R_Z(\lambda+90) \begin{bmatrix} X \\ Y \\ Z \end{bmatrix}_{CTS} \tag{24} $$

where R_X and R_Z are rotation matrices about the axes X and Z. The equation (24) can also be written as:

$$ \begin{bmatrix} E \\ N \\ U \end{bmatrix} = \begin{bmatrix} -\sin\lambda & \cos\lambda & 0 \\ -\sin\phi\cos\lambda & -\sin\phi\sin\lambda & \cos\phi \\ \cos\phi\cos\lambda & \cos\phi\sin\lambda & \sin\phi \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \end{bmatrix}_{CTS} \tag{25} $$

7.3 Generalized Local Geodetic System

The axes E' and N' (In Figure 12) are obtained by rotating the ENU-system (Figure 11) by an angle α in azimuth about the U-axis :

$$ \begin{bmatrix} E' \\ N' \\ U \end{bmatrix} = R_U(\alpha) \begin{bmatrix} E \\ N \\ U \end{bmatrix} \tag{26} $$

where rotation R_U is about the axis U.

or

$$ \begin{bmatrix} E' \\ N' \\ U \end{bmatrix} = R_U(\alpha) \; R_X(90-\phi) \; R_Z(\lambda+90) \begin{bmatrix} X \\ Y \\ Z \end{bmatrix}_{CTS} \tag{27} $$

or

$$ \begin{bmatrix} E' \\ N' \\ U \end{bmatrix} = \begin{bmatrix} C_{11} & C_{12} & C_{13} \\ C_{21} & C_{22} & C_{23} \\ C_{31} & C_{32} & C_{33} \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \end{bmatrix} \tag{28} $$

where the nine elements of the equation (28) are known as direction cosines. Then, comparing the corresponding elements of the two matrices in equations (27) and

14

(28), the following relationships are obtained:

$$
\begin{aligned}
C_{11} &= -\cos\alpha\sin\lambda - \sin\alpha\sin\phi\cos\lambda \\
C_{12} &= \cos\alpha\cos\lambda - \sin\alpha\sin\phi\sin\lambda \\
C_{13} &= \cos\phi\sin\alpha \\
C_{21} &= \sin\alpha\sin\lambda - \sin\phi\cos\alpha\cos\lambda \\
C_{22} &= -\sin\alpha\cos\lambda - \cos\alpha\sin\phi\sin\lambda \quad (29) \\
C_{23} &= \cos\phi\cos\alpha \\
C_{31} &= \cos\phi\cos\lambda \\
C_{32} &= \cos\phi\sin\lambda \\
C_{33} &= \sin\phi
\end{aligned}
$$

The generalized local topocentric E'N'U-system with its <u>arbitrary azimuth</u> is conveniently useable worldwide. Then, the geodetic coordinates (ϕ,λ) can be computed as:

$$
\begin{aligned}
\sin\phi &= C_{33} \\
\tan\lambda &= C_{32}/C_{31} \quad (30) \\
\tan\alpha &= C_{13}/C_{23}
\end{aligned}
$$

NOTE: In polar regions, the angles α and $\overline{\lambda}$ are meaningless or indeterminate and thus, the equations in (28) with direction cosines C_{ij} are to be used.

7.4 Horizon Coordinates

In the horizon or <u>local level</u> system, the G-axis is along the direction of the local plumb line (or gravity) and points upwards, the axis E" points towards the local east, and the axis N" is orthogonal to the axes G and E and points the local astronomical (true) north (Figure 13). The axes E" and N" both lie in the local horizon plane.

The relationship between the local geodetic (ENU) and local level (E"N"G) coordinate system is defined as:

$$
\begin{aligned}
\phi" &= \phi + \xi \\
\lambda" &= \lambda + \eta/\cos\phi \quad (31) \\
H &\approx h - N \text{ (meters)}
\end{aligned}
$$

where ξ,η are the two components of the deflection of vertical Θ in the meridian (N-S) and the prime vertical (E-W) planes (Section 1.6.5) and the height relation-ship for H is from equation (23).

7.5 Grid Coordinates

In many mapping and charting applications, the earth's surface is represented on a <u>flat</u> map through a projection with point positions defined in rectangular grid coordinate system. The grid coordinates (x,y) and geodetic positions (ϕ,λ) are interrelated. The unit of measurement is grid meter.

The most common map projections are the Universal Transverse Mercator (UTM), Transverse Mercator (TRM), Mercator (MER), Universal Polar Stereographic (UPS), and Lambert Conformal Conic (LCC) [16,17,18]. The first three projections are based on a cylindrical, the fourth on a plane, and the fifth on a conical surface. All these projections are conformal.

The UTM projection in secant version with 0.9996 as the scale factor along the central meridian [16] is the most popular over non-polar areas (See Section 7.6)

7.6 Military Grid Reference System (MGRS)

The Military Grid Reference System (MGRS) is used to extend the local grid system of coordinates to a large regional application [17,19]. MGRS is only used with the Universal Transverse Mercator (UTM) and Universal Polar Stereographic (UPS) map projections and their associated grids.

For MGRS numbering scheme, the UTM projection extends from latitude 80° S to latitude 84° N, and the UPS projection covers the remaining two polar caps.

8   OTHER GEODETIC DATUMS AND THEIR RELATIONSHIP WITH WGS 84

8.1 General

Currently, numerous local and regional geodetic datums (coordinate systems) are in use worldwide. Some of these datums,

covering only small remote islands and isolated areas, are very poorly defined; some datums demonstrate large errors and excessive distortions.

To map, chart, and navigate in the complexity of local or multiple datums is a challenge. One possible solution will be that all positions required to navigate are first transformed from their local and regional coordinate system into a globally available ECEF coordinate frame, e.g., the WGS 84 (Section 6).

One exception to the above local and regional geodetic datum complexity is the newly defined North American Datum (NAD) 1983.

8.2 North American Datum (NAD) 1983

Replacing its predecessor the NAD 27, the NAD 83 is based on a modern geocentric concept and also uses the GRS 80 parameters [13] for its definition.

a. Coverage Area - NAD 83 essentially extends over the North American continent from Alaska to Panama with extension over Hawaii and other American territories and some islands in the Pacific and Atlantic.

b. Relationship with WGS 84 - There is no difference between NAD 83 and WGS 84 as horizontal reference systems for mapping, charting, and navigation at 1 to 2 meter level.

8.3 Other Local/Regional Datums

The relationships between other local and/or regional geodetic datums with WGS 84 are very complex and also not available in all cases.

8.3.1 Availability of Constants

The transformation constants for the local and regional geodetic datums tied to WGS 84 are available in [8], wherein the number of such datums would be changing as more and more required data sets become

available. The accuracies of these datum transformation constants, except in a few cases, are very poor for geodetic applications. However, the available datum transformation constants in [8] are of adequate accuracies to support mapping, charting, and navigational applications.

All other local and regional datums and global reference frames, not listed in [8], can not be transformed to WGS 84 at this time due to lack of required common control data.

8.3.2 Transformation Methods

In avionics or navigation, the non-geodetic transformation methods, which can be used with necessary accuracy, are :

a. For Non-Polar Areas - In Table 12, the Standard Molodensky equations, which transforms the two geodetic coordinates (ϕ,λ) of any local and regional datum to WGS 84, are given. These equations need the use of the ellipsoidal and transformation constants listed in Appendices A, B, and C of [8]. In the absence of a consistent world height system, the present practical method assumes that all orthometric or msl heights are nonvariant during such datum transformations.

The Molodensky equations do not give satifactory results for $\Delta\lambda$ in the polar areas, i.e., from 89° N or S latitude to the respective pole.

b. For Polar Areas - Near the polar areas between 89° N or S latitude to their respective poles the local and regional datum coordinates should be transformed in three steps:

$$(\phi,\lambda,h)_{LD} \xrightarrow{\text{(Step\# 1)}} (X,Y,Z)_{LD}$$

$$\text{(Step\# 2)}\ (\Delta X, \Delta Y, \Delta Z)$$

$$(\phi,\lambda,h)_{WGS\ 84} \xleftarrow{\text{(Step\# 3)}} (X,Y,Z)_{WGS\ 84}$$
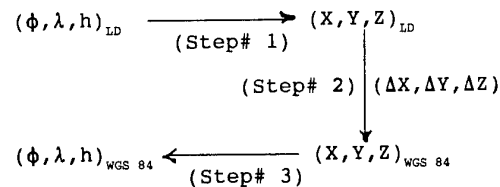
In this procedure, the coordinate conversion equations for steps 1 and 2 are tabulated in Table 11 while the shifts or transformation constants are in [8].

9 UNITS OF LENGTH AND OTHER CONSTANTS

9.1 General

In the present era of aviation over long distances, it is very important to use correct units. Unit measures of meter and foot can be significantly different (Table 10) and thus should be used with caution.

9.2 Units of Length

a. Meter - The length of international meter is linked with the velocity of light and thus has been a constantly changing unit over the years. In some country, a legal meter is in use and its value may differ significantly (about 1 in 75000 in Namibia) from the unit definition listed in Table 10.

b. Foot - The international foot is to used within the Department of Defense [20] while the US survey foot is used by the civil users; the difference between the two units is significant (Table 10).

9.3 Geodetic Constants

The old practice of using any geodetic constants in mapping, charting, and navigaional applications should be discontinued. To ensure correct mapping, charting, and navigation, use of WGS 84 constants is mandatory.

10 REFERENCES

1. Lieske, J., Lederle, T., Fricke, W., and B. Morando; "Expressions for the Precession Quantities Based upon the IAU 1976 System of Astronomical Constants", Astronomy and Astrophysics, Vol. 58, 1977.

2. Lieske, J.; "Precession Matrix Based on IAU 1976 System of Astronomical Constants", Astronomy and Astro and physics, Vol. 73, 1979.

3. Seidelmann, P.K.; "1980 IAU Theory of Nutation -The Final Report of the IAU Working Group on Nutation", Celestial Mechanics, Vol. 27, No. 1, May 1982.

4. Wahr, J.M.; "The Tidal Motion of a Rotating, Elliptical, Elastic,and Oceanless Earth", Ph.D. Dissertation, University of Colorado, Boulder, CO, 80309, 1979.

5. Wahr, J.M.; "The Forced Nutations of an Elliptical, Rotating, Elastic, and Oceanless Earth", Geophysics Journal of the Royal Astronomical Society, Vol. 64, 1981.

6. Aoki, S., Guinot, B., Kaplan, G.H., Kinoshita, H., McCarthy D., and P.K. Seidelmann; "The New Definition of Universal Time", Astronomy and Astrophysics, Vol. 105, 1982.

7. Kaplan, G.H.; "The IAU Resolutions on Astronomical Constants, Time Scales, and the Fundamental Reference Frame" US Naval Observatory Circular, No. 163, Washington, DC, 20392, 1981.

8. Defense Mapping Agency; "Department of Defense World Geodetic System 1984, Its Definition and Relationships with Local Geodetic Systems", DMA TR 8350.2, Second Edition, Fairfax, VA, 22031, 1 September 1991.

9. Fricke, W.; "On the Determination of the Equinox and Equator of the New Fundamental Reference Coordinate System, the FK5", Celestial Mechaics, Vol. 22, No. 2, August 1980.

10. Moritz, H., and I.I. Mueller; "Earth Rotation - Theory and Observation", Ungar Publishing Company, New York, NY, 1987.

11. US Government Printing Office; "The

Astronomical Almanac for Year 1986",
Washington, DC, 1985.

12. Mueller, I.I.; "Spherical and
    Practical Astronomy as Applied to
    Geodesy", Frederick Ungar Publishing
    Co., New York, NY.

13. Moritz, H.; "Geodetic Reference
    System 1980", Bulletin Geodesique,
    Vol. 54, No. 3, Paris, France, 1980.

14. Defense Mapping Agency; "Supplement to
    Department of Defense World Geodetic
    System 1984 DMA Technical Report:
    Part I - Methods, Techniques, and
    Data Used in WGS 84 Development", DMA
    TR 8350.2-A, Fairfax, VA, 22031,
    1987.

15. Heiskanen, W.A., and H. Moritz;
    "Physical Geodesy", W.H. Freeman and
    Co., San Francisco and London, 1967.

16. US Geological Survey; "Map Projections
    - A Working Manual", Professional
    Paper 1395, Reston, VA, 1987.

17. Defense Mapping Agency; "Datums,
    Ellipsoids, Grids, and Grid
    Reference Systems", DMA TM 8358.1,
    Fairfax, VA, 22031, 1991.

18. Defense Mapping Agency; "The
    Universal Grids : Universal
    Transverse Mercator (UTM) and
    Universal Polar Stereographic
    (UPS)", DMA TM 8358.2, Fairfax,
    VA, 22031, 1989.

19. North Atlantic Treaty Organization
    (NATO); "Geodetic Datums, Ellip-
    soids, Grids, Grid References",
    STANAG 2211, Fairfax, VA, 22031,
    1991.

20. Defense Mapping Agency; "Geodetic and
    Geophysical Sign Convention and
    Fundamental Constants", DMA INST
    8000.1, Fairfax, VA, 22031, 1991.

Figure 1. The Celestial Sphere of Epoch.

Figure 2. The Celestial Sphere with Precessing Equinox.

ζ, θ, z = Precession Parameters

$\bar{\gamma}$ to Q = $90^{\circ} - \zeta$

θ = Angle Between Equators

Q to $\bar{\gamma}$ = $-(90^{\circ} + z)$

Figure 3. The Nutating True Celestial Sphere of Epoch.
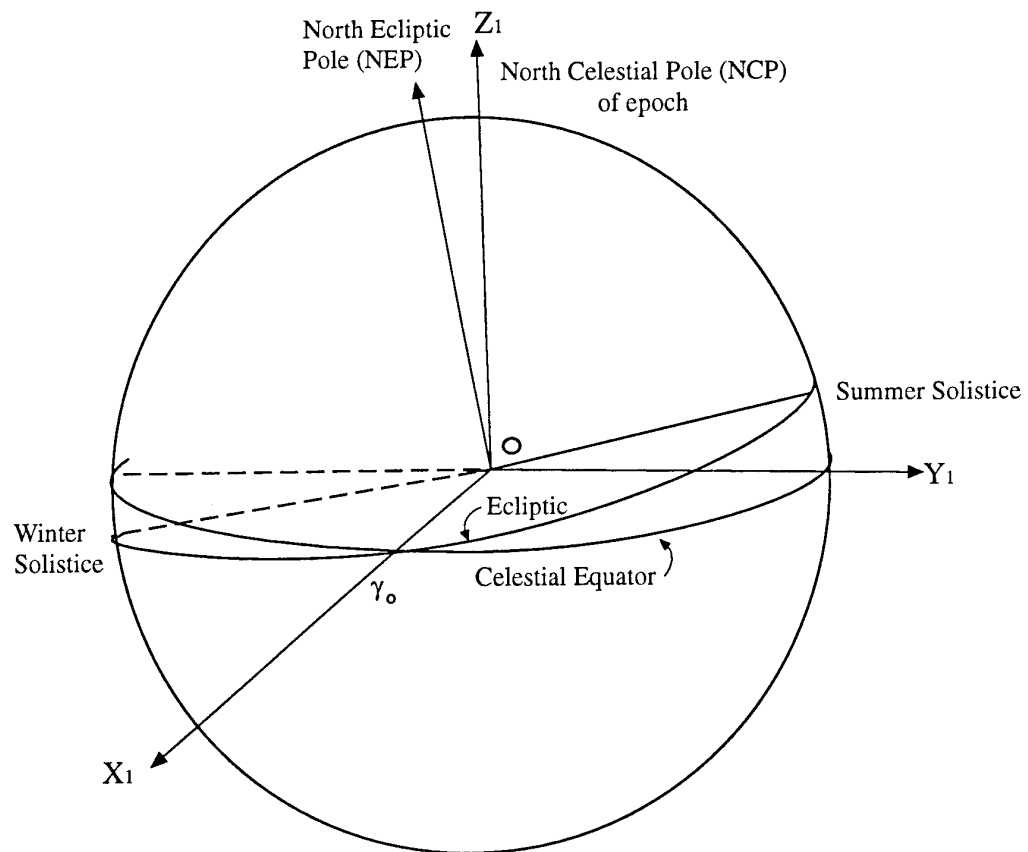
Figure 3. The Nutating True Celestial Sphere of Epoch.

Figure 4. The Astronomic Nutation

|| CEP or Z_3

|| Z_2 Axis

$\Delta\mu$

Ecliptic

Mean Equator of Date

$\bar{\gamma}$

$\bar{\varepsilon}$

Δv

$\Delta\psi$

ε

γ

$\Delta\varepsilon$

True Equator of Date ($\perp$ to CEP)
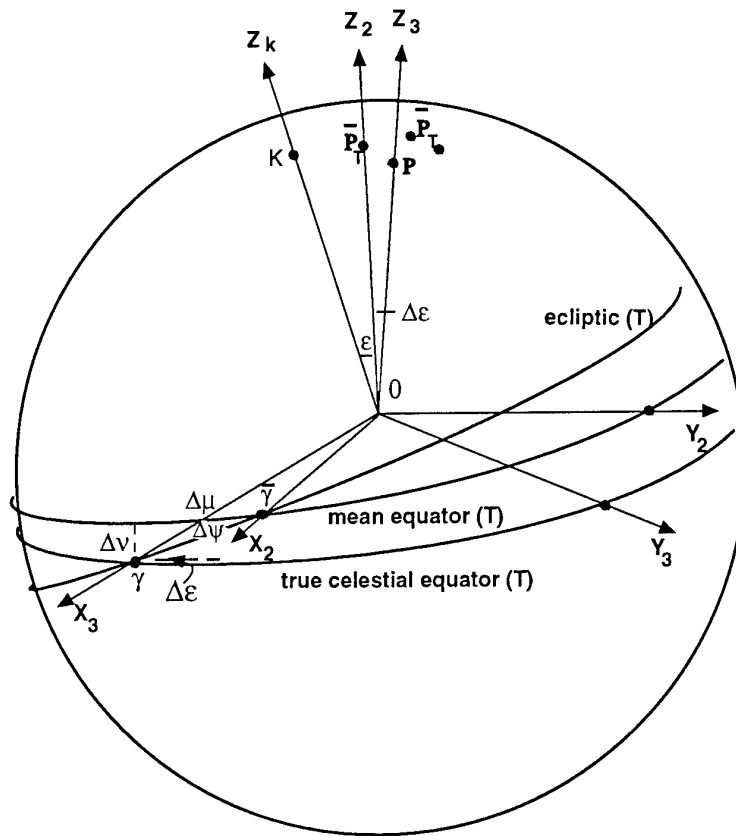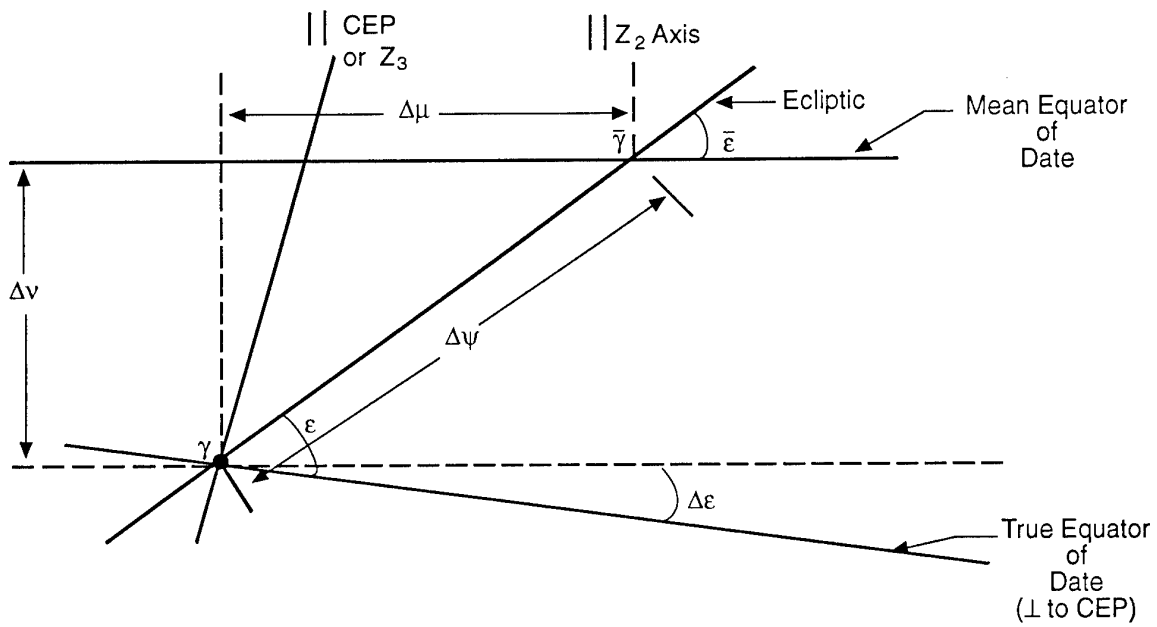
$\Delta\mu$ = Nutation in Right Ascension

Δv = Nutation in Declination

$\tan\Delta\mu = \tan\Delta\psi \ \cos\bar{\varepsilon}$

$\sin\Delta v = \ \sin\Delta\psi \ \sin\bar{\varepsilon}$

Nutation Parameters $\bar{\varepsilon}, \Delta\psi, \varepsilon$:

$\bar{\varepsilon}$ = Mean Obliquity of Ecliptic

$\Delta\psi$ = Nutation in Longitude

$\varepsilon = \bar{\varepsilon} + \Delta\varepsilon$ = True Obliquity of Ecliptic

$\Delta\varepsilon$ = Nutation in Obliquity

Figure 5. The Sidereal Time Transformation.
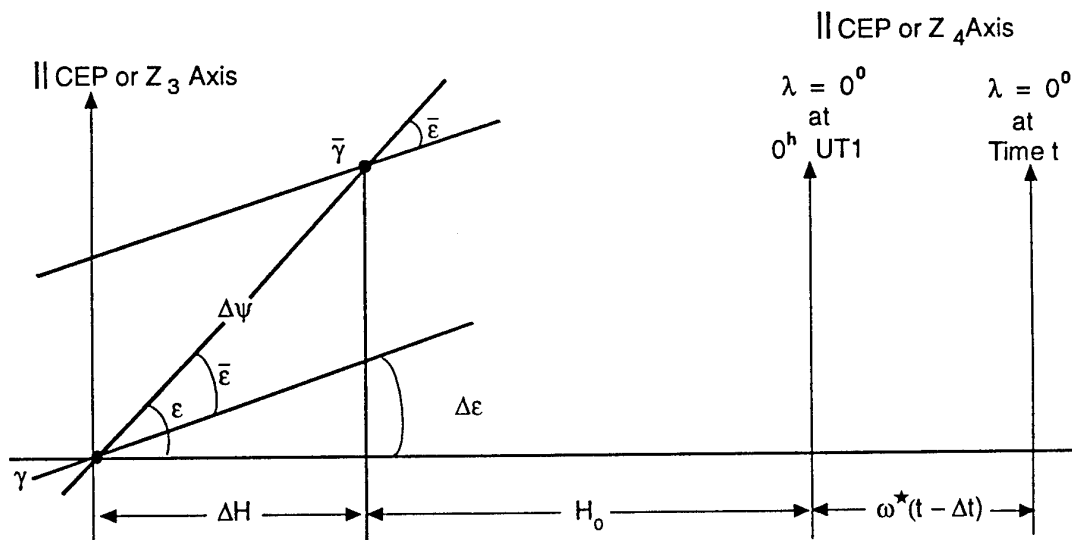
$$\Lambda = H_0 + \Delta H + \omega^\star (t - \Delta t)$$

$$\omega^\star = \omega' + m$$

$\omega^\star$ = Rotation Rate in Precessing Reference Frame
ω' = Earth's Inertial Rotation Rate
m = Rate of Precession in Right Ascension

Figure 6. The Polar Motion Transformation.

CTP = Conventional Terrestrial Pole
CEP = Celestial Ephemeris Pole
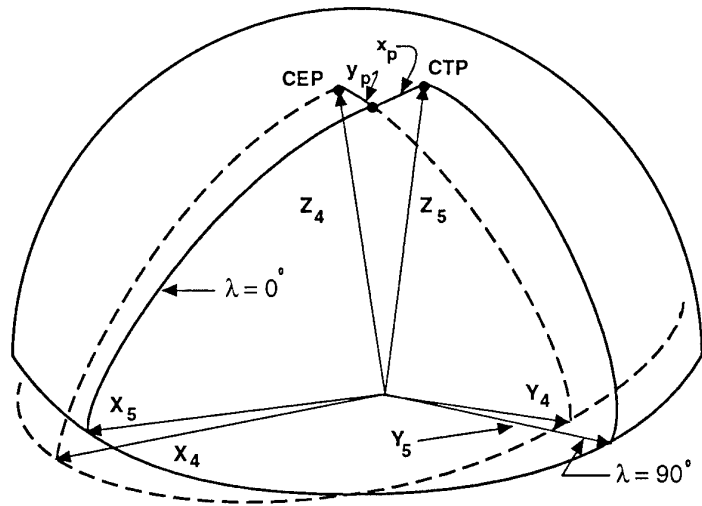
Figure 7. The WGS 84 Coordinate System Definition.

Orgin = Earth's center of mass

z-Axis = The direction of the Conventional Terrestrial Pole (CTP) for
polar motion, as defined by the Bureau International de
l' Heure (BIH) on the basis of the coordinates adopted for
the BIH stations.

x- Axis = Intersection of the WGS 84 Reference Meridian Plane and
the plane of the CTP's Equator, the Reference Meridian being
the Zero Meridian defined by the BIH on the basis of the
coordinates adopted for the BIH stations.

Y- Axis = Completes a right-handed, earth centered, earth fixed (ECEF)
orthogonal coordinate system, measured in the plane of the CTP
Equartor, 90° East of the x-Axis.

BIH-Defined CTP (1984.0)
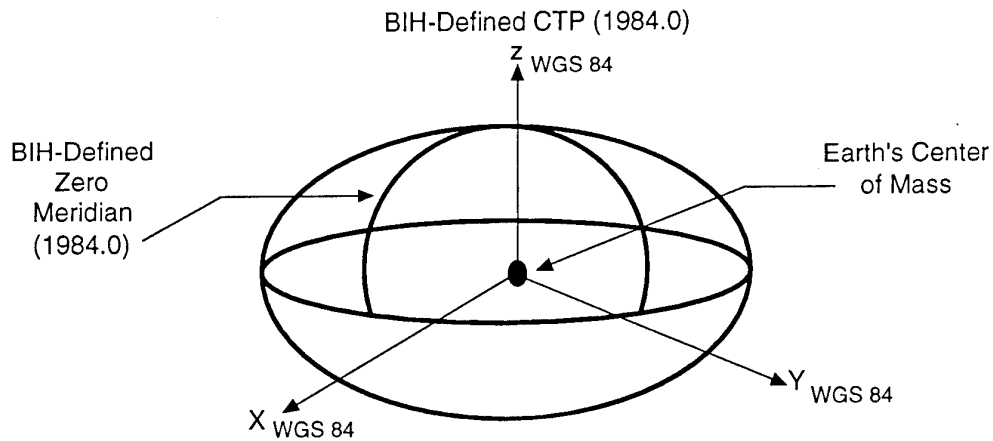
* Analogous to the BIN Defined Conventional Terrestrial System (CTS) , or BTS, 1984.0.

Figure 8. The Relationship between the Earth's Actual, Ellipsoidal, and Geoidal Surfaces.

Earth's Actual Surface

Deflection of the Vertical (Θ)

P (φ, λ, h)

Geoid

h

H

Geoid

Geoid Separation + N

Ellipsoid

Geoid Separation − N
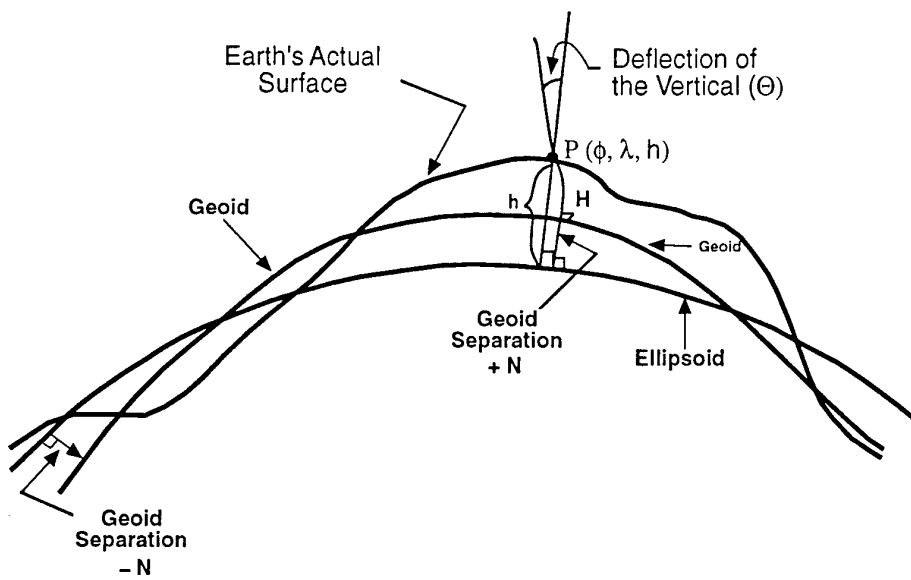
Figure 9. WGS 84 Geoid (n = m = 18 Truncation) Referenced to WGS 84 Ellipsoid (Units = Meters)
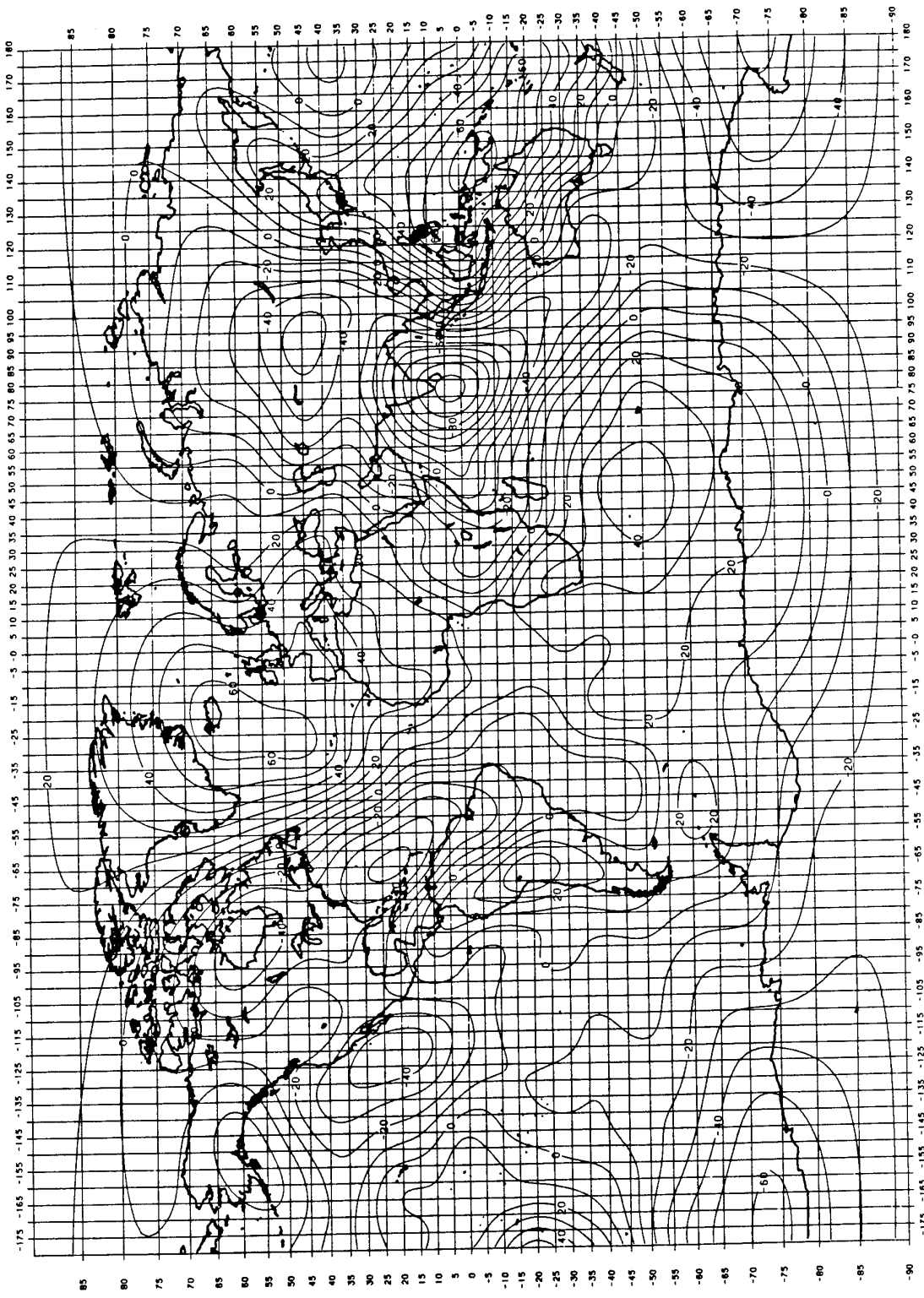
Figure 10. The Geocentric Spherical Coordinates (R,ψ, λ).
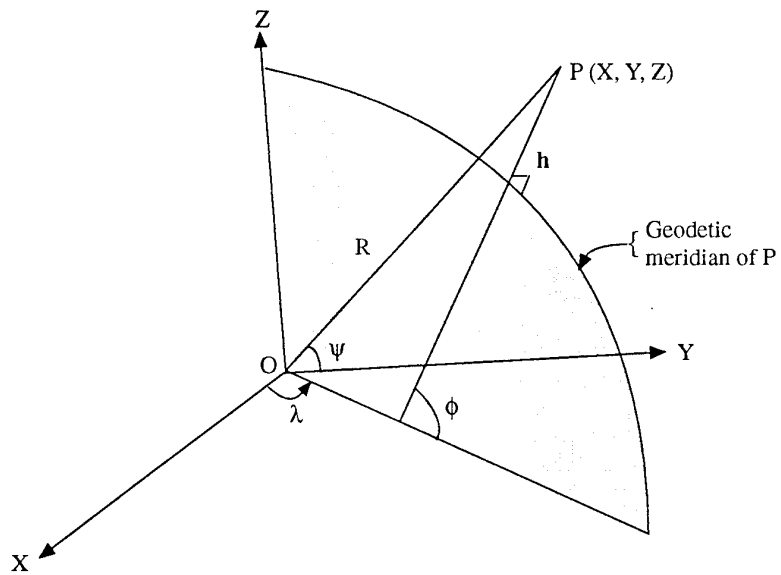
Figure 11. The Local Geodetic Coordinates (**E**, **N**, **U**).
(Topocentric)

Figure 12. The Generalized Local Coordinates (**E'**, **N'**, **U**).

α = alpha
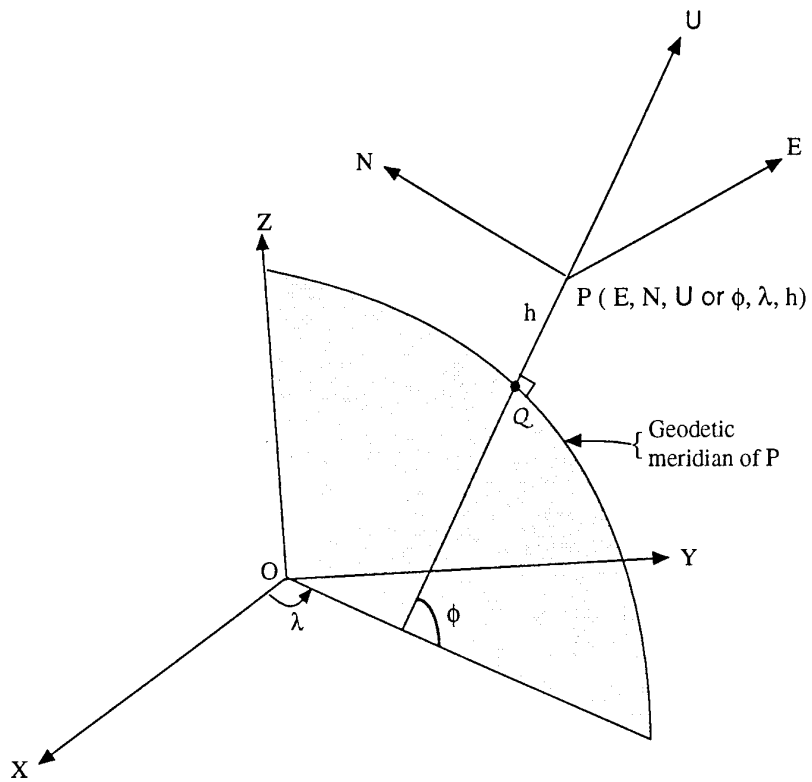
Figure 13. The Horizon Coordinates (E'', N'', H).

ϕ' = Astronomic
      Latitude
λ' = Astronomic
      Longitude
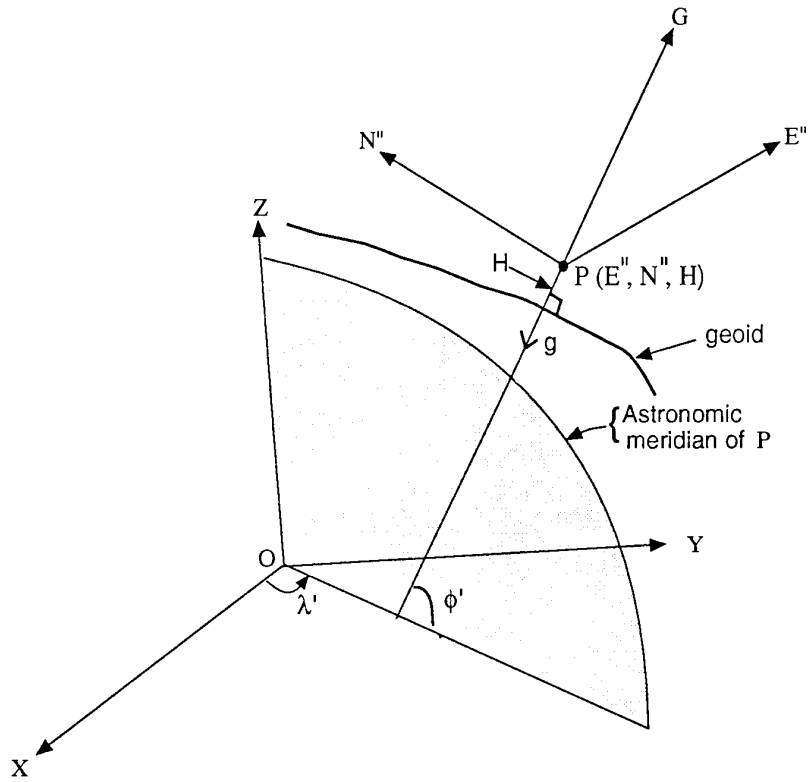
Table 1.
The Sidereal Time Equations **

$\Lambda = H_0 + \Delta H + \omega^\star (t - \Delta t)$

$H_0 = 24110.54841 + 8640184.8640184.812866\ T_u + 0.093104\ T_u^2 - 6.2 \times 10^{-6}\ T_u^3$
    Seconds of Time

   = Greenwich Mean Sidereal Time at 0^h UT1 of JED

Where,

$T_u = d_u / 36525$

$d_u = JED - 2451545$

$d_u \rightarrow 0.5, 1.5\ 2.5, ...$

$\Delta H = \arctan (\cos\varepsilon\ \tan\Delta\psi)$

   = (Apparent Minus Mean) Sidereal Time

ε = True Obliquity

$\Delta\psi$ = Nutation in Longitude

t = Time Within Day (UTC)

Δt = UTC – UT1

$\omega^\star$ = Rotation Rate in Precessing Reference Frame
    = $\omega' + m = 7.2921158553 \times 10^{-5} + 4.3 \times 10^{-15}\ T_u$ (Radians/Second)

ω' = Earth's Inertial Rotation Rate
    = $7.2921151467 \times 10^{-5}$ (Radians/Second)

m = Rate of Precession in Right Ascension
    = $7.086 \times 10^{-12} + 4.3 \times 10^{-15}\ T_u$ (Radians/Second)

** See Table 7 also.

Table 2.
The Precession Equations *

$$\left.\begin{array}{ll}\zeta'' & = 2306.2181\ T + 0.30188\ T^2 + 0.017998\ T^3 \\[2em] z'' & = 2306.2181\ T + 1.09468\ T^2 + 0.018203\ T^3 \\[2em] \theta'' & = 2004.3109\ T - 0.42665\ T^2 - 0.041833\ T^3\end{array}\right]$$

Where ,

$$
\begin{aligned}
T \quad &= [\ JED - 2451545\ ]/36525 \\
&= \text{Julian Centuries from Epoch} \\
&\quad\ \text{J2000.0 (2000 Jan 1.5)}
\end{aligned}
$$

JED = Julian Ephemeris Date

* Note: 1. See Reference [2, 7] .

   2. These three time equations are obtained by setting
      T = 0 in equation (7) of reference [2], _where_
      T has a different meaning than here.

Table 3.
The Precession Transformation Matrix

$$[D] = R_z[-(90+z)]\ R_x(\theta)\ R_z[90-\zeta]$$

$$[D] = \begin{bmatrix} \cos z\ \cos\theta\ \cos\zeta\ -\ \sin z\ \sin\zeta & -\cos z\ \cos\theta\ \sin\zeta\ -\ \sin z\ \cos\zeta & -\cos z\ \sin\theta \\[1.5em] \sin z\ \cos\theta\ \cos\zeta\ +\ \cos z\ \sin\zeta & -\sin z\ \cos\theta\ \sin\zeta\ +\ \cos z\ \cos\zeta & -\sin z\ \sin\theta \\[1.5em] \sin\theta\ \cos\zeta & -\sin\theta\ \sin\zeta & \cos\theta \end{bmatrix}$$

Table 4.
The Astronomic Nutation Equations And Arguments

$\bar{\varepsilon}$ $= \varepsilon_0 - 46.8150 \ T - 0.00059 \ T^2 + 0.001813 \ T^3$ Arc Seconds

= Mean Obliquity of Ecliptic

ε_0 $= 23° \ 26' \ 21.448''$

$= 84381.448''$

ε $= \bar{\varepsilon} + \Delta\varepsilon$

= True Obliquity of Ecliptic

T $= [\ JED - 2451545] / 36525$

$= d/36525$

= Julian Centuries from Epoch J2000.0

$\Delta\psi$ = Nutation in Longitude

$= \sum\limits_{i=1}^{106} \Delta\psi_i$

$= \sum\limits_{i=1}^{106} (A_i + B_i T) \ Sin(a_{1i}\ell + a_{2i}\ell' + a_{3i}F + a_{4i}D + a_{5i}\Omega)$

$\Delta\varepsilon$ = Nutation in Obliquity

$= \sum\limits_{i=1}^{106} \Delta\varepsilon_i$

$= \sum\limits_{i=1}^{106} (C_i + D_i T) \ cos(a_{1i}\ell + a_{2i}\ell' + a_{3i}F + a_{4i}D + a_{5i}\Omega)$

Table 4 (Contd.).
The Astronomic Nutation Equations and Arguments

ℓ = 485866.733 + (1325$^\Gamma$ + 715922.633)T + 31.310T^2 + 0.064T^3 Arc Seconds
   = Mean Anomaly of Moon

ℓ' = 1287099.804 + (99$^\Gamma$ + 1292581.244)T − 0.577T^2 − 0.012T^3 Arc Seconds
   = Mean Anomaly of Sun

F = 335778.877 + (1342$^\Gamma$ + 295263.137)T − 13.257T^2 + 0.011T^3 Arc Seconds
   = (Mean Longitude of Moon) − Ω

D = 1072261.307 + (1236$^\Gamma$ + 1105601.328)T − 6.891T^2 + 0.019T^3 Arc Seconds
   = Mean Elongation of Moon From Sun

Ω = 450160.280 − (5$^\Gamma$ + 482890.539)T + 7.455T^2 + 0.008T^3 Arc Seconds
   = Longitude of Ascending Node of Lunar Mean Orbit on
     Ecliptic Measured From Mean Equinox of Date

T = (JED − 2451545) /36525
1$^\Gamma$ = 1296000"

Table 5.
1980 IAU Theory of Nutation
- Series for Nutations in Longitude ($\Delta\psi$) and Obliquity ($\Delta\varepsilon$) -
(See Table 4)

i	a1	a2	a3	a4	a5	A	B	C	D	i	a1	a2	a3	a4	a5	A	B	C	D
1	0	0	0	0	1	-171996	-174.2	92025	8.9	54	1	0	2	2	2	-8	0.0	3	0.0
2	0	0	0	0	2	2062	0.2	-895	0.5	55	1	0	0	2	0	6	0.0	0	0.0
3	-2	0	2	0	1	46	0.0	-24	0.0	56	2	0	2	-2	2	6	0.0	-3	0.0
4	2	0	-2	0	0	11	0.0	0	0.0	57	0	0	0	2	1	-6	0.0	3	0.0
5	-2	0	2	0	2	-3	0.0	1	0.0	58	0	0	2	2	1	-7	0.0	3	0.0
6	1	-1	0	-1	0	-3	0.0	0	0.0	59	1	0	2	-2	1	6	0.0	-3	0.0
7	0	-2	2	-2	1	-2	0.0	1	0.0	60	0	0	0	-2	1	-5	0.0	3	0.0
8	2	0	-2	0	1	1	0.0	0	0.0	61	1	-1	0	0	0	5	0.0	0	0.0
9	0	0	2	-2	2	-13187	-1.6	5736	-3.1	62	2	0	2	0	1	-5	0.0	3	0.0
10	0	1	0	0	0	1426	-3.4	54	-0.1	63	0	1	0	-2	0	-4	0.0	0	0.0
11	0	1	2	-2	2	-517	1.2	224	-0.6	64	1	0	-2	0	0	4	0.0	0	0.0
12	0	-1	2	-2	2	217	-0.5	-95	0.3	65	0	0	0	1	0	-4	0.0	0	0.0
13	0	0	2	-2	1	129	0.1	-70	0.0	66	1	1	0	0	0	-3	0.0	0	0.0
14	2	0	0	-2	0	48	0.0	1	0.0	67	1	0	2	0	0	3	0.0	0	0.0
15	0	0	2	-2	0	-22	0.0	0	0.0	68	1	-1	2	0	2	-3	0.0	1	0.0
16	0	2	0	0	0	17	-0.1	0	0.0	69	-1	-1	2	2	2	-3	0.0	1	0.0
17	0	1	0	0	1	-15	0.0	9	0.0	70	-2	0	0	0	1	-2	0.0	1	0.0
18	0	2	2	-2	2	-16	0.1	7	0.0	71	3	0	2	0	2	-3	0.0	1	0.0
19	0	-1	0	0	1	-12	0.0	6	0.0	72	0	-1	2	2	2	-3	0.0	1	0.0
20	-2	0	0	2	1	-6	0.0	3	0.0	73	1	1	2	0	2	2	0.0	-1	0.0
21	0	-1	2	-2	1	-5	0.0	3	0.0	74	-1	0	2	-2	1	-2	0.0	1	0.0
22	2	0	0	-2	1	4	0.0	-2	0.0	75	2	0	0	0	1	2	0.0	-1	0.0
23	0	1	2	-2	1	4	0.0	-2	0.0	76	1	0	0	0	2	-2	0.0	1	0.0
24	1	0	0	-1	0	-4	0.0	0	0.0	77	3	0	0	0	0	2	0.0	0	0.0
25	2	1	0	-2	0	1	0.0	0	0.0	78	0	0	2	1	2	2	0.0	-1	0.0
26	0	0	-2	2	1	1	0.0	0	0.0	79	-1	0	0	0	2	1	0.0	-1	0.0
27	0	1	-2	2	0	-1	0.0	0	0.0	80	1	0	0	-4	0	-1	0.0	0	0.0
28	0	1	0	0	2	1	0.0	0	0.0	81	-2	0	2	2	2	1	0.0	-1	0.0
29	-1	0	0	1	1	1	0.0	0	0.0	82	-1	0	2	4	2	-2	0.0	1	0.0
30	0	1	2	-2	0	-1	0.0	0	0.0	83	2	0	0	-4	0	-1	0.0	0	0.0
31	0	0	2	0	2	-2274	-0.2	977	-0.5	84	1	1	2	-2	2	1	0.0	-1	0.0
32	1	0	0	0	0	712	0.1	-7	0.0	85	1	0	2	2	1	-1	0.0	1	0.0
33	0	0	2	0	1	-386	-0.4	200	0.0	86	-2	0	2	4	2	-1	0.0	1	0.0
34	1	0	2	0	2	-301	0.0	129	-0.1	87	-1	0	4	0	2	1	0.0	0	0.0
35	1	0	0	-2	0	-158	0.0	-1	0.0	88	1	-1	0	-2	0	1	0.0	0	0.0
36	-1	0	2	0	2	123	0.0	-53	0.0	89	2	0	2	-2	1	1	0.0	-1	0.0
37	0	0	0	2	0	63	0.0	-2	0.0	90	2	0	2	2	2	-1	0.0	0	0.0
38	1	0	0	0	1	63	0.1	-33	0.0	91	1	0	0	2	1	-1	0.0	0	0.0
39	-1	0	0	0	1	-58	-0.1	32	0.0	92	0	0	4	-2	2	1	0.0	0	0.0
40	-1	0	2	2	2	-59	0.0	26	0.0	93	3	0	2	-2	2	1	0.0	0	0.0
41	1	0	2	0	1	-51	0.0	27	0.0	94	1	0	2	-2	0	-1	0.0	0	0.0
42	0	0	2	2	2	-38	0.0	16	0.0	95	0	1	2	0	1	1	0.0	0	0.0
43	2	0	0	0	0	29	0.0	-1	0.0	96	-1	-1	0	2	1	1	0.0	0	0.0
44	1	0	2	-2	2	29	0.0	-12	0.0	97	0	0	-2	0	1	-1	0.0	0	0.0
45	2	0	2	0	2	-31	0.0	13	0.0	98	0	0	2	-1	2	-1	0.0	0	0.0
46	0	0	2	0	0	26	0.0	-1	0.0	99	0	1	0	2	0	-1	0.0	0	0.0
47	-1	0	2	0	1	21	0.0	-10	0.0	100	1	0	-2	-2	0	-1	0.0	0	0.0
48	-1	0	0	2	1	16	0.0	-8	0.0	101	0	-1	2	0	1	-1	0.0	0	0.0
49	1	0	0	-2	1	-13	0.0	7	0.0	102	1	1	0	-2	1	-1	0.0	0	0.0
50	-1	0	2	2	1	-10	0.0	5	0.0	103	1	0	-2	2	0	-1	0.0	0	0.0
51	1	1	0	-2	0	-7	0.0	0	0.0	104	2	0	0	2	0	1	0.0	0	0.0
52	0	1	2	0	2	7	0.0	-3	0.0	105	0	0	2	4	2	-1	0.0	0	0.0
53	0	-1	2	0	2	-7	0.0	3	0.0	106	0	1	0	1	0	1	0.0	0	0.0

Units: A = C = 0.0001"; B = D 0.0001" Per Julian Century (T from Epoch J2000.0)

Table 6.
The Astronomic Nutation Matrix

$$[\,C\,] \;=\; R_x(-\varepsilon)\; R_z(-\Delta\psi)\, R_x(\bar{\varepsilon})$$

$$[\,C\,] = \begin{bmatrix} \cos\Delta\psi & -\sin\Delta\psi\,\cos\bar{\varepsilon} & -\sin\Delta\psi\,\sin\bar{\varepsilon} \\[2mm] \cos\varepsilon\,\sin\Delta\psi & \cos\varepsilon\,\cos\Delta\psi\,\cos\bar{\varepsilon} + \sin\varepsilon\,\sin\bar{\varepsilon} & \cos\varepsilon\,\cos\Delta\psi\,\cos\bar{\varepsilon} - \sin\varepsilon\,\cos\bar{\varepsilon} \\[2mm] \sin\varepsilon\,\sin\Delta\psi & \sin\varepsilon\,\cos\Delta\psi\,\cos\bar{\varepsilon} - \cos\varepsilon\,\sin\bar{\varepsilon} & \sin\varepsilon\,\cos\Delta\psi\,\sin\bar{\varepsilon} + \cos\varepsilon\,\cos\bar{\varepsilon} \end{bmatrix}$$

Table 7.
The Sidereal Time Transformation Matrix **

$$[B] = R_z(\Lambda)$$

$$[B] = \begin{bmatrix} \cos\Lambda & \sin\Lambda & 0 \\[2mm] -\sin\Lambda & \cos\Lambda & 0 \\[2mm] 0 & 0 & 1 \end{bmatrix}$$

$$[\dot{B}] = \begin{bmatrix} -\omega^{*}\sin\Lambda & \omega^{*}\cos\Lambda & 0 \\[2mm] -\omega^{*}\cos\Lambda & -\omega^{*}\sin\Lambda & 0 \\[2mm] 0 & 0 & 0 \end{bmatrix}$$

** See Table 1 also.

Table 8.
The Polar Motion Transformation Matrix

$$[A] = R_y \, (-x_p) \; R_x(-y_p)$$

[Since x_p and y_p are small angles, it is technically permissible to use approximate transformation matrix]

$$[A] = \begin{bmatrix} 1 & 0 & x_p \\ 0 & 1 & -y_p \\ -x_p & y_p & 1 \end{bmatrix}$$

x_p = Angular displacement of CEP from mean terrestrial pole measured along Zero Meridian (positive south)

y_p = Angular displacement of CEP from mean terrestrial pole measured normal to Zero Meridian (positive west)

Table 9
WGS 84 Ellipsoid
- Four Defining Parameters -

Parameters	Notation	Magnitude	Accuracy (1 σ)
Semimajor Axis	a	6378137 m	$\pm$ 2 m
Normalized Second Degree Zonal Harmonic Coefficient of the Earth Gravitational Model (EGM)	C_{20}	$-484.16685 \times 10^{-6}$	$\pm 1.30 \times 10^{-9}$
Angular Velocity of the Earth	ω	7292115×10^{-11} rad s^{-1}	$\pm 0.1500 \times 10^{-11}$ rad s^{-1}
Earth's Gravitational Constant (Mass of Earth's Atmosphere Included)	GM	3986005×10^{8} m^3 s^{-2}	$\pm 0.6 \times 10^{8}$ m^3 s^{-2}
Parameter Values for Special Applications			
Earth's Gravitational Constant (Mass of Earth's Atmosphere Not Included)	GM'	3986001.5×10^{8} m^3 s^{-2}	$\pm 0.6 \times 10^{8}$ m^3 s^{-2}
Angular Velocity of the Earth (In a Precessing Reference Frame)	ω	$(7292115.8553 \times 10^{-11} + 4.3 \times 10^{-15} T_u)$ rad s^{-1}	$\pm 0.1500 \times 10^{-11}$ rad s^{-1}

T_u = Julian Centuries from Epoch J2000.0

Table 10
- Derived Parameters and Conversion Factors -

Derived Constant	Notation	Value
Geometric		
Flattening	f	1/298.257223563 (.00335281066474)
Semi-Minor Axis	b	6356752.3142 m
First Eccentricity Squared	e^2	0.00669437999013
Physical		
Ellipsoidal Gravity or Geoidal Constant	U_0 or W_0	62636860.8497 $m^2 s^{-2}$
Normal Gravity at the Equator(on Ellipsoid)	γ_e	9.7803267714 m s^{-2}
Normal Gravity at the Poles (on Ellipsoid)	γ_p	9.8321863685 m s^{-2}
Mean Normal Gravity	γ_m	9.7976446561 m s^{-2}
Conversion Factor		
1 Meter (m)		= 3.280833333 US Survey Feet
1 Meter (m)		= 3.28083989501 Int'l Feet
1 Int'l Foot		= 0.3048 m (Exact)
1 US Survey Foot		= 1200/3937 m (Exact)
		= 0.30480060960 m
1 Int'l Nautical Mile		= 1852 m (Exact)
1 Int'l Statute Mile		= 1609.344 m (Exact)
		= 5280 Int'l Feet (Exact)
Velocity of Light (In vacuum)		= 299792458 m s^{-1}

Table 11
Relationship Between Rectangular and Geodetic Coordinates

A.   Geodetic to Rectangular

$$X \;=\; (R_N + h)\, \cos\phi\, \cos\lambda$$

$$Y \;=\; (R_N + h)\, \cos\phi\, \sin\lambda$$

$$Z \;=\; \{R_N\,(1 - e^2) + h\}\, \sin\phi$$

where,

$$a \;=\; \text{semi-major axis}$$

$$R_M \;=\; a(1 - e^2)/(1 - e^2 \sin^2 \phi)^{3/2}$$

$$R_N \;=\; a/(1 - e^2 \sin^2 \phi)^{\frac{1}{2}}$$

B.   Rectangular to Geodetic

$$\lambda \;=\; \tan^{-1} (Y/X)$$

$$\phi_1 \;=\; \tan^{-1} \left[\{Z/(X^2 + Y^2)^{\frac{1}{2}}\} \left(1 - \frac{e^2 R_N}{R_N + h}\right)^{-1}\right]$$

$$h_1 \;=\; \frac{(X^2 + Y^2)^{\frac{1}{2}}}{\cos\phi} - R_N$$

NOTE: To solve for geodetic (ϕ, h), first put $h_1 = 0$ and compute ϕ_1, then use this value of ϕ_1 to compute h_2, and using h_2 iterate this sequence. In most cases, one or two iterations would produce required accuracy (for navigational applications) for ϕ and h.

Table 12
The Standard Molodensky Transformation Equations
- Local Geodetic Datum to WGS 84 -

<u>THE MOLODENSKY EQUATIONS</u>:

The corrections $(\Delta\phi, \Delta\lambda, \Delta h)$ for transforming local datum geodetic coordinates (ϕ, λ, h) to WGS 84 are:

$$\Delta\phi = [- \Delta X \sin\phi \cos\lambda - \Delta Y \sin\phi \sin\lambda + \Delta Z \cos\phi$$
$$+ \Delta a(R_N e^2 \sin\phi \cos\phi)/a + \Delta f \{ R_M (a/b)$$
$$+ R_N (b/a) \} \sin\phi \cos\phi]$$
$$\cdot [(R_M + h) \sin 1"]^{-1}$$

$$\Delta\lambda = [- \Delta X \sin\lambda + \Delta Y \cos\lambda]$$
$$\cdot [(R_N + h) \cos\phi \sin 1"]^{-1}$$

$$\Delta h = [\Delta X \cos\phi \cos\lambda + \Delta Y \cos\phi \sin\lambda$$
$$+ \Delta Z \sin\phi - \Delta a (a/R_N)$$
$$+ \Delta f (b/a) R_N \sin^2 \phi]$$

where,

$\Delta X, \Delta Y, \Delta Z$ are the datum shifts between local geodetic datum and WGS 84 [8].

ϕ, λ, h (to be used on the right hand side of the equations) are the local geodetic datum coordinates.

NOTE: 1. All Δ-quantities are formed by subtracting local geodetic datum quantities <u>from</u> WGS 84 quantities.

2. As ellipsoidalheights (h) are not available for local datums, the Δh correction is not applicable when transforming <u>to</u> WGS 84.

3. For transformation from WGS 84 to local datums, the above Δ-quantities and the coordinates on the right hand side of the equations would require an appropriate change.

SECTION III
NAVIGATION TECHNOLOGY
INTRODUCTION
DR. JOHN NIEMELA

The purpose of this section to describe the predominant navigation sensor technologies employed in modern aerospace navigation systems. A series of papers are presented which discuss in terms of physical principles and mathematical development the capabilities and limitations of each technology. Particular emphasis is given to the error sources and characteristics. Understanding of these errors is of great importance for a solid appreciation of the following section of this document on System Analysis and Synthesis.

Navigation systems can be viewed as three broad categories: self-contained, externally referenced and hybrid. The former category, self-contained, is frequently referred to as dead reckoning. A principal attribute of self-contained navigation systems is that they are not dependent on external radio navigation aids that may be vulnerable in a tactical environment. In dead reckoning system, position is determined by integrating sensed acceleration (in inertial system) or velocities (in Doppler systems). Typical sensor measurements are translational acceleration for inertial systems and translational velocity for Doppler systems. In the process, measurement errors are integrated as well. As a consequence, the accuracy of

computed position decreases with time and distance traveled, respectively for these two navigation technologies. Such systems are quite accurate for the initial time or distance interval of operation but generally require a manual or automatic position updating mechanism in the course of the mission. Hence a means is sought in most aerospace navigation systems to bound these time/distance growing errors. Externally referenced navigation systems offer such means.

Attention has been given to those navigation technologies that sense the aerospace vehicle's dynamic state. Such systems are often termed inertial navigation systems (INS) attitude and heading reference systems (AHRS), heading references and vertical gyros. To capture the underlying technology of these systems and not be encumbered by the wide variety of configurations possible, two complementary navigation technologies are discussed: inertial navigation systems are magnetic heading references.

Externally referenced navigation systems, however, offer highly complementary characteristics for aerospace navigation application, particularly when integrated with self contained systems. Radio navigation aids, the

most widely used externally referenced navigation system rely on the constancy of an electromagnetic wave propagation velocity and directivity of its propagation to determine distance and direction to remote known locations. Examples include non-directional radio beacons (NRB), VHF Omnidirectional Range (VOR), distance measuring equipment (DME), long range navigation (LORAN), DECCA, OMEGA and satellite navigation (GPS and GLONASS). This document provides a discussion of only two of these navigation technologies, OMEGA and satellite navigation. The technical literature is rich with descriptions of the remainder of these external radio navigation aids, most of which are expected to be phased out of operation.

Two other types of externally referenced navigation systems are discussed in some detail due to characteristics that make them suitable for military application: terrain referenced navigation systems and astroinertial navigation systems. Though each is in fact a hybrid navigation system - comprised of both self-contained and externally referenced elements, a unique attribute common to each deserves special attention: Their external reference is not susceptible to being denied by enemy ECM techniques.

The integration of two or more different types of navigation systems, typically a combination of self-contained and externally referenced navigation systems, results in a hybrid navigation system. The sensors which comprise such hybrid navigation systems typically have complementary error spectrums and/or data availability characteristics which yield a more accurate and reliable navigation system. Due to their importance for military aerospace navigation applications, a complete section of the document has been dedicated to the system analysis, design and synthesis of such systems.

Inertial Navigation

B. Stieler
DLR, Institute for Flight Guidance
Postfach 3267
W-3300 Braunschweig
GERMANY

1. Introduction

The development of this technology is truly exciting. It took place primarily in Germany, the United States and the former Soviet Union [Gi 71, So 76, Ma 90].

The gyrocompass indicating true north on a moving base as on ships can be regarded as the beginning of inertial navigation. The first seaworthy instrument was built by Anschütz with contributions made by Max Schuler and Albert Einstein and installed on the fast steamer "Imperator" in 1913 [Sc 62].

At the end of World War I the allies had in the Treaty of Versailles imposed restrictions to Germany for the maximum size of ships to be built. The so-called "vest-pocket battleships" were much more exposed to heavy seas than their bigger counterparts. These restrictions promoted in this country gun stabilization and inertial technology in general, which culminated at the end of World War II in a functioning air-supported gyrocompass with electronic Schuler tuning for the "One-Man Submarines", in the V2 guidance system and a true concept for an inertial navigation system (INS) [Gi 71]. After the war the development of this technology was taken over by the superpowers, the United States and the former Soviet Union [Ma 90].

The functional diagrams for a platform INS and for a so-called "strapdown INS (SDS)" are shown in Figs. 1.1 and 1.2. "Strapdown" comes from the fact that the sensors of the SDS are strapped down to the vehicle. Both figures show us that the inertial navigation system (INS) provides all information about the kinematics of a vehicle, namely attitude and heading, ground speed and position, and also angular rate and acceleration independent of any sources of reference from outside. No question about its role for military aviation, marine navigation and for missiles! A high inertial technology is nowadays a trademark for military independence of many countries. Inertial navigation is also widespread in use in civil aviation and in space flight. Any large civil aircraft is equipped with two or three inertial navigation reference systems. The advent of the "Global Navigation Satellite System (GNSS)" will in principle not change this situation in the time to come, especially under the consideration that a solution for its undisturbed availability in times of strained political situations for the countries running the system is not yet in sight.

Since the whole chain of information on the vehicle motion is contained in the INS output, these systems are also used as measuring instruments in flight tests and in other cases where accurate angle, velocity and position measurements have to be carried out in a difficult dynamic environment. Inertial sensors and systems are used above, on and under the ground. Examples are land surveying, borehole measurements, pipeline inspection in the Arctic and inspection of the tracks for highspeed trains and automobile tests, to name only a few.

Recent achievements in the high accuracy flight tests of "Microwave Landing Systems (MLS)" [Hu 74] and the high accuracy landing guidance of aircraft [Ja 90] are based on outputs from inertial systems which, together with outputs from other sensors, were combined into optimal information on attitude, velocity and position.

This draws attention to one aspect which, among others, should be one of the conclusions from this contribution. Like all technical instruments an INS has its specific error behaviour which is, however, of different nature compared to those of satellite and radio aids (e.g. GNSS, VOR/DME, TACAN, radar, laser tracker, Doppler radar, ILS). An optimum of information can be obtained by using information from different sources. In this respect mathematicians have provided technicians with a fairly efficient tool in the form of the Kalman filter algorithms [Ka 60]. The above mentioned achievements in precision navigation and flight testing have been obtained through the contribution of engineers in the form of hardware and of mathematicians in the form of software.

The advantage of blending INS data with those of different sources lies in the fact that the levels of information on angular rate, acceleration, attitude, heading, ground speed and position are connected in the system in a mathematically strict sense as to be seen from Figs. 1.1 and 1.2. Aiding the system by means of external references on the velocity level (e.g. Doppler radar) or position level (e.g. GNSS, VOR/DME), also increases the accuracy for the other levels. In Fig. 1.3 it is shown that the Kalman filter estimates the systematic INS errors down to the levels of gyro drift and accelerometer bias, i.e. based on the external measurements, an in-flight INS calibration is carried out by this algorithm. After a sufficient calibration time periods of missing external measurements are bridged with only slowly growing system errors. This is the reason why systems for high-accuracy simultaneous measurements of velocity and position or attitude and velocity, or even the three states together, are always centered around an INS. It is also advantageous to use an INS in those cases when the high-accuracy measurement of position is a main purpose of the test as indicated above with the example of MLS testing.

This chapter is arranged in the following train of thoughts. The directional reference in an INS is explained with the mechanical gyro and stabilized platform as examples. It is symbolized by the weathercock in the functional diagram for a platform INS in Fig. 1.1. The characteristics of the gyrostabilized platform as directional reference serve to visualize the characteristics of the "analytic platform" in the navigational computer of modern strapdown systems. Again this is symbolized by a weathercock in Fig. 1.2, the functional diagram for a strapdown system. Directional references and their error characteristics are discussed in Section 2.

Accelerometers as sensors for measuring the translational motion are fairly simple instruments in principle, but the formula for their output signal on the rotating earth is lengthy and it is the basis for programming the navigational computer of an INS. These aspects are treated in Section 3.

The integration of the accelerometer signal to ground speed and position and the control or computation of the directional reference is subject of the navigational computer. The interlinking of all signals within an INS causes error characteristics more benign than we would expect from our school learning. This is subject of Section 4. Also the INS for worldwide navigation, common features and differences of all mechanizations are discussed. The goal is to derive the error model appropriate for integrating the INS with the other sensors and systems discussed in this book.

The main part of this chapter contains only general outline which are essential from the system point of view. Special features as coordinate systems for inertial navigation supplementing the chapter "Navigation Coordinate Systems", digital data processing of inertial signals especially in strapdown systems, and optical gyros are treated in the appendices which are named correspondingly with C, D and O.

An excellent textbook about the material covered in this chapter was written by the author's teacher and friend at MIT, the late Kenneth R. Britting, Sc. D [Br 71].
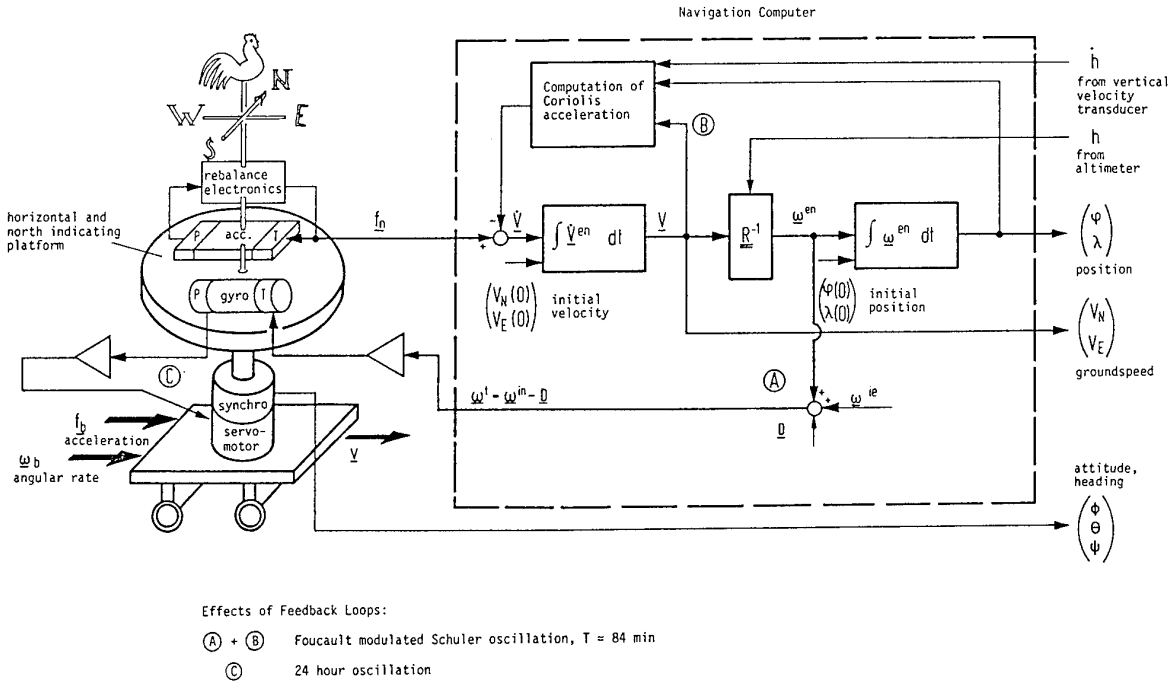
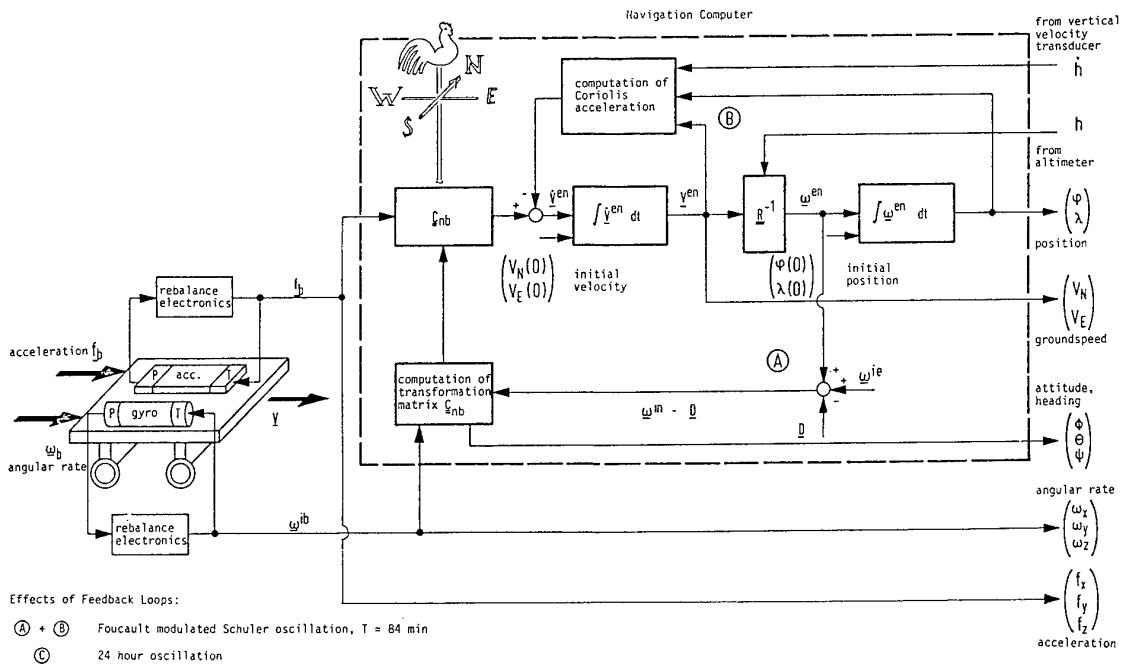Fig. 1.1 Functional Diagram of an Inertial Navigation System in Platform Mechanization
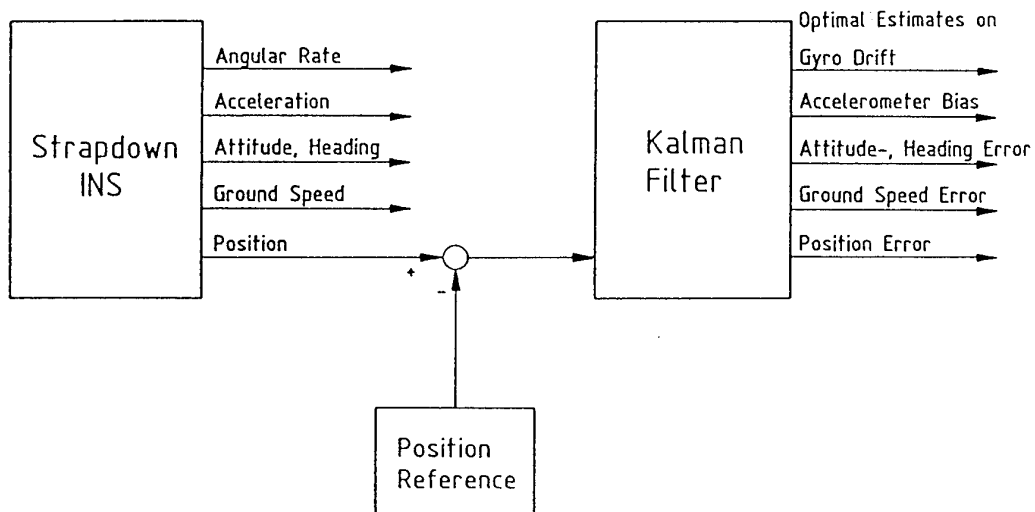

Fig. 7.4.8 Functional Diagram of an Inertial Navigation System (Strapdown System)

Fig. 1.2 Functional Diagram of an Inertial Navigation System in Strapdown Mechanization

Fig. 1.3 Aiding an Inertial Navigation System on the Position Level with the Use of a Kalman Filter

Literature Chapter 1

[Br 71] Britting, K. R.: "Inertial Navigation Systems Analysis."
Wiley-Interscience, New York, 1971

[Gi 71] Gievers, J.: "Erinnerungen an Kreiselgeräte."
Jahrbuch 1971 der DGLR. page 263 - 291

[Hu 74] Hurrass, K., Stieler, B.: "Zum Einsatz des hybriden Flugmeßsystems bei der Vermessung eines Mikrowellenlandesystems."
Proceedings of the DGON Symposium Gyrotechnology, Braunschweig 1976

[Ja 90] Jakob, Th.: "Integrated System for Automatic Landing Using Differential GPS and Inertial Measurement Unit."
Proceedings of Second International Symposium on Precise Positioning with the Global Positioning System, Ottawa, Canada 1990

[Ka 60] Kalman, R.E.: "A New Approach to Linear Filtering and Prediction Problems".
Journal of Basic Engineering (ASME), Vol. 82 D, pp. 35-45, 1960

[Ma 90] MacKenzie, D.: "Inventing Accuracy - A Historical Sociology of Nuclear Missile Guidance."
MIT Press, Cambridge, Mass., USA, 1990

[Sc 62] Schuler, M.: "Die geschichtliche Entwicklung des Kreiselkompasses in Deutschland." VDI-Zeitschrift, Vol. 104, No. 11, pp. 469 - 508, 1962

[So 76] Sorg, H.: "From Serson to Draper - Two Centuries of Gyroscopic Development."
Proceedings of the International Navigational Congress 1976, Boston, Mass., USA

2. Directional References

2.1 The Mechanical Gyro as Directional Reference

Mechanical gyros and gyro-stabilized platforms are the classical directional references in moving vehicles. Once their functioning and their error behaviour is well understood, it is not a major step to see the common features and differences to strapdown systems whose directional reference is called "analytic platform".

We will look at the mechanical gyro and platform in this chapter from the system point of view only, i.e. exclude major design specifics, high frequency characteristics and other details. The reader interested in these details is referred to the literature [Wr 69, De 70, St 82].

The rotating wheel is the sensing element of the mechanical gyro. To describe its characteristics mathematically, we define the angular momentum vector in the gyro element-fixed coordinate system shown in Fig. 2.1:

$$(2.1) \qquad \mathbf{H} = \begin{bmatrix} H_x \\ H_y \\ H_z \end{bmatrix} = (H_x \quad H_y \quad H_z)^T.$$

This vector is dominated by H_z, the rotor's angular momentum about the spin axis, whose magnitude is estimated from:

$$(2.2) \qquad H = I \cdot \omega$$

with I = rotor's moment of inertia about the spin axis, which for a ring and the measures of Fig. 2.2 is:

$$(2.3) \qquad I = (1/8) \cdot (m_a \, d_a^2 - m_i \, d_i^2) = 416 \ g \ cm^2.$$

With the rotor's angular rate about the spin axis:

$$(2.4) \qquad \omega = 2 \pi f \text{ and } f \approx 400 \text{ Hz, for platform gyros}$$

we obtain for the rotor's angular momentum magnitude:

$$(2.5) \qquad \begin{aligned} H &= 1.046 \cdot 10^6 \ g \ cm^2 \ s^{-1} \\ &= 1.046 \cdot 10^6 \ dyn \ cm \ s \\ &= 1.046 \ 10^{-1} \ N \ m \ s. \end{aligned}$$

Gyros mounted on a platform of an inertial navigation system (INS) have an angular momentum of this magnitude, since they stabilize the platform and rotate with respect to inertial space very slowly about their sensitive axes, i.e. the axes perpendicular to the spin axis. Gyros in a strapdown INS are hardmounted to the aircraft and and have to be torqued about these axes corresponding to the aircraft maneuver. Since the electric power for torquing is proportional to H^2, their rotor angular momentum is lower by 1 order of magnitude.

Because of the high magnitude of the rotor's angular rate and momentum about its spin axis, we can neglect the maneuver dependent components and approximate the total angular momentum vector by:

$$(2.6) \qquad \mathbf{H} \approx (0 \quad 0 \quad H)^T.$$

According to the Theorem of Coriolis [Wr 69] Newton's Law reads:

$$(2.7) \qquad \mathbf{M} = \left. \frac{d \, \mathbf{H}}{d \, t} \right|_i$$

$$= \left. \frac{d \, \mathbf{H}}{d \, t} \right|_g + \omega^{ig} \times \mathbf{H}$$

main contributor
"technical gyro
equation"

causing interior
gyro dynamics
"nutation" for instance.

The z-component of this equation is of no interest, since it expresses: motor torque $\cong$ friction torque.

In the low frequency range the gyro characteristics are described solely by the "Technical Gyro Equation":

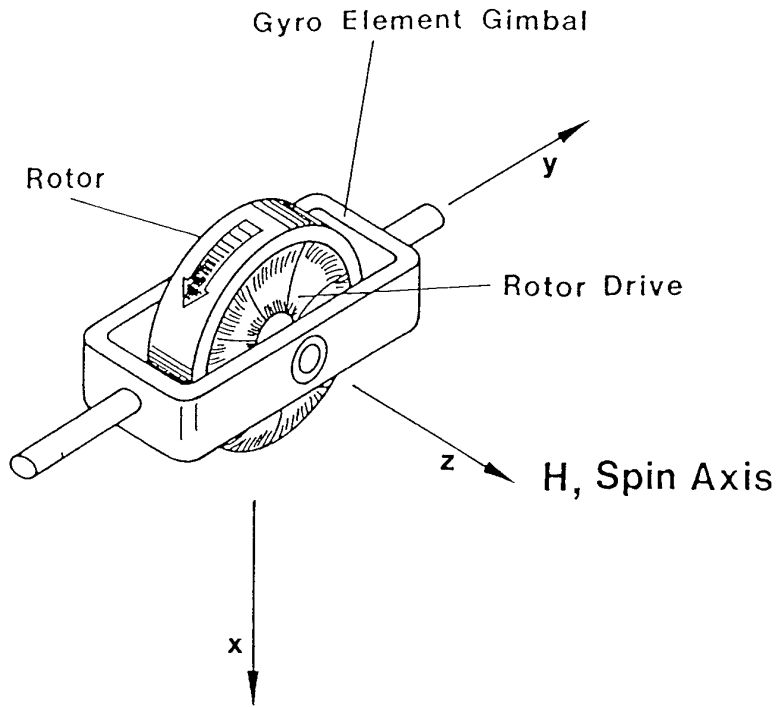$$(2.8a) \qquad \mathbf{M} \cong \omega \times \mathbf{H} ,$$

and in components:

Figure 2.1 Mechanical Gyro Notation

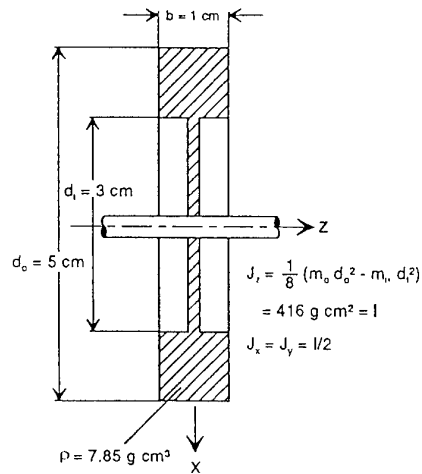

Figure 2.2 Gyro Rotor Cross Section

(2.8b)
$$\begin{bmatrix} M_x \\ M_y \end{bmatrix} \approx H \begin{bmatrix} \omega_y \\ -\omega_x \end{bmatrix}^{ig}$$

The Technical Gyro Equation can be interpreted by the "Rule of Spinvector Alignment" illustrated in Fig. 2.3.

"Rule of Spinvector Alignment"

The gyro reacts to a disturbance vector M or ω in such a way that its angular momentum vector H will align itself with the disturbance vector M or ω following the shortest path

M applied:
Direction of response angular rate:
H vector is "hunting" the input
torque vector M caused by mass m

ω applied:
Direction of response torque:
acts so as to take the H vector by the shortest
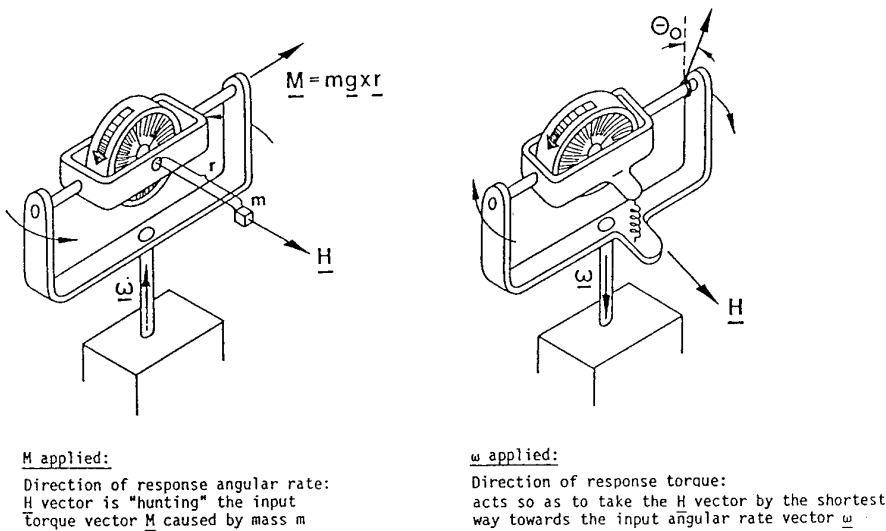way towards the input angular rate vector ω

Figure 2.3 Illustration of the "Rule of Spin Vector Alignment"

The setup in Fig. 2.3 right can in principle be dirctly used as a directional reference. This is the case for lower quality cockpit instruments and for vertical gyros (VG) and directional gyros (DG)) in air and ground vehicles and missiles. We will come back to these instruments.

For the derivation of the block diagram of Fig. 2.4a for such an instrument we introduce in Eq. 2.8b:

(2.9) $\dot{\Theta}^{ig} = \omega^{ig}$,

with Θ^{ig} = the spin vector's angle with respect to the inertial reference. For small angles Θ^{ig} can be regarded as a vector.

The torques on the gyro rotor comprise:
- the command torque M^t applied through the gyro torquer mounted at the gimbal axes and
- the disturbance torque M^d, discussed below.
Instead of the torques we introduce the corresponding angular rates due to the "Technical Gyro Equation":

(2.10a,b,c) $M^t = \omega^t \times H$, $M^d = d \times H$, $M = M^t + M^d$

with ω^t = command rate vector and d = gyro drift vector. Bearing this in mind we may derive from Eqs. 2.8 and 2.9:

(2.11) $\dot{\Theta}^{ig} = \omega^t + d$.

The corresponding block diagram of the gyro as directional reference with respect to inertial space is shown in Fig. 2.4b.

The inertial-fixed reference direction is now replaced by the reference direction in question (index r). For navigational purposes it is the earth-fixed reference direction (index r = n) with its components x = north (N), y = east (E) and z = down (D) (s. Appendix C, Fig. C1). For missile control the reference direction often is the line of sight (LOS, index r = l). For the corresponding block diagram we split up ω^{ig}:

(2.12) $\omega^{ig} = \omega^{ie} + \omega^{er} + \omega^{rg}$

with   earth rate
            transport rate
                misalignment rate = $\dot{\varepsilon}$ (s. Appendix C, Eqs. C 7 - C 9 and C 19).

We thus obtain for the rate of change of the misalignment of the gyro spin axis with respect to the reference direction:

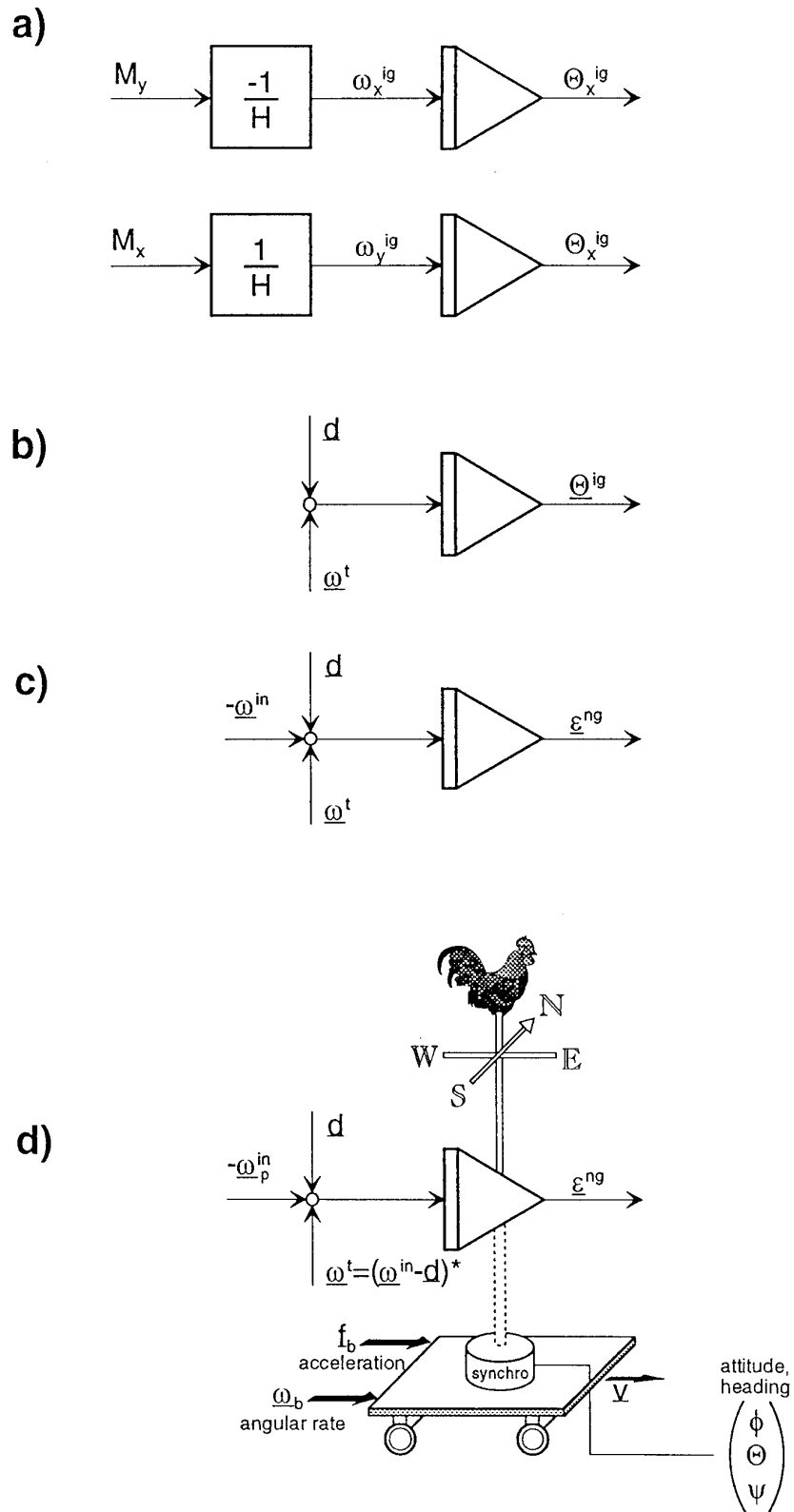(2.13) $\dot{\varepsilon} = \omega^t + d - \omega^{ir}$.

Figure 2.4 Block Diagrams of the Mechanical Gyro as Directional Reference with Respect to Inertial Space (a,b), with respect to North, East and Down (c) and Functional Diagram of the Directional Reference on a Moving Base (d).

The block diagram illustrating this equation, i.e. the use of a gyro as directional reference on the earth is shown in Fig. 2.4c. We have combined $\omega^{ir} = \omega^{in} + \omega^{nr}$ with $\omega^{ie} = $ earth rate and $\omega^{er} = $ reference direction rate with respect to the earth.. The small angle misalignment vector $\varepsilon = (\varepsilon_N \quad \varepsilon_E \quad \varepsilon_D)^T$ has its components with respect to north, east and down (s. Appendix C, Fig. C 4).

The gyroscopic directional reference with respect to the navigational frame on a moving base is finally illustrated in Fig. 2.4 d. The integrator with its 3 input signal vectors is the representation of the gyro - or the gyro-stabilized platform as we will in the next section. For proper functioning ($\varepsilon^{ng} \rightarrow 0$) this input has to be kept close to zero as indicated in this figure. It also shows as dotted lines the mechanical linking of the gyro to the base via the gimbal axes. Synchros mounted to the gimbal axes allow to measure attitude and heading of the base (aircraft, missile) with respect to this reference direction.

With the torquing signal ω^t in Fig. 2.4c being zero, $d + \omega^{in}$ cause the gyro spin vector to deviate from its initial direction. This is also true for $\omega^t = d = 0$ and for a geostationary gyro with its spin vector initially slightly misaligned with respect to the earth axis and as shown in Fig.2.5. The components of earth rate ω^{ie} in the gyro coordinate frame (index g) cause this sensor to carry out the motion expressed through Eq. 2.13: $\varepsilon = -\omega_g^{ie}$ and illustrated in this figure. We will come back to the analysis of this motion further below when we discuss the gyro-stabilized platform.
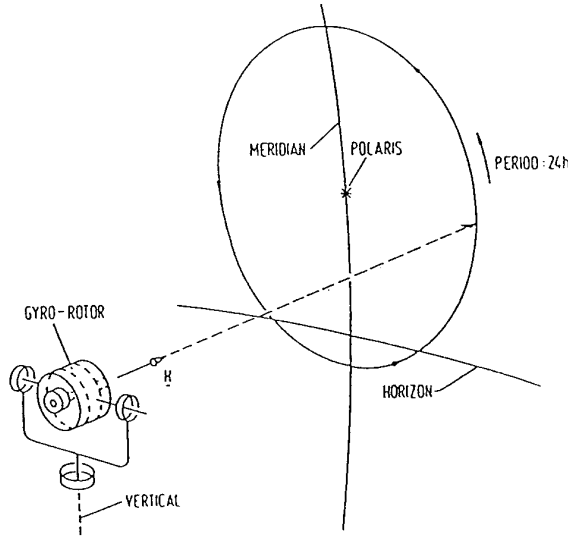

Figure 2.5 Motion of the Geostationary Gyro on the Rotating Earth

For inexpensive directional references of lower quality as the vertical gyro (VG) and the directional gyro (DG) the excursion of the misalignment angle with respect to the reference plumb line or to the reference magnetic north is measured by additional sensors and is driven close to zero by means of control loops to the gyro torquer as indicated in Fig. 2.6. Fig. 2.4c may be used for the derivation of the block diagrams in Fig. 2.7 where instead of the currents fed into the torquers the corresponding command rate vectors ω^t are shown. In the case of the vertical gyro (VG) bubble levels sense the horizontal misalignment components ε_h. In the case of the directional gyro (DG) a magnetic compass (flux valve, flux gate) measures ε_D. Though the long term directional reference is provided by the additional sensors, they cannot be used without a gyro. The reference direction sensors are too much affected by the aircraft maneuvers. The gyro's task is to avarage the high frequency components of the sensor's measurements. The following torquing rule applies for these instruments:

(2.14a) $\omega^t \sim \varepsilon.$

In the case of an INS the misalignment angles are kept close to zero by the following torquing rule:
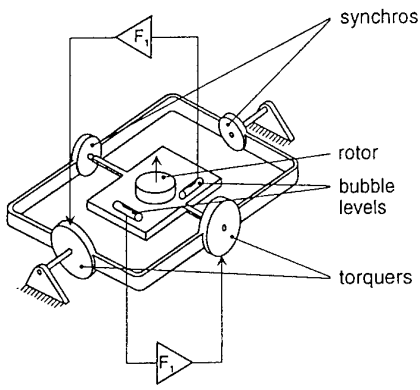
(2.14b) $\omega^t = \omega^{in}$,

i.e. by the computed sum of earth rate and transport rate (s. Section 4.2).

The so-called "Attitude and Heading Reference System (AHRS)" is settled in its accuracy as directional reference between the VG/DG and the INS. They are mechanized like an INS but equipped with inertial sensors of lower quality. For high quality attitude and heading indication they are aided by external sensors; in their most simple realization they are slewed to magnetic north by means of a flux valve, for instance, as it is done in the DG. We will come back to the AHRS in Section 6. Table 2.1 shows the achievable accuracy for attitude and heading, i.e. the uncertainty range of the misalignment angles $\varepsilon_{N,E,D}$ for all three applications.

		VG	DG	AHRS	INS
horizontal misalignment (plumb-line error)	$\varepsilon_N, \varepsilon_E$ [deg]	.4[1]	–	.25	.05
vertical misalignment (heading error)	ε_D [deg]	–	$3 + \delta\alpha$	$1 + \delta\alpha$	.4

[1] unaccelerated flight; $\delta\alpha$ = deviation + variation; deviation = difference between magnetic north and true north variation = difference between indicated north and magnetic north (s. Chapter "Magnetic Heading References").

Table 2.1 Misalignment Errors of Three Types of Direction References

synchros

rotor

bubble levels

torquers

F_1

F_1

F_1

F_2

heading synchro

remote magnetic compass
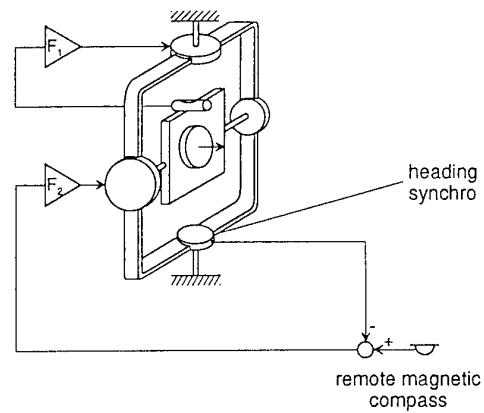
Principles of

Vertical Gyro

Directional Gyro

Figure 2.6 The Two-Degree-of-Freedom (TDF) Gyro for Use as Directional Instruments

vertical gyro

$\dot{V}_h$

f

K

f

K_2

K_1

accelero-meter

$-g$

$-\omega_h^{in}$

D

ε_h

integrator for
compensation
of quasistationary
disturbances

gyro

directional gyro

K_2

K_1

K_3

$-\omega_h^{in}$

D

ε_h

remote magnetic compass
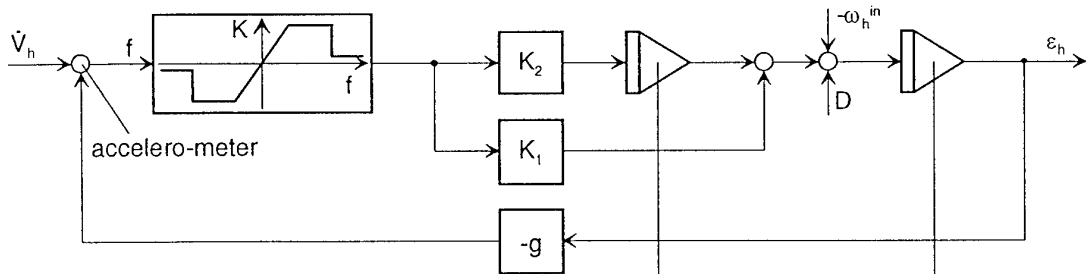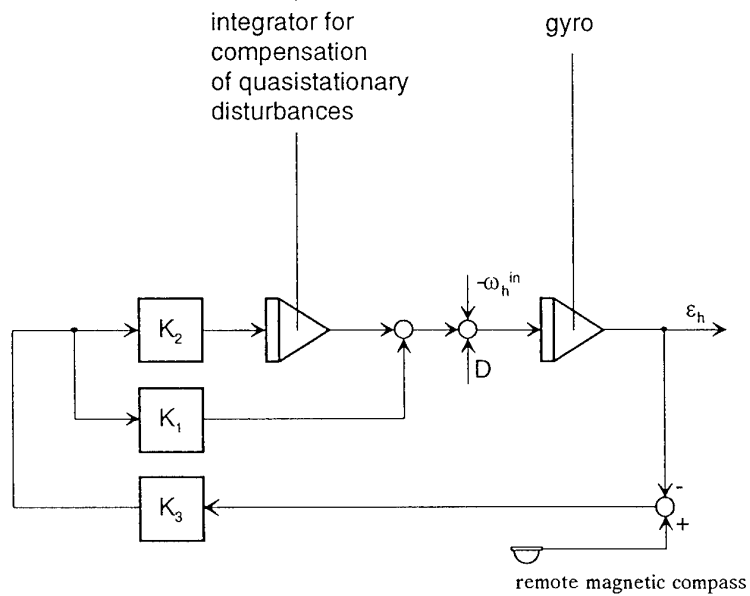
(index h = horizontal)

Figure 2.7 Control Loops for the Directional Gyro (CG) and the Vertical Gyro (VG)

2.2 The Gyro-Stabilized Platform as Directional Reference

As already mentioned, the VG's and DG's as fairly simple directional references are characterized by the fact that the gyro rotor has a high angular freedom with respect to the case - the angle between the rotor and the case being the aircraft's attitude or heading angle, for instance. Mechanical gyros for platform stabilization or angular rate measurements, shown in their principles in Fig. 2.8, have a very low angular freedom. The "Dynamically Tuned Gyro (DTG)" [St 82] is the modern version of such a sensor.

As to be seen from Fig. 2.8, one "Two-Degree-of-Freedom Gyro (TDF gyro)" can be used for stabilizing two axes.

The mechanical gyro for platform stabilization is a nullsensor. The angle $\Theta^{cg} = \Theta$ (pickoff angle) between the rotor and the case (subscript c) is closely held to zero by means of a control loop for properly slewing the case and platform. For the derivation of the block diagram for the TDF gyro in Fig. 2.9 we split up ω^{ig} in the Technical Gyro Equation 2.8 :

(2.15) $\omega^{ig} = \omega^{ic} + \omega^{cg} = \omega^{ic} + \dot{\Theta}^{cg}$

and solve it for the components of $\dot{\Theta}$. We introduce $\omega^{ic} = \omega$ for the input rate of the case with respect to inertial space. Due to the transfer function to be seen in Fig. 2.9:

(2.16) $\dfrac{\dot{\Theta}}{\omega} = \dfrac{\Theta}{\beta} = -1,$

the TDF gyro is often called "free rotor gyro (FRG)". We have put in this relationship $\beta = \int \omega \, dt$.

As to be seen from Fig. 2.10, the control loops for stabilizing the two axes of a platform connect the gyro pickoff output signals Θ via electronic networks F to the servo motors SM mounted on the gimbal axes parallel to the pickoffs. Since both axes of the TDF gyro can be regarded as uncoupled (s. Fig. 2.9), the block diagram for only one axis is shown in Fig. 2.11. We will use this for discussing the characteristics of the gyro-stabilized platform.

Three major subsystems are to be seen in this block diagram: the gyro, the electronics and the platform. The input into the latter one is the output current from the electronics. It is converted into a torque in the servo motor with the gain S_x^{SM}. The torquer time lag has been neglected. Disturbance torques M_x^{Pd} acting on the platform are due to friction in the gimbal bearings or due to mass unbalance of the platform, for instance. Both torques move the platform against its inertia. The platform output angular rate ω^{ip} with respect to inertial space is sensed by the gyro.

From the system's point of view we are only interested in the low frequency response of the gyro-stabilized platform, for which the inputs into the integrators of the block diagram must be zero.

At first we assume a disturbance torque M_x^{Pd} is acting on the platform and no command or disturbance torques M^t, M^d are acting on the gyro. Since for quasistationary conditions the input into the gyro as integrator must remain zero, this disturbance torque is compensated by a servomotor torque $i_x S_x^{SM}$, with the current generated by the control loop.

With similar arguments we can see, that a in the low frequency range command or disturbance torques acting on the gyro rotor are compensated by a gyroscopic torque $\omega_x \cdot H$, i.e. by a corresponding rotation of the platform. Making use of Eq. 2.10, we obtain as **Performance Equation of the Gyro-Stabilized Platform:**

(2.17) $\omega_p^{ip} = \omega^t + d$

which, in principle, is identical to Eq. 2.11 for the free gyro. This is not surprising, since the stabilization loop slews the platform with respect to the rotors of the 3 orthogonal gyros and the gyro-stabilized platform can be regarded as an ideally supported 3-axis gyro. Figs. 2.4 b to d can thus also be used for the representation of the gyro-stabilized platform in a block diagram or a functional diagram.

Stabilization of a 3-axis platform requires the use of 3 SDF gyros, as shown in Fig. 2.12 or 2 TDF gyros. In the latter case one gyro axis is redundant and often is used for fault detection, i.e. for checking whether the platform truly moves according to the command rate vector applied to the 3 gyro axes.

Fig. 2.12 shows 2 platform mechanizations, the "three- gimbal platform (TGP)" and the "four-gimbal platform (FGP)". The latter is a prerequisite for flight maneuvers, when the aircraft attitude passes a 90° pitch angle.

The definition of the Euler angles for roll, pitch and yaw for describing the aircraft's attitude and heading is given in Appendix C, Fig. C 2. The TGP is mechanized in a way, that these angles are directly measured at the gimbal axes via the resolvers RR, PR and YR. From this drawing and Appendix C, Fig. C 2 it can be deduced also, that at the pitch angle $\theta = 90^\circ$ the roll and yaw axes are parallel and the TGP has lost one degree of freedom. This state is called "gimbal lock".

The FGP is a TGP mounted in an additional outer gimbal OG', but with its roll and pitch axes transposed. The 3 innermost gimbals are gyro- stabilized as in the TGP, but the outer roll servo motor is controlled by the angle between the inner gimbal IG and the former outer gimbal OG. Its task is to keep the angle $\phi' = 0$ in normal flight. If the aircraft passes the critical region where gimbal lock would occur with the TGP, the control loop from the inner roll angle ϕ' to the outer servo motor changes its sign and becomes unstable. This causes the outer gimbal OG' to seek another equilibrium which is 180° away from the initial one. The FGP thus trades the "gimbal lock" problem for the "gimbal flip" problem which requires a servo loop fast enough so that the gyros do not hit their stops during the maneuver of the aircraft. In the Litton LN-3 Inertial Navigation System (INS) - mounted in the Starfighter and the Phantom Fighters - the gimbal flip requires less than .3 s to achieve 90° of the required 180° change. For more details see [St 82].

2.3 Misalignment Kinematics of an Earth-Referenced Gyro-Stabilized Platform

Since Eq. 2.17 is identical in principle with the characteristics of the free gyro, we can use the block diagram Fig 2.4c and Eq. 2.13 for the earth-referenced gyro-stabilized platform, too, and we can write:

(2.18) $\dot{\varepsilon}_n = \omega^t + d - \omega_p^{in}.$
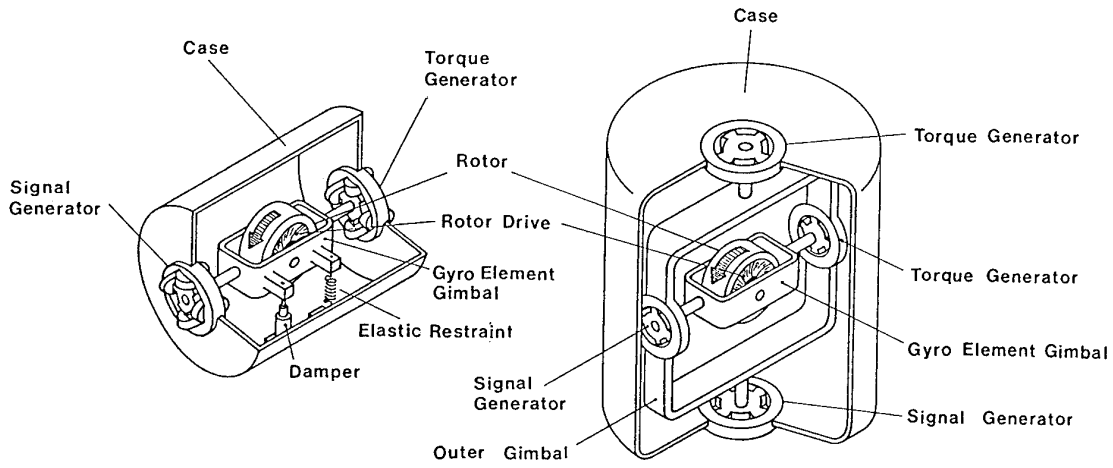
Figure 2.8 The Single-Degree-of-Freedom (SDF) and Two-Degree-of-Freedom Gyros in Principle
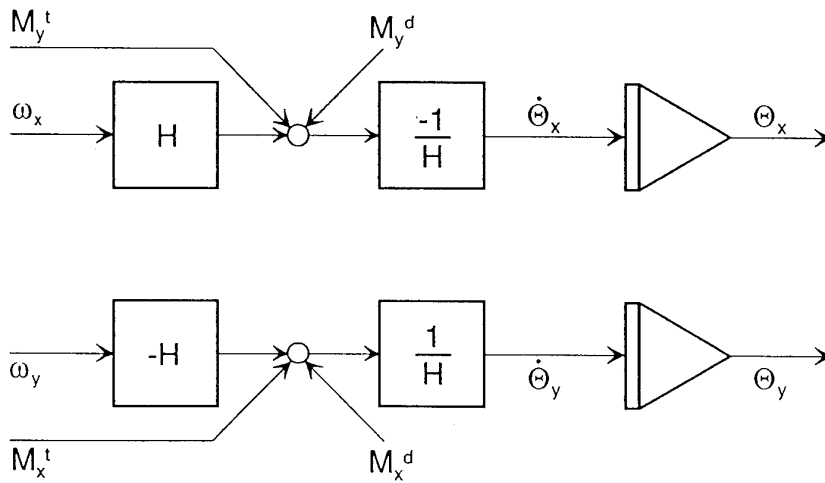
Figure 2.9 Block Diagram of the Two-Degree-of-Freedom Gyro, also Called Free Rotor Gyro
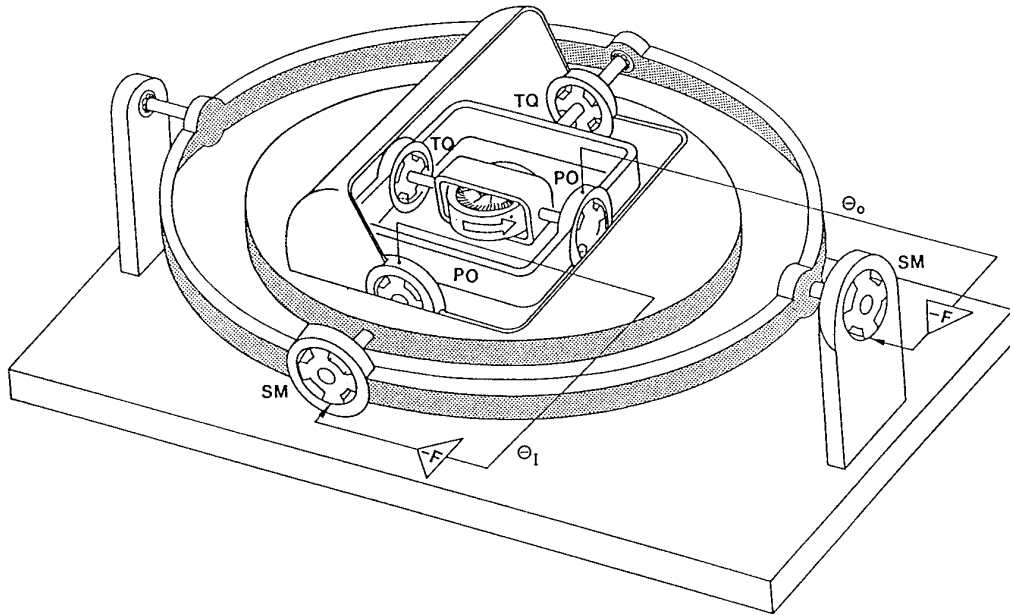
Figure 2.10 Stabilization of a Two-Axis Platform with a Two-Degree-of-Freedom Gyro
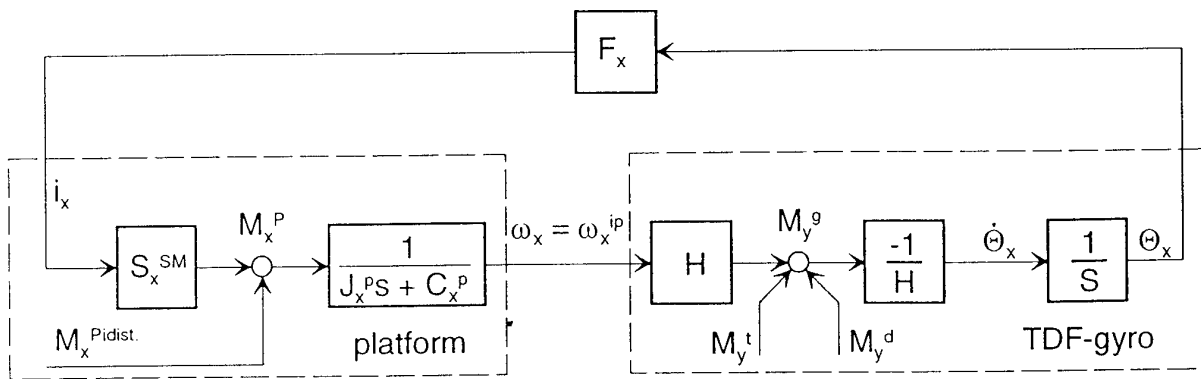
Figure 2.11 Block Diagram for the Platform Stabilization by Means of a Two-Degree-of-Freedom Gyro
(one axis shown only)

Figure 2.12 Schematic View of a Three-Gimbal Platform (TGP) and a Four-Gimbal Platform (FGP)

The vectors ω^t and $\mathbf{d}$ do not need the index p, since it is quite obvious that they are acting in the platform coordinate frame. The term ω^{in} is known in the navigational coordinate frame only. With the platform misalignment angles $\varepsilon = (\varepsilon_N \quad \varepsilon_E \quad \varepsilon_D)^T$ we may transform it into the platform frame (s. Appendix C, Eq. C 19 and C 20):

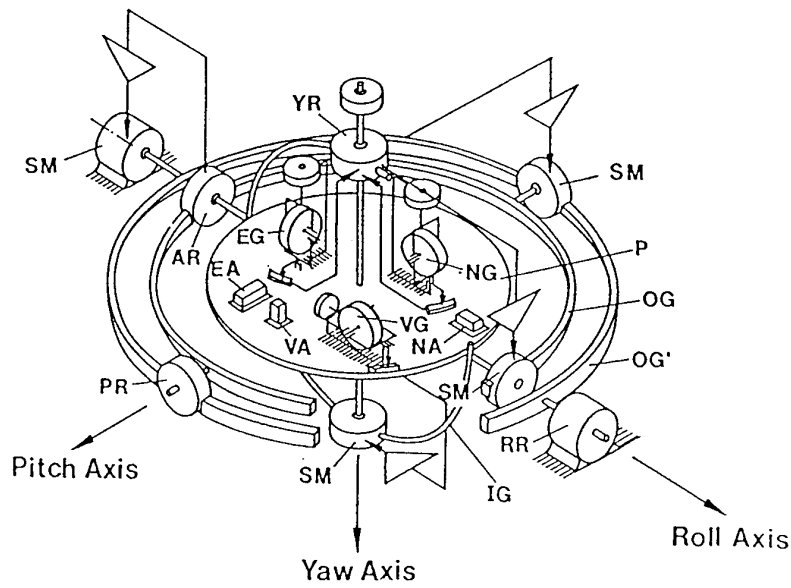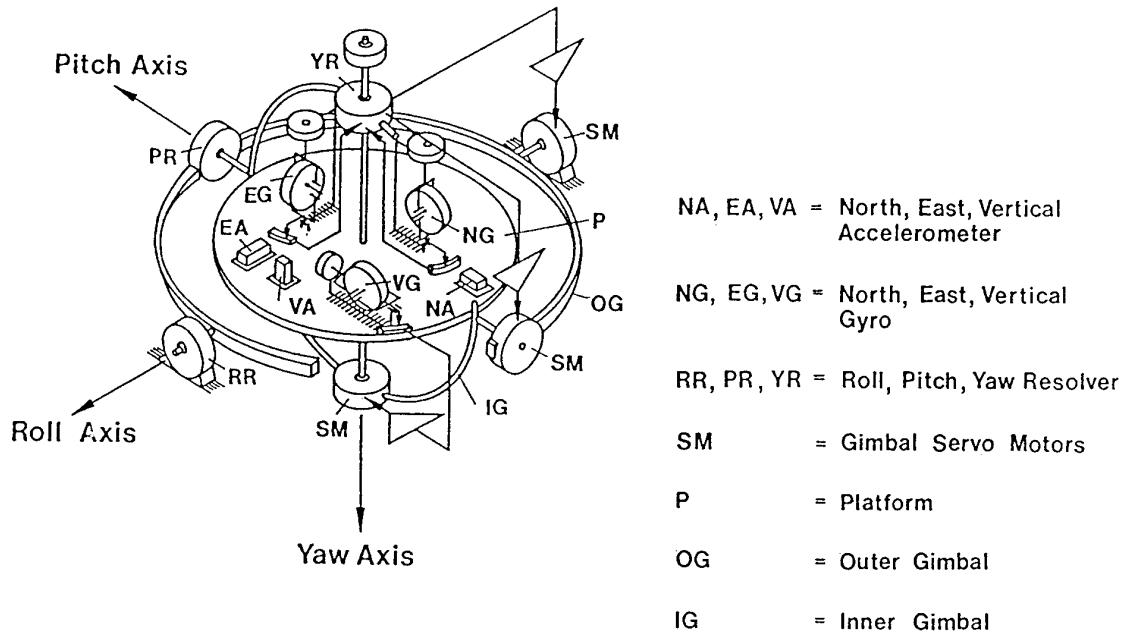$$(2.19) \qquad \omega_p^{in} = \omega_n^{in} + \varepsilon_n \times \omega_n^{in}.$$

For keeping ε_n to zero, i.e. the gyro-stabilized platform aligned with north, east and down, the right hand side of Eq. 2.18 must vanish, i.e. the gyros must in theory be torqued with $\omega^{in} - \mathbf{d}$. In practice this is never the case exactly and one has to live with deficiencies as indicated in Fig. 2.4 d by the asterix. There are firstly daily variations of the gyro drift. Secondly one has to rely on values for earth rate and transport rate compensation computed in the navigational computer and applied to the gyros as feedback signals in the so-called "Schuler feedback loop" (s. Section 4.2.2). They are corrupted by velocity and position errors resulting in in the error term $\delta\omega^{in}$. The platform slewing is also affected by the gyro scalefactor error κ and gyro reference axes alignment errors (α_{ij} = misalignment of the ith sensor reference with respect to the jth platform axis). The latter are comprised in the following matrix:

$$(2.20) \qquad \mathbf{K} = \begin{bmatrix} \kappa_x & \alpha_{xy} & \alpha_{xz} \\ \alpha_{yx} & \kappa_y & \alpha_{yz} \\ \alpha_{zx} & \alpha_{zy} & \kappa_z \end{bmatrix}.$$

We thus obtain for the true platform slewing rate:

$$(2.21) \qquad \omega^t = (\mathbf{I} + \mathbf{K}) \cdot (\omega^{ie} + \omega^{en})_n^* = (\mathbf{I} + \mathbf{K}) \cdot (\omega^{in} + \delta\omega^{in})_n \cong \omega_n^{in} + \delta\omega_n^{in} + \mathbf{K}\omega_n^{in}.$$

The gyro errors due to scalefactor and input axis alignment errors (elements of $\mathbf{K}$) and to drift $\mathbf{d}$ in Eq. 2.18 are combined to:

$$(2.22a,b) \qquad \begin{aligned} \delta\omega^g &= \mathbf{K} \cdot (\omega^{ie} + \omega^{en})_n + \mathbf{d} \\ &= \mathbf{K} \cdot \omega^t + \mathbf{d}, \end{aligned}$$

whereby $\mathbf{K}\omega^{en}$ generates roughly speaking groundspeed-dependent and $\mathbf{K}\omega^{ie} + \mathbf{d}$ time-dependent platform misalignment errors. The effect of $\mathbf{K}(\omega^{ie} + \omega^{en})_n$ is more dominant during easterly than westerly flights.

The variations of the platform misalignments with respect to north, east and down over time thus are governed by the following differential vector equation:

$$(2.23) \qquad \dot{\varepsilon}_n = \varepsilon_n \times \omega_n^{in} + \delta\omega_n^{in} + \delta\omega^g.$$

Assuming the geostationary case ($\omega^{in} = \omega^{ie}$), perfect platform slewing ($\delta\omega^{in} = 0$) and gyro bias drift only ($\delta\omega^g = \mathbf{D}$), we obtain the following relationship:

$$(2.24) \qquad \dot{\varepsilon} = \varepsilon \times \omega^{ie} + \omega^t - \omega^{ie} + \mathbf{D} = \Omega \begin{bmatrix} -\cos\phi - \varepsilon_E \sin\phi \\ \varepsilon_D \cos\phi + \varepsilon_N \sin\phi \\ \sin\phi - \varepsilon_E \cos\phi \end{bmatrix} + \omega^t + \mathbf{D},$$

which is shown in Fig. 2.13 as block diagram. All 3 platform axes are coupled within an oscillation network. Oscillations are excited by initial misalignment angles, for instance.

The following differential equation can be derived for the input of the east-west axis:

$$(2.25) \qquad \dot{\varepsilon}_E = \Omega \cos\phi \int \varepsilon_E (-\Omega \cos\phi) \, dt + \Omega \sin\phi \int \varepsilon_E (-\Omega \sin\phi) \, dt$$

$$(2.26) \qquad \ddot{\varepsilon}_E + \varepsilon_E \cdot \Omega^2 \cdot (\cos^2\phi + \sin^2\phi) = 0,$$

which shows that the frequency of this oscillating network is equal to earth rate ($\omega = \Omega$). This should not surprise us if we remember that a gyro and thus the gyro-stabilized platform keeps its attitude with respect to inertial space, for zero input driving functions. On the rotating earth it carries out a motion in the negative sense of the earth rate, as indicated in Fig. 2.5.

2.4 The Analytic Platform as Directional Reference in Strapdown Systems

In strapdown computers the reference direction is computed in form of the C_{nb} matrix - a 3×3 matrix for vector transformation from the body (index b) to the navigational (index n) frame, with its axes north (x = N), east (y = E) and down (z = D, s. Appendix C, Eq. C10). With the computed transformation matrix C_{nb} the acceleration f_b measured in the body-fixed coordinate system can be transformed to the navigational coordinate system according to:

$$(2.27) \qquad f_n = C_{nb} \cdot f_b.$$

The C_{nb} matrix thus is the directional reference for strapdown inertial navigation. It can be regarded as the "Analytic Platform" as indicated by the weathercocks in Fig. 1.2 in comparison with Fig. 1.1. There is no difference in principle between the acceleration and the computational process of inertial navigation following the weathercocks, i.e the directional references in strapdown and platform systems - regardless of the aircraft maneuver, the computed vertical acceleration will always point downwards in a strapdown system, too! From Appendix C, Eq. C11 we also see that the C_{nb} matrix allows to compute attitude and heading of the vehicle as already indicated in Fig. 1.2.

These considerations give an indication of the allowable overall drift of the reference direction computation. Its numerical integration must be of "inertial quality", i.e. its uncertainty << .01 deg/h. More details about the problems for computing this matrix are discussed in Appendix D.

Now we turn only to the principles of the C_{nb} matrix computation. It can be computed from the matrix differential equation:
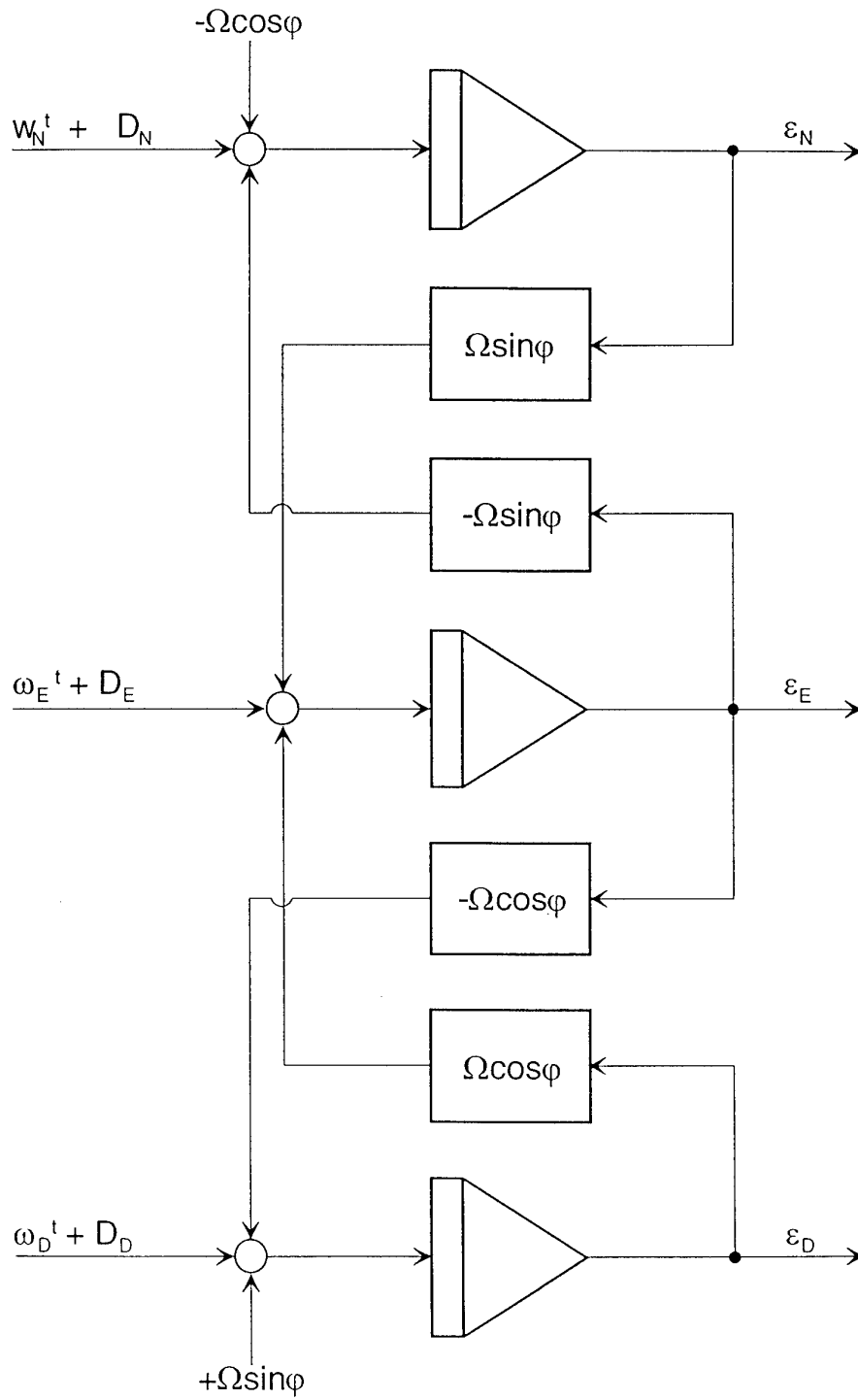
Figure 2.13 Kinematics of the Earth-Referenced Geostationary Gyro-Stabilized Platform

60

(2.28) $\qquad \dot{C}_{nb} = C_{nb} \cdot \Omega_b^{nb}$,

where

(2.29) $\qquad \Omega_b^{nb} = \begin{bmatrix} 0 & -\omega_z & \omega_y \\ \omega_z & 0 & -\omega_x \\ -\omega_y & \omega_x & 0 \end{bmatrix}_b^{nb} = \omega_b^{nb} \times .$

This skew symmetric Ω matrix can be regarded as the $\omega \times$ operator, i.e its matrix-vector product is identical to the vector cross product. Ω_b^{nb} is the result of the aircraft angular rate ω_b^{ib} measured by the gyros in the body frame b and the Schuler feedback ω_n^{in} computed from earth and transport rate in the navigational frame n:

(2.30) $\qquad \Omega_b^{nb} = \Omega_b^{ib} - C_{bn} \cdot \Omega_n^{in} \cdot C_{nb} .$

This relationship will be used in the error analysis further below. In practice the C_{nb} matrix is mostly derived from the product of the 2 matrices:

(2.31) $\qquad C_{nb} = C_{ni} \cdot C_{ib} = C_{in}^T \cdot C_{ib} ,$

i.e. the transformation matrices from the navigational to the inertial frame and from the inertial to the body frame. Again it should be mentioned that the former changes with earth and transport rate, i.e. very slow with regard to time and the latter with the aircraft maneuver, i.e. very fast. Its computation is most demanding as regards computer speed and algorithms. This is discussed in more detail in Appendix D.

Assuming the first order integration algorithm would deliver sufficient accuracy, we may write for the time increment t_n:

(2.32) $\qquad C_{ib}(t_n) = C_{ib}(t_{n-1}) + \int_{t_{n-1}}^{t_n} C_{ib}(\tau) \cdot \Omega_b^{ib}(\tau) \, d\tau \cong C_{ib}(t_{n-1}) \cdot (I + \Delta\Phi_b^{ib}),$

with the angular increment matrix :

(2.33) $\qquad \Delta\Phi_b^{ib} \cong \Omega_b^{ib} \cdot \Delta t \cong \begin{bmatrix} 0 & -\Delta\phi_z & \Delta\phi_y \\ \Delta\phi_z & 0 & -\Delta\phi_x \\ -\Delta\phi_y & \Delta\phi_x & 0 \end{bmatrix}_b^{ib} = \Delta\phi_b^{ib} \times.$

The elements of $\Delta\Phi_b^{ib}$ in this first order algorithm are the angle increments $\Delta\phi^P$ of the mechanical or optical gyros. They deliver a pulse train proportional to the input angular rate component ω^{ib}, whereby each pulse has an identical weight $\Delta\phi^P$ (s. Appendices D and O). The pulse train is counted in an up/down counter and read out at sampling times T^s to deliver the integral of the measured angular rate over the sampling interval. Written as a vector it is:

(2.34) $\qquad \Delta\phi^{ib} = \int_{t_n}^{t_n + T^s} \omega^{ib} \, dt.$

For the moment we have assumed ω^{ib} to be without an error.

The integration of $\dot{C}_{in}$ in Eq. 2.31 is carried out similarly but does not cause major numerical problems, since the elements of Ω^{in} contain the sum of earth rate and transport rate only. The latter is the so-called "Schuler feedback" discussed in Chapter 4.2.

2.5 Misalignment Kinematics of the Directional Reference in Platform and Strapdown Systems

Up to now we have discussed the computational problems only. For any gyroscopic measurement of ω^{ib} deficiencies $\delta\omega^{ib}$ due to scalefactor errors and drift (s. Appendix O for optical gyros or the cited references for mechanical gyros) and due to the digitizing of analog signals (s. Appendix D) have to be taken into account:

(2.35a,b) $\qquad \omega^{ib*} = \omega^{ib} + \delta\omega^{ib} = (I + K) \cdot (\omega^{ib} + d) \cong (I + K) \cdot \omega^{ib} + d$

$\qquad\qquad \delta\omega^{ib} = \delta\omega^g = K\omega^{ib} + d,$

with the scalefactor and input axes misalignment matrix K as shown in Eq. 2.20 and d = gyro drift vector. The superscript "g" refers more distinctly to a gyro error.

After integration, i.e. on the level of the directional reference, K causes maneuver-dependent errors and d time-dependent errors.

The rest of this section is devoted to the discussion of the common features of the gyro-stabilized and the analytic platforms with respect to their error characteristics. Once this has been achieved, we have found an essential basis for the common features of the navigational system error propagation within platform and strapdown systems.

The subject of the following train of thoughts is the error propagation of the C_{nb} matrix, whose computation is based on Eq. 2.28. Instead of the exact matrix C_{nb} we introduce the erroneous one:

(2.36) $\qquad C_{nb}^* = C_{n^*n} \cdot C_{nb} = (I - E) \cdot C_{nb},$

where $C_{n^*n} = (I - E)$ (s. Appendix C, Eq. C 20) comprises the small angle deviation elements ε_N, ε_E, and ε_D between the

computed (index n*) and the true (index n) navigational reference coordinate frame. The time derivative of C_{nb}^* can either be computed from $C_{nb}^* = (I - E) \cdot C_{nb}$:

(2.37a) $\qquad \dot{C}_{nb}^* = - \dot{E} \cdot C_{nb} + (I - E) \cdot \dot{C}_{nb} = - \dot{E} \cdot C_{nb} + (I - E) \cdot C_{nb} \cdot \Omega_b^{nb}$,

or it can be computed according to Eq. 2.28:

(2.37b) $\qquad \dot{C}_{nb}^* = C_{nb}^* \cdot \Omega_b^{nb*} = (I - E) \cdot C_{nb} \cdot \Omega_b^{nb*}$.

The angular rate matrix Ω_b^{nb*} is according to Eq. 2.30 computed from the difference of the gyro measurements Ω_b^{ib*} and the Schuler feedback Ω_n^{in*}. Due to velocity and position errors the latter has an error of $\delta\Omega_n^{in}$. The transformation of Ω_n^{in*} into the body frame is based on the computed C_{bn}^* matrix:

(2.38) $\qquad \Omega_b^{in*} = C_{bn}^* \cdot (\Omega_n^{in} + \delta\Omega_n^{in}) \cdot C_{nb}^*$

$\qquad\qquad\qquad = C_{bn} \cdot (I - E) \cdot (\Omega_n^{in} + \delta\Omega_n^{in}) \cdot (I - E) \cdot C_{nb}$.

The gyro measurement matrix $\Omega_b^{ib*} = \Omega_b^{ib} + \delta\Omega^g$ is already given in the body frame. According to Eq. 2.35 it is in error by drift, scalefactor error and misalignment. All are comprised in the skew symmetric matrix:

(2.39) $\qquad \delta\Omega^g = \begin{bmatrix} 0 & -\delta\omega_z & \delta\omega_y \\ \delta\omega_z & 0 & -\delta\omega_x \\ -\delta\omega_y & \delta\omega_x & 0 \end{bmatrix}^g$.

Subtracting Eq. 2.37a and 2.37b and using the relationship $C_{nb} \cdot \Omega_b \cdot C_{bn} = \Omega_n$, we obtain the following matrix differential equation:

(2.40a) $\qquad \dot{E} = E \cdot \Omega_n^{in} - \Omega_n^{in} \cdot E + \delta\Omega_n^{in} + C_{nb} \cdot \delta\Omega^g \cdot C_{bn}$.

This can be written as a vector differential equation:

(2.40b) $\qquad \dot{\varepsilon}_n = \varepsilon_n \times \omega_n^{in} + \delta\omega_n^{in} + C_{nb} \cdot \delta\omega^g$

This equation is identical in principle to Eq. (2.23) for the gyro-stabilized platform. It differs insofar only as the gyro drift affects the analytic directional reference computation via the C_{nb} matrix; in a strapdown system the gyros are "strapped down" to the vehicle, i.e. they are moving with respect to north, east and down according to the aircraft maneuver. This is visualized in Fig. 2.14; in a 360 deg turn the effects of constant sensor errors will thus partially cancel out. During this maneuver the error propagation of the strapdown system is similar to the one of Carousel INS. This is a platform system with continuous rotation of the horizontal sensors about the vertical axis for avaraging out the effect of their errors.

It is interesting to note that the directional accuracy of the strapdown system does not depend directly upon the aircraft maneuver - only indirectly through the measurement errors **d** which contain terms due to gyro scalefactor error and input axis misalignment.

Since the C_{nb} matrix is the only term in Eq. 2.40b hinting at the gyro axes coordinate frame, we may generalize its use for all inertial system mechanizations navigating with respect to the navigational coordinate frame. The index b is then replaced by the gyro measurement axes coordinate frame m:

(2.41c) $\qquad \dot{\varepsilon}_n = \varepsilon_n \times \omega_n^{in} + \delta\omega_n^{in} + C_{nm} \cdot \delta\omega^g$

and we may write:

- for the "North Indicating System (NIS)", i.e. the platform system with its axes pointing north, east and down:

$\qquad m = n$, i.e. $C_{nm} = I$,

(the system error dynamics derived in Section 4.2.2 are characteristic for all the subsequent system mechanizations);

- for the platform system measuring in the wander azimuth coordinate frame:

$\qquad m = a$;

(in Section 4.3 it is shown that this mechanization avoids the singularities of the NIS at the geographic poles);

- for the platform system measuring in a coordinate frame which is continuously rotating about the vertical axis (Carousel System):

$\qquad m = c$

(by means of the "carouseling" the effect of the horizontal sensor errors upon the navigational accuracy should be reduced);

- for the platform system measuring in the inertial coordinate frame:

$\qquad m = i$;

(electrostatic gyros are operated untorqued, i.e. the platform is space-stabilized),

- and finally for the strapdown system:

$\qquad m = b$

(this is the modern INS mechanization).
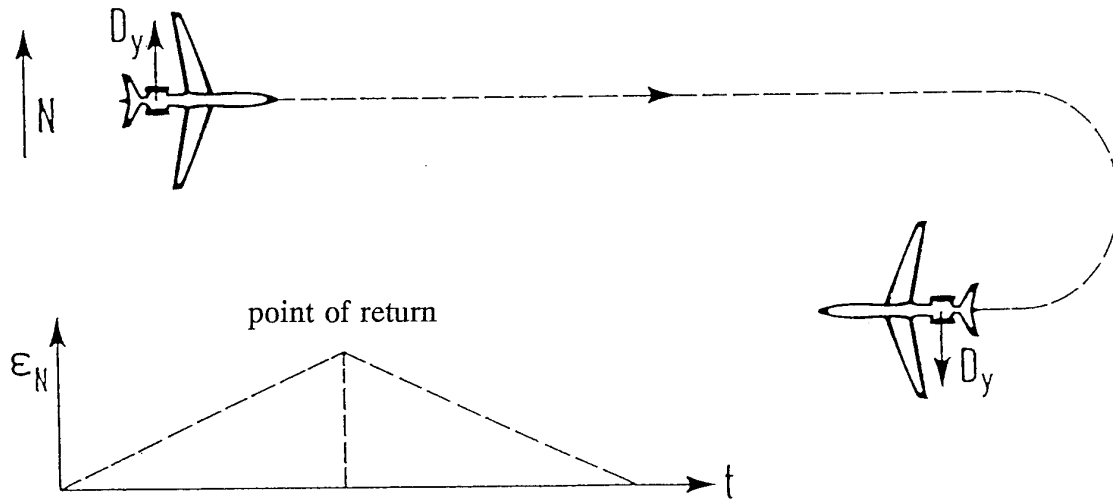
Figure 2.14 Horizontal Misalignment Error Growth During a Straight Flight with Subsequent Turn in a Strapdown System
Due to a Constant Drift

Literature Chapter 2

[De 70] Denhard W.G. (Ed): " Inertial Component Testing: Philosophy and Methods"
AGARDograph 128, AGARD 1970

[St 82] Stieler B., Winter H.: "Gyroscopic Instruments and their Application to Flight Testing"
AGARDograph No. 160, Vol. 15, 1982

[Wr 69] Wrigley W., Hollister W. M., Denhard W.G.: "Gyroscopic Theory, Design, and Instrumentation"
The M.I.T. Press Cambridge, Mass. USA, 1969

3. The Measurement of the Translational Motion
- Accelerometers and Acceleration with Respect to Ground

3.1 The Pendulous Accelerometer and its Output Signal

The principle of a conventional accelerometer is shown in Fig. 3.1. A proof mass is suspended in a case and confined to a zero position with the help of a spring or a rebalance loop as described below. In general, damping is added to give the spring/mass system a proper dynamic transfer function. Aircraft accelerations act upon the accelerometer case and cause the mass to react with a displacement with respect to the zero position so that the resulting spring force F compensates the acting acceleration. The displacement of the mass with respect to the case is then proportional to F.

For convenience we introduce the specific spring force $f = F/m$ and call it the accelerometer output signal as indicated in Fig. 3.1.

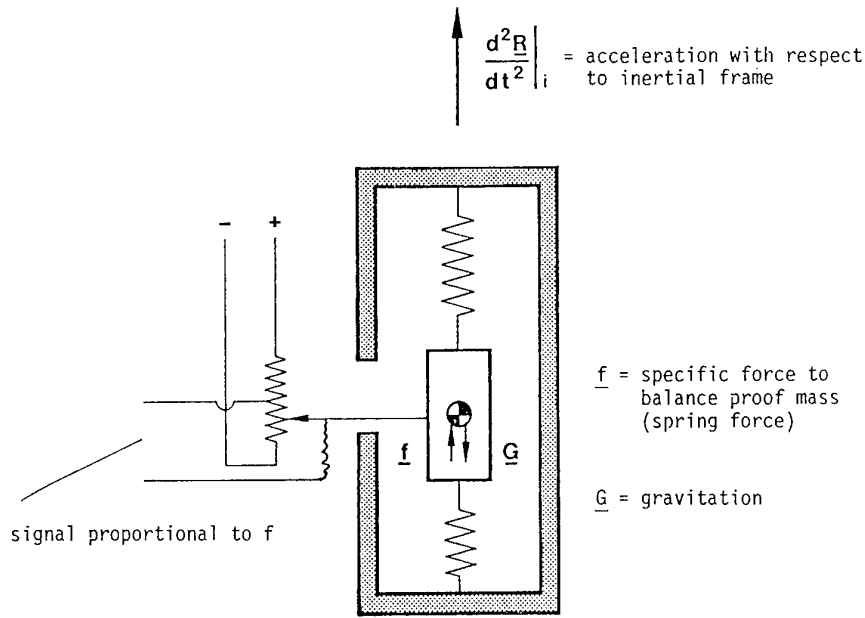Not only f but also gravitation G is acting upon the proof mass. Both cause its acceleration with respect to inertial space due to Newton's second law:

$$(3.1a,b) \qquad f + G = \left. \frac{d^2 R}{dt^2} \right|_i \qquad or \qquad f = \left. \frac{d^2 R}{dt^2} \right|_i - G,$$

where R is the radius vector from the earth's center of gravity which is the origin of the inertial coordinate system (index i), to the proof mass' center of gravity.

In some cases it is convenient to combine the right hand side of Eq. 3.1b into a vector a or $- g'$ when either the effect acceleration or of gravitation is of prime importance. The specific force vector is then:

$$(3.2) \qquad f = a \qquad or \qquad f = - g'.$$

The type of accelerometer that is presently used in most operational inertial navigation systems is the restrained pendulum accelerometer [QF]. Fig. 3.2 shows the construction principle of such a device. A pendulous mass is suspended and restrained to a zero position by a control loop. Optical, capacitive or inductive pickoffs detect a deflection of the pendulum which is forced back to its zero position by means of this control loop with an input current into the torquer. This current i that is necessary to compensate an acceleration and to bring the pendulum back to its null position is then a measure of the specific force f. It is converted into a voltage u by means of a precision resistor.

The true output signal of an accelerometer contains input signal-dependent errors comprised in the scale factor error κ and input signal-independent errors comprised in b:

$$(3.3) \qquad f^* = (1 + \kappa) \cdot (a + b) \approx (1 + \kappa) \, a + b.$$

As soon as both parameters can be modelled - as a function of temperature, for instance - they do not affect the measurement accuracy. What bothers is the stochastic variation, i.e. the day-to-day repeatability, for instance. This is a true indication of quality. Threshold is another one.

3.2 The Acceleration with Respect to Ground

The second time derivative of the radius vector in Eq. 3.1 is the acceleration of the point P of measurement with respect to the earth's center of gravity which is the origin O of the inertial frame (index i = Earth Centered Inertial Frame (ECI), s. Chapter "Navigation Coordinate Frames")). It can be expressed in any other reference coordinate frame (index r), for instance the line of sight (LOS) coordinate frame for missile application or the navigational frame (index n) for navigation with respect to the earth through application of the Theorem of Coriolis:

$$(3.4) \qquad \left. \frac{d R}{d t} \right|_i = \left. \frac{d R}{d t} \right|_r + \omega^{ir} \times R.$$

$$(3.5ab) \qquad \left. \frac{d^2 R}{d t^2} \right|_i = \left. \frac{d^2 R}{d t^2} \right|_r + \left. \frac{d \omega^{ir}}{d t} \right|_r \times R + 2 \, \omega^{ir} \times \left. \frac{d R}{d t} \right|_r + \omega^{ir} \times (\omega^{ir} \times R)$$

$$= \left. \frac{d^2 R}{d t^2} \right|_n + \left. \frac{d \omega^{in}}{d t} \right|_n \times R + 2 \, \omega^{in} \times \left. \frac{d R}{d t} \right|_n + \omega^{in} \times (\omega^{in} \times R)$$

with   vertical   horizontal   Coriolis   centrifugal accelerations.

The angular rate ω^{ir} or ω^{in} are the respective sums of earth rate plus transport rate (s. Appendix C, Eq. C 7 to C 9).

Eq. 3.5 simplifies considerably if we introduce the velocity of the point P of measurement with respect to the surface of the reference ellipsoid of the earth:

$$(3.6) \qquad \left. \frac{d R}{d t} \right|_e = V = \left. \frac{d R}{d t} \right|_n + \omega^{en} \times R,$$

which, expressed in the n-frame coordinates, is:

64

$$\frac{d^2\underline{R}}{dt^2}\bigg|_i = \text{acceleration with respect to inertial frame}$$

$\underline{f}$ = specific force to balance proof mass (spring force)

$\underline{G}$ = gravitation

signal proportional to f

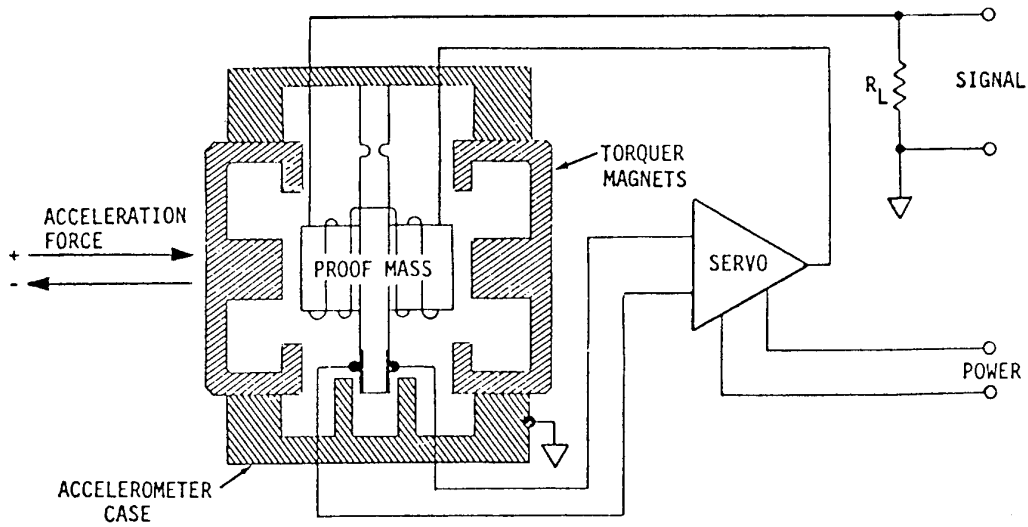Figure 3.1 Forces and Output Signal in an Accelerometer

Figure 3.2 Principles of the Pendulous Accelerometer with Caging Loop

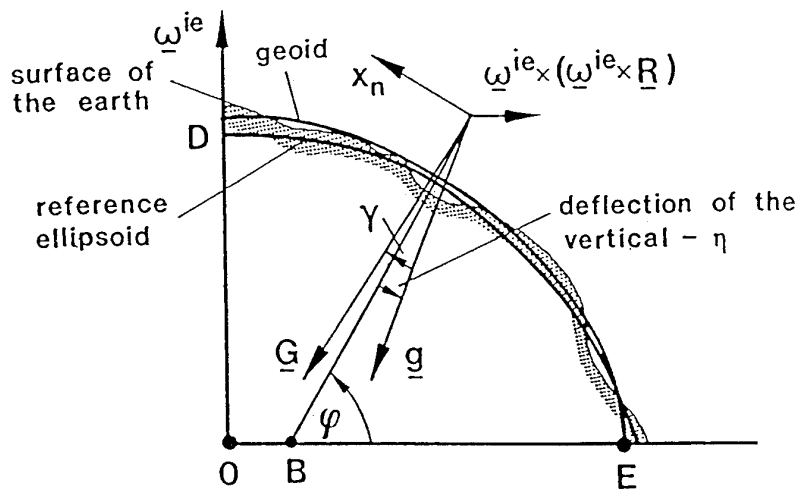$$(3.7) \qquad \mathbf{V}_n = \begin{bmatrix} V_N \\ V_E \\ V_D \end{bmatrix} = \begin{bmatrix} (R_N + h)\ \dot{\phi} \\ (R_E + h)\ \dot{\lambda}\ \cos\ \phi \\ -\dot{h} \end{bmatrix}.$$

The resulting acceleration of P is:

$$(3.8) \qquad \frac{d^2\ \mathbf{R}}{d\ t^2}\bigg|_i = \dot{\mathbf{V}} + (2\ \boldsymbol{\omega}^{ie} + \boldsymbol{\omega}^{en}) \times \mathbf{V} + \boldsymbol{\omega}^{ie} \times (\ \boldsymbol{\omega}^{ie} \times \mathbf{R}).$$

The last term on the right hand side is compounded with the **gravitation G** in Eq. 3.1 to give the **gravity g** as shown in Fig. 3.3:

$$(3.9) \qquad \mathbf{g} = \mathbf{G} - \boldsymbol{\omega}^{ie} \times (\boldsymbol{\omega}^{ie} \times \mathbf{R}) = \mathbf{G} - R \cdot \Omega^2 / 2 \cdot \begin{bmatrix} -\sin\ 2\phi \\ 0 \\ 1 + \cos\ 2\phi \end{bmatrix}.$$

$$\Upsilon = \frac{R\Omega^2}{2g}\sin 2\varphi \le \pm .6 \text{ arc min}$$

$$|\underline{G} - \underline{g}| = \frac{R\Omega^2}{2}(1 + \cos 2\varphi) \le 3.4 \cdot 10^{-3} g$$

Figure 3.3 Gravitation Vector **G** and Gravity Vector **g**

So we obtain for the ground acceleration:

$$(3.10a,b) \qquad \dot{\mathbf{V}} = \mathbf{f} - (2\ \boldsymbol{\omega}^{ie} + \boldsymbol{\omega}^{en}) \times \mathbf{V} + \mathbf{g}$$

$$= \mathbf{f} - \mathbf{c} + \mathbf{g}$$

or in components:

$$(3.10c) \qquad \begin{bmatrix} \dot{V}_N \\ \dot{V}_E \\ -\ddot{h} \end{bmatrix} = \begin{bmatrix} f_N \\ f_E \\ f_D \end{bmatrix} + \begin{bmatrix} -2\ V_E\ \Omega'\ \sin\ \phi + \dot{h}\ \dot{\phi} \\ 2\ V_N\ \Omega'\ \sin\ \phi - 2\ \dot{h}\ \Omega'\ \cos\ \phi \\ -2\ V_E\ \Omega'\ \cos\ \phi - V_N\ \dot{\phi} \end{bmatrix} + \begin{bmatrix} \xi\ g \\ -\eta\ g \\ g \end{bmatrix},$$

where these terms must be compensated in the computer: $\uparrow$ $\qquad\qquad$ $\uparrow$

The following abbrevations have been used:

$$(3.10d) \qquad \Omega' = \Omega + \dot{\lambda}/2$$

and the Coriolis acceleration vector:

$$(3.10e) \qquad \mathbf{c} = (2\ \boldsymbol{\omega}^{ie} + \boldsymbol{\omega}^{en}) \times \mathbf{V}.$$

A good example for understanding Coriolis acceleration is to think of a car driving with constant speed V_N from the equator to the northpole. Though not accelerating with respect to the ground an east-west accelerometer will sense an acceleration because the earth's tangential velocity of $\Omega \cdot R = 1668$ km/h has to be slowed down to zero at the northpole.

3.3 The Accelerometer Output on a Misaligned Directional Reference

In platform systems the acceleration is measured in the platform coordinate frame $\mathbf{f}_p{}^* = \mathbf{C}_{pn} \mathbf{f}_n{}^*$, with $\mathbf{C}_{pn}$ the transformation matrix containing the platform misalignment. According to Appendix C, Eq. C 20 it is composed of the small angles $\varepsilon_{N,E,D}$. In the strapdown system the acceleration is measured in the body frame and then transformed into the navigational frame via the $\mathbf{C}_{nb}^*$ matrix which is in error with respect to the accurate one by the small misalignment matrix $\mathbf{C}_{n^*n}$. Both misalignment matrices are identical (s. Eq. C 20). With the same arguments as in Section 2.5 we can make for all inertial navigation systems the following assumption for the acceleration in the navigational coordinate frame:

$$(3.11) \qquad \delta \dot{\mathbf{V}}_n = - \boldsymbol{\varepsilon}_n \times \mathbf{a}_n - \delta \mathbf{c}_n + \delta \mathbf{g}_n + \mathbf{C}_{nm} \cdot \mathbf{b}_m \, ,$$

where the index m stands for the coordinate frame, in which the accelerometers are measuring (s. Section 2.5).

Literature Chapter 3

[QF]: "QA-1000 Series, QA-1200-Series and QA-1300 Series Q-Flex Accelerometers."
Sundstrand Data Control, Inc. Document 012-0293-001

4. Inertial Navigation
Integration of the Directional Reference with the Translational Motion Measurement

4.1 Introduction

In Section 4.2 we will begin the discussion on inertial navigation systems (INS) and their error characteristics based on the mechanization indicated in Fig 1.1. The platform shown here is always kept horizontal and pointing north regardless of the vehicle's maneuver. Therefore an inertial navigation system (INS) with this characteristic is called "North Indicating System (NIS)" in the following. Special emphasis is also laid on the vertical channel in which as compared to the horizontal channels the physical law of mass attraction between two bodies (earth and proof mass of the accelerometer) leads to an instability of the error growthz. The aiding of the vertical channel with the barometric altitude will be discussed.

The NIS can certainly not be used for passing the geographic poles where all directions are pointing south or north. For worldwide inertial navigation another mechanization is required. It is the so-called "Wander Azimuth INS" discussed in Section 4.3. Some words will be said in Section 4.4 about the INS with space-stabilized coordinate frames used on earth and in space. As pointed out in Section 2.3, the functioning of the mechanical platform can be taken over by an analytic platform computed in the navigation computer. The data processing following this analytic platform computation is identical to that of the mechanical platform system as indicated in Fig. 1.2 for the NIS as an example. Section 4.4 deals with this modern INS mechanization, commonly known as "strapdown systems (SDS)". The characteristics of the error dynamics for all inertial navigation systems constitute the topic of Section 4.6. The error model derived therefrom can be used in a Kalman filter for the integration of any platform or strapdown INS navigating in terms of longitude and latitude with external sensors and systems.

As autonomous inertial navigation systems are expensive and other navigational information often lies at the wayside and can be blended with the inertial information, accuracy requirements on the inertial side can be reduced thus leading also to a reduction of costs. Inertial systems of this kind are often referred to as "attitude and heading reference systems (AHRS)". They will be treated in Section 4.7.

4.2 Inertial Navigation with Sensors Aligned to North, East and Down
- the North Indicating System (NIS)

4.2.1 Functioning of the NIS

Fig. 1.1 shows the diagram of an NIS which we use to obtain an understanding of the inertial navigation system (INS) functioning.

The platform represented as a round disc in Fig. 1.1 is suspended in the vehicle (airplane, missile, ship) with three degrees of rotational freedom. For simplification this figure shows only the vertical axis. The rotational freedom as well as the other indicated signals are understood to have 2 or 3 orthogonal components; 2 components for horizontal acceleration, velocity and position, 3 components for the angular rate.

On the platform one gyro can be seen - in place of the three gyros - which, with the control loop from gyro pickoff (P) via the electronics to the servo-motors at the gimbals, ensures the stabilization of the platform.

As we have seen in Section 2.2, the gyro-stabilized platform used as a directional reference with respect to the navigational frame, follows the performance equation: $\varepsilon = \omega^t + d - \omega^{in}$, regardless of the vehicle's maneuver. To keep the misalignment angle vector $\varepsilon = (\varepsilon_N \quad \varepsilon_E \quad \varepsilon_D)^T << 1$ deg in magnitude, the condition for torquing the 3 gyros $\omega^t = \omega^{in} - d$ has to be met, i.e. with $\omega^{in} = \omega^{ie} + \omega^{eh}$ (s. Appendix C, Eqs. C 8 and C 9) it has to be equal to the sum of earth rate plus transport rate plus drift as indicated in Fig. 1.1.

The computation of the earth rate and transport rate vectors are dependent on the computed ground speed and position, i.e. on the measured acceleration vector f^*. Eq. 3.10 is the basis for their computation.

Since with the NIS the geographic coordinate frame is stored in the platform, the aircraft attitude and azimuth can directly be measured at the gimbal axes (s. Fig. 1.1).

As already mentioned above, pure inertial navigation can only be carried out in the horizontal plane and Fig 1.1 is valid for this case only. Navigation in the vertical plane has to be aided by the barometric altitude, as will be discussed in Section 4.2.3.

4.2.2 The Error Dynamics in the Horizontal Channels of the NIS

4.2.2.1 The Error Dynamics of a Single Axis Inertial Navigation System

The signal flow of Fig. 1.1 for the two horizontal channels from the accelerometer, via the integrators and the division by the representative radius of curvature of the earth, to the torquers of the gyros are called Schuler loops. They are the electronic implementation of a two-axis mathematical pendulum whose length is equal to the radius R of the earth, having the frequency and period of:

$$(4.2.1a,b) \qquad \omega^s = \sqrt{g/R} = 1.235 \cdot 10^{-3} \text{ rad/s}, \qquad T^s = 84.4 \text{ min}$$

which are called "Schuler frequency" and "Schuler period".

A pendulum tuned to this frequency always indicates the vertical on a moving vehicle, once it has been aligned to it prior to the start. Neither a gyrocompass (s. Section 5) nor an INS tuned to this frequency will be excited to disturbance oscillations by the horizontal acceleration. These rules were revealed by Schuler in his publication of 1923 [Sc 23]. The Schuler period also governs the motion of a stone traversing a hole drilled through the earth regardless in direction - friction obviously neglected [Ma 78]. It governs also the motion of a satellite surrounding the earth when it is computed from the orbital radius and gravitation. For a "roof top" orbit it is again 84 minutes. We will come back to the satelite period in Section 4.4.

Neither a mathematical pendulum nor a physical pendulum can be implemented with this frequency. For the latter the following relationship would have to be satisfied:

(4.2.2) $i^2 / r = R.$

Assuming i = radius of inertia = 2.5 m, the pivot-to-center-of mass separation of r = 1 μm would have to be manufactured. In practice the Schuler tuning condition can only be met with the aid of a gyroscope - as in a gyrocompass or the INS.

The simplified single axis INS shown in Fig. 4.2.1 may serve as a basis for the following discussion.

The platform (P) is mounted on a vehicle heading north and gyro-stabilized about its east-west axis (perpendicular to the plane of the drawing) by means of the servo loop from the gyro (G) signal generator (S) to the servomotor (SM) of the platform. The accelerometer (A) has its sensitive axis in the direction of motion and its output signal f is coupled to the torquer (T) of the gyro via the integrator and the amplifier - 1/R. V is the north velocity and ε the angular deviation with respect to the vertical.

The system's error dynamics are described by the linearized state space equations:

$$(4.2.3) \quad \begin{bmatrix} \delta\dot{S} \\ \delta\dot{V} \\ \dot{\varepsilon} \\ \dot{D} \\ \dot{B} \end{bmatrix} = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & g & 0 & 1 \\ 0 & -1/R & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} \delta S \\ \delta V \\ \varepsilon \\ D \\ B \end{bmatrix} + \begin{bmatrix} w_{\delta V} \\ w_B \\ w_D \\ w_{\dot{D}} \\ w_{\dot{B}} \end{bmatrix}$$

with D = gyro and B = accelerometer bias assumed to be constant in the mean. The state vector is:

$$(4.2.4) \quad x = (\delta S \quad \delta V \quad \varepsilon \quad D \quad B)^T$$

The w-terms comprise random Gaussian distributed noise with zero mean. Instead of $w_{\dot{\varepsilon}}$ and $w_{\delta \dot{V}}$ we have introduced w_D and w_B to remind of the fact that these terms are caused by gyro and accelerometer noise.

Equation 4.2.3 is of the following form:

$$(4.2.5) \quad \dot{x} = F \cdot x + w.$$

Its solution has the general form:

$$(4.2.6) \quad x(t_n) = \Phi(t_n, t_{n-1}) \cdot x(t_{n-1}) + \int_{t_{n-1}}^{t_n} \Phi(t_n, \tau) \cdot w(\tau) \, d\tau,$$

with $\Phi(t_n, t_{n-1})$ = state transition matrix. For constant coefficients of the system matrix F the transition matrix is in the Laplace domain:

$$(4.2.7) \quad \Phi(s) = (s I - F)^{-1},$$

with s = Laplace variable. For zero random driving functions (w = 0) this leads to the following solution of Eq. (4.2.3) in the time domain:

$$(4.2.8) \quad \begin{bmatrix} \delta S \\ \delta V \\ \varepsilon \\ D \\ B \end{bmatrix}(t) = \begin{bmatrix} 1 & si/\omega & R(1-co) & R(\Delta t - si/\omega) & (1-co)/\omega^2 \\ 0 & co & R\,\omega\,si & R(1-co) & si/\omega \\ 0 & -si/(R\,\omega) & co & si/\omega & -(1-co)/g \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \delta S \\ \delta V \\ \varepsilon \\ D \\ B \end{bmatrix}(t_0).$$

We have used here the following abbrevations:

(4.2.9 a,b,c) $\Delta t = t - t_0$,   si = sin $\omega^s \, \Delta t$,   co = cos $\omega^s \, \Delta t$.

Horizontal platform misalignment ε, velocity error δV and position error $\delta S = R \cdot \delta\phi$ due to east-west gyro drift D and north-south accelerometer bias B are plotted in Fig. 4.2.2. All position errors begin with zero slope, indicating that for short periods of time the NIS is very accurate.

Distinctly to be seen is the Schuler oscillation on the 3 system levels. All NIS errors are bounded, except for the effect of gyro drift D on the position error. It confirms the importance of the gyro on the long-term system accuracy. The slope, i.e. the mean velocity error provides the **"Rule of Thumb for Inertial Navigation"**:

(4.2.10) $\delta\overline{V} = R \cdot D \Rightarrow$ 1 km/h navigational error per D = .01 deg/h

Physically this rule becomes understandable if we consider that the gyros keep the reference coordinate system for navigation, stored in the platform and that the gyro drift causes an analogous drifting of it (as a reminder: 1 degree of longitude = 111 km at the equator).

It is interesting to note that the effect of accelerometer bias B upon the NIS position error is not as severe as one could expect. The misalignment error and the position error are alike. The mean error is limited to:

(4.2.11) $\overline{\varepsilon} = \delta\overline{S}/R = \delta\overline{\phi} = B/g.$
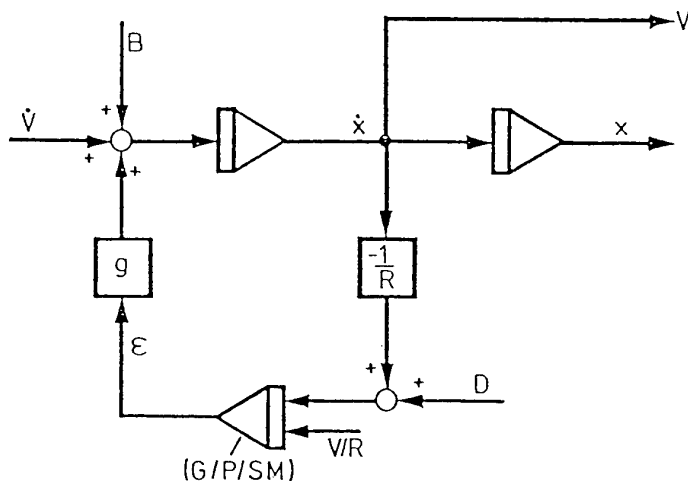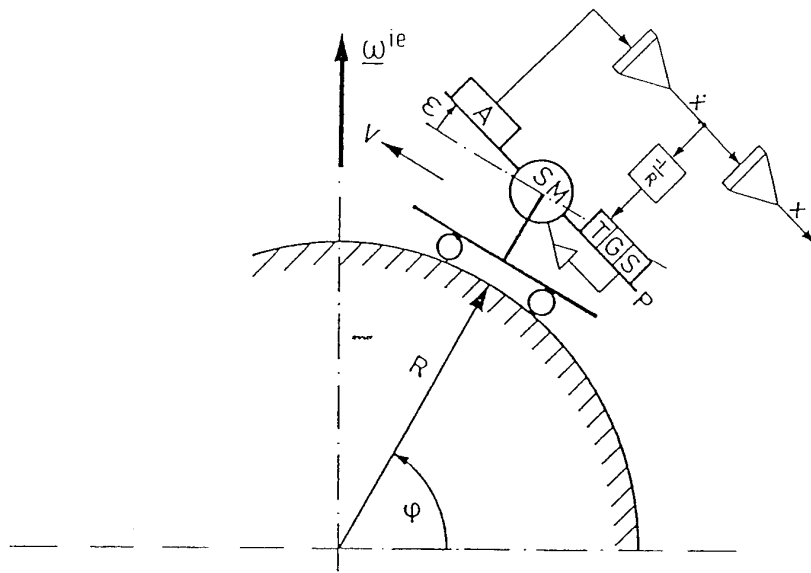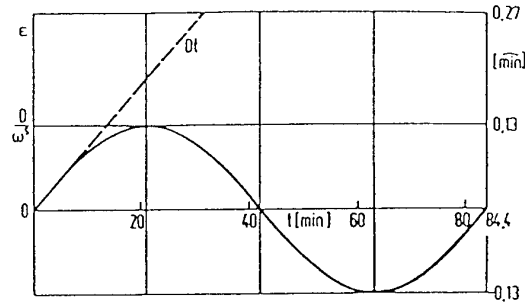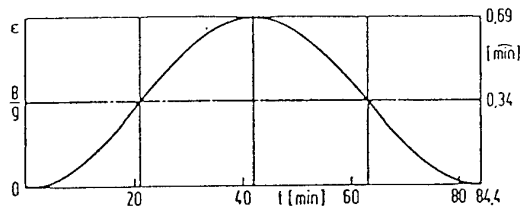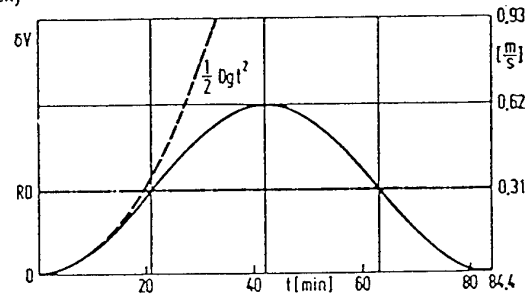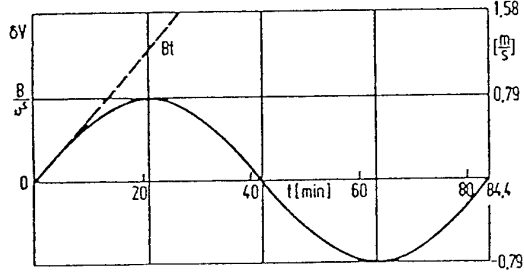
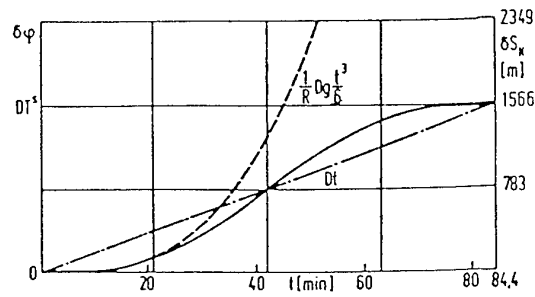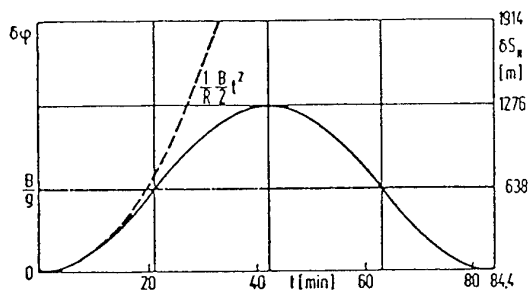Fig 4.2.1 Signal Flow Diagram and Block Diagram for the Single Axis Inertial Navigation System

Effect of

Accelerometer Bias $(B=10^{-4}g)$ Levelling Gyro Drift $(D=0.01°/h)$
on Attitude

ϵ [$\widehat{min}$]

0.69

$\frac{B}{g}$ 0.34

0   20   40   t [min]   60   80 84.4

ϵ ϵ_∞ 0

Dt

0   20   40   t [min]   60   80 84.4

0.27

[$\widehat{min}$]

0.13

-0.13

on Velocity

δV $\frac{B}{\omega_s}$ 0

Bt

0   20   40   t [min]   60   80 84.4

1.58

[$\frac{m}{s}$]

0.79

-0.79

δV RD   0

$\frac{1}{2}$ Dgt²

0   20   40   t [min]   60   80 84.4

0.93

[$\frac{m}{s}$]

0.62

0.31

on Position

$\delta\phi$ $\frac{B}{g}$ 0

$\frac{1}{R}\frac{B}{2}t^2$

0   20   40   t [min]   60   80 84.4

1914

δS_x
[m]

1276

638

$\delta\phi$ DT²   0

$\frac{1}{R}Dg\frac{t^3}{6}$

Dt

0   20   40   t [min]   60   80 84.4

2349

δS_x
[m]

1566

783

——— closed Schuler - loop

------ open Schuler - loop or navigation on plane earth with $\delta S_x = R\delta\phi$

—·— mean value

Fig. 4.2.2 Errors of a Single Axis Inertial Navigation System

This means that the accelerometer bias is compensated by a component of gravity due to the mean platform tilt.

For periods that are short with respect to the Schuler period, one can introduce the approximations in Eq. (4.2.8):

(4.2.12a,b) $si = \sin \omega^s \Delta t \cong \omega^s \Delta t - (\omega^s \Delta t)^3/6$
$co = \cos \omega^s \Delta t \cong 1 - (\omega^s \Delta t)^2/2.$

This leads to the curves marked in dotted lines in Fig. 4.2.2, which for a longer time period are also valid for the set-up of Fig. 4.2.1 without Schuler loop or for inertial navigation on a flat earth.

Another effect is worth mentioning, the effect of sensor scalefactor errors on the NIS performance. Let us begin with the gyro scalefactor error K^g (s. Eq. 2.22). It causes an error in the platform slewing $\delta\omega^t = K^g(\omega^{ie} + \omega^{en})$ and thus has a similar effect as the gyro drift. It is bigger for easterly flights when ω^{ie} and ω^{en} have the same sign than for westerly flights. The corresponding drift component about the north-south axis is then:

(4.2.13) $d' = \kappa^g \omega^t \cong .002$ deg/h   for $\kappa^g \cong 100$ ppm and V = 1000 km/h at the equator.

The assessment of the accelerometer scalefactor error's (κ^a) effect on system performance yields again a surprising result. It is active only when the vehicle is accelerating, but not during cruise. Assuming that the take-off acceleration is very short compared to the Schuler period, we may regard the velocity error $\delta V = \kappa^a \cdot V$ as an initial velocity error $\delta V(t_0)$. From Eq. 4.2.15 we obtain as position error:

(4.2.14a) $\delta S(t) = \kappa^a (V/\omega^s) \sin \omega^s t,$

i.e. it is zero in the mean! Only for short intervals of time compared to the Schuler period of 84 min, say 10 min, or for inertial navigation on a flat earth the position error is as expected:

(4.2.14b) $\delta S \cong \kappa^a \cdot S.$

In summary we may conclude that inertial navigation is only feasable because our Lord has made the earth spherical. The time exponent of the position error growth is lower by 2 as compared to our expectations from school physics!

It is not surprising that this was not recognized at the beginning of the inertial navigation development which took place in Germany before and during World War II. Outsiders did not believe in the feasability of inertial navigation due to their interpretation of the position error growth with school physics as a basis. The so-called "law of the third power" to which they referred says that the position error due to gyro drift grows $\sim t^3$. It was Reisch who proclaimed inertial navigation as the "principle of plumb line rotation" shown in Fig. 4.2.3 [He 80]. It indicates that position changes can be measured by the inclination of a space-stabilized platform with respect to a plumb line reference. Gyro drift causes a position error growing linearly with time in this navigation system which lacks the plumb line reference on a moving vehicle. The INS in Fig. 4.2.1 has the plumb line reference in form of the Schuler feedback. From today's point of view this was Reisch's valuable contribution to this development, but one cannot follow him in his demand to be the inventor of inertial navigation as proclaimed again in [Hi 92]. This honour is certainly due to Boykow [Bo 35]. The fairly benign INS error characteristics do **not** depend on the mechanization proposed by Reisch and before him by Boykow. It governs any INS mechanization as shown in Sections 4.4 and 4.6 and is due to the shape of the earth.
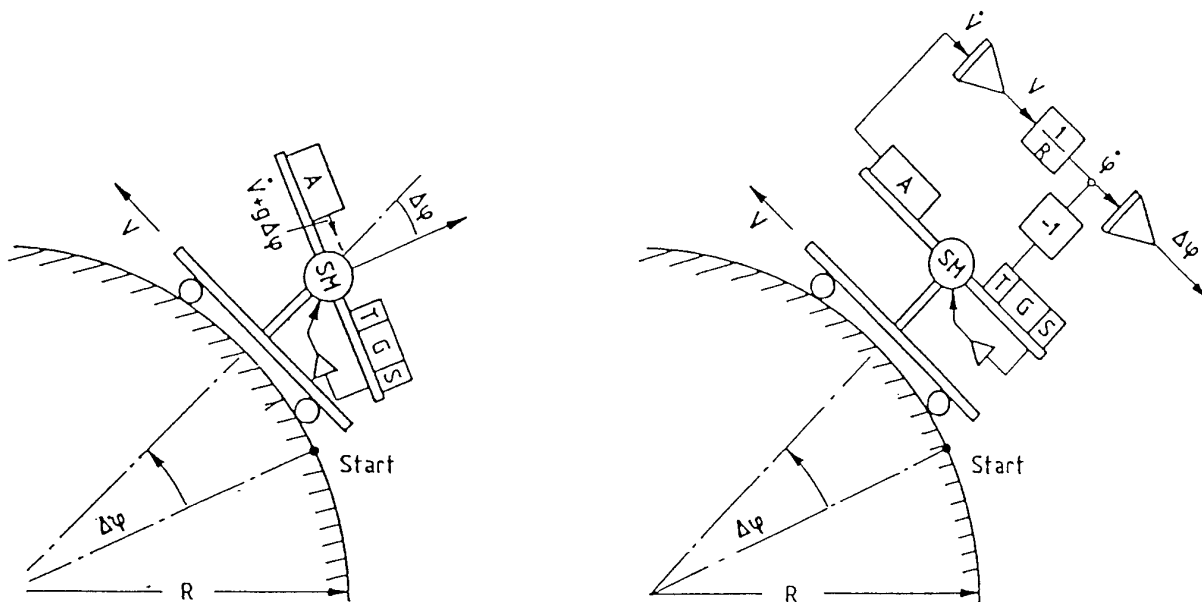

Figure 4.2.3 Reisch's "Principle of Plumb Line Rotation" for Position Measurement from the Inclination Measurement of a Space Stabilized Platform

So far we have discussed the effect of deterministic sensor errors on the NIS performance only. In modern sensors, especially in the ring laser gyro (RLG), the deterministic errors are fairly small and the effect of stochastic errors on system performance becomes more dominant (s. Appendix O). It is also of great importance for system modelling in a Kalman filter, and this will be the subject of our next discussion.. Tools for estimating their effect are the Monte Carlo simulation on a digital computer and the covariance analysis for predicting their $\pm$ 1 σ boundaries. The results of the former, shown in Fig. 4.2.4, are based on a simplified version of Eq 4.2.6:

$$(4.2.15) \qquad x(t_n) = \Phi(t_n, t_{n-1}) \cdot x(t_{n-1}) + w(t_{n-1}),$$

with the following state transition matrix for a first order integration:

$$(4.2.16) \qquad \Phi(t_n, t_{n-1}) \cong I + F(t_{n-1}) \cdot \Delta t.$$

In this simulation gyro noise was investigated with a random walk coefficient of $r_D = .01$ deg$/\sqrt{h}$.

From the different sample functions plotted in Fig. 4.2.4 the RMS values were calculated for a fixed time and plotted as dotted $\pm 1\sigma$-curves. In order to compare the simulation with the theory, we carry out a covariance analysis using the formular [Ge 74]:

$$(4.2.17) \qquad P(t_n) = \underset{\uparrow}{\Phi(t_n, t_{n-1}) \cdot P(t_{n-1}) \cdot \Phi^T(t_n, t_{n-1})} + \underset{\uparrow}{\int_{t_{n-1}}^{t_n} \Phi(t_n, \tau) \cdot q \cdot \Phi^T(t_n, \tau) \, d\tau},$$

where | deterministic propagation of covariance, i.e. propagation of uncertainty in knowledge of system error state between t_{n-1} and t_n. | added uncertainty due to integration of sensor and system noise.

The covariance matrix q specifying the sensor and system noise, has on its main diagonal the squares of the corresponding random walk coefficients and otherwise zeros. For noise the corresponding elements read on the accelerometer and gyro level (in Eq. 4.2.3 $w_B \neq 0$ and $w_D \neq 0$):

$$(4.2.18a,b) \qquad q_{22} = (r_B/60)^2 \qquad \text{and} \qquad q_{33} = (r_D/60)^2$$

with r_B [(m/s)/$\sqrt{h}$] and r_D [deg/$\sqrt{h}$].

The 1σ-values for the system uncertainty is the square root of the corresponding element on the main diagonal of the resulting covariance matrix, for instance:

$$(4.2.19) \qquad \sigma_{\delta s} = \sqrt{P_{11}}.$$

Using the state transition matrix from Eq. 4.2.8, we obtain the following result for the system uncertainty propagation due to gyro noise, i.e. a random walk on the misalignment level:

$$(4.2.20a,b,c) \qquad \sigma^2_\varepsilon = r^2_D \cdot [t/2 + 1/(4\omega_s) \cdot si2],$$
$$\sigma^2_{\delta v} = r^2_D R g \cdot [t/2 - 1/(4\omega_s) \cdot si2],$$
$$\sigma^2_{\delta s} = r^2_D R^2 \cdot [3 \cdot t/2 - 2/\omega_s \cdot si + 1/(4\omega_s) \cdot si2],$$

with the abbrevations:

$$(4.2.21a,b) \qquad si = \sin(\omega_s t) \quad \text{and} \quad si2 = \sin(2\omega_s t).$$

The results of the computed ± 1 σ values are plotted in Fig. 4.2.4 as dot-dash curves. They correspond fairly well with the simulation if it is kept in mind that the $\pm 1\sigma$ bands enclose 68% of the 8 Monte Carlo runs.

In a Kalman filter Eqs. 4.2.15 and 4.2.16 are used for the prediction of the system error state based on the last optimal estimate ($x(t_{n-1}) = \hat{x}(t_{n-1})$). Since the random driving function $w(t)$ cannot be predicted but its mean is zero per assumption, $w(t_{n-1})$ has to be assumed to be zero for the prediction.

4.2.2.2 The Error Dynamics of a Three-Axis Inertial Navigation System for a Short Period of Time (< 6 hours)

Fig. 4.2.5 shows in simplified form the block diagrams of the error models for a three-axis NIS in the lower half and the barometric aided altitude channel in the upper half. On the left side the sensor errors are listed and on the right side the system output errors.

As compared to the complete error model of an NIS [Br 71], which will be discussed in Section 4.2.2.3 (s. Fig. 4.2.7), the so-called 24-hour oscillation caused by the coupling of the platform angles N,E,D via the components of the earth rate plus the transport rate $\dot{\Lambda}$ (with $\dot{\Lambda} = \Omega + \dot{\lambda}$) or $\dot{\phi}$, respectively, has been cut down to the coupling of the azimuth misalignment ε_D into the rotation about the east-west axis. For the purpose of discussion this is justified because on the one hand ε_D is in general one order of magnitude higher than the horizontal misalignment angles $\varepsilon_{N,E}$. On the other hand, for flight periods of up to 2 hours we can assume with good accuracy: $\sin \dot{\Lambda} t \cong \dot{\Lambda} t$ and $\cos \dot{\Lambda} t \cong 1$ (error in this approximation < 10 %).

For an estimation of the errors in the NIS channels we assume that the slewing of the vertical gyro as well as the compensation of the Coriolis acceleration can occur without errors and there remain two Schuler loops as in Fig. 4.2.1. In addition to these Schuler loops the influences of the azimuth error ε_D as well as of the vertical gyro drift are indicated.

On the one hand the azimuth deviation ε_D causes acceleration errors in the horizontal channels which are orthogonal to the respective components $\dot{V}_{N,E}$ of the horizontal acceleration of the vehicle. This is indicated in Fig. 4.2.6a for an acceleration of $\dot{V}_N$.
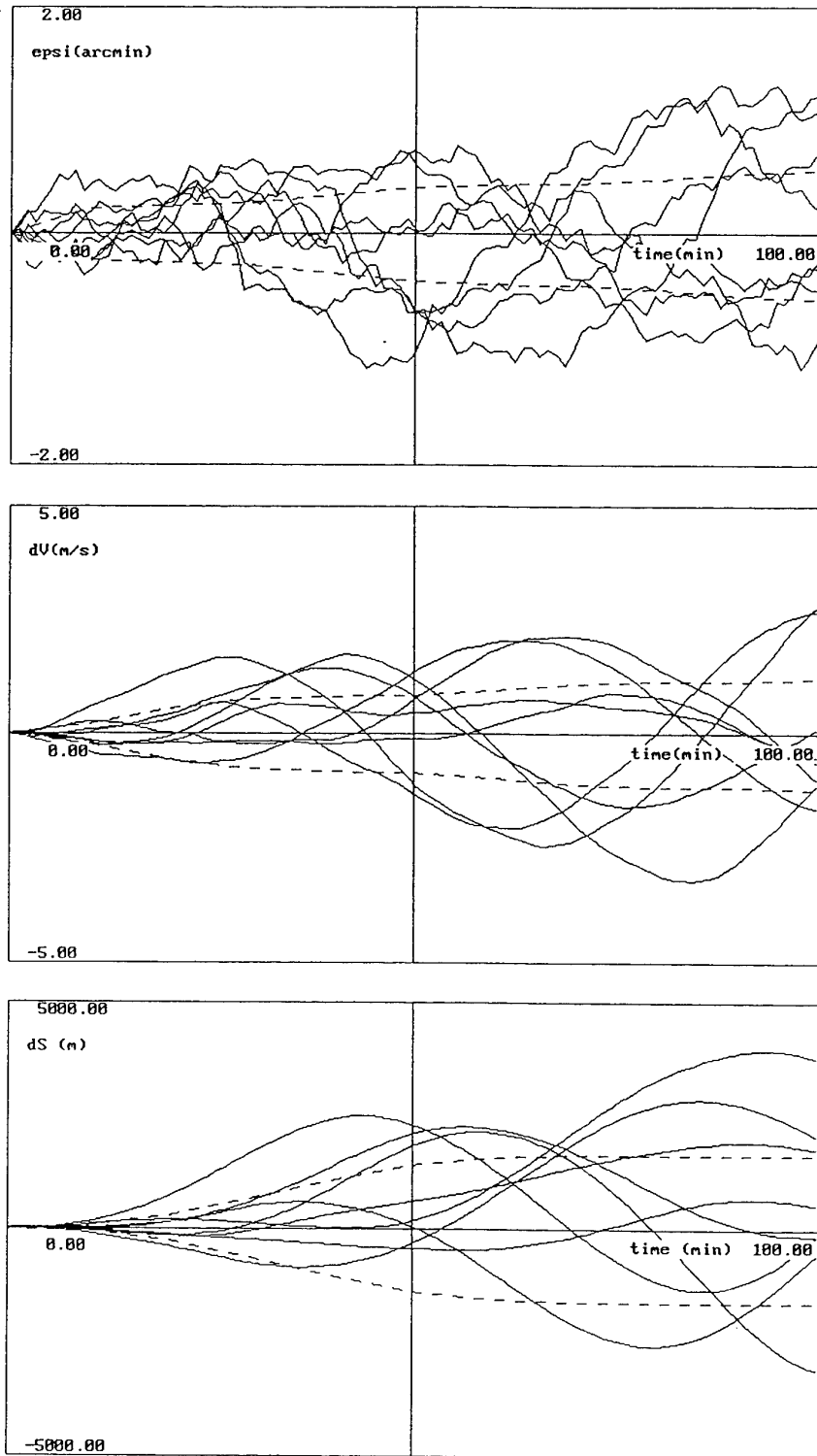
Figure 4.2.4 Errors and ±1σ Bands of a Single Axis Inertial Navigation System
Due to a Stochastic Gyro Drift (Random Walk of .01 deg/√h)
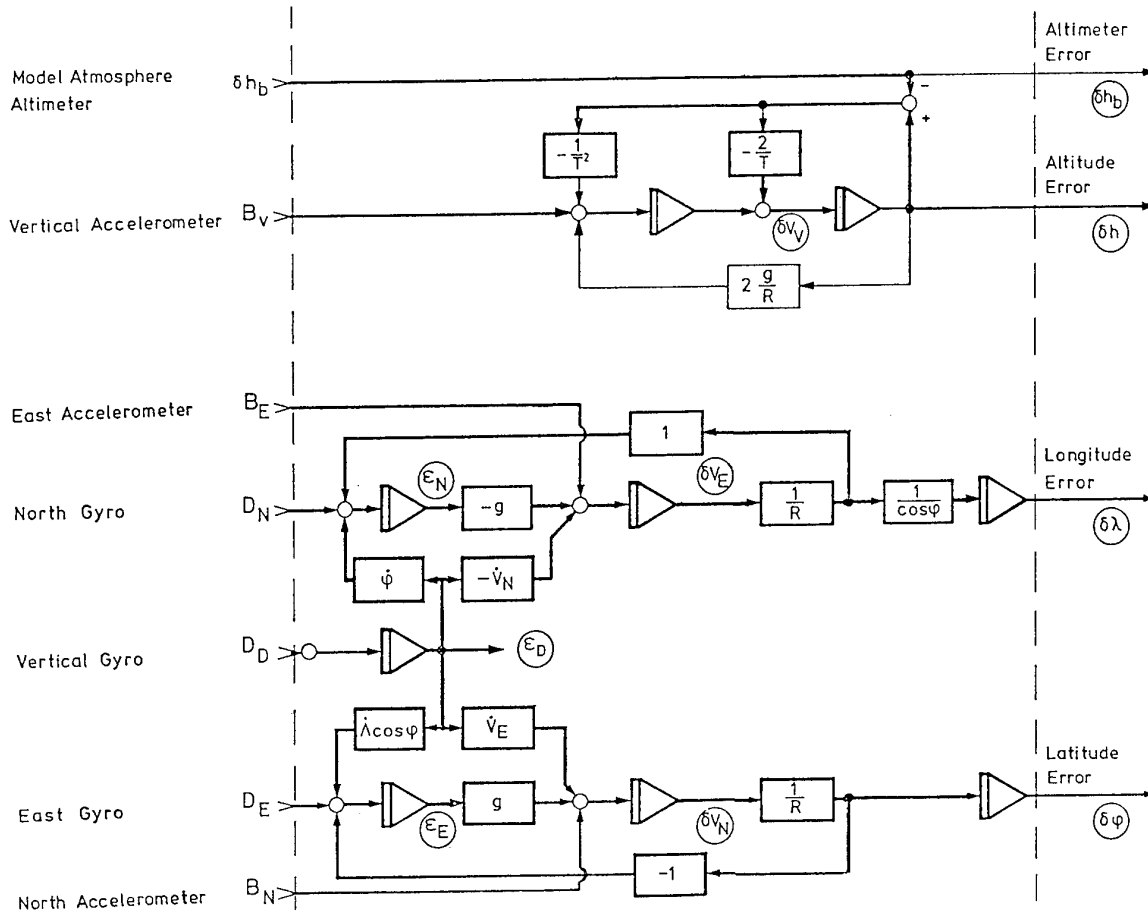
Sensor Errors

System Error Model

Figure 4.2.5 Simplified Block Diagram for the 3-Axis Error Model of a North Indicating INS Including Vertical Channel
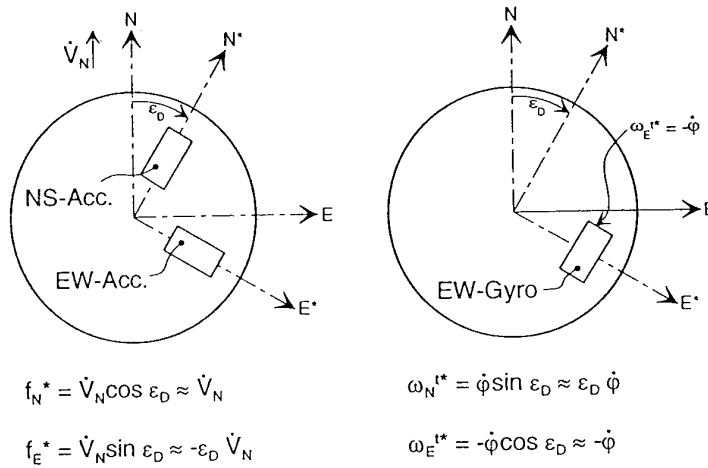

$$f_N^* = \dot{V}_N \cos \varepsilon_D \approx \dot{V}_N$$

$$f_E^* = \dot{V}_N \sin \varepsilon_D \approx -\varepsilon_D \dot{V}_N$$

$$\omega_N^{t*} = \dot{\varphi} \sin \varepsilon_D \approx \varepsilon_D \dot{\varphi}$$

$$\omega_E^{t*} = -\dot{\varphi} \cos \varepsilon_D \approx -\dot{\varphi}$$

Figure 4.2.6 View upon the Platform Misaligned about the Vertical Axis

Assuming that the acceleration of the vehicle from start to the constant cruise velocity is short compared to the Schuler period, the time integral of this cross-track acceleration error is equal to an initial velocity error:

$$(4.2.23a,b) \quad \delta V_N' = \delta V_N(0) + \varepsilon_D(0) \cdot V_E, \qquad \delta V_E'(0) = \delta V(0) - \varepsilon_D(0) \cdot V_N.$$

On the other hand, the azimuth error ε_D has an effect comparable to a north-south or east-west gyro drift through the coupling of the rates $\dot{\phi}$ and $\dot{\Lambda} \cdot \cos \phi$ into the respective orthogonal platform axes:

$$(4.2.24a,b) \quad D_N' = D_N - (\varepsilon_D + D_D \cdot t) \cdot \dot{\phi}, \qquad D_E' = D_E + (\varepsilon_D + D_D \cdot t) \cdot \dot{\Lambda} \cdot \cos \phi$$

The relationships compiled in Eq. 4.2.25 were obtained in a way similar to the one for the single axis inertial navigation system in Section 4.2.1; the state space equations of the two uncoupled horizontal channels were supplemented by Eqs. 4.2.23 and 4.2.24 and $\cos\phi$ was assumed to be constant.

(4.2.25)

$$\left[\begin{array}{ccccccc} \delta\phi(t) & \delta\lambda(t)\cdot\cos\ \phi & \delta V_N(t) & \delta V_E)t) & \varepsilon_N(t) & \varepsilon_E(t) & \varepsilon_D(t) \end{array} \right]^T =$$

$$\begin{bmatrix}
1 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 & 0 & 0 & 0 \\
si/(R\omega) & 0 & co & 0 & 0 & -si/(R\omega) & 0 \\
0 & si/(R\omega) & 0 & co & si/(R\omega) & 0 & 0 \\
0 & -(1-co) & 0 & -R\omega\ si & co & 0 & 0 \\
(1-co) & 0 & R\omega\ si & 0 & 0 & co & 0 \\
(\Delta\Lambda - \Omega/\omega\ si)\cos\phi & -\Delta\phi & R(\dot{\Lambda}-\Omega\ co)\cos\phi & -R\dot{\phi} & 0 & (\Omega/\omega)\cos\phi\ si & 1 \\
0 & -t+si/\omega & 0 & -R(1-co) & si/\omega & 0 & 0 \\
t-si/\omega & 0 & R(1-co) & 0 & 0 & si/\omega & 0 \\
\Delta\Lambda\cos\phi\ t/2 - (\dot{\Lambda}/\omega^2)\cos\phi(1-co) & \Delta\phi\ t/2 + (\dot{\phi}/\omega^2)(1-co) & R\dot{\Lambda}\cos\phi\cdot(t-si/\omega) & -R\dot{\phi}\cdot(t-si/\omega) & -(\dot{\phi}/\omega^2)\cdot(1-co) & (\dot{\Lambda}\cos\phi/\omega^2)\cdot(1-co) & t \\
(1-co)/g & 0 & si/\omega & 0 & 0 & -(1-co)/g & 0 \\
0 & -(1-co)/g & 0 & si/\omega & (1-co)/g & 0 & 0
\end{bmatrix}^T \times \begin{bmatrix} \delta\phi(0) \\ \delta\lambda(0)\cdot\cos\ \phi \\ \delta V_N(0) \\ \delta V_E(0) \\ \varepsilon_N(0) \\ \varepsilon_E(0) \\ \varepsilon_D(0) \\ D_N \\ D_E \\ D_D \\ B_N \\ B_E \end{bmatrix}$$

We have again made use of the abbrevations in Eq. 4.2.9, i.e. $\omega = \omega^s$, $si = \sin \omega^s t$ and $co = \cos \omega^s t$. The reader's attention is also drawn to the fact that the vector-matrix arrangement in this equation alleviates the identification of state vector dependencies.

Of the results we want to discuss only the position error equations and cast a glance at the main sources. The initial position error ($\delta\phi(0)$ and $\delta\lambda(0)$) can be assumed to be zero, as on any major aerodrome the geographic longitude and latitude are indicated at the aircraft parking positions. The initial velocity error can also be neglected if the switching of the NIS into the navigational phase takes place on a stationary aircraft, V(0) is then accurately known.

In comparison to Eq. (4.2.8) essentially only the effects of the azimuth alignment error ε_D and of the vertical gyro drift D_D have been added. It is interesting to note that the deviation with time of the "cross-track-error" is equal to that of a common dead reckoning navigational system in the east-west channel ($\delta\lambda(t) \cos \phi = - \varepsilon_D(o) \cdot \Delta\phi$). In the north-south channel this is true for short times only when $\delta\phi(t) = \varepsilon_D(0) \cdot [\Delta\Lambda - (\Omega/\omega_s) \cdot \sin \omega_s\Delta t] \cdot \cos \phi \cong \varepsilon_D \cdot \Delta\lambda \cdot \cos \phi$. After approximately 10 minutes, earth rate coupling becomes effective.

It is this difference in the error characteristics that allows in-flight gyrocompassing of a misaligned NIS based on a Doppler radar for measuring the body-fixed ground speed components (s. Chapter "Doppler Navigation") and the misaligned platform as heading reference. We will come back to this topic in Section 6.

If we start from the assumption that before the start the platform has aligned itself in the north direction and in the horizontal plane as a result of a gyrocompassing procedure, we shall under ideal conditions have the following relationships between the alignment and sensor errors (s. Section 5):

$$(4.2.26a,b) \quad \varepsilon_{N,E}(0) \cong \pm B_{E,N}/g \qquad (\cong .1 \ \text{mrad} \cong 20 \ \text{arc sec})$$

$$\varepsilon_D(0) \cong - D_E/(\Omega \cos \phi) \qquad (\cong 1 \ \text{mrad} \cong 3.4 \ \text{arc min}),$$

the numbers being valid for $B = 10^{-4}$ g and $D = .01$ deg/h as well as $\phi = 45$ deg. The position errors in Eqs. (4.2.25) are reduced in this case to:

$$(4.2.27a,b) \quad \delta\phi(t) \cong \varepsilon_D(0) \cdot S_E/R - D_D \cdot [\Delta\Lambda \cdot t/2 - (\dot{\Lambda}/\omega^{s2}) \cos \phi \cdot (1 - \cos \omega_s^s \cdot t)]$$

$$\delta\lambda(t) \cong -\varepsilon_D(0) \cdot S_N/R - D_N \cdot [t - \sin \omega^s \ t/\omega^s],$$

$S_{N,E}$ being the distance flown. For periods that are short in comparison with the Schuler period (6 min approximately for a 10 % error) these relationships will be reduced to:

$$(4.2.28a,b) \qquad \delta\phi(t) \quad \cong \quad \varepsilon_D(0) \cdot S_E/R \qquad\qquad \delta\lambda(t) \cdot \cos \phi \quad \cong \quad - \varepsilon_D(0) \cdot S_N/R \ ,$$

which means, that after fully completing self-alignment, the navigation error of the NIS under ideal conditions initially consists exclusively of the cross-track error, and is identical with that of a normal dead reckoning navigational system with ideal velocimeters!

Before we discuss in Section 4.6 the NIS horizontal errors over a longer period of time (> 6h), the INS vertical channel and other INS mechanizations should be described briefly. Then we are in a better position to understand common features and differences in their error models as a basis for Kalman filter design.

4.2.3 The Error Model for the Vertical Channel of an Inertial Navigation System

The decoupling of the vertical from the horizontal channels is justified, as will be shown in Section 4.6. Inertial navigation in the vertical channel consists of a mere double integration with gravity compensation. The latter includes the decrease of gravity as a function of height due to the physical law of mass attraction between two bodies - the earth and the accelerometer's proof mass. The block diagram for the vertical channel's error model is shown in Fig. 4.2.5, top, in thin lines. Due to the positive feedback of the gravity dependence upon height the autonomous inertial navigation in the vertical is unstable with the eigenvalues:

$$(4.2.29) \qquad \lambda_{1,2} = \pm \sqrt{2\omega^s} = \pm \sqrt{2g/R} \ .$$

They describe the following error growth:

$$(4.2.30) \qquad \delta h(t) = \delta h(0) \cdot \cosh \lambda t - B_v/\lambda^2 \cdot (1 - \cosh \lambda t),$$

where B_v is the accelerometer bias. This error growth is plotted in Fig. 4.2.7.

For navigating over a longer period of time, the vertical channel has to be aided by external measurements, as indicated in Figs. 4.2.5 and 4.2.8 for use in a civil aircraft. The difference between the baroinertial height h and the barometric height h^b is fed back to the baroinertial velocity and the acceleration via the gains K_1 and K_2. The shunting integrator with input gain K_3 serves to compensate quasiconstant acceleration errors due to sensor bias and gravity compensation errors. The gains quoted in this figure are valid for the Litton LTN 90. Due to the fairly weak barometric coupling, the baroinertial height follows in the low frequency domain the barometric height and in the high frequency domain the purely inertial height. The most important sideeffect of this baroinertial coupling is the generation of a signal for the vertical velocity. It can hardly be derived from pure barometric measurements.

Within a flying aircraft the static barometric pressure measurement is a function of:

- the magnitude of the air flow, i.e. the true air speed (TAS) and the Mach number,

- the direction of the air flow, i.e. the angles of attack and sideslip,

- and the flaps and throttle settings [Wu 80].

An optimal measurement position for the static pressure at the fuselage is selected during the design phase of the aircraft by means of windtunnel tests and by calibration flights, the results of which are "optimal" for a certain flight condition only. For other flight conditions an error model including calibration parameters is layed down in the "air data computer (ADC)" [NM 84].

The static pressure is thus a fairly accurate altitude reference during cruising speed in the subsonic range. It is true that the aircraft will thus not follow a certain geometric altitude, but the isobars of the atmosphere. This is a sound basis for air traffic control, since the controller knows that all aircraft will fly in their respective flight levels (FL) which he has assigned them to (FL = barometric mean sea level altitude [ft]/1000 above 3000 ft above mean sea level (MSL)).

The calibration of the static pressure for instationary flight conditions and the generation of a reliable reference for the vertical speed and the height is difficult and limited in its accuracy [Au 91]. Out of this reason a separate chapter is devoted to this topic in this book (s. Chapter "Vertical Channel Design Considerations").

For flight-test with accuracy requirements in the 1 m range it has also to be kept in mind that additional errors are due to the conversion of the analog pressure measurement into a digital signal and due to the conversion of this digital signal into a barometric altitude.

The former is affected by the volume of the pressure transducer, for instance, which causes the time lag in the order of magnitude of 1s [Re 85]. The digital data handling within the ADC includes a lowpass filtering with a time constant of the same order of magnitude [NM 84] again. These time lags T affect the barometric computation during ascent and descent with $\delta h \cong T \cdot \dot{h}$.

The latter, i.e. the conversion of the pressure into the barometric altitude, depends upon the ground pressure and temperature as well as the outside temperature.

The ground parameters are requested by the pilot from the controller and entered into the ADC when approaching an airport. Errors in these parameters cause additional barometric altitude bias and scalefactor variations.

If all error parameters mentioned above remained constant during the flight with respect to the runway, they could be modelled as a bias in an estimation process based on external measurements. Unfortunately this is quite often not the case as the flight test results in Fig. 4.2.9 demonstrate. They were obtained with the twin engine turboprop aircraft Do 228 of DLR with static pressure holes in the fuselage. The plotted differences between the barometric altitude obtained from the Nord Micro ADC and the reference obtained from the DLR "Avionic Flight Evaluation System (AFES)", i.e. a flight path based on microwave radar, laser tracker and INS, shows a very problematic time characteristic during the approach. Neither can this be removed by feeding the barometric altitude into the inertial system as indicated in Fig. 4.2.8, nor can this be estimated correctly by a Kalman filter with
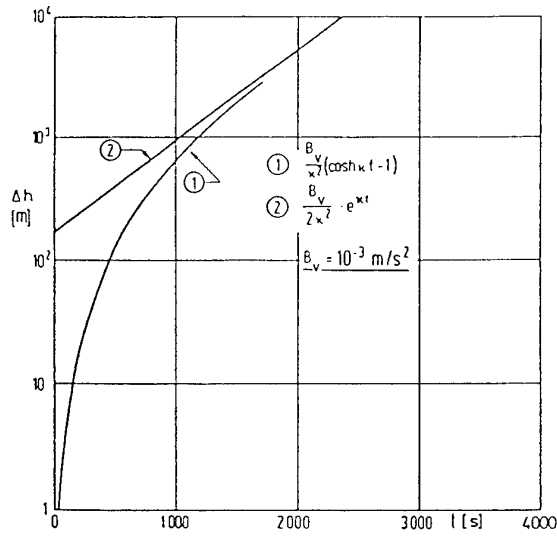
Figure 4.2.7 Unstable Error Growth of the Unaided Inertial Vertical Loop

LTN-90:

ALIGN: $K_1 = 3/20$ NAV: $K_1 = 3/100$
       $K_2 = 3/400$ $K_2 = 3/(100)^2$
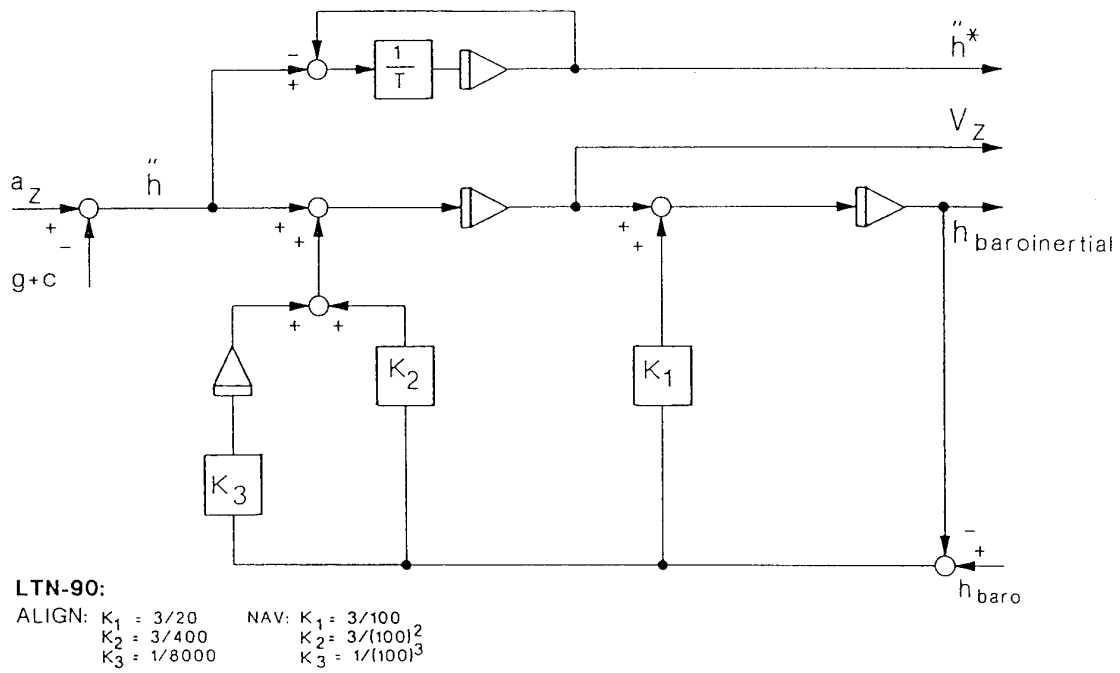 $K_3 = 1/8000$ $K_3 = 1/(100)^3$

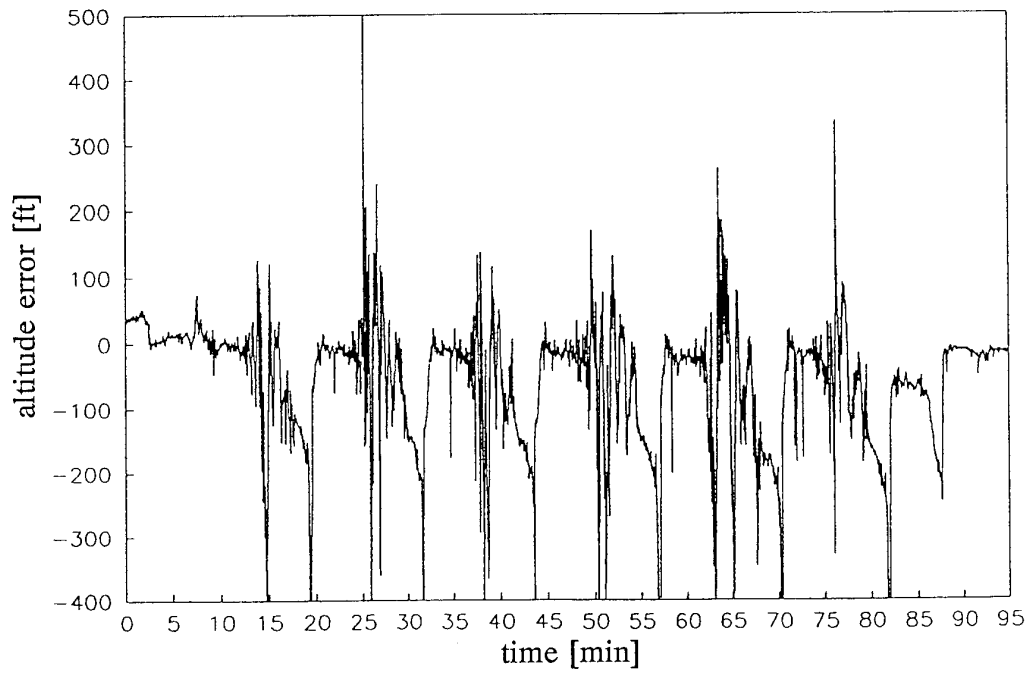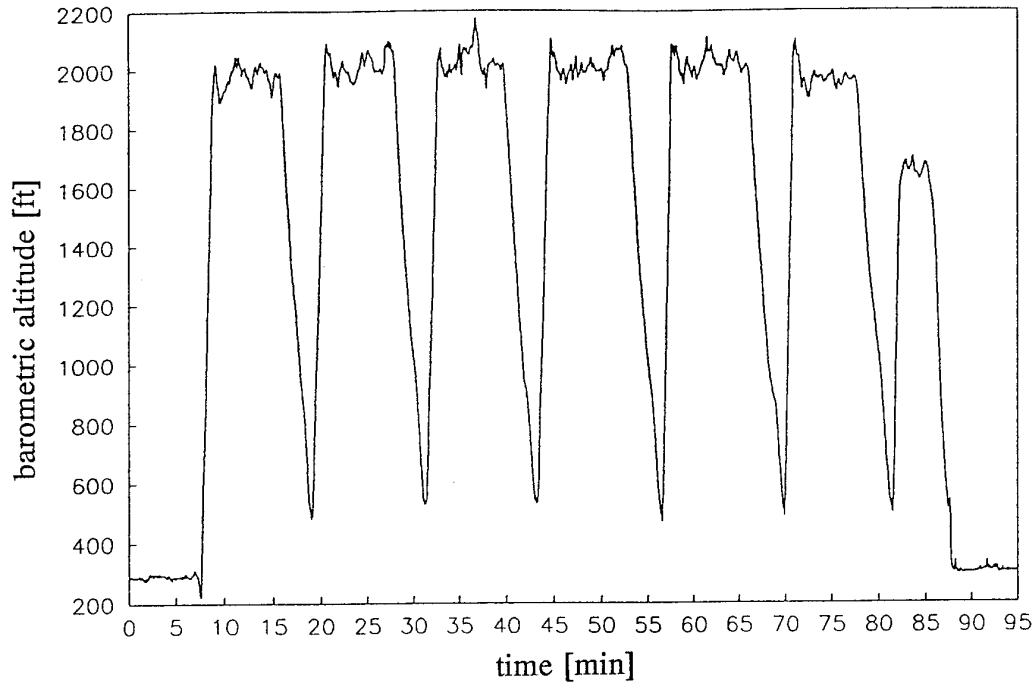Figure 4.2.8 Barometric Aiding of the Inertial Vertical Loop

Figure 4.2.9 Barometric Altitude and Altitude Error of the Twin Engine Aircraft Do 228
During Several Runway Approaches

the bias modelled as a state vector component.

The modelling and calibration of the barometric altitude errors during flight maneuvers is a fairly difficult task and their compensation or estimation accuracy with a Kalman filter is limited [Wi 75]. For obtaining a vertical flightpath reference over a limited period of time, it is advisable to rely on the inertial altitude aided only by external measurements as satellite navigation or the radar altimeter. This topic and problems connected when using a conventional INS for this purpose are covered in [St 94].

4.3 Inertial Navigation with Sensors Aligned to a Wander Azimuth Coordinate Frame

The wander azimuth coordinate frame (subscript a), defined in Appendix C, Section C 4 (s. Fig. C3), is obtained from the navigational frame through the "wander azimuth rotation α" about the vertical axis. Among different choices for the selection of the α [St 82] the most common stems from the fact that the a-frame is not slewed about the vertical axis with transport rate, i.e.:

$$(4.3.1) \qquad \dot{\alpha} = -\omega_D^{en} = [V_E/(R_E + h)] \tan \phi.$$

Prior to navigation the a-frame is aligned to true north, and the vertical in the same fashion as in an NIS (s. Section 5). The a-frame remains in these directions when the system stays stationary or moves exactly northward. When it then passes the north pole, there is not switching about the vertical by 180 deg as in an NIS, and its x-axis is pointing south. When the system moves eastward, the wander angle α increases; it decreases again, when it moves westward. Arriving at the initial meridian, the x-axis of the a-frame is pointing exactly north again. Only after a passage of the pole and returning to the starting point on a different way the x-axis of the a-frame is pointing in the opposite direction.

The computation of the position is based on the integration of a transformation matrix (s. App. C, Eq. C 17):

$$(4.3.2) \qquad C_{ea}(t) = C_{ea}(0) + \int_0^t C_{ea}(\tau) \cdot \Omega_a^{ea}(\tau) \; d\tau = \begin{bmatrix} C_{11} & C_{12} & C_{13} \\ C_{21} & C_{22} & C_{23} \\ C_{31} & C_{32} & C_{33} \end{bmatrix} ,$$

where

$$(4.3.3) \qquad \Omega_a^{ea}(t) = \begin{bmatrix} 0 & -\omega_z & \omega_y \\ \omega_z & 0 & -\omega_x \\ -\omega_y & \omega_x & 0 \end{bmatrix}_a^{ea} .$$

with the components of ω_a^{ea} from Appendix C, Eq. C 15.

The matrix C_{ea} can also be interpreted as being obtained from the rotations of the wander azimuth coordinate frame with respect to the earth-fixed frame via the angles λ, ϕ and α (s. Appendix C, Eq. C17).

Equating the elements of (4.3.2 and C17), the geographic position and true north can be computed using:

$$(4.3.4a \text{ to } c) \quad \sin\phi = -C_{13} , \qquad \tan\lambda = -C_{23}/C_{33} , \qquad \tan\alpha = -C_{12}/C_{11}.$$

Although singularities do exist in the computation of longitude λ and the wander angle α at the geographic poles, they do not exist for the computation of the matrix $C_{ea}(t)$, so that inertial navigation is assured on a polar flight and λ as well as α are again correctly computed shortly after passage of a pole.

According to [Ka 69] the integration of Eq. 4.3.6 has to be executed with 26-bit words and the integration of acceleration with 17-bit words, if the corresponding drift is to be kept below 0.001 deg/h.

4.4 Inertial Navigation with Sensors Aligned to an Inertial Coordinate Frame

Inertial navigation systems aligned to a space-stabilized coordinate frame (index "i") are used for terrestrial navigation using platform systems equipped with electrostatically supported gyros (ESG); those gyros remain untorqued in the navigation mode. These systems are also used for space navigation with platform and strapdown systems.

Figs. 4.4.1a,b show two possible configurations for inertial navigation with the acceleration measurement in the i-frame. It is quite obvious that the system in Fig. 4.4.1a navigates with respect to the earth after transforming the acceleration from the i-frame into the n-frame with the help of the transformation matrix in Appendix C, Eq. C 3b. The navigational computation is identical with that of the NIS. The Autonetics ESGM (electrostatically supported gyro monitor) system for marine application works in this fashion.

With the help of the transformation matrix $C_{ia} = C_{in}C_{na}$ (s. Appendix C, Eqs. C 3b and C 17) navigation could also be carried out in the wander azimuth frame, discussed in the previous section.

The configuration in Fig. 4.4.1b is the one for space navigation with strapdown and platform systems. Navigation is carried out within the inertial frame. This kind of navigation seems very simple on first sight due to the mere double integration of:

$$(4.4.1) \qquad \frac{d^2R_i}{dt^2}\bigg|_i = f_i + G_i(R_i).$$

The difficulty lies in the exact modelling of the earth's mass attraction $G_i(R_i)$ for which we can find assumptions in Chapter "Navigation Coordinate Frames".

On first sight again it is difficult to see that, when geostationary, this INS is governed in its interior error dynamics by the Schuler oscillation and the instability of the vertical channel similar to the conventional NIS. Fig. 4.4.2 may serve to explain this. It shows

the spherical earth and the equipotential surfaces of gravitation:

$$(4.4.2) \qquad G_i(R_i) = G_0 \frac{R_0}{R^3} \frac{R_i}{}$$,

with G_0 and R_0 the values at the surface of the earth. When the stationary INS computes a position error into the y-direction, for instance, the resulting computed but not acting horizontal y-component of G tries to pull this error back. The computed G-component due to an x-position error causes the instability discussed in Section 4.2.3. So we obtain as simplified error equation for the stationary INS:

$$(4.4.3) \qquad \begin{bmatrix} \delta \ddot{R}_x \\ \delta \ddot{R}_y \\ \delta \ddot{R}_z \end{bmatrix} = \begin{bmatrix} 2G/R & 0 & 0 \\ 0 & -G/R & 0 \\ 0 & 0 & -G/R \end{bmatrix} \begin{bmatrix} \delta R_x \\ \delta R_y \\ \delta R_z \end{bmatrix} + \begin{bmatrix} 0 \\ \varepsilon_z \cdot G \\ -\varepsilon_y \cdot G \end{bmatrix} + \begin{bmatrix} b_x \\ b_y \\ b_z \end{bmatrix},$$
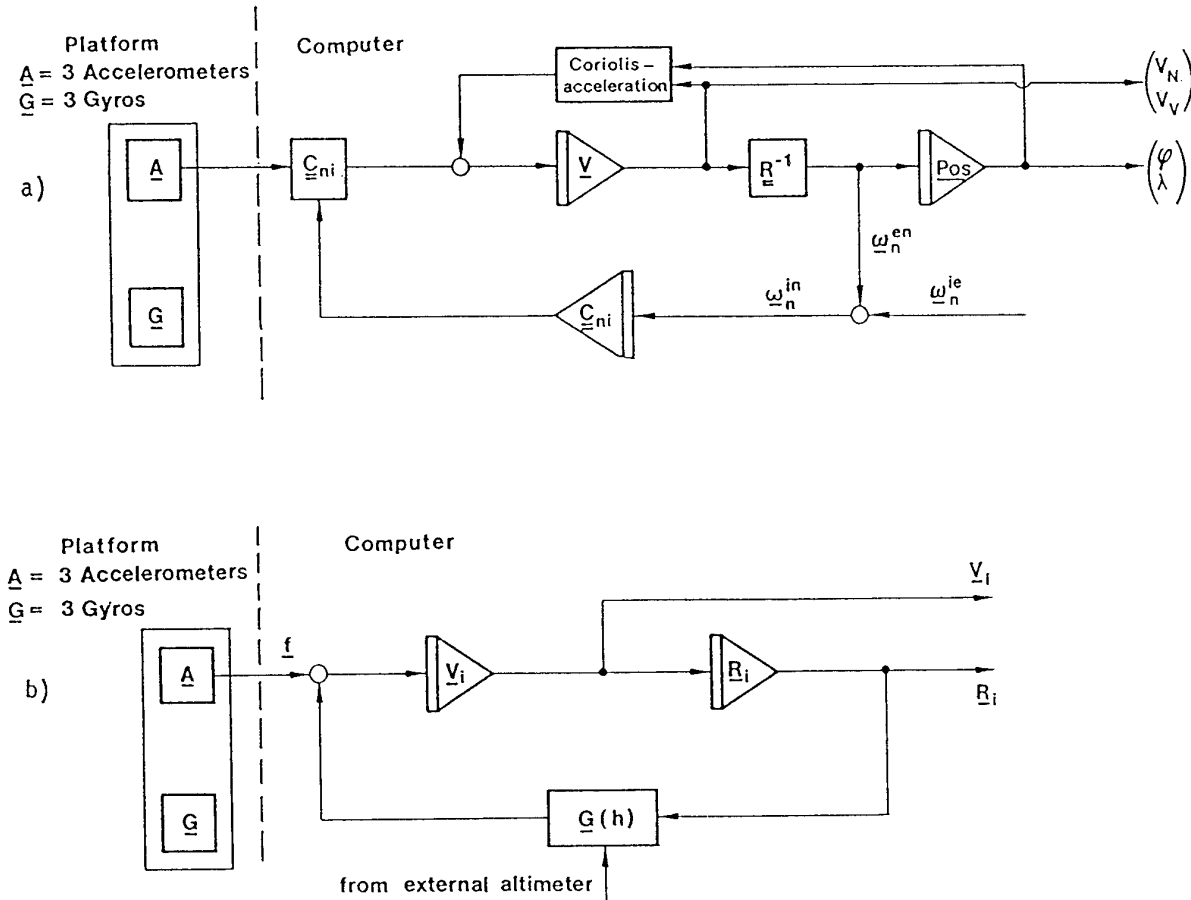
Figure 4.4.1 Two Concepts for Inertial Navigation in the Space-Stabilized Coordinate Frame

where ε is the respective misalignment angle including the effect of gyro drift and b is the respective accelerometer error. This equation shows that the horizontal channels are governed by the Schuler frequency (s. Eq. 4.2.1a) and the vertical by the instability known already from Section 4.2.3 (s. Eqs. 4.2.29 and 4.2.30). Fig. 4.4.2 visualizes these effects.

The Honeywell ESGN (electrostatically suspended gyro navigator) for marine application and the Honeywell GEANS (gimballed ESG aircraft navigation system) - both are space stabiliced platform systems - navigate in the inertial coordinate frame. The position components $R_{x,y,z}$ are transformed to geocentric latitude, longitude and altitude with the help of the transformation matrix C_{ni} in Eq. C 3b of Appendix C.

The next question is connected with the error dynamics of the space-stabilized INS in orbit. The main role of such a system in a satellite is thrust control during orbital changes and attitude determination. For a satellite in orbit the forces due to gravitation G and centrifugal acceleration $\omega^2 R$ are in equilibrium - thus its orbital frequency ω is equal to the Schuler frequency for this orbit $\omega^s = (G/R)^{1/2}_{orbit}$ (s. Eq. 4.2.1). This may give already an indication that the INS error dynamics will be governed by the local
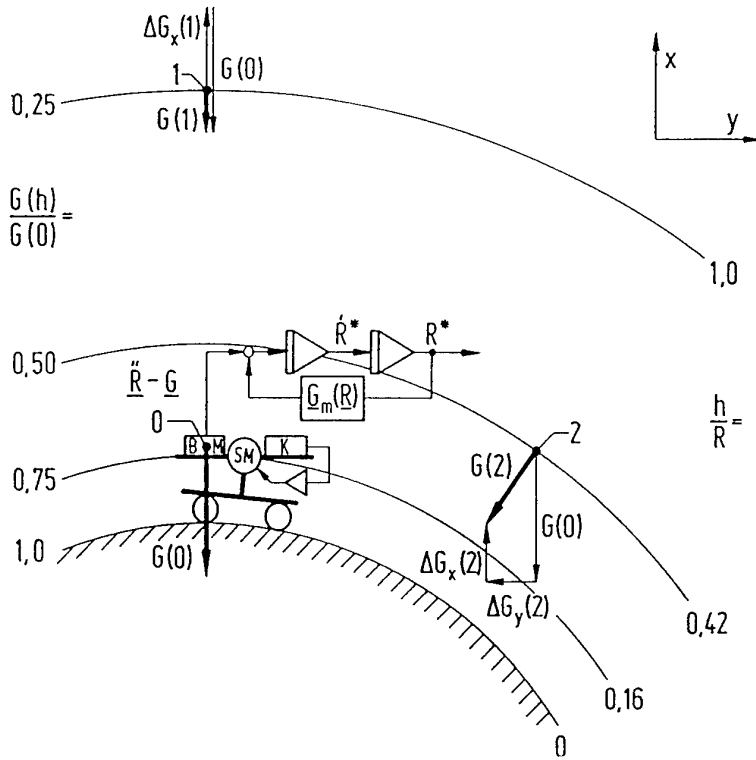
Figure 4.4.2 Physics of the Schuler Feedback and the Vertical Instability in a Space-Stabilized Inertial Navigation System
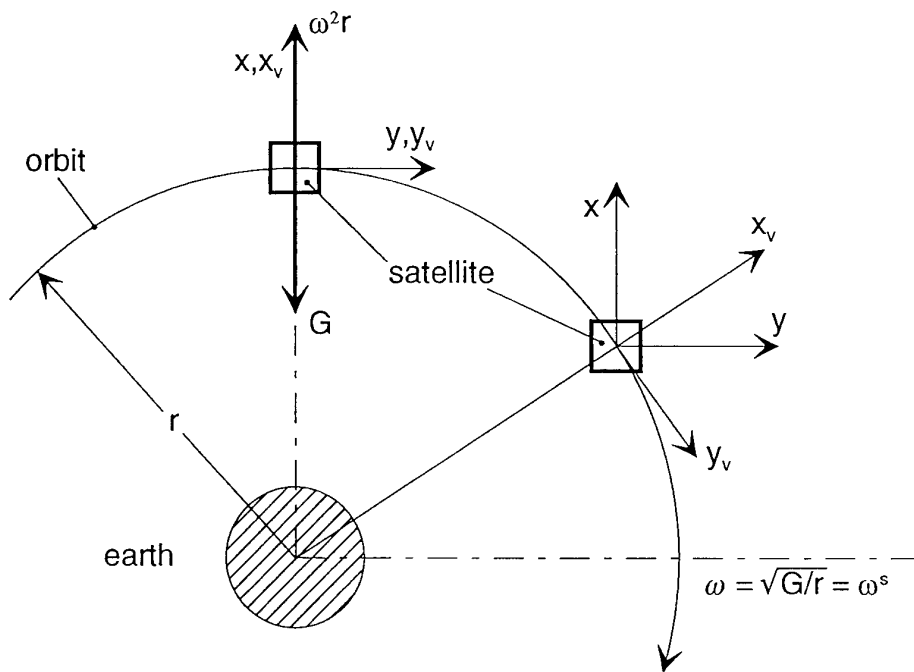

Figure 4.4.3 Physics of the Schuler Oscillation of a Space-Stabilized Inertial Navigation System in Orbit

Schuler frequency ω^s, too, as in all other INS mechanizations. This is indeed the case as may be shown for an INS whose computed coordinate frame is used in the satellite as vertical reference with respect to the earth. The local vertical coordinate frame (index v) has its x-axis pointing up parallel to local vertical, its y-axis pointing tangentially to the orbit righthanded in Fig. 4.4.3, and its z-axis pointing normal to the orbital plane. Making use of the block diagram in Fig. 2.4c for the kinematics of a computed reference direction with respect to the true direction and of the transformation matrix in Eq. C 20 of Appendix C between both, we arrive at the following state space equation of the misalignment angles:

$$(4.4.4) \qquad \begin{bmatrix} \dot{\varepsilon}_x \\ \dot{\varepsilon}_y \end{bmatrix} = \begin{bmatrix} 0 & \omega^s \\ -\omega^s & 0 \end{bmatrix} \begin{bmatrix} \varepsilon_x \\ \varepsilon_y \end{bmatrix}$$

with the following solution for an initial misalignment angle $\varepsilon_x(0)$:

$$(4.4.5a,b) \qquad \varepsilon_x(t) = \varepsilon_x(0) \cos \omega^s t, \qquad \varepsilon_y(t) = \varepsilon_y(0) \sin \omega^s t.$$

These results correspond in principle to the ones of the NIS on ground in Eq. 4.2.25.

4.5 Inertial Navigation with Body-Fixed Sensors – Strapdown Systems

Strapdown systems (SDS) have beeen treated in this section in one breath with platform systems, since the platform as inertial measurement unit (IMU) including the command rates for platform slewing can be fully replaced by a strapdown IMU plus computer for the corresponding transformation matrix computation. The latter is nowadays the preferred INS mechanization due to the obvious reasons concerning cost and reliability. Throughout this chapter the platform system is used primarily for visualization of system characteristics.

Besides the advantages mentioned above, strapdown systems offer additional ones:

- they deliver measurement signals not only for navigation and guidance, but also for flight control, namely angular rate and acceleration in the body-fixed coordinate frame (s. Fig. 1.1 and 1.2);

- they are much more flexible in their adaptation to different system concepts, since no restrictions exist for the choice of the IMU coordinate frame. This means that a strapdown IMU not hard-mounted to the vehicle can carry out its measurement in any coordinate frame and the SDS will still navigate correctly. Here are some examples:

For submarines the INS is the only navigational means for underwater missions. The accuracy requirements for these missions are with 1 NM/day allowable upper position error growth fairly difficult to achieve. Solutions developped by two competitive companies are based upon ring laser gyro (RLG) strapdown IMUs mounted again on a rotable platform.

In the model developed by Litton, the RLGs and the accelerometers are rotated with alternating directions about the vertical platform axis, thus raising the 3 RLG's input above the lock-in level (rate bias) and providing conditions for avaraging out the effect of sensor errors with respect to the navigational coordinate frame (carouselling) [Ko 90].

In the model developed by Honeywell, a two-axis indexer (roll on outside, azimuth on the inside) is employed for rotating the IMU with dithered RLGs periodically 180 degrees about the roll and the azimuth axes, thus providing again the conditions for avaraging out the effect of constant sensor errors upon the navigational error [Le 87].

In another proposal it was shown how to use a strapdown IMU for gyro-stabilizing an instrument and at the same time using it for inertial navigation [St 80].

Conventional SDSs used in civil and military aircraft are truely "strapped down" to the vehicle. They are equipped with RLGs in general and use the wander angle coordinate frame for inertial navigation. As mentioned in Section 4.3, their output signals are identical with the one navigating in the navigational coordinate frame (North Indicating System (NIS)), but they do not have any problems in passing the geographic poles.

The reader more interested in SDSs is referred to Section 1 for a brief opposition of platform and strapdown systems. Sections 2.4 and 2.5 give an introduction into the core of the SDS – the analytic platform as directional reference and its error behavior. The problems of SDS signal processing are outlined in Appendix D. Appendix O describes the optical gyros as sensors which are responsible for the widespread use of the SDS. Sections 3 and 4 are applicable to platform and strapdown systems as well.

4.6 Error Model for all Inertial Navigation System Mechanizations Navigating in Terms of Longitude and Latitude

The derivation of the vector error equations for the directional reference of a strapdown system in Section 2.5 and the acceleration measurement in Section 3.6 have enabled us to generalize these results for any INS mechanization system with respect to the earth. The idea behind this generalization is similar to the statement made in the previous section about the flexibility of a SDS for adaptation to different system concepts – no restrictions exist for the choice of the IMU coordinate frame (index m), i.e. a strapdown IMU not hard-mounted to the vehicle can carry out its measurements in any m-coordinate frame and the SDS will still navigate like the INS developed in this sense. The error propagation of this SDS will thus be the same as that of the INS.

The vector error equations for the directional reference and the acceleration measurement in Eqs. 2.36c and 3.12 read:

$$(4.6.1a,b) \qquad \dot{\varepsilon}_n = \varepsilon_n \times \omega_n^{in} + \delta\omega_n^{in} + \delta\omega_n^{im}, \qquad \dot{V}_n = -\varepsilon_n \times a_n - \delta c_n + \delta g_n + \delta f_n,$$

with

$$(4.6.2) \qquad \varepsilon_n = (\varepsilon_N \quad \varepsilon_E \quad \varepsilon_D)^T = \quad \text{misalignment angles of the platform or the transformation matrix about the north, east and down axes;}$$

(4.6.3) $\omega_n^{in} = \omega_n^{ie} + \omega_n^{en}$ = vector sum of earth rate + transport rate (s. App. C, Eqs. C8 andC9);

(4.6.4a) $\delta\omega_n^{im} = C_{nm} \cdot \delta\omega^g$ = error of IMU-slewing (platform slewing) or IMU angular rate measurement;

(4.6.4b) C_{nm} = matrix for vector transformation from the sensor measurement (index m) to the navigational (index n) coordinate frames;

m = n for the north indicating platform system, i.e. $C_{nm} = I$,

m = a for the platform system slewed to the wander azimuth coordinate frame, i.e. $C_{nm} = C_{na}$ (s. Eq. C 13),

m = c for the platform system continuously rotated about the vertical axis with 1 revolution per minute, for instance (Carousel System) for avaring out the effect of horizontal sensor errors upon navigational accuracy,

m = i for platform systems measuring in the inertial coordinate frame, i.e. $C_{nm} = C_{ni}$ (s. Eq. C 3b),

m = b for strapdown systems mounted to the vehicle, i.e. $C_{nm} = C_{nb}$ (s. App. C, Eq. C10);

(4.6.4c) $\delta\omega^g = K\omega_m^{im} + d$ = gyro errors (s. Eq. 2.22),

$\delta\omega^g \cong d$, since K is hardly observable, in general this approximation is valid,

(4.6.4d) $\omega_m^{im} = (\omega_x \ \omega_y \ \omega_z)^T$ = angular rate of IMU-slewing (platform-slewing = ω^t applied to the gyros, s. Eq. 2.21 for the case m=n) or IMU angular rate measurement;

(4.6.4e) $K = \begin{bmatrix} \kappa_x & \alpha_{xy} & \alpha_{xz} \\ \alpha_{yx} & \kappa_y & \alpha_{yz} \\ \alpha_{zx} & \alpha_{zy} & \kappa_z \end{bmatrix}$ = scalefactor error plus input axes misalignment matrix (s. Eq. 2.20),

(4.6.4f) $d = (d_x \ d_y \ d_z)^T$ = drift vector;

(4.6.5) $V_n = (V_N \ V_E \ V_D)^T = (V_N \ V_E \ -\dot{h})^T$;

(4.6.6a) $a_n = (a_x \ a_y \ a_z)^T$ = vehicle acceleration $\cong$ measured accelerion $f_n = C_{nm} f_m$;

(4.6.6b) $f_m = (f_x \ f_y \ f_z)^T$ = accelerometer output;

(4.6.7) c_n = Coriolis acceleration (s. Eq. 3.10e);

(4.6.8) $g_n = g(\phi) \cdot (\xi \ -\eta \ 1)^T$ with ξ and η = deflection of the vertical (s. Chapter "Navigation Coordinate Frames");

(4.6.9a) $\delta f_n = C_{nm} \delta f^a$ = acceleration measurement error in the n-frame;

(4.6.9b) $\delta f^a = Ka_m + b$ = acceleration measurement error due to scalefactor error plus input axis misalignment (s. Eq. 4.6.4e) and accelerometer bias,

$\delta f^a \cong b$, since K is hardly observable,

$\delta f^a = 0$, in the NIS (m = n), since it is not observable with respect to ε_n.

These relationships show us that for INS error modelling its mechanization enters the algorithm only through the transformation matrix C_{nm}. The internal coupling between the state vector components is the same for all of them and we can use the "North Indicating System (NIS)" as basis for further discussion.

Fig. 4.6.1 shows without simplifications the block diagram for the erroneous signal flow in an NIS. The lower feedback for slewing the vertical platform axis is proportional to $\tan\phi$, indicating that inertial navigation with this kind of mechanization is impossible for polar flights. In the discussion of this section we will stay in the allowable latitude ranges (e.g. $\phi < \pm 88.5$ deg for the Ferranti FE 500).

For flight times above 6 minutes, the drifts of the horizontal gyros become significant. The Coriolis acceleration errors and the earth rate couplings are of importance after approximately 2 hours.

If the NIS is standing on the ground, the dynamics of its errors for long periods are described by the following frequencies (s. [Br 71], Eq. (7-49)):

(4.6.11) $\omega_{1,2} = \omega^s \pm \Omega \sin \phi$, see Figures 1.1, 1.2 and 4.6.1 feedbacks A and
$\omega_3 = \Omega$, see Figures 1.1 and 4.6.1 feedbacks C and also Figure 2.14.

The first two frequencies characterize the Foucault-modulated Schuler oscillation, i.e. the oscillation of a freely swinging pendulum of length R on the rotating earth. The oscillation plane of this pendulum stays, as it is well known, inertially fixed, i.e. it rotates with respect to the earth with the negative vertical component of the earth rate.

The third frequency characterizes the 24-hour oscillation, i.e. the motion as indicated in Fig. 2.13, which is carried out by a misaligned gyro-stabilized platform without a Schuler loop on the rotating earth.
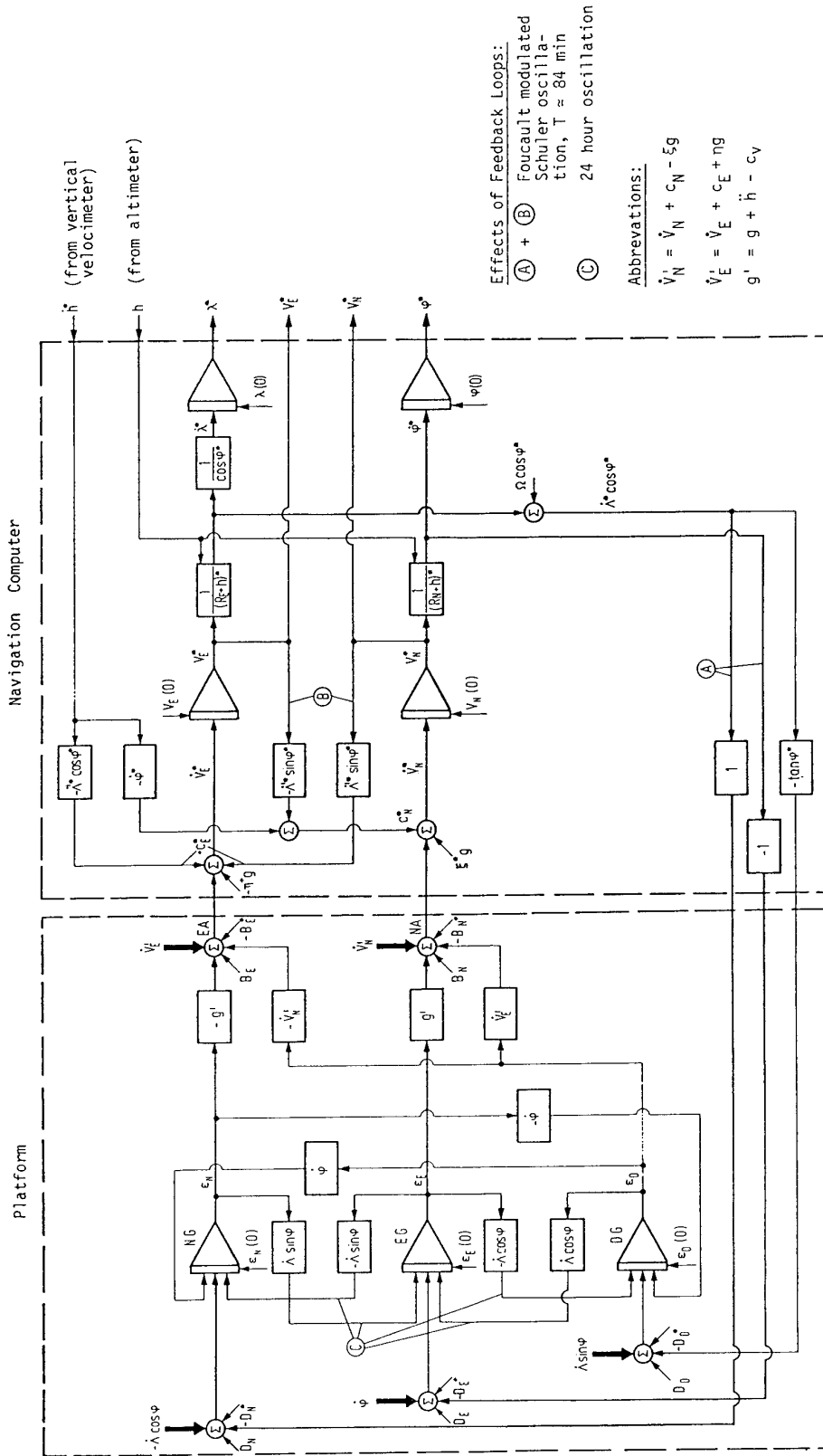
Navigation Computer

Platform

Effects of Feedback Loops:

Ⓐ + Ⓑ Foucault modulated
         Schuler oscilla-
         tion, T ≈ 84 min

Ⓒ   24 hour oscillation

Abbreviations:

$\dot{V}'_N = \dot{V}_N + c_N - \xi g$

$\dot{V}'_E = \dot{V}_E + c_E + \eta g$

$g' = g + \ddot{h} - c_v$

* for computed or
   measured value

$\dot{h}$ (from vertical velocimeter)

h (from altimeter)

Figure 4.6.1 Block Diagram for the Signal Flow in a North Indicating System (NIS) Subject to Errors

The linearized error Eq. 4.6.12 for the NIS is arranged in a way to alleviate the insight into the relationship between left and right hand side.

(4.6.12)

$$[\ \delta\dot\phi \quad \delta\dot\lambda \quad \delta\dot V_N \quad \delta\dot V_E \quad \dot\varepsilon_N \quad \dot\varepsilon_E \quad \dot\varepsilon_D\ ]^T =$$

$$
\begin{bmatrix}
0 & \dot\lambda\tan\phi & -V_E(2\Omega\cos\phi+\dot\lambda\cos\phi) & V_N(2\Omega\cos\phi+\dot\lambda/\cos\phi)+2\dot h\,\Omega\sin\phi & -\Omega\sin\phi & 0 & -(\Omega\cos\phi+\dot\lambda/\cos\phi) \\
0 & 0 & 0 & 0 & 0 & 0 & 0 \\
1/R & 0 & -\dot h/R & 2\Omega'\sin\phi & 0 & -1/R & 0 \\
0 & 1/(R\cos\phi) & -2\Lambda\sin\phi & (V_N\tan\phi-\dot h)/R & 1/R & 0 & -\tan\phi/R \\
0 & 0 & 0 & -g' & 0 & \Lambda\sin\phi & -\dot\phi \\
0 & 0 & g' & 0 & -\Lambda\sin\phi & 0 & \Lambda\cos\phi \\
0 & 0 & \dot V_E & -\dot V_N & \dot\phi & \Lambda\cos\phi & 0
\end{bmatrix}^T
x
\begin{bmatrix}
\delta\phi \\ \delta\lambda \\ \delta V_N \\ \delta V_E \\ \varepsilon_N \\ \varepsilon_E \\ \varepsilon_D
\end{bmatrix}
+
\begin{bmatrix}
-\Delta\dot\phi' \\ -\Delta\dot\lambda' \\ \Delta V_N' \\ \Delta V_E' \\ \Delta\dot\varepsilon_N' \\ \Delta\dot\varepsilon_E' \\ \Delta\dot\varepsilon_D'
\end{bmatrix}
+
\begin{bmatrix}
0 \\ 0 \\ \delta f_N \\ \delta f_E \\ \delta\omega_N^{im} \\ \delta\omega_E^{im} \\ \delta\omega_D^{im}
\end{bmatrix}
+
\begin{bmatrix}
w_{\dot\phi} \\ w_{\dot\lambda} \\ w_{b_N} \\ w_{b_E} \\ w_{d_N} \\ w_{d_E} \\ w_{d_D}
\end{bmatrix}
.
$$

In this equation we have made use of the following abbreviations:

(4.6.13 a to o)

$$\dot\phi = V_N/R \qquad\qquad \dot\lambda = V_E/(R\cos\phi)$$
$$\Lambda = \Omega + \dot\lambda \qquad\qquad R = R_0 + h$$
$$\dot V_N' = \dot V_N + 2\Omega'V_E\sin\phi + \dot h\,\dot\phi \qquad\qquad \dot V_E' = \dot V_E - 2\Omega'(V_N\sin\phi - \dot h\cos\phi)$$
$$\Omega' = \Omega + \dot\lambda/2 \qquad\qquad g' = g + \dot h - 2V_E\Omega'\cos\phi - \dot\phi V_N$$
$$\Delta\dot\phi' = \dot\phi\,\delta h/R \qquad\qquad \Delta\dot\lambda' = \dot\lambda\,\delta h/R$$
$$\Delta\dot\varepsilon_N' = -\dot\lambda\cos\phi\,\delta h/R \qquad\qquad \Delta\dot\varepsilon_E' = \dot\phi\,\delta h/R$$
$$\Delta\dot\varepsilon_D' = \dot\lambda\sin\phi\,\delta h/R$$
$$\Delta\dot V_N' = -\dot\phi\,\delta\dot h + (V_E\,\dot\lambda\,\sin\phi + \dot h\,\dot\phi)\,\delta h/R$$
$$\Delta\dot V_E' = -2\Omega'\cos\phi\,\delta\dot h - 2\dot\lambda\,(V_N\sin\phi - \dot h\cos\phi)\,\delta h/R$$
$$w = \text{uncorrelated noise.}$$

The third last column on the right hand side of Eq. 4.6.12 contains terms proportional to δh and $\delta\dot h$. They are separated from the horizontal components of position and velocity errors, since their contribution is fairly often negligible so that the NIS horizontal and vertical axes can be separated, as shown with the following numerical example.

The altitude error δh causes an error in the platform slewing corresponding to a gyro drift d' and both δh and $\delta\dot h$ a corresponding accelerometer error b'. For $\dot\lambda \cong \dot\phi \cong \Omega$ or $V \cong 1000$ km/h at $\phi = 45$ deg we obtain:

(4.6.14a,b,c) $d' \cong 3\cdot 10^{-3}$ deg/h   and $b' \cong 3\cdot 10^{-7}$ g   per $\delta h = 1000$ m.
$b' \cong 3\cdot 10^{-3}$ g   per $\delta\dot h = 10$ m/s.

Eq. 4.6.12 is of the type 4.2.5 and can be used in a Kalman filter for modelling the errors of any INS, i.e. the north indicating system (NIS), the wander azimuth platform system, the space stabilized platform system employing electrostatic gyros and the strapdown system (SDS). This statement is true as long as "north and south" are still meaningful, i. e. the regions of the geographic poles are exempted. The state vector consisting in this equation of the position errors ($\delta\phi$, $\delta\lambda$), the velocity errors (δV_N, δV_E) and the attitude and heading errors (ε_N, ε_E, ε_D) has to be supplemented by the model of the sensor errors, i.e. the accelerometer errors (δf_x, δf_y) and the gyro errors ($\delta\omega_x$, $\delta\omega_y$, $\delta\omega_z$) in the coordinate frame of the measurement axes. In case they are modelled as bias terms with additional uncorrelated noise, Eq. 4.2.3 for the single axis INS may serve as an example. It should also be emphasized that the transformation matrix C_{nm} between the sensor measurement axes and the axes of the navigational coordinate frame is properly considered as pointed out above. For a system in flight all components of the so-called system matrix F are maneuver-dependent coefficients. The state transition matrix Φ for predicting the INS errors thus can only be computed on-line.

For using the error model 4.6.12 some more practical hints. If an NIS has to be modelled, the accelerometer error $B_{N,E}$ should not be included in the state vector, because it is not observable. These sensor errors cannot be distinguished from a platform misalignment, i.e. the state vector components $\varepsilon_{E,N}$ will comprise them, too. In all other cases when the accelerometer axes are not always held parallel to the axes of the navigational coordinate frame, the accelerometer biases are observable and worth modelling.

At the end of an NIS selfalignment mode discussed in Section 5, the accelerometer errors $B_{N,E}$ are compensated by a platform tilt $\varepsilon_{E,N}$ about the horizontal axes and the east-west gyro drift D_E is compensated by an azimuth offset ε_D. These conditions remain when the NIS is switched from the alignment to the navigational mode, but still is geostationary. So theoretically these sensor errors do not directly cause a position error growth at the beginning of the navigational mode. They do it indirectly only due to

the cross-track error which according to Eq. 4.2.28 depends upon the azimuth misalignment ε_D which again depends upon D_E at the end of the alignment mode.

The situation is different for the north-south gyro drift. At the end of the alignment mode it is compensated by a platform tilt (s. Fig. 5.4) unless the gain K_N is shunted by an integrator (s. Fig. 5.4). In the former case the east-west position error increases immediately after switching into the navigational mode.

All items mentioned above are also valid for the geostationary strapdown system (SDS). But when this system is moved, the equilibrium between sensor error projection in the north-south, east-west and down directions and the misalignment angles is disturbed as indicated in Fig. 2.14. Due to this maneuver-dependency all sensor errors are observable in a Kalman filter and are worth modelling.

The theory discussed above is confirmed in the measurements of a stationary Litton LN-3A platform INS plotted in Fig. 4.6.2 for the velocity and position, and the Schuler oscillations as well as the initial part of a 24-hour oscillation are distinctly to be seen.

The INS in-flight system errors differ from the stationary system errors by the maneuver-dependency of several contributors, as to be seen from Fig. 4.6.3.

There are firstly the cross-track velocity ($\delta V_{N,E} \cong \pm\, \varepsilon_D \cdot V_{E,N}$) and position errors ($\delta S_{N,E} \cong \pm\, \varepsilon_D \cdot \Delta S_{E,N}$), as outlined in the previous section. They are common to all inertial navigation systems and are superposed to the Schuler oscillations.

There is secondly the fact that with strapdown systems the sensor errors are effective within the body-fixed axes, i.e. they act within the navigational coordinate frame via the maneuver-dependent C_{nb} matrix (s. Eqs. 2.41c and 3.11 as well as Fig. 2.14).

There is thirdly the fact that the sensor errors themselves can be maneuver-dependent as within mechanical gyros, for instance (s. [St. 82]).

In order to estimate the accuracy of a geostationary INS over more than 2 hours time of flight for a certain class of sensor and initial alignment errors comprised as 1-σ values, Eq. 4.6.12 was solved on a digital computer and its results have been plotted in Fig. 4.6.4 as 1-σ INS errors over a 6-hour and a 36-hour period. They describe the two-dimensional Gaussian distribution of the horizontal position error (s. Fig. 4.6.5 and 4.6.6). From the effect of the various uncorrelated error sources on the position error the following can be stated:

-As V = 0, the cross-track error does not become significant.

-For short periods of up to 1/2 h in the north-south channel or 2 h in the east-west channel the influences of the initial misalignment and of the accelerometer bias are identical, which can be explained by the fact that initially the azimuth misalignment does not become significant.

-In the north-south channel the initial misalignment causes a mean position error increasing linearly with time - according to Eq. (4.2.24b) the azimuth misalignment has the effect of an east-west gyro drift.

-The gyro drift causes a mean position error increasing linearly with time, which is due mainly to the horizontal gyros. According to Eq. (4.2.25) the vertical gyro drift of a stationary NIS after a time of:

(4.6.15) $\Delta t \cong 4/(\Omega \cos\phi) \,>>\, 2h,$

causes the same mean position error as that produced by an east-west gyro drift of the same amount. However at longer times the effect of the vertical gyro drift becomes dominant.

The two components of the horizontal velocity error or the horizontal position error have Gaussian distribution as indicated in Fig. 4.6.5. Horizontal sections through this distribution are error ellipses as described by the exponent of p(x) in Fig. 4.6.5 and as shown in Fig. 4.6.6 for the position. The projection of the ellipses on the north-south and east-west axes are the 1-σ values obtained from Eq. 4.6.6 for $S_{N,E}$. For system specification these 1-σ values in the two axes are not so convenient; the manufacturer often gives the overall specification in form of the CEP or SEP (circular error probable, spherical error probable) value which in a two-dimensional or three-dimensional case is the radius of a circle or sphere comprising the errors with 50 % probability unless otherwise mentioned (e.g. CEP_{95} or SEP_{95} corresponds to 95 % probability) [STA]. The CEP can be computed from the corresponding elements of the covariance matrix as pointed out in [Ka 69]. In the example of Fig. 4.6.6 we obtain, with σ_N = 5.62 km, σ_E = 3.81 km and ρ = - 0.8, the 50 % and 90 % probability circles CEP = 5.0 km. For equal σ-values in all three dimensions we obtain for instance: SEP = 1.5382 σ.
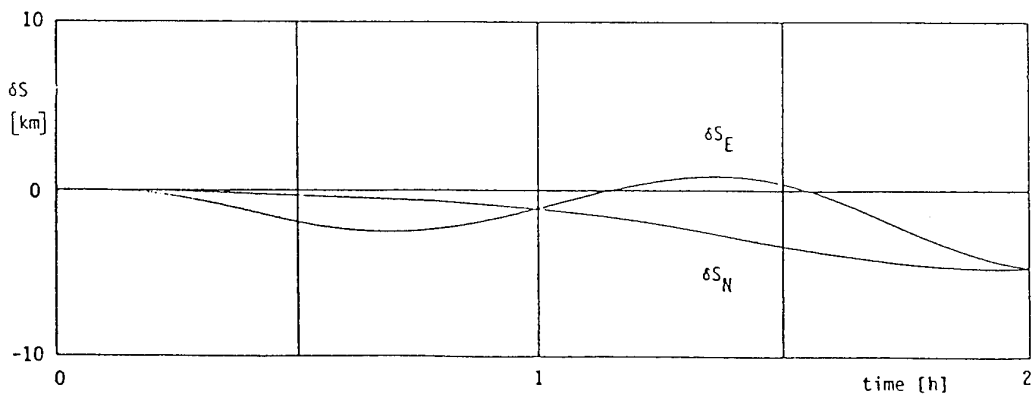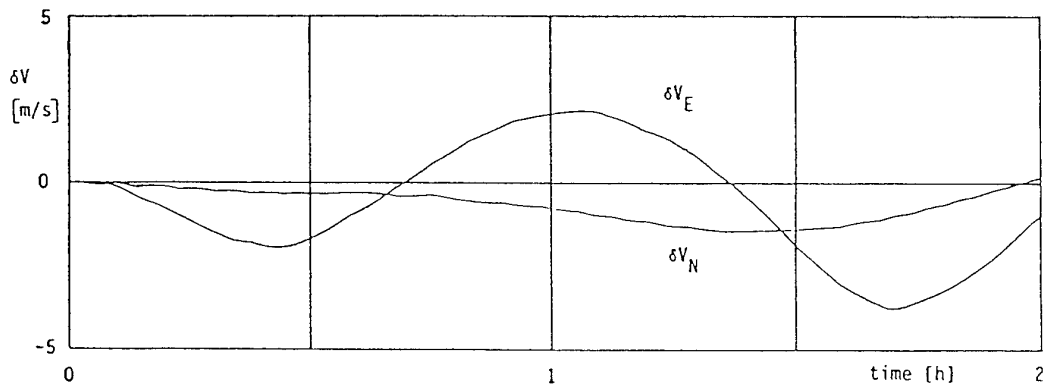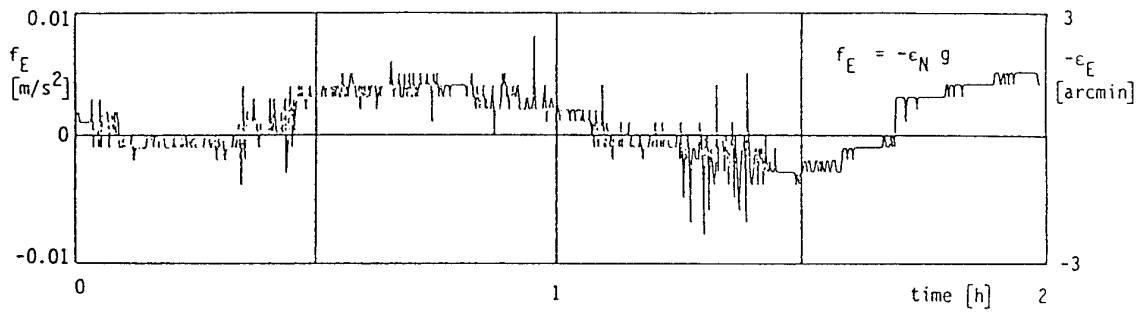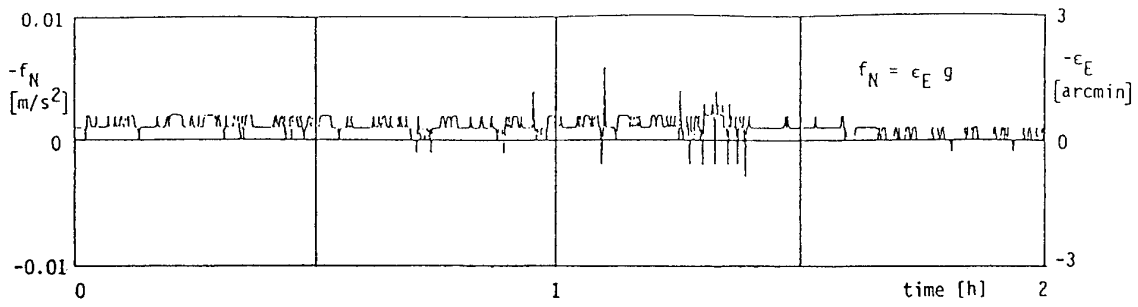
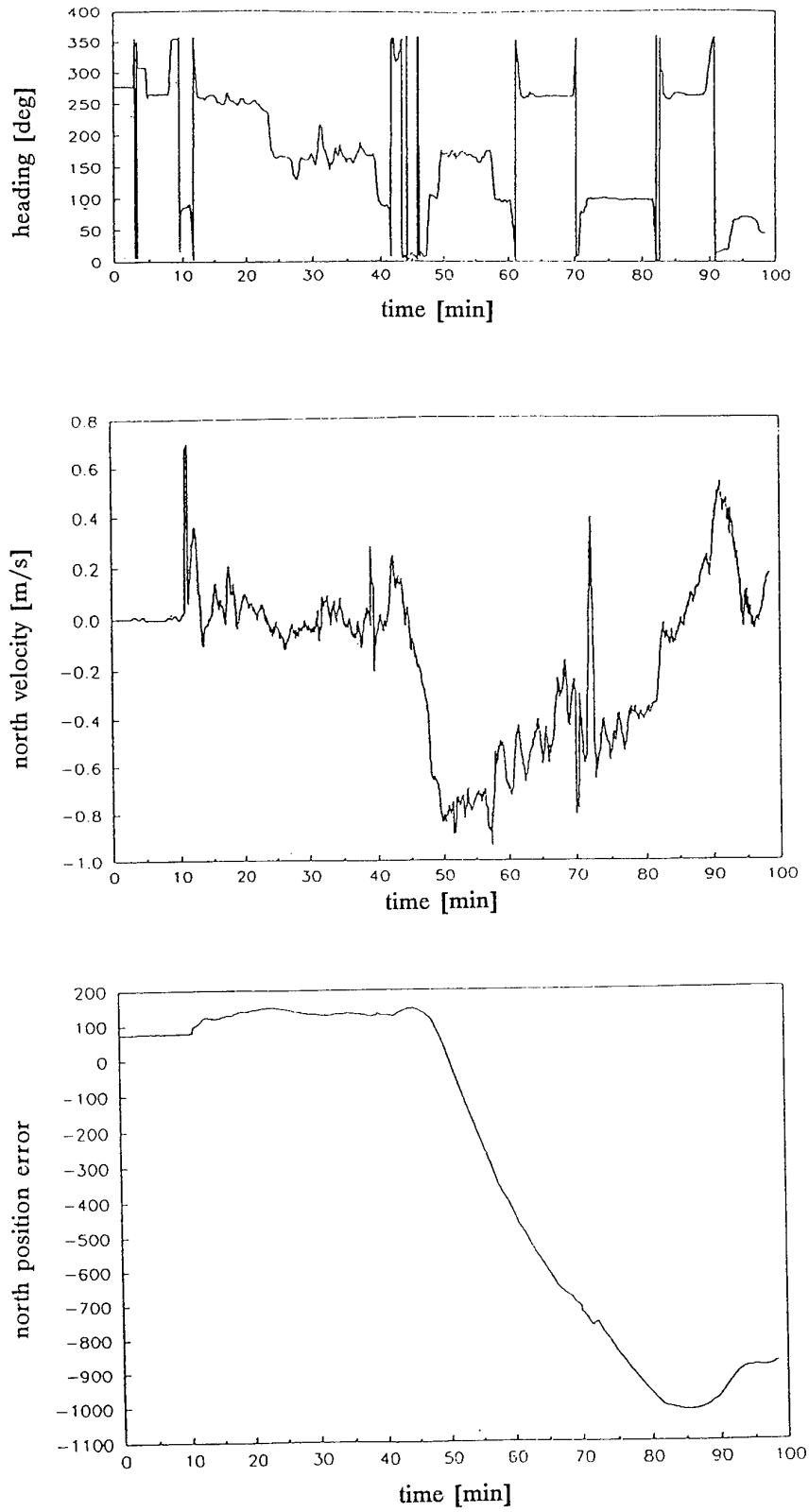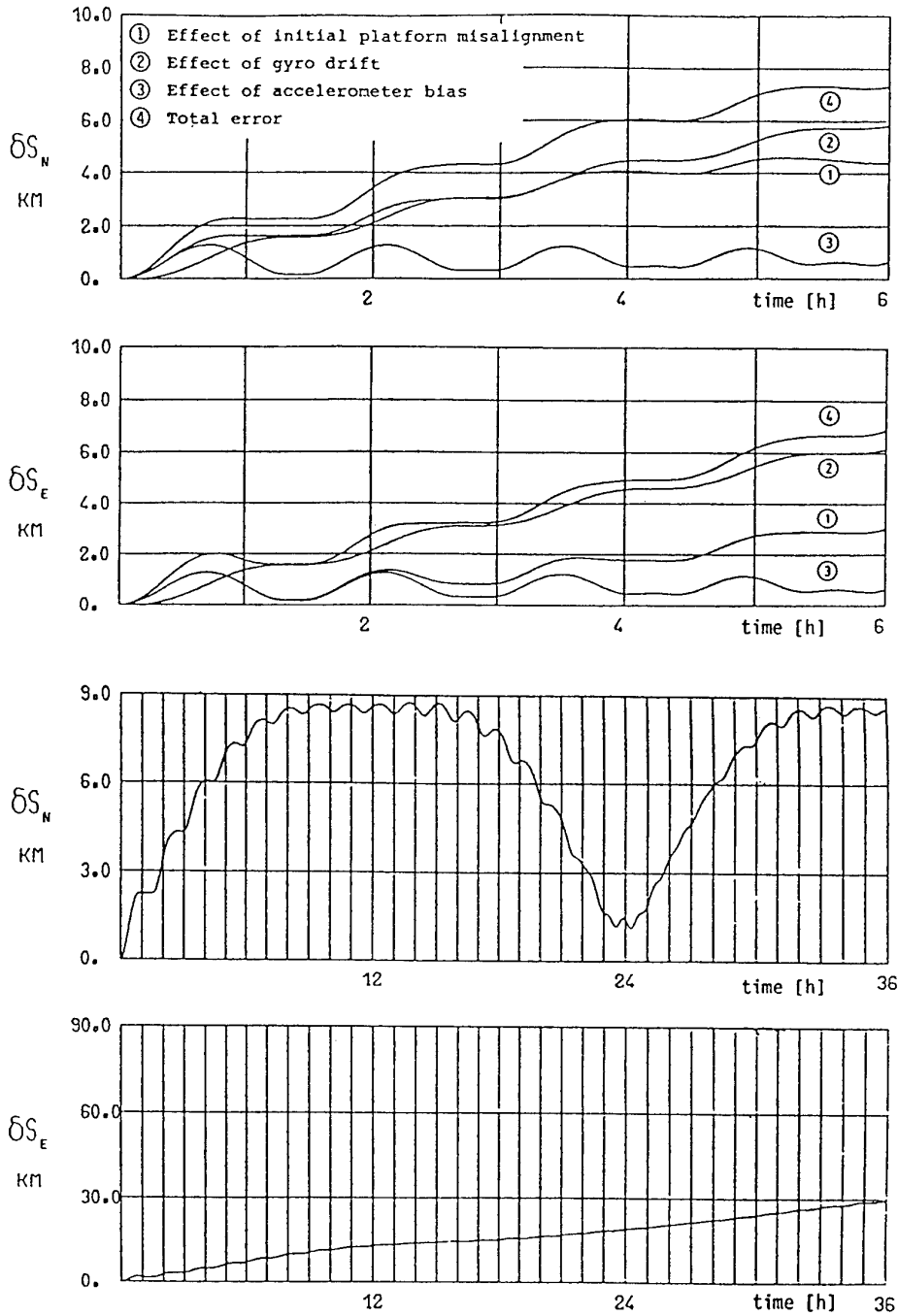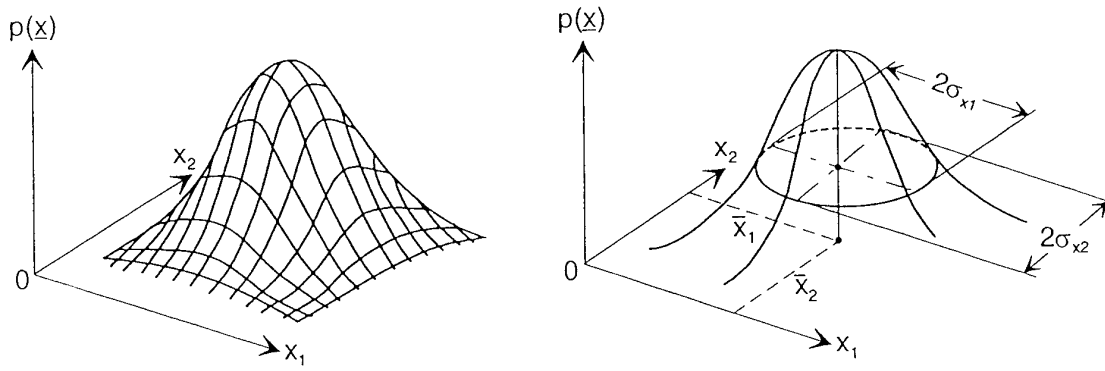Figure 4.6.2 Velocity and Position of the Stationary Litton LN-3A Platform Inertial Navigation System

Figure 4.6.3 Typical Flight-Test Results of the Litton LTN-90 Strapdown Inertial Navigation System

```
10.0
 8.0    ① Effect of initial platform misalignment
        ② Effect of gyro drift
 6.0    ③ Effect of accelerometer bias
        ④ Total error
```

δSₙ
KM

(curves labeled ④, ②, ①, ③)

time [h]   2   4   6

δS_E
KM

(curves labeled ④, ②, ①, ③)

time [h]   2   4   6

δSₙ
KM

time [h]   12   24   36

δS_E
KM

time [h]   12   24   36

Uncertainties for initial misalignment angles: $\sigma(\varepsilon_{N,E})$ = 0.1 mrad = 20 arc sec,

$\sigma(\varepsilon_D)$ = 1 mrad = 3.4 arc min,

for gyro drift: $\sigma(D_{N,E,D})$ = 0.01 deg/h,

for accelerometer bias: $\sigma(B_{N,E})$ = 0.0001 g.

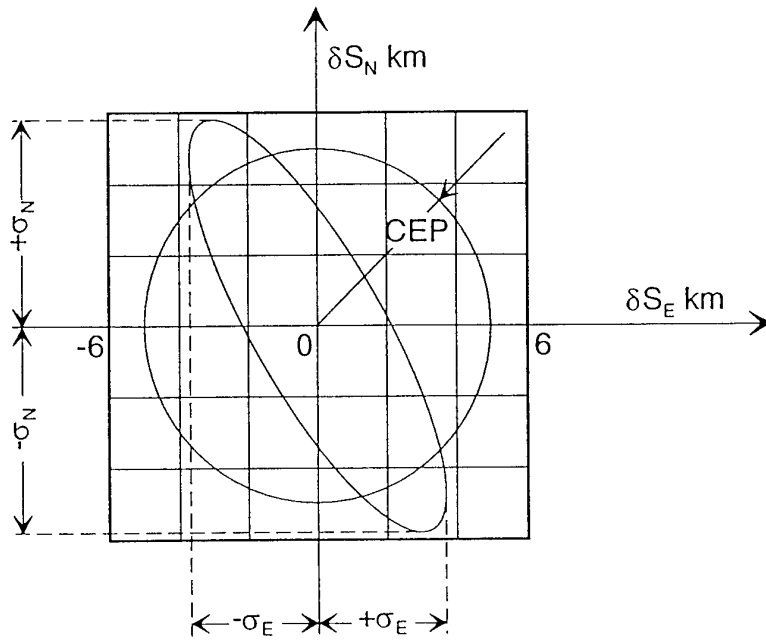Figure 4.6.4 1-σ Position Error of a Stationary North Indicating System

$$p(\underline{x}) = \frac{1}{(2\pi)^{n/2} |\underline{P}|^{1/2}} \exp[-\tfrac{1}{2}(\underline{x} - \underline{\overline{x}})^T \underline{P}^{-1}(\underline{x} - \underline{\overline{x}})]$$

$E(\underline{x}) = \underline{\overline{x}} = $ mean value of vector

$E[(\underline{x} - \underline{\overline{x}})(\underline{x} - \underline{\overline{x}})^T] = \underline{P} = $ corvariance matrix of vector

Figure 4.6.5 Two-Dimensional Gaussian Distribution

$$\underline{p} = \begin{bmatrix} \sigma_N^2 & \rho\sigma_N\sigma_E \\ \rho\sigma_N\sigma_E & \sigma_E^2 \end{bmatrix} = \begin{bmatrix} 5,62^2 & -0,8 \cdot 3,81 \cdot 5,62 \\ -0,8 \cdot 3,81 \cdot 5,62 & 3,81^2 \end{bmatrix}$$

Figure 4.6.6 Position Error Ellipsis and Circular Error Probable of a North Indicating System

Literature Chapter 4

[Au 91] Ausmann, J. S.: "Baro-Inertial Loop for the USAF Standard RLG INU."
Navigation: Journal of The Institute of Navigation, Vol. 38, No. 2, Summer 1991, pp. 205 - 220

[Br 71] Britting, K.: "Inertial Navigation System Analysis"
Wiley Interscience 1971

[Bo 35] Boykow, J. M.: "Einrichtung zum Messen von Wegstrecken."
Deutsches Patent 1935

[Ge 74] Gelb, A. (Editor): "Applied Optimal Estimation."
The M.I.T. Press, Cambridge, Mass., USA, 1974

[He 80] Heilbron, H., Klein, G.: " Ein Rückblick auf die Entstehung der Trägheitsnavigation."
DGON Zeitschrift Ortung und Navigation, 1/1980, pp 36 - 60.

[Hi 92] Hilscher, G.: "Flug ohne Sterne."
Olynthus Verlags Anstalt, Vaduz, 1992

[Ka 69] Kayton, M., Fried, W. R.: "Avionics Navigation Systems".
J. Wiley and Sons, Inc. , New York, 1969

[Ko 90] Kohl, K.W.: "A new High Accuracy Ship's Inertial Navigation System PL41 MK4."
Proceedings Symposium Gyro Technology 1990, DGON and Universitaet Stuttgart, 1990

[Le 87] Levinson, E., Majure, R.: "MARLIN - The next Generation Marine Inertial Navigator."
Proceedings Symposium Gyro Technology 1987, DGON and Universitaet Stuttgart, 1987

[Ma 78] Magnus, K.: "Kreisel als vielseitige Hilsmittel in Luft- und Raumfahrt."
Zeitschrift für Flugwissenschaften und Weltraumforschung 1978, Vol. 2, pp 217 - 227.

[NM 84] "Digital Air Data Computer, Part No. 12238-3, Maintenanca Manual with Illustrated Part List, 34-10-10"
Nord Micro, Oct. 15/84

[Re 85] Redeker, A., Voersmann, P.: "Precise Vertical Speed Reconstruction Based on Vertical Acceleration and Barometric Altitude."
Z. Flugwiss. Weltraumforsch. 9. 1985, Heft 4

[STA] " Standardization Agreement; Subject Method of Expressing Navigation Accuracies."
North Atlantic Treaty Organization, STANAG 4278

[St 80] Stieler, B.: "Fahrzeug mit einer Plattform fuer eine Ziellinienstabilisierung in Verbindung mit einem Inertialsystem."
Deutsch Patentschrift DE 3019743 C2, Anmeldetag 32.5.1980

[St 82] Stieler, B., Winter, H.: "Gyroscopic Instruments and their Application to Flight Testing."
AGARD-AG-160-Vol. 15, 1982

[St 94] Stieler, B.: "Generation of an Altitude Reference Using a Conventional INS.
DLR-Mitt. 94-xx, 1994

[Wi 75] Winter, H.: "Messung der Vertikalbewegung eines Flugzeugs."
Deutsche Luft- und Raumfahrt, Forschungsbericht 75-41

[Wu 80] Wuest, W.: "Pressure and Flow Measurement."
AGARD-AG-160-Vol.11, 1980

5. Initial Alignment and Calibration of Inertial Navigation Systems

5.1 Introduction

The initial self-alignment loop or gyrocompassing loop of a stationary inertial platform is in principle the electronic equivalent of a mechanical gyrocompass used as a heading reference on ships.

The gyrocompass is in principle a damped gyropendulum with a horizontal spin axis and a separation between center of gravity and center of support of about 1 mm. In the geostationary case this is a northseeking instrument, the input signals being the rotation of the earth and gravity. It has a settling time in the order of 4 hours.

In the following we will briefly review the principles of the gyrocompassing loop of an inertial platform and the accuracy achievable for the initial alignment.

5.2 The Gyrocompassing Loop of an Inertial Navigation System

In Section 2.1 we have discussed the effect earth rotation has upon a free gyro whose spin axis is slightly misaligned with respect to the earth axis. This gyro will carry out a 24-hour motion around Polaris (s. Fig. 2.5).

Fig. 5.1 shows in principle the mechanical gyrocompass differing from the free gyro in Fig. 2.5 through the pendulosity only. Its spin axis is nearly horizontal. Once disturbed from its equilibrium, namely true north, it will oscillate in an elliptic cone about this equilibrium with a frequency adjustable by the pendulosity or it will align itself with true north if it is damped as indicated in this figure.

Fig. 5.2 indicates how the pendulosity and the damping can be replaced by an accelerometer A_N and two feedback loops. The accelerometer input axis is aligned with the spin axis. Its output signal is fed via the gain $- K_D \cdot H$ to the horizontal torquer for raising the frequency, and via the gain $- K_E \cdot H$ to the vertical torquer for raising the damping of the sensor's oscillation.
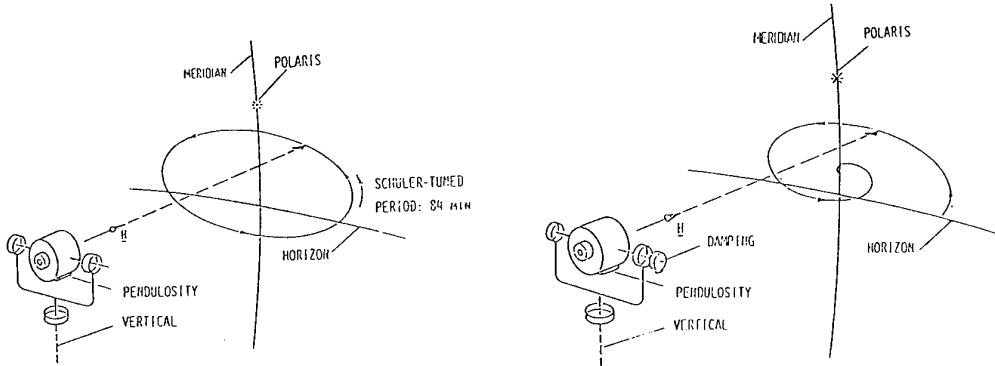

Fig. 5.1 The Motion of the Undamped and the Damped Gyrocompass
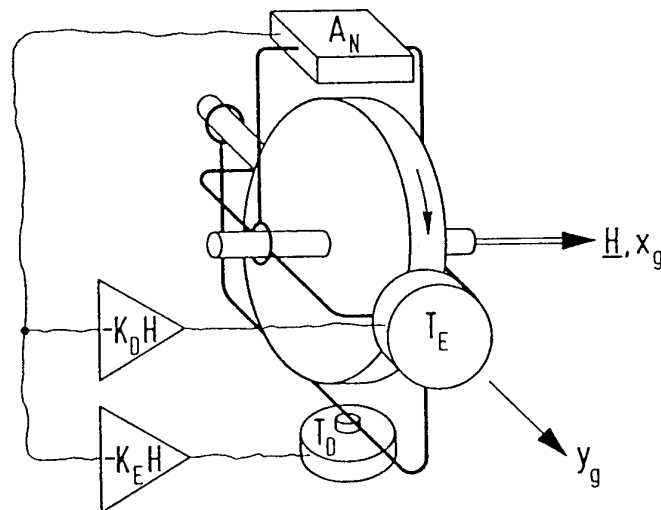

Fig. 5.2 The Gyrocompass Implemented with a Free Gyro, an Accelerometer and Two Feedback Loops

We find the same interconnections in the simplified gyrocompassing loop of an inertial platform whose block diagram is shown in the lower half of Fig. 5.3. The output of the north-south accelerometer A_N causes a platform rotation ω_E and ω_D about the east-west and the down axes via the gains $-K_E$ and $-K_D$. The upper half of the block diagram shows the levelling loop for the north-south axis.

One of the simplifications in this diagram concerns the model of the free platform (s. Fig. 2.13) which has been reduced to the coupling of the vertical misalignments into the east-west axis via the horizontal component of earth rate $\Omega \cdot \cos \phi$. The coupling between the horizontal axes via the vertical component $\Omega \cdot \sin \phi$ of earth rate has been neglected, which is acceptable since the horizontal misalignment angles are smaller by one order of magnitude than the azimuthal misalignment as indicated in Table 2.1 and as will be shown below.

5.3 Equilibrium and Sensor Calibration at the End of the Alignment Process

From the block diagram in Fig. 5.3 one can easily derive the following relationships for the equilibrium of the platform at the end of the alignment process when the inputs to the integrators are zero:

- for ideally compensated sensors, i.e. no accelerometer biases ($B_{N,E} = 0$) and no gyro drifts ($D_{N,E,D} = 0$) the platform will be ideally aligned with the vertical and true north:

(5.1) $\varepsilon_{N,E} = 0$;

- the horizontal accelerometer biases ($B_{N,E} \neq 0$) are compensated by corresponding gravity components generated by horizontal platform misalignment angles (remember the accelerometer output signals are null in the equilibrium phase in this case):

(5.2) $\overline{\varepsilon}_{N,E} = \pm\ B_{E,N}\ /\ g\ \ (=\ \pm\ .1\ \text{mrad}\ =\ 20\ \text{arc sec for}\ B_{N,E} = 10^{-4}\ g)$;

- the east-west gyro drift ($D_E \neq 0$) is compensated by a corresponding component of the horizontal earth rate generated by a vertical platform misalignment angle in the equilibrium phase:

(5.3) $\overline{\varepsilon}_D = -\ D_E\ /\ (\Omega \cdot \cos\ \phi)\ \ (=\ 1\ \text{mrad}\ =\ 3.4\ \text{arc min for}\ D_E = .01\ \text{deg/h and}\ \phi = 45^\circ)$;

- if in addition the north-south and the vertical gyros are corrupted by the drift $D_{N,D}$, the horizontal accelerometers furnish in the equilibrium phase additional control offset signals:

(5.4) $\overline{\varepsilon}_{N,E} = \pm\ (B_{E,N}\ +\ D_{N,D}\ /K_{N,D})\ /\ g$.

Whilst the accelerometer biases $B_{N,E}$ and the east-west gyro drift D_E cannot be measured - their effect is compensated by gravity or earth rate, respectively - the north-south and the vertical gyro drift $D_{N,D}$ cause an output voltage of the horizontal accelerometers and thus can be measured.

The additional tilt due to D_N, and thus the control offset voltage, can be reduced by setting the loop gain K_N high or preferably by shunting the amplifier of the levelling loop with an integrator as indicated in Fig. 5.4 for the north-south gyro drift. In practice only this sensor error is compensated in the manner shown. The vertical gyro drift can only be measured by means of the synchro mounted on the vertical gimbal axis, which requires the absence of any platform motion.

Eq (5.3) is applicable to all north-seeking methods with gyroscopic sensors. These methods are limited to 80 deg latitude, approximately ($\approx$ 76.5 deg for Delco Carousel IV platform system). Note that only the initial self-alignment of the INS is limited to this latitude; the actual navigation function can be performed at all latitudes (s. Section 4.3).

If we choose the following gains in the gyrocompassing loop of Fig. 5.4:

(5.5a,b) $K_N = K_E = 1/(2gT)$,

              $K_D = 1/(16gT\ \Omega\cos\ \phi)$

and the gain of the shunt integrator in the levelling loop:

(5.5c) $K_N' = K_D\ \Omega\cos\ \phi$,

both loops - the north-south levelling loop and the gyrocompassing loop - have the same dynamics, namely four equal roots with the time constant:

(5.6) $\tau = 2T = 60\ \text{s}$

(the figure was selected in laboratoy tests [St 78]) and the problems of compensating the north-south gyro drift and of vertical platform alignment are alike.

If the gyrocompassing alignment loop is implemented in this fashion, the system will carry out the elliptical motion as shown in Fig. 5.1; the semimajor axis of the ellipsis is then proportional to the vertical or azimuthal misalignment and the semiminor axis to the east-west or horizontal misalignment. With the time constant mentioned above we obtain as ratio of both axes 140, i.e an initial horizontal misalignment excites a much bigger disturbance motion of the gyrocompassing process than an initial vertical misalignment. In order to speed up the alignment process, it is advisable to split it up into two modes - the levelling and the gyrocompassing mode. During the former the horizontal misalignments in both axes are minimized with a levelling loop as indicated in Figs. 5.3 and 5.4 lasting 1 min approximately. During the subsequent gyrocompassing mode the azimuthal misalignment is minimized during 8 min approximately. The figures quoted are valid for the Litton LN-3A platform system.

It is shown in [St 78] that for strapdown systems another alignment concept has advantages as compared to the control loops shown in Figs. 5.3 and 5.4. Only two levelling loops are implemented here, and the steady state signal to clamp both horizontal axes to gravity are proportional to the north-south system drift D_N and the vertical misalignment ε_D. Both parameters are estimated with the use of a simple Kalman filter neglecting all system dynamics. The major advantage of this concept lies in the fact that the estimation of both parameters becomes with time increasingly insensitive to vehicle sway due to wind gusts or
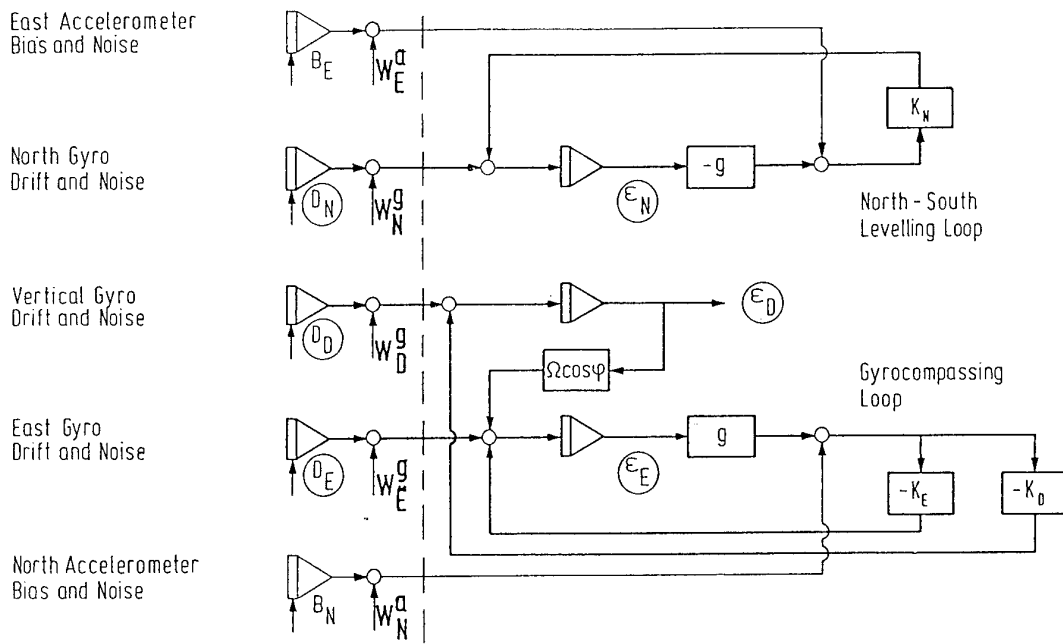
Figure 5.3 Simplified Block Diagram of the Gyrocompassing and the North Levelling Loops for an Inertial System
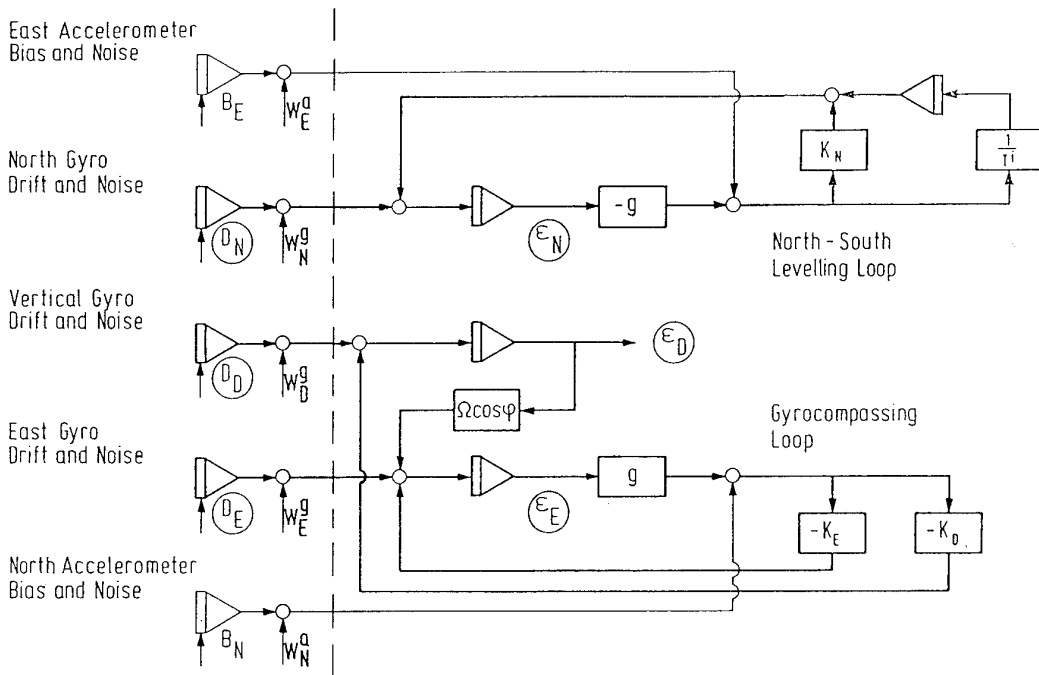

Figure 5.4 Simplified Block Diagram of the Gyrocompassing and the North Levelling Loops for an Inertial System Including Integrating Network for North Gyro Drift Compensation

baggage loading in an aircraft.

We have not yet mentioned the vertical accelerometer. Its output signal, which should be gravity, can certainly be used for calibration. For the field-calibration of all sensor biases ($D_{N,E,D}$ and $B_{N,E,D}$) and scale factors ($S^g_{N,E,D}$ for the gyros, $S^a_{N,E,D}$ for the accelerometers) the gyrocompassing mode is carried out several times with the input axes of each gyro once in the north and once in the south directions and the input axes of each accelerometer once in the up and once in the down directions, which requires 9 orientations of the platform. The sum and differences of the currents into the north or south gyro $i_{N,S}$ and into the up or down accelerometer $i_{U,D}$ are evaluated in the following way:

$$(5.7a,b,c,d) \qquad D = -S^g (i_N + i_S)/2 \qquad\qquad S^g = (i_N - i_S)/(2\ \Omega \cos \phi)$$

$$B = -S^a (i_D + i_U)/2 \qquad\qquad S^a = (i_D - i_U)/(2g).$$

5.4 Azimuth Alignment Accuracy Due to Random Walk Gyro Drift

So far we have discussed the initial alignment accuracy due to systematic sensor errors. We know that such errors can be compensated once they are known, but the user has to live with a certain day-to-day variation of the systematic errors which always limit the final alignment accuracy. Modern sensors, especially the ring laser gyros, have an unprecedented low systematic drift so that the stochastic error known as random walk dominates. One cannot predict the effect of stochastic errors as we have done for systematic errors. All one can predict is the expected value or the ± 1σ band which encloses 68% of all samples. In order to derive the ± 1 σ values for the azimuth alignment due to east-west gyro random drift, we use the integral of Eq. 5.3, i.e. we divide the ± 1σ values for all integrated random drift angles through the integral of earth rate. The result is:

$$(5.8) \qquad \sigma(\varepsilon_D) = r(d_E)\ t^{1/2} / (\Omega \cos \phi \cdot t)$$

$$= \sqrt{60}\ r(d_E)\ [deg/\sqrt{h}\ ] / (\Omega \cos \phi \cdot \sqrt{t\ [min]}).$$

This equation is evaluated in Fig. 5.5 for a random walk coefficient of $r(d_E) = .001\ deg/\sqrt{h}$ which is approximately the value measured at DLR for the Honeywell GG 1342 RLG (Ul 88). According to military requirements the alignment time is limited to below 4 min with an accuracy below 10 arc min.
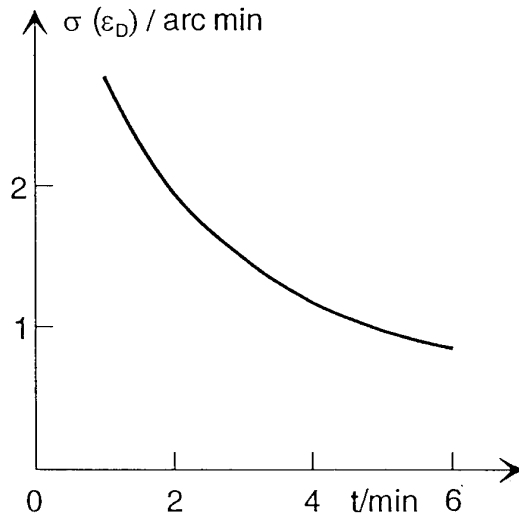
Fig. 5.5 Azimuth Alignment Accuracy Due to East-West Random Walk as a Function of Time

Literature Chapter 5

[St 78] Stieler, B., Zenz, H.P.,: "On the Alignment of Platform and Strapdown Systems."
Proceedings of the DGON Symposium Gyro Technology 1978, Stuttgart, Germany

[Ul 88] Ullrich, D., Lübeck, E., Wetzig, V.: "Static and Dynamic Measurements with a Ring Laser Gyro."
Proceedings of the DGON Symposium Gyro Technology 1988, Stuttgart, Germany

6. Attitude and Heading Reference Systems (AHRS)

6.1 Introduction

Attitude and Heading Reference Systems (AHRS) will play an increasing role in future aerospace guidance and control due to the increasing role of the Global Navigation Satellite System (GNSS) and its integration with inertial reference systems of reduced performance. The latter is the keyword for describing the AHRS system structure. As already mentioned in Section 2, the AHRS is no stand-alone system, but always integrated with other types of sensors or systems in a predefined control loop structure or with the aid of Kalman filters. The margin between both types of system integration is floating.

The principles of a fairly simple AHRS system were already outlined in Section 2.1 in connection with the Vertical Gyro (VG) and the Directional Gyro (DG). We will review them in this section briefly and describe some modern system implementations.

6.2 The AHRS with Magnetic Aiding

In Section 2 we have seen that the directional reference of the inertial navigation system (INS) and thus also of the AHRS can be implemented as a physical or analytic platform. It was shown that its characteristics can be described in both cases by that of a three axes free gyro with an ideal suspension. The AHRS signal flow is discussed in the following as platform implementation only.

Let us first discuss the basic functioning of the "attitude and heading reference" within the simplest AHRS application by looking again to that of the VG and DG in Figs. 2.6 and 2.7.

We recall that in the INS the accelerometer output signals were used in a first step to compute groundspeed and in a second step to compute the Schuler feedback for platform slewing. This resulted in an ideal vertical reference with an 84-minute Schuler period of the closed loop, but required gyros of "inertial quality" (s. Section 4.2.2.1).

The vertical reference in Figs. 2.6 and 2.7 for measuring attitude is based upon bubble levels or accelerometers for sensing the misalignment about the horizontal axes. Their output signals are directly used for gyro or for platform slewing for nulling the misalignments as shown in these figures.

Assuming in Fig. 2.8 the gain $K_2 = 0$, the gain K_1 defines the system's tuning about the horizontal axis. The time constant:

(6.1) $T = 1/(g\ K_1) < 10$ min

is much lower within the AHRS than within the INS. Due to the stronger coupling of the horizontal gyros to the horizontal accelerometers, the effect of gyro drift D_h upon horizontal misalignment:

(6.2) $\varepsilon_h = D_h \cdot T = D_h/(g\ K_1)$

is much smaller than in an INS and the requirements for the gyro performance can be reduced by at least one order of magnitude. Even this sensitivity to constant or slowly varying gyro drift is reduced by the additional integrator within the control loop.

These advantages are paid by the increased sensitivity of the misalignment angles to horizontal accelerations. As indicated in this figure, the most simple means to cope with this effect is the cutoff of the feedback at a certain acceleration level.

The directional reference for measuring heading is based in an AHRS upon a magnetic compass for sensing magnetic north, i.e. based upon a so-called flux-gate as indicated in Figs. 2.6 and 2.7 (s. Chapter "Magnetic Heading References") or based upon a 3-axis strapdown magnetometer. The difference between AHRS north and compass north is used for gyro or for platform slewing to magnetic north.

The flux gate has its built-in vertical reference which is not as accurate as the AHRS vertical reference. The integration of a 3-axis strapdown magnetometer with an AHRS is thus the system concept of higher performance.

Soft and hard iron induced magnetometer errors can fairly well be calibrated with the help of the gyros within the AHRS. For this purpose the aircraft standing on the airfield is rotated about the vertical axis with respect to magnetic north and the vertical gyro readings and magnetometer readings are evaluated with special software described in [Ba 85, St 90], for instance.

6.3 The AHRS with Doppler Aiding

The Doppler-aided AHRS is of higher performance than the one mentioned above. It has true navigational capability and is a preferred system for helicopters. It combines the advantages of the INS and Doppler, i.e. accurate attitude heading and groundspeed with reduced time-dependent position error growth.

The AHRS for this application is in principle identical in its signal flow to an INS (s. Fig. 4.2.5), but is equipped with sensors of lower performance. Its Doppler aiding consists of feeding the differences between the inertial velocities and the Doppler velocity back into the AHRS to correct the system attitude, heading and velocities.

Fig. 6.1 shows the error block diagram of a so-called third-order in-flight gyrocompassing loop, whereby the Doppler velocities measured in the body-fixed coordinate frame are resolved into the north and east directions by the attitude and heading angles as measured by he AHRS. This is indicated in the upper right corner of Fig. 6.1. Due to the feedback in the horizontal channels, the Schuler oscillations of the AHRS are damped and the frequency is augmented: the loops are N-times Schuler tuned and often critically damped.

A fairly simple analysis reveals the ground speed error δV due to a Doppler velocity error δV^d, to an accelerometer error b and to a gyro drift d ($\bar{\ }$ means "variable in the Laplace domain"):
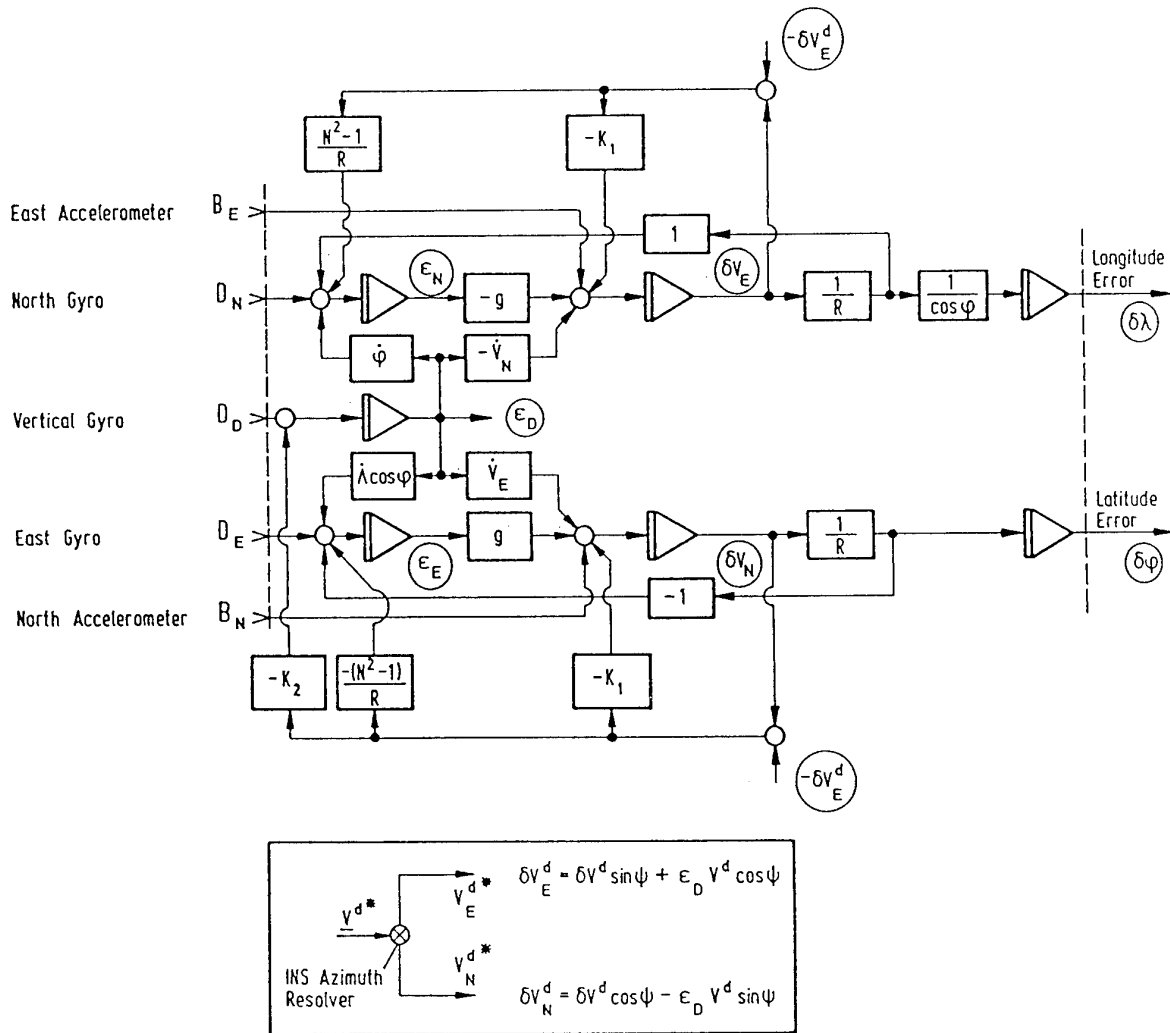
Figure 6.1 The Doppler-Aided AHRS or the Third-Order In-Flight Gyrocompassing Loop

(6.3a) $\quad \delta\tilde{V} = \dfrac{1}{[s^2 + K_2 s + (\omega^s)^2(1 + R\ K_1)]}\ [s\tilde{b} + g\tilde{d} + (sK_2 + gK_1)\ \delta\tilde{V}]$.

In the high frequency domain (s $\mapsto$ ∞) the ground speed error is:

(6.3b) $\quad \delta V \mapsto 0$

and in the low frequency domain it is:

(6.3c) $\quad \delta V \mapsto \dfrac{R}{[1 + R\ K_1]}\ [D + K_1 \delta V^d] \cong .01\ RD + \delta V^d$.

The latter is valid for the following gains used in the Singer Kearfott AHRS of the type SFIM 253 A:

(6.4 a) $\quad 1 + RK_1 \cong 100 \mapsto \omega \cong 10\ \omega^s$

with ω = closed loop frequency, ω^s = Schuler frequency (s. Eq. 4.2.1a) and:

(6.4b) $\quad K_2 \cong 14\ \omega^s \qquad \mapsto \quad \delta \cong .7$,

with δ = damping coefficient of the closed loop.

The Doppler noise is thus damped out and the navigation system error follows the low frequency Doppler error, but reduces the effect of inertial sensor errors upon navigation accuracy.

Fig. 6.1 also shows that the north velocity difference between the AHRS and the Doppler is fed back to the vertical gyro to correct the azimuth AHRS misalignment. Though neither the AHRS nor the Doppler radar "know" the direction of true north, both find it "hand in hand" by means of this feedback signal. This in-flight gyrocompassing becomes feasible, since due to earth rate the errors of this AHRS caused by an azimuth misalignment differ in the north- south channel from those of a Doppler system with external heading reference, i.e. a conventional dead reckoning system. In the east-west channel the INS error due to

the azimuth misalignment equals the conventional dead reckoning system through the well known crosstrack error, and feeding back the corresponding velocity difference is of no use. For verification compare in Eq. 4.2.5 $\delta\phi(t) = f(\varepsilon_D)$ and $\delta\lambda(t) \cos \phi = f(\varepsilon_D)$.

Eq. (4.2.28b) finally reveals that also in the north-south channel the INS error is equal to that of the dead reckoning system, if it is initially aligned using the stationary gyrocompassing procedure. From this we may deduce that Eq. (4.2.26b) also limits the in-flight gyrocompassing accuracy under the presumption that the Doppler velocities are measured accurately in the body-fixed coordinate frame.

Since the AHRS is equipped with lower quality gyros, this in-flight gyrocompassing is not applied in general and it is always integrated with a sensor for magnetic north.

6.4 Modern AHRS Implementations

From the previous sections we may deduce the following principal differences between the modern AHRS and the INS implementations:

-The AHRS is equipped with sensors of lower performance; fiber optic gyros (FOGs, see Appendix O) are very promising sensors at present, micromechanical gyros and accelerometers have the potential to penetrate this application in the future [Ma 92, El 91].

-The computation of the analytic platform (see Section 2.4 and Appendix D), i.e. the high frequency data processing is identical in principle in both systems, the AHRS and the INS.

-In contrast to the INS the AHRS always depends upon external sensors as for heading (magnetic compass), for ground speed (Doppler radar or air speed indicators) or for position (ground- or satellite-based radio stations) aiding.

For the development of the proper aiding software on the basis of Kalman filtering modern AHRSs can be treated in their error models like an INS (see Section 4.6), whereby the lower sensor performance is reflected in the initial covariance matrix and the system noise matrix.

In integrated AHRS-based navigation systems available on the market one can find fixed gain control loops again, as discussed in the previous sections [Ha 89], which often are the result of a laboratory Kalman filter simulation.

Literature Chapter 6

[Ba 85] Baeumker M.: "Verfahren zur Kurswinkelbestimmung mittels und zur automatischen Kalibration eines in einem Luftfahrzeug fest montierten Dreiachsen-Magnetometers."
Europäische Patentschrift, Veröffentlichungsnummer 0 226 653 B1, Anmeldetag: 20.12.85

[El 91] Elwell, J.: "Progress in Micromechanical Inertial Instruments."
Proceedings Symposium Gyro Technology 1991, DGON and Universitaet Stuttgart 1991

[Ha 89] Handrich, E. et al: "Fiber Optic Gyro Strapdown System."
Proceedings Symposium Gyro Technology 1989, DGON and Universitaet Stuttgart 1989

[Ma 92] Mattisek, A., Baeumker, M.: "Integration of a Fiber Optical Gyro Attitude and Heading Reference System with Differential GPS."
Proceedings Symposium Gyro Technology 1992, DGON and Universitaet Stuttgart 1992

[St 90] Stieler, B.: "Verfahren zur genauen Vermessung räumlicher Winkel, Trajektorien, Konturen und Bewegungsvorgängen sowie Schwereanomalien mit Kreiseln und Inertialsystemen."
Deutsches Patent Nr. DE 40 29 215 C 2, Anmeldetag: 14.09.1990

Appendix C

Coordinate Frames for Inertial Navigation

C1 Introduction

This appendix is written by the authors of Chapters "Navigation Coordinate Frames" [Ku 94] and "Inertial Navigation" in order to harmonize the coordinate frames with the practical requirements of inertial navigation and to introduce notations used in this chapter. Notes in { } refer to [Ku 94] and those in ( ) to this chapter.

C2 The Inertial, the Earth-Fixed and the Navigational Coordinate Frames

For inertial navigation, the required reference is the inertial coordinate frame in which the earth revolves around the sun. This yearly motion takes place in the Earth Centered Inertial {ECI} or the Conventional Inertial System {CIS} coordinate frame (index i) and is defined/based on the Fundamental Katalog {FK} 5 system [Ku 94, Section 2].

For navigation in a geodetic system, the required reference is the Earth Centered Earth Fixed {ECEF} coordinate system or the Conventional Terrestrial System {CTS} It is called earth-fixed coordinate frame (index e) in this chapterdefining positions in a three-dimensional frame on the surface of the earth and its adjoining space [Ku 94, Section 5].

At time zero, the basic Earth Centered Inertial {ECI} (index i) system and the Conventional Terrestrial System {CTS} (index e) are assumed to be coincident. Currently, the zero epoch is J2000.0, which is defined to start at noon on 1 January 2000 [Ku 94, Section 3].

The rigorous transformation from ECI to CTS frames are described in [Ku 94, Section 4]. However, for the purpose of navigation and the accuracies involved, we can make certain approximations.

To achieve the above, the matrix [D] for precession, matrix [C] for astronomic nutation, and the matrix [A] for polar motion defined in Chapter 1 can be omitted. This approximates the relationship between the X1 and X5 frames of [Ku 94, Section 4] to:

X5 = [B] X1 with [B] according to Table 7 and Λ according to [Ku 94, Table 1].

In this chapter we use the orientation of the i- and the e- frames as shown in Fig. C1, i.e. with their origin in the center of the ellipsoid, the x-axis pointing to geodetic north and the y- and z-axes in the equatorial plane. At t = 0 the z-axes of both systems are pointing to the Greenwich meridian. For inertial navigation we reduce the rotation angle Λ contained in [B] by $\Lambda = \Omega \cdot t$ with:

(C 1) $\qquad \Omega = 7.2921158553 \cdot 10^{-5}$ rad/s $\cong$ 15.041 067 deg/h.

The matrix cooresponding to [B] in [Ku 94, Table 7], for vector transformation from the e- to the i- frame is thus:

(C 2) $\qquad \mathbf{C}_{ie} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\Omega t & \sin\Omega t \\ 0 & -\sin\Omega t & \cos\Omega t \end{bmatrix}$.

The reference navigational coordinate system (index n) for inertial navigation is the local geodetic system defined in [Ku 94, Section 7.2]. Its origin is in the proof mass of the accelerometer or the origin of the inertial measurement unit (IMU). Instead of the sequence of axes east, north, up {E,N,U} used in [Ku 94] we use north, east and down (N,E,D $\cong$ x,y,z) in this chapter. The n-frame is the result of a first rotation with respect to the e-frame through the geodetic longitude about the x_c-axis, a second rotation through the geodetic latitude ϕ about the y_n-axis and a translational motion through the geodetic height h to the n-frame origin mentioned above. The matrix corresponding to $R_x(90-\Phi) \cdot R_z(\lambda+90)$ in [Ku 94, Eq. 24] for vector transformation from the n- to the e-frame is:

(C 3a) $\qquad \mathbf{C}_{en} = \begin{bmatrix} \cos\phi & 0 & -\sin\phi \\ -\sin\phi \, \sin\lambda & \cos\lambda & \cos\phi \, \sin\lambda \\ \sin\phi \, \cos\lambda & \sin\lambda & \cos\phi \, \cos\lambda \end{bmatrix}$.

For $\mathbf{C}_{in}$ replace in this relationship the geodetic longitude λ by the celestial longitude:

(C 4) $\qquad \Lambda = \lambda + \Omega\, t$.

(C 3b) $\qquad \mathbf{C}_{in} = \begin{bmatrix} \cos\phi & 0 & -\sin\phi \\ -\sin\phi \, \sin\Lambda & \cos\Lambda & \cos\phi \, \sin\Lambda \\ \sin\phi \, \cos\Lambda & \sin\Lambda & \cos\phi \, \cos\Lambda \end{bmatrix}$.

The actual coordinate frame for inertial navigation is the horizon or the local level system [Ku 94, Section 7.4]. Due to the "deflection of the vertical" gravity in the n-frame has small horizontal components:

(C 5) $\qquad \mathbf{g}_n = g(\phi) \cdot (\xi \quad -\eta \quad 1)^T$

which, when left uncompensated in the computer, are nulled during the initial alignment of the inertial navigation system (INS). The true rotation of the actual navigational frame with respect to the e-frame is thus:

Reference Coordinate Frames

x ,y ,z inertial frame i

x_e ,y_e,z_e earth-fixed frame e

x_n,y_n,z_n navigational frame n

present point of
measurement

Greenwich
meridian

local meridian

ω^{ie}

x,x_e

$\Omega = |\omega^{ie}|$

N,x_n

E,y_n

D

P

C

D,z_n

φ

φ

O

B

y

Ωt

λ

E

Λ

$-z$

$-z_e$

Earth Rate and
Transport Rate in
the n-Frame

x_n

$\Omega \cos \varphi$

Ω

φ

$\Omega \sin \varphi$

z_n

$$\dot{\varphi} = \frac{V_N}{R_N + h}$$

$$\dot{\lambda} = \frac{V_E}{(R_E + h)\cos \varphi}$$

$$\dot{\Lambda} = \Omega + \dot{\lambda}$$

Fig. C 1 The Inertial, the Earth-Fixed and the Navigational Coordinate Frames

(C 6a,b) $\phi' = \phi - \xi,$ $\lambda' = \lambda + \eta \cos \phi.$

In conventional inertial navigation systems with 1 NM/h position error growth these effects are neglected.

The vector for the sum of earth rate $\boldsymbol{\omega}_n^{ie}$ and transport rate $\boldsymbol{\omega}_n^{en}$ is in the n-frame:

(C 7) $\boldsymbol{\omega}_n^{in} = \boldsymbol{\omega}_n^{ie} + \boldsymbol{\omega}_n^{en}$
with

(C 8) $\boldsymbol{\omega}_n^{ie} = \Omega \cdot (\cos \phi \quad 0 \quad -\sin \phi)^T$

(C 9) $\boldsymbol{\omega}^{en} = (\dot{\lambda} \cdot \cos \phi \quad -\dot{\phi} \quad -\dot{\lambda} \cdot \sin \phi)^T$

$\qquad = [V_E/(R_E + h) \quad -V_N/(R_N + h) \quad -V_E \tan\phi/(R_E + h)]^T,$

with $R_N = \{R_M\}$ and $R_E = \{R_N\}$ the radii of curvature of the reference ellipsoid in the meridian (N-S) and of prime vertical plane [Ku 94, Section 6.3.1].

C3 The Body-Fixed Coordinate Frame

The body-fixed coordinate frame {YPR} (index b), i.e. the coordinate frame fixed to the aircraft and relates with the navigational or the local level system {E"N"G} through attitude and heading {Y as the yaw angle, P as the pitch angle and R as the roll angle}. In this chapter we use the international standards of flight mechanics [Std 70] (s. Fig. C2) with the Euler angles defined by the following sequence of rotations from the navigational into the body frame, i.e.:

- the first rotation about the z_n and z_b axis (both are parallel in the beginning) through the yaw angle Ψ,
- the second rotation about the y_b axis lying still in the horizontal plane through the pitch angle Θ and
- the third rotation about the x_b axis through the roll angle Φ.
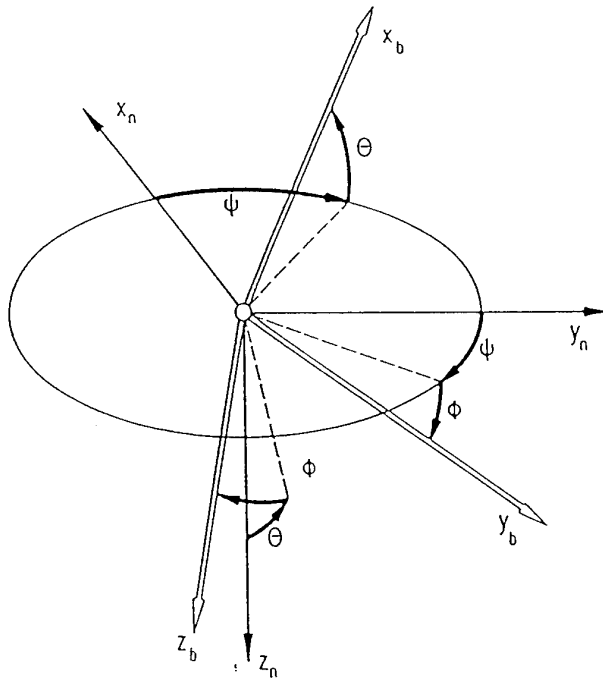

Figure C 2 The Body-Fixed Frame (Index b), the Local Geodetic or Navigational Frame (Index n) and the Euler Angles for Aircraft Roll (Φ), Pitch (Θ) and Yaw (Ψ)

The matrix C_{nb} for vector transformation from the b- to the n-frame is composed of 3 transformation matrices

(C 10a,b,c) $C_{n\,b} = C(\Phi) \cdot C(\Theta) \cdot C(\Psi).$

$$= \begin{bmatrix} \cos\Theta\,\cos\Psi & \sin\Phi\,\sin\Theta\,\cos\Psi & \cos\Phi\,\sin\Theta\,\cos\Psi \\ & -\cos\Phi\,\sin\Psi & +\sin\Phi\,\sin\Psi \\ \cos\Theta\,\sin\Psi & \sin\Phi\,\sin\Theta\,\sin\Psi & \cos\Phi\,\sin\Theta\,\sin\Psi \\ & +\cos\Phi\,\cos\Psi & -\sin\Phi\,\cos\Psi \\ -\sin\Theta & \sin\Phi\,\cos\Theta & \cos\Phi\,\cos\Theta \end{bmatrix} = \begin{bmatrix} C_{11} & C_{12} & C_{13} \\ C_{21} & C_{22} & C_{23} \\ C_{31} & C_{32} & C_{33} \end{bmatrix} .$$

In platform systems the 3 Euler angles are measured at the gimbal axes with angle encoders. In strapdown systems they are computed from the elements of the C_{nb} matrix:

(C 11a,b,c) $\Theta = -\sin^{-1} C_{31}$

$\Phi = \tan^{-1} C_{32}/C_{33}$

$\Psi = \tan^{-1} C_{21} /C_{11}.$

C4 The Wander Azimuth Coordinate Frame

The wander azimuth coordinate frame (subscript "a", s. Fig. C 3) or the generalized local geodetic system [Ku 94, Section 7.3] is rotated about the vertical with respect to the navigational coordinate frame through the "wander angle α". The angular rotation vector between both frames is:

(C12) $\omega_a^{na} = \omega_n^{na} = (0 \quad 0 \quad \dot{\alpha})^T.$

geographic-   wander angle-

coordinate system
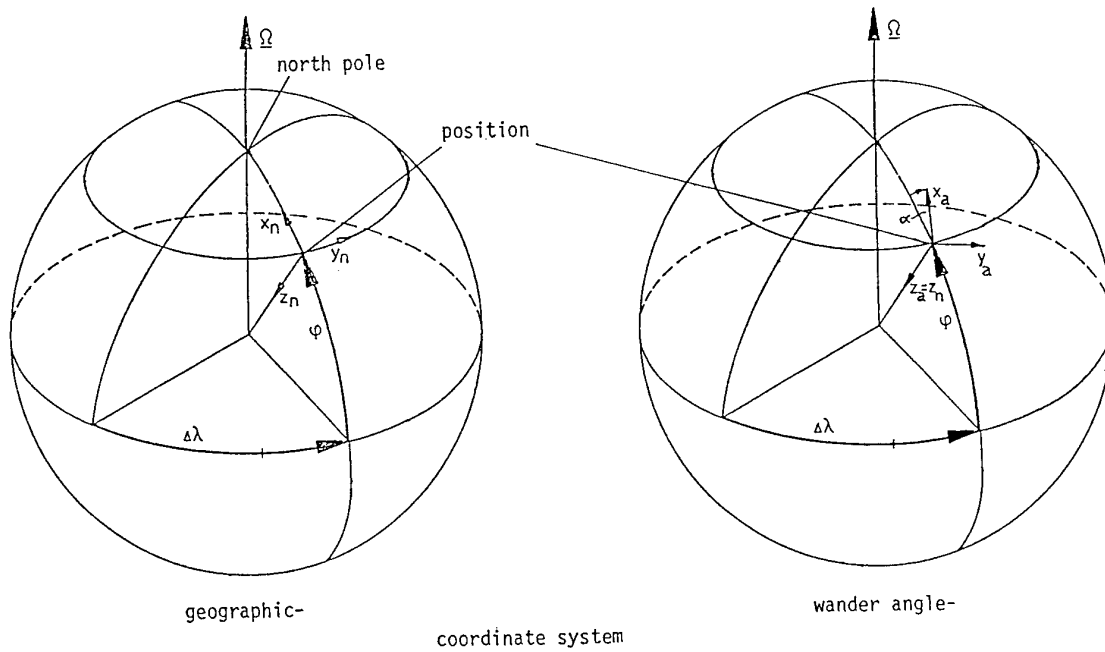
Figure C 3 The Navigational and the Wander Azimuth Coordinate Frames

The matrix for vector transformation from the a- to the n-frame is:

(C 13) $C_{na} = C_{an}^T = \begin{bmatrix} \cos\alpha & \sin\alpha & 0 \\ \sin\alpha & \cos\alpha & 0 \\ 0 & 0 & 1 \end{bmatrix}.$

Earth rate in the a-frame is thus:

(C 14) $\omega_a^{ie} = C_{an} \omega_n^{ie} = (\ \Omega\cos\phi\ \cos\alpha \quad -\Omega\cos\phi\ \sin\alpha \quad -\Omega\sin\phi)^T.$

The wander azimuth rate with respect to the earth, i.e. the transport rate is:

(C 15) $\omega_a^{ea} = \omega_a^{en} + \omega_a^{na} = C_{an} \omega_n^{en} + (0 \quad 0 \quad \dot{\alpha})^T =$

$$\begin{bmatrix} V_y \left(\dfrac{\cos^2\alpha}{R_E} + \dfrac{\sin^2\alpha}{R_N} \right) + V_x \dfrac{\sin 2\alpha}{2} \dfrac{R_N - R_E}{R_N R_E} \\ -V_x \left(\dfrac{\cos^2\alpha}{R_N} + \dfrac{\sin^2\alpha}{R_E} \right) - V_y \dfrac{\sin 2\alpha}{2} \dfrac{R_N - R_E}{R_N R_E} \\ 0 \end{bmatrix}$$

where $V_{x,y}$ are the ground speed components in the a-frame:

(C 16) $\mathbf{V}_a = (V_x \quad V_y \quad -\dot{h})^T = \mathbf{C}_{an}\mathbf{V}_n$

and $R_N = \{R_M\}$, $R_E = \{R_N\}$ are again the radii of curvature of the reference ellipsoid in the meridian and the prime vertical plane [Ku 94, Section 6.3.1].

The matrix $\mathbf{C}_{ea}$ for vector transformation from the a- to the e frame is essential for the computation of position λ, ϕ and the north direction α:

(C 17) $\mathbf{C}_{ea} = \mathbf{C}_{en} \cdot \mathbf{C}_{na} =$

$$
\begin{bmatrix}
\cos\phi\,\cos\alpha & -\cos\phi\,\sin\alpha & -\sin\phi \\[4pt]
\begin{array}{l}-\sin\phi\,\sin\lambda\,\cos\alpha \\ +\cos\lambda\,\sin\alpha\end{array} & \begin{array}{l}\sin\phi\,\sin\lambda\,\sin\alpha \quad -\cos\phi\,\sin\lambda \\ +\cos\lambda\,\cos\alpha\end{array} & \\[10pt]
\begin{array}{l}\sin\phi\,\cos\lambda\,\cos\alpha \\ +\sin\lambda\,\sin\alpha\end{array} & \begin{array}{l}-\sin\phi\,\cos\lambda\,\sin\alpha \quad \cos\phi\,\cos\lambda \\ +\sin\lambda\,\cos\alpha\end{array} &
\end{bmatrix}
=
\begin{bmatrix}
C_{11} & C_{12} & C_{13} \\
C_{21} & C_{22} & C_{23} \\
C_{31} & C_{32} & C_{33}
\end{bmatrix}
.
$$

Equating the corresponding elements we obtain:

(C 18 a to c) $\sin\phi = -C_{13}$, $\tan\lambda = -C_{23}/C_{33}$, $\tan\alpha = -C_{12}/C_{11}$.

Although singularities do exist in the computation of longitude λ and the wander angle α at the geographic poles, they do not exist for the computation of the matrix $\mathbf{C}_{ea}(t)$ so that inertial navigation is assured on a polar flight and λ as well as α are again correctly computed shortly after passage of a pole.

C5 The Actual Navigational Coordinate Frame
Including Errors of Platform Slewing of Transformation Matrix Computation

The navigational coordinate frame (s. Section C 2) is the reference frame for navigation with respect to the earth. In the North Indicating Platform INS (NIS) the platform is slewed to this direction (s. Fig. 1.1 and Section 2.2). In strapdown systems it is a computed reference frame (s. Fig. 1.2 and Section 2.4). In both cases we have to assume a small angle misalignment of the platform (index p) or the computed navigational coordinate frame (index n*) about all three axes:

(C 19) $\boldsymbol{\varepsilon} = (\varepsilon_N \quad \varepsilon_E \quad \varepsilon_D)^T$,

i.e. the small angle vector $\boldsymbol{\varepsilon}$ contains as its components the misalignments about the north, east and down axes as indicated in Fig. C 4. The small angle transformation matrix $\mathbf{C}_{np}$ in case of the platform system or $\mathbf{C}_{nn^*}$ in case of the computed frame can be written as:

(C 20a) $\mathbf{C}_{np} = \mathbf{C}_{nn^*} = (\mathbf{I} + \mathbf{E})$,

with $\mathbf{I}$ = unity matrix and $\mathbf{E}$ = skew symmetric containing the misalignment angles:

(C 20b) $\mathbf{E} = \begin{bmatrix} 0 & -\varepsilon_D & \varepsilon_E \\ \varepsilon_D & 0 & -\varepsilon_N \\ -\varepsilon_E & \varepsilon_N & 0 \end{bmatrix} = \boldsymbol{\varepsilon} \times .$

The right hand side of the relationship indicates that $\mathbf{E}$ may also be regarded as the vector cross product of $\boldsymbol{\varepsilon}$.

From both relationships it is obvious that:

(C 20c) $\mathbf{C}_{pn} = \mathbf{C}_{np}^T = \mathbf{C}_{n^*n} = \mathbf{C}_{nn^*}^T = (\mathbf{I} - \mathbf{E})$.
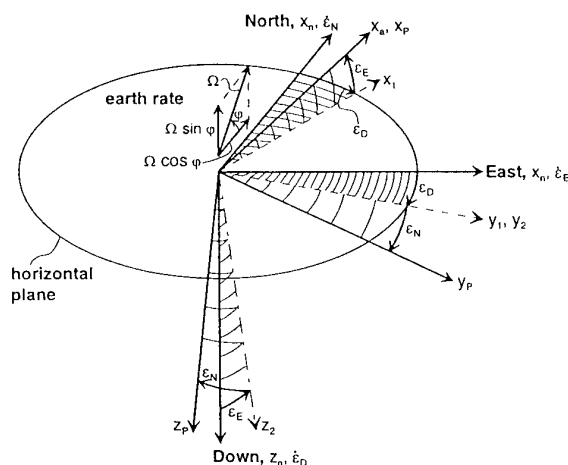

Fig. C 4 Misalignment Angles with Respect to the Navigational Frame

Literature Appendix C

[Ku 94] Kumar, M.: Chapter "Navigation Coordinate Frames" of this AGARDograph

[Std 70] "Flugmechanik, Begriffe, Benennungen, Zeichen, Grundlagen LN 9300."
Beuth Vertrieb GmbH, Köln, 1970

Appendix D

Digital Readout of Inertial Sensors and
Strapdown Algorithms for the "Analytic Platform"

D1 Introduction

In so-called "strapdown systems" the gyros and accelerometers are rigidly mounted to the vehicle and their output signals are processed in the computer for attitude, heading, ground speed and position computation. The sensor delivers analog signals in general which are digitized and sampled at time intervals $n \cdot T^S$ (T^S = sampling time increment) for further processing in the computer.

Three different possibilities for digitizing the analog readout signal of inertial sensors are discussed in this section under the aspect of subsequent high accurate integration. The digital integration process is the basis of attitude and heading computation and inertial navigation. Its accuracy must be adapted to that of the measurement process. For the reference direction computation with "inertial quality" with < .01 deg/h uncertainty, for instance, the integration error should be lower by at least 1 order of magnitude. This makes it understandable that the following considerations apply primarily to the gyro measurements.

The three digitizing methods discussed are:
- the analog-to-digital converter (ADC), s. Fig. D 1,
- the voltage-to-frequency converter (VFC), s. Fig. D 2,
- and the pulse rebalance loop (PRL), s. Fig. D 3.

Not only the digitizing method is of importance for the subsequent integration process, but also the computation speed, i.e. the sampling time and closely connected to it the algorithms for integrating the gyro and accelerometer readings into the computed spatial reference direction, velocity and position. They are called "strapdown algorithms". This aspect will be briefly discussed at the end of this section.

D2 Analog-to-Digital Conversion (ADC)

The ADC converts the analog signal into a whole number digital signal with a resolution limited by the "least significant bit (LSB)". Table D 1 lists some data of a high-accuracy ADC and Fig. D 1 shows the block diagram for a gyro with analog readout and AD conversion. It also includes the subsequent sampling and the scaling by S', which comprises the gyro and ADC scalefactor. The output signal ω^* is a whole number representation of the input signal ω, but contains the sensor errors and the ADC errors listed in Table D 1 plus the noise due to the limited resolution. Because of the roundoff process, information of LSB/2 may get lost.

Table D 1 Data of the High Accuracy Analog-to-Digital Converter Type DAS 1153 of Analog Devices [AD 90]

Resolution n:	16 Bits (negative sign with two's complement)
Throughput Rate:	25 kHz min
Nonlinearity:	± 30 ppm full scale range (FSR)
Nonlinearity Temperature coefficient (TC):	± 2ppm/K max
Gain TC:	± 8ppm/K max
Zero TC:	±80 μV/K max
Power Supply Sensitivity:	±10 ppm FSR/%V

For ω_{max} = 500 deg/s the peak value of this quantization error (uncorreleted noise) is:

(D 1) $\qquad \delta\omega_{max} = \pm LSB/2 = \pm \omega_{max}/2^n \quad (= \pm 55 \text{ deg/h });$

The figures in ( ) refer to the hardware examples.

The digital angular rate roundoff noise has zero mean and is uniformly distributed over the range of ± LSB/2. Its standard deviation is (s. [Be 66], Chapter 7):

(D 2) $\qquad \sigma_{\delta\omega} = \pm LSB/\sqrt{12} = \pm .289 \cdot LSB \quad (= \pm 32 \text{ deg/h}).$

In order to derive the time dependence of the so-called "random walk" angle, i.e. growth of the standard deviation of the integral of the angular rate roundoff noise over time, we assume a first order integration algorithm whereby the angular rate ω^* sampled at time intervals T^S delivers the angle increment of $\Delta\phi^* = \omega^* \cdot T^S$ for each cycle. The standard deviation of one integrated angle error increment $\Delta\varepsilon$ is obtained in the same way: $\sigma_{\Delta\varepsilon} = \sigma_{\delta\omega} \cdot T^S$. As the $\Delta\phi^*$s add up to the total angle so do the angle error increments $\Delta\varepsilon$s add up to the total error ε. Since the $\Delta\varepsilon$s are uncorrelated from cycle to cycle, their variances have to be summed up over the integration time. With $n = t/T^S$ the standard deviation of the integrated angle error is:

(D 3) $\qquad \sigma_\varepsilon = (\sum_{i=1}^{n} \sigma_{\Delta\varepsilon,i}^2)^{1/2} = \sigma_{\Delta\varepsilon} \cdot \sqrt{n} = LSB \cdot \sqrt{T^S/12} \quad (= 77.8 \text{ deg}/ \sqrt{h} \cdot \sqrt{t[h]}).$

This result holds for all integrated ADC output signals. It also holds for the first order integration process in any digital computer, where in critical cases, as with the strapdown algorithms discussed below, the word length has to be properly adapted. This characteristic is called "random walk", whose uncertainty increases with $t^{1/2}$. We define as corresponding "random walk coefficient":

(D 4) $\qquad r_{ADC} = r_{first \ order \ integration \ process} = LSB \cdot \sqrt{T^S/12} \quad (= 77.8 \text{ deg}/ \sqrt{h}).$

Applied to the ADC-digitized gyro signal, we obtain for an integration cycle of 50 Hz, which is a common value in strapdown inertial navigation systems, a random walk coefficient of .075 deg/$\sqrt{h}$. Since this value is intolerable for inertial navigation, the ADC is not recommendable for this application. This stochastic error is avoided by using one of the other two digitizing methods
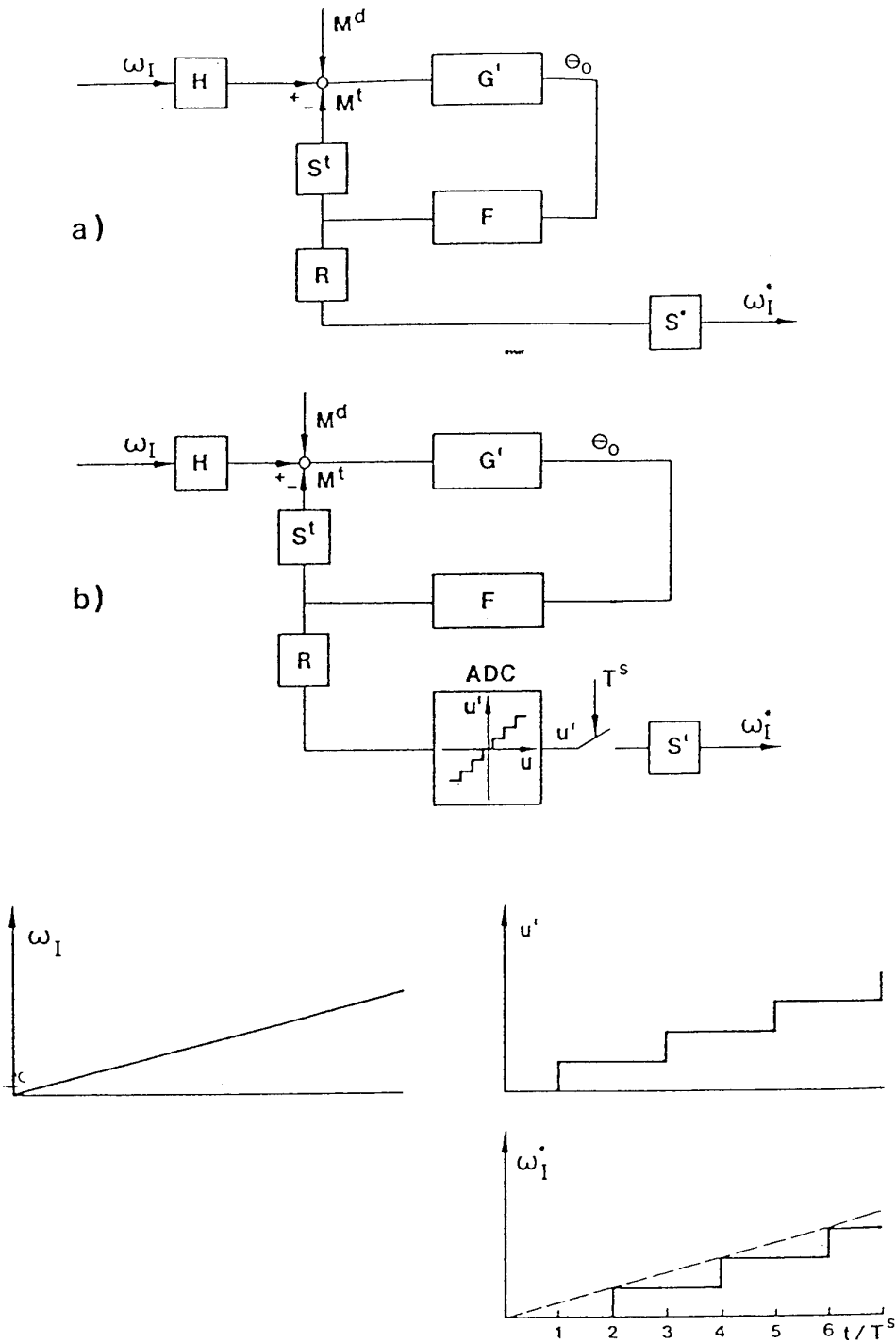
Fig. D 1 Block Diagrams and Signals for Analog (a) and Digital Gyro Readout (b) Using an Analog-to-Digital Converter (ADC)
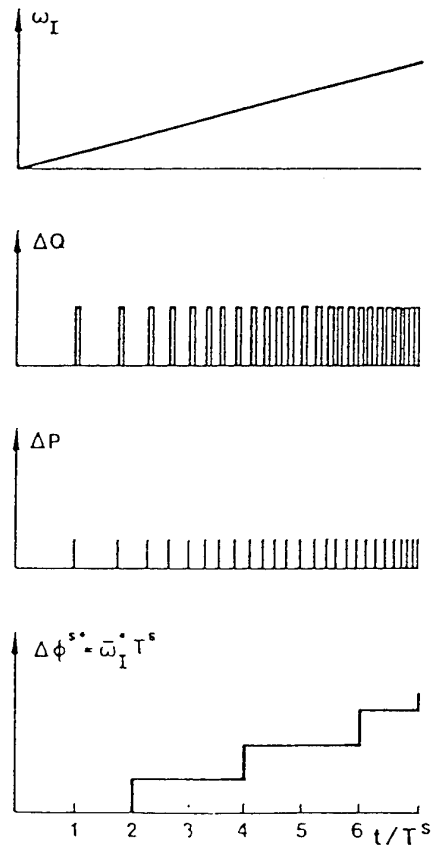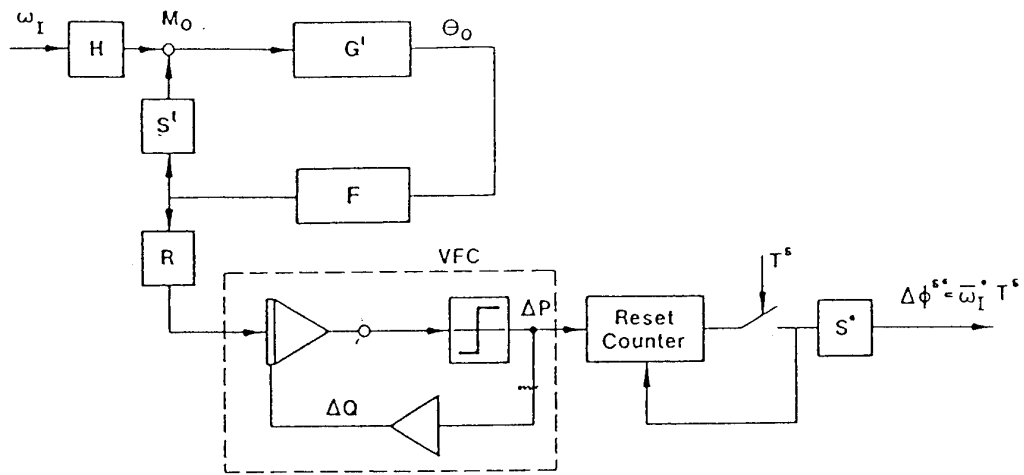
Fig. D 2 Block Diagram and Signals for Digital Gyro Readout Using a Voltage-to-Frequency Converter (VFC)

ω_I — H — M_O — G' — Θ_O

M_t

S^t

ΔQ — Quantizer — F

ΔP

T^s

Reset Counter — $S^\cdot$

$\Delta\phi^{s\cdot} = \bar{\omega}_I^\cdot T^s$

Pulse Frequency Modulation

ω_I

ΔQ

ΔP

T^P

$\Delta\phi^{s\cdot} = \bar{\omega}_I^\cdot T^s$

1 2 3 4 5 6 t/T^s

Binary Pulse Width Modulation

ω_I

ΔQ

$2T^P$

switching of current pulse ΔQ synchronized with voltage pulse ΔP

ΔP

$\Delta\phi^{s\cdot} = \bar{\omega}_I^\cdot T^s$
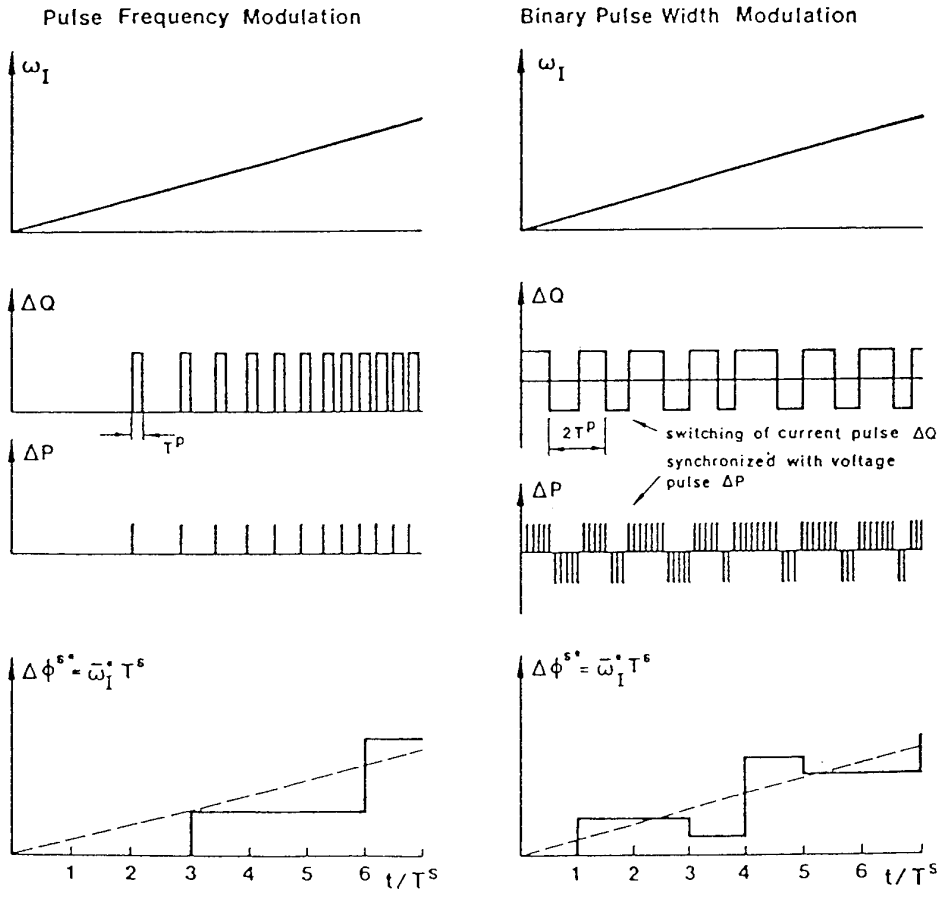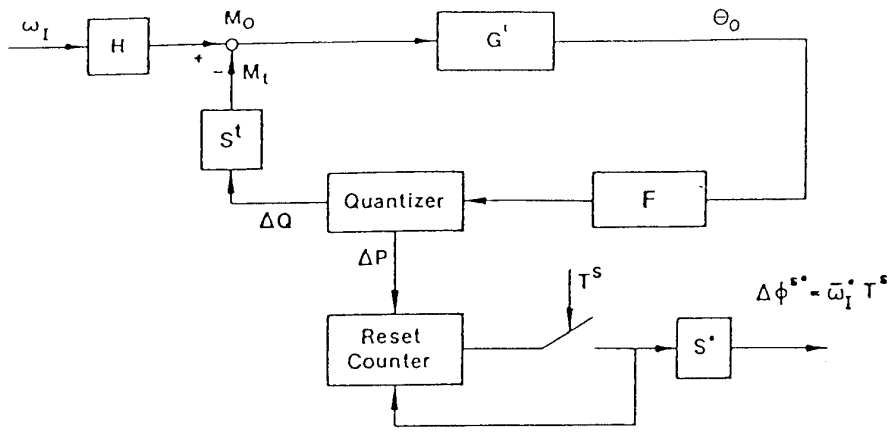
1 2 3 4 5 6 t/T^s

Fig. D 3 Block Diagram and Signals for Digital Gyro Readout Using a Pulse Rebalance Loop (PRL)

discussed below.

In Fig. D 4 simulation results are plotted proving the theory derived above. An ADC with LSB $= 10^{-3}$ receives as input signal a sine wave and the difference between the truncated and the exact sine wave was integrated over a time of 300 s. Ten runs were simulated with different initial conditions for the input signal. The single input sample function shows the computed mean value and standard deviation as well as the theoretical standard deviation according to Eq. D 2. The ten output sample functions are plotted together with their $\pm 1\ \sigma$ bands computed for fixed times. The theoretical standard deviation according to Eq. D 3 can thus be compared with the calculated one - the correspondance is satisfying if it is kept in mind that it is based on ten sample functions only.
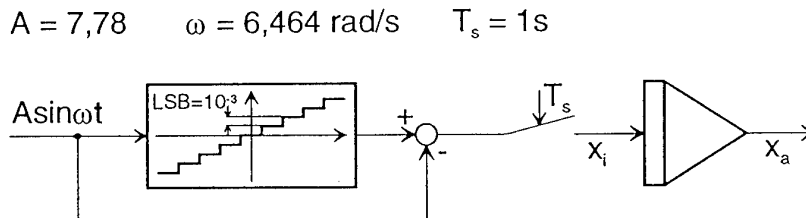
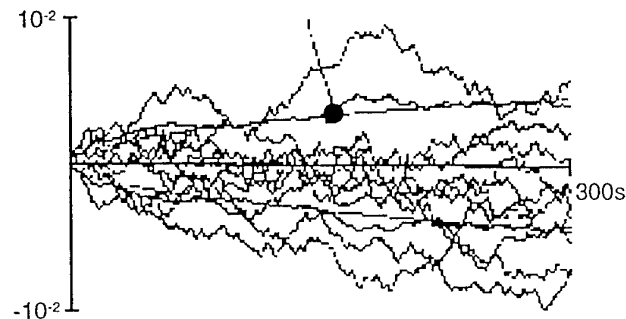$$A = 7{,}78 \qquad \omega = 6{,}464 \text{ rad/s} \qquad T_s = 1s$$
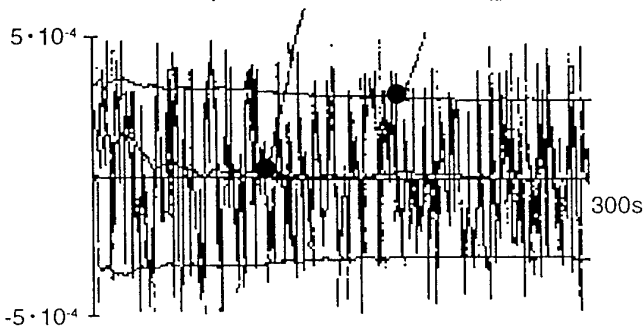
Fig. D 4 Simulation Results for the Random Walk Error in the Integrated Output of an Analog-to-Digital Converter

D3 Voltage-to-Frequency Conversion (VFC)

In the VFC the analog output voltage of a gyro is converted into a pulse train, whereby each pulse ΔP carries the same information proportional to the input voltage. In the case of a gyro ΔP it is an angle increment $\Delta \Phi^P$. Its weight is derived from the ratio of voltage to frequency:

(D 5a) $\Delta P = u_{max}/f_{max}$

(D 5b) $\Delta \Phi^P = \omega_{max}/f_{max}$ (= .18 arc sec for $\omega_{max} = 500\ ^o/s$ and $f_{max} = 1$ MHz).

The frequency in the numerical example equals the high-accuracy synchronous VFC whose data are listed in Table D 2.

Table D 2 Data of a High Accuracy Synchronous Voltage-to-Frequency Converter Type AD652KP/BQ of Analog Devices [AD 90]

Output Frequency:	1MHz
Initial Gain Error:	$< \pm .5\ \%$
Gain Temperature Coefficient (TC):	$< \pm 25$ ppm/K
Linearity Error:	$< \pm 50$ ppm
Offset:	$< \pm 2$ mV
Offset:	TC $< 25\ \mu V/K$
Response Time:	One period of new output frequency plus one clock period

This pulse train is fed to a buffer or reset counter which is read out at the sample time T^s by the computer and multiplied with the scalefactor for processing by the integration algorithm. As indicated in Fig. D 2, the sampled angle increment $\Delta \Phi^s$ is proportional to a mean angular rate $\bar{\omega}$.

The sampled angular increment $\Delta\Phi^s$ is also corrupted by noise, known as "quantization noise" - at each sample time increment just one pulse increment $\Delta\Phi^p$ may have got lost. In reality this information is not lost, but saved for the next sampling cycle, as can be seen from the simplified VFC block diagram in Fig. D 2. The input voltage u is integrated and the output of the integrator is fed to a comparator. Each time the output voltage of the integrator exceeds a certain threshold, an electric charge increment ΔQ of a precisely known area is released to reduce the integrator's net charge. The voltage pulse ΔP is fed to a reset counter The roundoff error remains stored in the integrator. The functioning of the VFC in Table D 2 is different, but here again no information is lost.

D4 The Pulse Rebalance Loop (PRL)

A different approach for accurate readout of inertial sensors, the so called "pulse rebalance loop (PRL)" or "incremental caging loop", is shown in Fig. D 3. A quantizer is placed into the rebalance loop releasing a current i of constant amplitude to flow through the torque generator in positive or negative directions for the pulse time increment T^p. The maximum input angular rate is thus defined by the current level i flowing in one direction only:

(D 6) $\omega_{max} = S^t\, i,$

but reduced by the fact that the current is not applied continuously but as a flow of electric charges Q, which need time to reach the maximum level. The resulting torquer scalefactor S^t is thus reduced by a certain amount, 10 % for instance.

With the so-called "binary pulse width modulation" - this will be discussed only - the current pulse is applied in the positive and negative directions during one cycle, whereby the switching of the current direction is controlled by the input signal.

Due to the inductivity of the torquer coil the pulse rebalance torquing frequency is fairly low. A common value is:

(D 7a) $f^p = 1$ kHz.

Each positive and negative pulse is measured with readout counting pulses of much higher frequency:

(D 7b) $f^r = 256 \cdot f^p,$

(the figure is taken from [Ra 77]), thus improving the resolution of one pulse to

(D 8) $\Delta\Phi^p = \omega_{max}/f^r$ (≈ 7 arc sec for $\omega_{max} = 500$ ^{o}/s).

The pulse evaluation is the same as mentioned for the VFC digitizing method.

Compared to the analog rebalance techniques and external digitizing process, the pulse width modulation torquing has the following advantages:

- the dissipated rebalance energy within the sensor is constant and not proportional to ω^2 as with the analog rebalance method employed with the other two methods; this is an optimal basis for keeping the torquer at constant temperature and thus its scalefactor at a constant value;

- it works only at two points of the torquer characteristic line (torque = function of current), which is favorable for linearity;

- the scalefactor stability depends upon the stability of the constant current source only and not upon additional error sources of the digitizing element (ADC or VFC).

Nonlinearities may arise in pulse-rebalanced sensors from eddy currents varying with the input signal. These effects are minimized by observing certain rules in the design of the torquer [Ra 77].

Fig. D 5 shows two test results for the stability of a SDF gyro (Ferranti M 2519) and an accelerometer (Litef B 250) plus pulse rebalance electronics designed at DLR [Ra 77].

The accelerometer readings were taken in 1976 during a period when earthquakes in China may have caused movements in the foundation of the laboratory. The digital accelerometer readout, compensated for these movements, has a mean slope of $3 \cdot 10^{-6}$ g/day.

The gyro measurements were taken with the axes in the optimal orientation (output axis up). The readings prove that the stability of the sensor in this position plus rebalance loop is better than 0.001 deg/h.

The pulse rebalance method may be used not only with inertial sensors but also with all sensors with compensation readout (e.g. flow and pressure sensors).

D5 Computation Cycle and Algorithms for Spatial Direction, Velocity and Position Integration from Gyro and Accelerometer Data

Due to the well known sampling theorem, the maximum frequency f_{max} contained in a signal to be sampled may only be half the sampling frequency:

(D 9) $f_{max} \leq 1/(2 \cdot T^s).$

In that case only the original signal can be recovered from the sampled signal. Once this sampling theorem has been violated, the frequency spectrum of the sampled signal is contaminated by the so-called "aliasing error", which is impossible to correct afterwards.

Fig. D 6 shows the aliasing error for the two cases that the signal was digitized by the ADC and by the VFC or PRL method. For illustration the sampling time has been chosen in both cases to be 20% longer than the sine wave period T. All sampled data contain the aliasing error, i.e. they pretend a signal of much lower frequency, whereby the amplitude of the ADC-digitized signal
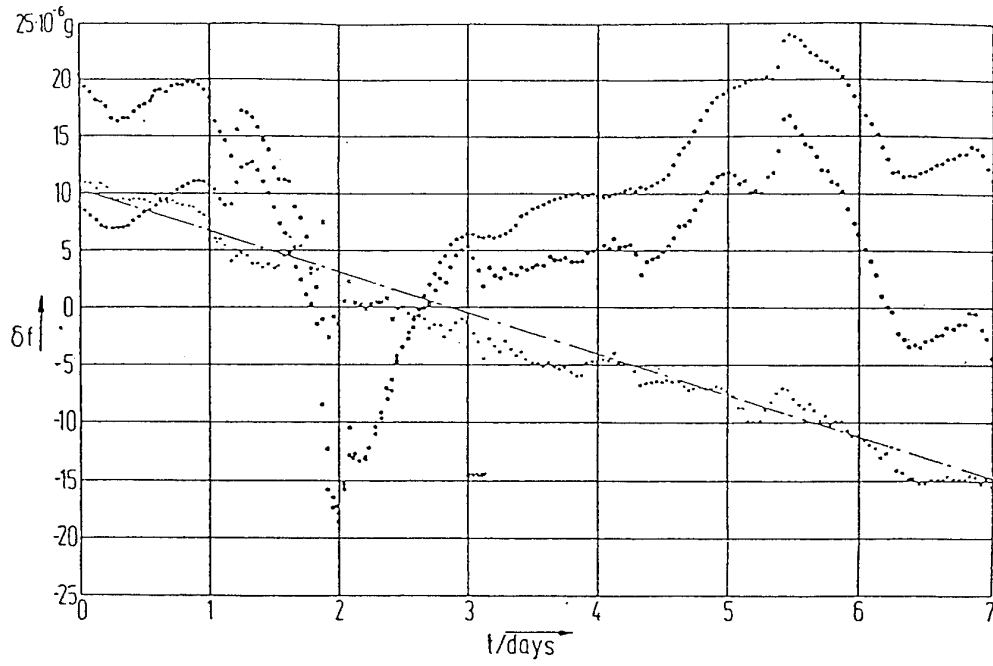
a) Test Results for Litef Accelerometer B 250

ooo uncompensated test result
... error due to inclination of foundation
xxx compensated test result
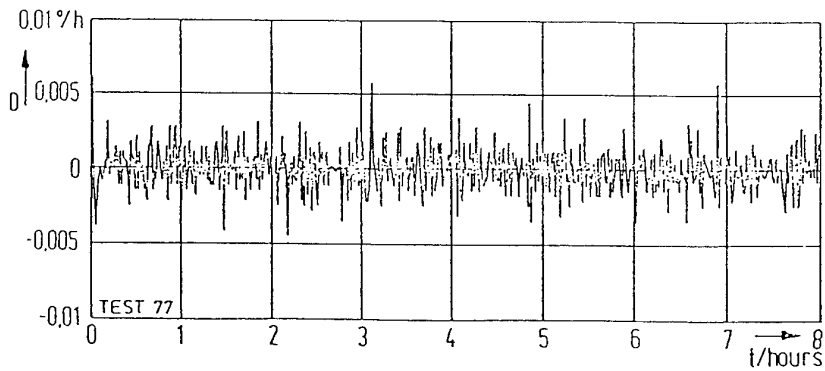
b) Test Results for Ferranti Gyro M 2519 (OA up)

Figure D 5 Test Results of a Rate Integrating Gyro Type Ferranti M 2519
and of a Pendulous Accelerometer, Type Litef B250 Including a Pulse Rebalance Loop [Ra 77]

Figure D 6 Aliasing Error Due to Sampling of an Analog Signal without Avaraging (ADC) and with Avaraging over the Sampling Period (VFC, PRL) when the Sampling Frequency is Too Low

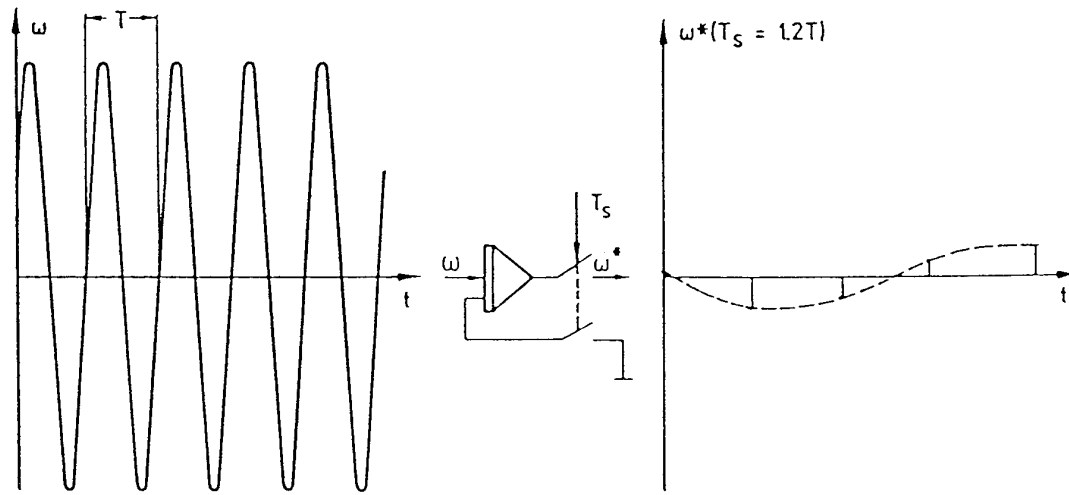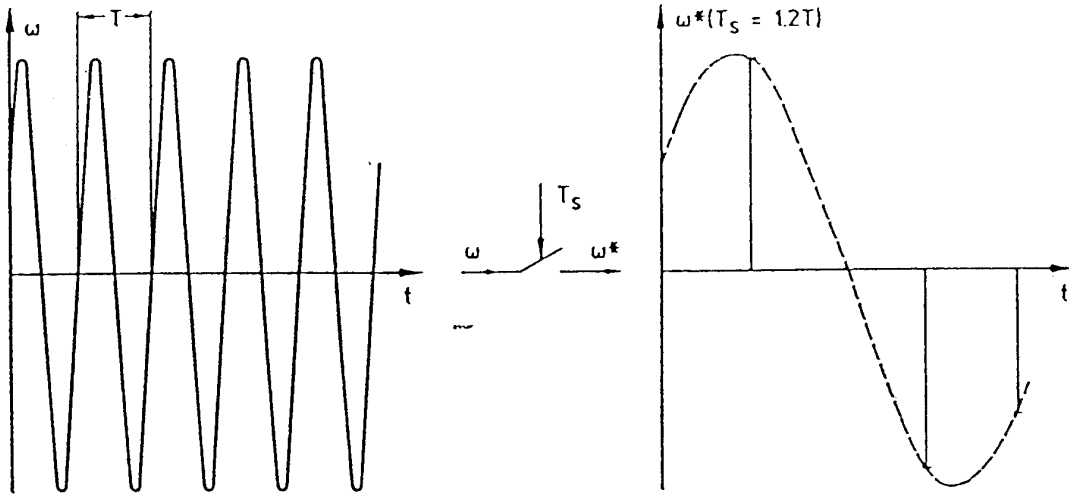is identical to the input amplitude and the VFC and PRL digitized amplitudes are much lower. The latter ones are integrating digitizing methods!

A common source for the high frequency content in the measurement signal of inertial sensors is the vibration due to the turbines of an aircraft, for instance, or due to the piston motor of vehicles. In inertial navigation systems (INS) employing ring laser gyros (RLG), the dithering mechanism of these sensors is another vibration source. The RLG is dithered with a 400 Hz frequency and an amplitude of a few arc minutes, for instance. Neither for navigation nor for vehicle control the tracking of the high frequency vibration is of importance, especially since its amplitude is fairly small in general. For inertial navigation another aspect is of highest importance and because of this the sampling theorem has to be met: the core of the navigation computation is the integration of the gyro and accelerometer measurements to the spatial direction reference, the velocity and position (s. Fig. 1.2). Algorithm errors much smaller than the "inertial quality" mentioned above are required.

Let us first concentrate on the computation of the direction reference, i.e. the computation of the C_{nb} matrix in Eq. 2.21 and Fig. 1.2.

For motions in one plane only, the angle between two directions is the integral of a single gyro measurement - regardless of the sampling time its accuracy at sampling intervals is limited by the quantization noise only. For spatial motions this is no longer true, as visualized by Fig. D 7a. The so-called "coning motion" is shown here, i.e. a spatial sinusoidal motion about the two axes x and y which are phase-shifted by $90°$. The analysis of the kinematics indicate, that the z-axis senses a continuous rotation, though the coordinate system stays inertially fixed. If these angular rotations are correctly measured by the gyros and the algorithms for computing the direction of the coordinate system are correct, the computed direction will stay inertially fixed as required. If the sampling theorem is violated or the algorithms are of low quality, the reference direction computation will drift.

The so-called "non-commutativity, indicated in Fig. 7 b is another example of the requirements for spatial rotation integration. A 1st rotation about the x-axis and a 2nd about the y-axis bring the coordinate system into a position differing from that after a reversed sequence of rotation. This shows us that in a vibrating environment the sampling theorem has to be met in a strict sense, so that the computer can accurately follow the sequence of the spatial rotation.

The problem of strapdown algorithms touched off in Section 2.4 can mathematically be visualized after partially integrating Eq. 2.26 a:

$$(D\ 10) \qquad \int_0^{t_{N+1}} \dot{C}_{ib}(t)\ dt\ =\ C_{ib}(0)\ +\ \sum_0^N n\ [C_{ib}(\tau)\cdot\Delta\Theta^{ib}(\tau)]\Big|_{t_n}^{t_{n+1}}\ -\ \int_{t_n}^{t_{n+1}} \dot{C}(\tau)\ \cdot\Delta\Theta^{ib}(\tau)\ d\tau],$$

where $\Delta\Theta^{ib}(\tau)$ is the integral of the gyro measurements. The update of the first term on the right hand side of this equation is not so problematic. It comprises the first order integration algorithm. The integral on the right hand side indicates the contribution of the higher order terms. It depends not only on the integral of the gyro measurements, but also on the rate of change of the transformation matrix, indicating that Eq. 2.26 can only be solved in the computer by means of approximations which should be accurate enough so that:

- there is no drifting of the computed reference directions,

- the computed transformation matrix is "normal", i.e. its unity vectors remain unity in magnitude and

- the computed matrix is "orthogonal", i.e. its unity vectors keep their orthogonality.

If Eq. D 10 is chosen for the computation of the C_{ib} matrix, only 3 of its elements can be derived independently of 3 gyro measurements. The other 6 can be computed by the requirements of normality and orthogonality. Due to the former the quadratic sum of one column's or row's elements must be unity:

$$(D\ 11a) \qquad C_{i1}^2\ +\ C_{i2}^2\ +\ C_{i3}^2\ =\ 1$$

$$C_{1i}^2\ +\ C_{2i}^2\ +\ C_{3i}^2\ =\ 1,$$

and due to the latter the dot product of one row and one column must vanish:

$$(D\ 11b) \qquad C_{i1}C_{j1}\ +\ C_{i2}C_{j2}\ +\ C_{i3}C_{j3}\ =\ 0$$

$$C_{1i}C_{1j}\ +\ C_{2i}C_{2j}\ +\ C_{3i}C_{3j}\ =\ 0.$$

Though the direct integration of the C_{nb} matrix according to Eqs. D 10 and D 11 is not the preferred way for integrating the gyro data, this matrix is the core of the strapdown algorithms. We have called it "analytic platform" in Section 2.4 because the acceleration measured in the body frame is transformed with its help into the navigational frame. Neglecting Coriolis and gravity compensation, the latter is:

$$(D\ 12) \qquad \dot{V}_n(t)\ =\ C_{nb}(t)\cdot\dot{V}_b(t)$$

$$(D\ 13) \qquad \int_0^{t_{n+1}} \dot{V}_n(t)\ dt\ =\ V(0)\ +\ \sum_0^N n[C_{nb}(\tau)\cdot\Delta V_b(\tau)\,]\Big|_{t_n}^{t_{n+1}}\ -\ \int_{t_n}^{t_{n+1}} \dot{C}_{nb}(\tau)\cdot\Delta V_b(\tau)\ d\tau].$$

The transformation of the accelerometer output increments into the navigational frame and their summing up is an integration algorithm of first order only neglecting higher order terms due to the integral on the right hand side of this equation. A sampling frequency adapted to the vibration environment and higher order algorithms are prerequisite for an acceleration transformation and integration with sufficient accuracy. Otherwise the so-called "sculling errors", resulting from in-phase transformation and acceleration errors as illustrated in Fig. D 8a, may arise and act like measurement signal errors.

Still another acceleration error source - the so-called "size effect" - has to be taken into account in a vibrating environment. It is not due to an algorithm error, but due to the fact that the three accelerometers have different distances to the axis of rotation. This is illustrated with the help of Fig. D 8b. Choosing in Eq. 3.5 the body-fixed coordinate frame (index n = b), neglecting earth rate and setting R = r, we may derive the following output signals of the north and down accelerometers for a vibration with Θ =

114

a) The Coning Motion

$$\omega_x = \alpha\Omega \sin\Omega t$$

$$\omega_y = \beta\Omega \cos\Omega t$$

$$\omega_z = \frac{1}{2}\alpha\beta\Omega$$

b) Non Commutativity of Rotations

initial
position

after 90
1st rotation
about :

x-axis

y-axis

after 90
2nd rotation
about :

y-axis
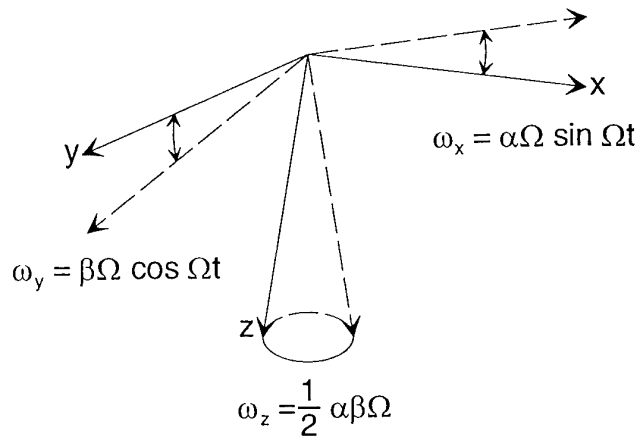
x-axis
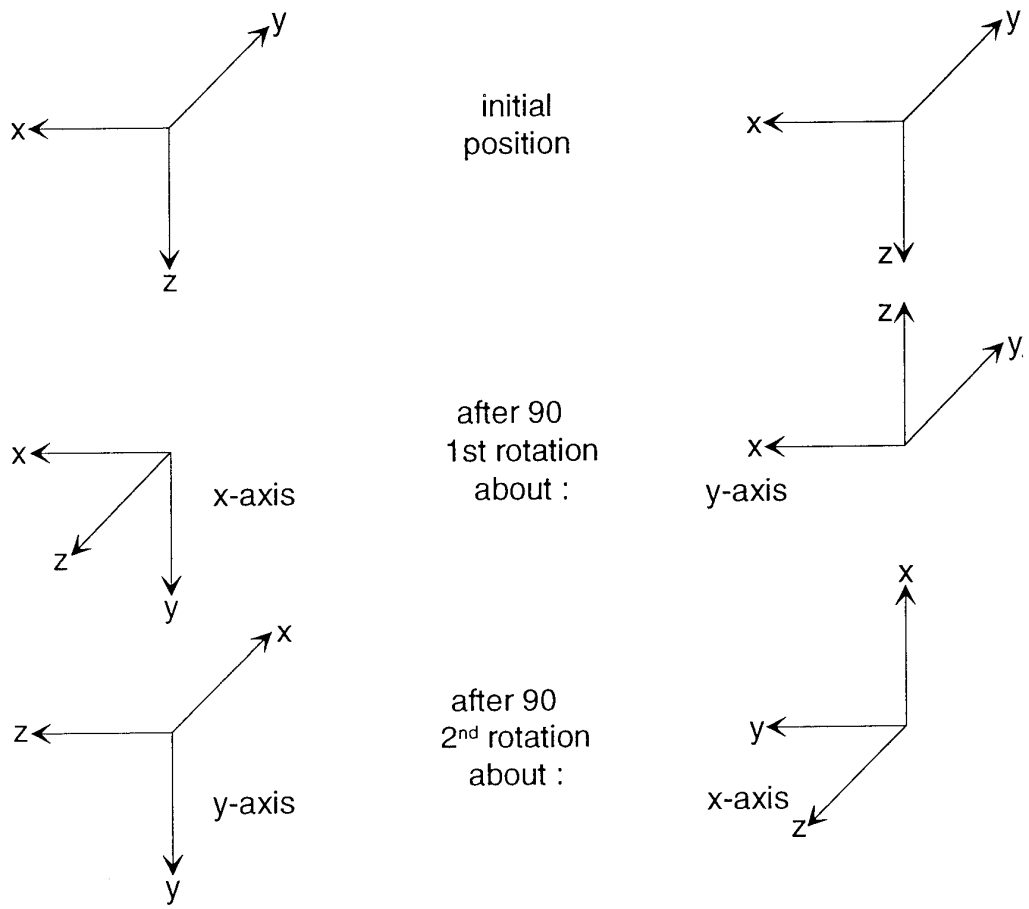
Figure D 7   The Coning Motion and an Illustration for the Non-Commutativity of Rotational Sequences

$f_0 \approx 400\,Hz$
$\alpha_0 \approx 3.6\,arc\ min$
$\omega_0 \approx 150°/s$

$\Delta\Theta\cdot$

$\Delta V\cdot$

The Sculling Error

oar

man

boat

V

sculling a boat

$\delta\Delta\Theta$

t

$\delta\Delta V$

t
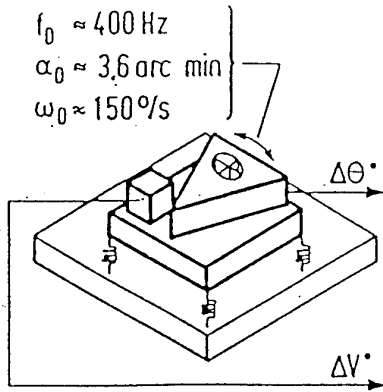
product in navigation
computer causes
rectified
sculling
error
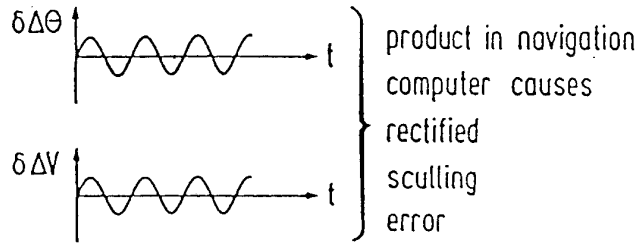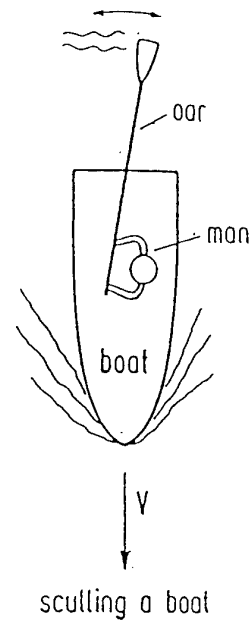
Acceleration Error Due to Size Effect

Δr^z

r^x

r^z

a_x

a_z

Fig. D 8 Acceleration Errors due to Sculling (a) and Size Effect (b)

$\hat{\Theta}$ sin ωt about the east axis:

(D 14a) $a_x = -(\dot{\Theta})^2 \cdot r^x = - r^x (\hat{\Theta} \ \omega \ \cos \ \omega t)^2$

(D 14b) $a_z = \vec{\Theta} \times r^z = (r^x + \Delta r^z) \ \hat{\Theta} \ \omega^2 \ \sin \ \omega t.$

Transformed into the horizontal north-south axis, we obtain:

(D 15) $a_N = a_x \ \cos \Theta + a_z \ \sin \Theta$

$= (\hat{\Theta} \ \omega)^2 \ r^x \ [-\cos^2\omega t + (1 + \Delta r^z/r^x) \ \sin^2\omega t].$

For $\Delta r^z = 0$ we obtain $a_N \sim \cos 2\omega t$, i.e. zero in the mean, otherwise it is rectified. It is compensated in a RLG-INS. A survey of computation and compensation problems in such a system is to be found in [Ma 82].

Two other ways of transformation matrix computation should briefly be mentioned - the integration of the Euler angles and of the so-called "quaternions".

The rate of change of the Euler angles is governed by the differential equation [Std 70]:

(D 16) $$\begin{bmatrix} \dot{\Phi} \\ \dot{\Theta} \\ \dot{\Psi} \end{bmatrix} = \begin{bmatrix} 1 & \sin\Phi \ \tan\Theta & \cos\Phi \ \tan\Theta \\ 0 & \cos\Phi & -\sin\Phi \\ 0 & \sin\Phi/\cos\Theta & \cos\Phi/\cos\Theta \end{bmatrix} \cdot \begin{bmatrix} \omega_x \\ \omega_y \\ \omega_z \end{bmatrix}^{nb} ,$$

with:

(D 17) $\omega^{nb} = \omega^{ib} - \omega^{in}$

$\uparrow \qquad \uparrow$

with   "Schuler" feedback (s. Section 4.2)

gyro measurements.

The integration of the Euler angles and the computation of the C_{nb} matrix according to Eq. 2.24 has the following properties:

Advantage: it remains orthogonal and normal;
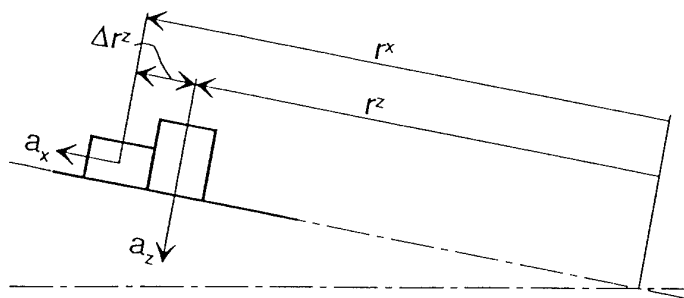Disadvantage: Eq. D 16 is highly nonlinear and has a singularity at $\Theta = \pm 90^\circ$, which corresponds to "gimbal lock" discussed in Section 2.3.

The transformation matrix computation based on the so- called "quaternions" is most commonly used. The orientation of the body-fixed frame with respect to the navigational coordinate frame can be described by one rotation through an angle δ about one axis defined with respect to the reference frame through the 3 direction cosines [Br 55] (s. Fig. D 9). These 4 parameters are comprised in the 4 quaternions, written as a column matrix:
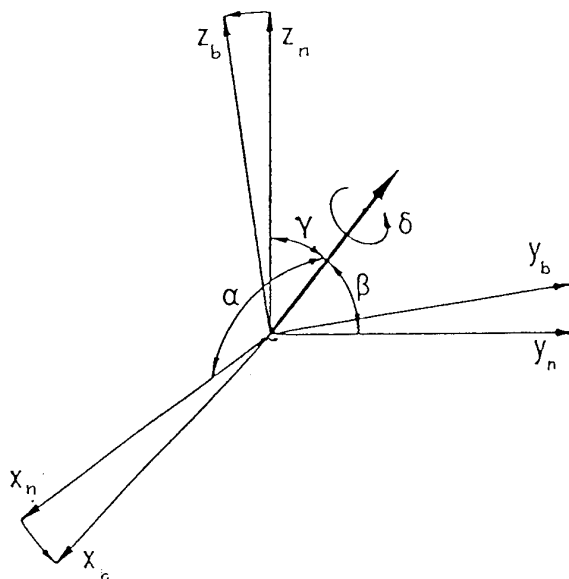

Fig. D 9 Parameters Describing Quaternions

(D 18) $\quad$ $\mathbf{q}$ $=$ $(q_0 \quad q_1 \quad q_2 \quad q_3)^T$

with $\quad$ $q_0 = \cos(\delta/2)$

$q_1 = \cos\alpha \, \sin(\delta/2)$

$q_2 = \cos\beta \, \sin(\delta/2)$

$q_3 = \cos\gamma \, \sin(\delta/2)$.

Though the vector transformation can directly be based on quaternions using special algorithms, it is carried out with the transformation matrix $\mathbf{C}_{nb}$ in the strapdown computer. $\mathbf{C}_{nb}$ is composed of the quaternion elements in the following way:

$$(D\ 19) \quad \mathbf{C}_{nb} = \begin{bmatrix} (q_0^2 + q_1^2 - q_2^2 - q_3^2) & 2(q_1 q_2 - q_0 q_3) & 2(q_1 q_3 + q_0 q_2) \\ 2(q_1 q_2 + q_0 q_3) & (q_0^2 - q_1^2 + q_2^2 - q_3^2) & 2(q_2 q_3 - q_0 q_1) \\ 2(q_1 q_3 - q_0 q_2) & 2(q_0 q_1 + q_2 q_3) & (q_0^2 - q_1^2 - q_2^2 + q_3^2) \end{bmatrix}.$$

The quaternion update is based on the differential equation:

(D 20) $\quad$ $\dot{\mathbf{q}} = \mathbf{Q} \cdot \boldsymbol{\omega}'/2$,

where:

(D 21) $\quad$ $\boldsymbol{\omega}' = (\ 0 \quad \omega_x \quad \omega_y \quad \omega_z\ )^T$

and

$$(D\ 22) \qquad \mathbf{Q} = \begin{bmatrix} q_0 & -q_1 & -q_2 & -q_3 \\ q_1 & q_0 & -q_3 & q_2 \\ q_2 & q_3 & q_0 & -q_1 \\ q_3 & -q_2 & q_1 & q_0 \end{bmatrix}$$

When initially both coordinate frames b and n coincide and the Euler angles (Φ_0, Θ_0, Ψ_0) are zero, the quaternions are:

(D 23) $\quad$ $\mathbf{q} = (\ 1 \quad 0 \quad 0 \quad 0)^T$.

For nonzero initial Euler angles the initial quaternions are derived by means of a successive rotation with Ψ_0 (i = 1), Θ_0 (i = 2) and Φ_0 (i = 3), using Eqs. D 20 to 23:

(D 24) $\quad$ $\mathbf{q}(i) = \mathbf{Q}(i\text{-}1) \cdot \Delta\mathbf{q}(i,\ i\text{-}1)$,

where $\Delta\mathbf{q}(i, i\text{-}1)$ is taken from Eq. (D 18). For the rotation with Ψ_0, when $\alpha = \beta = 90^\circ$ and $\gamma = 0$ it reads, for instance:

(D 25) $\quad$ $\Delta\mathbf{q}(1,\ 0) = (\cos(\Psi_0/2) \quad 0 \quad 0 \quad \sin(\Psi_0/2))^T$.

The transformation matrix computed with quaternions is always orthogonal. The normalisation is based on the fact that the so-called "quaternion norm" has to be unity ($q_0^2 + q_1^2 + q_2^2 + q_3^2 = \sum q_i^2 = 1$), its approximation reading:

(D 26) $\quad$ $\mathbf{q}_{new} = \mathbf{q}_{old} \, /[\sum q_i^2\ ]^{1/2}$.

The most severe environment for strapdown algorithms is within a ring laser gyro INS, where the whole inertial measuring unit (IMU) is vibrating due to the dithering of the RLGs with a few arc minutes amplitude and 400 Hz frequency. The sampling frequency in these systems has to be 3 to 4 times this dither frequency, i.e. it is above 1 kHz. The transformation matrix computation and the velocity transformation are in general split up into 2 parts [Sa 84]. The results of a high frequency (> 1 kHz) precomputation are the corrected angle and velocity increments. At a lower frequency (50 to 100 Hz) the integration of the transformation matrix and the acceleration transformation are carried out. This can be seen from Fig. D 10, the functional diagram of the Litton LTN-90 INS.

References [Mc 68, We 78, Sa 84] give a good insight into the problems of strapdown algorithms.
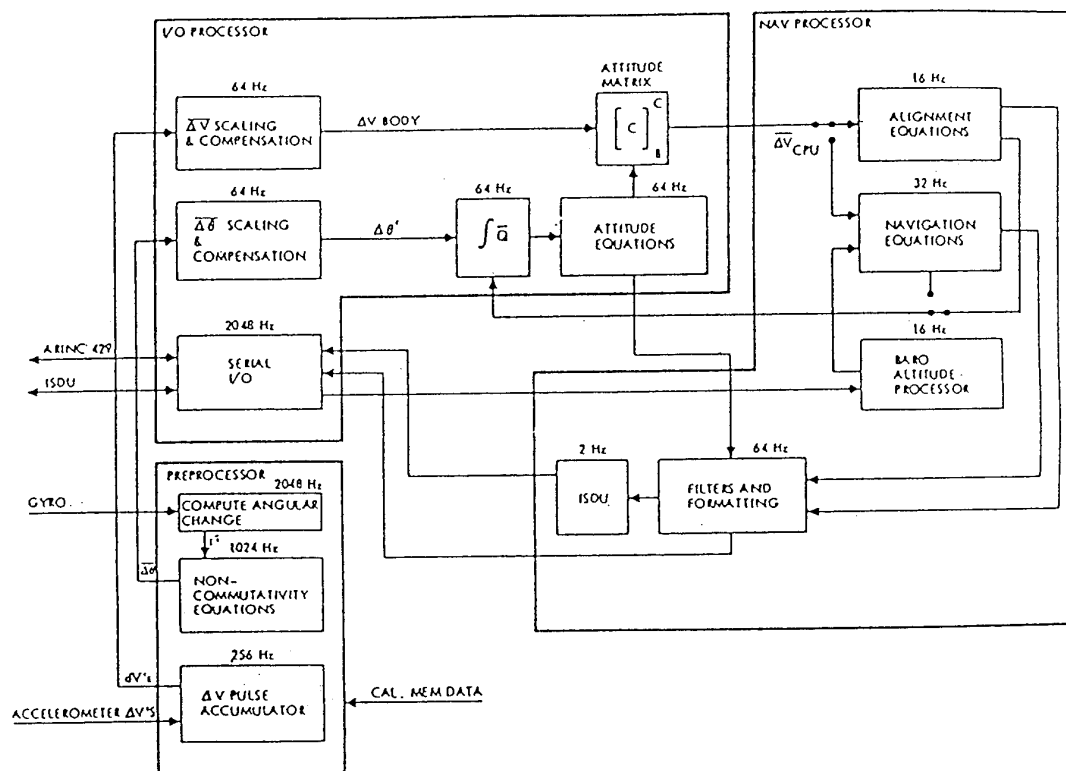
Fig. D 10 The Functional Diagram of the Litton LTN-90

Literature Appendix D

[AD 90] "1989/90 Data Conversion Products Databook"
Analog Devices, , Inc., One Technology Way, P.O. Box 9106, Norwood, Ma. 02062-9106, USA

[Be 66] Bendat, J.S., Piersol, A.G.: "Measurement and Analysis of Random Data."
John Wiley & Sons, Inc. New York, USA, 1966

[Br 55] Brandt, L.: "Vector and Tensor Analysis."
J. Wiley & Sons, Inc. 1955

[Ma 82] Mark, J.G. et al: "Design of RLG Inertial Systems for High Vibration".
CH1820-0/82/0000-0379 $00.75 1982 IEEE

[Mc 68] McKern, R.A.: "A Study of Transformation Algorithms for Use in a Digital Computer"
MIT Master Thesis, No. T-493, 1968

[Ra 77] Rahlfs, D.: "On the Accuracy Improvement of Incremental Readout Techniques for Inertial Sensors".
ESA-TT-573, Translation of DFVLR Report No. DLR-FB 77-70, 1977

[Sa 84] Savage, P. G.: "Strapdown System Algorithms"
AGARD Lecture Series No. 133, Advances in Strapdown Inertial Systems, 1984

[Std 70] "Flugmechanik, Begriffe, Benennungen, Zeichen, Grundlagen LN 9300 (Flightmechanics Standards)"
Beuth-Vertrieb GmbH, Köln, 1970

[We 78] Wetzig, V.: "Computation of Attitude and Heading in a Strapdown Computer"
DFVLR Internal Report No. IB 153-78/32, 1978

Appendix O

Optical Gyros

O1 Introduction

The so-called "Sagnac Effect" is the basis for optical gyros. It describes the effect of rotation upon 2 light beams travelling in opposite directions on the same optically closed path. Their phaseshift is proportional to the rotation component about the axis normal to the plane set up by the beams. This phenomenon was first observed by Harres in 1912 and by Sagnac in 1913 [Ha 12, Sa 13].

The sensing element of the optical gyro thus is massless and in contrast to the mechanical gyro is not affected by the dynamic environment. This feature appears to offer enormous advantages in a strapdown system (Section 2 and App. D). Therefore, optical gyros seem to be the ideal sensors for use in navigation and flight test systems.

Optical gyros in use today are fiber optic gyros (FOG) and ring laser gyros (RLG). In the following sections a brief review of the physics of the Sagnac ring interferometer and its use in the FOG and RLG is provided. The main sources of errors are also discussed and examples are given for both of them.

O2 The Sagnac Ring Interferometer (SRI) as Basis for Optical Gyros

Around the turn of the century physicists discussed intensely the nature of light. Michelson contributed to the discussions in 1881 with his investigations about the effect of translatory motion of a light source upon the speed of the emitted light. For his observations he used the interference of lightbeams travelling in the direction of motion and perpendicular to it. They did not show any measurable effect, thus proving that the speed of light (c) is not affected by the motion of its source, but remains constant with respect to inertial space. However, for the same reason, the investigations of Sagnac with lightbeams in rotary motion, see Fig. O 1, showed an effect [So 78].

The SRI setups in Figs. O 1a and b consist of a light source (LS), a beam splitter (BS), three fully reflecting mirrors M1, M2 and M3 in Fig O 1a, or a glass fiber in Fig O 1b for guiding the two light beams I and II on a square or circular path back to their common origin on BS, where part of their power is combined and projected onto the screen (SC). If the optical pathes for both beams are identical, a photon emitted from BS on paths I or II will reach the origin again after time $T = L/c$, where L is the length of the path and c the speed of light.
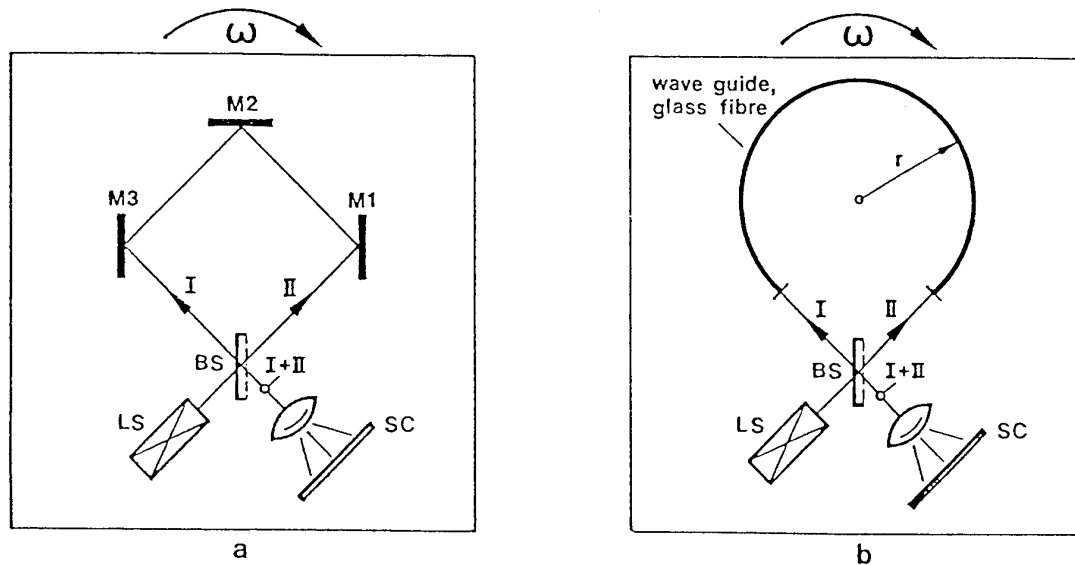
Figure O 1 Sagnac Ring Interferometer (a) and its Circular Substitute (b)

For derivation of the effect of rotation upon both beams we follow [Ro 85]. Any rotation – positive as shown in Fig. O 1 – about the axis normal to the plane of both beams affects the travelling length of both beams:

$$(O\ 1) \qquad L_{I,II} = L \pm v \cdot T_{I,II} = u_{I,II} \cdot T_{I,II},$$

with T = signal transmission time, $v = \omega \cdot r$ = tangential velocity of the ring and u = signal transmission velocity. In another

arrangement:

(O 2) $\quad\quad T_{I,II} = L/(u_{I,II} \bar{+} v)$.

An observer in the center of rotation will measure the signal transmission velocities according to the relativistic addition theorem as:

(O 3a,b) $\quad\quad u_{I,II} = (u \pm v)/(1 \pm u \cdot v/c_0^2)$

with c_0 = vacuum light speed. The transmission time difference between both beams is:

(O 4a,b,c) $\quad \Delta T = T_I - T_{II} = \dfrac{2 \; L \; v}{c_0^2 \; (1 - \beta^2)} \cong \dfrac{2 \; L \; v}{c_0^2} = \dfrac{2 \; L \; r}{c_0^2} \, \omega,$

with $\beta = v/c_0$. The time difference can also be interpreted as if the beams had travelled optical paths differing by the length. This interpretation helps us to understand the ring laser gyro very easily.

It is interesting to note that this formular is valid for any signal travelling in a closed path in opposite directions, i.e. as well for sound and for light! The technical problem lies in the measurement of the minute time difference - remember c_0^2 is in the denominator! In practice it can only be measured with light as transmitting signal. When the two light beams are brought to interference, the time shift causes a noticeable Sagnac phaseshift of:

(O 5) $\quad\quad \phi_s = 2\pi \; \nu \Delta T = \dfrac{4\pi \; Lr}{c_0 \lambda_0} \, \omega = \dfrac{8\pi \; A}{c_0 \lambda_0} \, \omega = 3.57$ arc min for $\omega = 1$ rad/s.

with $\nu = c_0/\lambda_0$ = frequency of light and $A = Lr/2$ = area of the circle. The figures in this equation are valid for the following assumptions: a circular path of 10 cm diameter, a light source of He-Ne (= 6.33 10^{-5} cm) and c = 3 10^{10} cm/s. The numbers reveal how weak the signal is.

It has been shown in [Po 67] that A is the area enclosed by the beams travelling on any path.

O3 The Fiber Optic Gyro (FOG)

Though the FOG is in its principle closer to the Sagnac ring interferometer than the ring laser gyro discussed in the next section, its development began at a time when the latter was commercially already available. This is due to the fact that the proper technology and readout concepts became available only in the late 1970ies.

Its basic principles are shown in Fig. O 1b with the output signal according to Eq. O 5.

Three principle obstacles had to be overcome in order to make the FOG a serious competitor to existing gyros:
- the low scalefactor according to Eq. O 5,
- the readout insensitivity at low input rates and
- readout nonlinearity as well as ambiguity.

O3.1 Scalefactor Increase and Error Reduction

With the advent of fibers as optical wave guides the increase of the Sagnac ring interferometer sensitivity became easily accessible. Each turn of a fiber optic coil increases the sensitivity in Eq. O 5. This idea was first suggested in [Kr 68]. But along with this benefit went adverse effects, since light was increasingly brought in contact with matter.

The resulting increase of temperature sensitivity and magnetic sensitivity was reduced by the use of polarization preserving fiber and designing the sensor setup in a reciprocal fashion [Ul 79, Ul 80], in which both lightbeams suffer the same fate on their path from light source to light sink or they follow the same number of reflections and transmissions. This is accomplished by using 2 beam couplers so that both beams are reflected and transmitted twice (s. Fig. O 2).
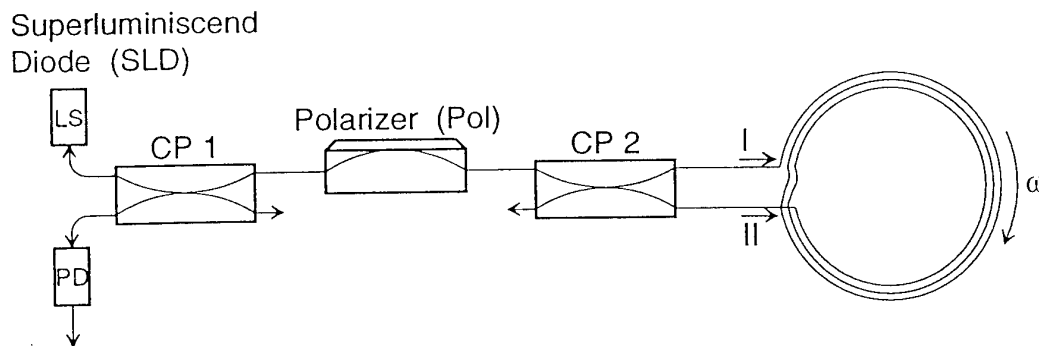

Figure O 2 Functional Diagram of a Reciprocal Fiber Optic Gyro Setup

The increasing pathlength of both optical beams increases the effect of backscattering and output noise, since the backscattered light causes a phaseshift of the light beams in a similar fashion as the the Sagnac effect. Light from a luminiscent diode with very short correlation length brought a drastic improvement [Pe 82].

O3.2 Increasing Sensitivity for Low Input Rates

Fig. O 3 may serve to explain the problem. It shows the superposition of the two countertravelling beams and the corresponding light intensity as a $(1 + \cos\phi_s)$- function of the Sagnac phaseshift ϕ_s with zero scalefactor at zero input rate. Only a phaseshift of $\pm\pi/2$ would bring the sensor into the desired range of highest sensitivity at low input rate. Such a "nonreciprocal" phaseshift ϕ_{nr}, i.e. a phaseshift acting upon both beams in a different manner, is generated by the phase modulator PM in Fig. O 2 at one end of the fiber coil.

The phase modulator PM causes a photon of beam I to be delayed after having travelled through the coil and a synchronous photon of beam II to be delayed before entering the coil. PM is stimulated with a squarewave voltage u_{PM} for an individual phaseshift of $\pi/4$ and a period of twice the light travelling time nL/c with n = refraction index of the fiber. The resulting phaseshift ϕ_{nr} is the difference of both beams, i.e. equal to $\pm\pi/2$. It is superposed to the Sagnac phaseshift and causes a light-intensity-function proportional to $[1 + \cos(\phi \pm \pi/2)]$ as indicated in Fig. O 4 left. It is demodulated and results in an output voltage $u \sim \sin\phi$. For low input rates the sensitivity of this "phasemodulated" FOG is thus proportional to ω, as expected. For high input rates it is both nonlinear and ambiguous. The proper signal evaluation in so-called "phasemodulated FOGs" (s. Fig. O 5a) is then carried out digitally.

Phase modulation can be accomplished by simply stretching the fiber with piezo elements or better by applying an electric field perpendicular to the fiber. These phase modulators are manufactured in integrated optics based on the use of lithium niobade.

O3.3 Scalefactor Linearization with a Control Loop for Sagnac Phaseshift Compensation

The scalefactor of the Sagnac ring interferometer can only be linearized through a control loop for nulling the Sagnac phaseshift ϕ_s, i.e. a method must be found for generating a controllable constant nonreciprocal phaseshift between both beams. This can be accomplished by a controllable nonreciprocal frequency difference $\dot{\phi}_m = f_m$ between both beams, because its corresponding phasedifference is $\phi_{nr} = f_m \cdot nL/c$.

The breakthrough for such a nonreciprocal "frequency" difference came by the proposal to stimulate the phasemodulator with an increasing voltage ramp as indicated in Fig. O 4 right, causing a linearly increasing phaseshift in both beams, with the one in beam I time-delayed by $n \cdot L/c$ [Le 85]. The phaseshift difference ϕ_{nr} of both beams is controllable by the ramp's inclination as indicated above and can be used to compensate the Sagnac phaseshift ϕ_s.

The steps superposed to the ramp serve to sense the Sagnac phaseshift as discussed with the phasemodulated FOG mentioned above, i.e. their demodulated output is controlled to zero by the ramp's inclination as indicated in this figure.

The voltage ramp for increasing phaseshift has to be set to zero again at a certain limit. This is done when the phaseshift has nominally reached 2π as indicated in Fig. O 4b. At that instant a pulse is released to the computer whose weight is:

$$(O\ 6) \qquad \Delta\phi = \frac{n\ \lambda}{2\ r}.$$

It is identical in principle to the one of the ring laser gyro (RLG) discussed in the next section.

Differences between the actual phase release and its nominal value remain stored in the fiber and serve to stabilize this "closed-loop" FOG scalefactor.

This closed-loop FOG (s. Fig. O 5b) has for the time being the best performance data, namely ≤ 5 deg/h drift and ≤ 500 ppm scalefactor variation. As compared to the RLG it has the following advantages:

- it is not subject to the so-called "lock-in effect" of the RLG and no mechanical dither is required to compensate it;
- the use of integrated optics promises that it will become cheaper;
- the solid state laser instead of a gas laser required for the RLG promises a longer lifetime;
- the pulse weight and the bandwidth is limited by the speed of the electronics only.

Though these advantages are of great importance for many applications, it does not seem likely that the FOG with conventional dimensions (< 10 cm radius) will reach the performance of the RLG. Single pieces have been developped with bigger diameter and superb performance as regards drift [Sc 92].

O4 The Ring Laser Gyro (RLG)

The setup for the ring laser gyro (RLG) in Fig. O 6 differs in principle from the one for the Sagnac ring interferometer in Fig. O 1 in the following way. The light source - a gas laser - is mounted into the optical path replacing the exterior light source plus beam splitter; the optical path layed down by mirrors M1, M2, M3 acts as a resonance cavity for the laser.

In this particular device the optical path is composed of three mirrors only. This is pursued at present by most of the manufacturers with the exception of Litton, which selected four mirrors. The RLG used by St. Petersburg Electrotechnical University for high accuracy goniometry has 4 prisms instead of mirrors [Fi 94].

In Fig. O 6 the mirror (M1) is partially transmitting and the prism (P) is part of the interference optics.

Like in the linear laser the ring laser is tuned to an oscillation with a wave length equal to an integer of the cavity length L. Due to the Sagnac effect, the cavity length of both light beams I and II will apparently be displaced from L by an amount $\Delta L_I = c \cdot \Delta t_I$ and $\Delta L_{II} = c \cdot \Delta t_{II}$, or relative to each other by an amount ΔL (s. Eq. O 1). This results in the frequency difference:

Field strength of beams I and II :

$$E_{I, II} = \hat{a} \exp j \left(2\pi \nu t + \frac{2\pi x}{\lambda} \pm \frac{\varphi_s}{2}\right)$$
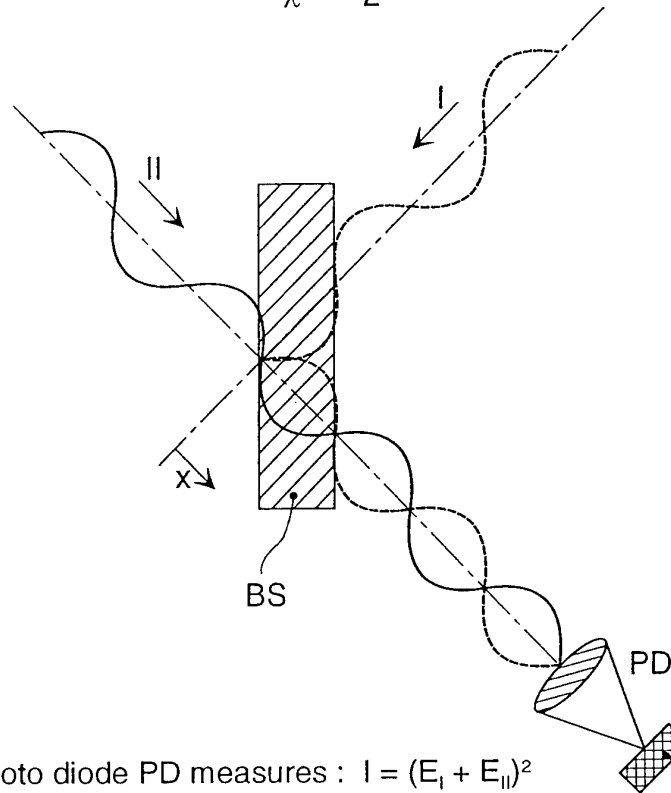
Photo diode PD measures : $I = (E_I + E_{II})^2$

$$= \frac{I_o}{2} (1 + \cos \varphi_s) \sim u(\varphi_s)$$
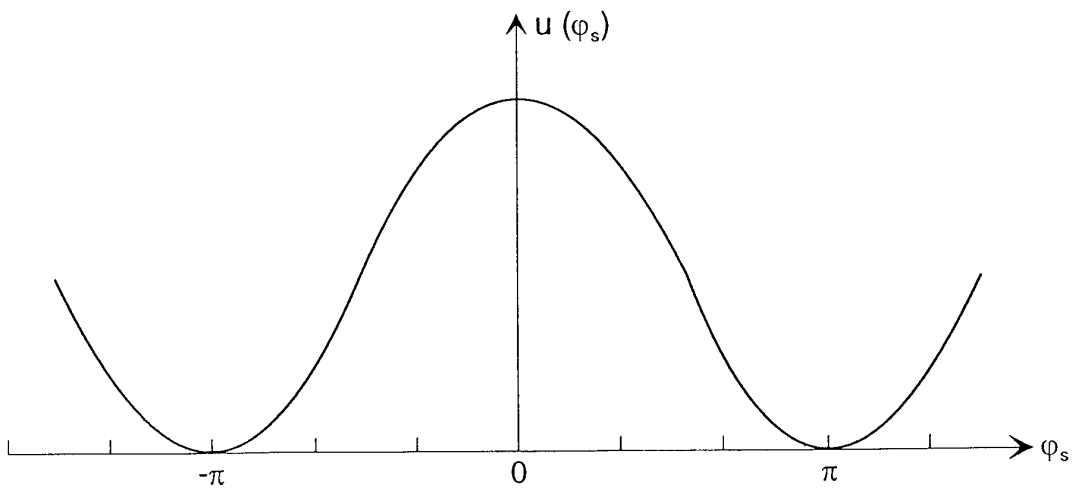
Figure O 3 The Interference of the 2 Light Beams and the Corresponding Light Intensity as a Function of their Phase Shift

Phase-Modulation with

Changing Direction

Digital Ramp plus
Square-Wave Bias
Modulation

U_m

$\rightarrow T \leftarrow$

t

U_m

$\rightarrow T \leftarrow$

t

$\varphi_{nr,II}$

$\varphi_{nr,II}$

$\varphi_{nr,I}$

$\varphi_{nr,I}$

$\rightarrow T \leftarrow$

φ_s

φ_s

φ_{nr}

$u(\varphi_s)$

$K\varphi_s$

$u(\varphi_s)$

$\leftarrow \varphi_{nr} \approx \dfrac{\pi}{2} \rightarrow$
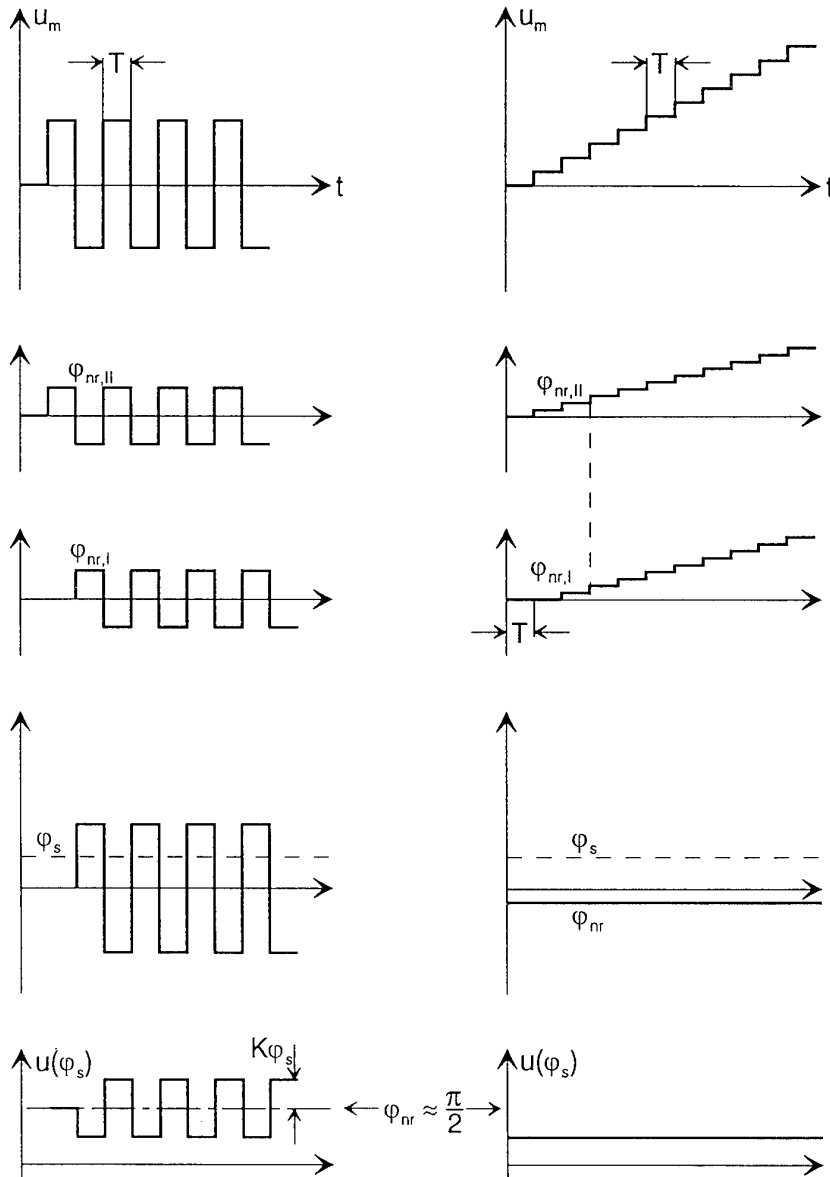
Figure O 4 Two Kinds of Phase Modulation in a Fiber Optic Gyro (FOG)

a)

Superluminiscend
Diode (SLD)

LS

CP 1

Polarizer (Pol)

CP 2

I

PD

Demodu-
lator

$u(\varphi_s)$

PM

II

ω

$u_a \sim I_o \sin(\varphi_s + \delta\varphi)$

$\varphi_s \sim \omega$

u_m

t

b)

Superluminiscend
Diode (SLD)

LS

CP 1

Polarizer (Pol)

CP 2

I

PD

Demod.1

Control
Electr.1

$u(\varphi_s)$

(A)

Voltage Controlled
Amplifier

PM

II

ω

Demod.2

Control
Electr.2

Reset-Ramp
Generator

(B)

u_m

t

Reset Pulses $\Delta\phi$
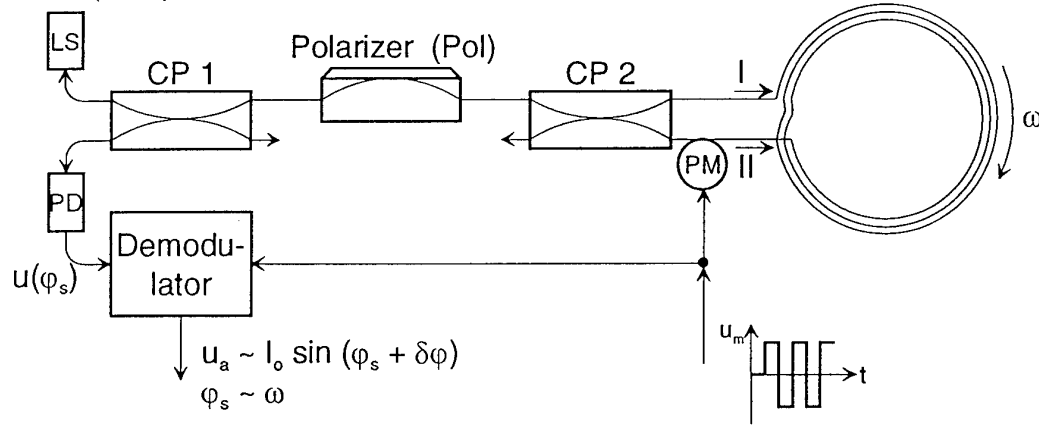
Figure O 5 Functional Diagrams for Fiber Optic Gyros with Two Kinds of Phase Modulation:
Square-Wave Modulation (a) and Digital Ramp Modulation (b)
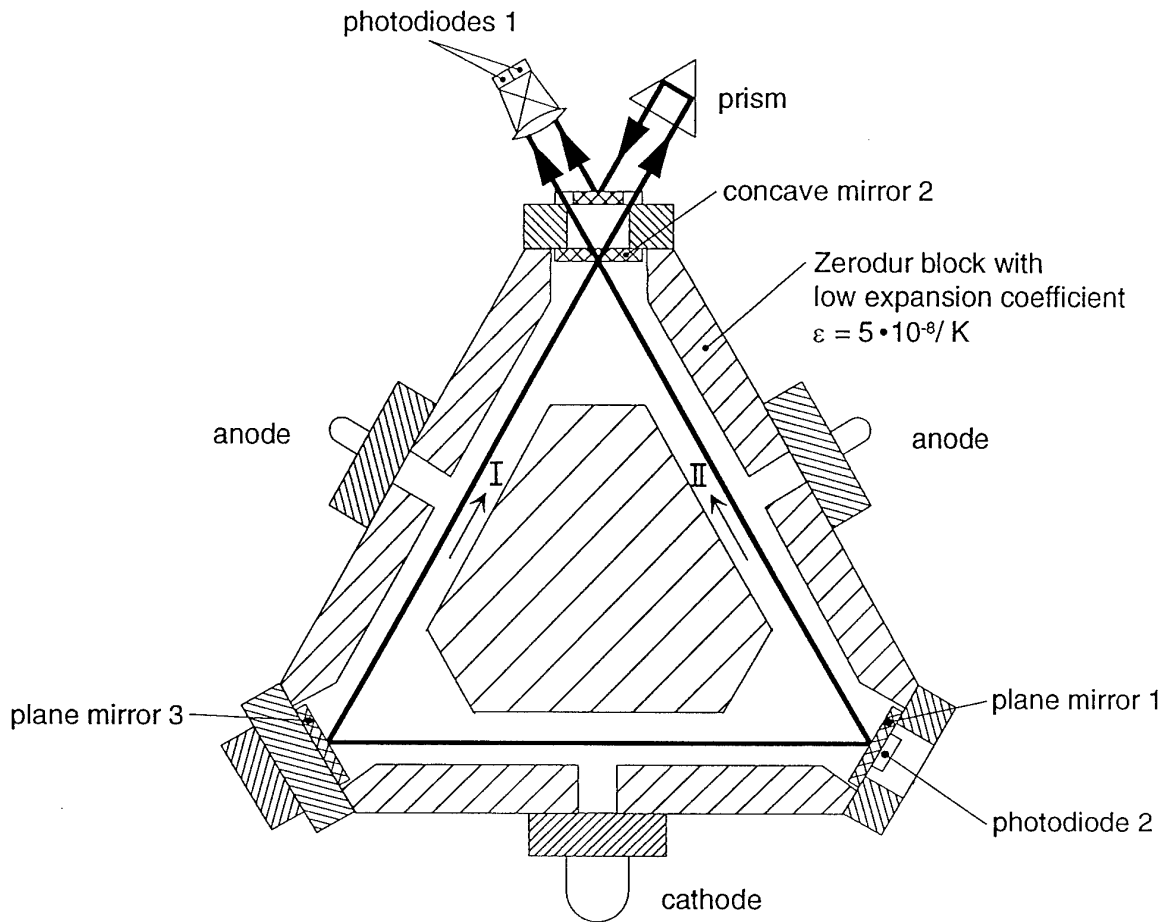
Figure O 6 The Ring Laser Gyro (RLG) in Principle
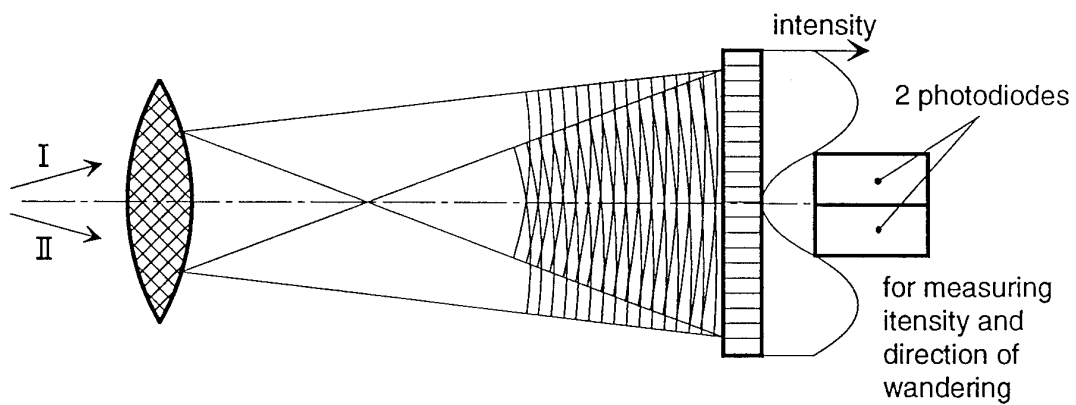
Figure O 7 Interference of the 2 Light Beams of the Ring Laser Gyro and Interference Pattern Readout

(O 7) $\Delta\nu[\text{Hz}] = \dfrac{4\ A}{L\ \lambda}$ $\omega[\text{deg/s}] = 1,588$

for a triangular path with a side length of 10 cm and $\lambda = 6.33\ 10\text{-}7$ m (He-Ne laser).

This frequency difference causes a wandering of the interference pattern at the screen (SC) proportional to the angular rate with respect to inertial space. This is indicated in Fig. O 7.

Two photodiodes mounted on the screen (SC) will count the number of fringes that pass it and allow the definition of the sense of rotation. Each fringe may then easily be converted into a pulse which indicates the angle increment of the sensor with respect to inertial space. The number N of the pulses for the total input angle displacement:

(O 8) $N = \dfrac{4\ A}{L\ \lambda}\ \Delta\phi$

defines the RLG nominal scale factor S, which for $= 6.33\ 10\text{-}7$m and a triangular RLG of 10 cm side length has the following magnitude:

(O 9) $S = \dfrac{\Delta\phi}{N} = \dfrac{L\ \lambda}{4\ A} = 2.25$ arc sec/pulse.

In Fig. O 8 a setup is shown indicating the RLG characteristics. It seems as if the mirrors were passing alongside the space- fixed light pattern [Sc 66]. This setup indicates that the RLG is an integrating gyro with digital output and ideally suited for use as a flight test sensor and in strapdown navigation systems.

We will discuss in the following some design criteria and error sources common for all RLGs and compare them with those of the FOGs, if possible.

standing light wave
fixed in space
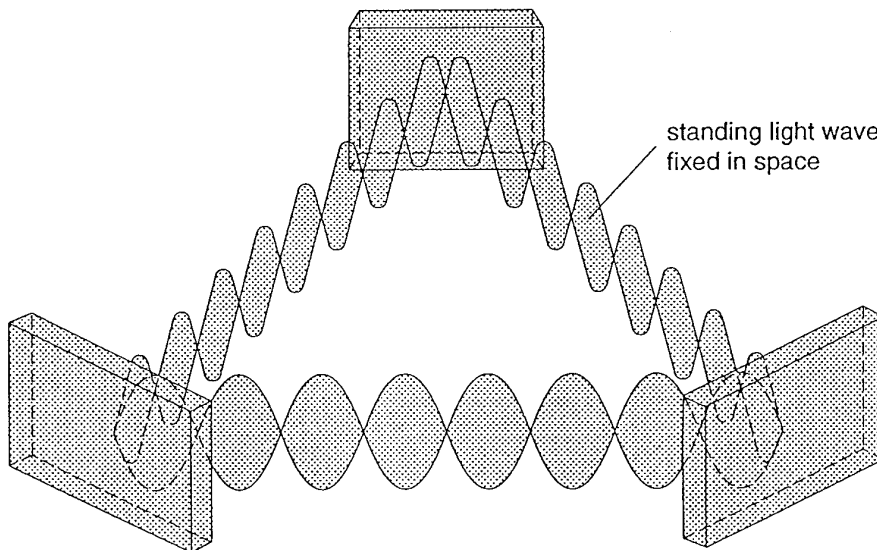
Figure O 8 Model for the Functioning of a Ring Laser Gyro

O4.1 Design Criteria Common to all Ring Laser Gyros

Let us first comment on the selection of the laser - gas laser or solid state laser.

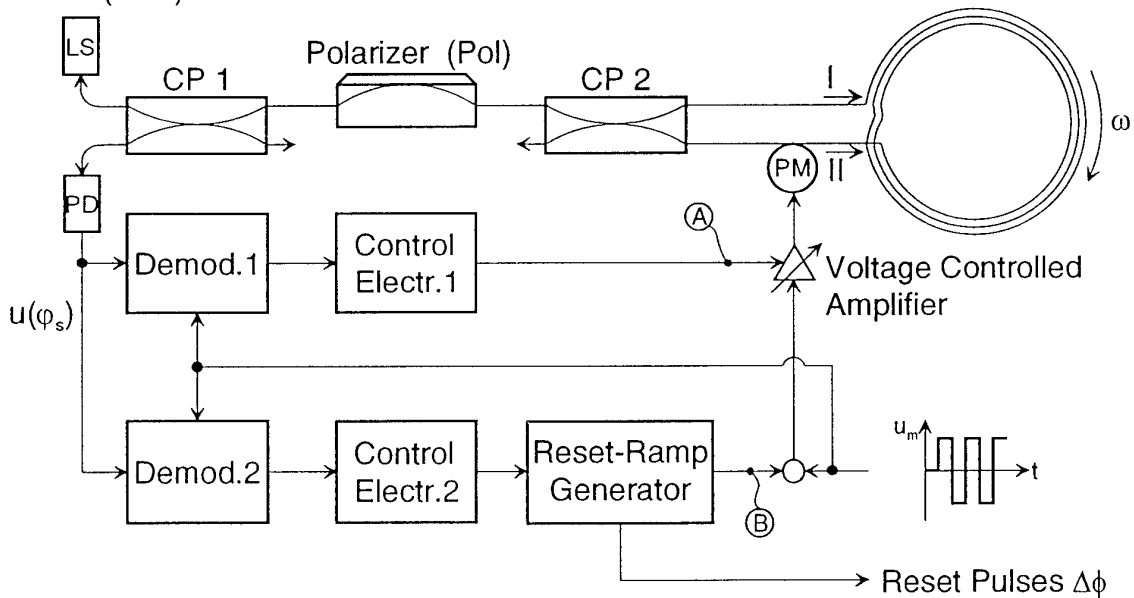In difference to the FOG which operates at the fixed frequency of its exterior light source, the RLG operates at two frequencies both of which have to be amplified inpendently in one laser. This can be accomplished in a gas laser only in which due to the Doppler effect of its atoms the intensity of the amplified light as a function of resonator length or lasing frequency, respectively, is bell- shaped with a bandwidth of 1500 MHz, approximately corresponding to 1 wave length λ.

Gas flow within the laser tube caused by the voltage between anode and cathode (Langmuir flow) and temperature gradients affects the lasing frequency (Fizeau effect). For its overcoming RLG's are always equipped with two anodes and one cathode, or vice versa, mounted symmetrically within the tube, as indicated by Fig. O 6. If only one anode were to be used, the result would be an apparent rotation of the sensor proportional to a multiple of 100 deg/h [Po 68]. The sensor may be biased to compensate for some of these effects by adjustment of the high voltage (ca. 1500 V) in both branches: The number just quoted indicates that the voltage should be properly controlled and temperature gradients should be prevented.

The natural He-Ne laser wave length is $\lambda = 633$ nm. The exact lasing wave length is adjusted due to the geometric path length, and we may assume that its mean value for both beams varies linearly within one λ-period. Within that path length range the RLG scalefactor is insensitive to path length variations, since in Eqs. O 7 to O 9 the variations of nominator and denominator cancel!
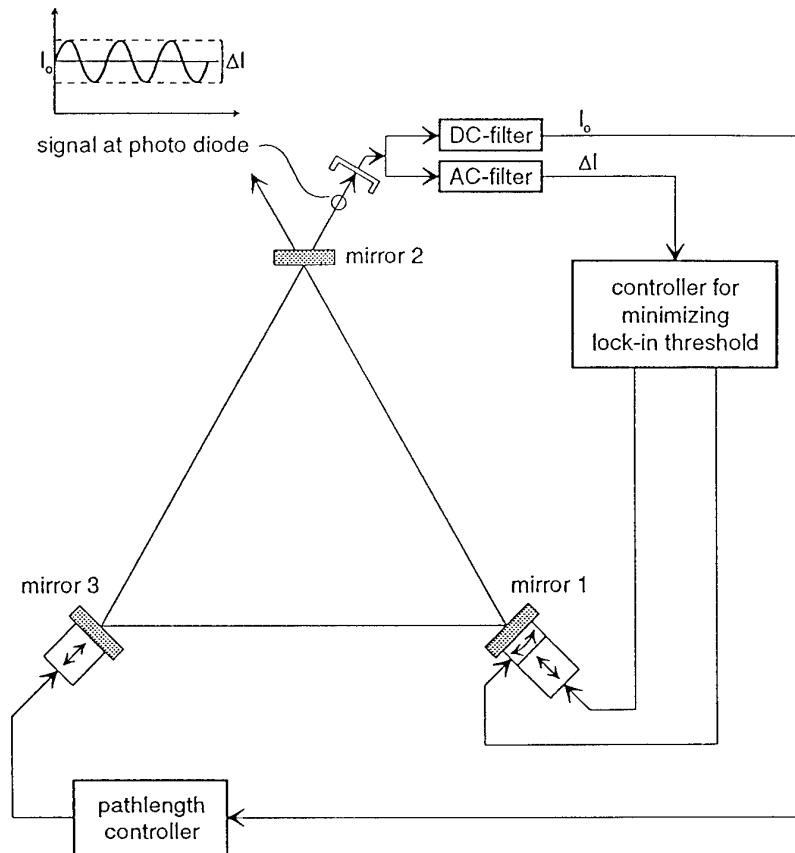
Figure O 9 Control Loops for Stabilizing the Beam Geometry of a Ring Laser Gyro

For preventing the path length to swap the λ-period and for scale factor stabilization, the RLG is firstly built of material with a very low thermal expansion coefficient (Cer-Vit, Zerodur) and secondly is equipped with a path length control loop as indicated in Fig O 9. The input signal into this loop is the mean intensity of both beams measured at the backside of one of the mirrors (all of them are partly transmitting). Its control to its relative maximum controls at the same time the path length to a constant value! This is accomplished by positioning one of the remaining mirrors with a piezo element. The mean light intensity maximum definition is done by dithering this piezo element with a frequency ω and filtering out the 2ω- term.

The path lenght control to the maximum of the mean light intensity makes any other temperature control of the sensor unnecessary. We may thus state two essential advantages of the RLG over other gyroscopic sensors: **digital output and temperature insensitivity** - at least in theory.

## O4.2 Error Source Common to all Ring Laser Gyros - the Lock-In Effect

A RLG incorporating the above design features will work satisfactorily for high input rates limited only by the readout electronics. But the RLG has a lower limit on input rates of a few hundred degrees per hour called "lock-in" range (s. Fig O 10). Above this range its actual input-output relationship is approximated by (s. also Eq. O 7):

$$(O\ 10)\quad \Delta \nu \quad = \quad \frac{4\ A}{L\ \lambda}\ [\ 1\ -\ (\omega^L/\omega)^2\ ]^{1/2}\ \omega,$$

with ω^L = "lock-in rate".

This lock-in effect is experienced by any two weakly coupled oscillators with slightly natural frequencies and thus known in other fields of physics, too. If this oscillating system is excited near their mean natural frequency, both subsystems will oscillate at the same frequency. This phenomenon is of benefit in acoustics (tuning of different violins in an orchestra) or electronics (mechanical and electronic coupling in a quartz watch), for instance yet, it is the main problem in a RLG. The beat frequencies between both beams and sensed at the output diode will be experienced only, if the oscillation frequencies are distinctly different as to be seen from Fig. O 10.

The coupling of both light beams is caused primarily by the scattering of light at the mirror surfaces (s. Fig O 11 top) and within the laser tube. It cannot be completely overcome. Its reduction requires on the one hand mirrors with low backscattering and high reflectivity. It requires on the other hand an additional control loop for the light beam geometry (s. Fig. O 9), since ω^L is the result of the superposition of the backscattering of all sources within the resonator, especially at the mirrors (s. Fig. O 11 bottom). Due to the phase-relationship of all backscattered lightsources the superposition of all of them can be driven to a minimum [Ro 92]. For present RLGs of inertial quality and with a beam length of 10 to 15 cm ω^L is in the order of 100 deg/h. The control loop for minimizing ω^L stabilizes this parameter at the same time, thus stabilizing the RLG scalefactor as well (s. Eq. O 10 and [Ro 92]).

The sensitivity threshold mentioned above is still too high for use in an inertial navigational system. Among the different

Δv [kHz] ⊢ 50

⊢ 10

– ωl

–10   +ωl   . 10   ωi [°/s]

–10

$A = 0.017\ m^2$

$L = 0.6\ m$

$\lambda = 6.33 \cdot 10^{-7}\ m$

–50
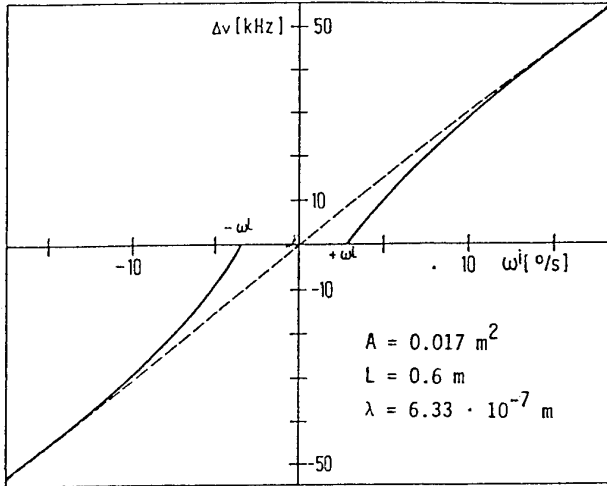
The Lock-In Effect of the Ring Laser Gyro

Lock-In Compensation: Rotational Dither

Input-Output with Dither

Δv

ωi

–ωl   ωl

peak dither rate ωD

–ωD   ωD

typically: ωD ≈ 100 to 300 deg/s

Δv

without dither

ωi

ωl

ωD   ωl   ωD

with dither
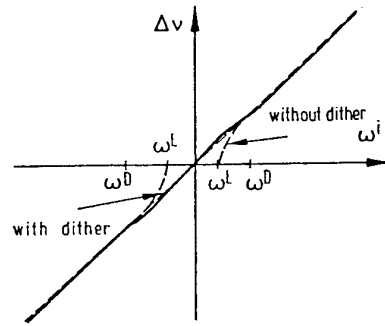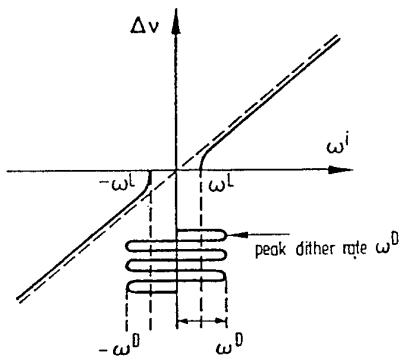
d) Scale Factor for Dithered Ring Laser Gyro

$\mathcal{K} = \dfrac{\delta S}{S}$

with dither

$0,43 \times 10^{-9}$

$1,3 \times 10^{-6}$

0

ideal

$\dfrac{1}{2}\left(\dfrac{\omega l}{\omega i}\right)^2$

without dither

ωi

0 ωl   ωD
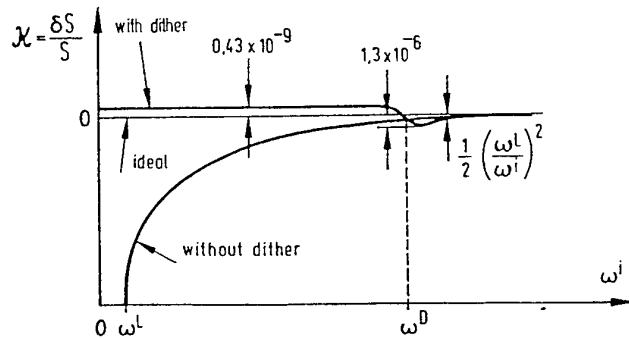
$\omega^L = 300\ deg/h$

$\omega^D = 200,000.\ deg/h = 56\ deg/s$

Figure O 10 The Lock-In Effect of the Ring Laser Gyro and Bias Techniques to Overcome it

reflected and backscattered
light energy of beam I;
backscattered energy visible
as red point on mirror
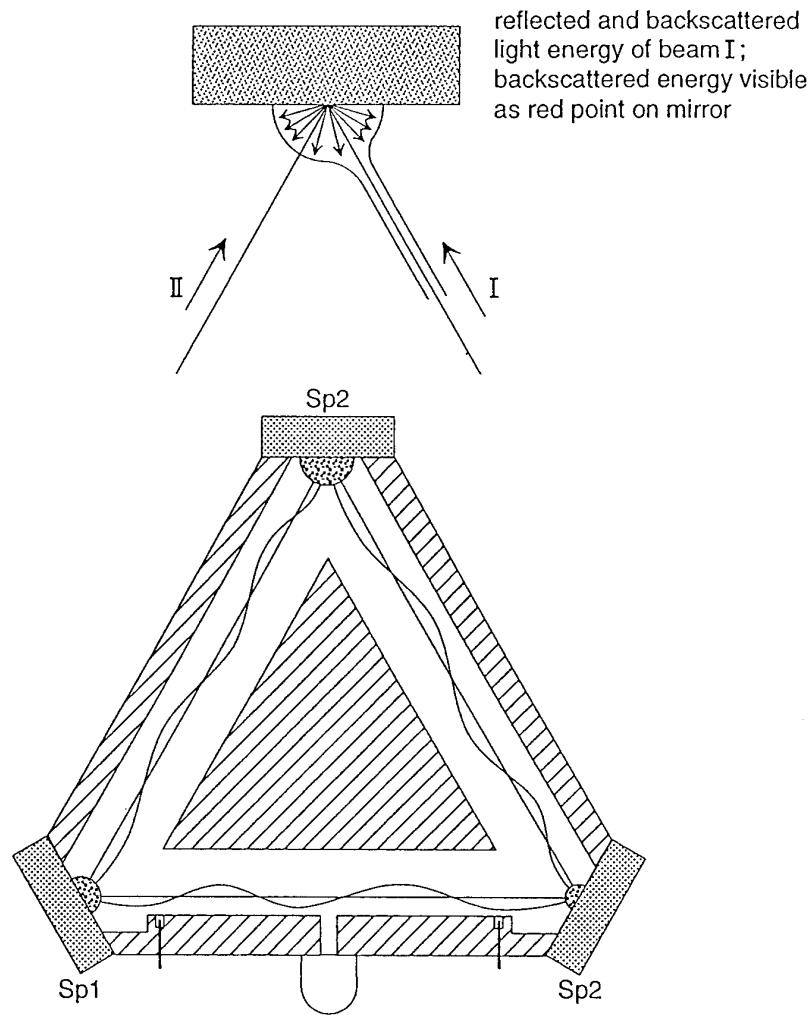
Figure O 11 Backscattering at one Mirror and within the Resonator

techniques for "biasing" the RLG, i. e. making it sensitive to low input rates (mechanical dither, constant rotation, magnetic mirrors, Faraday cell) the mechanical dither is most commonly in use. As indicated in Fig. O 10, the RLG is rotated within its case periodically with a frequency of 400 Hz and an amplitude of a few arc minutes approximately. Distinctly to be seen in Fig. O 12 is the dither spring of the Honeywell RLG in the center of the lasing block. For its compensation at the output signal this RLG has a case-fixed prism and readout as indicated in Fig. O 12 bottom.

Since this bias is applied periodically, the gyro will enter the lock-in region twice per dither cycle and each beam will loose the lock of phase to some extent. The resulting noise accumulates in a so-called "random walk" after integration of the measured rate to an angle. The uncertainty of this error expressed as 1σ band increases with the square root of the elapsed time. This is particularly serious when short measurement times are of interest as with the INS alignment (s. Section 5) or with the extraction of the sensor's drift.

Another error source with the dither bias is due to the scalefactor nonlinearity at input rates near the maximum dither rate (s. Fig. O 10c). It arises from the fact that at this point the highly nonlinear maximum dither sine wave dives into the lock-in region. This scalefactor nonlinearity shows an almost square law dependence on the lock-in rate amounting to 50 ppm at a lock-in rate of approximately 0.4 deg/s. For its overcoming the dither frequency in the order of 400 Hz and a few arc minutes amplitude are varied randomly.

Among the other bias techniques mentioned above the ones based upon magnetism (magnetic mirrors, Faraday cell) seem to open the most elegant way to cope with the lock-in effect. They are hampered by the fact that both of them increase backscattering within the resonator and from this point of view are not promising for high-quality sensors. For the time being dither bias is most widely spread for inertial systems of 1 NM/h performance. In special system implementations as in inertial systems for submarines we also find rate bias. These systems have proven already navigation accuracies in the range of 1 NM/d [Ko 90] and seem to have a growth potential [Ro 92].

Commercially available RLGs have excellent performance data: <0.01 deg/h drift, < 5 ppm scalefactor variation and < 0.001 deg/$\sqrt{h}$ random walk [Ul 88].
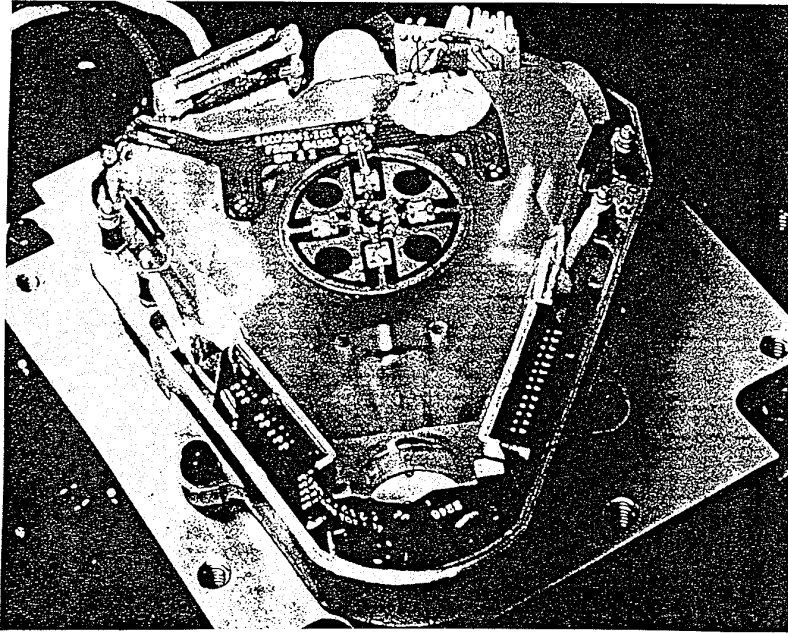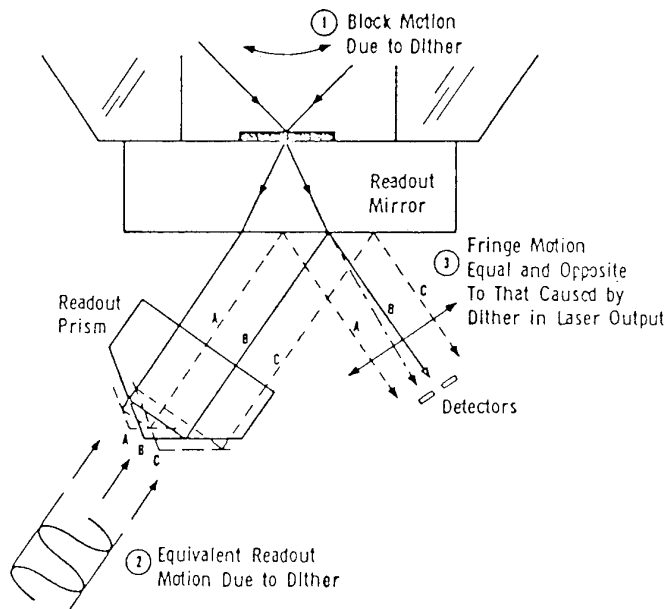
Figure O 12 The Honeywell GG 1342 Ring Laser Gyro and its Optomechanical Readout Compensation for Dither

Literature Appendix O

[Fi 94] Filatov, Y.U. et al: "Laser Goniometric System for Measurment of Object Angular Position."
Proceedings of the 3rd International IMEKO Symposium on Laser Metrology for Precision Measurement and Inspection in Industry, Heidelberg, March 1994,VDI-Report No. 1118

[Ha 12] Harres, F.: "Die Geschwindigkeit des Lichtes in bewegten Körpern."
Dissertation Universität Jena, Erfurt (1912), no longer available

[Ko 90] Kohl, K.W.: "A new High Accuracy Ship's Inertial Navigation System PL41 MK4."
Proceedings Symposium Gyro Technology 1990, DGON and Universitaet Stuttgart, 1990

[Kr 68] Kroy, W., Mehnert, W.E.: "Anordnung zur Messung von Drehgeschwindigkeiten."
DBP Nr. 1 807 247, 1968

[Le 85] Lefevre, H.C. et al: Double Closed-Loop Hybrid Fiber Gyroscope Using Digital Phase Ramp."
Proceedings of the 3rd International Conference on Optical Fiber Sensors(OFS '85), San Diego, PDS 7.1-7.8, 1985

[Po 67] Post, E.J.: "Sagnac-Effekt."
Review of Modern Physics, Vol. 39, No. 2, (1967), pages 475 to 493

[Pe 82] Petermann, K,: "Intensity-Dependent Nonreciprocal Phase Shift in Fiber-Optic Gyroscopes for Light Sources with Low Coherence."
Optics Letters, Vol. 7, No. 12, pp 623-625, 1982

[Po 68] Podgorski, T.J.: "Langmuir Flow Effects in the Laser Gyro"
IEEE Journal of Quantum Electronics, Vol. QE-4, No. 1 pp 11-18, 1968

[Ro 85] Rodloff, R.: "Der Laserkreisel - Einfluß der Ringlaser-Geometrie auf die Kreiseleigenschaften."
Laser und Optoelektronik Heft Nr. 2, 1985

[Ro 92] Rodloff, R.: "Lasergyro - the Next Generation."
Proceedings Symposium Gyro Technology 1992, DGON and Universitaet Stuttgart, 1992

[Ro 93] Rodloff, R. et al: "Hochaufloesendes Kreiselsystem zur Präzisions-Winkelmessung und - Navigation"
Deutsche Patentanmeldung P 42 31 935.8, Anmeldetag 18. 09.1992

[Sa 13] Sagnac, G.: "Comptes Rendus, 157, (1913)

[Sc 92] Schroeder, W. et al: "Telescope Pointing and Tracking with Optical Gyros."
Proceedings Symposium Gyro Technology 1992, DGON and Universitaet Stuttgart, 1992

[Sc 66] Schulz-Du Bois, E.O.: " Alternative Interpretation of Rotation Rate Sensing by Ring Lasers."
IEEE Quantum Electronics. Vol QE-2,pp 299-305, 1966

[So 78] Sommerfeld, A.: "Optik."
Verlag Harri Deutsch, Thun, Frankfurt/M, 1978

[Ul 79] Ulrich, R.: "Fiber-Ring Interferometer Polarization Analysis."
Optics Letters Vol. 4, pp152-154, 1979

[Ul 80] Ulrich, R.: "Fiber-Optic Rotation Sensing with Low Drift."
Optics Letters Vol. 5, pp 173-175, 1980

[Ur 88] Ullrich, D.: "Static and Dynamic Angular Measurements with a Ring Laser Gyro."
Proceedings Symposium Gyro Technology 1988, DGON and Universitaet Stuttgart, 1988

Doppler Radar Navigation
Heinz Buell
GEC-Marconi Electronic Systems Corporation
164 Totowa Road
Wayne, NJ 07474-0975
U.S.A.

1. INTRODUCTION

The use of the Doppler principle to measure velocity has been underway since the early 1950's. The Doppler principle or effect has been applied to many systems in which velocity is an important variable. The discussion herein will concentrate on the use of a self-contained radar in an airborne vehicle to measure the velocity of that vehicle. Fixed-wing aircraft and helicopters have used Doppler radars successfully for many years, and drone aircraft and missiles are beginning to exploit the low cost and high reliability of Doppler radars.

The Doppler effect or frequency shift was first described by the Austrian physicist, Christian Doppler (1803-1853) for the case of sound waves. This effect is also characteristic of light and microwave energy. In each case a shift in frequency is observed when the source of radiation is moving toward or away from an observer of that radiation. The frequency shift increases when the source moves toward the observer, and decreases when the source moves away from the observer.

Airborne Doppler radars are both the source (transmitter) and the observer (receiver) of the microwave energy utilized to measure vehicle velocity. A typical Doppler radar transmits a small amount of energy toward the ground, and measures the Doppler shift in that portion of the transmitted energy reflected or back-scattered toward the receiving antenna. The resultant Doppler frequency shift is directly proportional to the vehicle's velocity relative to the ground. Note that two frequency shifts occur in this case; one between the energy (source no. 1) and the ground (receiver no. 1) and another between the energy reradiated from the ground (source no. 2) and the receiving antenna (observer no. 2).

The frequency shift is proportional to that component of vehicle velocity which is parallel to the direction of the radiation. Radiation that is transmitted perpendicular to the direction of the vehicle's velocity, or velocity vector, will not experience any Doppler shift. Radiation transmitted parallel to the velocity vector will experience the maximum frequency shift. One beam of radar energy will therefore provide a measurement of one component of vehicle velocity. In general, three or more non-coplanar beams must be used if the three orthogonal components of vehicle velocity are to be measured.

Doppler radars to measure velocity of aircraft have been in production and in use since the early 1950's, and over 20,000 systems have been produced since then by several manufacturers. Some of these systems provide only velocity outputs whereas others also perform navigation, guidance and control and display functions by using data obtained from the vehicle's attitude sensors. Some Doppler radars measure not only velocity but also the

vehicle's altitude above the ground, and thereby eliminate the need for a separate radar altimeter.

The size, weight and cost of Doppler velocity sensors and navigation system has decreased dramatically since 1950; weight, for example, has dropped from approximately 160 Kg for the AN/APN-81 to less than 4.5 Kg for Dopplers now in production for drone aircraft, and to less than 5.9 Kg for helicopter Dopplers.

Performance of a Doppler radar is determined by a number of basic system design characteristics. The major design trade-offs will be discussed, including the resultant performance variations, and a typical Doppler design will be described. The key error contributors will be listed and summarized; errors of a particular system can be described in detail by the manufacturer of that system.

Cost of a Doppler radar depends upon its performance, and also upon the quantity being purchased and the rate of production. In general Doppler velocity sensors are inexpensive alternatives to high-quality inertial systems, and in fact, a Doppler radar and a low-cost inertial system are less expensive than a high-quality INS.

2.0 Principles of Operation of a Doppler Radar

2.1 The Doppler Equation

When the Doppler effect is utilized in an airborne Doppler velocity sensor, the source is a microwave transmitter located in an aircraft moving with a velocity V relative to the earth's surface (see Figure 1).

It can be shown that

if V = aircraft's speed relative to the ground
c = speed of propagation of electromagnetic energy
f_t = transmitted frequency
L = "looking angle" - (angle between the line of flight and the transmitted beam direction)

and f_d = the Doppler frequency shift,

then $f_d = \dfrac{2V}{c} f_t \cos L$ \qquad (1)

Note that V cos L is the component of velocity in the direction of propagation. The wavelength λ is related to f_t and c by

$$\lambda = c/f_t \qquad (2)$$

Equation (1) can therefore be written as:

$$f_d = \frac{2V}{\lambda} \cos L \qquad (3)$$

Equations (1) and (3) constitute the basic expression for measurement of velocity by applying the Doppler principle to the propagation of electromagnetic radiation.
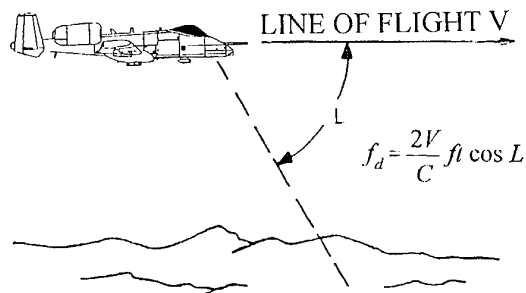
Figure 1. The Basic Doppler Equation

134

2.2 Determining the Total Velocity Vector

A single beam such as that shown in Figure 1 could be used to determine the aircraft's speed relative to the ground, provided that the aircraft speed were horizontal with no vertical velocity and no sideways "drift" (assuming that λ and L of equation (3) were known). Normally, however, the aircraft's speed does contain components of vertical velocity and drift. To determine the total velocity vector, V, it is therefore necessary to employ at least three (non-coplanar) beams. The velocity so determined is relative to a frame of reference attached to the antenna assembly providing the beams; i.e., relative to the aircraft's coordinate system, (x,y,z).

A Doppler radar velocity sensor is ideally suited for use in helicopter hover control, fire-control and weapons delivery and in aiding inertial navigation systems (INS) since these systems require velocity in air-frame coordinates. Velocity sensors such as gimballed INS and GPS receivers measure velocity in geographic coordinates, and true airspeed sensors measure velocity relative to the air mass. Doppler radars are thus a unique source of velocity data for many functions.

Consider an antenna assembly generating four beams A, B, C and D as shown in Figure 2. The Doppler frequency of the signal received in the Nth beam is (from equation (3):

$$f_d,N \quad = \underline{2} \, V \cos L_N \qquad (4)$$
$$\lambda$$
$$= \underline{2} \, (V_x \cos \gamma_N + V_y \cos \sigma_N$$
$$\lambda$$
$$+ V_z \cos \psi_N)$$

Any three of the equations can be solved for V_x, V_y, and V_z, assuming that the Doppler frequencies, the wavelength and the beam direction in the x, y, z coordinate system are known.

The fourth equation provides redundant information that can be used to test overall performance of the system. For example, if V_x, V_y and V_z as computed from N = 1, 2, 3 agree with V_x, V_y and V_z computed from, say, N = 1, 2, 4, then the overall operation of the system is valid. Such "redundant" velocity checks are an important part of the BITE function of modern Dopplers.
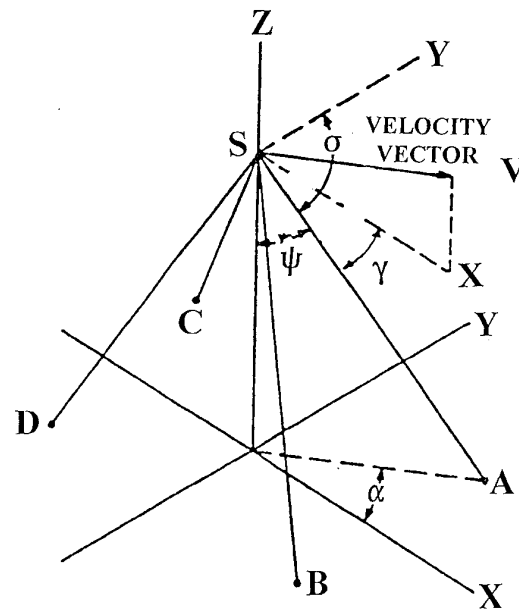
Figure 2. Total Velocity Vector Geometry

2.3 Antenna Stabilization

For most applications of Doppler velocity sensors it is necessary to find the horizontal (V_H, V_D) and vertical (V_V) components of velocity. The horizontal components, in particular are needed to perform dead-reckoning navigation. The (x,y,z) velocity components must be transformed through vehicle pitch and roll to obtain V_H, V_D and V_V. Pitch and roll are usually obtained from the vehicle's attitude sensor (vertical gyro or inertial navigation system). Almost all modern Doppler radars utilize antennas that are fixed or body-mounted, and this transformation is performed in a digital computer that may be in the Doppler itself, or in a separate "mission" computer. Some older Dopplers had antennas that were mounted on pitch and roll gimbals so that their outputs of V_x, V_y and V_z were the same as V_H, V_D and V_V.

Such mechanically stabilized antennas eliminated the need for "data" stabilization of V_x, V_y and V_z and also provided better performance during maneuvers over water. These systems were heavier and required considerably more volume for the gimbal structure as well as a separate radome. For these reasons most modern Dopplers use fixed antennas.

2.4 Transmitter Frequency Considerations

Doppler radars have been allocated the frequency band of $13,325 \pm 75$ MHz. Some earlier Dopplers operated in the X-band, in the vicinity of 8,800 MHz but most of these have been retired from service.

Equation (1) shows that the Doppler frequency shift depends on the transmitter frequency f_t. This sensitivity is virtually eliminated by mixing or heterodyning the received signal with a portion of the transmitted signal. The major remaining sensitivity to frequency is that of the antenna beam geometry or (γ, σ, ψ), which depends on the antenna design itself.

2.5 Antenna Beam Geometry Considerations

Equation 4 shows that f_d depends on beam geometry, that is, upon (γ, σ, ψ) of each beam. These equations can be combined to obtain V_x, V_y and V_z as a function of the 12 beam angles. It can be shown that the velocity errors are given approximately by

$$\frac{EV_x}{V_x} = \text{Tan } \gamma_o \text{ } \overline{E\gamma} \quad (5A)$$

where γ_o = average γ angle of the 4 beams, and

$\overline{E\gamma}$ = average uncertainty of γ.

$$\frac{EV_y}{V_y} = \text{Tan } \sigma_o \text{ } \overline{E\sigma} \quad (5B)$$

where σ_o = average σ angle of the 4 beams and

$\overline{E\sigma}$ = average uncertainty of σ

$$\frac{EV_z}{V_z} = \text{Tan } \psi_o \text{ } \overline{E\psi} \quad (C)$$

where ψ_o = average ψ angle of the 4 beams and

$\overline{E\psi}$ = average uncertainty of ψ.

The direction of a beam in space can be defined in terms of two angles; the angle ψ

can be expressed as a function of γ and σ. Selection of γ and σ is one of the fundamental trade-offs of a Doppler radar design. A large γ and a large σ (small ψ) results in beams pointed nearly straight down, and thus provide a relatively high returned signal strength. This is particularly important when flying over water where the radar back-scattering decreases with increasingly smooth water. Figure 3 shows the radar back-scattering versus incidence angle. (The beam angle ψ during level flight, is the same as the angle of incidence). It can be seen from Figure 3 that the back-scattered energy is higher for land and rough water, but diminishes sharply with increasing incidence angle over smooth water. Extensive operation over water therefore favors a smaller ψ and thus larger γ and σ angles.

SCATTERING
CROSS SECTION
($10 \log \sigma_o$)

WOODED LAND

GPL4

GPL3

GPL2

GPL1

GPL½
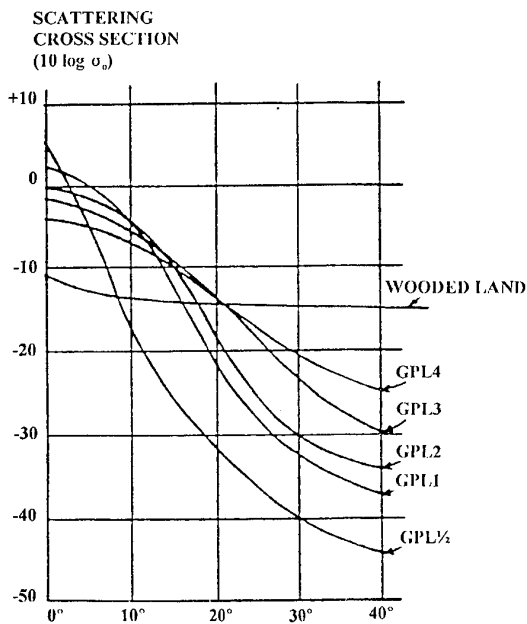
Figure 3. Scattering Cross Section vs. Incidence Angle (γ)

Equations (5A) and (5B) show that the sensitivity of EV_x and EV_y to beam angle uncertainty (Eγ and Eσ) increases sharply as γ and σ increase. Thus velocity accuracy favors smaller γ and σ angles and a larger ψ angle. Doppler radars have been designed with γ and σ angles as small as 65°, but were intended primarily for over land and low altitude flight. Most Doppler radars have larger γ and σ angles - of the order of 75° - which results in somewhat greater sensitivity to beam uncertainties but better overwater performance. Some Dopplers have angles as large as 80° but are intended for operation primarily over water. Beam angles in the 72° to 75° range are a reasonable compromise for Dopplers on aircraft that may experience a variety of missions.

Equations (5A) and (5B), when evaluated for beam angles of 73° show a sensitivity of velocity error of 3.9% per degree of beam uncertainty. Beam geometry must therefore be known to better than 0.06 degrees if velocity accuracy is to be better than 0.25%. Normal production tolerances make it very difficult to achieve these beam accuracies; this problem is overcome by measuring the actual beam geometry of each antenna on a radar range and then applying the resultant "calibration" factors to the velocity outputs.

2.6   Doppler Velocity for Navigation

A Doppler dead-reckoning navigation system integrates the north and east components of velocity to determine the change in present position, and adds this change to initial position to obtain present position. The horizontal velocity components, V_H and V_D, must therefore be transformed through the aircraft's heading angle, H, to obtain V_N and V_E. The vehicle's heading reference is used to provide H and can be a magnetic compass

or a more sophisticated and costly inertial navigation system. The transformation is usually performed in the Doppler itself, when it is configured as a navigator, or in a separate mission or navigation computer when the Doppler is used as a velocity sensor.

2.7 Types of Transmission Modulation

Four types of transmission systems have been used in Doppler radar sets: continuous wave (CW), frequency-modulated CW (FW-CW), coherent pulse, and incoherent pulse. A free-running highly stable CW oscillator is required to supply the transmitter energy in CW, FM-CW, and coherent pulse systems. In each case, the received signal is mixed with a portion of the stable transmitted signal (or reference oscillator) to achieve coherence, and the beat signal is detected.

In an incoherent pulsed system, succeeding pulses are emitted with random phase (as in the case wherein a magnetron is used for transmission). In such systems, the requisite coherence for recovery of Doppler information is achieved by mixing two echoes derived from the same transmitted pulse (Janus mixing); such a system is frequently called self-coherent. In an FM-CW system, the transmitter is frequency modulated, as the name implies. The received signal is mixed with a portion of the transmitted energy (thereby achieving coherence) and the Doppler shift in a selected sideband is used. The first or J_1 sideband is frequently used since its amplitude versus altitude is nearly flat at low altitudes. This system is relatively unresponsive to unwanted reflections from nearby structures such as radomes and airframe structure and thus is not likely to experience a false lock-on.

2.8 The Doppler Spectrum

The beamwidth of a beam of radar energy is inversely proportional to the antenna size. Aircraft installation constraints, and cost and weight considerations limit the size of practical antennas. Doppler radars with antennas that are 0.3 meters to 0.6 meters long have beamwidths of 8° to 4°. Since the beam from a microwave antenna cannot be made to have zero width, it overlaps a number of equal-frequency shift lines in the narrow direction. As the beam moves across the randomly situated scatterers, the result is a Doppler return consisting of a spectrum of frequencies, rather than a discrete frequency. The Doppler shift which must be measured is the center frequency of this spectrum. Since the illumination of the target surface is most intense at the center of the transmitted beam, and diminishes in power toward the edges, the result is a spectrum whose amplitude has an essentially normal (or Gaussian) distribution, as shown in Figure 4.

The Doppler spectrum of Figure 4 is idealized, since the actual nature of the Doppler return is complex and capable of description only in statistical terms. At any instant, the Doppler spectrum can be quite unlike the figure and may present a highly skewed picture. This can be seen by considering a microwave signal transmitted from the antenna and incident upon some small physical object. The object scatters a portion of the incident signal in every direction including the direction back toward the transmitter. The intensity of the back scattered signal depends on the size, shape, orientation, and electrical properties of the scatterer. In actuality, the reflector involved is not a single scatterer; a reasonably large area of the ground or of the surface of the sea is illuminated by the transmitted RF energy. Thus the target contains a large

number of randomly positioned, physically independent scattering centers. The net return signal available is then the sum of a large number of waveforms reflected by many scatterers. Each constituent waveform has an amplitude and a phase determined by the corresponding scattering center in the target area. Since these amplitudes and phases are randomly distributed quantities, the return signal can be described adequately only in statistical terms. The Doppler spectrum is statistically equivalent to narrow-band noise; the band center frequency is the desired mean Doppler frequency.
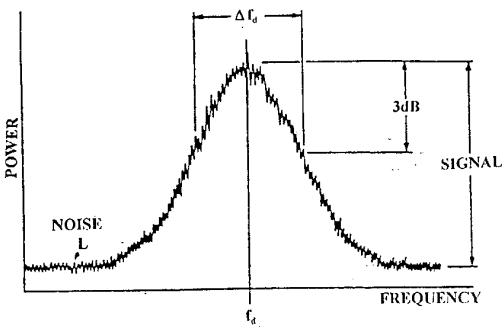
The Doppler spectrum width, Δf_d, at the half-power points is given by

$$\Delta f_d = \frac{2V}{\lambda} (\sin \gamma_o) \, \Delta \gamma. \qquad (6)$$

$\Delta \gamma$ is the round-trip half-power width of the antenna pattern in the γ coordinate.

A narrower spectrum width Δf_d results in a smaller instantaneous error in measuring the band center frequency. Equation (6) indicates that the spectrum width is reduced if γ_o is smaller, and if the beamwidth is smaller (a larger antenna).

Because of the appreciable spectrum width resulting from the use of a practical-sized antenna, the instantaneous frequency of the Doppler signal is therefore subject to random fluctuations about its mean value. The values of speed and distance measured by a Doppler system are, even with an ideal instrumentation, subject to error. This error will be discussed in Para. 3.1.

2.9 Measurement of the Doppler Frequency Shift

Frequency measurement is accomplished in a device called a frequency tracker. The frequency tracker must provide a single-valued output representative of the spectrum of frequencies comprising the Doppler signal. The frequency tracker can be a frequency discriminator, an autocorrelator, or an axis-crossing frequency meter. Most trackers are of the frequency discriminator type, and measure the centroid of power of the spectrum of Doppler frequencies. It has been found that frequency trackers generally measure this centroid with great accuracy except possibly for a small bias or constant offset of about 0.2 Km/hr or less. The effects of this bias offset are most noticeable in helicopters where velocity is usually low.

Figure 4. The Doppler Spectrum

As vehicle speed increases, the bias offset has relatively less effect.

In addition to determining the Doppler frequency shift, the frequency tracker generally does the following:

a. It tracks the signals; i.e., it provides a continuous measure of a frequency characterizing the input. It does this as the input frequency changes and in the presence of background noise. Time constants of typical frequency trackers are generally less than 25 m.s. to assure adequate tracking during maneuvers. When the effects of beam switching are taken into account then the effective time constant is somewhat greater. For example, a tracker with a 20 m.s. time constant and operating in a system with a 8 Hz beam switching rate has an effective time constant of about 100 m.s.

b. It acquires the signal when the system is turned on or after a period of signal loss. This is usually wholly automatic in modern systems; however, approximate velocities can be provided by external velocity sensors such as inertial navigation systems to reduce the acquisition time.

c. It provides a measure of the adequacy of the signal-to-noise ratio. Below some critical value, tracking is not reliable. In the presence of a substandard S/N, the operator receives a warning and certain functions are performed by other means.

d. When the S/N is below a minimum

acceptable value (or when the operator chooses to observe Doppler radar silence), it indicates memory and switches into its acquisition mode. When the system is in the memory mode, either the last measured velocity, or velocity computed from true airspeed and the last measured wind, can be provided as a substitute output (the last measured wind is computed, during normal Doppler operation, from Doppler ground velocity and air velocity; normally, the wind is averaged over some time period to smooth out gust effects).

2.10 Over Water Operation

Doppler system operation over water differs from operation over land in three important respects: (a) signal strength over water is less than over land, (b) the centroid of power of the returned signal is shifted by the radar back-scattering properties of water, and (c) the mean velocity of the scatterers which give rise to the sea return appears as a velocity error.

The back-scattering properties of an extended target medium can be specified by the scattering coefficient, as a function of incidence angle, ψ. (The scattering coefficient is defined as the ratio of power back-scattered by the target to that which would be back-scattered by a perfectly reflecting hemispherically isotropic scatterer). The scattering cross section per unit area, σ_o, is equal to $2 \cos \psi$ times the scattering coefficient, $f(\psi)$. (See reference 1, the chapter on "Sea Echo") Figure 3 gives the scattering cross section (σ_o) as a function of incidence angle (ψ).

The curves of Figure 3 are based on the GPL

sea scale (see reference 1) developed for use in Doppler radar set design, since the conventional Beaufort and Douglas sea scales were not sufficiently fine for adequate use. A comparison of the relationships among the GPL, Hydrographic Office Publication 606F, Beaufort, and Douglas scales is shown in Figure 5.

Note that over land, σ_o is essentially constant. Over water, however, σ_o decreases rapidly as the incidence angle, ψ, increases; the specific values of $\sigma_o (\psi)$ are functions of the sea state.

The nature of the first two phenomena, the reduction in signal strength over water and the change in calibration results in an error proportional to speed. In practice, this shift results in an error on the order of 1% to 5% in magnitude. Its exact value depends on beam shape, beam direction, and on sea state.

GPL	CODE PER HYDROGRAPHIC OFFICE PUB.606F	BEAUFORT	DOUGLAS
		0	0
½	0	1	1
1			
2			
3	1	2	2
4			
	2	3	3
	3	4	4
	4	5	5
	5	6	6
	6	7	7
		8	
		9	
	7	10	
	8	11	8
	9	12	

Figure 5. Approximate Relationships Among Various Water Surface Condition Scales

Several techniques have been developed to minimize the overwater shift. (a) The error varies substantially as the square of the γ beam width, so that narrow beams are desirable. This technique has the disadvantage that it increases the antenna size. (b) The use of a manual land-sea/switch to change the system calibration as a function of the terrain being traversed. The over-water calibration value corresponds to the required calibration over the most frequently expected sea state. (c) If the signal power in each slice of the beam at the same ψ angle had the same Doppler spectral composition as all other slices, then the calibration shift would be zero. Certain beam shaping techniques can result in approximating this situation. (d) A beam-lobing technique can be used to substantially reduce the over-water calibration error. Each antenna beam is switched a small amount in the γ direction at a low frequency (about 20 hz). The return signal then consists of two Doppler spectra existing alternately in time at the switching rate.

The cross-over of these two spectra is invariant with the sea-state since the return at the crossover point for each spectrum was derived from the same group of scatterers or incidence angle. Thus, even though the return signal spectra are distorted over water, the crossover point remains unchanged and the overwater shift is essentially eliminated.

Techniques (b) and (c) are most frequently used in modern Dopplers.

The Doppler frequency shift provides a measure of the velocity of the aircraft relative to the scattering surface. If the surface is moving, this velocity will include the surface velocity, thus producing an erroneous measure of aircraft velocity.

Water surface movements result from current flow and surface wind. Random sea current flow is generally no larger than 0.7 Km/hr. with random direction; such random velocities tend to average out for long flights. The Gulf Stream, whose direction and velocity are generally known, has a surface speed less than 6 Km/hr. Surface winds can cause water surface particle movements on the order of 4 to 8 Km/hr. or higher, depending on the surface wind speed. In general, if W = the surface wind, and ϵW = velocity of the surface particles as a result of W, then

$$\epsilon W \approx 1.28 W^{1/3} \ (W>2) \qquad (7)$$

Most of the effects of water surface motion can be compensated for in the navigation computer by manual insertion of known surface current velocity or surface wind conditions.

2.11 Doppler Signal-to-Noise Ratio

A Doppler radar set measures the Doppler frequency by processing the signal and noise within the Doppler spectrum bandwidth Δf_d;

$$\Delta f_d = \frac{2V}{\lambda} \sin \gamma_o \cdot \Delta \gamma \qquad (6)$$

where γ_o is the mean value, and $\Delta \gamma$ is round-trip half-power width, of the antenna pattern in the γ coordinate.
There is a minimum acceptable signal-to-noise ratio (S/N), within the Doppler bandwidth, below which the frequency tracker can no longer acquire the signal and determine the Doppler frequency shift within acceptable limits. This figure defines the frequency tracker sensitivity, and is a function of the tracker design. The altitude capability of a Doppler radar can be expressed as a function of the available

Doppler signal-to-noise ratio at the highest altitude, highest velocity, and over the worst terrain, as compared to the frequency tracker sensitivity.

Using the radar equation an expression for the Doppler S/N can be derived. The equation is valid for coherent systems using post-tracker mixing:

$$\frac{S}{N_d} = \frac{P_t \cdot G_o \cdot W_{rf} \cdot E \cdot f(\psi) \cdot \lambda_L \cdot F \cdot L \cdot \cos^2 \psi}{8\pi^2 \cdot NF \cdot K \cdot T \cdot \Delta f_d \cdot h^2} \qquad (8)$$

where

$\dfrac{S}{N_d}$ = Doppler signal-to-noise ratio (Doppler S/N).

= Ratio of the total Doppler signal power to the noise power in the bandwidth of the Doppler spectrum. The bandwidth of the Doppler spectrum is measured at the 3 dB points.

P_t = Average transmitted power per beam.

G_o = One-way maximum antenna gain relative to an isotropic radiator. Note that $G_o = \dfrac{kA}{\lambda^2}$, where

k is a constant, and A = antenna area.

W_{rf} = RF attenuation in the plumbing of the transmitter and receiver paths, including duplexing and wave guide losses.

E = Efficiency factor. (Ratio of available Doppler S/N to the Doppler S/N which would be available if all of the received signal were converted to Doppler information. This includes spectrum utilization in pulsed and

FM-CW systems, gating improvements, gating losses, and noise-fold-over losses.)

$f(\psi)$ = Scattering coefficient at the given ψ - angle. $(f(\psi) = 0.5\ \sigma_o \sec \psi$ where σ_o is the scattering radar cross-section per unit area.

λ = Transmitter wavelength

F = Antenna pattern parameter accounting for loss of power outside the main lobe 3 dB limits (normally between 0.5 and 0.67)

L = Attenuation in the atmosphere

NF = RF noise figure

K = Boltzmann's constant $(1.38 \times 10^{-23}$ watt-seconds/degree Kelvin)

T = absolute temperature (290° Kelvin)

Δf_d = Doppler bandwidth

h = Altitude above the terrain.

Since

$$f(\psi) = \frac{\sigma_o}{2 \cos \psi}$$

and $h = r \cos \psi$, where r=range to the terrain, equation (8) can be written

$$\frac{S}{N_d} = \frac{P_T \cdot G_o \cdot W_{rf} \cdot E \cdot \sigma_o \cdot \lambda_L \cdot F \cdot L}{16\pi_2 \cdot NF \cdot K \cdot T \cdot \Delta f_d \cdot r^2 \cdot \cos\psi} \qquad (9)$$

The term E in equation (9) for an FM-CW Doppler is given by

$$E = J_n^2\ (M) \qquad (10)$$

where M is the Modulation number:

$$M = 2\ m \sin \pi f_m \cdot \tau \qquad (11)$$

$$= 2\ m \sin 2\ \pi\ \frac{f_m \cdot r}{C} \qquad (12)$$

where τ is the round-trip time and r is the one-way distance to the ground.

Several of the parameters in equation (9) are of interest. First, the product $G_o \cdot \lambda^2$ is directly proportional to antenna area hence Doppler S/N is proportional to antenna area.

The Doppler S/N is a direct function of the scattering coefficient, $f(\psi)$ (and hence of the scattering cross-section, σ_o)

The Δf_d bandwidth is given by:

$$\Delta f_d = \frac{2V}{\lambda} \sin \gamma_o \cdot \Delta\ \gamma \qquad (6)$$

Thus, Doppler S/N is inversely proportional to the aircraft speed.

From equation (9), it is apparent that the parameters W_{rf} and NF must be kept as low as possible.

Frequency tracker sensitivity is expressed as acquisition sensitivity (the minimum S/N at which signal can be acquired and tracked), and drop-out sensitivity (the minimum S/N at

which the system can no longer track, and hence goes into memory). Typical values are 5 dB acquisition sensitivity and 3 dB drop-out sensitivity.

Typical values for S/N ratios over land at 300 meters (1000 ft.) AGL are 25 to 30 dB, and thus such Dopplers have a margin of 20 to 25 dB. Overwater operation can result in a 10 to 15 dB loss in S/N due to poor backscattering resulting in less margin.

Dopplers that will frequently fly over water should therefore have a higher S/N - typically 35 - 40 dB over land - to assure adequate margin over water.

2.12. Doppler ECM Considerations

Doppler radars transmit very low amounts of power - generally less than 250 m.w. - in narrow beams directed nearly straight down. The low power combined with very low side-lobes results in a low RF signature and thus considerable difficulty in detection by enemy ECM gear.

3.0 Doppler Radar Accuracy

A number of sources contribute to the total Doppler radar set error in measurement of velocity. These errors are either bias or systematic errors (those which remain essentially constant during operation) or random or non-systematic errors (errors which vary during operation). Some of the systematic errors can be reduced by calibration. Errors can also be classified as either slope type (those which vary as a percentage at speed) or zero type (those which are constant). The major sources of velocity error are discussed below.

3.1 Fluctuation Error

The Doppler frequency shift is a result of back-scattering from a large number of randomly positioned, physically independent scattering centers. The return signal is indistinguishable in its statistical properties from the signal obtained by passing white noise through a band-pass filter having a power transfer characteristic of the same functional dependence on frequency as the Doppler spectrum (white noise is defined as noise having a uniform power density at all frequencies of interest). The spectrum, to a first approximation, has a Gaussian shape, with a bandwidth proportional to the speed.

This noise-like nature of the Doppler spectrum gives rise to the Doppler fluctuation error. The relative fluctuation error, e_f can be expressed as

$$e_f = \frac{\sigma_v}{V} = \frac{1}{2 \cos \gamma_0} \cdot \frac{\lambda \cdot \Delta \gamma \cdot \sin \gamma_0}{VT} \quad (13)$$

$$= K_1 / VT = K_1 / D$$

where σ_v is the RMS velocity error, and $\frac{\sigma v}{V}$ is the relative velocity error.

Equation (13) shows that, since V and T appear only as a product, the relative or percentage error in velocity is a function only of the distance travelled, and is independent of speed. That this is true, and that the percentage error varies inversely with the square root of distance travelled, can be seen intuitively: as D increases, the Doppler information is derived from an increasing number of independent scatterers, and the amount of refinement of measurement (smoothing) increases.

3.1.1 Power Spectral Density

Many applications use Doppler velocities to

aid other sensors such as inertial navigation systems. It is useful, in listing Doppler requirements for such systems, to specify the power spectral density per unit of speed, call the Doppler P_o number.

The instantaneous difference between the true ground speed and the measured ground speed is a random function of time. Suppose this representation of ground speed error as a function of time is transformed into the frequency domain, with amplitudes of the error given as a function of the frequency components making up the error. Then the mean square error density in each frequency interval can be determined in units of $(knots)^2/$ (frequency interval). If the frequency intervals are made sufficiently small, a continuous distribution is approached. This distribution is called the ground speed error power spectral density, and its units are $(knots)^2/Hz$. If all of the error frequency components are summed (integrated) from $-\infty$ to $+\infty$, the result is the variance (square of the standard deviation) of the error distribution.

Since the true ground speed varies during a mission, the error power spectral density is normalized" and defined in units of $(knots)^2/Hz/knot$ of ground speed. The error power spectral density at any speed can be determined by multiplying the normalized error power spectral density function by the ground speed.

It can be shown that the error power spectral density is essentially constant in amplitude from virtually zero frequency up to the bandwidth of the Doppler radar frequency tracker, normally greater than ten Hz. P_o is defined as the error power spectral density in the vicinity of zero frequency, normally expressed as $(knots)^2/[(radian/sec/knot]$ (using angular frequency).

It can be shown that

$$P_o = \frac{D\,\sigma^2}{\pi \cdot V^2} \qquad (14)$$

Hence, P_o and e_f are related by the equation

$$P_o = \frac{D\,e_f}{\pi}^2 \qquad (15)$$

Note that P_o has the dimension $[(knots)^2/(rad/sec)]$ knot, or distance.

In a typical fixed-antenna Doppler radar, each of the three orthogonal velocity outputs is characterized by an associated P_o:

$$\left(\frac{EV_x}{V}\right)^2 = \frac{\pi\,P_{ox}}{D} \qquad (16)$$

$$\left(\frac{EV_y}{V}\right)^2 = \frac{\pi\,P_{oy}}{D} \qquad (17)$$

$$\left(\frac{EV_z}{V}\right)^2 = \frac{\pi\,P_{oz}}{D} \qquad (18)$$

Typical values for P_{ox}, P_{oy} and P_{oz} are 0.005, 0.005 and 0.002 respectively.

Equations (16) through (18) describe the effect of P_o on long-term velocity or position error. It can be shown that the equation for the "instantaneous" velocity error is $(EV_x/V)^2 = \pi\,P_{ox}/V/2\,\tau_v$,

where τ is the correlation time of the velocity error. The correlation time is effectively the frequency tracker time - constant as modified by the beam-switching effect. Similar equations occur for ϵV_y and ϵV_z.

3.2 Antenna Geometry Errors and Calibration

The frequency shift error, Δf_d, caused by an error in beam direction or $\Delta\gamma$ is given by

$$\Delta f_d = -\frac{2V}{\lambda} \sin\gamma \cdot \Delta\gamma, \qquad (19)$$

or

$$\frac{\Delta f_d}{f_d} = -\tan\gamma \cdot \Delta\gamma. \qquad (20)$$

The velocity component V_x can be derived from Equation 4, and is given by

$$V_x = \frac{\lambda}{2} \frac{[f_{d,A} + f_{d,B} + f_{d,C} + f_{d,D}]}{[\cos\gamma_A + \cos\gamma_B + \cos\gamma_C + \cos\gamma_D]} \qquad (21)$$

where it is assumed that all σ and all ψ angles are equal.

The EV_x is given by, approximately;

$$\frac{EV_x}{V_x} = -\frac{\tan\gamma_o}{4}[\Delta\gamma_A + \Delta\gamma_B + \Delta\gamma_C + \Delta\gamma_D] \qquad (22)$$

Equation (22) shows that errors in pitch stabilization tend to cancel. A positive (nose-up) error in pitch would increase the forward beam γ angles (γA and γB) and decrease the aft beam γ angles (γC and γD). Errors in roll stabilization would not affect V_x accuracy. Similar results occur for EV_y, except that the role of pitch and roll angles is reversed. Equation (22) is particularly important since it quantifies the dependence of Doppler performance on the calibration and stability of the beam geometry. As is the case with most radars, performance of a Doppler radar is very dependent upon the quality of the antenna - both design and fabrication. Errors in measuring or calibrating the geometry of each beam result in the velocity errors defined by Equation (22).

3.3 Transmission Frequency Errors

Equation (21) shows that transmission frequency directly affects the value of the measured Doppler frequency. However, the frequency stability of existing solid-state microwave sources is so tight that this error is normally quite small. Moreover, phased linear array and planar array antennas can be so designed as to make the Doppler beam geometry independent of transmission frequency.

3.4 Errors in Frequency Measurement

This error is a function of the instrumentation of the frequency tracker. The error can normally be expressed as a fraction of the Doppler spectrum width and hence as a percentage of velocity. Frequency trackers having errors on the order of 0.02% to 0.05% are within the current state of the art. Trackers can be in analog or digital form. Analog frequency trackers usually have a bias offset as well due to the analog components that are used. Digital trackers usually have a smaller bias error, and also can be programmed for faster acquisition times. Digital trackers are now being used in most new Dopplers.

3.5 Errors in Data Conversion

The measured Doppler frequencies representing the desired velocity components must be converted into some other form in order that the information can be used in a computer or by other equipments in the vehicle. This data conversion can deteriorate the accuracy of the velocity information. The error has been particularly significant in the earlier Doppler radars, which used relatively inaccurate conversion devices. Airborne digital computers accept Doppler

frequency data in digital form with negligible error, i.e., with an error which is limited only by the number of binary bits used.

3.6 Errors in Stabilization (or in Conversion from Vehicle to Ground Coordinates)

Since the basic Doppler velocity information is in vehicle coordinates and is generally required in ground or some other coordinates system, a conversion from the former to the latter is usually necessary. This conversion amounts to stabilization about the pitch and roll of the vehicle. In antenna stabilized systems, the Doppler outputs are in the desired coordinate system, and the stabilization error is primarily a function of the errors of the vertical reference and of the servo or servos controlling the antenna. In fixed antenna systems, Doppler velocity is in vehicle coordinates and is corrected for the pitch and roll of the vehicle in a computer which receives these angles from a vertical reference. In this case, the resulting error consists of the error of the vertical reference and the error of the stabilization computer. The error resulting from pitch and roll uncertainties is quite small - 0.014% per degree of uncertainty in pitch. Reduction of this error to extremely small values can be achieved through the use of an accurate vertical reference system and a high precision servo or stabilization computer. Note that the sensitivity of velocity errors to pitch and roll errors is fundamentally different from that of errors in beam direction. Pitch and roll errors affect all four beams equally and hence there is cancellation of "lst-order" effects. Residual effects are 2nd-order and of the magnitude given above. Beam direction errors - if identical in sign and magnitude - would behave similarly. Beam direction errors that differ from beam to

beam can cause significantly larger errors- of the order of 4% per degree.

3.7 Terrain Errors

The error resulting from the backscattering characteristics of the target area is called "terrain error". Over land, this error is insignificant. Over water, as explained in para. 2.10, the terrain error can be appreciable, as a result of changes in scattering coefficient (for various sea state conditions), as a function of beam incidence angle.

Water surface droplet motion and water current motion also cause Doppler velocity errors; however unlike the calibration shift error, these errors decrease in percentage as the vehicle velocity increases, and therefore tend to become acceptably small for most fixed wing aircraft. Helicopters, however, fly at lower speeds and water motion is an important error source.

3.8 Altitude Hole Errors

The effect of altitude holes for pulsed and FM-CW systems is to distort the Doppler spectrum and change its effective centroid of power. Since this phenomenon is related to the overwater calibration shift error, it can be seen that techniques that reduce the overwater shift also reduce the altitude hole errors.

3.9 Installation Errors

Errors in installation of the Doppler radar antenna relative to the aircraft reference axis contribute to the total system error. This reference axis is normally the heading and attitude reference system(s). The primary error in horizontal velocity results from

misalignment relative to the heading reference, and the primary error in vertical velocity is due to misalignment relative to the vertical reference.

3.10 Radome Errors

Most modern Doppler radars have integral or built-in radomes and are mounted so that these radomes are flush with the aircraft's surface. In these applications, radome errors are removed during the antenna calibration process. In a few special cases a separate radome is used and its effect on beam geometry, and hence velocity accuracy must be included in the overall error budget.

3.11 Total Doppler Error

On the assumption that systematic errors have been removed or reduced by some form of calibration procedure, the total Doppler velocity error can be obtained by combining the Doppler fluctuation error (for a desired distance or velocity and smoothing time), the non-systematic components of the other errors discussed above and the residual portion of the systematic errors.

Reference 1 contains a detailed discussion of Doppler radar errors.

4.0 Doppler Radar Design Configuration

Doppler radars are currently in production in two basic forms: as a velocity sensor, and as a self-contained dead-reckoning navigation system. The latter system configuration sometimes includes some form of external velocity aiding from sensors such as True Airspeed and inertial navigation systems. Doppler velocity sensors are externally aided to speed up acquisition time, to enable computation of wind speed and direction,

and for performance validation or BITE.

4.1 Typical Doppler Velocity Sensor Designs

Current Doppler radar velocity sensors are all solid-state designs, and use body-mounted antennas to minimize size, weight and cost. An integral radome simplifies installation in the airframe and eliminates the need for a separate radome. The antennas generally consist of some form of a printed circuit board with radiating elements printed on a dielectric substrate. The dielectric substrate is bonded to a metal ground plane thereby forming a very thin, lightweight and low cost antenna. One form of such an antenna utilizes microstrip-based elements to generate the beams. A typical antenna generates four beams from a single aperture. Separate transmit and receive antennas, or space-duplexing, are used when isolation between these two functions is critical, such as Doppler radars intended for operation over water where higher transmitter power is used to assure reliable operation. Doppler radars intended primarily for over land operation sometimes use a single antenna to simultaneously transmit and receive RF energy. Isolation between these two functions is achieved via RF circulators, and results in a lower cost design, but reduced S/N due to transmitter coupling or leakage into the receiver.

Today's transmitters are all solid-state devices such as Impatt Diodes, Gunn Diodes, or Dielectric Resonating Oscillators (DRO). The latter are usually excited by Gunn Diodes or FET devices. Typical power outputs are 50 milliwatts to 250 milliwatts. The higher power units can also have higher AM and FM noise so that the signal-to-noise ratio is not necessarily

increased in direct proportion to the increase in transmitter power output.

Solid-state microwave switches are used to direct the transmitter output to each of the ports on the antenna to form the desired beams. Two sets of switches are needed for space-duplexed antennas, whereas only one set is needed for a single antenna design. The microwave switches are usually combined or integrated with the receiver function. The latter consists of a device such as an RF circulator that taps off part of the transmitter output to serve as an input to an RF mixer. The other input to the mixer is the received signal from the ground.

The mixer output is amplified, filtered, converted to digital form and then sent to a digital signal processor to extract the Doppler frequency shift. The mixer output includes not only the spectrum of Doppler frequency shifts that is a measure of vehicle velocity, but also a signal that results from leakage of transmitter output directly into the receiver path. Leakage signals can be up to 40 to 50 dB higher than the spectrum to be measured, and must first be eliminated to enable processing of the desired signal. A Leakage Elimination Filter, or LEF, centered at zero frequency shift is usually employed to reduce leakage by 30 to 40 dB. This filter can be fairly broad for fixed-wing aircraft Dopplers, but must be extremely narrow (a few Hertz) for helicopter Dopplers to enable operation at hover.

The digital signal processor computes the centroid of power of the Doppler spectrum, and also the signal-to-noise ratio (S/N) of this spectrum. A low S/N results in a "no-track" or memory indication and the signal processor is switched into its signal acquisition mode. When the S/N increases above some predetermined level, a "Track"

state is indicated and normal operation resumes. During Memory, the Doppler radar outputs a Memory indication and its velocity data should be ignored. It is usually advisable to wait one to two seconds after Track is indicated before using the velocity data. Memory is usually indicated when S/N drops below +3 dB and acquisition or Normal mode is indicated when S/N is above +5 dB. Acquisition usually takes one to two seconds after the S/N of the received signal increases above 5 dB.

Doppler radars are available in one or two box configurations. In a 2-box system, the antenna, transmitter, receiver, and microwave switches are mounted into one unit, and the LEF, digital signal processor, interface, power supply and timer are mounted in a second electronics unit. A single-box configuration, of course, contains all of these elements in one box. Doppler radars are available with many types of interfaces although digital is most commonly specified.

Current Dopplers require less than 30 watts and often operate from 28VDC, although some systems use both 28VDC and 115V, 400 Hz A.C. Many current Doppler radars use frequency-modulated continuous wave (FM-CW) modulation of the carrier. The FM is sinusoidal with a modulation index of approximately one, and at a frequency of 30 KHz, resulting in an altitude hole at approximately 4600 meters. A second FM, usually at 40 KHz, is applied alternately with the first one to assure continuous operation through and above this altitude hole.

A Built-in-Test mode, or BIT, is incorporated to test Doppler functional operation during flight and to provide the flight crew with system status. A BITE function unique to Doppler radars is the

goodness check provided by the 4th or redundant beam velocity. This test is used during normal 4-beam operation, and also to verify proper lock-on during acquisition mode.

4.2   Doppler Radar Navigation Systems

The advent of very powerful, low-cost single-chip computers has enabled the development of low-cost Doppler navigation systems that provide present position and also guidance signals to destination or waypoint coordinates stored in non-volatile memory. Left-right steering signals, distance and time-to-go to these destinations are typically provided. The Doppler navigation system is provided with heading, pitch and roll from the vehicle's heading and attitude sensors. These sensors in moderate performance aircraft are usually a magnetic compass and a vertical gyro. More sophisticated aircraft may have inertial sensor-derived heading and attitude that have greater accuracy, but also at a higher cost. Figure 6 shows the position error in terms of CEP as a function of heading error, for 3 values of Doppler velocity error. It can be seen from this figure that the dominant error source is actually the heading reference and not the Doppler radar. Selection of the type of heading and attitude sensors is a function of the missions to be performed. In most cases these three angles are in synchro format although they are being provided in digital format in many new aircraft. Doppler velocity (three-dimensional) is then transformed through these angles into North, East and vertical components, and the North and East components are integrated to determine change in present position.
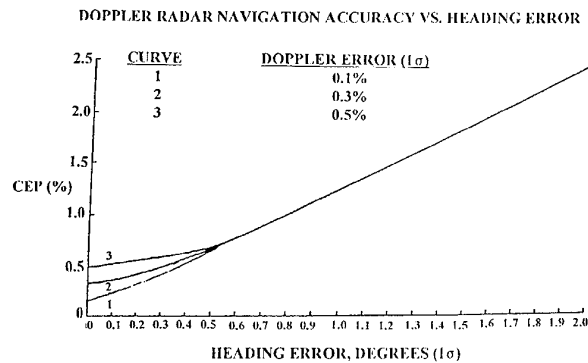
Figure 6.   Doppler Radar Navigation Accuracy vs. Heading Error

A Doppler navigation system computes change in present position. Computation of the actual present position therefore requires that some initial value of present position be inserted into the system. This is usually performed just prior to takeoff, or shortly afterwards when the aircraft flies over a known checkpoint.

Compensation for the three errors introduced when flying over water is performed as part of the navigation routine. Sea current data are entered manually, usually as magnitude and direction relative to North. Wind-induced water motion is removed by the computer using the empirically-derived relationship discussed in Para. 3. Wind data are obtained by either manually inserting wind magnitude and direction, or wind data is automatically computed from the difference of True Airspeed (if available) and Doppler velocity. In each case the flight crew must indicate to the navigation computer when the aircraft is flying over water via a "Land/Sea" switch. The use of a land-sea switch to indicate flight over water can also be used to apply an approximate correction for the over-water shift. Another approach is beam-shaping to automatically

compensate for this shift using the techniques discussed in Para. 2.10. Many Dopplers are now using this method successfully, and it will probably become the standard technique in virtually all future Dopplers.

Navigation during periods when the Doppler radar has lost track and is in the "Memory" mode, is normally performed using the last valid velocity data prior to loss of track. This "remembered" velocity is used until the Doppler resumes normal operation, or is replaced by a manually entered estimate of velocity by the flight crew if Memory is too long. Another method is to compute, during normal Doppler operation, the wind vector using True Airspeed (TAS) and Doppler velocity. During Doppler memory, an average value of the computed wind vector is added to True Airspeed to arrive at an estimate of vehicle velocity. Many airspeed sensors are single-axis and thus do not provide the cross-axis or transverse component of airspeed. The effect of this "missing" component of airspeed is an error in the change in present position during maneuvers and particularly during turns where the transverse component of wind changes with aircraft heading. Use of True Airspeed plus computed wind during Doppler memory will compensate for most of the error caused by loss of valid velocity data even when a single-axis TAS sensor is used.

Display and control of Doppler-derived navigation and guidance data has been provided in many cases by a Control/Display Unit or CDU dedicated to the Doppler navigator. Increasingly, the Control and Display function is being provided by a central or shared unit that is also used for other avionic systems such as fuel management and radios. The availability of map data bases resident in a few high density memory chips will now enable the CDU to also display the terrain, major man-made features, the aircraft's present location, and the optimum flight path to the destination.

The computational power of modern high-speed single chip computers enables present position to be provided in many different mapping coordinate systems. A typical Doppler navigator provides data in Latitude/Longitude and also in Military Grid Reference Systems, or MGRS. The latter is particularly important for Army aircraft since they must interact with ground forces that use MGRS for their own position determination and for artillery pointing and weapons delivery. Other coordinate systems can be provided as required by specific missions.

Current Doppler navigation systems can store large numbers of destinations or waypoints, and some systems now permit entry of sequences of waypoints to form routes. The flight crew enters only the predetermined route number and the Control/Display unit provides steering signals to each waypoint of that route, including automatic transition to the next waypoint. The flight crew can store present position coordinates while over a target, and recall these later in the flight for review and for taking further actions.

The CDU can display wind magnitude and direction, can indicate time-to-go to destinations, can display groundspeed, track-mode-good and other flight parameters. When the Doppler navigation system is operating in conjunction with a magnetic compass, the local magnetic variation and the vehicle's magnetic deviation effects can be entered for subsequent use in converting magnetic heading to true heading.

The Control and Display functions described above are being performed in a number of aircraft and helicopters by CDUs dedicated to Doppler navigation. Advanced and/or sophisticated aircraft are beginning to use central Control/Display systems in which the Doppler functions are but one subset of a large number of functions including radios, weapons delivery, fire-control and fuel management. These systems can also perform the navigation and guidance calculations, and the Doppler radar is then used only to provide vehicle velocity. The cost of digital processors may be low enough to allow a redundant navigation function to be performed in the Doppler for back-up purposes.

References

1. Myron Kayton and Walter R. Fried, Avionics Navigation Systems, John Wiley and Sons, Inc., New York, 1969.

2. M. I. Skolnick, Introduction to Radar Systems, McGraw-Hill Book Co. Inc., New York, 1962.

3. H. Buell, "Doppler Radars for Low Cost Medium Accuracy Navigation", Agard Conference Proceedings No. 176 on Medium Accuracy Low Cost Navigation, Sandefjord, Norway, September, 1975.

TERRAIN REFERENCED NAVIGATION

Drayton D. Boozer
Sandia National Laboratories
Albuquerque, NM, 87185-5800 USA

ABSTRACT
Terrain referenced navigation is a technique for improving the accuracy of a navigation system by correlating a sensed elevation profile of terrain beneath a vehicle with stored terrain elevation data. Position estimates are referenced to the terrain data and are insensitive to position bias errors in the terrain data. Because of this characteristic, terrain referenced navigation systems are especially useful in applications that require accurate navigation relative to targets, obstacles, structures, and other features whose locations are derived from the same source as the stored elevation data. Example applications include low-emission terrain following/terrain avoidance, target queuing for standoff weapon terminal sensors, indirect ranging and ground proximity warning. System navigation accuracy depends primarily on the ratio of terrain roughness to terrain data vertical accuracy and secondarily on navigation system accuracy, vehicle ground clearance, ground cover, vehicle maneuvers, and update frequency. Terrain referenced navigation systems are often considered for use with terrain masking for covert, low-altitude ingress into hostile areas. Low probability of intercept radar altimeters may be used in these applications. A key issue is the availability and quality of terrain elevation data. This chapter provides avionics system developers an overview of terrain referenced navigation system capabilities and characteristics, and an outlook for future applications.

ACRONYMS

AHRS	Attitude Heading Reference System
CEP	Circular Error Probable
DTED	Digital Terrain Elevation Data
DLMS	Digital Landmass System
DMA	Defense Mapping Agency
GLONASS	Global Orbiting Navigation Satellite System
GPS	Global Positioning System
INS	Inertial Navigation System
SITAN	Sandia Inertial Terrain Aided Navigation
TERCOM	Terrain Contour Matching
TERPROM	Terrain Profile Matching
TF/TA	Terrain Following/Terrain Avoidance
TRN	Terrain Referenced Navigation
WGS	World Geodetic System

1. INTRODUCTION
Two generic types of navigation systems are often used in military vehicles; inertial navigation systems (INS) and doppler/attitude heading reference systems (AHRS). INSs are typically used in fixed wing aircraft and doppler/AHRS in rotary wing aircraft.

Once aligned, an INS operates autonomously of external stimuli, but errors drift with time. INS stochastic error characteristics are well understood and can be accurately modeled. Doppler/AHRS navigation systems produce a navigation solution by integrating doppler velocity measurements transformed via the AHRS-determined vehicle attitudes. Errors drift with distance traveled rather than time, thus, doppler/AHRS systems are often preferred to INSs for rotary wing aircraft. Both types of navigation systems can use position updates from external sources to remove accumulated drift. Terrain referenced navigation (TRN) is a technique for updating the position of a navigation system by correlating a sensed elevation profile of terrain beneath an aircraft with stored digital terrain elevation data (DTED). TRN updates may occur over pre-planned mapped areas, or as they become available over large mapped areas. For simplicity, an INS-type navigation system used in an aircraft and updated over large geographical areas is emphasized in this presentation. The concept has been extended to standoff weapons, land vehicles, doppler/AHRS navigation systems, and limited geographical areas but these applications will not be discussed here.

2. CONCEPT
TRN consists of sensing a terrain elevation profile beneath an aircraft and correlating the profile with stored DTED to produce an estimate of aircraft position. For tractability, the approximate shape of the trajectory must be known and the correlation limited to areas of reasonable size. An INS, usually with barometric altimeter aiding, provides the approximate trajectory. TRN systems provide three dimensional position updates to the navigation system by estimating INS trajectory errors. A radar or laser altimeter measures ground clearance and the DTED gives terrain elevation above mean sea level (MSL). Implementation requires an INS, an altimeter, DTED, and a flight computer for executing the TRN algorithm. Figure 1 illustrates the sensing of the terrain profile along the true ground track and the terrain profile along the INS ground track.

3. THEORETICAL OVERVIEW
Estimating errors in an INS trajectory by measurements functionally related to the terrain profile beneath the aircraft requires application of nonlinear estimation theory. As shown in Figure 2, the sensed profile obtained along the true ground track is compared to DTED-derived profiles for possible horizontal locations of the aircraft within a position uncertainty region and a

fit-error surface generated. The sensed terrain elevation profile is obtained by subtracting the ground clearance measurements from aircraft altitude. The terrain elevation profiles from the DTED are produced by interpolating the DTED at the appropriate horizontal position along the translated INS ground track. The location with the best fit is the estimate of the aircraft's horizontal position.

Even in the simplest case of an INS with only position errors, error-free DTED, and an error-free altimeter, the correlation process may yield multiple solutions due to identical terrain profiles. Thus there is a fundamental issue of ambiguous solutions located in the aircraft initial position uncertainty region. Mean squared difference is the metric most often used for fit-error. In the error-free case the minimum of this surface is zero for both the correct and ambiguous solutions. Errors in the DTED and/or altimeter cause the surface to be positive at the minimum and increase the possibility of a position estimate far from the true position. The possibility of a "false fix" results from the nonlinear aspect of the estimation problem. Usually TRN performance specifications separate nonlinear (false fix) and linear (update accuracy) effects. The characteristics of system errors, the size of the uncertainty region, terrain characteristics and the aircraft ground track over the terrain are primary contributors to the probability of false fix [1].

The accuracy of the TRN position estimate for a simple case is derived from application of linear estimation theory. If one assumes independent profile measurement errors, the circular error probable (CEP) horizontal accuracy, where position biases between the actual and sensed trajectory in each of three dimensions are estimated, and terrain slopes are independent in the down-range and cross-range directions, is [2],

$$CEP = 1.17\ \sigma_n/(\sqrt{N}\ \sigma_s) \qquad (1)$$

where

CEP is the circular error probable of the update (m)

σ_n is the standard deviation of the profile measurement errors (m)

σ_s is the standard deviation of local terrain slopes at the measurement locations in both down-range and cross-range directions (unitless)

N is the number of independent measurements

σ_s parameterizes both terrain correlation length and elevation variation effects. While (1) suggests that CEP can approach zero for some choice of N given any σ_n, INS velocity errors make arbitrarily small error unachievable in practice. Because of the measurement geometry, aircraft altitude (vertical position) errors in DTED coordinates are estimated more accurately than horizontal position errors. For this case, altitude error standard deviation is $\sigma_n/\sqrt{N}$.
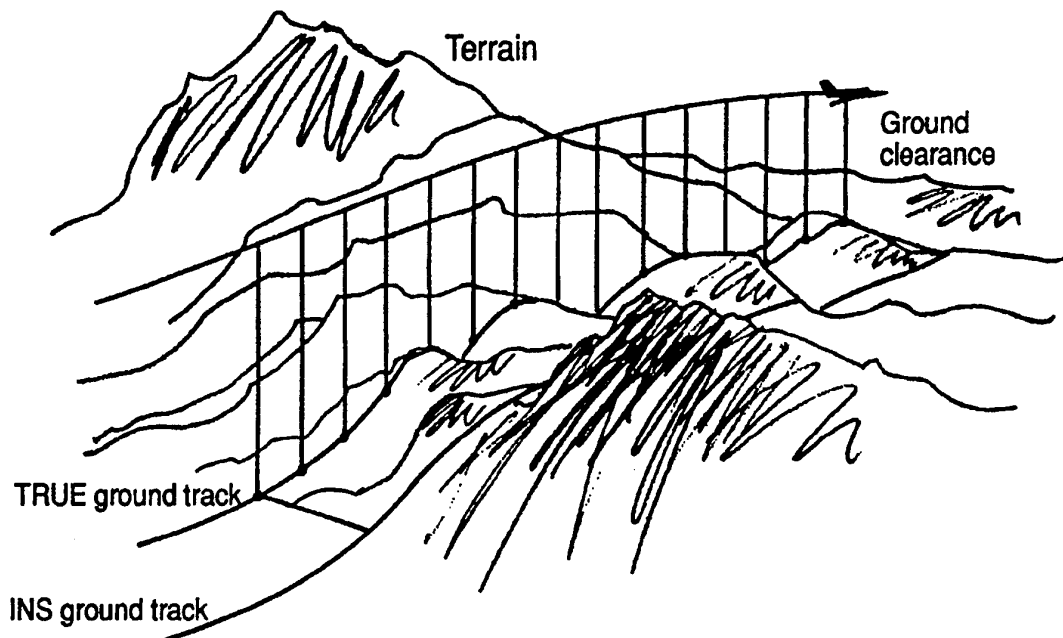
Figure 1. TRN Process
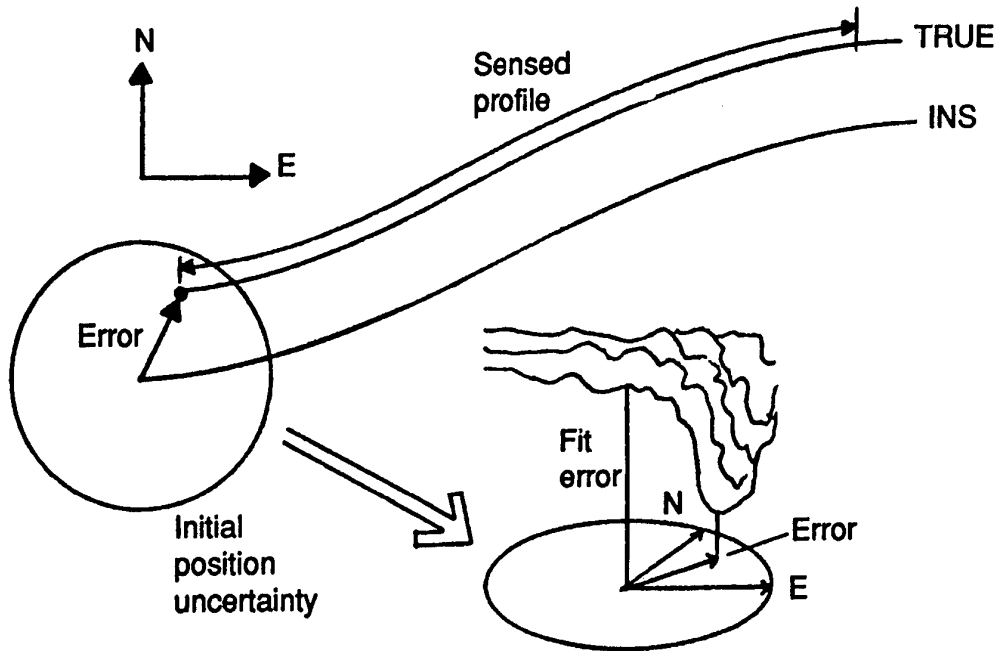
Figure 2. Horizontal Position Estimation

Using horizontal INS position errors modeled as independent random walks, no altitude error and first-principle covariance analysis [3] lead to the expression,

$$CEP_{ss} = 0.57 \, \delta V^{1/4} \, (\Delta d/s)^{3/8} \, (\sigma_n/h)^{3/4} \qquad (2)$$

where

CEP_{ss} is the steady state circular error probable of horizontal position updates (m)

σ_n is the standard deviation of the profile measurement errors (m)

h   is the deterministic local terrain slope at the measurement locations in both down-range and cross-range directions (unitless)

Δd is the distance between profile measurements (m)

s   is aircraft ground speed (m/s)

δV is INS maximum velocity error (m/s)

The primary value of (2) is that it shows the sensitivities of accuracy to implementation parameters. Steady state CEP is most sensitive to the ratio σ_n/h, least sensitive to δV, and nominally sensitive to the time between profile measurements $\Delta d/s$. Using typical values of

δV	=	1 m/s (1 nm/hr-class INS)
s	=	250 m/s
Δd	=	100 m
h	=	.05 (moderately rough terrain)
σ_n	=	15 m

results in a CEP_{ss} of 29 m. Because of the assumptions leading to (2), predictions should be treated as approximations, a conservative lower bound for TRN accuracy.

DTED errors are not explicitly considered in (1) and (2). Lower accuracy is achieved when the DTED vertical (elevation) errors have the same correlation length as the terrain itself. This is conceptually reasonable since in this case the correlation process must accommodate errors that undulate statistically like the terrain itself. The lack of statistical error models for the only broad-area database available, U.S. Defense Mapping Agency (DMA) Digital Landmass System (DLMS) Level 1 DTED, is the key reason TRN has not been developed to a theoretical level comparable to other navigation updating techniques like ground- and satellite-based ranging systems.

4. IMPLEMENTATION CONSIDERATIONS

Aircraft TRN systems typically consist of four basic elements: an INS, an altimeter, a DTED database, and an algorithm executed by a flight computer. The function of the INS is to provide a stable three-dimensional approximate trajectory whose primary errors are horizontal position and altitude bias errors. A 1 nm/hr-class INS is a common choice. Either strapped-down or gimbaled INSs can be used. A barometric altimeter is traditionally used to stabilize the vertical dimension of the INS and can be retained in TRN

systems. In either case, the navigation system vertical channel must be accurately modeled since TRN requires accurate measurement of the terrain profile.

DTED is a static database of terrain elevations over a prescribed geographical area. It may be produced from topographic maps of various scales or stereo photographs of different resolutions. The DTED with the widest geographic coverage is produced by the U.S. DMA. DLMS Level 1 DTED have been produced for about 70% of the world's landmass and are distributed on compact disk read-only memory (CD-ROM) media [4]. Level 1 DTED have an elevation value for each 3″ in latitude and longitude within ± 50° latitude of the equator. Larger horizontal post-spacings (in degrees longitude) are used at latitudes greater than 50° to maintain approximately constant spatial sampling of the terrain. Post-spacing is constant at 93 m in latitude but varies around this value in longitude. Algorithm design and implementation are simplified and computer throughput minimized when DTED have constant horizontal post-spacing. When using Level 1 DTED, this requires *a priori* preparation of the flight DTED. As flight computers become more capable, the trend toward using Level 1 DTED as produced by DMA directly in TRN algorithms will likely continue.

The function of the altimeter is to measure nadir ground clearance, the distance between the aircraft and the terrain at the same horizontal coordinates as the aircraft. Radar altimeters measure the distance to the closest point in the antenna beam. Unless modeled, the difference between the closest point and the nadir ground clearance is a measurement error whose magnitude is a function of aircraft ground clearance and attitude, antenna beam pattern, and terrain characteristics. These effects can be eliminated with a laser altimeter gimbaled to point at nadir. Radar altimeter antennas used on aircraft typically have broad beam patterns that accommodate significant aircraft maneuvers while continuing to measure the closest point beneath the aircraft. Such altimeters have been used in most TRN applications. Radar altimeter antennas with narrow beam patterns require either gimbals or limiting terrain sensing to specific aircraft pitch and roll attitude limits. Attitude limits are a function of aircraft ground clearance. Whenever radar altimeters are used, modeling of beam pattern effects in the TRN algorithm will improve performance, especially at higher ground clearances.

Altimeters typically provide ground clearance measurements to the flight computer at a rate significantly greater than the iteration rate of the TRN algorithm and spacing along the ground track that is small with respect to the DTED horizontal post-spacing. Depending on the algorithm design and DTED source, measurements may be either averaged or simply used as needed by the algorithm. Trees, foliage, ice, structures and other ground cover, as well as soil type, affect the altimeter measurements. TRN system performance depends on the altimeter measuring whatever is contained in the database. For example, if the DTED contain elevations of tree tops rather than the ground beneath the trees, the altimeter should range on treetops for best TRN correlation performance.

TRN algorithms must solve a nonlinear estimation problem in real time. Practical algorithms require a number of simplifying assumptions and in most cases were developed for specific aircraft avionics and for use in specific scenarios. Given this situation it is not surprising that a number of TRN algorithms have been proposed and developed over the last thirty years. The TERCOM approach developed for U.S. cruise missiles [5,6,7] is a batch-processing design that was developed for accurate, but limited, geographical coverage DTED, small flight computer processing throughput and memory, and an accurate INS. The missile flies a constant heading course over the DTED. TERCOM provides one position update for each DTED "patch" or matrix. Thus INS updates are infrequent. TERCOM was extended to accommodate arbitrary shaped trajectories over wide-area DTED [8]. Algorithms of this ilk must deal with aircraft maneuvers and lower quality DTED. Sandia Inertial Terrain Aided Navigation (SITAN) [9] was initially developed for a specific weapon application that required vehicle maneuvering, short flight distances, and relatively small initial position errors. These requirements led to an extended Kalman filter, recursive algorithm design that processes altimeter ground clearance measurements and produces an estimate of aircraft state every 100m along the aircraft ground track. A key idea is use of stochastic terrain linearization to permit use of a single extended Kalman filter when horizontal position errors are much larger than the local terrain correlation length. The initial application led to "track-mode" SITAN. Later, a bank of parallel filters was used with decision logic (an "acquisition mode") to accommodate larger initial position errors [10]. The SPARTAN algorithm [11], originally designed by A. R. Runnalls, uses maximum likelihood estimation in such a way that there is more measurement smoothing before incorporating measurement information into the Kalman filter than in SITAN but more frequent INS updates than in TERCOM. All current implementations use variations on one of these three basic approaches, e.g. terrain profile matching (TERPROM) [12] is based on an extended TERCOM acquisition mode coupled with a SITAN track mode.

Key implementation issues are the balance between on-board processing versus preflight mission planning and access to the stored DTED during flight. To minimize requirements on flight systems, the DTED may be formatted into a grid with constant horizontal post-spacing. Data compression may be used to minimize the amount of on-board storage at the expense of requiring in-flight reconstruction of the DTED. Storing only the DTED on the aircraft that may be used in a given mission lowers the required DTED storage capacity at the expense of requiring more pre-mission effort. Trends are to store the DTED of large areas on the aircraft and to use the DTED in the format in which it is supplied by DMA. During flight the DTED of the area being over flown are extracted from the large capacity storage device and placed in a buffer where they can be rapidly accessed by the TRN algorithm. How this is done is often determined by DTED access requirements of other system functions like low-emission TF/TA. The flight computer interfaces to the altimeter, INS, and cockpit displays are not usually affected by TRN.

The primary determinants of incremental flight computer resources needed to implement TRN beyond accessing the DTED and placing it in a buffer are the speed of the aircraft and the size of the largest horizontal position errors that the system must accommodate. Higher aircraft speeds require greater computer throughput. The required maximum position error depends primarily on the longest anticipated flight without updates and INS quality.

5. SYSTEM CONSIDERATIONS
TRN position estimates are referenced to the stored terrain data and are insensitive to bias errors in the terrain elevation data. Because of this characteristic, TRN systems are especially useful in applications that require accurate navigation relative to targets, obstacle, structures, and other features whose locations are derived from the same sources as the stored terrain data. Example applications include low-emission TF/TA, target queuing for standoff weapon terminal sensors, ground proximity warning and indirect ranging [13,14,15]. Three dimensional position bias errors made in the DTED production process can be estimated by a combined satellite navigation (GPS and/or GLONASS) and TRN system. This is especially important in systems using the DTED for TF/TA since DTED bias errors cannot be estimated by satellite navigation alone. Satellite navigation is performed in world-wide coordinates and has no reference to errors in the DTED used for TF/TA. Thus, even though the absolute accuracy of a satellite system may be superior to a TRN system in world-wide coordinates (e.g. World Geodetic System (WGS)), a TRN system may be more accurate with respect to obstacles and features within the

DTED because of errors made in registering the DTED to world-wide coordinates. Used alone, TRN systems can only estimate the total bias in each position dimension with respect to the INS position but cannot estimate the constituent parts of the total bias in each coordinate; DTED position bias errors and INS errors.

To improve underlying INS accuracy, TRN systems must be used in areas for which terrain of sufficient roughness and DTED of appropriate accuracy and quality are available. The area over which the TRN system searches for updates is an important consideration since it determines the TRN system's ability to correct INS drift error accumulated during periods of no updating. The larger the search area the greater the possibility of false fix. A false fix that is recognized by the TRN algorithm logic is not of particular concern, but large search areas do require greater care in design and testing of the TRN algorithm to ensure that no unrecognized false fixes are used as INS updates.

The robustness of the TRN system design to DTED error processes is perhaps the most important consideration for system developers. As used here, robustness is the ability of the design to navigate in the presence of a variety of DTED errors. Because of its broad-area coverage, Level 1 DTED is used by most TRN systems. TRN system performance using Level 1 DTED is very repeatable over the same geographical area but tends to vary from one geographical region to another. This is caused in part by terrain and ground cover differences but primarily by the varying characteristics of DTED errors. As one gains experience using the DTED in TRN systems, some areas may need to be reworked or the TRN algorithm changed to accommodate the errors discovered. This learning process is inevitable until world-wide DTED derived from the same source materials become available.

With very accurate and expensive-to-produce DTED, TRN system horizontal position accuracies rivaling those of GPS can be achieved. In TRN systems using Level 1 DTED over broad areas, accuracies in the range of 50-200 m CEP are typical for low-flying aircraft. Since TRN systems require terrain roughness and operate best at low ground clearance, they are especially attractive for use in covert attack TF/TA systems. Conversely, these are the flight regimens that cause satellite-based systems the most problems; satellite line-of-sight masking by the airframe or terrain, and signal jamming. Low probability of intercept radar altimeters are often considered for use in these applications.

Most TRN system implemented to date are add-on's to existing avionics systems. These systems are termed

"loosely coupled" because TRN INS error estimates are simply added to the output of the standard INS solution in software. Future systems will likely be more tightly integrated with real time estimation and correction of INS inertial sensor parameters.

6. OUTLOOK

The coincidence of commercial and military requirements in combination with advances in computer, communications and satellite technology will ensure availability of very accurate, high-quality, world-wide databases of both DTED and feature data early in the twenty-first century. By using satellite positioning, the absolute position of terrain and features will be known to sub-meter accuracy. These databases will be used pervasively in military aircraft to enable no-emission TF/TA, very accurate indirect ranging and ground proximity warning. TRN will be used in integrated avionics systems to provide navigation redundancy and to improve position estimation accuracy. The availability of significantly greater computer throughput, memory and mass storage will lead to implementation of more capable TRN algorithms; performance improvements will be greatest at higher ground clearances and over very rough terrain. Altitude position estimation accuracy will be improved over that available from satellite-based sensors alone. This is because altitude error is greater than horizontal error in satellite-based sensors, TRN is especially sensitive to altitude errors, and the DTED will be accurately positioned with respect to world-wide coordinates.

Over the next 10-to-20 years, TRN will be integrated with GPS/INS systems to improve low-emission TF/TA, indirect ranging and ground proximity warning performance when using Level 1 DTED. TRN will enable conversion of local DTED coordinates into world-wide coordinates in areas of sufficient terrain roughness.

Operational use of TRN by the U.S. is currently limited to the TERCOM system used for cruise missile guidance. TRN integration with GPS/INS in fixed- and rotary-wing attack aircraft is an area of current development in Europe, Australia and the U.S. Fielding of several operational systems during the decade is anticipated.

REFERENCES

1. Hinrichs, P R., "Advanced Terrain Correlation Techniques", IEEE Position Location and Navigation Symposium, 1972, p 94. [also in Kayton, Myron, NAVIGATION, LAND, SEA, AIR, & SPACE, IEEE Press, 1990, pp 288-295.]

2. Hostetler, L. D., and Andreas, R. D., "Nonlinear Kalman Filtering Techniques for Terrain-Aided Navigation", IEEE Transactions on Automatic Control, Vol. AC-28, No. 3, March 1983, p 320. [also in Sorenson, H. W., KALMAN FILTERING: THEORY AND APPLICATION, IEEE Press, 1985, pp 325-333.]

3. Maybeck, P. S., STOCHASTIC MODELS, ESTIMATION, AND CONTROL, VOL. 1, Academic Press, 1979, p 300.

4. U.S. Defense Mapping Agency, DIGITIZING THE FUTURE, 3rd Edition.

5. Sorrels, C A., U.S. CRUISE MISSILE PROGRAMS, McGraw-Hill, 1983, p 9.

6. Werrell, K. P., THE EVOLUTION OF THE CRUISE MISSILE, Air University Press, 1985, p 136.

7. Betts, R. K., ed., CRUISE MISSILE, TECHNOLOGY, STRATEGY, POLITICS, The Brookings Institution, 1981, p 37.

8. Bialecke, E. P. , and R. C. Lewis, "A Digital Terrain Correlation System for Tactical Aircraft", IEEE/AIAA 5th Digital Avionics Systems Conference Proceedings, Oct. 31-Nov. 3, 1983, p 14.1.1.

9. Hostetler, L. D., "A Kalman Approach to Continuous Aiding of Inertial Navigation Systems Using Terrain Signatures", IEEE Milwaukee Symposium on Automatic Computation and Control Proceedings, April 1976, pp 305-309.

10. Boozer, D. D., Lau, M. K., Fellerhoff, J. R., "The AFTI/F16 Terrain-Aided Navigation System", Proceedings of the IEEE National Aerospace and Electronics Conference Proceedings, May 1985, pp 351-357.

11. Henley, A. J., "Terrain Aided Navigation - Current Status, Techniques for Flat Terrain and Reference Data Requirements", IEEE Position Location and Navigation Symposium Record, March 20-23, 1990, pp 608-615.

12. Robins, A. J., "Recent Developments in the "TERPROM" Integrated Navigation System", Proceedings of the Institute of Navigation 44th Annual Meeting, June 21-23, 1988, pp 58-66.

13. Beal, C., "World in a box: Air Navigation Leaps Forward", International Defense Review, Vol. 25, May 1992, pp 417-422.

14. Priestley, N., "Terrain Referenced Navigation", IEEE Position Location and Navigation Symposium Record, Mar 20-23, 1990, pp 482-489.

15. Bennett, P. J., "Enhanced Navigation and Displays from Passive Terrain Referenced Avionics", Proceedings of the IEEE National Aerospace and Electronics Conference Proceedings, May 1988, pp 209-216.

SATELLITE NAVIGATION

Richard B. Langley
Geodetic Research Laboratory
Department of Geodesy and Geomatics Engineering
University of New Brunswick
Fredericton, N.B., E3B 5A3
CANADA

1. INTRODUCTION

Satellire navigation provides unprecedented accuracy and world wide coverage to aerospace vehicle. This section reviews in some detail the NAVSTAR Global Position System (GPS) followed by a survey of other satellite navigation systems:

2. SATELLITE NAVIGATION SYSTEMS

3. NAVSTAR GLOBAL POSITIONING SYSTEM

3.1 Introduction

The Navstar Global Positioning System (GPS) is a satellite-based positioning system currently in operation by the United States Department of Defense. Work on the system began in 1973 as a result of the merger of the U.S. Air Force's 621B Project. Both of these programs had been established in the mid-1960s to develop a passive navigation system using measured ranges. GPS consists of three segments: the satellites, the control system, and the users.

3.2 Satellites

When fully deployed by late 1993 or early 1994, the satellite segment will consist of a constellation of 21 Block II or second-generation satellites plus 3 in-orbit operating spares. The satellites are arrayed in 6 orbital planes inclined 55° to the equator. Each orbit is circular with a nominal altitude of 20 183 km. The corresponding orbital period is 12 sidereal hours, one half of the earth's period of rotation.

The deployment of the Block II satellites was preceded by a program of testing using prototype, or Block I, satellites. The Block I satellites were launched from Vandenburg Air Force Base in California using Atlas F rockets. 10 of these satellites were successfully launched between February 1978 and October 1985. They were placed in nominally circular orbits with semimajor axes of about 26,560 km. The satellites were positioned in two orbit planes with inclinations of about 64° to provide maximum coverage for the main military testing area for GPS, the Yuma Proving Grounds in Arizona.

Whereas the Block I satellites were launched using an expendable launch vehicle, it was originally intended to launch the Block II satellites, up to three at a time, using the Space Transportation System – the Space Shuttle. But after the Challenger accident, the decision was made to use an expendable launch vehicle and a new rocket, the Delta II, was developed for this purpose. The Block II satellites are launched from the Cape Canaveral Air Force Station next door to the Kennedy Space Center.
-- Block IIA
-- Block IIR
-- Block III
The satellites achieve their final orbits in steps. The first and second stages of the rocket together with the nine solid-fuel strap-on booster engines put the third stage of the rocket and the attached satellite into an elliptical orbit with a perigee height of about 180 km and an apogee height of about 870 km. (The *apogee* is the point on the orbit furthest from the earth.) The third stage, called a Payload Assist Module (PAM), is then used to increase the apogee height of the orbit so that it matches the height of the desired final orbit, about 20,200 km,

above the earth's surface. At this point, the satellite is in a highly elliptical orbit, called a transfer orbit. The PAM is jettisoned and the satellite's orbit insertion engine is fired to put the satellite into an approximate final orbit. Small adjustments are then made using the satellite's thrusters. The orbits of the Block II satellites are nominally circular (maximum eccentricities are about 0.01) with a semimajor axis of about 26,560 km and an inclination of about 55°. The resulting orbital periods of the satellites are within a minute or so of exactly one half of a sidereal day. In order to achieve global, 24 hour a day coverage, four satellites will be placed in each of 6 orbital planes, named A through F. The right ascensions of the ascending nodes of orbits in adjacent planes are separated by 60°. This arrangement of the satellite constellation results in at least six satellites being visible at all times.

3.3 Signals

3.3.1 *The Carriers.* Each GPS satellite transmits signals centred on two microwave radio frequencies, 1575.42 MHz, referred to as Link 1 or simply L1, and 1227.60 MHz, referred to as L2. These channels lie in a band of frequencies known as the L band which starts just above the frequencies used by cellular telephones. Within the L band, the International Telecommunications Union, the radio regulation arm of the United Nations, has set aside special sub-bands for satellite-based positioning systems. The L1 and L2 frequencies lie within these bands.

Such high frequencies are used for several reasons. The signals, as we have said, consist of a number of components. A bandwidth of about 20 MHz is required to transmit these components. This bandwidth is equal to the whole FM broadcast band! So a high, relatively uncluttered part of the radio spectrum is required for GPS-type signals. The GPS signals must provide a means for determining not only high accuracy positions in real-time, but also velocities. Velocities are determined by measuring the slight shift in the frequency of the received signals due to the Doppler effect   In order to achieve velocities with centimetre-per-second accuracies, centimetre wavelength (microwave) signals are required.

A further reason for requiring such high frequencies is to reduce the effect of the ionosphere. The ionosphere affects the speed of propagation of radio signals. The range between a satellite and a receiver derived from measured signal travel times will therefore be in error. The size of this error gets smaller as higher frequencies are used. But at the L1 frequency it can still amount to 30 metres for a signal arriving from directly overhead. For some applications, an error of this size is tolerable. However there are applications that require much higher accuracies. This is why GPS satellites transmit on two frequencies. If measurements made simultaneously on two well-spaced frequencies are combined, it is possible to remove almost all of the ionosphere's effect.

Although high frequencies are desirable for the reasons just given, it is important that they not be too high. For a given transmitter power, a received satellite signal becomes weaker the higher the frequency used. The L band frequencies used by GPS are therefore a good compromise between this so-called *space loss* and the perturbing effect of the ionosphere.

GPS signals, like most radio signals, start out in the satellites as pure sinusoidal waves or *carriers*. But pure sinusoids cannot be readily used to determine positions in real-time. Although the phase of a particular cycle of a carrier wave can be measured very accurately, each cycle in the wave looks like the next so it is difficult to know exactly how many cycles lie between the satellite and the receiver. This ambiguity can be resolved using the differential technique pioneered by surveyors [Wells and Kleusberg, 1990] but can be time consuming.

In order for a user to obtain positions independently in real-time, the signals must be modulated; that is, the pure sinusoid must be altered in a fashion that time delay measurements can be made. This is achieved by modulating the carriers with *pseudo-random noise (PRN) codes*.

These PRN codes consist of sequences of binary values (zeros and ones) that at first sight appear to have been randomly chosen. But a truly random sequence can only arise from unpredictable causes which, of course, we would have no control over and could not duplicate. However, using a mathematical algorithm or special hardware devices called *tapped feedback registers*, we can generate sequences which do not repeat until after some chosen interval of time. Such sequences are termed *pseudo*-random. The apparent randomness of these sequences makes them indistinguishable from certain kinds of noise such as the hiss heard when a radio is tuned between stations or the "snow" seen on a television screen. Although noise in a communications device is generally unwanted, in this case the noise is very beneficial.

Exactly the same code sequences are independently replicated in a GPS receiver. By aligning the replicated sequence with the received one and knowing the instant of time the signal was transmitted by the satellite, the travel time, and hence the range can be computed. Each satellite generates its own unique codes, so it is easy for a GPS receiver to identify which signal is coming from which satellite even when signals from several satellites arrive at its antenna simultaneously.

3.3.2 *The C/A-code.* Two different PRN codes are transmitted by each satellite: the C/A or coarse / acquisition code and the P or precision code. The C/A-code is a sequence of 1,023 binary digits or *chips* which is repeated every millisecond. This means that the chips are generated at a rate of 1.023 million per second and that a chip has a duration of about 1 microsecond. Each chip, riding on the carrier wave, travels through space at the speed of light. We can therefore convert a time interval to a unit of distance by multiplying it by this speed. So one microsecond translates to approximately 300 metres. This is the *wavelength* of the C/A-code.

Because the C/A-code is repeated every millisecond, a GPS receiver can quickly lock onto the signal and begin matching the received code with the one it generates.

3.3.3 *The P-code.* The precision of a range measurement is determined in part by the wavelength of the chips in the PRN code. Higher precisions can be obtained with shorter wavelengths. To get higher precisions than are afforded by the C/A-code, GPS satellites also transmit the P-code. The wavelength of the P-code chips is only 30 metres, one-tenth the wavelength of the C/A-code chips; the rate at which the chips are generated is correspondingly 10 times as fast: 10.23 million per second. The P-code is an extremely long sequence. The pattern of chips does not repeat until after 266 days or about 2.35×10^{14} chips! Each satellite is assigned a unique one-week segment of this code which is re-initialized at Saturday/Sunday midnight each week.

The GPS PRN codes have additional useful properties. When a receiver is processing the signals from one satellite, it is important that the signals received simultaneously from other satellites not interfere. The GPS PRN codes have been specially chosen to be resistant to such mutual interference. Also the use of PRN codes results in a signal that is essentially impervious to unintentional or deliberate jamming from other radio signals. The possibility of deliberate jamming of GPS signals is something that the U.S. Department of Defense, the owners of the system, have to worry about.

At the present time, the C/A-code is modulated onto the L1 carrier whereas the P-code is transmitted on both L1 and L2. This means that only users with dual frequency GPS receivers can correct the measured ranges for the effect of the ionosphere. Users of single frequency receivers must resort to models of the ionosphere which account for only a portion of the effect. It is access to the lower accuracy C/A-code which is provided in the GPS *Standard Positioning Service* (SPS), the level of service authorized for civilian users. The *Precise Position Service* (PPS) provides access to both the C/A-code and the P-code and is designed (primarily) for military users. The SPS incorporates a further intentional degradation of accuracy, called *Selective Availability* (see 3.3.7).

3.3.4 *The Broadcast Message.* In order to convert the measured ranges between the receiver and the satellites to a position, the receiver must know where the satellites are. To do this in real-time requires that the satellites broadcast this information. Accordingly there is a message superimposed on both the L1 and L2 carriers along with the PRN codes. Each satellite broadcasts its own message which consists of orbital information (the *ephemeris*) to be used in the position computation, the offset of its clock from GPS system time, and information on the health of the satellite and the expected accuracy of the range measurements. The message also contains *almanac* data for the other satellites in the GPS constellation as well as their health status and other information. The almanac data, a crude description of the satellite orbit, is used by the receiver to determine where every satellite is. It uses this information to quickly acquire the signals from satellites which are above the horizon but are not yet being tracked. So once one satellite is being tracked and its message is decoded, acquisition of the signals from other satellites is quite rapid.

The broadcast message also contains another very important piece of information for receivers that track the P-code. As we mentioned, the P-code segment assigned to each satellite is 7 days long. A GPS receiver with an initially unsynchronized clock has to search through its generated P-code sequence to try to match the incoming signal. It would take many hours to search through just one second of the code, so the receiver needs some help. It gets this help from a special word in the message called the *hand-over word* (HOW) which tells it where in the P-code to start searching.

The GPS broadcast message is sent at a relatively slow rate of 50 bits per second, taking 12.5 minutes for all the information in the message to be transmitted. To minimize the delay for a receiver to obtain an initial position, the ephemeris and satellite clock offset data are repeated every 30 seconds.

The C/A code and P code chip streams are separately combined with the message bits using *modulo 2 addition*. This is just the binary addition that computers and digital electronics do so well. If the code chip and the message bit have the same value (both 0 or both 1) the result is 0. If the chip and bit values are different, the result is 1. The carriers are then modulated by the code and message composite signal. This is readily done with the L2 channel as it only carriers the P-code. But the L1
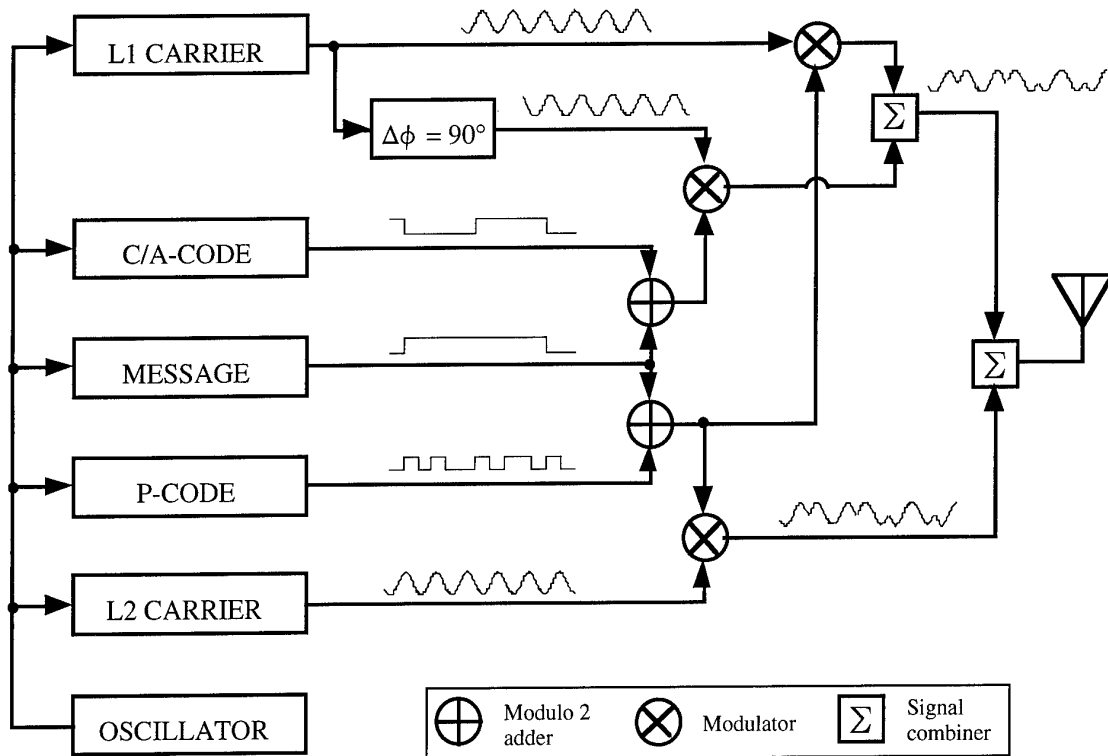
Figure 3.1 How the different components of the GPS signal are combined. Note that the various waveforms are not to scale.

channel has to carry both the P-code and the C/A-code. This is achieved by a clever technique known as *phase quadrature*. The P-code signal is superimposed on the L1 carrier in the same way as for the L2 carrier. To get the C/A-code signal on the L1 carrier, the unmodulated carrier is tapped off and this tapped carrier is shifted in phase by 90°. This quadrature carrier component is mixed with the C/A-code signal and then combined with the P-code modulated in-phase component before being transmitted by the spacecraft antenna.

3.3.5 *Binary Biphase Modulation.* Carrier waves can be modulated in a number of ways. The amplitude of the carrier can be varied, the frequency can be varied, or the phase can be varied. Phase modulation is the approach used for the GPS signals. Because the PRN codes and the message are binary streams, there must be two states of the phase modulation. These two states are the normal state, representing a binary 0, and the mirror image state, representing a binary 1. The normal state leaves the carrier unchanged. The mirror image state results in the unmodulated carrier being multiplied by -1. Therefore a code transition from 0 to 1 (normal to mirror image) or from 1 to 0 (mirror image to normal) each involves a phase reversal or a phase shift of 180°. This technique is known as *binary biphase modulation*. An interesting property of binary biphase modulation was exploited by one of the first commercially available GPS receivers, the Macrometer™. By electronically squaring the received signal, all of the modulation is removed leaving a pure carrier. The phase of the carrier could then be measured to give the ambiguous range measurements used by surveyors. Of course, the broadcast message was lost in the process and so orbit data had to be obtained from an alternate source.

The composite GPS signal consists then of carriers modulated by the PRN C/A and P codes and the broadcast message. The combining of these different components is illustrated in Figure 3.1. Forgetting for a moment that GPS is a ranging system, we could consider the satellites to be simply broadcasting a message in an encoded form. The bits of the message have been camouflaged by the PRN code chips. The effect of this camouflaging is to increase the bandwidth of the signal. Instead of occupying only a fraction of one kiloHertz, the signal has been spread out over 20 MHz. Inside a GPS receiver, the code matching operation de-spreads the signal allowing the message to be recovered. Clearly this can only be done if the receiver knows the correct codes. The de-spreading operation conversely spreads out any interfering signal considerably reducing its effect. This is a common technique, especially in military circles, for ensuring security and combatting interference and is known as *direct sequence spread spectrum communication*. Spread spectrum signals have the additional property of limiting the interference from signals reflected off nearby objects (*multipath*).

The features which the GPS signal is called upon to provide are summarized in Table 3.1.

3.3.6 *GPS Time.* The timing and frequency control for the carriers, the PRN codes, and the message all comes from an atomic oscillator on board the satellite. For redundancy each Navstar GPS Block II satellite carries four oscillators — two cesium beams and two using rubidum vapour cells. One of the oscillators is selected by the spacecraft controllers to provide the frequency and timing requirements for generating the satellite's signals.

The signals transmitted by the GPS satellites are referenced to GPS (System) Time. Until June 1990, this was the time kept by a single atomic clock at one of the monitor stations. However, the practice now is to obtain GPS Time from a composite or "paper" clock consisting of all operational monitor station and satellite clocks.

GPS Time is steered over the long run to keep it within about 1 microsecond of UTC ignoring leap seconds. So, unlike UTC, GPS Time has no leap second jumps. At the integer second level, GPS Time was equal to UTC in 1980 but at the present time, due to the leap seconds which have been inserted into UTC, GPS Time is ahead of UTC by 7 seconds plus a fraction of a microsecond that varies day to day.

A particular epoch is identified in GPS Time as the number of seconds that have elapsed since the previous Saturday / Sunday midnight. Such a time measure is, of course, ambiguous so that one must also indicate which week the epoch is in. GPS weeks are numbered consecutively with week 0 starting on 6 January 1980. Epochs may also be identified by Julian Date (JD), the number of days and fractional days elapsed since noon UT on 1 January 4713 B.C., or the Modified Julian Date (MJD) which is equal to JD minus 2400000.5 JD and MJD are frequently used by astronomers, navigators, and others for compactly and unambiguously identifying a particular epoch in time.

Atomic clocks perform best if they are not continuously adjusted. So the clocks in the GPS satellites are only approximately synchronized to GPS Time. The offsets of the satellite clocks from GPS Time, which can be a millisecond or so are carefully monitored by the GPS Operational Control System and the United States Naval Observatory (USNO). For each satellite clock an offset at an initial epoch, a linear drift term, and for rubidium clocks a drift rate of change term are determined. The parameters are uploaded to the corresponding satellite and subsequently included in its navigation message. A GPS receiver uses the satellite clock data to convert the measured pseudoranges from the satellite time scale to GPS Time. Also included in the satellite message is the offset of GPS Time with respect to UTC.

When a GPS receiver initially acquires signals, its clock will, in general, have a large unknown offset with respect to GPS Time. This offset, however, will contribute the same timing bias to all pseudorange measurements made at any particular epoch and can be solved for along with the receiver coordinates. Once determined, the bias can be used to synchronize the receiver clock to GPS Time. GPS Time or UTC can then be displayed by the receiver, used to time-tag recorded data, or used to generate a one pulse per second electrical signal for controlling other equipment. The relationship among the satellite, receiver, and system time

Table 3.1 Required GPS Signal Properties

- Work with one-way measurements (receive only)

- Serve tens of thousands to millions of both military and civilian users

- Provide accurate, unambiguous, real-time range measurements

- Provide accurate Doppler shift measurements

- Provide accurate carrier phase measurements

- Provide a broadcast message

- Provide ionospheric delay correction

- Allow simultaneous measurements from many satellites

- Have interference protection

- Have multipath tolerance

scales and pseudorange measurements is shown schematically in Figure 3.2. The raw measurement made by the receiver is the time interval $d\tau$ which, multiplied by the speed of light, c, gives the measured pseudorange, p. Correcting this measurement for the offsets with respect to GPS Time of the satellite clock, dt, and receiver clock, dT, gives the true geometric range ρ (ignoring propagation delays and other potential biases).
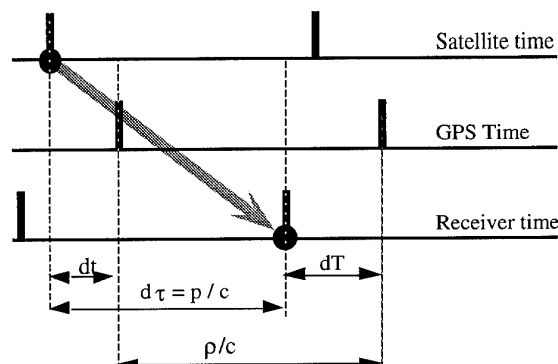
Figure 3.2 Relationships among GPS Time and satellite and receiver time scales.

Several manufacturers offer GPS receivers specifically designed to be used as sources of precise time information. Generally operated from a fixed site, such a receiver once it has accurately determined its location can provide synchronized time signals even when only one satellite is in view. Accuracies to within about 100 ns are readily achievable.

The USNO, BIPM, and others have developed sophisticated techniques for using GPS to synchronize clocks to a precision of 10 ns or better even when the clocks are on different continents.

3.3.6.1 *Relativistic Effects.* The frequency output of the atomic clocks in the satellites is 10.23 MHz. This fundamental frequency corresponds to the chipping rate of the pseudorandom noise P-code and, when divided by 10, gives the rate of the C/A-code. The fundamental frequency is multiplied by 154 to produce the L1 carrier frequency and by 120 to produce the L2 carrier frequency. Actually, in order to account for the effects of relativity, the fundamental frequency of the satellite clocks is not set exactly at 10.23 MHz but very slightly less.

Einstein showed in his Special Theory of Relativity published in 1905 that a clock moving with a constant speed relative to another clock will appear to run more slowly. A clock in a satellite travelling in a circular orbit about the earth would accordingly lose time compared to one on the ground. But 11 years later, in 1916, Einstein in his General Theory of Relativity deduced that clocks in different gravitational potentials will also appear to run at different rates. Due to the difference in gravitational potential, a clock in a satellite would appear to run faster than one on the ground. The net effect on a satellite clock is the combination of the two effects. A clock in a GPS satellite in a circular orbit with a nominal radius of 26,560 km, gains 38.4 μs per day compared to one on the ground. This time difference corresponds to a relative frequency offset of its oscillator of 4.45×10^{-10}. In order to compensate for this, the fundamental frequency of the satellite clocks is reduced by 0.00455 Hz to 10.22999999545 MHz.

If the GPS satellites were in circular orbits no further action would be required to compensate for relativity to give ranging

162

accuracies at the metre level. However, the orbital eccentricity of a GPS satellite can be up to 0.02 which means that both its speed and the gravitational potential it experiences change with time. The result is an oscillating time offset which is proportional to eccentricity and varies sinusoidally with the position of the satellite in its orbit. The magnitude of this effect can be as large as 45.8 ns which corresponds to a ranging error of 13.7 metres. A GPS receiver must correct its measured pseudoranges for this variation using the satellite orbit description contained in the navigation message.

3.3.7 *Selective Availability and Anti-spoofing.* GPS was initially developed to serve the needs of the DoD and the military forces of U.S. allies. In keeping with its national security mandate, the DoD plans to limit the real-time absolute positioning capability of GPS afforded to most civil users. The service to be provided to all GPS users on a continuous, worldwide basis without direct charge is called the Standard Positioning Service (SPS). SPS users will have access to the coarse acquisition (C/A) code and navigation message on the L1 frequency. The accuracy afforded by SPS has been intentionally degraded through a process known as Selective Availability [Georgiadou and Doucet, 1990]. The planned predictable horizontal positioning accuracy to be provided by SPS is 100 metres 2 drms at the 95% probability level with a corresponding vertical accuracy of 156 metres (2 sigma). At the 99.99% probability level, the accuracy will be 300 metres. During the buildup of the GPS constellation, the precision (P) code transmitted on both the L1 and L2 frequencies will generally be available to all users. However, when GPS is declared fully operational, access to the P code by most civil uses will be denied through the use of an encryption technique known as Antispoofing.

A higher accuracy service, the Precise Positioning Service, will be available to DoD-authorized users. The PPS provides access to the P code on both the L1 and L2 frequenies and the undegraded navigation message. The predictable accuracy of the SPS is 17.8 metres horizontal (2 drms) and 27.7 metres vertical (2 sigma).

The PPS community includes civil and military users in the U.S. federal and allied governments and a limited number of other users who can demonstrate to the DoD that their use of the PPS is in the U.S. national interest, that certain security requirements can be met, and that a reasonable alternative to the use of the PPS is not available.

Through the use of differential techniques, SPS users can achieve positioning accuracies that are higher than those otherwise available. In differential GPS (DGPS), a monitor station operating at a known location provides range corrections to users' receivers using some sort of communication link. It is also feasible to transmit the actual data collected by the monitor station rather than range corrections. The user's receiver combines the monitor station's data with its own to obtain a high accuracy relative position [DeLoach, 1991]. A non-real-time implementation of this technique was pioneered by surveyors using the GPS prototype constellation in the early 1980s.

3.4 Control Segment
In order for a GPS receiver to compute its position in real time, it must know the positions of the satellites. These positions are established by the GPS Operational Control System (OCS) and provided in the navigation messages broadcast by the satellites. The OCS, which is operated by the Air Force Space Command, includes five tracking stations, spaced in longitude around the globe. Three of these stations are on small islands, primarily used for military purposes: Ascension Island, Diego Garcia, and Kwajalein. The other two stations are in Hawaii and at Colorado Springs, CO. The site at Colorado Springs also acts as the Master Control Station (MCS). (A back-up MCS is located at the Onizuka Air Force Base in Sunnyvale, CA.) The MCS collects the pseudorange and carrier phase data obtained by the tracking stations and, with sophisticated software models, predicts the future orbits of the satellites. These computed orbits, called an *ephemeris*, are then uploaded to the corresponding satellite using facilities at the island tracking stations, and are subsequently broadcast by the satellite.

The OCS computes its orbits by fitting Keplerian elements to the tracking data together with some additional parameters to account for the perturbations of the orbits. The sixteen parameters of the broadcast ephemeris are listed in Table 3.2 and illustrated in Figure 3.3. The square root of the semimajor axis is used rather than the semimajor axis itself to speed up the GPS receiver's calculation of the satellite's position. Likewise, the longitude of the ascending node rather than its right ascension is used in the message. The three parameters Δn, i-dot, and Ω-dot account for the linear changes in the orbit with time; the six C-values are the amplitudes of sinusoidally-varying correction terms.

Table 1. Ephemeris Parameters in the Navigation Message

|	
$\sqrt{a}$	Square root of the semimajor axis
e	Eccentricity
i_0	Inclination angle at the reference time
Ω_0	Longitude of the ascending node at the beginning of the GPS week
ω	Argument of perigee
M_0	Mean anomaly at the reference time
Δn	Correction to the mean motion computed using $\sqrt{a}$
i-dot	The rate of change of the inclination with time
Ω-dot	The rate of change of the right ascension of the ascending node with time
C_{uc}, C_{us}	Amplitude of correction terms for the computed argument of latitude
C_{rc}, C_{rs}	Amplitude of correction terms for the computed orbit radius
C_{ic}, C_{is}	Amplitude of correction terms for the computed inclination angle
t_{oe}	Ephemeris reference time
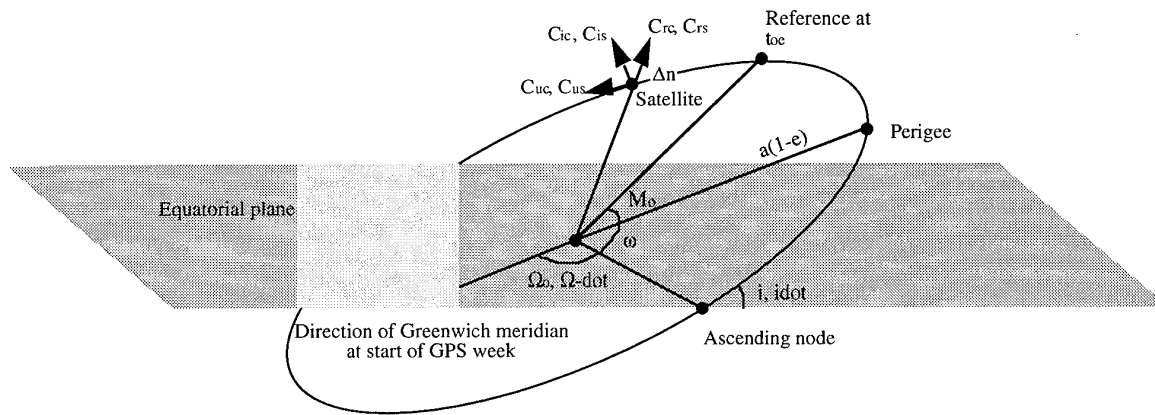
Figure 3.3 The parameters of the GPS broadcast ephemeris.

The GPS receiver takes the ephemeris parameters and computes the coordinates of the satellite in an earth-centred, earth-fixed (ECEF) coordinate system. The details of this computation have been carefully spelled-out by the designers of the Global Positioning System and can be found in a number of reference publications including the *Guide to GPS Positioning* published by Canadian GPS Associates [Wells et al., 1987]. The particular ECEF system used by GPS is the World Geodetic System 1984 (WGS 84) of the Defense Mapping Agency (DMA). For all intents and purposes, the reference frame of WGS 84 and the new North American Datum of 1983 (NAD 83) are identical.

A new set of orbital parameters is computed for each one hour period using over-lapping data spans of four hours. A GPS satellite broadcasts the appropriate set of parameters during a particular one hour interval.

Fresh orbital data is uploaded to the satellites up to three times daily. The initial Block II satellites can store the navigation messages for the following 14 days and therefore are afforded a certain degree of autonomy should something happen to the OCS. This autonomy capability has been extended to 180 days for the Block IIA satellites.

The broadcast ephemeris is computed with sufficient accuracy to guarantee the design goal of horizontally positioning a GPS receiver with an accuracy of 16 meters. This is the accuracy of the Precise Positioning Service, the service afforded to authorized (primarily military) users. When GPS is fully operational, the accuracy of the broadcast ephemerides will be intentionally degraded as one of the mechanisms for implementing the policy of Selective Availability (S/A) [Georgiadou and Doucet, 1990]. Already, extensive tests of S/A have been carried out. PPS users will have access to the undegraded orbits through a decryption process. Unauthorized users will have to be content with the lower accuracy Standard Positioning Service unless they use the technique of differential or relative positioning (see 3.6.5). Differential positioning is far less sensitive to orbit errors than point positioning.

For some applications even the undegraded broadcast ephemerides are not accurate enough. Very high accuracy geodetic surveys and surveys carried out to study deformations of natural and man-made structures require orbits with the highest accuracy possible. Such surveys are carried out using the differential mode of positioning with a network of receivers. Each receiver records the phase of the GPS carrier signals for subsequent analysis. Since the data does not have to be processed in real time, the data analysts have the luxury of using post-computed rather than predicted ephemerides. These ephemerides are computed using data from a tracking network spanning the same time interval as that of the survey.

The DMA operates a global network of five tracking stations for the purpose of computing precise ephemerides. Data from these stations is combined with that from the five OCS stations to produce the orbits. Several civilian agencies also operate tracking networks for orbit determination. The largest of these networks is the Cooperative International GPS Network (CIGNET). This global network currently consists of about 20 stations which are operated by various national agencies. The collected data is transmitted to the CIGNET Data Center in Rockville, MD, where it is archived and made available for orbit computation and research.

In addition to its ephemeris, each satellite transmits approximate orbit information for all of the other satellites in the GPS constellation. Known as the almanac, this information can be used by a GPS receiver to determine the location of each satellite so that it can quickly acquire the signals from satellites that are above the horizon but are not yet being tracked. The almanac consists of values for computing the Keplerian motion of the satellites and only the most significant of the correction terms is provided – the change in the right ascension of the ascending node with time. The almanac can also be used with various software planning tools to predict satellite availability at a given location and time. The almanac is updated by the OCS at least once every six days.

3.5 GPS Receivers
3.5.1 *The Antenna.*   The job of the antenna is to convert the energy in the electromagnetic waves arriving from the satellites into an electric current which can be handled by the electronics in the receiver. The size and shape of the antenna are very important as these characteristics govern in part the ability of the antenna to pick up the very weak GPS signals. The antenna may be required to operate at just the L1 frequency or at both the L1 and L2 frequencies. Also because the GPS signals are circularly polarized, all GPS antennas must be circularly polarized as well. Despite these restrictions, there are several different types of antennas that are presently available for GPS receivers. These include monopole or dipole configurations, quadrifilar helices (also known as volutes), spiral helices, and microstrips.

Perhaps the most common antenna is the microstrip because of its ruggedness and relative ease of construction. It can be circular or rectangular in shape and is similar in appearance to a small piece of copper-clad printed circuit board. Made up of one or more patches of metal, microstrips are often referred to as *patch* antennas. They may have either single or dual

frequency capability and their exceptionally low profile makes them ideal for airborne and some hand-held applications.

Other important characteristics of a GPS antenna are its gain pattern which describes its sensitivity over some range of elevation and azimuth angles; its ability to discriminate against multipath signals, that is, signals arriving at the antenna after being reflected off nearby objects; and for antennas used in very precise positioning applications, the stability of its phase centre, the electrical centre of the antenna to which the position given by a GPS receiver actually refers.

Some antennas, such as the microstrip, require a ground plane to make them work properly. This is usually a flat or shaped piece of metal on which the actual microstrip element sits. In high precision surveying, the ground plane of the antenna is often extended with a metal plate or plates to enhance its performance in the presence of multipath.

The GPS signals when they arrive at the earth are very weak; they have roughly the same strength as those from geostationary TV satellites. The reason a GPS receiver does not need an antenna the size of those in some people's backyards has to do with the structure of the GPS signal and the ability of the GPS receiver to de-spread it. The power to extract a GPS signal out of the general background noise of the ether is concentrated in the receiver rather than the antenna. Nevertheless, a GPS antenna must generally be combined with a low noise preamplifier that boosts the level of the signal before it is fed to the receiver itself. In systems where the antenna is a separate unit, the preamplifier is housed in the base of the antenna and receives power from the same coaxial cable along which the signal travels to the receiver.

3.5.2 *The RF Section.* The job of the RF section of a GPS receiver is to translate the frequency of the signals arriving at the antenna to a lower one, called an intermediate frequency or IF which is more easily managed by the rest of the receiver. This is done by combining the incoming signal with a pure sinusoidal signal generated by a component in the receiver known as a local oscillator. Most GPS receivers use precision quartz crystal oscillators, enhanced versions of the regulators commonly found in wristwatches. The IF signal contains all of the modulation that is present in the transmitted signal; only the carrier has been shifted in frequency. The frequency of the shifted carrier is simply the difference between the original received carrier frequency and that of the local oscillator. It is often called a beat frequency in analogy to the beat note that is heard when two musical tones very close together are played simultaneously. Most receivers employ multiple IF stages, reducing the carrier frequency in steps. The final IF signal passes to the work horse of the receiver, the signal tracker.

3.5.3 *The Signal Trackers.* The omni-directional antenna of a GPS receiver simultaneously intercepts signals from all satellites above the antennas' horizon. The receiver must be able to isolate the signals from each particular satellite in order to measure the code pseudorange and the phase of the carrier. The isolation is achieved through the use of a number of signal channels in the receiver. The signals from different satellites may be easily discriminated by the unique C/A-code or portion of the P-code they transmit and are assigned to a particular channel.

The channels in a GPS receiver may be implemented in one of two basic ways. A receiver may have dedicated channels with which particular satellites are continuously tracked. A minimum of four such channels tracking the L1 signals of four satellites would be required to determine three coordinates of position and the receiver clock offset. Additional channels

permit tracking of more satellites or the L2 signals for ionospheric delay correction or both.

The other channelization concept uses one or more sequencing channels. A sequencing channel "listens" to a particular satellite for a period of time, making measurements on that satellite's signal and then switches to another satellite. A single channel receiver must sequence through four satellites to obtain a three-dimensional position "fix". Before a first fix can be obtained, however, the receiver has to dwell on each satellite's signal for at least 30 seconds to acquire sufficient data from the satellite's broadcast message. The time to first fix and the time between position updates can be reduced by having a pair of sequencing channels.

A variation of the sequencing channel is the multiplexing channel. With a multiplexing channel, a receiver sequences through the satellites at a fast rate so that all of the broadcast messages from the individual satellites are acquired essentially simultaneously. For a multiplexing receiver, the time to first fix is 30 seconds or less, the same as for a receiver with dedicated multiple channels.

Receivers with single channels are cheaper but because of their slowness are restricted to low speed applications. Receivers with dedicated channels have greater sensitivity because they can make measurements on the signals more often but they have inter-channel biases which must be carefully calibrated. This calibration is usually done by the receiver's microprocessor.

The GPS receiver uses its tracking channels to make pseudorange measurements and to extract the broadcast message. This is done through the use of *tracking loops.* A tracking loop is a mechanism which permits a receiver to 'tune into' or track a signal which is changing either in frequency or in time. It is a feedback device which basically compares an incoming (external) signal against a locally-produced (internal) signal, generates an error signal which is the difference between the two, and uses this signal to adjust the internal signal to match the external one in such a way that the error is reduced to zero or minimized. A GPS receiver contains two kinds of tracking loops: the delay lock, or code tracking, loop and the phase lock, or carrier tracking, loop.

The delay lock loop is used to align a pseudorandom noise (PRN) code sequence (from either the C/A or P code) that is present in the signal coming from a satellite with an identical one which is generated within the receiver using the same algorithm that is employed in the satellite. Alignment is achieved by appropriately shifting the receiver-generated code chips in time so that a particular chip in the sequence is generated at the same instant its twin arrives from the satellite.

A correlation comparator in the delay lock loop continuously cross-correlates the two code streams. This device essentially performs a multiply and add process that produces a relatively large output only when the code streams are aligned. If the output is low, an error signal is generated and the code generator adjusted. In this way the replicated code sequence is locked to the sequence in the incoming signal. The signals from other GPS satellites will have essentially no effect on the tracking process because the PRN codes of all the satellites were chosen to be orthogonal to each other. This orthogonality property means that a very low output is always produced by the correlator whenever the code sequences used by two different satellites are compared.

Because the P-code sequence is so long, a P-code tracking loop needs some help in setting its code generator close to the right spot for obtaining lock with the satellite signal. Its gets this

help from information included in the broadcast message which is available to the receiver by first tracking the C/A-code.

The time shift required to align the code sequences is, in principle, the time required for a signal to propagate from the satellite to the receiver. Multiplying this time interval by the speed of light gives us the distance or range to the satellite. But because the clocks in a receiver and in a satellite are, in general, not synchronized and run at slightly different rates, the range measurements are biased. These biased ranges are called *pseudoranges*. Since the chips in the satellite code sequences are generated at precisely known instants of time, the alignment of the receiver and satellite code sequences also gives us a reading of the satellite clock at the time of signal generation.

Once the code tracking loop is locked, the PRN code can be removed from the satellite signal by mixing it with the locally generated one and filtering the resultant signal. This procedure de-spreads the signal, shrinking its bandwidth down to about 100 Hz. It is through this process that the GPS receiver achieves the necessary signal to noise ratio to offset the gain limitation of a physically small antenna.

The de-spread IF signal then passes to the phase lock loop which demodulates or extracts the satellite message by aligning the phase of the receiver's local oscillator signal with the phase of the IF or beat frequency signal. If the phase of the oscillator signal is not correct, this is detected by the demodulator in the phase lock loop and a correction signal is then applied to the oscillator. Once the oscillator is locked to the satellite signal, it will continue to follow the variations in the phase of the carrier as the range to the satellite changes.

Most implementations of carrier tracking use the Costas Loop, a variation of the phase lock loop designed for binary biphase modulated signals such as those transmitted by the GPS satellites.

The carrier beat phase observable is obtained in principle simply by counting the elapsed cycles and by measuring the fractional phase of the locked local oscillator signal. The phase measurement when converted to units of distance is then an ambiguous measurement of the range to the satellite. It is ambiguous because a GPS receiver cannot distinguish one particular cycle of the carrier from another and hence assumes an arbitrary number of full cycles of initial phase when it first locks onto a signal. This initial ambiguity must be solved for mathematically along with the coordinates of the receiver if phase observations are used for positioning. Because this ambiguity is constant as long as the receiver maintains lock on the received signal, the time rate of change of the carrier phase is freed from this ambiguity. This quantity is related to the Doppler shift of the satellite signal and is used, for example, to determine the velocity of a moving GPS receiver such as that in an aircraft.

After the carrier tracking loop locks onto a satellite signal, the bits in the broadcast message are subsequently decoded using standard techniques of bit synchronization and a data detection filter.

There is another way to measure the carrier beat phase other than the code tracking / Costas Loop combination, but it comes with a penalty. This is the so-called signal squaring technique. The GPS signal is simply a constant carrier who's phase is shifted by exactly 180° more than a million times each second as a result of modulation by the PRN codes and the broadcast message. These phase reversals can be considered as a change in the amplitude of the signal from +1 to –1 or from –1 to +1 and the instantaneous amplitude is therefore either plus or minus one. Electronically squaring the signal results in a signal with a constant amplitude of unity, although with a frequency equal to twice the original. However, the phase of this signal is easily related to the phase of the original carrier. Of course, in the squaring process both the codes and the broadcast message are lost so code-derived pseudorange measurements are not possible and the information describing the orbits of the satellites as well as their health and the other details in the message must come from another source. There is also some inherent signal to noise loss in the squaring process compared to code tracking which may result in slightly noisier phase measurements.

One of the first commercially available GPS receivers, the Macrometer®, used the squaring technique and a number of currently available dual frequency receivers use this approach for measurements on the L2 frequency. A variation of this technique has been used in receivers which measure the *phase* of the code modulations without having to know the actual code sequences.

If anti-spoofing is turned on in the satellites resulting in an encryption of the P-code, then multiplying the signal by itself will be the only way to make measurements on the L2 frequency. Pseudorange measurements and the broadcast message would still be available from the code and carrier tracking on the L1 frequency.

3.5.4 *The Microprocessor.* Although the bulk of a GPS receiver could be built using analogue techniques, the trend in receiver development has been to make as much of the receiver digital as possible resulting in smaller, cheaper units. In fact, it is possible for the IF signal to be digitized and to perform the code and carrier tracking with software inside the microprocessor. So in some respects, a GPS receiver may have more in common with a compact disc player than it does with an AM radio. Because it has to perform many different operations such as initially acquiring the satellites signals as quickly as possible once the receiver is turned on, tracking the codes and carriers of the signals, decoding the broadcast message, determining the user's coordinates, and keeping tabs on the other satellites in the constellation, a GPS receiver's operation is controlled by a microprocessor. The microprocessor's software, that is the instructions for running the receiver, is imbedded in memory chips within the receiver.

The microprocessor works with digital samples of pseudorange and carrier phase. These are acquired as a result of analogue to digital conversion at some point in the signal flow through the receiver. It is these data samples that the receiver uses to establish its position and which may be recorded for future processing. The microprocessor may run routines which do some filtering of this raw data to reduce the effect of noise or to get more more reliable positions and velocities when the receiver is in motion.

The microprocessor may also be required to carry out the computations for waypoint navigation or convert coordinates from the standard WGS 84 geodetic datum to a regional one. It also manages the input of commands from the user, the display of information, and the flow of data through its communication port if it has one.

3.5.5 *The Command Entry and Display Unit.* The majority of self-contained GPS receivers have a keypad and display of some sort to interface with the user. The keypad can be used to enter commands for selecting different options for acquiring data, for monitoring what the receiver is doing, or for displaying the computed coordinates, time or other details. Auxiliary information such as that required for waypoint navigation or weather data and antenna height for geodetic

166

surveying may also be entered. Most receivers have well integrated command and display capabilities with menus, prompting instructions, and even "on line" help. It should be mentioned that some receivers have a basic default mode of operation which requires no user input and can be activated simply by turning the receiver on.

Some GPS receivers are designed as sensors to be integrated into navigation systems and therefore don't have their own keypads and displays; input and output is only via data ports.

3.5.6 *Data Storage and Output.* In addition to a visual display, many GPS receivers including even some hand-held units provide a means of saving the carrier phase and/or pseudorange measurements and the broadcast messages. This feature is a necessity for receivers used for surveying and for differential navigation.

In surveying applications the pseudorange and phase observations must be stored for combination with like observations from other simultaneously observing receivers and subsequent analysis. Usually the data is stored internally in the receiver using semiconductor memory. Some receivers store data on magnetic tape or directly on a floppy disk using an external microcomputer.

Some receivers, including those which store their data internally for subsequent analysis and those used for real-time differential positioning, have an RS-232 or some other kind of communications port for transferring data to and from a computer, modem or data radio. Some receivers can be remotely controlled through such a port.

3.5.7 *The Power Supply.* Most GPS receivers have internal DC power supplies, usually in the form of rechargeable nickel-cadmium (NiCd) batteries. The latest receivers have been designed to draw as little current as possible to extend the operating time between battery charges. Most receivers also make a provision for external power in the form of a battery pack or AC to DC converter.

3.6 Observation Equations
3.6.1 *Determining Positions from Pseudoranges*
The basic measurement made by a GPS receiver is the time required for a signal to propagate from a GPS satellite to the receiver. Since the signal travels at the speed of light, c, this time interval can be converted to a distance simply by multiplying it by c. Let's assume that the clock in the receiver is synchronized with the clock in the satellite, and that the ionosphere and troposphere, which slightly delay the arrival of the signal, do not exist. Furthermore, let's assume there is no measurement noise; that is, no random perturbation to the measurement, something which invariably affects all measurements to a greater or lesser degree. With a single such measurement of the distance or range to the satellite we can determine something about the position of the receiver: it must lie somewhere on a sphere, centred on the satellite, with a radius equal to the measured range, call it p_1.

If we simultaneously make a range measurement to a second satellite then our receiver must also lie on a sphere, of radius p_2, centred on this satellite. The two spheres will intersect, with the loci of intersection points forming a circle. Our receiver must lie somewhere on this circle which is therefore called a line of position. A third simultaneous range measurement, p_3, gives us a third sphere which intersects the other two at two points. One of these points can be immediately dismissed as being the location of our receiver, since it far out in space. So the simultaneous measurement of the ranges to three satellites is sufficient to determine a position fix in three dimensions -- at least in principle.
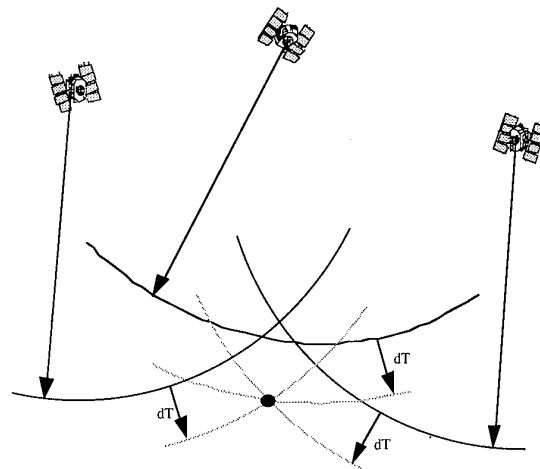
Figure 3.4 Determination of receiver clock offset (dT) and true user position (intersection of shaded lines) from the intersection of spheres centered on the satellites; pseudoranges are shown by arcs of solid lines.

When we started our analysis, we assumed that the clock in the GPS receiver was synchronized with the clocks in the satellites. This assumption, however, is fallacious. When a GPS receiver is switched on, its clock will in general be mis-synchronized with respect to the satellite clocks, by an unknown amount. Furthermore, the atomic clocks in the satellites are synchronized with each other and to a master time scale, called GPS time, only to within about a millisecond. The range measurements the receiver makes are biased by the receiver and satellite clock errors and are therefore referred to as *pseudoranges.*

A timing error of a millisecond would result in an error in position of about 300 km, clearly an intolerable amount. It would be possible to better synchronize the satellite clocks by frequently sending them adjustment commands from the ground, but it has been found that clocks actually keep better time if they are left alone and the readings of the clock corrected. The United States Naval Observatory monitors the GPS satellite clocks and determines the offsets and drifts with respect to GPS time. These parameters are subsequently uploaded to the satellites and transmitted as part of the navigation message broadcast by the satellites. A GPS receiver uses these satellite clock offset values to correct the measured pseudoranges.

However, we still have the receiver clock error to deal with. Because of this error, the three spheres with radii equal to the measured pseudoranges corrected for the satellite clock offsets will not intersect at a common point. However, if the receiver clock error can be determined, then the pseudoranges can be corrected and the position of the receiver determined. The situation, compressed into two dimensions, is illustrated in Figure 3.4.

So we actually have four unknown quantities or parameters that we must determine: the three coordinates of our position (say latitude, longitude, and height) and the receiver clock offset. Now it is a mathematical impossibility to uniquely determine the the values of four parameters, given only three measurements. The way out of this conundrum is to simultaneously measure an additional pseudorange to a fourth satellite.

But just how does the GPS receiver actually extract the position coordinates and the clock offset from the measurements? In the software embedded within the GPS receiver is an algebraic model that describes the geometrical arrangement we've just looked at. For each pseudorange measurement, an equation can be written which relates the measurement to the unknown quantities:

The pseudorange measurement made by the receiver, in units of distance, is on the left hand side of each of the equations. The expression under the square root sign is the true range to the satellite. It is actually a representation of the sphere centred on coordinates x,y,z, the position of the satellite. The satellite coordinates are obtained from the navigation message. The coordinates X,Y,Z represent the position of the receiver. The term c dT is the contribution to the pseudorange from the receiver clock offset, dT. The set of four equations must be solved simultaneously to obtain the values for X,Y,Z together with the clock offset, dT. Although the equations are written in terms of geocentric Cartesian coordinates, the resulting X,Y,Z values can easily be converted to latitude, longitude, and height in any geodetic datum or into map grid coordinates.

Linearization of the pseudorange equations. Because of the squares and square roots in the equations, the pseudorange measurements are dependent on the receiver coordinates in a nonlinear way. Consequently, the equations cannot be solved in the usual fashion we all learned in high school. Instead a procedure known as Newton-Raphson iteration is used. In this procedure, each of the equations is expanded about a set of trial values or guesses for X,Y,Z, and dT into an infinitely long polynomial. Then each series is truncated after the first degree term resulting in an equation which is linear in increments or corrections to the trial values. The four linearized equations can then be solved to determine the values of these increments, and the trial values adjusted accordingly. Since the linearized equations are an approximation of the nonlinear ones, this process, in general, must be iterated, with subsequent iterations yielding smaller and smaller increments. The final solution is the one that satisfies the original nonlinear equations to within an acceptable tolerance. Several iterations may be required to converge to the final solution. However, if the initial position estimate is close to the actual position, the GPS pseudorange equations may be solved in just one iteration.

If one or more of the receiver coordinates is already accurately known, then the remaining coordinates and the receiver clock offset can be determined using fewer than four pseudoranges. For example, say that the height of the GPS receiver is known. Then pseudoranges to three satellites will suffice to determine the two horizontal coordinates and the clock offset. In using GPS for synchronizing a clock at a site with known coordinates, only one pseudorange measurement to a single satellite is actually required.

Overdetermined solutions. What if more than four satellites are above the GPS user's horizon? If the user's receiver can only track four satellites at a time, then the receiver will have to choose which four satellites to track. We'll have something to say about a possible selection method a little later on. But if the receiver can track five or more satellites simultaneously, then we have the situation where we have more measurements than unknowns; that is, we have five or more equations, like those above, but still with only four unknown parameters. It is impossible to solve such a set of equations in the same way as we did for the case of four observations. Why? So far we have neglected the fact that there are other errors in our measurements in addition to the satellite and receiver clock offsets. The presence of these errors means that any subset of

four measurements taken from the full set will produce slightly different solutions. We say that the solution is *inconsistent* .

What do we do? We could discard the extra observations but, although expedient, that seems wasteful of data. The best approach is to use a method that was devised in the early 1800s by the great German mathematician and father of modern geodesy, Karl Friedrich Gauss: *the method of least squares.* In this method, we obtain a unique solution for the unknown parameters which best fits all of the measurements. This solution is the one which when substituted into the pseudorange equations gives the smallest discrepancies with respect to the measurements in a sum-squared sense. That is, the sum of the squares of the discrepancies is a minimum. Without going into the mathematical reasons for adopting this criterion, we can see qualitatively that it assumes that positive and negative discrepancies are equally likely to occur and that smaller discrepancies are more likely to occur than larger ones.

3.6.2 The Pseudorange. Before discussing the pseudorange, let's quickly review the structure of the signals transmitted by the GPS satellites Each GPS satellite transmits two signals for positioning purposes: the L1 signal, centred on a carrier frequency of 1575.42 MHz, and the L2 signal, centred on 1227.60 MHz. Modulated onto the L1 carrier are two pseudorandom noise (PRN) ranging codes: the 1 millisecond-long C/A-code with a chipping rate of about 1 MHz and a week-long segment of the P-code with a chipping rate of about 10 MHz. Also superimposed on the carrier is the navigation message, which among other items, includes the ephemeris data describing the position of the satellite and predicted satellite clock correction terms. The L2 carrier is modulated by the P-code and the navigation message -- the C/A-code is not present.

The PRN codes used by each GPS satellite are unique and have the property that the correlation between any pair of codes is very low. This characteristic allows all of the satellites to share the same carrier frequencies.

The PRN codes transmitted by a satellite are used to determine the pseudorange -- a measure of the range, or distance, between the satellite and the antenna feeding a GPS receiver. The receiver makes this measurement by replicating the code being generated by the satellite and determining the time offset between the arrival of a particular transition in the code and that same transition in the code replica. The time offset is simply the time the signal takes to propagate from the satellite to the receiver. The pseudorange is this time offset multiplied by the speed of light. The reason the observable is called a pseudorange is that it is biased by the lack of time synchronization between the clock in the GPS satellite governing the generation of the satellite signal and the clock in the GPS receiver governing the generation of the code replica. This synchronization error is determined by the receiver along with its position coordinates from the pseudorange measurements. The pseudorange is also biased by several other effects including ionospheric and tropospheric delay, multipath, and receiver noise. We can write an equation for the pseudorange observable that relates the measurement and the various biases:

$$p = \rho + c \cdot (dt - dT) + d_{ion} + d_{trop} + \varepsilon_p$$

where p is the measured pseudorange, ρ is the geometric range to the satellite, c is the speed of light, dt and dT are the offsets of the satellite and receiver clocks from GPS time, d_{ion} and d_{trop} are the delays imparted by the ionosphere and troposphere respectively, and ε_p represents the effect of multipath and receiver noise. The receiver coordinates are
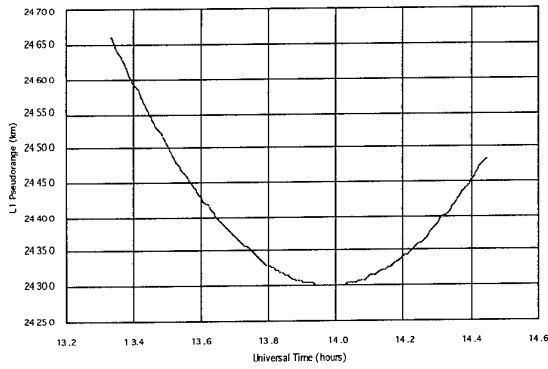
168

Figure 3.5 Typical variation in L1 pseudorange measurements over approximately a one-hour period.

hidden in the geometric range along with the coordinates of the satellite. The objective in GPS positioning is to mathematically describe all of the terms on the right-hand side of the equation -- including the initially unknown receiver coordinates in the geometric range term -- so that the sum of the terms equals the measurement value on the left-hand side. Any error in the description of the terms will result in errors in the derived receiver coordinates. For example, both the geometric range term and the satellite clock term may include the effects of Selective Availability (SA) which, if uncompensated, introduce errors into the computed position of the receiver.

Figure 3.5 illustrates the variation in the pseudorange of a particular satellite as measured by a stationary GPS receiver. The large variation is of course dominated by the change in the geometric range due to the satellite's orbital motion and the rotation of the earth.

Pseudoranges can be measured using either the C/A-code or the P-code. Figure 3.6 shows typical C/A-code pseudorange noise. This "noise record" was obtained by subtracting the geometric range, clock, and atmospheric contributions from the pseudorange measurements illustrated in Figure 3.5. What remains is chiefly pseudorange multipath and receiver measurement noise. Because of its higher chipping rate, the P-code generally provides higher precision observations. However recent improvements in receiver technologies have resulted in higher precision C/A-code measurements than were previously achievable.

3.6.3 *Carrier Phase.* Even with the advances in code measurement technology, a far more precise observable than the pseudorange is the phase of the received carrier with respect to the phase of a carrier generated by an oscillator in the GPS receiver. The carrier generated by the receiver has a
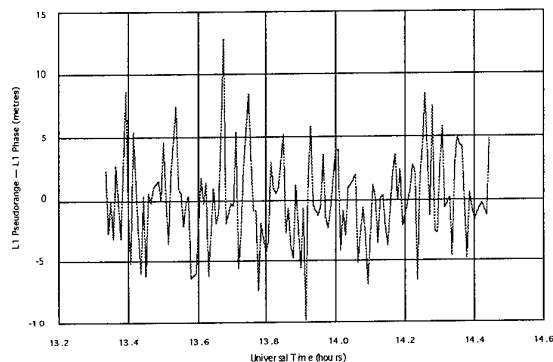

Figure 3.6 The difference between the L1 pseudorange measurements shown in Fig. 3.5 and the corresponding phase measurements.

nominally constant frequency whereas the received carrier is changing in frequency due to the Doppler shift induced by the relative motion of the satellite and the receiver. The difference between the received carrier and the receiver-generated one is sometimes referred to as the carrier beat phase since the phase difference relates to the difference or *beat* frequency of the two carriers. Such a beat frequency is well known to musicians who tune their instruments by listening for the beat note generated when they play two notes slightly different in frequency. The phase of the received carrier is related to the phase of the carrier at the satellite through the time interval required for the signal to propagate from the satellite to the receiver.

So, ideally, the carrier phase observable would be the total number of full carrier cycles and fractional cycles between the antennas of a satellite and a receiver at any instant. The problem is that a GPS receiver has no way of distinguishing one cycle of a carrier from another. The best it can do, therefore, is to measure the fractional phase and then keep track of changes to the phase; the initial phase is undetermined, or ambiguous, by an integer number of cycles. In order to use the carrier phase as an observable for positioning, this unknown number of cycles or *ambiguity*, N, must be estimated along with the other unknowns -- the coordinates of the receiver.

If we convert the measured carrier phase in cycles to equivalent distance units by multiplying by the wavelength, λ, of the carrier, we can express the carrier phase observation equation as

$$\Phi = \rho + c \cdot (dt - dT) + \lambda \cdot N - d_{ion} + d_{trop} + \varepsilon_\Phi$$

which is very similar to the observation equation for the pseudorange -- the major difference being the presence of the ambiguity term. In fact, the carrier phase can be thought of as a biased range measurement just like the pseudorange. Note also that the sign of the ionospheric term in the carrier phase equation is negative whereas in the pseudorange equation it is positive. This comes about because the ionosphere, as a dispersive medium, slows down the speed of propagation of signal modulations (the PRN codes and the navigation message) to below the vacuum speed of light whereas the speed of propagation of the carrier is actually increased beyond the speed of light. Don't worry, Einstein's pronouncement on the sanctity of the speed of light has not been contradicted. The speed of light limit only applies to the transmission of information and a pure continuous carrier contains no information.

Although all GPS receivers must lock onto and track the carrier of the signal in order to measure pseudoranges, they may not measure or record carrier phase observations for use in navigation or positioning. Some however, may internally use carrier phase measurements to smooth -- reduce the high frequency noise -- on the pseudorange measurements.

Incidentally, in comparison with the carrier phase, pseudoranges when measured in units of the wavelengths of the codes (300 meters for the C/A-code and 30 meters for the P-code) are sometimes referred to as code phase measurements.

3.6.4 *Point Positions.* Most civilian receivers intended primarily for navigation exclusively use C/A-code pseudorange measurements to establish the position of a point or the trajectory of a moving platform. The accuracy of point positions afforded civilians is primarily limited by SA rather than receiver measurement precision. In fact, before SA was
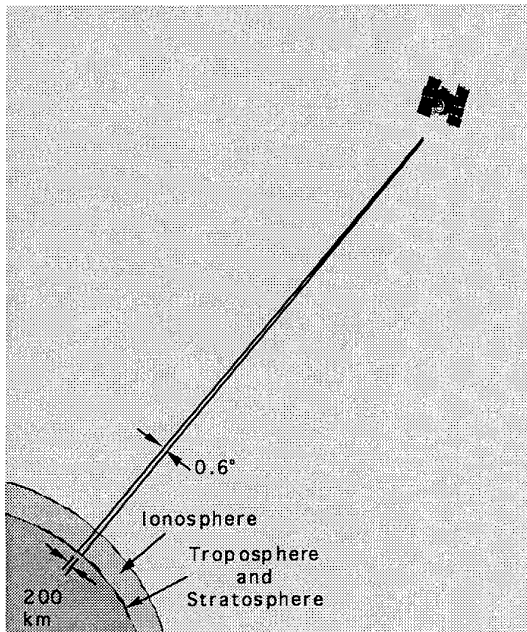
Figure 3.7 Common-mode error. Distances are approximately to scale.

implemented it was shown that low cost civilian C/A-code receivers could obtain positions almost as accurate as those provided by military receivers using the P-code.

3.6.5 *Relative Positions.* The accuracy of point positions is limited by unmodellable or residual errors in the pseudorange observation equation as well as by measurement noise. In order to get higher accuracy, we need to take a different approach.

The pseudoranges and carrier phases measured simultaneously by a pair of GPS receivers tracking a particular satellite will share to a large degree the same satellite ephemeris and satellite clock errors (including the effects of SA), and atmospheric errors (see Figure 3.7). The closer together the receivers are the more similar the errors. If we set up a receiver at a site with a known location, we can monitor the accuracy of the receiver-computed positions. Any difference between the known and computed positions can be attributed to pseudorange modelling errors. These errors can be computed and transmitted to another receiver tracking the same set of satellites and used to correct the measured pseudoranges before computing its coordinates. In this fashion, real-time position accuracies better than 5 meters are possible. The corrections could also be stored in the receiver or an attached computer and applied in post-processing of the collected data for applications not requiring positions in real time. A similar approach with carrier phase measurements is possible (difficulties associated with the integer ambiguities not withstanding) giving real-time position accuracies at the centimeter level.

3.6.5.1 *The Single Differences.* Rather than computing and transmitting or storing pseudorange or carrier phase corrections, we can form what are known as between-receiver single difference observables -- new observables with significantly reduced errors. Although both pseudoranges and carrier phases can be used to form single differences, we will concentrate on the use of the carrier phase. The observation equation for the between-receiver single difference is

$$\Delta\Phi = \Delta\rho - c \cdot \Delta dT + \lambda \cdot \Delta N - \Delta d_{ion} + \Delta d_{trop} + \varepsilon_{\Delta\Phi}$$
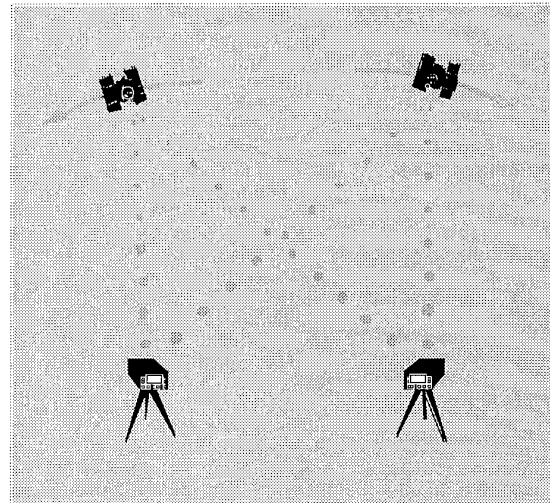
Figure 3.8 The double difference.

where Δ denotes the operation of forming differences between receivers. Note that the satellite clock term has disappeared from the equation since the effect of satellite clock errors on the phase measurements of the two receivers is essentially identical. We could use the between-receiver single differences for determining the relative coordinates of the receivers but in addition to modelling the residual satellite orbit and atmospheric errors and estimating the relative integer ambiguities, we would have to carefully model the relative behaviour of the receiver clocks -- not an easy task. But, if differencing between receivers is useful in removing satellite clock errors then it stands to reason that differencing between satellites should remove receiver clock errors. This is just what happens. If we take the carrier phases (or pseudoranges) measured by a single receiver tracking two satellites and difference them, we form the between-satellite single differences:

$$\nabla\Phi = \nabla\rho + c \cdot \nabla dt + \lambda \cdot \nabla N - \nabla d_{ion} + \nabla d_{trop} + \varepsilon_{\nabla\Phi}$$

The symbol ∇ denotes the operation of differencing between satellites. Notice that ∇, as an upside down triangle has two vertices on the top (in the sky) whereas Δ had two vertices on the bottom (on the ground).

3.6.5.2 *The Double Difference.* The advantages of the between-receiver and between-satellite single differences can be combined by forming the double difference. Consider two receivers tracking two satellites at the same time (see Figure 3.8). We can form two between-receiver single differences each involving a different satellite, and two between-satellite single differences each involving a different receiver. We can difference either the between-receiver or the between-satellite difference pairs to generate a double difference:

$$\nabla\Delta\Phi = \nabla\Delta\rho + \lambda \cdot \nabla\Delta N - \nabla\Delta d_{ion} + \nabla\Delta d_{trop} + \varepsilon_{\nabla\Delta\Phi}$$

The great advantage of the double difference is that essentially all of the satellite and receiver clock errors are removed from the observations. Notice that the ambiguity term in this equation is still an integer number of wavelengths as $\nabla\Delta N$ is simply the difference of integers and is therefore an integer itself. The double difference observable has become the standard observable in precise differential positioning.

The ambiguities for all the satellite pairs forming an observation set must be estimated in the least squares

169

adjustment along with the coordinates of one of the receivers (the coordinates of the other receiver are held fixed at *a priori* known values). Usually these ambiguities are initially estimated as real or floating point numbers rather than integer numbers. If the data quality is good enough and the model used in the adjustment accurately describes the observations, the real number estimates from the float solution will turn out to be very close to integers. The estimates can then be rounded off to the nearest integer and then held fixed at these integer values in a second adjustment of the data. The so-called fixed solution will in general provide more accurate results than those afforded by the float solution.

3.6.5.3 *The Triple Difference.*

A potential problem with the double difference observable is associated with the integer ambiguity term. As long as the receivers do not lose carrier lock on the signals, the integer ambiguities remain constant for the whole data set. If, however, one or both receivers loses lock due to a satellite passing behind an obstacle, low signal to noise ratio, rapid motion of the receiver, or a severe ionospheric disturbance, then one or more carrier phase cycles will be lost or slipped. The receiver essentially loses track of the continuous cycle count. This introduces a discontinuity or cycle slip into the data. If any data gap accompanying the cycle slip is short and any noise corrupting the data is minimal then it is generally possible to determine the correct number of slipped cycles and to correct the phase measurements to produce a continuous phase record. If this cannot be done, then a new set of ambiguities must be adopted for the measurements following the cycle slip.

An observable which is insensitive to both the initial integer ambiguities and cycle slips is the triple difference. Triple differences are formed by sequentially differencing double differences in time. If we have double differences at epochs one, two, and three for example, then we can create two triple differences: double difference two minus double difference one and double difference three minus double difference two. The triple difference observation equation is written as

$$\delta\nabla\Delta\Phi = \delta\nabla\Delta\rho - \delta\nabla\Delta d_{ion} + \delta\nabla\Delta d_{trop} + \varepsilon_{\delta\nabla\Delta\Phi}$$

where δ is the time difference operator. The differencing in time results in an observable which has less information content and as a result receiver coordinates estimated from triple differences tend not to be as accurate as those obtained from double differences especially when the latter are used in fixed ambiguity solutions. Nevertheless, triple difference results may be accurate enough for certain applications. Also, the triple difference observable is very useful for obtaining initial estimates of receiver coordinates which can then be used as *a priori* coordinates in a double difference solution. Triple differences are also very useful in spotting and correcting cycle slips. In a triple difference data series, a cycle slip usually appears as an easily-identified spike. If the data is particularly noisy -- for example when it is corrupted by severe ionospheric irregularities, identifying and repairing cycle slips can be quite difficult.

3.6.6 *Other Linear Combinations.*

The single, double, and triple difference observables are known in mathematical parlance as *linear combinations* of the measured carrier phases (or pseudoranges). Several other linear combinations of the basic GPS observables have also found utility in GPS navigation, positioning, and time transfer. Foremost among these perhaps is the ionosphere-free linear combination of the raw or undifferenced L1 and L2 carrier phases measured by a single receiver. The ionospheric delay term in the carrier phase observation equation is, with negligible error, inversely proportional to the square of the carrier frequency. Therefore,

by combining the equation written for the L2 observation with the equation for the L1 observation, we can create an equation and hence an observable which is essentially free of the ionospheric effect. This observable is sometimes referred to as Lc or L3 although the latter term should be deprecated because of possible confusion with the L3 signal associated with the nuclear burst detection package on the GPS satellites. We can remove the ionospheric delay from pseudorange data with a similar operation on the L1 and L2 pseudoranges. Ionosphere-free observations may be combined into single, double, or triple differences and processed in almost the same manner as single frequency data. We say "almost" because in the ionosphere-free observations, the ambiguities are no longer integers and ambiguity resolution is more complex than in the single frequency case. Another distinction is that for baselines shorter than about 20 kilometers or so, ionosphere-free observations are noisier than their single frequency counterparts. This characteristic arises from the fact that the dominant error source on short baselines is multipath and receiver noise which in large part is uncorrelated between the L1 and L2 signals. Therefore for short baselines it may be preferable to use single frequency observations.

Instead of combining the L1 and L2 phase measurements to remove the ionospheric delay, we can simply difference the L1 and L2 phases measured in distance units and determine changes in the ionospheric delay. Unfortunately, we cannot determine the absolute ionospheric delay from phase measurements because of the unknown integer ambiguities. Nevertheless, since orbit, clock, and tropospheric effects are removed, this linear combination of the L1 and L2 phases is particularly useful for locating and potentially removing cycle slips.

Other linear combinations of the L1 and L2 phase measurements have been devised to help with the ambiguity resolution problem. Differencing the L1 and L2 phases measured in cycles results in an observable with an effective wavelength of about 86 cm whereas summing the L1 and L2 phases in cycle units gives an observable with a wavelength of about 10.7 cm. Algorithms have been developed to use the so-called wide-lane and narrow-lane ambiguities to help resolve the L1 and L2 ambiguities. Still other linear combinations of both phase and pseudorange observations are being investigated for data cleaning, particularly under ionospherically noisy conditions.

We can also combine together carrier phase and pseudorange measurements. In fact you have seen such a combination in Figure 3.6. Differencing the carrier phase and C/A-code pseudorange measurements on L1 removes all of the common effects: geometric range, clock terms, and tropospheric delay. What remains is the effect of the ionosphere (doubled because of the sign difference), multipath, receiver noise, and the carrier phase ambiguity. An estimate of the ambiguity and the constant part of the other effects can be removed by subtracting the arithmetic mean of the carrier phase — pseudorange difference. We are left with an observable which is usually dominated by the pseudorange multipath and the pseudorange measurement noise. Also remaining is the variation in the ionospheric delay. Over the one-hour period of the measurements shown in Figure 3.6, the ionospheric delay changed by only about 1 meter in a very smooth fashion. Of course, if we had used data from a dual-frequency P-code receiver, we could have removed the effect of the ionosphere ab initio by using the Lc carrier phase linear combination and the corresponding P-code pseudorange combination.

With the advent of receivers providing low-noise L2 pseudorange observations, we can effectively use the pseudoranges to help determine the integer ambiguities of the

carrier phase measurements. This synergistic combination of pseudoranges are carrier phases has spawned the rapid static surveying technique with which we can determine the ambiguities with observation times of mere minutes.

3.7 Measures of GPS Position Accuracy.

As we have mentioned, the pseudorange measurements are contaminated by the satellite and receiver clock offsets. Even after solving for the receiver clock offset and correcting the pseudoranges for the satellite clock offset using the parameters in the navigation message, errors still remain in the measurements. These errors will, of course, affect the accuracy of the position determination. Since these errors will, in general, change with time, repeated determinations of the position of a fixed location will give slightly different results.

The pseudorange errors come from several sources. The parameters in the navigation message describing the behaviour of the satellite clock account for almost all of the clock offset with respect to GPS time. However because the model used to describe the clock behaviour is quite simple and the parameters of the model are predicted ahead of time, there are some small residual clock errors remaining in the pseudoranges. The positions of the satellites as computed from the predicted ephemerides in the navigation messages are also slightly in error. Other errors in the pseudoranges include unmodelled effects of the ionosphere and troposphere, multipath, receiver measurement errors, and (for the civilian user) the additional clock and orbit errors due to Selective Availability (S/A) when it is in effect.

3.7.1 *User Equivalent Range Error*.

Each of these errors, regardless of their origin, can be expressed as an error in the range between the user and the satellite. When an error is expressed in this way, it is known as a *user equivalent range error* (UERE) or just *user range error* (URE). Over a sufficiently long period of time, these errors can be considered to be random in nature, with negative and positive errors being about equally probable giving a mean value of zero. Also, smaller errors are more probable than larger errors. A graph of the frequency of occurrence of an error of a certain size,is referred to as a *probability density function*. The probability that an error will occur with a value between e_1 and e_2 say, is just the area under the curve between these two values on the horizontal axis.

The shape of the probability density curve depends on the particular parameter being measured. However, it is often the case in science and engineering that the probability density curve is of a particular shape known as a Gaussian or normal distribution. To quantify such a distribution or dispersion of possible errors with a single number, we use the *standard deviation* and usually represent it by the Greek letter sigma (σ). It is possible to determine σ experimentally by making a large number of observations. It is just the square root of the sum of the squares of the errors in the observations divided by one less than the number of observations made. It is from this method of computation that gives σ its alias of root-mean-square (r.m.s.) error. For the Gaussian distribution, there is a 68% chance that the magnitude of the error we actually get will be smaller than the standard deviation. There is a 95% chance that it will be smaller than twice the standard deviation and a 99.7% chance that it will be smaller than thrice the standard deviation.

3.7.2 *Dilution of Precision*.

Since these errors originating from the different sources are independent of each other, a combined error can be obtained by taking the square root of the sum of the squares of the individual errors. This value is the total user equivalent range error.

The total UERE is clearly not the error in the position determined by a GPS receiver. It is only a measure of the error in the distance to one of the satellites. In order to determine the three dimensional position error, we must also take into account where the satellites are in the sky with respect to the receiver. Because we can only receive signals from GPS satellites that are above our horizon, the effect of the satellite geometry is always to dilute the precision of the position determination. It is possible to quantify this dilution through the position dilution of precision (PDOP) parameter. PDOP is the number which when multiplied by the r.m.s. UERE gives the r.m.s. position error, that is, the square root of the sum of the squares of the standard deviations in latitude, longitude, and height.

PDOP is a mathematical function involving the relative coordinates of the receiver and the satellites and can easily be computed for a particular satellite arrangement. PDOP can also be visualized geometrically using four satellites by looking at the tetrahedron formed by the end points of vectors of unit length pointing from the receiver to each of the satellites. PDOP is inversely proportional to the volume of this tetrahedron. The more spread out the satellites are in the sky, the larger the volume of the tetrahedron, the lower the PDOP, and hence the lower the r.m.s. position error. If more than four satellites are in view, the GPS receiver can select the four which give the lowest PDOP. The lowest value of PDOP that could be obtained is with one satellite at the user's zenith and three satellites evenly spread out in azimuth on the user's horizon. The maximum value PDOP could theoretically have is infinity. This would occur if the four satellites were to lie in the same plane. The final GPS constellation has been designed to provide users anywhere in the world with a PDOP of less than 6 (except for occasional very brief periods of time), assuming four satellites are used with a minimum satellite elevation angle of 5°. Users can expect the PDOP to be less than 3 most of the time.

Several other related DOP factors have been defined. HDOP is the dilution of precision in the two horizontal coordinates; VDOP is the dilution of precision in the vertical coordinate; and TDOP is the dilution of precision in the range equivalent of the receiver clock offset. A factor which combines the effect of geometry on both position and clock offset is the geometric dilution of precision, GDOP.

3.7.3 *Other Accuracy Measures*.

In general, the three coordinates of a three-dimensional position fix will have different error probability distributions and hence different standard deviations. Also the errors between any two coordinates may be mutually correlated; that is, an error in one coordinate will have an effect on the other. If we trace out a contour of equal probability density in all three coordinates, we get an ellipsoid centred on our position fix. The shape of the ellipsoid is determined by the standard deviations of the coordinates and their correlations. Note that in general, due to the correlations, the ellipsoid axes are not oriented in the same directions as the coordinate axes. There is a certain probability that the true position lies within the ellipsoid. If this probability is 20%, then this ellipsoid is referred to as the *standard* error ellipsoid.

If the standard deviations in the three orthogonal directions are identical, the ellipsoid degenerates into a sphere. The radius of such a sphere inside of which there is a 50% probability of the true position fix being located is called *spherical error probable* (SEP). The term SEP is also used to indicate the radius of a sphere inside of which there is a 50% probability of being located even when the actual error figure is an ellipsoid. If we make a large number of position fixes, we can say that

the SEP is the radius of the sphere containing 50% of the individual fixes. The Department of Defense's accuracy goal for GPS is to have a worldwide SEP of 15 metres.

If we forget about the height coordinate for the time being and consider just the horizontal coordinates, we can construct the two-dimensional analogue to the error ellipsoid: the *error ellipse*. It is defined as the contour of equal probability density in the two horizontal dimensions. There is a certain probability that the true horizontal coordinates lie within the ellipse. For the *standard error ellipse*, this probability is 39%. As with the error ellipsoid, the semi-axes of the error ellipse are not in general equal to the standard deviations. However, given the standard deviations in the horizontal coordinates and their correlation, the semimajor and semiminor axes of the ellipse can be calculated.

The two-dimensional analogue of SEP is circular error probable (CEP). CEP is the radius of a circle inside of which the true horizontal coordinates of a position have a 50% probability are being located.

Another accuracy measure frequently used in navigation is twice the root-mean-square of the horizontal distance error or 2 d_{rms} for short. It is equal to twice the square root of the sum of the squares of the semimajor and semiminor axes of the error ellipse. A circle of radius 2 d_{rms} will contain the true horizontal position with a certain probability. Unfortunately a drawback of 2 d_{rms} as a measure of error is that it does not correspond to a fixed value of probability for a given value of error. The probability varies with the eccentricity of the error ellipse, ranging from 95.4% (ellipse collapses to a line) to 98.2% (ellipse becomes a circle). Because of this variation in probability, there is not a constant relationship between values of 2 d_{rms} and CEP. The ratio of 2 d_{rms} to CEP varies with the eccentricity of the error ellipse from 2.4 to 3.

Because of its wide use in navigation, 2 d_{rms} is used to specify the designed level of horizontal positioning accuracy for the GPS Standard and Precise Positioning Services. The latest issue of the Federal Radionavigation Plan (FRP) states that when GPS is declared operational, the horizontal accuracy for SPS is planned to be 100 m 2 d_{rms} at 95% probability. This means that 95% of all horizontal position fixes should be within 100 m of the true position. But what about the other 5%? Theoretically, if the position errors due to the various UEREs including S/A are from a Gaussian distribution, we could occasionally get extremely large errors. However, the Department of Defense will control S/A such that excursions will not exceed 300 m 99.99% of the time. The corresponding designed 2 d_{rms} (95%) horizontal accuracy for PPS is planned to be 17.8 m.

The FRP describes the designed accuracy of the vertical component of a GPS-derived position at the 2σ level. As this corresponds to a 95% probability level it is consistent with the accuracy quoted for the horizontal position. For SPS, the designed vertical 2σ is 156m; for PPS it is 27.7m.

The FRP gives the designed accuracy of receiver clock synchronization at the 1σ level. For SPS, σ_t is planned to be 167 nanoseconds. For PPS, σ_t is given conservatively as 100 nanoseconds.

It should be pointed out that the stated PPS position and time accuracies are designed estimates of GPS capabilities and superior results have already been obtained in practice. Significantly greater accuracies can be obtained for both PPS and SPS users by operating in a differential mode with two or more receivers used simultaneously. In fact, almost all of the effects of S/A can be removed when operating in this mode.

3.8 Coordinate Systems
3.8.1 *WGS 84.*
With the requirement to tie different regional datums together for military and other purposes and with the advent of satellite-based positioning systems, there was a clear need for a global geodetic reference system. One of the first such systems was the U.S. Department of Defense World Geodetic System (WGS) introduced in 1960. WGS 60 was created from a global data base of conventional geodetic measurements as well as satellite observations and Hiran (High Precision Short Range Navigation), an airborne trilateration or range measuring system developed during and after World War II. In the years following the introduction of WGS 60, the accuracy and number of satellite observations greatly increased and led to the development of WGS 66 and subsequently WGS 72.

WGS 72 was adopted as the CTS for describing the orbits of the GPS satellites in their navigation messages. But like that of WGS 60 and WGS 66, the accuracy of WGS 72 eventually was found wanting. WGS 72 was superseded by WGS 84 and has been used for the GPS navigation messages since January 1987.

The reference ellipsoid of WGS 84 is essentially that of the International Union of Geodesy and Geophysics (IUGG) Geodetic Reference System 1980 (GRS 80) with some minor changes. This ellipsoid was adopted by the international geodetic community at the 17th quadrennial meeting of the IUGG in Canberra in 1979 as best representing the size and shape of the earth. The WGS 84 ellipsoid is specified by the value of 6,378.137 km for its semimajor axis, a form factor describing the earth's equatorial bulge from which a flattening of 1/298.257 223 563 is derived as well as values for the mean rotation rate of the earth and a value for the product of the mass of the earth and the gravitational constant. How well does the WGS 84 ellipsoid represent the actual earth? The equatorial radius is probably in error by no more than 1 or 2 metres and the value for the flattening is accurate to about 3 parts in a million.

The coordinate system of WGS 84 is a realization of the CTS as established by the Bureau International de l'Heure (BIH) on the basis of coordinates adopted for BIH stations (the BIH was a forerunner of the International Earth Rotation Service). The Z axis of the coordinate frame is parallel to direction of the CTP; the X axis lies at the intersection of the CTP's equatorial plane and the zero meridian; and the Y axis completes the system. The centre of the coordinate frame coincides with the centre of the WGS 84 ellipsoid and the coordinate axes coincide with the rotational axes of the ellipsoid.

The WGS 84 coordinate system was established using a set of coordinates for a worldwide network of U.S. Navy Navigation Satellite System (Transit or Doppler) stations. However, it had been learned from co-siting Doppler stations with VLBI and satellite laser ranging (SLR) stations that the Doppler coordinate system had metre-level errors. For example, it was discovered that the origin of the coordinate system was about 4.5 metres above the earth's centre of mass. So the Doppler station coordinate set was modified in origin, scale, and orientation to agree in the mean with the VLBI and SLR results.

WGS 84 also includes a description of the earth's gravity field. Knowledge of the gravity field is needed, for example, for modelling the orbits of satellites. The field is described by a series of coefficients or values which account for smaller and smaller features with each additional term. These terms are

called *spherical harmonics*. Spherical harmonics are identified by a pair of indices called the *degree* , m, and *order*, n. The harmonics with m = 0 are independent of longitude and are called *zonal harmonics*. Those with m = n are independent of latitude and are called *sectorial harmonics*. The others are functions of both latitude and longitude and are called *tesseral harmonics*. The WGS 84 gravity field is complete up to degree and order 180. This means that variations in the field originating from an area of roughly 200 km by 200 km in extent are described. Becuase of the military importance of this model of the gravity field, only the coefficients up to degree and order 18 are unclassified. The field coefficients were obtained from the analysis of Doppler satellite tracking data, SLR data, surface gravity data, oceanic geoid heights from satellite altimetry, and GPS data as well as results from the analyses of the orbits of a number of other satellites.

The coordinates directly computed by a GPS receiver refer to the same system as the satellite coordinates that are used in the computation. Since these normally refer to WGS 84, so do the receiver coordinates. These coordinates may be expressed as Cartesian coordinates (X, Y, Z) or geodetic ellipsoidal coordinates (ϕ, λ, h). However, most GPS receivers also provide an option to transform the coordinates to one of a number of different regional datums such as those listed in Table 3.3. In the table, the datum shifts give the mean offsets of the reference ellipsoids from the centre of the WGS 84 ellipsoid, the reference frame rotations represent the misalignment of the axes the coordinate systems with respect to that of WGS 84, and the scale parameter accounts for differences in the scale of lengths in each datum with respect to the scale of WGS 84. The parameters in Table 3.1 are mean values from a theoretical study and should not be used for everyday mapping, charting, and survey applications. Such mean values do not account for variations in scale and the distortions in the datums which might exist. For transformations between datums, DMA provides a set of multiterm polynomials for specific regions that it has derived from extensive least squares analyses.

The receiver can also display orthometric height rather than geodetic height if the geoidal undulation is known.

3.8.2 *NAD 83.*
GPS users in North America currently have a slight advantage over users in other regions as a result of the recent introduction of a new datum, the North American Datum of 1983 (NAD 83). NAD 83 replaces the North American Datum of 1927 (NAD 27). NAD 27 coordinates had become inadequate for many purposes. Many of the published coordinates of survey control stations were unreliable due to errors and distortions in the system. In fact, relative coordinates in NAD 27 were sometimes in error by as much as 1 part in 15,000. A further disadvantage of NAD 27 was that its reference ellipsoid was non-geocentric and was not precisely oriented with respect to the CTS as established by the BIH.

The need for a readjustment of North American networks was realized in the late 1960s and work on the new datum officially began in 1975. NAD 83 was obtained by a least squares adjustment of over one and three quarter million geodetic observations at sites in the United States, Canada, Greenland, Mexico, Central America and the Caribbean. In additional to a large number of observations obtained using traditional surveying techniques, Doppler and VLBI observations were also used. Originally scheduled for completion in late 1982, the new datum was christened NAD 83. But due to various delays and extensions the first NAD 83 coordinates were not published by the National Geodetic Survey until March 1987.

The reference ellipsoid and coordinate system of NAD 83 are almost identical to those of WGS 84; the two systems agree at about the 0.1 mm level. So, WGS 84 coordinates provided by a GPS receiver can be taken to be NAD 83 coordinates. However, the GPS surveyor should realize that if he occupies a particular reference marker for which published NAD 83 coordinates exist, the coordinates he computes for the marker from GPS observations may occasionally differ from the published coordinates by a metre or more due to remaining errors and distortions in the datum. Nevertheless, the vast majority of published first- and second-order NAD 83 coordinates are accurate relative to one another at the 3 to 4 parts-per-million level.

To provide geodetic reference coordinates at a higher accuracy than afforded by standard NAD 83 coordinates, many states have established special GPS-derived "NAD 83 high precision" networks.

3.8.3 *UTM.*
The GPS receiver may also be able to project the ellipsoidal coordinates onto a mapping plane, that is, a flat map. It is impossible to project an ellipsoidal surface onto a flat surface without some distortion taking place. However, projections have been developed which minimize these distortions. One such projection is the Universal Transverse Mercator (UTM). The UTM projection, which can trace its lineage back to Karl Friedrich Gauss, has been adopted by the IUGG, NATO and other military organizations and many civil administrations worldwide for their various mapping needs.

The UTM projection divides the world between 80°S and 84°N into 60 zones each with a width of nominally 6 degrees of longitude onto which is superimposed a grid. Each of these zones, which constitutes a segment of a reference ellipsoid, is projected onto a cylinder whose axis is parallel to the earth's equator and whose radius is chosen to keep the scale errors of the projection within acceptable limits. Coordinates of points on the ellipsoid within a particular zone can then be transformed to coordinates on the UTM grid. The UTM coordinates are generally referred to as eastings and northings and are expressed in metres. Eastings are reckoned from the central meridian of a zone and have 500,000 m added to them so that all coordinates remain positive. Northings are reckoned from the equator which has a coordinate value of 0 m for work in the northern hemisphere and 10,000,000 m for work in the southern hemisphere. An important property of the UTM projection is that is uses a conformal mapping which means that the magnitude and sense of angles measured on the ellipsoid are preserved when coordinates are transformed to the mapping plane.

The U.S. State Plane Coordinate System uses a transverse Mercator projection or another conformal projection, the Lambert conic map projection, in one or more zones, to map each state of the Union, Puerto Rico and the U.S. Virgin Islands, onto a plane rectangular coordinate system. (The panhandle of Alaska is a unique case with its own special projection.) The transformations from NAD 83 geodetic coordinates to grid coordinates yield errors less than about 1 cm for points within the boundaries of the appropriate zone so that either geodetic coordinates or the corresponding grid coordinates of a point may be used depending on the application.

4. OTHER SATELLITE NAVIGATION SYSTEMS
4.1 INTRODUCTION
The following discussion briefly summarizes the characteristics of several other satellite navigation systems.
4.2 GLONASS

The former Soviet Union has developed a system very similar to GPS. Called Glonass, for Global Navigation Satellite System, it uses satellite orbits, transmission frequencies, and signal structure which bear a striking resemblance to the U.S. system. However, the signals are different enough so that presently-available commercial GPS receivers are unable to access Glonass signals. The first three Glonass satellites were launched on 12 October 1982 on a single rocket from the Baikonur Cosmodrome near Tyuratam. Since then a further nine trios have been launched, including one set that went into undesirable elliptical orbits. As of early 1988, only five of these prototype satellites were fully functional. The full operational constellation scheduled for deployment by about 1995 will consist of 24 satellites, with 3 in standby mode [Anodina, 1988].

The Glonass satellites are placed into circular orbits with a nominal semi-major axis of 25 507 km giving them an orbital period of about 675.7 minutes. These satellites have ground tracks which repeat every 17 orbits, i.e., 8 sidereal days. However, not all of the prototype satellites are presently in such synchronized orbits. Glonass orbital planes are separated from one another by 120° and have target inclinations of 64.8°. The spacings of satellites within a plane are 30° or 45°.

Like the GPS satellites, Glonass satellites transmit two pseudo-random noise signals with binary phase-shift keying. The clock rate of the Glonass signals is about one-half that of GPS: 5.11 MHz for the equivalent of the P-code and 0.511 MHz for the equivalent of the C/A-code. A 50 bits per second message lasting 2.5 minutes is also superimposed on the signals. The signals are transmitted within two bands: L1, 1597 - 1617 MHz and L2, 1240 - 1260 MHz, with different satellites using different carrier frequencies. The channel spacing for L1 is 0.5625 MHz and for L2, 0.4375 MHz. As with GPS, the P-code is transmitted on both L1 and L2, whereas the C/A-code is only present on the L1 signal. The C/A-code is 511 chips long. The length of the P-code is not presently known. Unlike GPS satellites, all Glonass satellites appear to transmit the same codes. Signal timing and frequencies are derived from one of two on-board atomic clocks [Klass, 1988]. The signals are right-hand circularly polarized, like GPS signals, and have spectral power densities of -44 dbW/Hz in the L1 frequency band and -57 dbW/Hz in the L2 band. Some of the prototype satellites transmit signals which include narrow spectral lines or tones. It is not clear if these are intentional or spurious. If intentional, they could be used for instantaneous velocity determination.

Accuracies of positions determined from the C/A-code of 100 m horizontal, and 150 m vertical are claimed, with velocity component and time accuracies of 15 cm s^{-1} and 1 μs respectively [Anodina, 1988].

Some of the details about Glonass have been sleuthed through monitoring of the satellite signals, primarily by Daly and his co-workers at the University of Leeds [Dale and Daly, 1986; 1988; Klass, 1987a].

4.3 Transit Satellite Positioning

Experiments performed by scientists at the Johns Hopkins University Applied Physics Laboratory (ARL) following the launch of Sputnik 1 showed that it was possible to determine the satellite's orbit by analysing the Doppler shift of the satellite's radio transmissions. It was subsequently realized that if the position of a satellite could be determined by measuring the Doppler shift at a station of known position, then it should be possible to determine the position of the station if the position of the satellite is known. This realization led to the development by ARL of the U.S. Navy Navigation Satellite System, commonly known as Transit.

The first prototype Transit satellite was launched in 1961. Following deployment of a number of test satellites, the system was declared operational in 1964 and became classified. However, in 1967 Transit was declassified and became available for civilian use.

The Transit system consists of three components: the satellite tracking and control facilities, the satellites themselves, and the users. The Transit satellites are controlled by the U.S. Naval Astronautics Group (NAG) at Point Mugu, California. Four tracking stations in the U.S. record Doppler measurements on each pass of every operational satellite. These data are sent to Point Mugu where the orbit of each satellite is determined and extrapolated into the future. This ephemeris is transmitted by one of two injection stations to the satellite where it is stored in memory for subsequent rebroadcast.

Six Transit satellites are presently fully operational, including one launched in May 1967. Of these six, four are of the older Oscar type and two are of the advanced NOVA class. Another four Oscars and one NOVA have been stored in orbit for future use. The satellites are in circular, polar orbits with altitudes of roughly 1100 km and their nodes more or less evenly spaced around the equator. Their corresponding orbital period is about 107 minutes. Each satellite transmits two harmonically-related carrier frequencies, one at 400 MHz, the other at 150 MHz. Both frequencies are actually offset from these nominal values by -80 ppm to make receiver operation simpler. The use of two frequencies permits correction for the dispersive effect of the ionosphere (see Chapter 9). Superimposed on both carriers by balanced phase modulation is a broadcast message containing the orbit information previously injected into the satellite by the NAG. The message is continuously transmitted as a series of two-minute paragraphs, each paragraph consisting of 6103 binary bits. From the message, a new instantaneous orbit can be computed every two minutes.

The received signals are compared with the local oscillator frequency generated in the user's receiver and the Doppler shift, or frequency difference, is integrated to yield the observable. By combining these Doppler counts with the satellite orbit data, accurate coordinates of the receiver can be determined.

The Soviet Union operates a system similar to Transit, called Tsicada.

4.3.1 *Principle Of Doppler Positioning.*

The Doppler effect, discovered by Christian Doppler a nineteenth century Austrian physicist, is familiar to anyone who has waited patiently at a railway level crossing for a train to pass. The pitch of the train's horn or whistle changes as the train passes. It starts out high, changing imperceptibly as the train approaches, then drops noticeably as the train goes through the crossing, and maintains a lower pitch as the train recedes in the distance. This same phenomenon which is so readily apparent at audio frequencies also affects electromagnetic waves. The frequencies of both radio and light waves are shifted if the source (transmitter) and the observer (receiver) are in relative motion.

The classical explanation of the effect is that the observer receives more wave crests per second, i.e., the frequency is increased if the source and the observer are moving closer together, whereas fewer wave crests per second are received, i.e., the frequency is decreased, if the source and the observer are moving farther apart. If the relative speed of the source and observer is much less than the speed of light, then the received frequency is given approximately as

$$f_r \approx f_s \left(1 - \frac{1}{c} \frac{dS}{dt} \right)$$

where f_s is the frequency at the source, c is the speed of light, and S the distance or *range* between the source and the observer; dS/dt is the *range rate*.

Returning to the train at the level crossing, you may have noticed that the closer you are to the track, the faster the change in pitch of the horn. And even if you could not see or feel the train, you can tell when it passes the crossing (*the point of closest approach*) by noting the instant when the pitch of the horn is mid-way between the high and low extremes (f_s). Therefore by monitoring the frequency of the received sound as the train passes and knowing its schedule (position and speed), you can establish your position in a two-dimensional coordinate system where the x-axis, say, runs along the track and the y-axis runs perpendicular to it. The origin may be assigned arbitrarily. This is the principle of Doppler positioning.

In the case of a Transit satellite (or any other satellite for that matter), the position of a receiver can be established by continuously recording the Doppler shift of the received signals (or the number of cycles of the Doppler frequency which is a more precisely obtained observable). Subsequently these data are combined with accurate coordinates of the satellite to determine the position of the receiver. As with the passage of a train, a single satellite pass can provide at most only two coordinates of the receiver's position. Whereas this may be satisfactory for navigation at sea where the height above the reference ellipsoid is approximately known, three-dimensional positioning requires observing multiple satellite passes.

4.3.2 *Transit Doppler Measurements.*
The approximate frequency of a received satellite radio signal (ignoring relativistic effects) is given by

$$f_r \approx f_s \left(1 - \frac{1}{c} \frac{dS}{dt} \right)$$

where f_s is the frequency of the signal measured at the satellite, c is the speed of light, and dS/dt is the range rate. The *Doppler shift frequency*, $f_r - f_s$, is approximately proportional to the range rate, the component of the satellite's velocity vector along the line of sight from the receiver. The maximum range rate of a Transit satellite is about 6.3 km s^{-1} implying a maximum Doppler shift when the satellite rises or sets of 21 ppm of the transmitted frequency. This corresponds to 8.4 kHz at a frequency of 400 MHz.

The Doppler shifts may be measured by differencing the received frequencies from constant reference frequencies in the receiver. For most Transit receivers, these frequencies are 400 MHz and 150 MHz precisely. The satellite transmitter frequencies are approximately 80 ppm lower than the receiver reference frequencies in order that the frequency difference does not go through zero. If the transmitter frequencies were not offset, the receiver would have difficulty distinguishing between positive and negative Doppler shifts.

Most Transit Doppler receivers count the number of accumulated cycles of the Doppler frequency (actually, $f_0 - f_r$) rather than measure the instantaneous Doppler frequency itself, since counting cycles can be carried out more precisely than measuring the instantaneous frequency. The counter is read out at intervals and the data stored. The counter is reset either after each two minute paragraph or at the end of the pass. Sequential differences in counter readings actually constitute a series of biased range differences.

Suggested further readings are Bomford [1980] and Stansell [1978].

4.4 ARGOS
Another satellite system which uses the Doppler effect for positioning is Argos [Service Argos, 1984], a cooperative project of the French Centre National d'Etudes Spatiales (CNES), NASA, and the U.S. National Oceanic and Atmospheric Administration (NOAA). However, in contrast to the Transit system, the transmitters are operated by the users and the receivers are in the satellites. An Argos transmitter on an instrumented 'platform' of some sort (oceanographic or navigation buoy, radiosonde balloon, remote weather station, etc.), periodically emits a 401.65 MHz signal carrying information from the platform's sensors. One of two passing U.S. TIROS/N-class weather satellites picks up this signal and records its Doppler shift along with the sensor data. These data are subsequently played back when the satellite is in range of one of three tracking stations: Wallops Island, Virginia; Gilmore Creek, Alaska; or Lannion, France. The tracking stations relay their data to NOAA's National Environmental Satellite and Data Information Service in Suitland, Maryland, where they are sorted and then passed on to the Argos Data Processing Centre at CNES in Toulouse, France. CNES computes the position of the platform from the recorded Doppler shifts. This information along with the sensor data can be conveyed to the operator of the platform by a variety of means including packet switching data networks, telex, or letter. The two-dimensional (latitude and longitude) positions can be as accurate as ±150 m (1σ). The actual accuracy obtained depends to a very large degree on the stability of the user's transmitter. A new Argos service is being planned that would provide even more accurate positions. Up to 5000 platforms requiring location service can be handled by the Argos system, assuming these are uniformly distributed over the earth's surface.

The Argos receiving system was first implemented on the prototype TIROS/N spacecraft, orbitted in 1978. Subsequent TIROS/N satellites, NOAA-6 through NOAA-10, have also carried the Argos Data Collection System.

Position determination from signals uplinked to a satellite is also utilized in the COSPAS-SARSAT search and rescue system [McPherson, 1981; Elliot and Exter, 1987].

5. SUMMARY
The electronic components of a satellite navigation receiver are very amenable to high volumn - low cost production. The impact of adding satellite navigation to an aerospace vehicle in terms of size, weight and space is minimal. Hence the major technical challenge in application of satellite navigation is in the design to functionally and physically integratae its manifold capabilities in the most efficient manner.

5. FEDERAL RADIONAVIGATION PLAN

Acknowledgements

References

Ackroyd, N. and R. Lorimer (1990). Global Navigation, A GPS User's Guide. Lloyd's of London, 202 pp.

Advisory Group for Aerospace Research and Development (1988). The NAVSTAR GPS System. AGARD Lecture Series 161. Neuilly-sur-Seine, France, September.

anon. (1987). GPS: NAVSTAR GPS Space Segment - Navigation User Interface, Report ICD-GPS-200, November.

· anon. (1987). System Specifications for Navstar GPS Space Segment/Navigation Users Interface. Interface Control Document - Global Positioning System - 200 (ICD-GPS-200), 10 November.

ARINC Research Corporation (1991). GPS Interface Control Document. ICD-GPS-200, 3 July, 115 pp.

Canadian GPS Associates (1993). GPS Biliography - July 93 edition.

Georgiadou, Y., and K.D. Doucet (1990). The issue of selective availability. GPS World, September/October, Vol. 1, No. 5, pp. 53-56.

Hofmann-Wellenhof, B., H. Lichtenegger, and J.Collins (1992). Global Positioning System, Theory and Practice. Springer-Verlag, 326 pp.

Institute of Navigation, The (1980). Global Positioning System. Vol. I. Collected reprints of papers published in Navigation. The Institute of Navigation (U.S.), 246 pp.

Institute of Navigation, The (1984). Global Positioning System. Vol. II. Collected reprints of papers published in Navigation. The Institute of Navigation (U.S.), 257 pp.

Institute of Navigation, The (1986). Global Positioning System. Vol. III. Collected reprints of papers published in Navigation. The Institute of Navigation (U.S.), 293 pp.

Kleusberg, A. (1990). *Comparing GPS and GLONASS.* GPS World, November/December, Vol. 1, No. 6, pp. 52-54.

Kleusberg, A. (1992). Precise differential positioning and surveying. *GPS World*, July/August, Vol. 3, No. 7, pp. 50-52.

Kleusberg, A., and R.B. Langley (1990). The limitations of GPS. *GPS World*, March/April, Vol. 1, No. 2, pp. 50-52.

Klobuchar, J. A. (1991). Ionospheric effects on GPS. *GPS World*, April, Vol. 2, No. 4, pp. 48-51.

Langley R.B. (1992). The Federal Radionavigation Plan. *GPS World*, March, Vol. 3, No. 3, pp. 50-53.

Langley, R. B. (1990). Why is the GPS signal so complex? *GPS World*, May/June, Vol. 1, No. 3, pp. 56-59.

Langley, R. B. (1991). The GPS receiver: An introduction. *GPS World*, January, Vol. 2, No. 1, pp. 50-53.

Langley, R. B. (1991). The mathematics of GPS. *GPS World*, July/August, Vol. 2, No. 7, pp. 45-50.

Langley, R. B. (1991). The orbits of GPS satellites. *GPS World*, March, Vol. 2, No. 3, pp. 50-53.

Langley, R. B. (1991). Time, clocks, and GPS. *GPS World*, November/December, Vol. 2, No. 10, pp. 38-42.

Langley, R. B. (1992). Basic geodesy for GPS. *GPS World*, February, Vol. 3, No. 2, pp. 44-49.

Leick, A. (1990). GPS Satellite Surveying. John Wiley and Sons, Inc., New York, Toronto, 352 pp.

Logsdon, T. (1992). Navstar Global Positioning System. Van Nostrand Reinhold, 249 pp.

May, M. B. (1992). Measuring velocity using GPS. *GPS World*, September, Vol. 3, No. 8,

McDonald, K. D. (1991). GPS in civil aviation. *GPS World*, September, Vol. 2, No. 8, pp. 52-59.

Seeber, G. (1989). Satellitengeodasie. Verlag de Gruyter.

Seeber, G. (1992). Satellite Geodesy.

Wells, D., and A. Kleusberg (1990). GPS: A multipurpose system. *GPS World*, January/February, Vol. 1, No. 1, pp. 60-63.

Wells, D.E., N. Beck, D. Delikaraoglou, A. Kleusberg, E.J. Krakiwsky, G. Lachapelle, R.B. Langley, M. Nakiboglu, K.P. Schwarz, J.M. Tranquilla, and P. Vanříček (1987). Guide to GPS Positioning. 2nd printing with corrections. Canadian GPS Associates, Fredericton, N.B., 600 pp.

Yunck, T. P. (1992). A new chapter in precise orbit determination. *GPS World*, October, Vol. 3, No. 9, pp. 56-61.

AN OVERVIEW OF OMEGA
RADIO NAVIGATION   SYSTEM

by
D.F. LIANG
Directorate Research and Development Communications and Space
National Defence Headquarters
Ottawa, Ontario,
Canada K1A OK2

1.   SYSTEM DESCRIPTION

Omega is a very low frequency (VLF) navigation system using frequencies allocated for radio navigation in the band between 10 KHz and 14 KHz. Such a low frequency band enables Omega navigation system to achieve the long operating ranges required for global coverage, at the same time provides a stable and predictable propagation environment. This also makes it the only radio navigation system applicable to completely submerged submarines.

Omega is supported by 8 transmitting stations located in Norway, Liberia, Hawaii, North Dakota, La Reunion, Argentina, Japan and Australia. These stations are, in general 5000-6000 nm (nautical miles) apart. They are designated alphabetically with the letters A to H. These transmitters are independent of each other in the sense that they do not operate in a master-slave configuration. Each station transmits three frequencies, (10.2, 11.33 and 13.6 kilohertz (KHz), time sequenced within a basic repetition rate of 10 seconds. A fourth frequency at 11.05 KHz as well as one discrete frequency for each station, have been added to facilitate station identification, conventional navigation and also to support special applications such as

buoy-tracking, and search and rescue operations.

Omega is a hyperbolic position fixing system, the hyperbolae are loci of constant time difference between the arrival of signals from two transmitting stations. The time difference is measured as a difference in phase of the two received signals. Each hyperbola is known as a line of position (LOP).

Signals must be received from at least three stations with one of them serving as a common station, to obtain a position fix from the two LOPs. The hyperbolic mode is attractive, since it removes the need to use a precise and expensive local oscillator. It is also possible to work with two stations using a circular mode, which requires a precision local oscillator. In such a case, the position fix is obtained from the intersection of circular rather than hyperbolic LOPs.

Figure 1 shows two transmitting stations A and B radiating phase-synchronized signals simultaneously. The concentric circles around each transmitter are used to represent wavelengths at the transmitted frequency. For a receiver located at point X on the baseline joining the two transmitters, the signals are received from the two

transmitters in phase. A full cycle difference occurs every half wavelength in distance, since in moving from point X to Y, the receiver has moved by exactly a half wavelength towards B. The area between two adjacent zero phase difference LOPs is called a lane. The phase measurements are repeated through 360° cycle. Since the lane is subdivided into 100 centilanes, each centilane (cel) is then equivalent to 3.6° of phase difference. At 10.2 KHz, the lane width, which is at half a wavelength is approximately 8 nautical miles.

2.   LANE AMBIGUITY

Since Omega stations are far apart, on the average there are 600-700 lanes between any of them. To reduce ambiguity, each lane is numbered for its identification. The lane number must be initialized to an accuracy of half a lane and the lane count must be properly maintained. The lane counting can be manually maintained or performed automatically in most modern Omega receivers. When the Omega receiver is unable to keep the correct lane count, this is referred as a lane slip or lane jump, which can be in the multiples of its lane width.

Single frequency receivers use the 10.2 KHz signal which has a lane width of about 8 nm on the baseline between stations. Because of the lane ambiguity, the receiver must be preset to a known location at the start of a voyage. The accuracy of that position must be known to sufficient accuracy to be within the lane that the receiver is capable of appropriate lane identification. This has to be achieved to within a mere 8 nm for a single frequency receiver at 10.2 KHz. To effectively use a single frequency

receiver, it is essential to carefully maintain a DR plot, and count the number of lanes it crosses in the course of a voyage. It must be periodically compared to its Omega position so that any lane ambiguities can be detected and corrected. Unless the Omega position is occasionally compared to the position fix of another navigation system or to a carefully maintained DR plot, the possibility of lane count error increases with time and distance. However, two or three - frequency receivers can be used to reduce the position fix ambiguity. Navigational frequencies of 13.6 KHz and 10.2 KHz have a frequency difference of 3.4 KHz, the phase of which can be measured by some receivers. 3.4 KHz LOPs coincide with every third 10.2 KHz one, which has exactly a lane width of 8 nm to give a lane width of 24 nm. With this, the operator only need to know the platform location to within an accuracy of 12 nm along the baseline without any ambiguity. Having established the lane count, it should then be possible to determine its centilanes without ambiguity.

The lane resolution process can be further extended by using even lower frequency differences. The frequency difference between 11.33 KHz and 10.2 KHz is 1.13 KHz, which can provide a lane width of 72 nm. The frequency difference between 11.33 KHz and 11.05 KHz can be used to generate LOPs with a lane width of 288 nm, which is 36 times the normal lanes width of 10.2 KHz. The use of the latter frequency difference is aimed primarily at air navigation where the high speed of the platform makes the use of wider lane width imperative.

3. PROPAGATIONAL PERFORMANCE

The Omega position fixing depends on the differences in phase measurements of computed LOPs. The signals are propagated, in effect, by a waveguide, with the earth's surface and D layer of the ionosphere as the boundaries. Therefore, ionospheric conditions greatly affect Omega signal propagation. A key factor is the variation of ionospheric height which changes from seasons to seasons and from day to night. The major error sources then are due to the inability to precisely determine atmospheric phenomena and signal conductivity. Meteorological effects, however, are known to introduce negligible effects to VLF propagation.

3.1 Diurnal Variations

Fundamental to any VLF propagation is the propagation velocity, which can be used to relate a phase or time difference to the geography of the hyperbolic position lines. The height of the ionosphere changes from approximately 70 Km by day to 90 Km by night, such a variation is a considerable change in the dimension of the waveguide, which affects the signal propagation. The resulting variation can be of the order of one whole lane. Fortunately, such diurnal changes are cyclic, and therefore predictable over a 24 hour period using skywave corrections. For an all-daylight propagation path, the correction can be accurate to five centicycles, whereas for an all-night-time path, the accuracy can be up to ten centicycles. The necessary corrections are presented in US Defence Mapping Agency Omega Propagation Tables. They take into account changes in propagation velocity due to ionospheric height variations, earth magnetic field orientation, and to surface conductivity. In general, the navigational accuracy of the Omega system depends on the quality of these published corrections. However, there are other unpredictable disturbances that can affect the propagation conditions.

3.2 Modal Interference

Omega transmissions are launched from vertically polarized antennas which generate transverse-magnetic (TM) propagation modes within the waveguide. At a distance of 1,000 Km from the transmitters, only the first and second modes, TM 1 and TM 2 (see Figure 2) are large enough to warrant consideration. Close to the transmitters, the TM 2 mode predominates, but its attenuation rate is greater than that of the TM 1 mode. The TM 1 and TM 2 modes propagate with different phase velocities, and therefore in areas where their amplitudes are comparable, the two modes will produce interferences that can advance or retard the phase propagation. It is generally unwise to use Omega within 200 nm of any Omega transmitters.

Modal interference from the Liberia transmitter has also been severely experienced as far away as the east coast of the US during night time. It can cause large errors, especially when each mode predominates separate day - night time paths.

Some receivers have the software algorithms to prevent near-field modal interference for both day and night-time operations and automatically disable the selection

of stations whose ranges are less than a preset minimum range.

3.3 Solar Flare Effects

Solar Flare Effects are known to have major impacts on Omega propagation, especially Sudden Ionospheric Disturbances (SID) and Polar-Cap Absorption (PAC) event.

3.3.1 Sudden Ionospheric Disturbances (SID)

Sudden ionospheric disturbances are primarily caused by an increase in the energy of x-ray flux emancipated by the sun due to solar eruptions or "flares".

Increased energetic x-ray radiation in the D-layer, can reduce the waveguide height affecting the VLF signal propagation. This causes a phase advance on propagation paths that are exposed to the sun. These phase changes are known as sudden phase anomalies (SPA).

Such a disturbance can happen suddenly peaking within 5-6 minutes followed by a slow decay of over a period of 50 minutes. A typical phase shift is about 10-20 centicycles and occurs mostly in lower latitude regions. In the year of sunspot activity, the SID may occur as much as once per day.

3.3.2 Polar Cap Absorptions (PCA)

This is due to the precipitation of solar particles, primarily protons, in the vicinity of the earth's magnetic poles shortly following significant solar flares. It causes an increase in ionization of the D-region, resulting in an effective lowering of the D-layer ionosphere. The amount of propagation anomaly depends on the extent of the

transmission path crossing the magnetic pole region. Therefore, a propagation path away from the polar region will not be affected by PCA. For users in high-latitude regions, it may be difficult to completely ignore signal paths over polar routes, especially when some transmitters are off the air. The PCA effect does not occur as frequent as the SID. The effect of PCA tends to come on over several hours and slowly decay over a period of 1-2 weeks. The maximum effect can vary from 10-50 centicycles.

4. PHASE PROPAGATION CORRECTIONS

A most effective way of accounting for Omega propagation errors is the application of published Phase Propagation Corrections (PPC). The 1980 PPC Model [1] is widely employed through the Omega user community. The application of propagation corrections reduces the positioning errors to the 2-4 nm range. Figure 3 presents the error of an Omega LOP, over a 4 day period in November 1987, which shows significant repeatability of the phase error as a function of the time of the day. The PPCs initially was intended to be updated every 5 years, which unfortunately has not been fulfilled.

Recent examination of the PPC [2] has led to the conclusion that the 1980 Model contains a term which grows exponentially with time. The retention of the time term into the 1990s should be a cause for concern. For this reason, the US Coast Guard has eliminated it from its "Official" PPC Model.

5. OEMGA ERROR MODEL

A significant part of Omega phase errors can be compensated by using

Phase Propagation Correction (PPC) tables. The residual error can be represented by the following four major random components:

a. A slowly varying component representative of the spatially correlated phase error due to slowly varying factors such as ground conductivity;

b. A 24-hour periodic diurnal component used to represent errors in phase propagation corrections;

c. A short time constant Gauss-Markov process phase error;

d. A white noise representing receiver dependent measurement noise.

The auto correlation function of the Omega phase errors are represented in Section 4.8 of [3]. A four state error equation to model each Omega phase error is also represented in the same reference.

Since an Omega line of position error is the difference between two independent phase errors, the statistical properties of the LOP error will be basically the same as that of a phase error. For the sake of simplification, a two state representation can be used as a substitute for the four state model. This was accomplished by omitting the periodic states and adjusting the parameters of the weakly correlated Markov processes to maintain the correct mean values and correlation time. The phase error of each Omega signal is therefore modelled by the following two Markov processes:

$$\frac{d}{dt}\begin{bmatrix} BIAS \\ MARKOV \end{bmatrix} = \begin{bmatrix} -1/T1 & 0 \\ 0 & -1/T3 \end{bmatrix}\begin{bmatrix} B \\ MA \end{bmatrix}$$

Where T1 and T3 are correlation time constants.

6.   DIFFERENTIAL OMEGA

Differential Omega is a method of providing increased relative position fixing accuracy with respect to a local reference station. Normal Omega system is subject to errors in diurnal corrections and propagation effects, therefore limiting the accuracy of the system. Even though Omega transmits VLF signals at long distance, its propagation errors are highly correlated over a small geographic region. Therefore, a reference receiver can be established at a precisely known location to measure the LOP errors for compensation to users in the neighbourhood of about 200 nm.

Studies have shown that the position accuracy can be reduced to 200 meters at a few miles of the reference stations, and up to 400-600 metres at 100 nm for a full power operational system [4]. However, at a distance of 200-300 nm, the accuracy deteriorates to that of basic Omega performance. The Differential Omega has not been widely adopted because the corrections can only be applied over a relatively small area, and that reference stations must be installed and maintained together with the associated communication links to transmit the corrections. In spite of this, there are currently 22 Differential Omega reference stations installed around the world,

and some 2000 receivers have been in
service worldwide.

The policy statement of 1988 Federal
Radionavigation Plan (FRP) had
generated considerable concerns
within Omega User community that the
Omega was headed for oblivion by the
end of the century. The FRP
reported that the normal transition
period of 10-15 years might be
shortened.

In the 1990 FRP, even though the DOD
requirement did not change from the
1988 version, it states that "Omega
aviation service requirements will
remain until there is a suitable
replacement". Some in the aviation
community can argue that there will
never be a suitable substitute for
Omega in low cost global coverage.
The recognition of aviation service
requirement is significant in light
of DOD's termination of its
requirement beyond 1994. Without
such a substantiated aviation
requirement, the need for Omega
beyond the end of this century
becomes uncertain. With the
widespread use of Omega in aviation,
Omega can be expected to serve as a
global supplementary aid to GPS and
INS.

References:

[1]   E.R. Swanson and P.B. Morris, "New Coefficients for the Swanson Propagation
       Correction Model," *Proceedings of the Fifth Annual Meeting, International Omega
       Association*, Bergen, Norway, August 1980, pp. 26-1 to 26-4.

[2]   R.J. Wenzel etc, "The 1980 Omega PPC Model Time Term Should Not Be used,"
       Proceedings of the Fifteen Annual Meeting, International Omega Association, Bali,
       Indonesia, September 1990, pp. 10-1 to 10-10.

[3]   D.F. Liang "An Overview of Generic Multi-sensor Integrated Navigation System
       Design," This ACARDOGRAPH.

[4]   Nard, "Results of Recent Experiments With Differential Omega," *Navigation Journal of
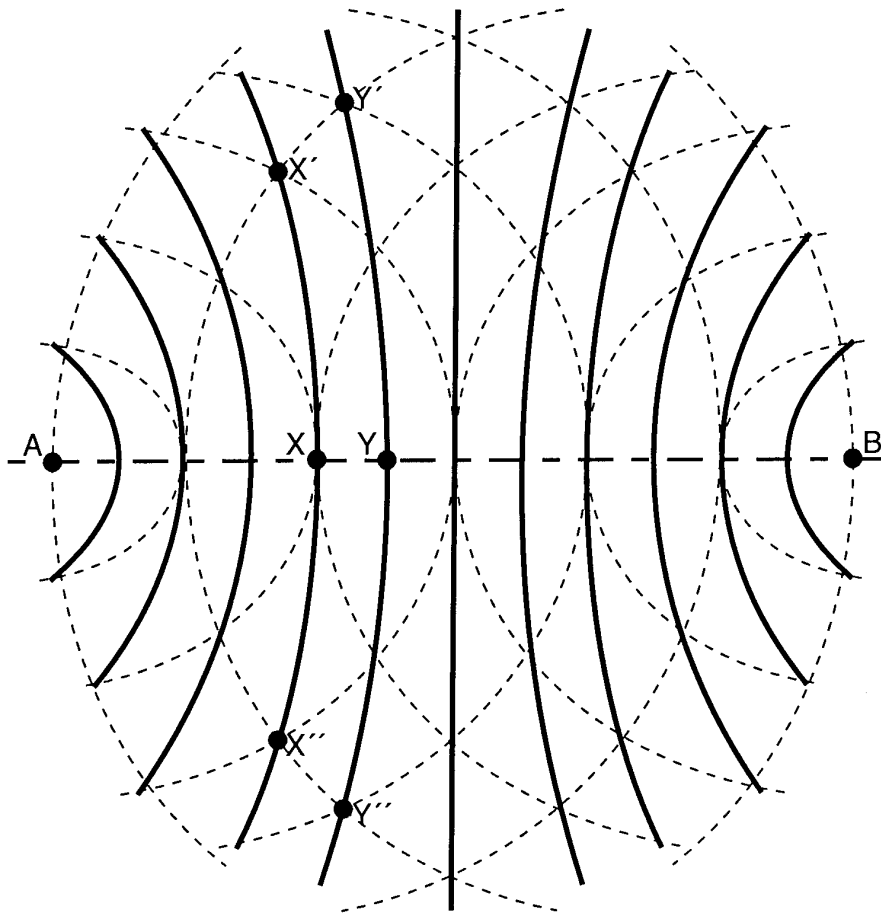   the Institute of Navigation*, Vol. 19, No. 2, Summer 1972.

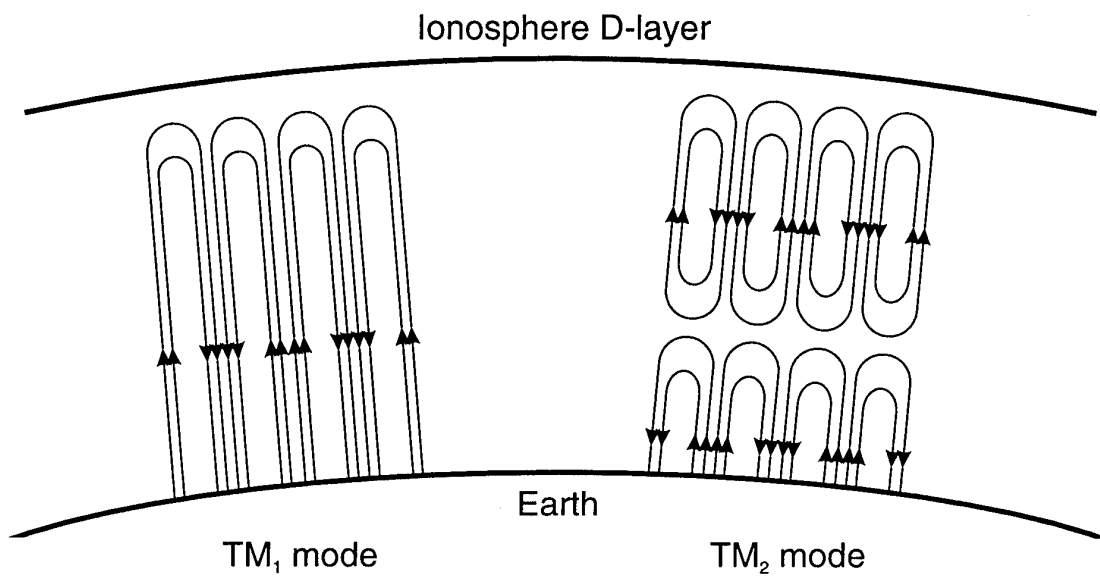Figure 1. Simplified Omega Hyperbolic Lattice.

Ionosphere D-layer

Earth
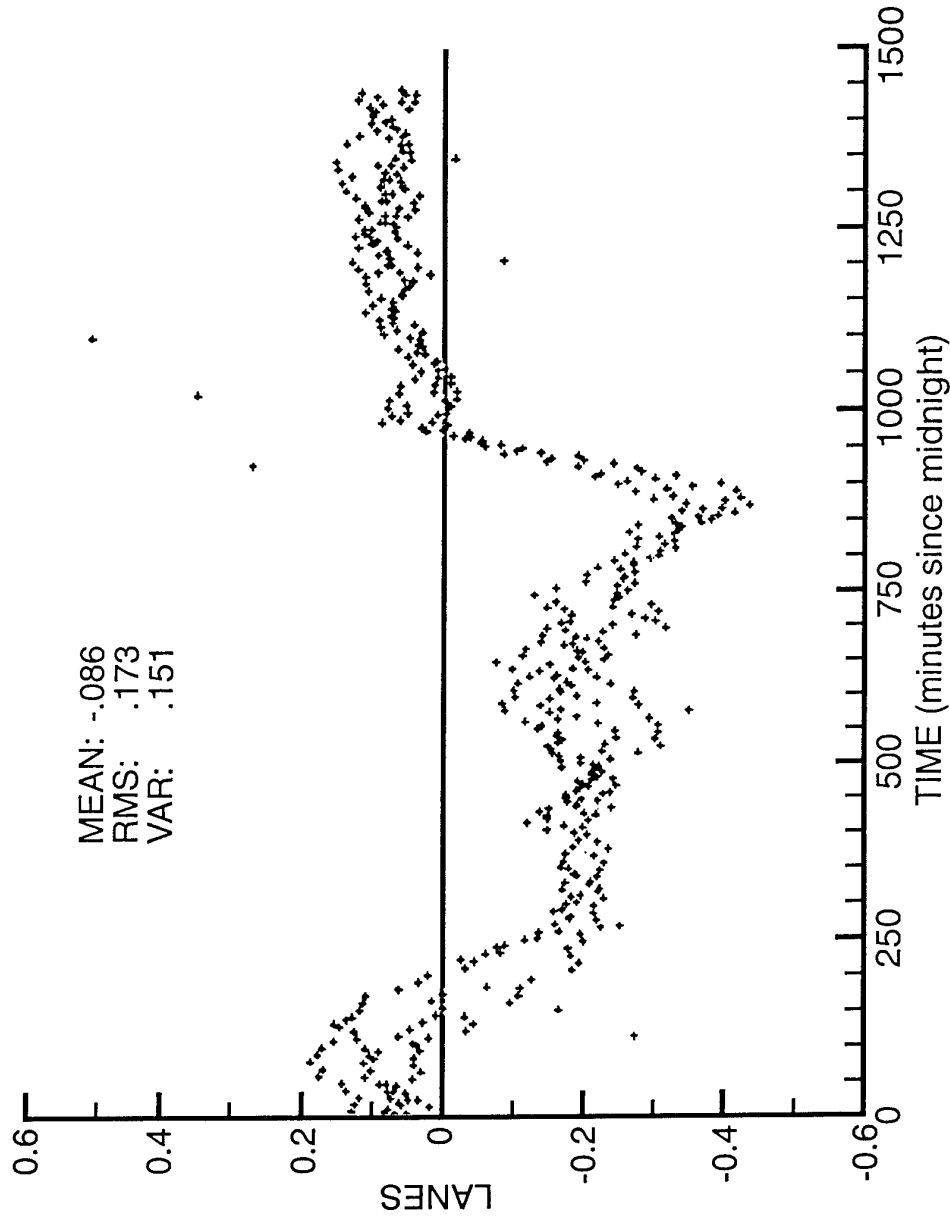
TM₁ mode

TM₂ mode

Figure 2. TM Modes Propagation.

Figure 3. Omega LOP errors without PPCs.

AGARDograph on Advanced Astroinertial Navigation Systems

Seymour (Sy) Levine
Northrop Corporation
Electronics Systems Division
2301 West 120th Street
P.O. Box 5032
Hawthorne, CA 90251-5032
USA

ABSTRACT

This paper discusses modern Strapdown Astroinertial Navigation (SAIN) systems as autonomous navigators for manned aircraft, ships, missiles, and remotely piloted vehicles. These systems, which approach Global Positioning System (GPS) accuracy, do not depend upon man-made electromagnetic radiating devices that may be intentionally shut down, destroyed, or become unreliable in a hostile environment. The paper analyzes the gyroscopic accuracy, artificial stellar image stabilization, star density, sky visibility, and sky background irradiance effects on system performance. It concludes that a high-precision, reliable, low-cost stellar inertial system can be achieved by eliminating gimbals and combining a strapdown Inertial Navigation System (INS) with an Optical Wide-Angle-Lens Startracker (OWLS).

1.0 INTRODUCTION

For several hundred years, man successfully used star positions to reset his long-range navigation systems. During his earthly journeys, the stellar augmentation subsystem-the sextant-provided updates to minimize the accumulated errors in his continuous Dead Reckoning (DR) navigation. As he approached his destination or wartime target, he resorted to a Relative Navigation (RN) solution. His eyes took on the function of a terminal seeker. They focused on the terminus of his mission (e.g., dock, etc.) or on the position of his enemy. By the nineteenth century, man had an autonomous precision-navigation suite that utilized DR, stellar updating, and terminal seeking to achieve a Circular Error Probable (CEP) measured in feet.

Although present autonomous navigation suites bear little physical resemblance to their predecessors, the system elements are the same (see Table 1). They consist of a DR continuous navigator, i.e., an Inertial Navigation System (INS), periodically updated or reset by a DR reference system such as a stellar tracker and altimeter. After the navigation system guides the vehicle into an operational area or basket of the terminal seeker, the seeker provides precise closure for docking, landing, or interdiction of fixed and maneuvering targets. This autonomous suite is free from radio aids, which may be unreliable, jammable, or unavailable during a twenty-first century wartime encounter.

Since the 1950s, inertial systems have been the dominant DR navigators aboard ships and air vehicles. During the 1950s, these devices were made with gyroscopes and accelerometers mounted on one gimbal and surrounded by two or three other gimbals to isolate the inertial sensors from the vehicle's angular rates. The inertial element assembly was usually called the stable member, or yaw gimbal. To provide high precision, a Narrow Field of View (NFOV) startracker, embedded in train and elevation pointing gimbals, was mounted on top of the yaw gimbal. In this configuration, there were usually five gimbals, multiple gimbal resolver angular readouts, a servo-mechanism, and associated electronics. A limited quantity of two-gimbal versions, the minimum number of gimbals necessary to point a telescope at any point in three-dimensional space, were also developed. These gimballed astroinertial navigators were successfully used as the navigation reference for reconnaissance, precision rendezvous, and indirect, coordinated firing of munitions.

Figure 1 shows the performance of pure, unaided inertial systems, stellar inertial systems, and GPSs, starting in the year 1950 and projected to the beginning of the 21st century. As the curves indicate, the error of a terrestrial inertial system operating for 10 hours is measured in miles whereas the stellar system error is measured in hundreds of feet. Figure 1 further shows that the strapdown system position accuracy is rapidly approaching the precision of the gimballed system. In addition, the strapdown system's attitude accuracy (e.g., roll, pitch, and azimuth) is now superior to its gimballed predecessor. This strapdown attitude superiority is primarily due to the attitude being a computer-generated number as compared to an inaccurate gimbal angular readout such as from a resolver or optical shaft encoder.

These strapdown INSs not only provide position, and the Euler angles of roll, pitch, and heading, but they also supply the timely, accurate angular rates required for the vehicle's flight control and autopilot. The angular rate data, flight-critical to dynamically unstable vehicle control systems, is relatively simple to obtain in a strapdown INS as compared to the gimballed INS process, which required the costly and noisy differentiation of the gimbal resolver readouts.

In 1990, work proceeded on the development of a Strapdown Astroinertial Navigation (SAIN) system capable of producing the same autonomous high-precision navigation that exists on a gimballed stellar inertial system.

2.0 SYSTEM PERFORMANCE

Figure 2 presents the canonical model of a stellar INS; Figure 3 shows a simplex model. The strapdown inertial system, utilizing its gyroscopes and accelerometers, provides estimates of the vectors of position (e.g., latitude, longitude, and altitude); velocity (e.g., north, east, and vertical); and Euler angles (roll, pitch, and heading).

The raw inertial vectors are combined with a self-contained ephemeris and chronometer (an error of 1 s of time in the chronometer represents a 1,500 ft position error) to predict a star fix angle. The astrotracker then uses this estimated coordinate location to find the actual star and to provide an error vector representing the difference between the predicted star's angular coordinates and the angular coordinates of the astrotracker's measured star. A single star observation does not provide enough information to compensate three gyroscopic orthogonal axes of rotation. This is because the single stellar observation is unable to sense rotations around its line of sight. Thus, to measure all three gyroscopic axes, a second noncoincident observation, preferably orthogonal to the first, must be made. Additional measurements are used to further characterize and refine the error vector. This error vector is then processed optimally in a Kalman filter to yield the best estimates

TABLE 1. AUTONOMOUS NAVIGATION SUITE CHRONOLOGY

ELEMENT \ YEAR	NAVIGATION COMPONENTS (1800 — 1940 — 1950 — 1960 — 1970 — 1980 — 1990 — 2000)
DEAD RECKONING	MAGNETIC COMPASS → FLUX VALVE →; MAGNETOMETER →; GYROCOMPASS → GIMBALLED INERTIAL NAVIGATION → STRAPDOWN INERTIAL NAVIGATION →; DOPPLER RADAR →; LOG → PITOT TUBE/EM LOG →; INCREMENTAL CLOCK →
STELLAR TRACKER	SEXTANT →; CHRONOMETER →; RUBIDIUM TIME STANDARD →; GIMBALLED NARROW FOV STAR TRACKER → STRAPDOWN WIDE FOV STAR TRACKER →
TERMINAL SEEKER	VISUAL/OPTICAL AIDS →; LOW LIGHT LEVEL TV →; IR THERMAL →; MAGNETIC RADAR →; MILLIMETER WAVE →; ACOUSTICAL (AIR) →; SONAR (SEA) →
PROCESSOR	MANUAL → MECHANICAL → ANALOG → DIGITAL →; OPTICAL →

92M-121-17

1,000,000

100,000

10,000

1,000

100

10

3D RADIAL POSITIONING ERROR IN FEET

1950 1960 1970 1980 1990 2000 2010

UNAIDED STRAPDOWN INERTIAL NAVIGATION SYSTEM [1]

GIMBALLED TECHNOLOGY UNAIDED INERTIAL NAVIGATION SYSTEM

GIMBALLED ASTROINERTIAL NAVIGATION SYSTEM

STRAPDOWN STELLAR/INERTIAL SYSTEM

SATELLITE GLOBAL POSITIONING SYSTEM (GPS)

4 HOUR COVERAGE

3D 24 HOUR COVERAGE

PROJECTION OF NAVIGATION SYSTEM PERFORMANCE FOR GPS, ASTROINERTIAL, AND UNAIDED INERTIAL NAVIGATION SYSTEMS FOR 10-HOUR FLIGHT

[1] NOTE: THE STRAPDOWN SYSTEM ERROR PROJECTION ASSUMES A MILITARY VEHICLE WITH A MIX OF BENIGN AND SERPENTINE MOVEMENTS. FOR BENIGN ONLY THE STRAPDOWN APPROACHES THE GIMBAL SYSTEM ERROR.

92M-121-01

Figure 1. Startrackers Enhance Inertial Navigator Performance
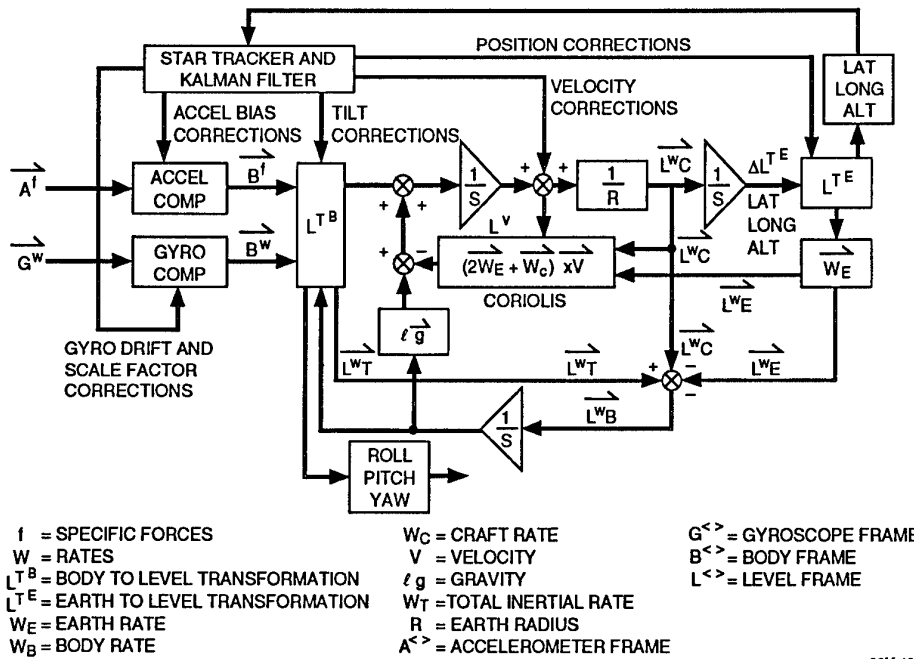
Figure 2. Strapdown Stellar Inertial System, Canonical Model
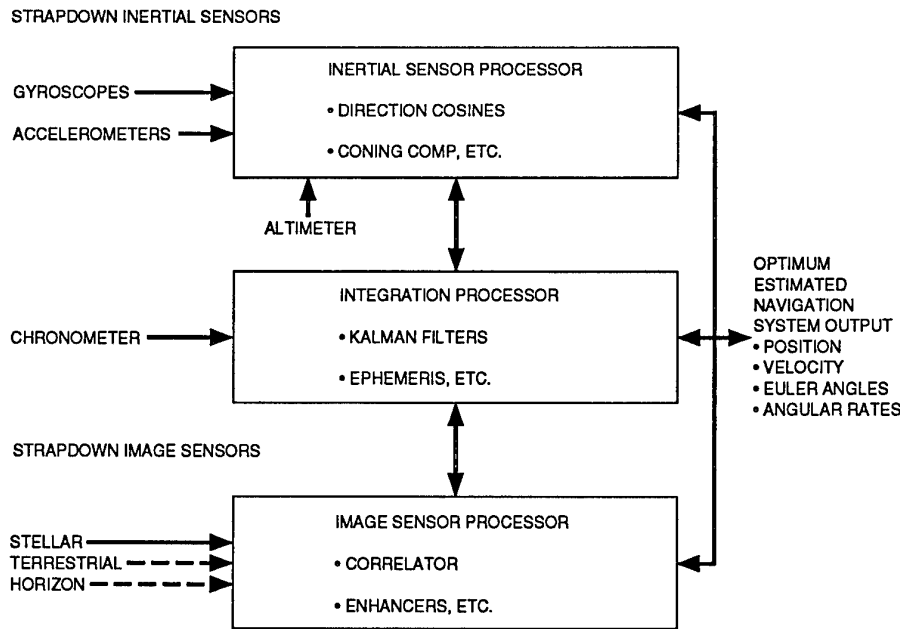
Figure 3. Simplex Strapdown Stellar Inertial System Model with Terrestrial and Horizon Image Options

of the vectors of position, velocity, and Euler angles, as well as gyroscopic errors, such as bias, scale factor, and misalignment, and observable accelerometer errors.

After the inertial stellar observations are made, every additional measurement continues to minimize the error vectors.

It should be pointed out that the astrotracker is similar to a marine sextant in that it looks at the star angle with respect to the horizontal level, but, unlike the sextant's observed horizon, the astroinertial system uses the inertial system's estimated vertical. This deviation leads to an inability to observe certain errors. Figure 4 is a heuristic error model of an astroinertial system. As can be seen, the
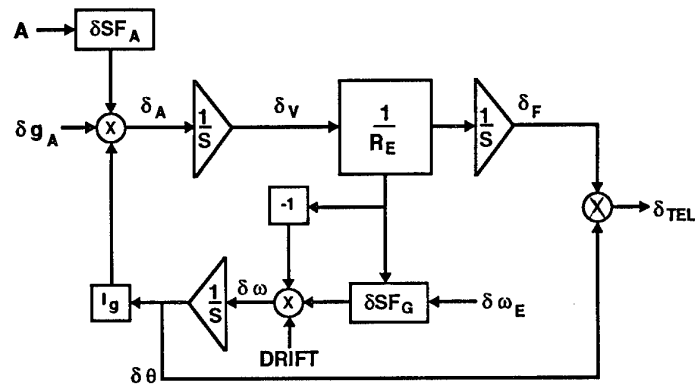
Note: LaPlacian Operator 1/S is used for an Integrator.

Where: I_g = Acceleration of Gravity $\delta\omega_E$ = Error In Earth Rate Compensation

δ_A = Acceleration Error δg_A = Error In Gravity Horizontal Component

δ_V = Velocity Error δSF_A = Accelerometer Scale Factor

δ_P = Position Error δSF_G = Gyroscope Scale Factor Error

$\delta\theta$ = Tilt Error δ_{TEL} = Telescope Angular Error

$\delta\omega$ = Angular Rate Error

92M-121-04

Figure 4. Heuristic Astroinertial Error Model (the astroinertial system cannot compensate for some acceleration errors)

startracker cannot distinguish between a tilt in the inertial system and an error in indicated position when both are equal and opposite. Thus, the tracker is unable to observe, and therefore cannot aid in damping, the INS's 84 min Schuler oscillations. Uncompensated gravity horizontal deflections have correlated tilt and position errors. Since this correlation produces a position error opposite to the tilt errors, the astrotracker is unable to sense (and therefore compensate for) these differences. Also for uncompensated gravity horizontal deflections, the astroinertial systems exhibit the identical error response as do the pure inertial navigators. The errors can be minimized by accurate gravity compensation that takes into account spherical harmonics, terrain data, and geological information. Horizon sensors and velocity augmentation reference systems can be used to minimize these errors, as shown in Figure 5. This figure shows a simulation of a strapdown astroinertial navigator aided by a velocity reference system after an 18 h flight. The velocity reference augmentation is used to damp the navigator's Schuler oscillations and thereby minimize the vertical error. When the vertical oscillations are damped, the stellar sightings are able to observe and minimize position errors. The less than 200 foot simulated errors shown in Figure 5 were made using a 1.5 arcsec astrotracker. When the simulations are run with a 0.5 arcsec stellar sensor, the Circular Error Probability (CEP), is only 60 ft.

Gyroscopic drift terms, on the other hand, are easily distinguished, since the one integration response to tilt is much faster than the three integration processes required for the error to go into position. Figure 6 is an example of how stellar augmentation reduces the effect of gyroscopic drift on navigation errors. It illustrates the

performances of four different Ring Laser Gyroscopes (RLGs) in a Strategic Penetrator scenario. Free inertial system navigation errors are compared with a stellar augmentation. As can be seen, the stellar augmentation reduces the errors significantly. By isolating a 40 cm RLG (i.e., the most accurate gyroscope at 0.00025 deg/√h), a graphical high-resolution comparison can be made both with and without stellar augmentation. Figure 7 shows that the error after 8 h of flight with the 40 cm RLGs is 9,000 ft unaided and only 600 ft with stellar augmentation (a 15 to 1 improvement factor).

As the flight extends past 8 h, the improvement factor increases further. Figure 8 compares 10, 20, 30, and 40 cm RLG stellar inertial performance. The stellar navigation performance just begins to degrade with 10 cm RLGs. It should be noted that the 40 cm gyroscopic random walk angular drift rate is 16 times lower than that of a 10 cm RLG or Resonant Fiber Optic Gyroscope/Interferometric Fiber Optic Gyroscope (RFOG/IFOG); yet even with this significant disparity in gyroscopic drift performance, astroinertial system accuracy is essentially the same. The astrotracker compensates for gyroscopic drift by monitoring the navigator's tilt and azimuth angles and by providing tilt, azimuth, and gyroscopic drift corrections. These curves illustrate one of the values of astroinertial systems: they minimize the system's performance dependence on ultra-precise, large, costly gyroscopes. It should be pointed out that the strapdown astroinertial simulation responses have higher navigation errors that the present class of high-precision gimballed astroinertial instruments. This difference is expected to disappear over the next three years as the strapdown acceleration inertial sensors (i.e., accelerometers) improve.
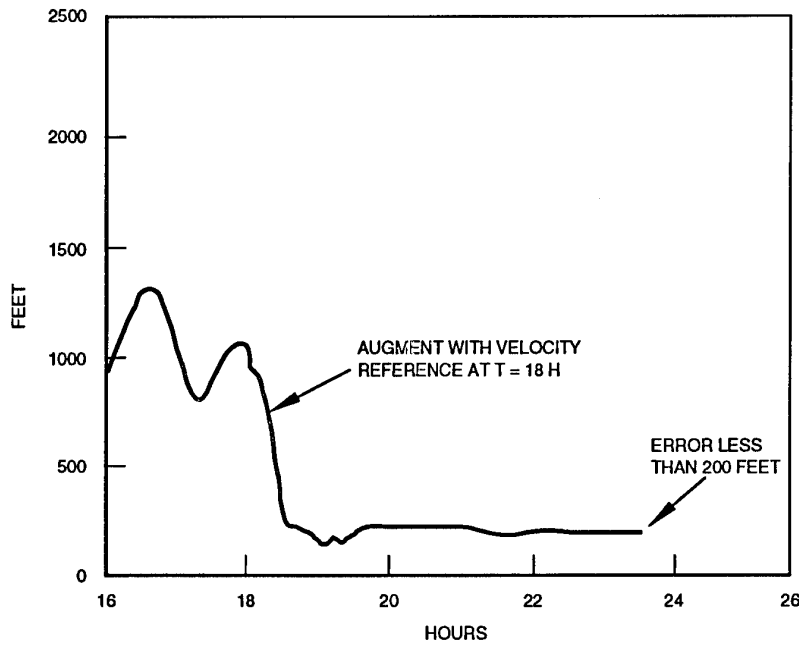
Figure 5. Simulated Results: Strategic Penetrator Position CEP of Inertial Navigator with Low-Rate Stellar and Velocity Reference System Augmentation
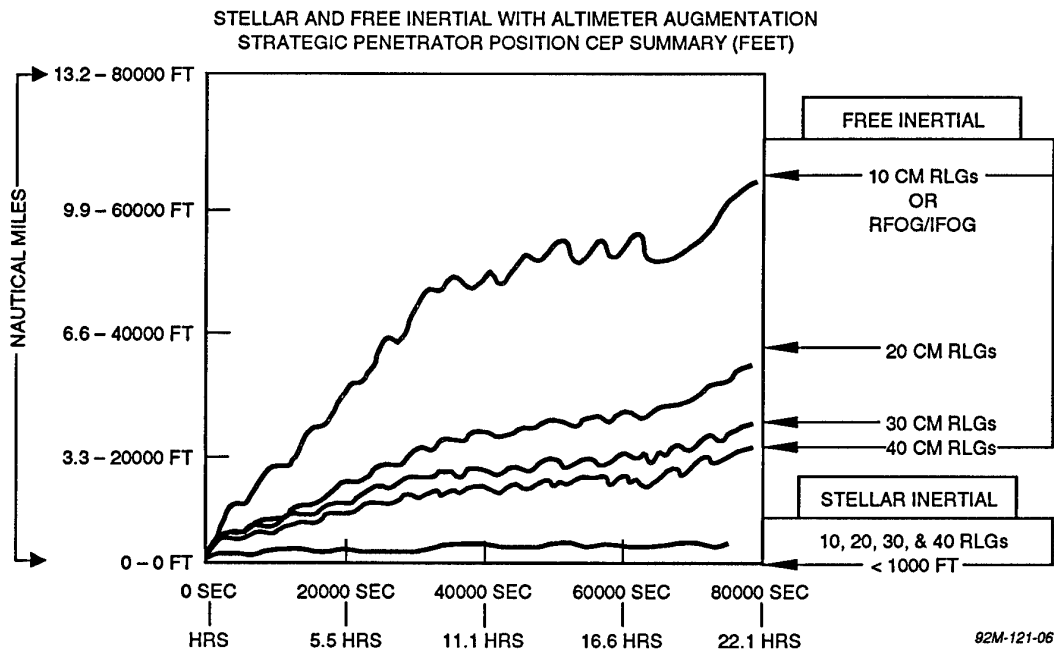
Figure 6. Simulation Results: Astrotracker Inertial Augmentation Essentially Eliminated the RLG as a Dominant Error Source (CEP vs Mission Time)
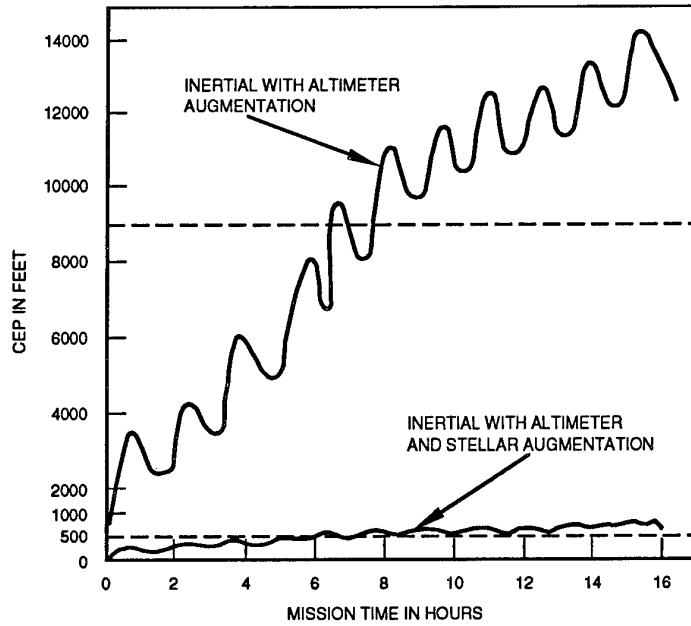
92M-121-07

Figure 7. Strapdown Stellar Inertial System Performance--Strategic Penetrator Mission with 40 cm RLGs
(CEP vs Mission Time)
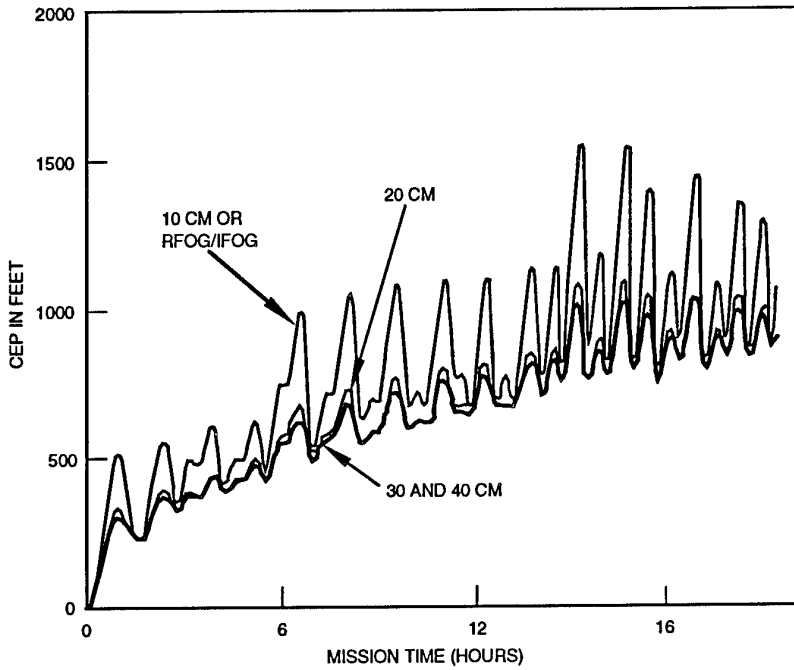
92M-121-08

Figure 8. Strategic Penetrator Strapdown Stellar Inertial System Performance: 10, 20, 30, and 40 cm RLGS or
RFOG/IFOG (CEP vs Mission Time)

Another major attribute of astroinertial systems is the ability to observe north directly. In inertial systems, the initial azimuth determination process is corrupted by the effective east gyroscopic drift. Furthermore, the azimuth error vector can be changed by effective azimuth gyroscopic drift. Startrackers can observe and correct these error vectors. This endows astroinertial systems with excellent response times for accurate performance, since they can correct these quantities in the sky. Stellar inertial systems also provide a high level of self-monitoring and self-healing. For example, a pure inertial system cannot observe an initial or updated longitude insertion error during flight, but a stellar inertial system can detect this potentially catastrophic problem. This is because the pure inertial navigator's dynamic response is transparent to longitude data, while the stellar inertial star ephemeris depends on

longitude. When the star detection probability is high, the stars should be in the Field of View (FOV). If the star is not in the FOV, the flight computer could trigger either a self-corrective action response or a mission safety abort mechanism. The choice of action can be programmed into the flight computer and depends on subsequent self-tests.

Thus, an astrotracker provides an excellent augmentation device for the inertial navigator. Furthermore, by the addition of a velocity reference (see Figure 5) or horizon sensor, SAIN can approach GPS/inertial performance (60 vs 45 ft) without the use of or dependence on man-made radiation devices. If GPS were available, it would be integrated into the navigation suite as another inertial navigator augmentation device. This augmented navigation suite provides a level of navigation position and velocity redundancy. Nevertheless, the mission's performance goals need not be jeopardized by a dependence on satellite data availability. Therefore, the astroinertial system yields the high-performance autonomous mission capability necessary for the 21st century.

3.0 STAR IMAGE CONSIDERATIONS

Image stabilization is a major problem with solid-state astrotrackers. Figure 9 shows how a vehicle's motion causes a star's image to traverse a fixed-to-vehicle Focal-Plane Array (FPA). Because of the vehicle's motion, a high-speed snapshot of the star's image must be taken to prevent blurring.

After the image snapshot is taken, the data in the image processor is transformed into stabilized coordinates with the aid of the inertial navigator's direction cosine matrices. The image data integration for signal-to-noise enhancement, as shown in Figure 10, is a complex process in which the shutter opening is operated at 100 to 1,000 μs to prevent star image blurring. This high-speed shuttering of data is necessary even for vehicle attitude rates as low as 1 deg/s.

4.0 STELLAR TRACKER PERFORMANCE PARAMETERS

The driving factor in the design of shipboard and airborne stellar trackers that augment inertial systems is their capability to accurately measure star positions in daylight at sea level. For simulated Hyper Velocity Vehicle (HVV) missions, the Signal-to-Noise Ratio (SNR) improves by a factor of 2.5 (one star magnitude) as the vehicle ascends from sea level to a 20,000 ft altitude. This is primarily due to the effects of the atmosphere. As the air density decreases, the sky darkens because of a decrease in light scattering. Figure 11 shows the sky background as a function of wavelength for sea level and at a 100,000 ft altitude. The spectral radiance is a hundred times brighter at sea level than at the 100,000 foot altitude. In addition, atmospheric molecular absorption degradation factors decrease at higher altitudes. Figure 12 shows the optical atmospheric transmittance as a function of wavelength and zenith angle. The atmospheric transmission is attenuated at wavelengths where there is O_2, O_3, and H_2O molecular absorption. For an astrotracker that utilizes an FPA as its photon detector, the SNR can be approximated by the following expression:
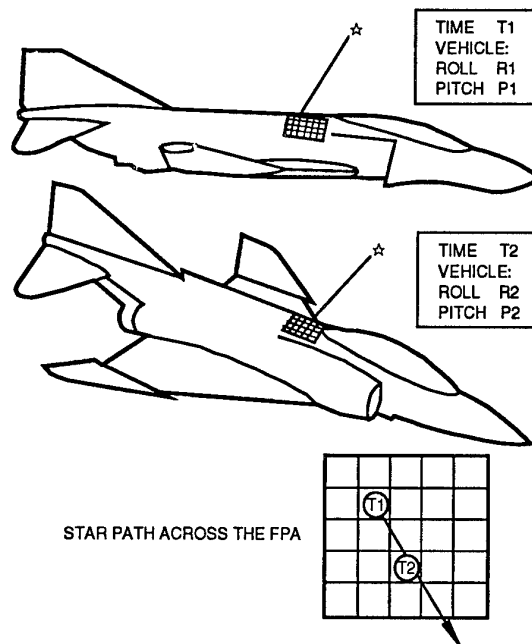
Figure 9. Image Trajectory Across an Unstabilized FPA Star Detector

194
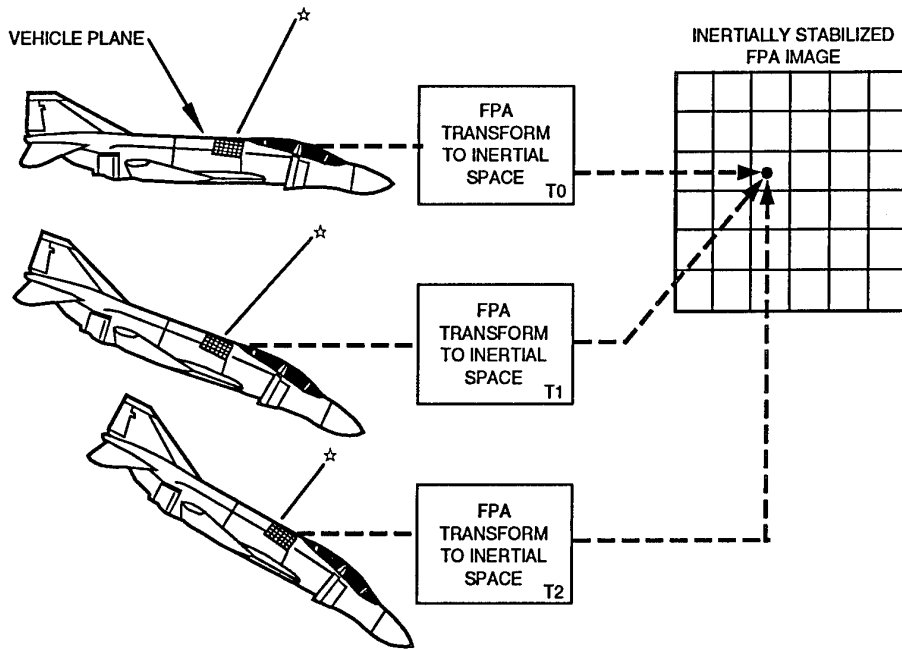
Figure 10. Strapdown FPA Charge Enhancement
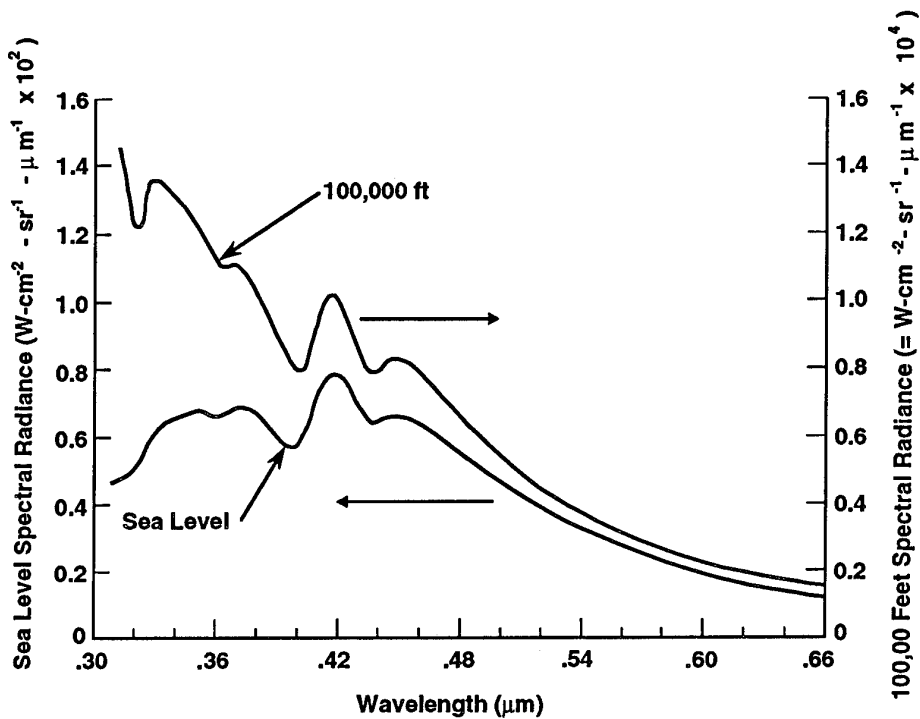
Figure 11. A Typical Sky Background at Sea Level and 100,000 ft (Sun Zenith Angle 53 deg, Observer Angle 58 deg, Sun-to-Observer Azimuth 90 deg)SEYRAY Electro-Optical System Analysis, page 139, Pub. "EO RESEARCH CO."(AFAL Report of NEMD)
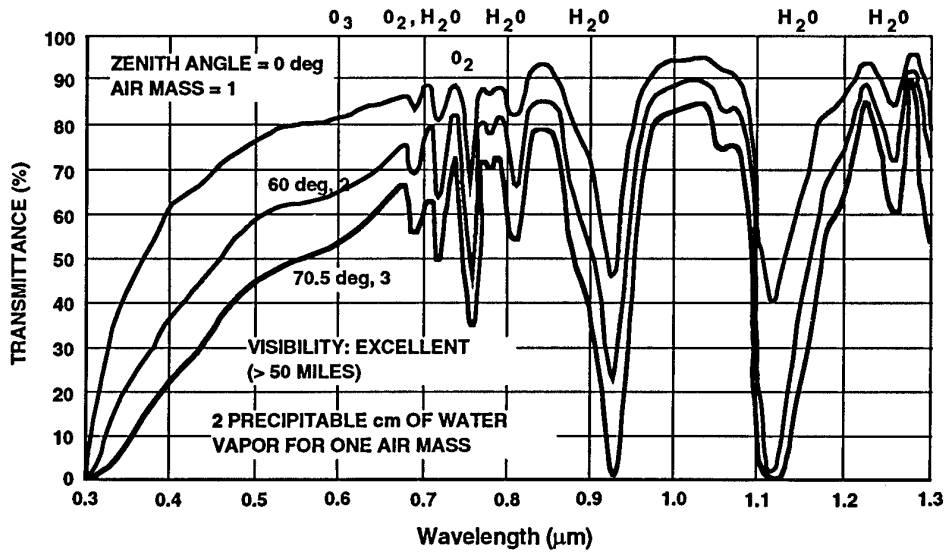
92M-121-12

Figure 12. Atmospheric Tranmission for Various Air Masses or, Equivalently, From
the Ground to Space at Various Angles from the Normal; One Air Mass
Corresponds to the Amount of Air Through Which a Beam Must
Pass When Going Straight Up to Space (Courtesy RCA)

Signal-to-Noise Ratio (SNR)

$$Q_{SP} = q \cdot \phi_S \cdot \Delta\lambda \cdot A_T \cdot A_a \cdot T_T \cdot \eta \cdot I_t \cdot N_f/N_S$$

where:

Q_{SP}	=	Star signal charges per pixel
q	=	Electronic charge
ϕ_S	=	Stellar spectral radiance
$\Delta\lambda$	=	Wavelength bandwidth of telescope system
A_T	=	Atmosphere transmission
A_a	=	Aperture area
T_T	=	Telescope transmission
η	=	Quantum efficiency
I_t	=	Integration time of a frame (stellar snapshot)
N_f	=	Number of frame snapshots utilized
N_S	=	Number of pixels containing the star image (star blur factor)

$$Q_{NP} = \sqrt{(Q_{SP} + Q_{BP} + Q_{DP})\, Q_{ES}}$$

where:

Q_{NP}	=	Star signal charges per pixel
Q_{SP}	=	Star photo charges per pixel
Q_{BP}	=	Background photo charges per pixel
Q_{DP}	=	Dark detector charges per pixel
Q_{ES}	=	Electronic system bandwidth noise coefficient

$$SNR = \frac{Q_{SP}}{\sqrt{(Q_{SP} + Q_{BP} + Q_{DP})\, Q_{ES}}}$$

$$Q_{BP} = \frac{q \cdot \phi_B \cdot \Delta\lambda \cdot A_a \cdot T_T \cdot FOV \cdot \eta \cdot I_T \cdot N_f}{N \cdot O_F}$$

where:

Q_{BP}	=	Background photo charges per pixel
ϕ_B	=	Sky background spectral radiance
FOV	=	Field of view
N	=	Total number of pixels in sensor array
O_F	=	Sky background attenuation of optical filter

For sea level daytime tracking in a non-nuclear
event environment, the SNR equation simplifies to:

$$Q_{BP} >> Q_{SP} + Q_{DP}$$

$$SNR \approx \frac{Q_{SP}}{\sqrt{Q_{BP} \cdot Q_{ES}}}$$

$$Q_{DP} = q \cdot ni/\tau \cdot V_B \cdot I_t \cdot N_f$$

where:

Q_{DP}	=	Detector dark current charges per pixel
ni	=	Detector intrinsic carrier concentration
τ	=	Detector dark current charge generation time
V_B	=	Detector charge generation bucket volume

$$SNR \approx \frac{\phi_S \cdot A_T}{N_S} \cdot \sqrt{\frac{q \cdot \Delta\lambda \cdot A_a \cdot T_T \cdot \eta \cdot I_t \cdot N_f \cdot N \cdot O_F}{\phi B \cdot FOV \cdot Q_{ES}}}$$

Unfortunately, the tracker's SNR provides only a part of the insight necessary for the design of a stellar inertial system. Knowledge of the detected stellar density is also of major importance. For example, the number of stars brighter than a given magnitude, M_r, is a function of the spectral contents of the stars. Stars radiate in a similar fashion to a blackbody at different maximum temperatures based on their age and composition. Table 2 is a comparison of the stellar densities for the R band (red) stars and that of B band (blue) stars.

A few simple terms dominate the astrotracker design thought process. The SNR is proportional to the square root of the tracker's aperture and the star photon integration time (stellar exposure time). The reason for only a square-root improvement in the SNR as a function of aperture is that the expansion of the aperture not only increases the star signal photon collection but also has an associated increase in the collection of sky background noise flux. Therefore, a 6.25 times increase in aperture area is required in order to observe a single magnitude increase (i.e., 2.5) in dimmer stars. Similarly, the 6.25 increase factor holds for the stellar exposure time, since extension of the integration increases the collection of both the star photons and the sky background noise flux. In the opposite direction, a wide FOV detracts from the stellar observation by allowing more sky brightness background noise flux into the optical system. For example, when the FOV is very small, as when looking at the sky from the bottom of a deep mine shaft (illustrated in Figure 13), the unaided eye can observe stars in the daytime. The sky brightness noise increases as the inverse of the square root of the FOV. In order to achieve satisfactory performance, a balance is struck between the benefit of a wide FOV increasing the number of bright stars observed in the tracker at a given time vs. the deteriorative effects of the sky brightness noise. Figure 14 summarizes some of these considerations for a silicon FPA and a 3 deg FOV Optical Wide-Angle Lens Startracker (OWLS). It is a plot of star magnitude capability versus altitude. A space or satellite startracker with only a 0.72 in.^{2} aperture and a 1 ms exposure time has the same capability as the much larger sea-level tracker with a 36 in.^{2} aperture utilizing a 100 ms exposure time. Figure 15 shows a system size comparison for both gimballed and strapdown astroinertial navigators that have the same daytime sea-level tracking capability. The dramatic disparity shown in Figures 14 and 15 illustrates how the tracker technology and application dictate the physical startracker design.

Figure 16 shows the probability of seeing the sky from sea level. Above 13.7 km (45,000 ft) in altitude, the probability of having an unobstructed view of the stars is essentially 100 percent. Although at an altitude of 13.7 km, an unaided observer's view of the sky is unobstructed by cloud cover, the sky background light will prevent observing the stars in daylight. Yet at night, the unaided observer's view of the stars from that altitude is spectacular. Above 13.7 km, the daytime sky background, not in the direct vicinity of the sun, grows progressively darker with increases in altitude until it essentially turns black and is indistinguishable from the nighttime sky. Thus, at altitudes above 60 miles, a startracker the size of an eyeball provides sufficient signal to reset an inertial guidance platform.

The quality or pure inertial performance of the gyroscopes needs to be high when operating for sustained periods under cloud cover. Under these conditions the gyroscopes may only get stellar updates a few times an hour. On the other hand, when operating above 13.7 km with a high-speed startracker, the system navigation performance will be outstanding even with low cost, modest performance gyroscopes. Furthermore, when operating at zero velocity on the ground, while tracking stars, the stellar inertial system can calibrate many of the inertial instrument errors including the accelerometer bias. This feature reduces some at the accelerometer long-term error sources.

TABLE 2. COMPARISON OF STELLAR DENSITIES FOR THE R & B BAND STELLAR WAVELENGTHS

Number of Stars Equal to or Brighter than Magnitude M_r		
M_r	B Band 0.4 ± 0.049μ	R Band 0.70 ± 0.11μ
	Density per Deg^2	
1	0.0002	0.0006
2	0.0006	0.0022
3	0.0022	0.0077
4	0.0075	0.0250
5	0.0240	0.0750
6	0.0710	0.2300

STARLIGHT

SUNLIGHT

SCATTERED
LIGHT

APERTURE

FOV = $\dfrac{\text{Aperture}}{\text{Depth}}$

MINE SHAFT
DEPTH

SUNLIGHT

STARLIGHT

SCATTERED SUNLIGHT

*Note: The starlight is
also scattered but the
Sun is a billion times
(25 magnitudes) brighter
than the brightest star
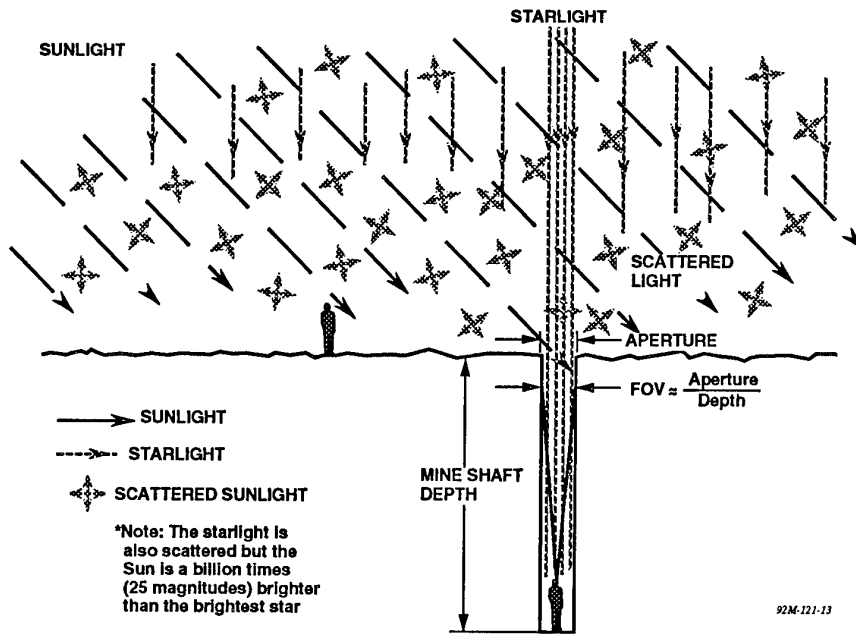
92M-121-13

Figure 13. Miner Observing Starlight in the Daytime

STAR MAGNITUDE CAPABILITY

t_I = INTEGRATION TIME

t_I = 100ms

10ms

1ms

100ms

10ms

1ms

ALTITUDE IN MILES

A = APERTURE AREA

TERRESTRIAL OWLS
A = 36 sq in.

SPACE
Mini-OWLS
A = 0.72 sq in.

TERRESTRIAL
OWLS
IN DAYLIGHT

SPACE Mini-OWLS
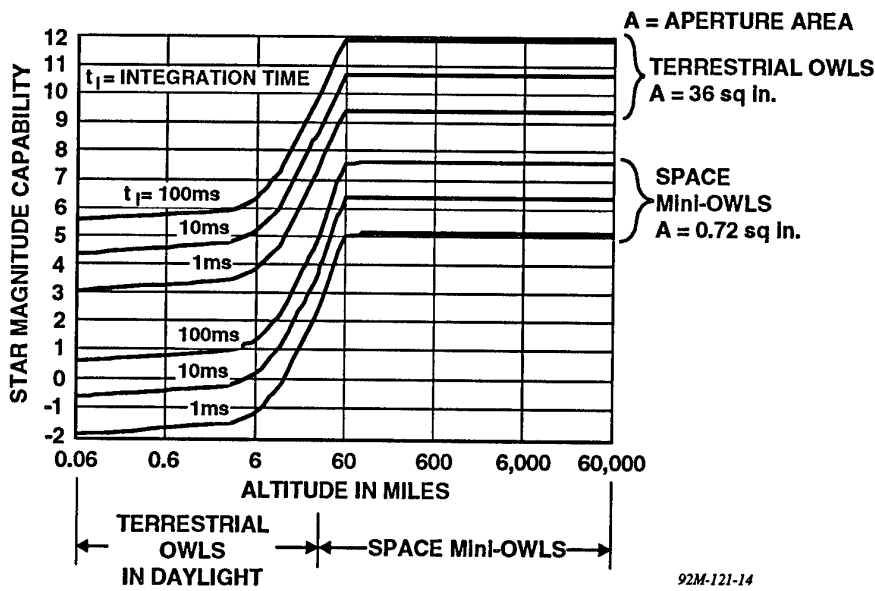
92M-121-14

**Figure 14. Startracker Sensed Magnitude vs Altitude for a
Silicon Detector and a 3 Deg FOV**

1992 - Gimballed

1996 - Strapdown Including Power
Supply and Processor

4:1
Reduction

$7\frac{1}{2}$ in. W x $7\frac{1}{2}$ in. H x 20 in. D
45 lb

Satellite and Spacecraft System

120:1
Reduction

2 in. H x 1 in. W x 4 in. H
250 Grams

14 in. W x 15 in. D x 20 in. H
Including Power Supply & Processor (165 lb)
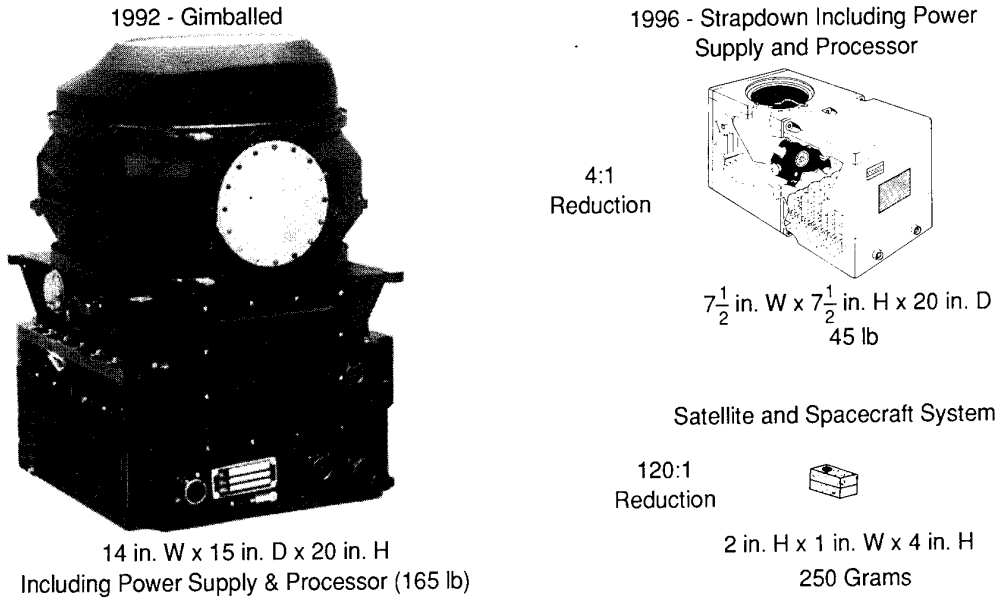
92M-121-15

Figure 15. Comparison of Gimballed and Strapdown Stellar Inertial Navigators for
Aircraft Along with a Spacecraft Stellar Inertial System

ALTITUDE
(FT x 10^3)

ALTITUDE
(KM)

O = GROUND

H = HORIZON

S = SKY

PERCENT PROBABILITY
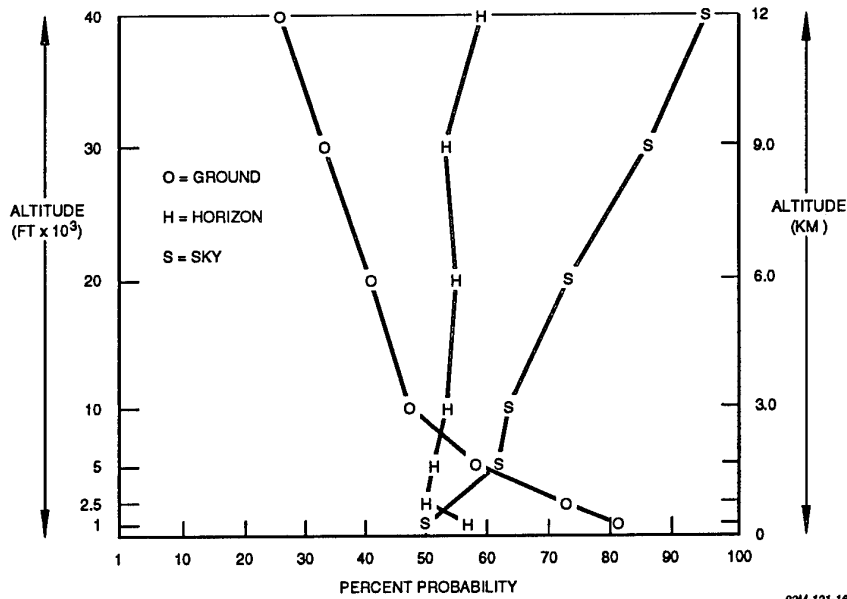
92M-121-16

Figure 16. Probability of Clear Lines of Sight Over the Northern Hemisphere
for All Seasons Combined (72,000 Observations)

5.0 CONCLUSIONS

Stellar inertial capability is a precision autonomous navigation strategic asset. The technology encompasses gravity compensation techniques, stellar catalog, telescope design, and inertial navigation. It provides the user with a unique surgical strike advantage in an area where, for a host of reasons, radio navigation has been eliminated or temporarily denied. The SAIN enables reducing the cost, size, and reliability penalty for autonomous precision strike to an affordable option in future vehicles. When sensor fusion is taken into account, the low-cost stellar option provides in-air alignment regardless of the vehicle's flight path or speed, and a level of navigation redundancy to enhance mission success. Thus, as in the past, modern man should depend on the stars for navigation.

BIBLIOGRAPHY

1. Joseph F. Caligiuri, "The Navigation System for the Polaris Submarine," Navigation Journal of the Institute of Navigation, p. 3, Spring 1960.

2. Vern A. Blumhagen, "Stellar Inertial Navigation Applied to Cruise Vehicles," IEEE Transactions of Aerospace and Navigation Electronics, p. 235, September 1963.

3. Kenton L. Bachman, Ronald Dennis, and Seymour Levine, "Strapdown Astro-Inertial Navigation (SAIN) – An Idea That Has Come of Age," Central Inertial Guidance Test Facility (CIGTF) – Fourteenth Biennial Guidance Test Symposium, p. 181, October 1989.

4. Kenton L. Bachman, Ronald Dennis, and Seymour Levine, "Strapdown Astro-Inertial Navigation Utilizing the Optical Wide-angle Lens Startracker," Navigation, Journal of the Institute of Navigation, p. 347, Winter, 1990-91.

5. Fredrick R. Nadeau and Seymour Levine, "Strapdown Astro-Inertial Navigation as a Secure Method of Satisfying the Twenty-first Century Precision Mission Requirements," Central Inertial Guidance Test Facility (CIGTF), Fifteenth Biennial Guidance Test Symposium, p. 19, September 1991.

6. Major Rick Miller, Fredrick R. Nadeau, and Seymour Levine, "Mini-OWLS in Autonomous Stellar Angular Reference Systems (STARS) for Hypervelocity Strategic Systems and Space-Based Interceptors," IEEE Aerospace and Electronics Systems Society, Plans Symposium, March 1992.

Magnetic Heading References

by
Donald L. Moore
Avionics Systems Engineering

Smiths Industries
Aerospace & Defense Systems Inc.
Grand Rapids Division
4141 Eastern Avenue, MS 240
Grand Rapids, MI 49518
USA

Summary

This paper addresses the most common system for providing an aircraft heading reference: the magnetic heading reference system. It begins briefly explaining the importance of a magnetic heading reference (section 1). It then addresses the fundamental characteristics of the earth's magnetic field and explains the concept of "magnetic variation" (section 2). It discusses two of the more common styles of magnetic heading sensors along with their inherent errors (section 3). After explaining how these different compasses operate, the paper explains the different types of heading errors that occur. Section 4 addresses the in-flight errors, while section 5 addresses the magnetic disturbances caused by the aircraft itself. Section 6 discusses calibration techniques that provide for the correction against these magnetic disturbance induced errors. Finally, this paper concludes with a brief discussion of future trends (section 7).

1 Introduction

Many pieces of information must be available to a navigator (either a person or an on-board computer) to reliably fly an aircraft from point A to point B. Obviously, one very useful piece of information is the correct direction of flight to eventually arrive at point B. But, before turning the aircraft in the desired direction, the navigator must know the current heading of the aircraft. The magnetic heading reference system provides this critical piece of information.

2 The Earth's Magnetic Field

A heading is based on a coordinate system that uses a pre-defined direction as a reference. The angle created between the vectors pointing in the reference direction and the desired direction is known as the heading. To facilitate the job of air traffic controllers, all aircraft navigators should use the same reference direction. The earth's magnetic field can provide that reference.

The properties of the earth's magnetic field are similar to that of a powerful bar magnetic located at the earth's center (see figure 1). The field consists of horizontal and vertical components of the total intensity of the magnetic field. When the field is parallel to the earth (near the equator), the horizontal intensity is the same as the total intensity of the field. There is no vertical component to the field at the equator. Likewise, at the magnetic poles of the earth, the vertical intensity is the same as the total intensity of the field. There is no horizontal component to the field
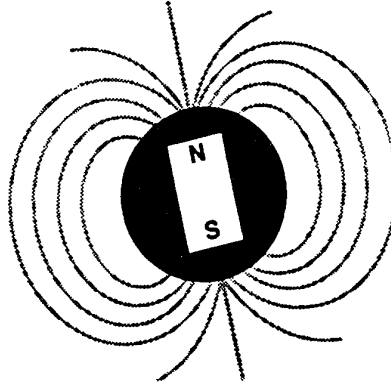
Figure 1: Earth's Magnetic Field

at the poles. All points on the globe between the equator and the poles have some magnitude of a horizontal and vertical component.

As previously stated, the earth's magnetic field is only "similar" to a bar magnet. The earth's crust does contain materials that affect the earth's magnetic field. There is a diurnal change and a slow drift over the years [1]. The reasons for these anomalies are not within the scope of this text but it is important to note that the crust's effect is fairly stable and well known. The diurnal change is only $\pm 0.1°$ but in certain regions, such as Alaska, variations of 3 to 4 degrees per minute have been observed. Also, during a magnetic storm, variations up to $12.8°$ have been observed over a 19 hour period. The slow drift, which in on the order of $\pm 0.1°$ per year, is cumulative over a number of years but eventually reverses and drifts in the other direction. Reference [2] provides additional information pertaining to the earth's magnetic field.

The poles of the earth's magnetic field are referred to as the north and south "magnetic" poles. As previously stated, to determine a heading, an appropriate heading reference must be chosen. The north magnetic pole provides this reference in a magnetic heading reference system. That is, all derived headings are based on magnetic north, which is defined as $0°$.
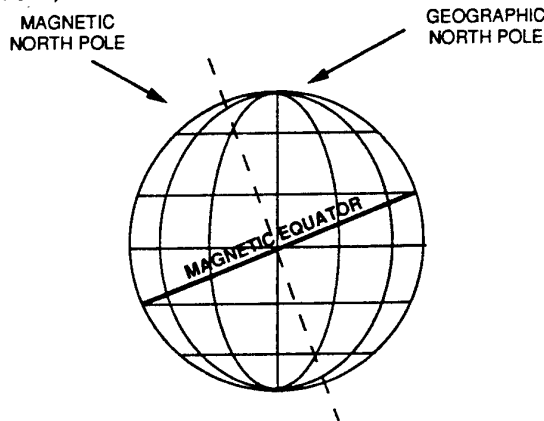
Figure 2: Magnetic and Geographic Poles

The earth's position-locating "grid" system (latitude and longitude) is based on the earth's rotational or polar axis. The poles of the earth's rotational axis are referred to as the north and south "geographic" poles. Unfortunately, the magnetic poles and the geographic poles do not coincide (see figure 2). In fact, they are over 1000 miles apart.

When plotting a course from point A to point B, the navigator (which may be an on-board computer) determines the desired heading based on the earth's grid system. This heading references the north geographic pole, and is referred to as the "true" heading (as opposed to "magnetic" heading). Course plotting is usually performed using true heading, and course navigating is performed using magnetic heading.

The difference between the true heading (which references the geographic north pole) and the magnetic heading (which references the magnetic north pole) is known as magnetic variation. Aeronautical charts are available that provide the easterly or westerly magnetic variation for any point on the earth. These charts are known as "Magnetic Variation" charts and are published by the Defense Mapping Agency Hydrographic/Topographic Center [3]. It closes the gap between course plotting to true heading and navigating to magnetic heading. For example, if the plotted course is 210° true heading, and the chart indicates a westerly variation of 5°, the magnetic course to fly would be 215°.

For systems without magnetic sensors, an algorithm exists which creates a MAG VAR look-up table. The magnetic variation algorithm (GEOMAG) is based on a 12th order spherical harmonic expansion of the Earth's magnetic field, the coefficients of which comprise the World Magnetic Model (WMM). These coefficients are produced jointly by the U.S. Naval Oceanographic Office and the British Geological Survey and are distributed by the U.S. Defense Mapping Agency and the British Hydrographic Office. The World Magnetic Models are usually produced at 5-year intervals and are composed of two parts: a main field model, which describes the Earth's magnetic field at some base epoch, and a secular variation model, which accounts for the slow temporal variations in the main geomagnetic field from the base epoch to a maximum of 5 years beyond the base epoch. For example, the base epoch of the WMM-90 magnetic field model is 1990.0. This Model is therefore considered valid between 1990.0 and 1995.0 and will subsequently be replaced at 1995.0 by the WMM-95 magnetic field model.

3 Magnetic Sensors

The main function of a magnetic heading reference sensor is to provide the actual magnetic heading of the aircraft. Sensors capable of providing this data are often referred to as "magnetometers". Magnetometers must be capable of detecting the direction of the horizontal component of the earth's magnetic field. From this, they can derive the actual aircraft heading. Two of the most common methods for determining this are the "simple magnetic compass" and the "flux-gate style compass".

3.1 The Simple Magnetic Compass

The simple magnetic compass, better known as the "whiskey" compass, is found in most aircraft today. It is used as the sole heading reference in some small aircraft and as a backup heading reference in larger aircraft with an advanced cockpit design.

The "whiskey" compass consists of a disk (also know as a card) attached to a buoyant material, with two long magnets mounted underneath. The card contains markings corresponding to all points of the compass. Between the magnets is a pivot point allowing the disk to rotate upon a bearing thus allowing the two magnets to freely align themselves with the magnetic north pole. This "floating" assembly is then encased in a chamber that is filled with a liquid such as alcohol or white kerosene (hence the name "whiskey" compass). The liquid inhibits the erratic swing of the compass due to turbulence or abrupt course changes [4].

Some advantages in using the whiskey compass are:

1. They are relatively inexpensive.

2. They do not require electrical power to operate.

3. They do not require inputs from other on-board sensors.

4. They are very durable and reliable.

Some of the disadvantages of a whiskey compass are:

1. The aircraft heading display is reliable only when the aircraft is in less than a 18° bank [4] due to the physical limitation of most units.

2. The liquid causes a noticeable damping effect during turns.

3. The heading data from the compass cannot be electrically transferred to other aircraft computers that may need magnetic heading information.

3.2 The Flux-gate Compass

Most, if not all, of the technically advanced aircraft today use a flux-gate style sensor as the primary source of magnetic heading detection. There are several variations of this type of sensor, but all are based on the electrical principle of "flux-gating".

Flux-gating is a very complex electrical concept. Its detailed explanation is outside the scope of this document. For a detailed description of flux-gating, refer to [5]. A top-level description, however, is provided to give the reader a basic understanding of how the flux-gate magnetometer detects the aircraft's current heading.

The flux-gate style magnetometer consists of a highly permeable material (the core) such that the earth's magnetic flux lines are drawn into one end of the core and released from the other end of the core (see figure 3). A wire is wrapped around this core material, and an alternating current is then applied to the wire. This wire is called the "primary" coil. A second wire is also wrapped around the core and is used to sense the voltage peaks generated from the AC current. This wire is often referred to as the "sense" coil or "secondary" coil.
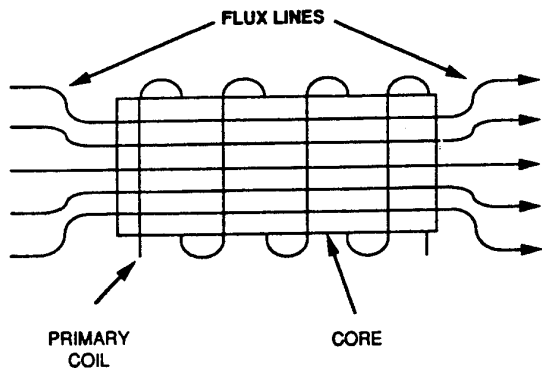
Figure 3: Flux Lines - Unsaturated Core

When the current is applied to the primary wire, there is a point at which the core becomes saturated. When saturated, the core is no longer permeable to the earth's magnetic flux lines (the earth's magnetic field is no longer drawn into one end of the core and out the other end—see figure 4). The key to the entire "flux-gate" design is that the amount of current needed to saturate the core is proportional to the horizontal component of the earth's magnetic field. In other words, more current is required to saturate the core when the detected horizontal component of the earth's magnetic field is strong (heading north or south: parallel to the horizontal component). Likewise, less current in needed when the detected horizontal component of the earth's magnetic field is weak (heading east or west: perpendicular to the horizontal component).
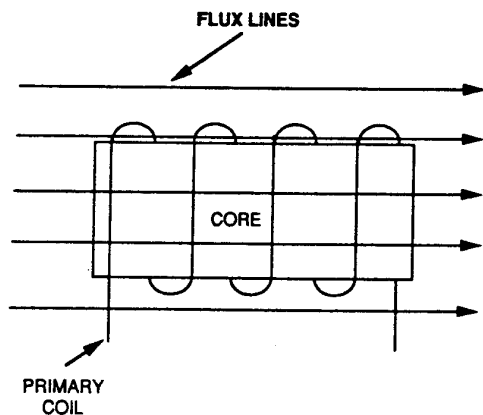
Figure 4: Flux Lines - Saturated Core

The amount of current needed to saturate the core can be detected by the output of the secondary coil (the sense coil). The angle of the core (which is the same as the aircraft's) relative to the earth's magnetic field can be derived by reading the peak voltages on this secondary coil (see figure 5 as taken from [6]).

In the flux-gating principle style just described, two possible headings exist for every current needed to saturate the core. In the example given in figure 5, the current of 0.707 amps implies that the angle of the core relative to the earth's magnetic field is either 45° or 315°. To resolve this conflict, another core (with a primary and secondary coil) is offset such that the ambiguity can be resolved. The number of additional cores, the physical placement of the additional core(s), the type of material used as the core, and the shape of the core is where the flux-gate magne-
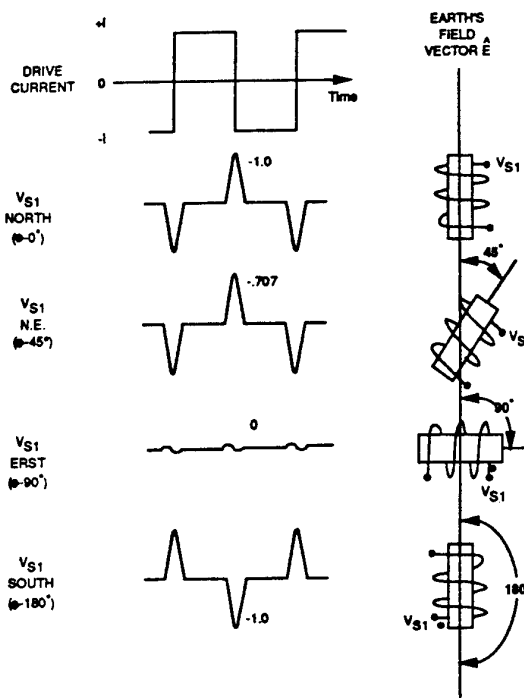
Figure 5: Peak Voltages in Secondary Coil

tometer becomes divided into types.

Recall that the sensed heading from a flux-gate magnetometer is a function of the horizontal component of the earth's magnetic field. If the aircraft is banked, the sensed horizontal component of the field is not of the same magnitude as when the aircraft is flying in a level attitude. This is due to the coupling of the vertical field component and results in an error in the calculation of the aircraft's heading. To correct this error, the compass keeps the "cores" in the horizontal position in unaccelerated flight by either placing the unit in a "gimbal", or by setting it in a fluid similar to that used in the "whiskey" compass. The "fluid" approach poses the same damping problems as the "whiskey" compass. Magnetometers that are "gimbaled" or set in fluid are referred to as pendulous magnetometers.

4 In-Flight Errors

Two of the most common in-flight errors are the "northerly turning" and "acceleration/deceleration" errors, and these are attributed to what's known as "magnetic dip". As explained earlier, the earth's magnetic field has both a horizontal and vertical component. The earth's "magnetic dip angle" (or inclination) is the angle between the earth's total magnetic field vector and the horizontal. At the magnetic equator this angle is 0°, and at the north and south magnetic poles it is 90°. When the dip angle is large, the ratio of the earth's vertical to horizontal field strength is large. Tipping the magnetic heading sensing element causes a component of the vertical vector to be coupled into the plane of the sensing element. This component adds to the sensed horizontal vector, resulting in a heading error.

The large dip angle encountered near the magnetic poles renders the typical magnetic heading system useless. For this reason, the use of the magnetic heading system is usually limited from 70°N to 60°S.

4.1 Northerly Turning Error

When in a coordinated turn, the centrifugal force causes the gimbaled or pendulous magnetometer to swing in a direction parallel with the floor of the aircraft as opposed to remaining parallel with the horizontal plane of the earth. This deviation from the horizontal introduces the vertical component of the magnetic field which, as stated earlier, results in an error in the calculation of the aircraft's heading (see figure 6 as taken from [7]).

As seen in figure 6, the error is at its greatest when the perpendicular to the horizontal field component (ϕH_v) is greatest. Since a turn direction from a northerly or southerly direction causes this "perpendicular tip", the er-
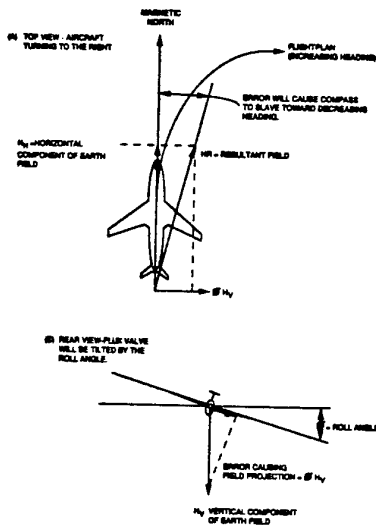
Figure 6: Generation of North Turning Error

ror is often called a northerly turning error. While in the turn, the error causes the heading indicator to either show that the aircraft is turning in an opposite direction or that it is turning at a greater rate than the actual turn rate. The effect on the heading indicator is a function of the hemisphere in which the aircraft is being flown, and whether the turn is away from a northerly or southerly direction.

Most magnetometers are designed to correct for this error by utilizing the sensed roll angle from a vertical gyroscope. Using this roll information, the compass knows and compensates for the vertical component of the magnetic field being introduced.

4.2 Acceleration/Deceleration Errors

When an aircraft accelerates or decelerates, the gimbaled or pendulous compass tilts from the horizontal. This deviation from the horizontal plane introduces the vertical component of the magnetic field, which causes an error in the calculation of the aircraft's heading.

The geometry behind this error is the same as explained in the "northerly turning error" section. That is, the error is at its greatest when the sensor tips in a direction such that the perpendicular to the horizontal component of the field is greatest. An acceleration/deceleration while flying in an easterly or westerly direction causes this "perpendicular tip". This causes the heading indicator to show that the aircraft is in a turn, even though it is in a straight and level attitude.

Most magnetometers are designed to correct for this error by utilizing the sensed acceleration/deceleration from an accelerometer. Using this information, the magnetometer knows and compensates for the vertical component of the magnetic field being introduced.

5 Aircraft Magnetic Disturbances

Anything that affects the direction of the horizontal component of the earth's magnetic field at the magnetic heading sensor is a disturbance. The magnetometer's housing itself may cause a disturbance to the field. There are four major sources of magnetic interferences associated with an aircraft. The magnetometer is therefore usually located in an area which suffers the least influence from these sources of interference (e.g., the wing tips or the tail of the aircraft).

5.1 Hard Iron Disturbances

The first major source of magnetic interference is the permanent magnetism of various ferromagnetic structural parts from the aircraft, or the cargo it contains. This type of field turns with the vehicle, thereby changing its relation to the earth's field vector and causing a change in the magnetic field surrounding the sensor. These disturbances are usually referred to as "hard iron" disturbances. For one revolution of heading through a 360° turn, the hard iron generates a heading error with one cycle of variation (see figure 7). Hard iron disturbances are therefore said to induce a "one-cycle" error.
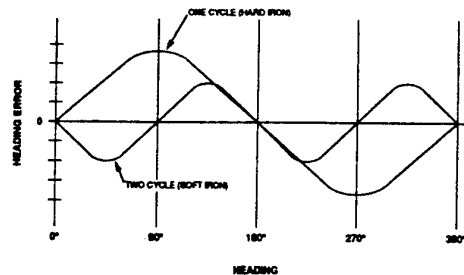
Figure 7: Hard/Soft Iron Errors

5.2 Soft Iron Disturbances

A second major source of magnetic interference, produced by the aircraft, is the induced magnetic fields created in the vehicle's ferromagnetic structures by the earth's magnetic field. These fields are generally associated with soft iron parts, although they also occur, to a limited extent, in hard iron parts. Unlike the permanent field, the induced field polarity and magnitude are determined by the relative direction and magnitude of the earth's field. These disturbances are generally referred to as "soft iron" disturbances and are more prominent in smaller aircraft [1]. For one revolution of heading through a 360° turn, the soft iron generates a heading error with two cycles of variation (see figure 7). Soft iron disturbances are therefore said to induce a "two-cycle" error.

5.3 Eddy Current Disturbances

The third major source of interference is that produced by eddy-current magnetic fields. These occur in all air-

craft skin, ribs, frames, and other structural units without respect to the magnetic materials from which they are constructed. The eddy currents require only the existence of electrical conductivity and aircraft maneuvers, and are generated in the same fashion as the currents produced in a coil of wire rotating in a magnetic field. An electrical conducting path is formed around the outside edge of the aircraft's conducting sheets or closed-loop structures. As the aircraft maneuvers in the magnetic field of the earth, electrical currents are produced in different metallic parts proportional to time changes of the fluxes of the earth's magnetic field through these parts. The flow of current creates a magnetic field in a direction perpendicular to the plane of the sheet or conducting loop.

5.4 DC Conductor Disturbance

The fourth major source of interference is that produced by conductors or cables carrying DC currents. When current is passed through a conductor, lines of flux are generated in circles concentric with the conductor and in a plane perpendicular to it. These types of disturbances are two-cycle in nature. They can be minimized by using a twisted pair of wires to counteract the disturbance, as opposed to using one wire with the aircraft structure as a ground return.

5.5 Index Errors

Index error is an error which is due to misalignment of the flux valve or compass system synchros. This error has the same magnitude at all headings.

In the past, index error has been corrected by rotating the flux valve in its mount until the average error on the cardinal headings was zero. This technique usually left the flux valve misaligned with respect to the aircraft centerline. Present day techniques call for accurately aligning the flux valve with the aircraft centerline. The residual system index error is then removed by providing an electrical differential between the flux valve and its control transformer, adjustment of which permits compensation of index error to less than 0.1°.

6 Calibration Techniques

The aircraft magnetic disturbances, as defined in the previous section, affect the detected direction of the horizontal component of the earth's magnetic field. To account for these disturbances and prevent an induced error, the compass must be "calibrated". Two of the more common methods used to calibrate the magnetic heading reference sensor are "manual swinging" and "electrical swinging".

These two methods provide the actual error caused by the aircraft itself. The actual compensation of the errors can be performed by:

1. Installing degaussing coils or permanent magnets at the location of the sensor to counteract the aircraft field.

2. Installing soft iron pieces in a position that will compensate the soft iron (2-cycle) errors.

3. Compensating for the soft iron (2-cycle) errors by adding appropriate impedances in series with the

magnetic sensor transmission wires.

4. Digitally compensating for the error in an aircraft computer.

6.1 Manual Swinging

The "manual swinging" method of calibration is normally performed on the ground by physically swinging or turning the aircraft 360°, and, at different points on the compass, measuring the error between the actual heading and the heading as detected by the sensor.

When performed on the ground, the swing must be accomplished on flat terrain in an area free from magnetic field disturbances such as nearby buildings or aircraft. Aircraft swinging ground sites are available that meet these criteria and provide a surveyed compass rose for an actual heading reference.

If an on-board gyro-based inertial heading reference is available (e.g., an Inertial Navigation System), then it can be used as the heading reference in place of the surveyed points on the compass rose. This approach allows the manual swing to be performed either on the ground (still in a magnetic disturbance free environment) or in-flight. It also facilitates automation of the data collection and processing functions, thereby simplifying the calibration procedure and greatly shortening the time required to accomplish it. An airborne swing requires the aircraft to fly several straight and level, unaccelerated flight legs in different directions. The more flight legs used, the better the accuracy in the error calculation.

6.2 Electrical Swinging

The "electrical swinging" method of calibration is performed on flux-gate style compasses by applying a controlled DC current into the flux valve to simulate a manual 360° swing of the aircraft [9]. In essence, this simulates the earth's magnetic field rotating around the aircraft. As in "manual swinging", this method of calibration requires that the measurements be taken on a flat terrain and in an area free from external magnetic field disturbances.

Before the electrical swing can be performed, the characteristics of the flux-gate sensor must be known. That is, the amount of DC current required in each coil to simulate any heading must be known for a magnetic disturbance free environment (including the aircraft disturbances). This information may be provided by the vendor of the flux-gate sensor, or the information can be determined independently. In the latter method, the flux-gate sensor is physically removed from the aircraft and placed in a magnetic disturbance free environment that provides a surveyed compass rose. The characteristics of the sensor are then recorded as the unit is manually rotated to several different headings.

When performing the "electrical swing" on-board the aircraft, DC currents are applied to the coils to simulate a heading in a disturbance free environment. The difference between the displayed heading and the desired heading are then recorded, and another data sample is taken at a different heading. Again, the more data points taken, the more accurate the error analysis will be.

Note that the "electrical swinging" method records only the aircraft's hard-iron (1-cycle) errors. This method of calibration does not account for soft-iron (2-cycle) errors since they are a function of the relative direction and magnitude of the earth's field. Soft-iron error analysis still requires that the aircraft be manually rotated. However, since soft iron errors tend to be a function of the aircraft design, they typically are stable. Thus, once calibrated using the manual swing technique, soft iron errors usually do not need to be recalibrated, and subsequent electrical swings therefore provide adequate results.

7 Future Trends

The magnetometers discussed in this paper are very reliable, relatively inexpensive, and sufficiently accurate for most uses today. Some military applications, however, require more accuracy, especially during accelerated flight and aggressive turns. Developers have therefore created a magnetometer that is strapped down as opposed to gimbaled. To detect the roll angle of the aircraft, it uses a separate flux-gate coil placed at an angle that senses the vertical component of the earth's magnetic field. Coupling this roll information with acceleration data from an accelerometer, the magnetometer can then provide very accurate heading information under all flight conditions. This "3-axis strapdown magnetometer" [8] is already in use today in some aircraft, and should become more widely used in the future.

The most dramatic change with magnetometers will probably be in the way they are calibrated. Currently, on-ground manual swinging is the most common method used today and it takes at least 3 hours to perform. The time used for calibration using the manual swing technique may soon be reduced to no more than 10 minutes by including comparisons against an inertial heading reference such as an INS or Attitude Heading Reference System (AHRS). Either an on-board computer or a carry-on computer will collect and process the information during the swing and digitally relay the appropriate calibration coefficients back to the unit for storage. A developmental model of such a system (an improved version of the U.S. Army's AN/ASN-43) calibrated a compass to 0.5° RMS within 6 minutes using 1975 technology.

References

[1] M. Kayton and W. R. Fried, *Avionics Navigation Systems*, John Wiley & Sons, Inc., New York, 1969.

[2] R. T. Merrill and M. W. McElhinny, *The Earth's Magnetic Field*, Academic Press, New York, 1983.

[3] Nathaniel Bowditch, *American Practical Navigator*, Defense Mapping Agency Hydrographic/Topographic Center, Pub. No. 9, Washington, D. C., 1984.

[4] *Private Pilot Manual*, Jeppesen Sanderson, Inc., 1983.

[5] H. D. Gardner, *Improved Flux-Gate Magnetometer* NASA Langley Research Center, LAR–13560.

[6] J. G. Farrar, U.S. Patent 4,918,824, *Electronic Digital Compass*, April 24, 1990.

[7] Technical Newsletter, Sperry Flight Systems Division, TA-595-TN-1, March 1963, Revision 2, Oct 31, 1969.

[8] W. A. Shapiro and C. E. Roemer, "A Strapdown Magnetic Azimuth Detector", *Navigation: Journal of The Institute of Navigation*, Vol. 17, No. 4, Winter 1970-1971.

[9] D. Baker and M. Skaar, *Means For Compass Swinging Without Rotating the Aircraft*, Proceedings of the International Air Transport Associations Sixteenth Technical Conference, Miami, FL, April 1965.

SECTION IV
SYSTEM ANALYSIS, DESIGN
AND SYNTHESIS METHODOLOGIES

THE INTRODUCTION
by

Dr. D.F Liang
Directorate Research and Development
Communications and Space
National Defence Headquarters
Ottawa, Ontario
Canada
K1A OK2

INTRODUCTION

For aerospace navigation systems, the rapid advances of sensor technology, estimation techniques, high speed processors as well as the necessity to satisfy mission requirements in a cost/weight/reliability/performance conscious environment have led to increasing applications of integrated multi-sensor navigation systems. In general, stand alone dead-reckoning navigation subsystems require periodic updates, which may be time consuming and can affect the operational effectiveness of the aerospace platforms. Radio navigation systems are inherently noisy over the short term, but the errors are normally bounded. Even though the GPS is an accurate radio navigation system, which can act as a "stand alone" navigation system, it cannot on its own provide attitude information of sufficient interest to most aerospace navigation users. In contrast, the synergism and fault tolerance realized by automated complementary sensor data blending provides performance accuracy and reliability far beyond the capabilities of individual system components.

This section deals with system analysis, design and synthesis methodologies of multi-sensor system integration. As an introduction to this Section, a general overview of multi-sensor integrated navigation system design is presented. In most aerospace navigation system applications, the INS and GPS are considered the two most important navigation sub-systems, therefore the emphasis of our discussions on the system analysis, design and synthesis methodology is on GPS/INS integration looking at deep integration, general integration overview and federated fault-tolerant integration. For real-time suboptimal filter design, a practical implementation of an airborne transfer-of-alignment system aiming at a challenging Synthetic Aperture Radar Motion Compensation application is presented. This is followed by a detailed discussion on the vertical channel design of an inertial system.

To be more specific, in Dr. D. Liang's Chapter on "An Overview of a Generic Multi-Sensor Generic Integrated Navigation System Design", he presented a general overview of the design of a

208

Multi-Sensor Generic Integrated Navigation System. It provides a tutorial of navigation system error models of interest in forming robust, complementary integrated systems. Considerable attention is placed on the Kalman filter design and on the selection of error states to reduce computational requirements without compromising performance accuracy and stability. Lessons learned from his development of a number of integrated navigation systems have been summarized to assist the integrated system designer so that practical development can be completed with minimum risk.

In Col Lewantowicz and Major Paschall's Chapter on "Deep Integration of INS, GPS and Other Sensor Information", the authors presented the methodology to perform optimal deep integration of information provided by the GPS, INS and other sensors. Their chapter discusses the need to perform vigorous error modelling and validation in support of system level analysis and filter synthesis. High fidelity simulation analysis provides a sound basis for understanding parameter sensitivities in Kalman filter integration schemes. The simulation analysis is essential for sound engineering trade-offs in Kalman filter implementations. Introductory material is provided at the outset to discuss various sensor operational concepts, error modellings, linear system tools and estimation theory. Integration issues are illustrated by way of analysis case studies looking at single versus multiple filter integration and stability. The analysis of differential GPS integration with INS and other sensors is the final example presented. The final section

presents a broader discussion from a mission context of the role of INS/GPS in future avionics and weapon systems.

In Dr. Carlson's Chapter on "Federated Filter for Fault-Tolerant Integrated Navigation", he presented an efficient, federated Kalman filtering methodology, based on rigorous information-sharing principles. The method applies to decentralized navigation systems in which one or more sensor-dedicated local filters feed a larger master filter. The local filters operate in parallel, processing unique data from their local sensors, and common data from a shared inertial navigation system. The master filter combines local filter outputs at a selectable reduced rate, and yields estimates that are globally optimal or subset optimal. The method provides major improvements in throughput (speed) and fault tolerance, and is well suited to real-time implementation. Practical federated filter examples are presented, and discussed in terms of structure, accuracy, fault tolerance throughput, data compression, and other real-time issues.

In Dr. Greenspan's Chapter on "GPS/Inertial Integration Overview", he presented a rather thorough overview of GPS/INS Integration. It presents an overview of the benefits and trade-off issues of integration; an overview of integration architectures and algorithmic concerns; and sample results from performance evaluations of integration case studies.

In Mr. DiFilippo's Chapter on "Design Considerations for a Suboptimal Kalman Filter", he dealt with the practical implementation of a transfer alignment filter for a

Synthetic Aperture Radar Motion
Compensation System.

He started with the
optimal filter design and proceeded
at length to show how a practical
suboptimal filter can be designed to
accommodate state deletion and
decoupling. The criteria used in
making the filter model
simplification provide useful
guidelines for general suboptimal
filter design. Simulation results
were presented to confirm that the
suboptimal filter provided similar
performance to the optimal filter.

In Dr. Ausman's Chapter on
"Vertical Channel Design
Considerations", he first discussed
barometric and inertial errors,
followed by a detailed look at the
LN93/94 conventional vertical
channel mechanization, and finally
described the design and performance
of a 5-state Kalman filter using
barometric altitude updating. It is
shown to have achieved superior
performance by modelling the baro
scale factor error as a linear
function of altitude, a relationship
which the real atmosphere tends to
follow.

AN OVERVIEW OF A GENERIC MULTI-SENSOR INTEGRATED NAVIGATION SYSTEM DESIGN

Dr. D.F. Liang
Directorate Research and Development
Communications and Space
National Defence Headquarters
Ottawa, Ontario
Canada K1A OK2

1. INTRODUCTION

Modern avionics systems are becoming increasingly sophisticated as the demands for better mission performance and higher reliability continue to escalate. Many of the missions must be carried out at ultra-low altitude under all weather and visibility conditions. The increased range, speed and accuracy of modern weapon systems, impose stringent accuracy and reliability requirements upon the aircraft navigation system. To enhance mission success in a hostile environment, the pilot amongst other things needs to operate weapon systems, target acquisition and designation systems, radar detection, night vision systems and perhaps engage in air-to-air combat. This paper describes the application of Kalman filtering technology to the design and development of a multi-sensor Generic Integrated Navigation System (GINS).

2. SYSTEM OVERVIEW

A typical structure of the GINS can be illustrated by Figure 1. The primary system component is the integrated sensor unit which consists of all the selected navigation sensors (subsystems). They in general, can be classified into two main categories, dead reckoning and radio navigation systems. The dead reckoning (DR) systems are the inertial navigation system (INS), doppler radar and air data computer. These systems are self-contained and provide velocity information which can be integrated to obtain the position of the aircraft. Position estimates so obtained have a tendency to drift, slowly accumulating a large error over an extended flight. DR systems (except doppler) are generally non-radiating and non-jammable and hence attractive for operation in a hostile environment. On the other hand, radio navigation systems such as Omega, TACAN and GPS (Global Positioning System) rely upon radio signals transmitted from several external sources which make them vulnerable to enemy actions in any serious confrontation. These systems provide absolute position information with error characteristics which do not grow with the passage of time, because errors are not accumulated through an integration routine. In the short term however, radio aids excepting the GPS generally do not produce as stable and accurate a track as the DR systems.

In order to take advantage of the short-term stability of the DR system and the long term stability of the radio navigation system, one can apply the theory of optimal estimation in the form of a Kalman filter (K.F.) to combine all the available measurements and provide a statistically "optimal" estimate of the aircraft position, attitude and velocity, etc. The workload of optimal integration, navigation and

guidance functions can be handled by the data processing computer. This means that large amounts of data can be handled at high rates, and the computer can be entrusted to carry out very sophisticated computations to provide much more reliable navigation and control output data.

To achieve the desired performance, the GINS processor must perform signal conditioning and sensor compensation functions (Figure 2), and it must contain efficient K.F. integration algorithms, error control and performance monitoring routines. The signal conditioning function controls the sampling rate of incoming data and works as a prefilter to average the higher frequency data input. The sensor compensator compensates for predictable deterministic errors intrinsic to each sensor, and also provides compensation to remove the effects of unstable positive feedback in the GPS rate aiding loop. The performance monitoring routines serve as a watchdog to keep track of the overall system performance, and also provide CEP estimates to indicate the expected performance accuracy of the automated system.

The GINS integration filter will optimally weigh and combine all available sensor data to estimate aircraft position, velocity and attitude with greater accuracy than available from any of the sensors individually. This means that mission requirements can be met by using sensors which are individually less accurate and less expensive than would otherwise be required. In the event of subsystem failures, the system performance degradation will be graceful and the integrated system will automatically configure itself to the next optimum operational mode.

The use of a digital Data Bus (MIL-STD-15538) facilitates configuration selection and data interface of different brands and types of equipment, and provides a higher data flow capacity, self check on each data transmission and reduces susceptibility to electromagnetic interference. It also reduces weight and improves reliability because less wiring and fewer connectors are needed.

3. SYSTEM DEVELOPMENT APPROACH

For the system design of a generic multi-sensor integrated navigation system, a large number of equipment configurations are possible. The typical approach is to use previous experience in selecting a candidate configuration in an ad hoc manner. This has the potential danger of eliminating good alternatives early in the project and could result in a suboptimal configuration. It is therefore more prudent to dedicate some resources to simulate and study two or three potential configurations with the aim of identifying, developing and testing an integrated navigation system which best satisfies the mission requirements. The prototype of the selected system configuration can then be developed and tested with the aim of subsequent engineering development for use on board the chosen mission platform. The development effort can be divided into at least the following two phases:

3.1 Phase I - System Definition and Design.

To properly analyse the performance of various candidate configurations, a versatile simulation package must be developed. Since the fidelity and performance prediction of these configurations are of prime importance, a substantial portion of the development effort should be dedicated to generating complete and

212

accurate error models. These activities are described as follows:

a. Survey existing or soon to be available navigation sensors to determine their suitability.

b. Identify several cost-effective candidate system configurations that can potentially satisfy the performance, weight, size and reliability requirements.

c. Generate the navigation sensor error models.

d. Generate a set of mission profiles to be used for trajectory generation.

e. Develop the integration algorithms to blend the sensor outputs.

f. Develop control and display software.

g. Develop diagnostic software to detect sensor failure.

In the detailed simulation analysis both covariance analysis and Monte Carlo simulation software should be used. Covariance analysis software is effective for the design of integration filters because it provides ensemble statistical data. Because ensemble statistics are the outputs, covariance analysis can (in many circumstances) reduce the need for computationally intensive Monte Carlo simulations. This statistical information can be used to assess candidate Kalman filter designs and to project the performance of a particular navigation system configuration. Covariance analysis is very useful for assessing the effects of mismatch between the filter design model and the "real

world" or truth model. In addition, it can be readily used to establish error contribution tables and error budgets which let the filter designer focus in on the major error contributors.

However, Monte Carlo simulation also has its place in the design of suboptimal Kalman filters. Simulation can be particularly effective in assessing the effect of nonlinearities which are difficult to address in the covariance analysis framework. Also, certain types of mismatches between the Kalman filter model and the truth model are more conveniently addressed with the simulation program than with the covariance program. An example of this is the sensitivity assessment of the effect of sea current correlation time mismatch.

Another area in which the Monte Carlo simulation can be of more use than the covariance analysis program has been in the investigation of the effects of unmodelled manoeuvre-dependent sensor errors. As well, covariance analysis software will never be able to replace the function of simulation software for final checkout of Kalman filter code.

3.2 **Phase II Development and Testing**
· Acquire the sensor hardware for the preferred configuration and complete the development of the design produced in Phase I.

· Develop the data base and interface electronics to connect the sensors.

· Refine and convert the preliminary forms of the Phase I Kalman filter and

integration software into flyable programs. Develop the control and display software and transfer these programs into the airborne processor.

- Conduct static tests of the completed system in an integration laboratory and low dynamics tests in a mobile van.

- Conduct flight tests in a test aircraft to evaluate system performance in a realistic vibration environment.

- Assess the navigation system's performance by conducting extensive flight tests in a test aircraft on an instrumented range.

The Phase II development can be aimed at realizing in hardware the Phase I system design by constructing and developing through ground and flight testing the Advanced Development Model (ADM) to produce a fully developed and flight validated integrated navigation system prototype.

4. **SYSTEM ERROR MODELLING**
Detailed simulation and error models for all the relevant navigation equipment and environmental disturbances must be developed for the performance evaluation and sensitivity analysis.

A wide variety of types and brands of navigation sensors can contribute to meeting a specific set of requirements. The following list of generic navigation subsystems can be considered having merit in any GINS applications:

Global Positioning System (GPS),

Inertial Navigation System (INS),
Attitude and Heading Reference System (AHRS),
Doppler Radar,
TACAN,
Omega,
Air Data System,
Strapdown Magnetometer,
Radar Altimeter.

However, many of the generic equipment categories can be further subdivided. For example, in the INS category there are high and medium accuracy inertial equipments available. The inertial sensors can be mechanized either as a gimballed platform or a strapdown configuration. Strapdown configurations may use conventional or ring laser gyroscopes. Similarly, Doppler velocity sensors and Doppler navigation systems are available in the Doppler category.

Now that the GPS system is fully deployed, the utility of Omega has become rather limited, however for the sake of completeness of treatment, Omega has been included in this study as an alternative. TACAN can supply absolute navigation data from ground stations, and its transmitter can be used for relative navigation and as a homing signal. The air data system and magnetometer are also considered as part of the standard airborne equipment. Furthermore, due to the relative inaccuracy of the barometric altitude information, a radar altimeter is considered necessary for low altitude flight. Therefore, a radar altimeter is also included in the evaluation of all candidate configurations.

4.1 **Inertial Measurement Unit (IMU) Models.**

The models described in the following subsections are basically developed for an inertial sensor assembly. The same IMU error model is utilized for AHRS systems with appropriate changes in the magnitudes of the sensor error parameters. This approach is reasonable, should an AHRS be selected as the inertial reference, it is expected that strapdown computations would be implemented in the processor. Thus the inertial measurements made by the AHRS would be processed just as if the AHRS were an inertial sensor assembly for an inertial navigation system.

4.1.1 IMU Error

Several mechanizations can be employed to represent errors of Inertial Measurements Units (IMU). One such error model is the "Ψ - angle" formulation, which is a classic set of navigation error equations applicable to either strapdown or platform systems. The error model is described by the following set of vector differential equations:

$$\dot{\psi} = \tau_g - (\rho + \Omega) \times \psi ,$$

$$\dot{\delta V} = \tau_a - \psi \times A^L - (2\Omega + \rho) \times \delta V - \omega_s^2 \, \delta R + 3 \, \omega_s^2 (\delta R - \frac{R}{r}) \frac{R}{r} + \delta g.$$

This has the advantage of simpler attitude error propagation provided that the position error is propagated via

where

$$\dot{\delta R} = \delta V - \rho \times \delta R,$$

ψ	=	attitude error vector,
δV	=	velocity error vector,
δR	=	position error vector,
ρ	=	angular velocity vector of local vertical reference frame relative to an earth-fixed frame,
Ω	=	angular velocity vector of the earth relative to an inertial frame,
A^L	=	vehicle non-gravitational acceleration
R	=	radius vector from earth's center to vehicle,
r	=	distance from earth's center to vehicle,
ω_s	=	Schuler frequency = $\sqrt{g/r}$,
τ_g	=	gyro output error vector,
τ_a	=	accelerometer output error vector,
δ_g	=	gravity error vector.

The total gyro and accelerometer error vectors are expressed in terms of individual sensor errors as

$$\tau_g = CHX_g,$$

$$\tau_a = CGX_a,$$

where

X_g	=	vector of gyro error (bias plus random noise),
X_a	=	vector of accelerometer error (bias plus random noise),

C = transformation matrix relating inertial sensor assembly reference frame to the local-vertical frame,

H = coefficient matrix which transforms the collection of gyro errors into a net angular rate error in the sensor assembly frame,

G = coefficient matrix which transforms the collection of accelerometer error in the sensor assembly frame.

In the event the IMU is strapdown in nature, the transformation matric C is defined to be the direction cosine matrix relating the vehicle body axes to the local-vertical frame. When the IMU is of the gimballed platform type, the matrix C is an identity matrix, since the inertial sensor assembly is at all times coincide with the local vertical reference frame.

4.1.2 IMU Simulation Model.
The IMU simulation model implements the classic set of navigation equations

$$\dot{T} = T[\omega] - [\rho + \Omega]\, T,$$

$$\dot{V} = TA^L - (2\Omega + \rho)\, x\, V + g,$$

$$\dot{D} = D[\rho].$$

Where,

T = transformation matrix from body frame to the local-vertical frame,

D = transformation matrix from the local-vertical frame to the earth reference frame,

ρ = angular velocity vector of the local vertical frame relative to the earth reference frame,

V = velocity vector of the vehicle relative to the earth reference frame,

Ω = angular velocity vector of the earth relative to an inertial frame,

ω = angular velocity vector of the vehicle relative to an inertial frame,

g = gravity vector,

A^L = vehicle nongravitational acceleration,

$[\cdot]$ = skew-symmetric realization of the enclosed vector.

4.1.3 IMU Sensor Errors.
Most of the gyro and accelerometer sensor errors can be treated as random constants or slowly varying Gauss Markov processes. They are:

Gyro

G^2 - drift coefficient,
Bias drift,
Scale factor error,
Input axis misalignment,
G - sensitive drift coefficient.

Accelerometer

Bias error,
Scale factor error,
IA misalignment,
Vibration induced Bias.

216

The remaining errors can be treated in the following manner:

Gyro random drift (random walk),
Gyro attitude random walk (white noise drift),
Gyro and accelerometer turn-on transients (exponential decay).

The typical RMS values for gimballed and strapdown INS errors are shown in Table 1.

4.2   Doppler Radar.

Doppler radar measures the aircraft velocity with respect to the fixed-antenna coordinate system.   The errors include:

a.   Doppler fluctuation noise;

b.   Scale factor error in the along heading velocity component;

c.   Cross coupling scale factor errors; and

d.   Speed offset errors.

4.2.1 Doppler Fluctuation Error.

This error represents the noiselike character of the Doppler spectrum. It is due to variations in the backscatter of the terrain over which the Doppler measurements are being made. Typical correlation times for this error are on the order of 0.25 to 1.0 second, with power spectral density of the order of 0.002 - 0.005 kt^2/rad/sec/kt.

4.2.2 Scale Factor Error.

Typical values of scale factor error due to antenna calibration and transmitter frequency deviation are in the range of 0.1 to 0.25% of total velocity.

4.2.3 Speed Offset Error.

This error is intended to model a bias in the Doppler frequency tracker. The RMS value for this error is in the order of 0.05 to 0.1 kt.

4.2.4 Ocean Current Error.

The effects of ocean currents can be accounted with north and east ocean current states represented as first order Gauss-Markov processes. Typical RMS values of the current states range from 0.8 to 2.5 kts.

$$\dot{v}_{cn} = \frac{1}{\tau_c} + \eta_{cn}$$

and

$$\dot{v}_{ce} = \frac{1}{\tau_c} v_{ce} + \eta_{ce}$$

where

τ_c =   correlation time constant of 1 to 2 hours for ocean current,

η_{cn} =   north ocean current white noise,

and

η_{ce} =   east ocean current white noise.

4.2.5 Wind-Induced Error

The wind-induced water motion velocities are commonly represented by

$$v_{wm} = 1.3 \ (v_w)^{\frac{1}{3}} \ \frac{v_{wm}}{v_w},$$

$$v_{we} = 1.3 \ (v_w)^{\frac{1}{3}} \ \frac{v_{we}}{v_w} ,$$

where

v_w =   magnitude of wind,

v_{wn} =   north surface wind velocity,

and

v_{we} =   east surface wind velocity.

The doppler velocities in the aircraft coordinates are given by

$$v_{xd} = v_{xt} + v_{xc} + N_x + S_x v_{xt} + B_{xy} v_{yt} - (v_{cn} + v_{wn}) C_{11} - (v_{ce} + v_{we}) C_{12},$$

$$v_{yd} = v_{yt} + v_{yc} + N_y + S_y v_{yt} + B_{yz} v_{zt} + B_{yx} v_{xt} - (v_{cn} + v_{wn}) C_{21} - (v_{ce} + v_{we}) C_{22},$$

and

$$v_{zd} = v_{zt} + v_{zc} + N_z + S_z v_{zt} + B_{zy} v_{yt} + B_{zx} v_{xt} - (v_{cn} + v_{wn}) C_{31} - (v_{ce} + v_{we}) C_{32},$$

where

v_{it} =   true velocity along i-axis,

C_{ij} =   north-east-down to x-y-z frame transformation matrix,

v_{ic} =   speed offset along i-axis,

N_i =   fluctuation noise along i-axis,

S_i =   scale factor along i-axis,

B_{ij} =   cross coupling scale factor error,

for which i = x, y, z.

The typical (1RMS) Doppler error parameters are presented in Table 2.

4.3 GPS

GPS three dimensional position and velocity errors can be modeled as independent first order Gauss-Markov processes with relatively long correlation times of one or more hours.

$$\dot{x}_{ig} = - \frac{x_{ig}}{\tau_{gp}} + \eta_{igp},$$

$$\dot{v}_{ig} = - \frac{v_{ig}}{\tau_{gv}} + \eta_{igv},$$

where

x_{ig} =   i-axis position correlated error,

v_{ig} =   i-axis velocity correlated error,

τ_{gp} =   position error correlation time,

τ_{gv} =   velocity error correlation time,

η_{igp} =   i-axis position correlated noise,

η_{igv} =   i-axis velocity correlated noise,

for which i = N, E, D.

218

Typical 1 σ values for η_{igp} and η_{igv} are 15 metres and 0.1 m/s, respectively. It should be noted that GPS velocity errors generally have correlation times short enough to be represented as uncorrelated white noise processes.

4.4 Magnetometer Model
The magnetometer provides a true heading measurement referenced to true north. The heading error can be modeled as

$$\psi_m = \psi_b + \eta_1 \sin(\psi_t) + \eta_2 \cos(\psi_t) + \eta_3 \sin(2\psi_t) + \eta_4 \cos(2\psi_t) + \eta_5$$

where

ψ_t = true heading,

ψ_b = slowly varying instrument bias error with typical value of 0.5 - 2 degrees

η_1, η_2 = single cycle Gauss - Markov heading errors with typical RMS of 0.5 - 1.0 degree,

η_3, η_4 = two cycle Gauss-Markov heading errors with typical RMS of 0.5 - 1.0 degree,

η_5 = uncorrelated white noise in the order of 0.1 degree.

4.5 Air Data Model
The air data subsystem provides air speed and barometric altitude measurements. The barometric altitude contains four potential sources of error. They are:

a. Error due to altitude variation in a constant pressure surface;

b. Temperature dependent scale factor error;

c. Static pressure measurement error; and

d. Instrument errors.

Depending on flight profiles, the dominant source of altitude errors could vary. For a maritime aircraft with low altitude flight profile the dominant source of error can be reasonably modeled as a first order Gauss-Markov process

$$\dot{h}_b = -\frac{1}{\tau_b}h_b + \eta_b$$

where

h_b = correlated barometric altitude error,

τ_b = barometric altitude correlation time which equals to the correlation distance divided by the vehicle velocity.

η_b = barometric altitude white noise error.

The air speed can be modeled as the groundspeed corrupted by winds. They are given by

$$v_{ha} = \cos \psi_t\, v_{na} + \sin \psi_t\, v_{ea}$$

and

$$V_{la} = \cos \psi_t \ V_{ea} - \sin \psi_t \ V_{na},$$

where

V_{na} = northerly airspeed,

V_{ea} = easterly airspeed,

V_{ha} = airspeed along the heading,

V_{la} = lateral airspeed,

The northerly and easterly airspeeds are given by the equations.

$$V_{ea} = v_{et} - (1 - \frac{h_t}{h_a}) \ v_{es} - \frac{h_t}{h_a} \ v_{ea}$$

and

$$v_{na} = v_{nt} - (1 - \frac{h_t}{h_a}) \ v_{ns} - \frac{h_t}{h_a} \ v_{na}$$

where

h_t = true altitude,

h_a = wind aloft altitude,

v_{es} = easterly surface wind velocity,

v_{ea} = easterly aloft wind velocity,

v_{ns} = northerly surface wind velocity,

v_{na} = northerly aloft wind velocity,

The surface and aloft wind velocities can be represented by the combinations of slowly and rapidly varying first order Gauss-Markov processes with corresponding correlation time.

4.6   TACAN Error

The TACAN measurement is corrupted by range and bearing errors. The errors sources can be modelled by:

a.   Random white noise ranging error of the order of 0.1 and 0.2 nm;

b.   First order Markov bias ranging error to account for time-delay between signal reception and transmission with typical correlation time of 1 to 2 hours and RMS value of 0.1 to 0.2 nm;

c.   Random white noise bearing error of the order of 0.1 to 0.5 degree;

d.   First order Markov bearing calibration error; and

e.   First order Markov slowly varying bearing correlation error.

4.7   Radar Altimeter Error

The radar altimeter measures altitude above ground level. Thus for overland operation the largest error in altitude will be due to variations in the terrain itself. Radar altimeter errors are modeled as a random scale factor error plus

an uncorrelated (white noise) altitude error.

The radar altimeter simulation model uses the error model to produce simulated radar outputs. The radar altimeter error in altitude is added to true value of altitude obtained from the trajectory generator data to produce simulated radar altimeter outputs.

4.8   OMEGA Error

Omega is a hyperbolic system that utilizes lines of position based on phase difference measurements from at least three transmitters. It operates in the 10 to 14 KHz band and, since each station has a range of 13,000 Km, only eight stations are sufficient to provide world-wide coverage.

Diurnal changes in the ionospheric propagation characteristics and the inhomogeneity of the earth's magnetic field and surface conductivity cause anomalous variations in the Omega phase measurements. These deterministic errors can be partially compensated for by using phase propagation correction (PPC) tables. However, there remain four major random components of residual error. The statistical properties of this residual error have been studied in detail by various groups and it is generally agreed that when PPC tables have been used, the remaining phase error at the Omega receiver has an autocorrelation function of the form.

$$\Phi(\tau) = A^2 e^{-\tau/T1} + B^2 e^{-\tau/T2} + C^2 e^{-\tau/T1} \cos(\omega\tau)$$

where auto-correlation times T1 and T2 are respectively 180,000 and 4800 seconds, ω is the earth rate and A, B, C are respectively 1520, 500 and 850 metres. This phase error can therefore be adequately described by the sum of three stochastic processes: two first order Markov processes and a periodic process with a period of 24 hours. Since the periodic process is second order, a total of four states are required to properly model each Omega phase error. Determining three independent Omega lines of position (LOPs) requires reception of Omega signals from 4 distinct stations.

4 states are required to model each Omega phase error which propagates according to

$$\frac{d}{dt}\begin{bmatrix} \text{BIAS} \\ \text{MARKOV} \\ \text{PERIODIC} \\ P_2 \end{bmatrix} = \begin{bmatrix} -1/T1 & 0 & 0 & 0 \\ 0 & -1/T2 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & -\alpha^2 & -2/T1 \end{bmatrix} \begin{bmatrix} \text{BIAS} \\ \text{MARKOV} \\ \text{PERIODIC} \\ P_2 \end{bmatrix} + W_B$$

where W_B is a zero mean Gaussian white (ZMGW) noise process, and P_2 is α the extra state needed to model the periodic error. T1 and T2 are the Markov process correlation times and α is a constant. Here the strongly correlated Markov process is referred to as a bias to distinguish it from the more weakly correlated one (i.e. T1>>T2).

5. THE KALMAN FILTER DESIGN

The dynamics of the error state vector

$$\dot{x}(t) = F(t)x(t) + w(t)$$

where $F(t)$ is the fundamental matrix and $w(t)$ is a zero mean Gaussian white noise process with spectral density matrix Q. The dynamic system model defines the manner in which the system errors propagate in time and the geometry relating estimated error states to error observations constructed from the primary and auxiliary sensors. There are many possible suboptimal filter models that could be derived from simplified truth models. The major problem in Kalman filter design is to determine which error states can be ignored or grouped together to reduce the size of the filter (and hence the computational burden), without significant performance degradation.

The measurement model describes the relationship between the inputs to the Kalman filter, $\underline{z}$, called measurements, and the state vector $\underline{x}$ that is to be determined from these inputs. For linear systems this is generally of the form

$$\underline{z}(t) = H(t)\underline{x}(t) + \underline{v}(t)$$

where $H(t)$ is the measurement matrix,

and $v(t)$ is the measurement noise with spectral density R.

The discrete form Kalman filter equations may be written as

$$\hat{X}_k(+) = \phi_{k-1}\hat{X}_{k-1}(+) + K_k[Z_k - H_k\phi_{k-1}\hat{X}_{k-1}(+)]$$

and

$$P_k(+) = [1 - K_k H_k] P_k(-)$$

where

$$K_k = P_k(-)H_k^T / (H_k P_k(-)H_k^T + R_k)$$

The error covariance propagation is given by

$$P_{k+1}(-) = \phi_k P_k(+) \phi_k^T + Q_k$$

The filtering algorithm most commonly implemented is Bierman's U-D factorised Kalman filter. This algorithm avoids the explicit and computation of the estimation error covariance matrix P_k by propagating in terms of its factors U and D:

$$P_k = UDU^T$$

where U is a unit upper triangular matrix and D is a diagonal matrix. The UD factors are calculated by the modified weighted Gram-Schmitt (MWGS) algorithm.

The U-D algorithm is efficient and provides significant advantages in

222

numerical stability and precision. Specifically, the factorization of P_k provides an effective doubling in computer word length in covariance-related calculations, and avoids filter divergence problems which can arise in more conventional filter mechanizations due to loss of the positive (semi-) definite property of P_k through the accumulation of round-off and truncation errors.

For the practical Kalman filter design of any inertially based integrated navigation systems, two categories of modelling simplifications are possible. The first one deals with inertial error dynamics. In Section 4.1 we have commented that the Ψ formulation has become very popular because of simpler attitude error propagation, in addition, some of the following simplifications should be considered.

5.1 Elimination of Vertical Axis Modelling
For most aircraft navigation system, it is possible to eliminate vertical axis error modelling. Conventional fixed-gain error control mechanizations using altitude measurements might provide adequate vertical axis performance. As well, there is normally little cross coupling from the vertical to horizontal axes to demand the modelling of vertical axis error state. However the vertical velocity error can affect horizontal error propagation through the horizontal coriolis acceleration components.

5.2 Elimination of Horizontal Coriolis Acceleration
In most aircraft application, the Coriolis acceleration error due to the vertical velocity is substantially less than the uncertainty in gravity model. Such an observation allows the Kalman

filter designer to eliminate vertical axis dependencies from his dynamic error model. In most cases, the dynamical coupling of errors associated with the system computed coriolis acceleration errors can be eliminated.

5.3 Kalman Filter Error States
In view of the above, a Kalman filter design for a generic multi-sensor integrated navigation system might adequately be represented by the following error states:

5.3.1 INS

- INS horizontal position and velocity errors (4 states)

- INS attitude errors (3 states)

- 3 gyro drift rates and 2 acceleration bias errors (5 states)

5.3.2 Doppler

- Doppler boresight and along heading scale factor errors (2 states)

- Sea surface currents for doppler over water application (2 states).

5.3.3 GPS

- No GPS error states are necessary except when GPS pseudo range and range rate data are used instead of position and velocity measurements.

5.3.4 Omega

- 8 Omega states (1 long term and 1 short term

error state for each of
four stations)

5.3.5 For most applications, it is
not necessary to dedicate special
error states to estimate errors
associated with the air data
strapdown magnetometer, and radar
altimeter, etc. The vertical
channel design can normally take
care of errors in the vertical
channel, and this is fully addressed
in S. Ausman's Chapter on Vertical
Channel Design Considerations.

5.4   GPS Rate Aiding
The INS data can be used to rate aid
the GPS receiver. The aiding
permits narrower bandwidth code
tracking loops in the GPS receiver
to improve its anti-jamming
performance. It also reduces the
satellite acquisition time by using
a priori antenna velocity data to
account for Doppler shift of the
carrier frequency.

The closed loop system design of
using GPS measurement based Kalman
filter estimates to reset INS was
found to be unstable at high Kalman
gains in the noncoherent mode.
Widnall [3] has suggested an
alternate approach of using reduced
control gains for correcting INS
errors. Such an approach was also
found to be insensitive to tracking
loop bandwidth.

6.0   Covariance Analysis
Covariance analysis is an efficient
and powerful tool for sensitivity
performance analysis to determine
the contributions of distinct error
sources. It is essential for
developing a robust filter design of
minimum state size. Its simulation
analysis consists of three major
components [4].

a.   an aircraft trajectory
     generator which provides
     nominal flight data,

b.   a reference sensor error truth
     model which characterizes all
     the sensor errors, and

c.   the Kalman filter reduced-
     order design to be evaluated.

To conduct sensitivity analysis, a
covariance analysis is performed
with the reduced order Kalman
filter, and the gain history is
recorded. Then another covariance
analysis is performed using the
truth model for all sensor errors,
with the Kalman filter gain computed
from the earlier step. The
performance obtained in the second
covariance analysis represents the
predicted performance of the reduced
order filter design.

Figure 3 illustrates the real-world/
filter-world analysis flow diagram
for integrated navigation system
performance assessment. The upper
half of the diagram represents the
iterative solution of the filter
covariance equations. These are
solved in order to generate the
sequence of filter gains, K_k^*, which
is a necessary input to the lower
half of the diagram, representing
the iterative solution of the truth
model covariance equations. Thus
two passes through the covariance
program are required to assess the
performance of a suboptimal filter.

7.   DESIGNER BEWARE
For any integrated system design, it
might be useful to consider some of
the following design tips.

a.   Be generous on size of truth
     model,

b.   Be precise on noise level of
     truth model,

c.   Minimize filter states,

d.   Be pessimistic on noise level
     of filter model,

e.   Implement easily modifiable
     computational algorithms
     (Don't take short cuts too
     early), and

f.   Potential causes of
     divergency:

     .   Numerical error (negative
         diag. P can be eliminated with
         appropriate choice of P
         propagation equation),
     .   Programming error,
     .   Nonlinearity,
     .   Unmodelled error,
     .   Optimistic noise level,
     .   Incorrect model.

8.   DESIGNER'S ROAD MAP

For the designer, a simple and easy
to understand road map is offered
for reference:

a.   Determine system truth error
     model,

b.   Determine true measurement
     model,

c.   Determine filter error model,

d.   Determine noise model and
     statistics,

e.   Determine initial conditions,
     noise and error variance,

f.   Simulate truth model and
     physical environments,

g.   Verify truth model simulation
     (How realistic?),

h.   Verify simulation tool: set
     all error sources to zero to
     verify that there are no
     unintended system errors when
     system output is compared to
     the input, or/ and signifi-
     cantly increase system errors,
     verify that the output errors
     are reasonable and behave as
     expected,

i.   Implement filter algorithm
     (could waste lots of time if
     this is done before #h),

j.   Review filter results,

k.   Tuning: good agreement
     between state estimate error
     and covariance, and between
     measurement residual and
     residual variance (through
     Monte Carlo and covariance
     analysis),

l.   Determine achievable optimal
     performance,

m.   Robustness test (filter should
     perform well with increased
     noise level of truth model),

n.   Filter sub-optimization
     (reduction of filter states to
     improve efficiency while
     increasing other noise levels
     to absorb unmodelled errors),

o.   Substitute simulated data by
     real data whenever possible,

p.   Conduct controlled test of
     partial operation, and

q.   Redesign or retune the filter
     if necessary.

9.   CONCLUSION

This paper has presented a brief
overview of the design of a multi-
sensor Generic Integrated Navigation
System (GINS). Lessons learned from
the development of a number of
integrated navigation systems have
been summarized to assist the
integrated system designer so that

practical development work can be
completed with minimum risk.

REFERENCES

[1]   Benson, D.O., "A comparison of
       Two Approaches, to Pure
       Inertial and Doppler-Inertial
       Error Analysis," IEEE Trans on
       Aerospace and Electronics
       Systems, Vol. AES-II, No.4,
       July 1975.

[2]   Bierman, G.J., "Factorization
       Methods for Discrete
       Sequential Estimation,"
       Academic Press, N.Y., 1977

[3]   Windall, W.S., "Alternate
       Approaches for Stable Rate
       Aiding of Jamming - Reistant
       GPS Receivers, "National
       Aerospace and Electronics
       Conference, May 15-17 1979.
       Dayton, Ohio.

[4]   Gelb, A., _Applied Optimal
       Estimation_, MIT Press,
       Cambridge, 1974.
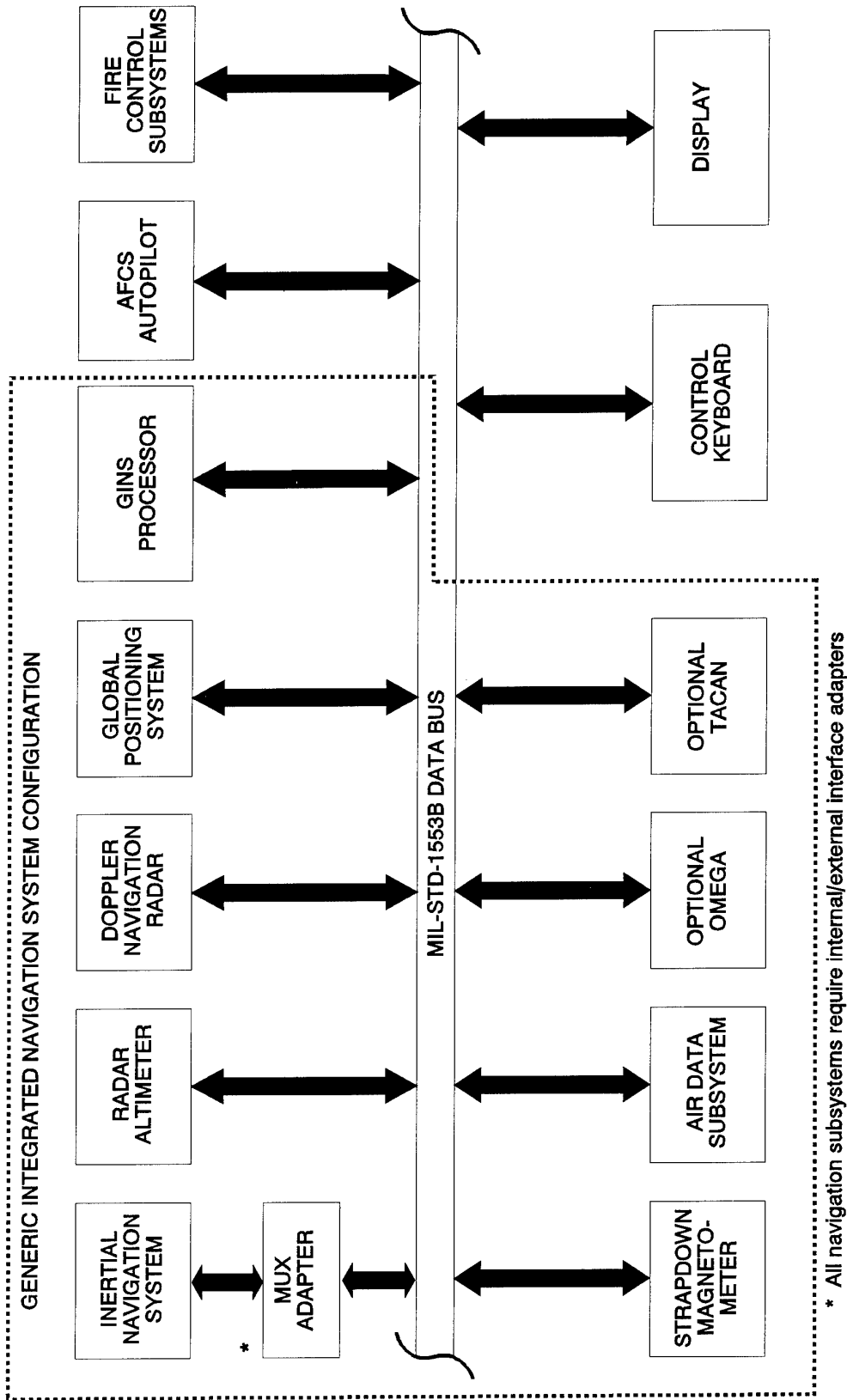
FIGURE 1: TYPICAL GINS CANDIDATE CONFIGURATION

GENERIC INTEGRATED NAVIGATION SYSTEM CONFIGURATION

FIRE CONTROL SUBSYSTEMS

AFCS AUTOPILOT

GINS PROCESSOR

GLOBAL POSITIONING SYSTEM

DOPPLER NAVIGATION RADAR

RADAR ALTIMETER

INERTIAL NAVIGATION SYSTEM

MUX ADAPTER

MIL-STD-1553B DATA BUS

DISPLAY

CONTROL KEYBOARD

OPTIONAL TACAN

OPTIONAL OMEGA

AIR DATA SUBSYSTEM

STRAPDOWN MAGNETO-METER

* All navigation subsystems require internal/external interface adapters

FIGURE 2: GINS INTEGRATION FUNCTIONS
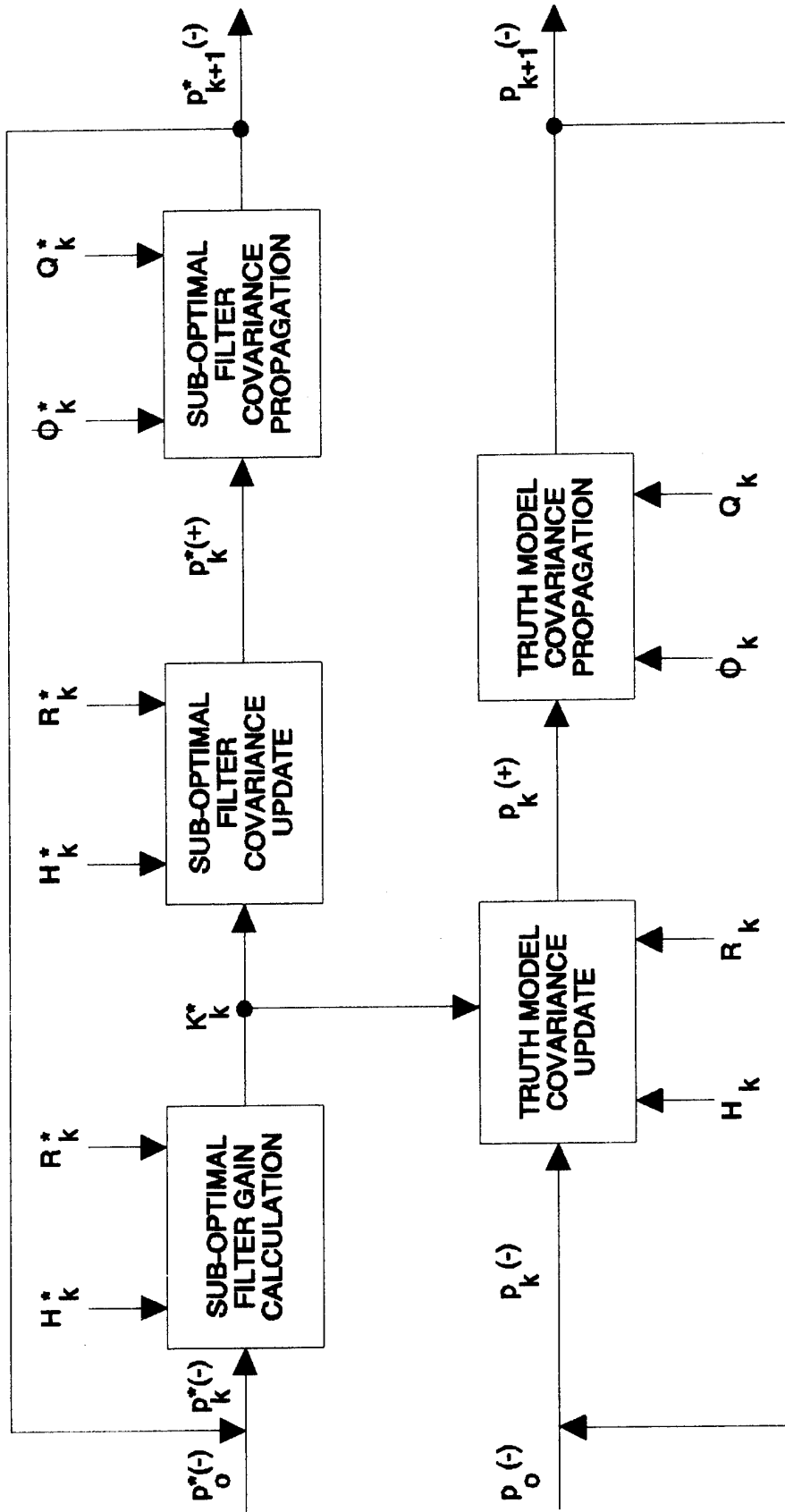
FIGURE 3: COVARIANCE ANALYSIS

TABLE 1

TYPICAL INS ERROR PARAMETERS

Parameter	Typical (1 σ) RMS Value Tuned Rotor Gyro (Gimballed INS)	Typical (1 σ) RMS Value RLG Gyro (Strapdown INS)
x, y, z gyro bias	.005, .005, .01 deg/hr	.008 deg/hr all axes
x, y, z Markov	.003, .003, .01 deg/hr for 2hr cor time	Markov not modelled
x, y, z gyro scale factor	200, 200, 1000 ppm	2 ppm all axes
x, y, z gyro misalignment	100 arc sec	5 arc sec
x, y, z gyro random walk	random walk not modelled	.002°/$\sqrt{hr}$ random walk all axes
x, y accelerometer bias	100, 100 μg	40 μg both axes
x, y accelerometer Markov	25, 25 μg/1hr	3 μg/1 hr both axes
x, y accelerometer scale factor	200, 200 ppm	60 ppm both axes
x, y accelerometer misalignment	20 arc sec	10 arc sec
x, y vertical deflection Markov (gravity anomaly)	5 arc sec/25 nm (uncompensated)	5 arc sec/25 nm (uncompensated)

TABLE 2

TYPICAL DOPPLER ERROR PARAMETERS

PARAMETER	CAUSE OF ERROR	TYPICAL (1 σ) RMS VALUE
along/across heading velocity scale factor	precision of manufacturer	.2% of ground speed
along/across heading velocity bias	precision of manufacturer	.25 kts along/across heading
along/across heading white noise	precision of manufacturer	.001 kt²/rad/sec/kt PSD along heading .01 kt²/rad/sec/kt PSD across heading .2 sec correlation time along/across heading
boresight bias	precision of manufacturer	.05 deg
heading bias	INS/magnetic compass	.05 deg
heading rate	INS	.02 deg/hr
along/across heading ocean current Markov	environmental	.4 kts/15 nm along/across heading
along/across heading wind-water surface motion Markov	environmental	3 3 kts/300 nm along/across heading

Deep Integration of GPS, INS, SAR, and Other Sensor Information

Zdzislaw H. Lewantowicz, Colonel, USAF
Deputy Director, Avionics Directorate, Wright Laboratory
WL/AA Bldg. 22
2690 C St., Ste 2
Wright-Patterson AFB, OH 45433-7409
USA

Randall N. Paschall, Ph.D., Major, USAF
Chief, Advanced System Integration Research Team
System Avionics Division, Avionics Directorate, Wright Laboratory

1. INTRODUCTION

This discussion is presented in the context of avionics sensor integration methodology. However the concepts developed and illustrations presented are in no way limited in application to military avionics systems. There is a growing number of civil applications, where information from multiple sensors is combined to improve performance, provide redundancy management, increase robustness, or achieve graceful degradation when sensor failures (or outages) occur. We are rapidly moving beyond the classical examples of sensor information integration such as in aircraft navigation, or in control of chemical processes, nuclear plants, and jet engines. For example, the automotive industry is using several sensors for engine emissions and fuel control, vehicle active suspension, and yes, *vehicle electronics* (*vetronics*). Communication systems, position, velocity, and attitude (rotation) sensors, and other information are rapidly appearing in the consumer automobiles. These sensors provide, at affordable cost, functional capabilities, which until recently were reserved for the more expensive commercial air and ground systems. Other applications of sensor integration are either being implemented, prototyped, or considered in toys, electronics, consumer products, such as the household washers and dryers, communications, and home environment monitoring and control systems. Although the sensor integration possibilities are expanding into these other domains, this discussion focuses on deep integration[1] of Global Positioning System (GPS), inertial navigation systems (INS), synthetic aperture radar (SAR), and other sensors which are a subset of modern aerospace systems.

Historically, the aircraft avionics have evolved from dedicated, single function, mechanical sensor assemblies to, more recently, sensor systems which have become sophisticated in their functionality and accuracy. These earlier sensors were developed, refined, and added to the aircraft as stand-alone (federated) devices which provided crew members with displayed information which increased their mission performance. Fusion of this information provided by a multitude of these avionics subsystems has historically been one of the required pilot (or crew member) skills. For example, instrument flying requires the pilot to methodically, rapidly, and effectively *cross-check* several instruments that make up the *performance* and *control* instrument groups. He must absorb this dynamically changing information and continually make real-time decisions about the type and quantity of control input(s) to achieve the desired/required flight path precision. In this case the pilot mentally performs the fusion of the various elements of information required for safe instrument flight. Incidentally the autopilot, when engaged, performs similar fusion and control decisions for a subset of the flying tasks. The fusion of displayed information by the human operator is an increasing challenge due to the increases in quantity and complexity of information provided in the cockpit.

In recent past either adequate computational and data bus technologies did not exist or the mission did not require the deep integration of information that the various sensors provided. The sensor performance and the computational and data bus capabilities have increased, and the mission requirements warranted some exchange of information outputs among the avionics black boxes. For example, INS velocity information was provided to the synthetic aperture radar (SAR) to enhance its motion compensation performance. These were the earlier stages of integration and could be called *limited integration* or *output integration*. Numerous constraints limited the quality, quantity, type, and other attributes of information in that limited integration information exchange.

In addition to the various computational and data bus constraints, standardization constraints also affect avionics sensor subsystem. For example, the USAF Standard INS has stringent specifications for not only the required physical fit and function but also for the digital output format, rate, and content, including the data word length, precision, structures, time tagging, and other related parameters. The output specifications, which were developed primarily to drive cockpit displays, are often not suitable for deep integration. Some constraints are imposed by the availability of technology, such as the speed limitation (capacity) of the MILSTD 1553 data bus while others are driven by the near-term focused specification process.

An early example of limited sensor integration is the central air data computer (CADC) for aircraft applications. The CADC combines several pitostatic and dynamic measurements of atmospheric conditions with other information, and provides altitude, airspeed, and rate-of-climb information corrected for several nonlinear distorting effects. The integration exemplified by the CADC is in combining of the dynamic and static atmospheric pressure-sensor information and enhancing it with relatively sophisticated, but open loop, compensation algorithms to provide higher accuracy outputs than those possible from the basic sensors. In other instruments the combining of information is nonexistent or limited. The GPS/INS integration, described in a later section is an example of limited integration. Of significance here is the fact that each sensor evolved primarily from the early necessity for stand-alone operation, that is without the necessity for communication of information to or from other sensors. Where information was needed to be shared with other sensors, the format of that information was typically constrained to that available at the post-processed output format specified by the primary sensor function. The more recently developed sensors have imbedded in them powerful computing capabilities which enhance sensors' output performance. In these cases, functional outputs from other sensors may be combined to enhance this performance.

Continuing rapid advances in processors, processing, detectors, networking, and other technologies are motivating the reexamination of the sensor information integration architectures and the integration of navigation information with that from other sensors. These advances in technology and rapid reduction in component cost provide opportunities for integration at sensor

---

1. Deep integration, for the purpose of this paper, is defined as a process which applies the system theory at integrated system level to the ensemble of individual sensor raw measurements.

signal levels deeper (closer to the *raw* signals) than just recently possible. This deeper integration is an unexplored area, especially when considered from a broader system level perspective, where the system might contain a variety of sensors, including those classically dedicated to navigation. While this assessment is based on information available in the professional literature, it is important to recognize that substantial work in this area has likely been accomplished and not yet reported.

In many avionics sensors the output signal is a result of substantial processing of raw signals. An example of a raw signal in this discussion might be the digitally coded radio frequency (RF) in a communication system such as the phase shift keying (PSK). Examples of *raw measurements* are: GPS receivers, pseudo- and delta-range measurements derived from the binary phase shift keying (BPSK) L_1 and/or L_2 RF signals, outputs of gyros and accelerometers in inertial measurement, or pixel level image elements in imaging devices. The acceptance of the GPS as a prominent system for navigation and the explosion of the receiver technologies, even before the full (24 space vehicle) constellation is deployed in 1994, are first motivators for examining deep integration. Additionally, the rapid processing capability advances, with the simultaneous reduction in cost, are no longer barriers to implementation of complex algorithms requiring substantial processing capacity.

Deep integration, for the purpose of this paper, is defined as a process which applies the system theory at integrated system level to the ensemble of individual sensor raw measurements. Essential ingredients of this deep integration process are proper (adequate) modeling, system synthesis and simulation, error analysis, and engineering trade-offs. A variety of approaches to modeling and analysis are available. Most frequently useful, in these types of analysis, are the error model approaches. One critical step in this deep integration process is the access to (or development of) validated *truth* error models. This *truth* error model becomes the foundation of system level analysis. [1, 2, 3, 4, 5, 6, 7, 8, 9].

Linear algebra is a rich linear system toolbox well suited for system level modeling, analysis, synthesis, and engineering trade-off analysis [7, 10, 11, 12, 13, 14]. In many system applications, linear system theory is widely used in the modelling, analysis, and synthesis purposes. For nonlinear systems, where linearization is practical, the same linear system tools are applied. One of the profound impacts on system level integration is the development of the recursive linear estimation theory. The most significant recent contribution to estimation theory is the publication by R. E. Kalman, in early 1960s, of his work in various articles, such as: *A New Approach to Linear Filtering and Prediction Problems* [15]. This work in discrete-time recursive estimation theory, combined with the advent of wider use of digital computers, is known as the Kalman filter theory and application. It has provided a powerful tool useful in system level integration. Although the Kalman filter is most widely used, other methods of integration, or sensor fusion, are used. This discussion is limited to linear estimation theory applications.

Estimation theory is an essential ingredient for optimally extracting the desired information given the random nature of the measurement information received by the sensors and which is inherent in processing of the information. The Kalman filter, in its various forms, is the most useful estimation algorithm, which is well suited for operation with linear systems tools [7, 16, 17, 18]. In this endeavor, it is important to properly (adequately) characterize the dynamic and stochastic error behavior of the sensor signals. Wherever possible, the least processed (least correlated) information should be operated on by the estimation (integration) algorithm, because sound error modeling becomes difficult for signals which have been modified with various filtering and decision algorithms. Theoretically it is possible to use processed (correlated) information, provided that the processing, which had modified the basic signal, is accounted for analytically. However this is not practical for a variety of reasons, including the difficulty of obtaining the actual processing algorithms which operated on (correlated) the particular signals. Frequently these algorithms include nonlinear processing, which can be troublesome, especially where processing jump discontinuities are present.

There exist various factors which cause barriers to deep integration in the aerospace vehicles. Federated integration architectures are barriers to deep integration since these architectures do not provide sufficient communication protocols and bandwidth and sufficiently precise timing required for deep integration. The federated architectures and sensor outputs were not designed to provide signals (measurements) with deeper integration in mind and constrain progress in deeper integration. The sensor outputs, of the current inventory sensors, are designed to drive cockpit displays and are inadequate for deep integration. The sensor output information is typically processed, (functionally correlated) through a variety of filtering, compensation, and transformation algorithms. The unprocessed (uncorrelated) signals are often available in the sensor "box." They are often created as intermediate variables in the implemented algorithms and stored in the sensor digital memory, but not made available on the output/data bus ports. These barriers to deep integration prevent achievement of maximum performance, robustness, and cost benefits possible with deep sensor benefits. Later sections illustrate these benefits.

This section discusses the historical evolution of sensor and computational technology. The analytic process and the tools required for deep integration are identified. This process consists of necessary modeling, simulation, analysis, and synthesis, and the linear systems and estimation theory domain tools. In later sections, several examples of modeling, simulation, and analysis of deep integration are presented. These illustrate the methodology with comments and interpretations. Simulation results, with empirical data comparisons (where available) are presented in later sections.

2. ANALYSIS, ESTIMATION, AND SIMULATION

In analyzing various engineering systems, it is normally necessary to represent the physical quantities involved with mathematical functions. There are many different types of functions one can use, and choosing an appropriate function usually is determined by the type of problem under study, and the dominant characteristics of the quantities involved. For example, a process may be either deterministic or random. A deterministic process is one whose future realization can be exactly predicted based on knowledge of past behavior, whereas a random process is one in which there exists a degree of uncertainty with regard to future activity. Typically, the more one knows about the governing factors of a phenomena, the more likely they are to use a deterministic description, and conversely, the less one knows about a process, the more likely they are to use a random description. The advantage of using deterministic models is the ability to use straightforward explicit mathematical formulas.

Another distinction in representing a physical quantity is the dimension of the problem. Depending on the complexity of the problem, and the level of detail required, the model may describe a single entity or a group of related quantities. This leads to the use of scalar versus vector representations. However, since all scalar problems are a subset of vector representations, we will only consider vector representations.

Many excellent texts are available which present linear system fundamentals [10, 11, 13], and the material found in this section can be found in any one of these references. This section will present only those concepts used in later sections to design integration algorithms for GPS, INS, and other sensors.

2.1 Linear Systems

A general dynamical system mathematical model is written

$$\dot{x}(t) = f(x, u, t) \tag{1}$$

with associated general measurement equation

$$y(t) = h(x, u, t) \tag{2}$$

The time-rate of change of state vector $x(t)$, $\dot{x}$, is a vector valued function f (not necessarily linear) of x, input u, and time t. The system states may not all be physically meaningful quantities, but any physical quantity of interest can be obtained from a linear combination of the states.

For any system, an infinite number of linear models exists, but the models are constructed so that the equations describing the system dynamics are combined into a unifying structure expressed by a matrix differential equation of the form

$$\dot{x}(t) = A(t)x(t) + B(t)u(t) \tag{3}$$

In many problems, the elements of the A and B matrices are constants, implying that the system is time invariant in addition to linear. This is not the case for the navigation systems considered in this paper. In particular, the elements of the A matrix are a function of the position, velocity, and acceleration of the systems.

The biggest reason for using linear models to approximate real world systems is the ability to solve the system of dynamic equations using relatively simple linear algebra techniques, and to achieve a closed form solution whose form does not depend on the nature of the inputs or initial conditions (as is often the case with nonlinear system models). For linear systems, the output equation can be formulated as the matrix equation

$$y(t) = C(t)x(t) + D(t)u(t) \tag{4}$$

In many problems of interest, $D(t)$ is often zero as is the case for the problems considered in this paper.

In the case where the system model is nonlinear, a commonly used approach is to linearize the system of equations about a nominal using Taylor series expansion techniques, and to keep perturbations from this nominal condition small so as to maintain the validity of the linear approximation [13]. When the nominal Equation (1) and the higher order terms (HOT) are subtracted, the linearized error perturbation equation provides

$$\delta\dot{x} = A(t)\delta x + B(t)\delta u + w \tag{5}$$

where $A(t) = \partial f(x, u, t)/\partial x$ and $B(t) = \partial f(x, u, t)/\partial u$ are Jacobian matrices which model the linearized error dynamics A and the control distribution B matrices. This perturbation differential equation in δx, for notational convenience is often expressed in terms of x, that is x is used in lieu of δx. When this substitution is made, then Equation (5) is in the form of the linear vector differential Equation (1). When both the whole value and perturbation quantities are discussed, the symbology ambiguity exists, and caution (or variable redefinition) is required.

Similarly, the linearized measurement Equation (2) is

$$\delta y = H(t)\delta x + D(t)\delta u \tag{6}$$

where $H(t) = \partial h(x, u, t)/\partial x$ and $D(t) = \partial h(x, u, t)/\partial u$ evaluated at the nominal conditions provide the linearized measurement function. The H matrix contains the geometric, and other relationships, between the sensor measurements and the state vector variables. Thus Equation (6) is in the form of linear measurement Equation (2). In many applications, the control matrix D is zero, stating that the control has no direct effect on the measurement.

Because all linear models can be put into a common form, and thus the form of the solution is the same in every case, one can use standard computer aided design packages to perform system analysis regardless of the system under study. This allows integrators to use a variety of commercially available products to conduct design and analysis with. The form of the equations

are also easy to transform into discrete algorithms for microprocessor implementation. Computer programs to perform analysis for nonlinear systems are not as easy to use nor as simple to implement. Performing parametric studies becomes extremely challenging when nonlinear models are used.

2.2 Example of Dynamic Analysis Utility

Once a model is in the form of Equation (3), the stability of the system can be easily analyzed by computing the complex eigenvalues of the A matrix. If the eigenvalues all have negative real parts, the linear system is asymptotically stable. The stability of a linear system is not a function of the model chosen, but rather the eigenvalues remain the same between any models related by a similarity transformation. This provides additional design freedom to the integration engineer. Stability may be time varying as in the case when the A matrix (and thus the eigenvalues) is a function of time. In this case one may need to study the eigenvalue migration as was done by Cunningham [19] where he explored the eigenvalue migration characteristics of the INS during various maneuvers. The resulting eigenvalue migration is depicted in the next three figures.
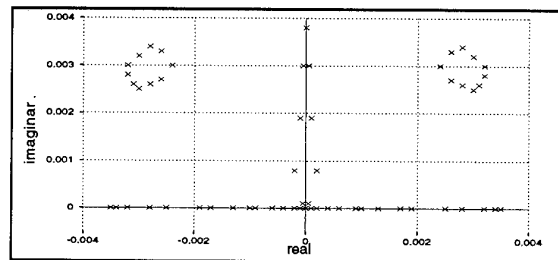
Figure 1. Horizontal (Level) Turn Eigenvalue Migration
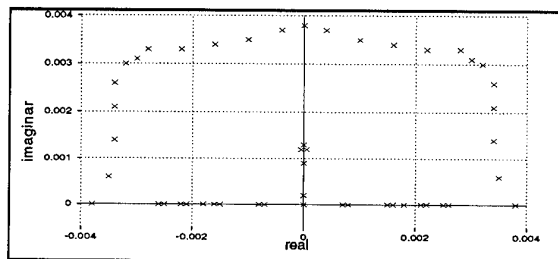
Figure 2. East Heading Loop Eigenvalue Migration
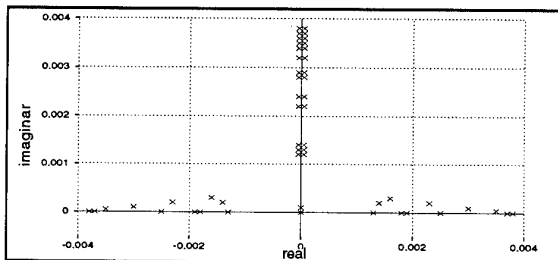
Figure 3. North Heading Loop Eigenvalue Migration

In Figure 1 the important fact is the general location of the eigenvalues during the maneuvers. Note that at least one complex pair of eigenvalues is always in the right half plane. The eigenvalue on the positive real axis is due to the vertical channel instability. The eigenvalues for the vertical maneuvers have different characteristics. In the maneuver which has an initial east heading, there is a complex pair of eigenvalues which breaks away from the imaginary axis, migrates to the real axis, and returns to the imaginary axis. For the maneuver with an initial north heading, only eigenvalues associated with the vertical channel appears in the right half plane. This indicates that east/west maneuvers result in an additional pair of unstable eigenvalues. This type of analysis lends useful insight to an engineer as to which type of maneuvers will cause the greatest problems to

the systems under test. This information allows one to fully test conditions that might otherwise be overlooked until too far into the development cycle. A nonlinear system's stability is a much more complicated issue and cannot be analyzed as easily. One is often forced to talk about stability in the sense of Lyapunov or bounded input bounded output stability. This is another reason to use a linear model whenever possible.

In addition to stability, one can easily compute the eigenvectors of a linear system to gain geometric insights into the system behavior. Eigenvectors form a basis vector set which spans the n-dimensional state space. There is an eigenvector $\zeta(t)$ associated with each eigenvalue λ_i, for $i=1,...,n$. This, when coupled with Equations (3) and (4), yields insights into the controllability and observability directions of a given system. Knowledge of the eigenvectors coupled with the system matrices gives one insight into the limitations imposed by a given linear model. One must realize that observability and controllability are not physical properties of a system as are eigenvalues and stability. Controllability and observability are determined by the form of linear model chosen to represent the physical system [13]. When using a Kalman filter, it is necessary to use a model where the states to be estimated are observable. States which are not observable cannot be estimated accurately by the Kalman filter.

For time varying systems, the observability and controllability of a system is determined by the rank of the Grammians defined

$$W_o(t, t_0) = \int_{t_0}^{t} \Phi^T(\tau, t_0) H^T(\tau) H(\tau) \Phi(\tau, t_0) \, d\tau \quad (7)$$

$$W_c(t, t_0) = \int_{t_0}^{t} \Phi(\tau, t_0) B(\tau) B^T(\tau) \Phi^T(\tau, t_0) \, d\tau \quad (8)$$

The rank may change as a function of time which is useful to determine. For example, one may wish to address the question of what type of geometry or dynamics may enhance the observability of states of interest. For large dimensional problems which are time varying, the evaluation of these Grammian functions is computationally intensive. For this reason one often looks at specific instances in time, evaluates the Φ, B, and H matrices of the system and for those times the invariant Grammian matrix. This is an easier computation to perform, given the available software products. By evaluating the rank of this matrix, one can determine the observability (or controllability) of the system at a given time. To get some idea of how the observability changes with time, several of these evaluations must be made at different time instances.

Another property of linear systems that aids in the analysis is the property of linear superposition which says that the response of linear system to multiple inputs is simply the sum of the responses to individual inputs. This allows one to study the response of a system to one input at a time for parametric evaluations, and then to make performance evaluations to combinations of inputs without performing additional analysis. This is not the case for nonlinear systems where the response depends on the exact form of the input and changes unpredictably depending on what input combinations are applied.

One can also analyze linear systems using frequency domain techniques as well as time domain techniques. Using Laplace transforms, a system described by Equations (3) and (4) can also be described in the Laplace domain

$$x(s) = H[sI - A]^{-1} Bu(s) \quad (9)$$

The matrix $H[sI-A]^{-1}B$ is the transfer function matrix of the system. Using Bode, Nyquist, or Nichols charts one can graphically evaluate items such as stability margins, bandwidth, settling time, peak overshoot, and crossover frequency. These issues are important when one starts to question how well a system performs in the presence of external disturbances, in various dynamic environments, and over frequency ranges of interest. As an example, in a GPS receiver one must decide on

the tracking loop bandwidth which affects the receiver's ability to reject jamming signals. If one chooses a low tracking bandwidth to give maximum jam resistance, then the dynamic range over which one can maintain track is reduced. Similar questions for nonlinear systems are very difficult to analyze.

If a linear model cannot be obtained, which accurately represents the behavior of the system being modelled, then one is forced to use nonlinear techniques, and the problem becomes much more difficult to manage. On the other hand, even if a good linear model can be achieved, it is critically important to perform system level evaluations using simulations which incorporate nonlinear effects and thus more accurately reflects the real system's behavior. This is important to recall when one is developing a truth model as opposed to a design model or an operational filter model. This is the real reason behind system testing using real hardware. The actual system contains all of the nonlinear affects that were not modelled in a linear system design and analysis. It is common for engineers to not use nonlinear truth models since they are difficult to understand and develop. This leads to a linear truth model which ignores the nonlinear affects, often leading to unexpected outcomes in system testing. One must always keep in mind the limitations of the linear models used.

2.3 Stochastic Models

Section 2.1 discusses mathematical models useful in the analysis of deterministic dynamical systems. In the physical world there are no known deterministic phenomena. Thus the mathematical model developed in that section must be modified to account for the stochastic variable behavior and the corruption of the measurements by random processes.

In general, the geometric information from each (or all) sensor outputs can be modeled by a set of *stochastic differential equations* of the form of

$$\dot{x} = f(x, u, t) + w(t) \quad (10)$$

where w is an additive random driving noise vector which is a zero-mean, white Gaussian noise process with covariance kernel [7, 16, 17, 18]

$$E\{w(t) w^T(t+\tau)\} = Q(t)\delta(\tau) \quad (11)$$

The measurement vector y, here in the discrete-time sampled measurements form, follows the form of Equation (2)

$$y(t_i) = h[x(t_i), t_i] + v(t_i) \quad (12)$$

These output measurements are corrupted by additive white Gaussian noise vector v of mean zero and covariance kernel

$$E\{v(t_i) v^T(t_j)\} = \begin{cases} R(t_i), t_i = t_j \\ 0, t_i \neq t_j \end{cases} \quad (13)$$

The pair of Equations (10) and (12) is sufficient to describe the dynamic and stochastic behaviors of the sensors. The noise statistics of w, and v, which represent the actual sensor error behavior must be provided *a-priori*. This set of stochastic dynamic and measurement equations provide the framework for modeling and integration of sensors from a system level perspective.

2.4 Estimation Concepts

Whenever one attempts to integrate two or more navigation sensors, a choice must be made as to the type of integration algorithm to use. A Kalman filter is an optimal recursive algorithm (estimator) that processes measurements from sensors in order to compute a minimum error estimate (in a well defined statistical sense) of the state of a system. It does so by taking advantage of knowledge of systems dynamics, the statistics of the system error sources and measurement errors, and available initial condition information. A Kalman filter can also be operated in a batch mode, but recursive operation is the mode presented

in this section. This section briefly presents the fundamentals of the Kalman filter in order to enhance the readers understanding of the integration results presented in later sections. For a more complete presentation of the theory, the reader should refer to any of several excellent texts on the subject [7, 16].

A Kalman filter may be used to perform one of three different types of estimation operations. The operation of filtering refers to the situation where the time at which an estimate is computed corresponds to the time when the last measurement is provided. The operation of smoothing is when the filter estimate corresponds to a previous time when measurements were provided, and prior to the present measurement. The operation of prediction involves performing an estimate corresponding to a time after the last available measurement. This section limits its discussion to the filtering operation.

A Kalman filter computes estimates of states which cannot be accurately measured (perhaps not measured at all), but which is observable (as defined in Section 2.1). The Kalman filter has several underlying assumptions. First it is assumed that the system and measurement equations are linear. When the system and/or measurement equations are not linear one uses a linearized or extended Kalman filter. Second, it is assumed that the noises that act as forcing functions to the system are white. This implies that the noises are uncorrelated, and have equal power at all frequencies. Third, it is assumed that the noises have Gaussian probability distributions. None of these assumptions are completely satisfied for actual systems, but in many cases they are very good approximations. Because of concepts such as the central limit theorem, shaping filters, systems bandwidths, and because one often knows the mean and variance of the noise processes, the assumptions underlying a Kalman filter are not very restrictive in application [7].

The general problem for a Kalman filter to solve can be cast as follows. Suppose there are certain states associated with a system which one cannot determine directly and exactly, but for which there are measuring sensors that provide measurements corrupted by white Gaussian noise and which are functionally related to the states of interest. The Kalman filter computes an estimate of the states based on the measurements available, but to do so requires basically three elements. First one must have a mathematical model which describes the relationship between the states and the measurements. Measurement errors are modeled by a noise vector v. When measurement uncertainties are non-Gaussian, shaping filters driven by white noise are used to augment the state vector. Second, a dynamics model is required that describes the time varying nature of the states. This model uses a noise term Gw to account for model uncertainty and random behavior of the states.

These measurement and dynamic models describe the a-priori knowledge of the states as well as the statistics of the noises. In developing these models it is assumed that the dynamics noise w and the measurement noise v are uncorrelated. The final element associated with the Kalman filter is the performance index to be minimized. The Kalman filter is based on a Bayesian viewpoint. The estimates of the states are based on all measurements which is reflected in the conditional probability density function written as

$$f_{x|z}(\xi|z) \qquad (14)$$

There are three possible measures of optimality: median, mode, and mean. If the density function described by Equation (14) is Gaussian, then all three measures of optimality are satisfied by the Kalman filter. When the measurement noises are statistically independent (i.e. uncorrelated) then the matrix R defined by

$$E\{v(t), v^T(t-\tau)\} = R\delta(t) \qquad (15)$$

is diagonal. If however the noises are correlated, then R is not diagonal. This degree of freedom is usually avoided in order to simplify the computations. However, by accounting for correla-

tions which are present, the estimation performance of the filter is enhanced. The same is true when one considers the system dynamic noises.

The Kalman filter does not have knowledge of the actual noises except for the statistics provided. This is especially important to recall when filter models are being tested against truth models in simulation analysis. The truth model will include all known noises, and should be of a higher order than the Kalman filter. Along with the state propagation equation, there is a propagation equation for the covariance matrix. It is this equation one evaluates when performing covariance analysis.

2.5 Linearized and Extended Filters

In many cases, especially those involving navigation sensors, a linear model does not provide an adequate model of the system under study. In this case the system is described by a set of nonlinear differential equations as discussed in Section 4. Given a nonlinear model it is still desired to produce an optimal estimate of the states. To do this, one can assume that a nominal state trajectory can be generated. Associated with the nominal trajectory is a nominal measurement

$$z(t_i) = h[x(t_i), t_i] + v \qquad (16)$$

The measurement may be a nonlinear function of the states. Using Taylor series perturbation techniques (see Section 4.1) the perturbation equations are derived

$$\delta\dot{x} = A[x, u, t]\delta x + B[x, u, t]\delta u + Gw \qquad (17)$$

$$\delta z = H[x, t]\delta x + D[x, u, t](\delta u) + v \qquad (18)$$

Using these linearized equations one can now apply linear filter theory. The input measurement for such a filter is the difference between the actual measurement and the nominal measurement. The output of such a filter is an optimal estimate of δx. To establish an estimate of the state, one simply forms the total state estimate as

$$\hat{x}(t) = x_n(t) + \delta\hat{x}(t) \qquad (19)$$

This form of the Kalman filter is called a linearized Kalman filter [7]. It is computationally efficient compared to a nonlinear filter, but it can experience large errors if the nominal state trajectory differs greatly from the true trajectory. This leads to the use of an extended Kalman filter.

The basic idea of the extended Kalman filter is to evaluate the Jacobian matrices about the last estimate once it is computed. As soon as a new estimate is produced, a new state trajectory is computed and included in the estimation process. In this way the validity of the small perturbation assumptions inherent to the Taylor series approximation is maintained. The extended Kalman filter is defined by the update equations given as follows [7]

$$\hat{x}(t_i^+) = \hat{x}(t_i^-) + K(t_i)[z_i - h[\hat{x}(t_i^-), t]] \qquad (20)$$

$$P(t_i^+) = P(t_i^-) - K(t_i)H(\hat{x}, t_i)P(t_i^-) \qquad (21)$$

and the propagation equations are

$$\dot{\hat{x}}(t) = f[\hat{x}(t), u(t), t] \qquad (22)$$

$$\dot{P} = F[\hat{x}, t]P + PF^T[\hat{x}, t] + GQG^T \qquad (23)$$

Notice that in the extended Kalman filter, the propagation is done using the nonlinear equations evaluated about the latest estimate. Also, notice that the update equation for the state involves the nonlinear relation described by $h[x,t]$. These are important differences between the extended filter and the linearized filter. The extended filter is used in this paper to integrate GPS with other sensors. This is due to the fact that the measurement is nonlinear for the navigation sensors.

236

2.6 Simulation Tools

Software which explicitly calculates time-varying satellite position for a full GPS constellation was written, and analyses were conducted using the *Multimode Simulation for Optimal Filter Evaluation* (MSOFE) [20]. It is referred to as multimode in the sense that it is capable of simulating a variety of system problems. Its two most widely used modes are Monte Carlo and Covariance simulations. MSOFE implements the upper-diagonal (U-D) form of the extended Kalman filter algorithm [20, 7]. MSOFE is used to establish a performance baseline against which other full-order and reduced-order filters may be compared. Time histories of variables, such as aircraft position in R^3 space, attitude, and acceleration, are created and serve as the *true* aircraft trajectory in all simulations conducted and presented in this research.

Profile generator (PROFGEN) is a trajectory generation program developed by the Avionics Directorate [21]. It computes position, velocity, acceleration, attitude, and attitude rates for a maneuvering air vehicle profile based on user provided inputs. This flight trajectory is directly interfaceable to MSOFE simulation. These data are used in the MSOFE simulation as *nominal* quantities about which the truth model is relinearized after each integration step [7]. The extended Kalman filter is relinearized about its best estimate of the trajectory.

3. GEOMETRIC SENSORS AND ERROR MODELS

Among the avionics sensors aboard a typical aircraft, most contain common information in the geometric domain, that is these sensors are geometrically related. Their measurements contain some, or all, the elements of geometric information which is defined as a set consisting of position vector (r), orientation (attitude) parameter array (ϕ), and time scalar (t). The orientation of the vehicle can be represented in any of several orientation parameter sets; the direction cosine matrix (DCM), quaternions, Euler angles, and *rotation vector*[2]. Two most commonly used are the DCM and the quaternion because these offer singularity-free representations [22]. The orientation array $\underline{\phi}$, can represent any chosen orientation parameter set. These are related and can be transformed between the sets. The rotation vector is a useful geometric concept and is the choice for this discussion. Thus the symbol ($\underline{\phi}$) represents the rotation vector in the following discussion. Temporal and/or spatial derivatives of these information elements may also be directly or indirectly available. For many applications it is convenient to use the derivative of position, that is the velocity vector v explicitly.

3.1 Geometric Vector Concept

A ten-element array G of geometric parameters is defined:

$$G = \begin{bmatrix} r & v & \underline{\phi} & t \end{bmatrix}$$
$$= \begin{bmatrix} r_x & r_y & r_z & v_x & v_y & v_z & \phi_x & \phi_y & \phi_z & t \end{bmatrix} \quad (24)$$

Although this array does not always satisfy all vector properties as discussed above, it is referred to as the *geometric vector G* and denotes the represented physical variables. To be useful in mathematical and algorithmic calculations, this vector must be coordinatized in some reference frame, such as body *b*-frame, navigation *n*-frame, earth *e*-frame, inertial *i*-frame, etc. For definitions and discussion of reference frames see [1, 2, 3, 4, 5, 6, 7, 8, 9, 23]. Of significance, for this discussion, is the fact that this geometric vector can be used as an absolute or a relative reference. An example of an absolute geometric vector might be the position and velocity of a vehicle with respect to

the earth center, attitude with respect to the earth surface, and the Greenwich Meridian Time (GMT). When two systems, each represented by a geometric vector G_1 and G_2 respectively, operate relative to each other, we can define a *relative geometric vector* ΔG

$$\Delta G = G_1 - G_2 \quad (25)$$

Most every sensor aboard an aerospace system, provides an output containing some or all elements of G or ΔG. For example, an INS provides r, v, and $\underline{\phi}$. The GPS provides r, v, and t, while a crude approximation of $\underline{\phi}$ can be extracted for a receiver which is not maneuvering rapidly. In another example, a modern aircraft radar system has the capability of providing various forms of information accumulated along the line of sight, however the basic information available from most radar systems consists of the LOS azimuth and elevation (ψ,θ) and/or range R. This information is derived from elements Δr and $\Delta\underline{\phi}$ of the basic ΔG, where G_1 is that of the radar carrying platform and G_2 is the target. The presence of G and/or ΔG components in these sensors is the foundation for deep integration. Because these sensors refer to the same physical geometric elements, the measurements that these sensors provide are thus spatially correlated.

3.2 Error Model Format

The spatial correlation, between the various sensor outputs, discussed in Section 3.1 is mathematically modelable in the context of linear systems theory and a stochastic process. This model can be processed in a context of a Kalman filter, where the off-diagonal elements of the filter covariance matrix contain the cross-correlations representing the spatial correlation. This cross-correlation between the sensors is the cornerstone of deep integration benefits. By proper modelling of all relationships and processing in a single (joint[3]) Kalman filter, the optimal (best possible) estimate of the cross-correlated errors is possible. This can be thought of as a mutual (cross-) calibration process of the various sensor errors which share the same physical geometric domain (or have a known relationship with the same geometric vector). Again, this cross-calibration is limited by the quality of the dynamic and stochastic error models, the relative magnitudes of the uncorrelated noises among the sensors, and the noise (both measurement and driving noises) magnitudes relative to the signal strengths. The truth model analysis establishes the best possible cross-calibration performance. The performance of the subsequently synthesized reduced-order filter can be compared to that of the truth model. The next discussion presents key linear algebra tools and concepts in modeling of the sensors and their error behavior.

3.3 Sensor Classes

The quantities *position, velocity, attitude, range, line-of-sight, and time* form the basic set of variables measured by sensors. These quantities are either elements of the geometric vector G, or the relative geometric vector ΔG. The sensors aboard an aircraft can be basically classified into the following categories: *inertial, direct position/velocity/attitude,* and/or *relative position/velocity/attitude.*

A particular sensor may provide one or more of the G or ΔG elements. The differences between sensors providing these variable outputs are in the quality of the measurements and their environmental capabilities. For example an RF radar penetrates atmospheric fluids, however a laser radar, with more restricted propagation due to moisture and other attenuating matter, provides a much more accurate measure of both the LOS and range. The key point is that for deep integration, the structure of the geometric and error models for these LOS sensors are

---

2. Caution is required in interpreting the orientation parameter array as a vector, as it does not satisfy the vector property of commutation, although, for small rotations, the *rotation vector* does satisfy vector properties, to first-order approximation [22].

3. Joint in the joint probability density function sense [7, 16, 17, 18].

similar. The following discussion presents the basic principal of operation for each sensor class, followed by a structure for the error model. The emphasis is on those aspects of error models that relate to the elements of absolute or relative geometric vector. This is not limiting nor exclusive. If other, non-geometric but related, quantities are of interest in a particular system design, then the models need to be expanded to those domains and appropriate analytic tools applied.

The goal in integration of a set of sensors is to take advantage of the fact that all of the sensors share a common geometric reference. They all experience the same reference position, velocity, and attitude at any instant of time. Even when the sensors may be "looking" in different directions, and their lines-of-sight might be varying, their origins are either fixed with respect to each other, or are varying at some rates. These relative sensor orientations are known, or measured, except for relative errors, such as installation errors, boresight errors, randomly varying errors, and others.

3.3.1 Sensor Truth Models

Sensor errors can be characterized as random processes, representable in the context of system level models. If linearization and Gaussian uncertainty model assumptions are valid, then linear estimation theory is applicable. When all of these sensors are completely (adequately) modeled in their dynamic and stochastic domains, then these models can be assembled in a single, joint error model. This joint model then becomes a *truth model*. Estimation theory is then applied at the truth model level where Kalman filter is the tool of choice for linear/linearized error models. Various desirable outcomes are immediately available.

First, the Kalman filter computes the joint estimate of the mean $\hat{x}$ and covariance matrix P which contains the cross-covariances in the off-diagonal elements [7, 16, 17, 18]. These off-diagonal elements represent the statistically (and dynamically) derived error dependencies among the various sensors. These cross-correlations, through the sharing of the common geometric vector, produce in the filter the error estimates, or *sensor cross-calibrations*.

Second, the truth model system-level analysis provides the baseline for the best possible system level performance. This is useful for various trade-off studies, engineering designs, sensor level parameter trade-offs, baselining of contributions of each sensor to the system level solution, and other analyses.

Third, the truth model becomes the reference system infrastructure for suboptimal filter design. Various system level trade-offs in the filter design, filter tuning, and other sensitivity studies are well performed at this level.

3.3.2 Sensor Error Truth Models

The most important aspects of this process is the availability of sufficiently correct sensor error models represented in the dynamic and stochastic domains. The error model development must apply sound scientific method. The necessity for sufficient correctness of these models can not be understated.

3.4 Inertial Navigation System (INS)

A key sensor available in many airborne systems is the inertial navigation system (INS) [1]. It is an amazing example of advances in various technologies from improvements in gyros and accelerometers, and gimbal platform performance, to computational capability explosion, strapdown systems, and system level integration. These improvements were driven by two key historical events requiring accurate navigation. The ballistic missile and the space programs of the 1960's era played a key role. The endeavor to the moon by the Apollo program was a noteworthy milestone in that evolution. Scientific process applied to error model development produced essentially what

is known today about inertial system error behavior structure [1, 2, 3, 4, 5, 6, 7].

Navigation, in general, is the determination of a physical body's position and velocity relative to some reference coordinate frame or coordinate grid. The case of terrestrial navigation involves the determination of a vehicle's position and velocity relative to the earth. The grid coordinates usually used for this application consist of the spherical coordinates, latitude, longitude, and altitude. An INS utilizes the inertial properties of sensors mounted aboard the vehicle to execute the navigation function. The system accomplishes this task through appropriate processing of the data obtained from specific force and inertial angular velocity measurements.

All INSs must perform the following functions:

  1) *Instrument a reference frame, e.g. n-frame*
  2) *Measure specific force, f*
  3) *Have knowledge of the gravitational field, model* g
  4) *Time integrate specific force to obtain velocity and position*

The first function is accomplished by the use of gyroscopic instruments. Typically three gyros are mounted, with their sensitive axes mutually orthogonal, on a rigid platform whose orientation is measured and tracked by either mechanical or digital computer closed loop integration and control. The gimbaled inertial platform can be mechanized to instrument, for example, the local geographic coordinates, such as the commonly used north, east, and down navigation n-frame. If three structurally mounted gyros are used, the relative orientation between the initial and present vehicle coordinate frame can be determined. Systems which analytically instrument a reference frame are popularly referred to as *strapdown* (strapped-down) systems [1, 2, 3, 4, 5, 6, 8, 9]. The current inertial navigation systems utilize the strapdown configuration, where typically ring laser gyros (RLG), hemispheric resonating gyros (HRG), fiber optic gyros (FOG), or other gyros, provide the angular orientation information.

The second function, the specific force measurement, is accomplished with devices commonly called accelerometers. Although there are many ways of measuring specific force, most of the devices in common use are sophisticated variations of the simple pendulum. The motion of the pendulous element is related to the motion of the platform, or structural element upon which the accelerometer is mounted, via Newton's second law of motion. Three accelerometers with sensitive axes typically in mutually orthogonal orientation are required to measure the three components of specific force.

3.4.1 INS Sensor Model

The inertial sensors provide measurements of inertially referenced Newtonian motion. The specific force f and angular velocity $\underline{\omega}$ vectors are measured by accelerometers and gyroscopes, respectively. The specific force vector f, consists of gravitational force vector g and the reactive force vector due to inertial acceleration $p_i^2 r$

$$f = p_i^2 r - g$$
$$= p_i v - g \qquad (26)$$

where p is the time derivative operator, $p_i^2 r$ is the acceleration (second time derivative of position) with respect to inertial frame i. Gravity g must be computed and subtracted from f in order to extract the inertial acceleration. This satisfies the third required function of an INS. With proper computations, information from the gyros provides the platform orientation, where the platform can now be mechanized to represent a desired reference frame. Equation (26) is coordinatized in a selected reference frame, e.g. the navigation frame n, and solved for earth referenced acceleration (rate of change of earth referenced velocity) v^n,

238

$$v^n = \begin{bmatrix} \dot{v}_N^n \\ \dot{v}_E^n \\ \dot{v}_D^n \end{bmatrix} = \begin{bmatrix} f_N - v_e\left(\dot{l} + 2\omega_{ie}\right)\sin L + \dot{L}v_D \\ f_E + v_N\left(\dot{l} + 2\omega_{ie}\right)\sin L + v_D\left(\dot{l} + 2\omega_{ie}\right)\cos L \\ f_D - v_E\left(\dot{l} + 2\omega_{ie}\right)\cos L - \dot{L}v_N + g \end{bmatrix} \quad (27)$$

The product terms in Equation (27) represent the accounting for the nonlinear Coriolis and centrifugal force terms. The processing includes double integration of the acceleration information, starting with appropriate initial conditions, to provide the desired position and velocity. The gyro outputs are given in terms of an attitude variable differential equation as illustrated by the direction cosine matrix (DCM) representation of a strapdown system

$$\dot{C}_p^c = C_p^c\left(\Omega_{ip}^p - \Omega_{ic}^p\right) \quad (28)$$

where C_p^c represents the relative orientation of the platform p-frame with respect to the computation c-frame, Ω_{ip}^p is the skew-symmetric form of the angular velocity vector between the inertial i-frame and the platform p-frame coordinatized in the platform p-frame, and the Ω_{ic}^p represents the angular velocity between the computation and inertial frames [1]. Similar differential equations exist for the other orientation parameter sets. For detailed discussion see any of the following: [1, 2, 3, 4, 5, 6, 8, 9], with [1] providing the best analytic discussion with complete error analysis methodology. Equations (27) and (28) are in the nonlinear stochastic differential equation form of Equation (10). These two vector differential equations form a set of nonlinear, coupled, second and first order differential equations, respectively, which are easily convertible to a set of nine coupled, first order differential equations and form the foundation of the INS relationships. The initialization of an INS includes the initialization of the orientation parameter sets through a self alignment process using the information provided by the measurements of the gravity vector and the earth rotation vector [1]. These equations, when integrated from a set of initial conditions, provide continuous navigation solution in the desired reference frame. However due to errors in the initial alignment, gyros and accelerometer, and other sources, the INS errors are nonzero and grow in time.

3.4.2 INS Stochastic Error Model

Britting [1], and others [2, 3, 4, 5, 6], develop error models for various INS mechanizations. The basic INS error model is represented by nine fundamental variables, three position, three velocity, and three orientation (attitude) errors. The errors form an intricately coupled set of nine, linear differential equations which are characterized in their dynamic behavior be nine eigenvalues, corresponding to the dynamic modes of the system. These error dynamics are the linearized representation (the $A(t)$ matrix of Equation (5)) of the nonlinear, coupled differential Equations (27) and (28). The gyro, accelerometer, instrument misalignments, component transients, and other errors are added to the basic nine error equations.

When the full error truth model is assembled it can exceed 90 states. There exist various versions of error models, with primary differences resulting from the choices of the coordinate frame. For additional discussion on the error model development see references such as [1, 3, 23]. The Litton Guidance and Control Systems LN-93 INS error model is used in each illustrative analysis described in later sections.

3.4.2.1 Litton LN-93 INS Error Model

In the class of navigation sensors, the INS error model is likely the most complex in structure and in the extensive modelling work that has been required. Significant physical error sources were identified and modeled and the components were extensively tested to validate the error models. This process is the most significant step in any sensor error model development.

The generalized INS error state equation is of the form [1, 7] (with symbols defined in Section 9.)

$$\delta\dot{x}(t) = A(t)\,\delta x(t) + Gw(t) \quad (29)$$

The nine *general* error states describe the fundamental INS behavior [24]. Britting [1], (and others) analytically demonstrates that an invariant property of the error dynamics is represented by three fundamental oscillation modes. Errors of a stationary INS consist of the classical oscillations called Schuler (84.4 minute period), Foucault (24 hour period), Foucault-modulation of Schuler (projection of earth rate onto Schuler), and the vertical channel instability. Elegant discussion of these topics is found in the referenced texts and literature. Other error states coupling into these *general* states, grouped by categories, are added.

The Litton Guidance and Control Systems LN-93 INS error model is used in the analyses presented. It defines 93 error states (see [24] and Appendix). The Litton LN-93 documentation defines the earth e-frame, the true t-frame, the computer c-frame, the platform p-frame, the sensor s-frame, the gyro g-frame, the accelerometer a-frame, and the body b-frame [24]. For discussion of INS errors, the reader is referred to any good text on inertial navigation systems, such as [1, 2, 3, 4, 5, 6, 8, 9, 22].

The LN-93 is a strapdown, local-level, wander azimuth INS. Position is specified in terms of latitude, longitude, altitude, and wander angle. Generally, wander angle is the angle about the navigation frame z-axis which results when the INS platform is not torqued to maintain precise (ENU) orientation. In this mechanization, the platform is *torqued* only in the *level* axes in order to maintain the local level orientation. Since the LN-93 is a strapdown mechanization the *"torquing"* is performed in the INS computer software. The WGS-84[4] model is used as the ECEF reference for this effort.

Random constants (biases) represent variables which may take on values within some specified probability distribution and thereafter retain this value [7]. Approximately 64 bias error states are classically lumped into the four subcategories: gyro bias, accelerometer bias, trending, and gyro compliance states.

First-order Markov processes describes the gyro and accelerometer drifts and thermal transient errors characterized by random, but time correlated, *wandering*. This process is modeled by a first order lag driven by zero-mean white Gaussian noise [7]. Litton partitions the error states into the six subvectors [24]

$$\delta x = [\delta x_1^T, \delta x_2^T, \delta x_3^T, \delta x_4^T, \delta x_5^T, \delta x_6^T]^T \quad (30)$$

δx_1 represents the *general* error vector containing position, velocity, attitude, and vertical channel errors. δx_2 consists of gyro, accelerometer, and baro-altimeter correlated errors, and *trend* (first-order Markov) states. δx_3 represents gyro bias errors modeled as random constants in the truth model and are (later) modeled in the Kalman filter as random walks with small magnitude pseudo-noises. δx_4 is the accelerometer bias error state vector. δx_5 depicts accelerometer and gyro initial thermal transient (first-order Markov) states. δx_6 models the gyro compliance errors as biases in the system model and as random walks in the Kalman filter.

The Litton LN-93 error model is inadequate in its depiction of the baro-altimeter error sources in that it contains only a single state for the barometric altimeter. Other authors discuss several error sources normally linked with baro-altitude [25]. A revised baro-altitude error model, developed and embedded in the INS error model, includes states for correlated noise effects, bias error, and scale-factor error [25, 26]. These error sources are

---

4. GPS positions are based on an earth-centered, earth-fixed (ECEF) Cartesian coordinate frame designated as the WGS84. This reference frame designates the origin at the center of earth mass having the z-axis passing through the instantaneous pole at the year 1984.

combined in a more complete error model. For the analyses performed and reviewed in the following sections the 98 states are reduced to 68 states, with the added baro states retained in tact. These states are listed in the Appendix.

The vertical INS channel instability requires aiding by an additional altitude (or height) sensor. Classically, the baro-altimeter has been used successfully in bounding this instability. Although the baro-altimeter will remain aboard the aircraft for a long period of time, today the GPS can also superbly perform the vertical channel aiding function. For completeness, the baro altimeter model is discussed in Section 3.5.

3.4.3 *Other Inertial Sensor Applications*

Inertial components, particularly accelerometer and gyros, are used in a variety of applications other than the inertial navigation systems. They may be mounted at various locations throughout the aircraft to provide the vehicle acceleration or rotation with respect to the earth frame. In these applications the gravitational force components must be subtracted from the accelerometer outputs, while the centripetal, coriolis, and tangential acceleration forces may be ignored due to their small relative magnitudes. Often a flat, non-rotating earth model is assumed as in an aircraft flight control system. This paper does not elaborate on such applications, only mentions them for completeness. The current trend is to include more of such information for purposes such as boresight error reduction and flexure control. When information from such sensors is used in system-level integration appropriate sensor models are required. The reader is referred to various texts, e.g. [23], and articles, e.g. [27, 28], for these sensor models.

3.5 Barometric Altimeter

Barometric altimeter provides an indication of altitude by measuring the static air pressure. Altimeter output is a robust bounding signal to the INS vertical channel. Like most pressure sensors, the altimeter transducer typically consists of an aneroid bellows with mechanical linkages to a dial or drum display. This sensor is rather inaccurate due to several factors. The relationship between the static air pressure and altitude depends on current weather conditions. The altimeter must be compensated with a bias input (*altimeter setting*). Transducer time lag, instrument bias, and scale factor are significant error sources. The pilot and aircraft system need for accurate air data (speed, Mach number, altitude, rate of climb) drives the trend to use air data computers.

3.5.1 *Barometric Sensor Model*

Typical central air data computer (CADC) provides a more accurate static error corrected pressure altitude encoded for automatic altitude reporting and in analog form for pilot display. Even the CADC corrected altitude is of limited accuracy. The operational tolerance for the altimeter output error on the ground is ±75 feet, and ±300 feet at altitude. Figure 4 illustrates a typical CADC altitude and error model. The variables in Figure 4 are defined in the next section and Section 9.
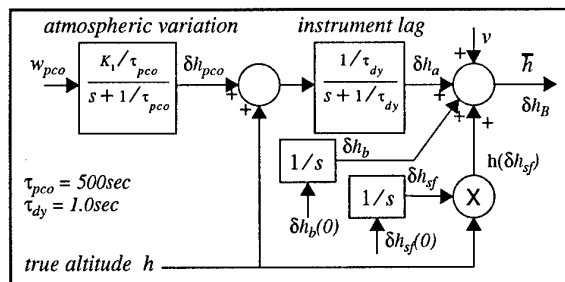

Figure 4. CADC Stochastic Model

3.5.2 *Barometric Sensor Error Model*

Because the altimeter error affects many of the INS states, the fidelity of the baro error model is important. The single baro state included in the Litton error model is adequate to represent only one of the major sources of baro altitude error. Other baro altimeter error sources are statistically independent and should be included in the truth model. The baro output error state, δh_B, is directly coupled to four states in the INS dynamics equations, and is indirectly coupled to several others. An adequate baro altimeter error model, includes the time/position varying bias (function of atmospheric pressure-altitude spatial gradient and vehicle speed) δh_{pco}, instrument dynamic lag δh_a, scale factor error δh_{sf}, and an instrument bias δh_b.

The instrument dynamic lag error is typically modeled by first order lag with a time constant on the order of *1 second.* The instrument bias is modeled as a random constant whose initial 1-σ value is chosen to be 15 feet. The choice is based on a baro altimeter model contained in [26]. The scale factor error is a multiplicative combination of aircraft altitude and the random constant scale factor state. The initial *1*-σ value for the scale factor state is *0.01* [26]. Although the error contribution from this state is negligible at low altitudes, it can be a significant source of error at higher altitudes. The last error state in the revised baro altimeter model is a first order lag which sums and filters the previously described errors.

An example of a CADC (barometric) altitude error model, in the form of a first order stochastic differential equation, is

$$
\begin{bmatrix} \delta \dot{h}_{pco} \\ \delta \dot{h}_a \\ \delta \dot{h}_{sf} \\ \delta \dot{h}_b \end{bmatrix} = \begin{bmatrix} -\dfrac{1}{\tau_{pco}} & 0 & 0 & 0 \\ \dfrac{1}{\tau_{dy}} & -\dfrac{1}{\tau_{dy}} & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} \delta h_{pco} \\ \delta h_a \\ \delta h_{sf} \\ \delta h_b \end{bmatrix} + \begin{bmatrix} K_1/\tau_{pco} \\ 0 \\ 0 \\ 0 \end{bmatrix} w_{pco} \quad (31)
$$

The output error equation of this filter represents the total, lagged baro-altimeter error δh_B is

$$
\delta h_B = \delta h_a + \delta h_b + h(\delta h_{sf}) + v \quad (32)
$$

This error model represents a CADC corrected barometric altimeter measurement error reflected in Figure 4. This model framework is adequate for most air vehicle applications. For those cases where the CADC internal compensation computations can affect either the dynamic or stochastic error, this model can be enhanced appropriately to account for such processing.

3.6 Global Positioning System (GPS)

The Global Positioning System (GPS) is a navigation system which transmits signals from which a user can compute navigation information. The three elements of GPS are the space, user, and the control segments. The space segment, when completed in 1994, will consist of a group of 24 satellites in semi-synchronous orbits around the Earth. Each satellite, known as a GPS space vehicle (SV), transmits on two microwave frequencies L_1 and L_2. Each frequency has binary phase-shift keyed (BPSK) coded information as well as a 50 *bps* navigation message.

The user segment consists of all users of GPS information with suitable equipment to receive the electromagnetic energy transmitted from a subset of *visible* SVs. Using a combination of phase lock loop (PLL) and delay lock loop (DLL) techniques, the receiver measures the apparent propagation time of the signal from each SV. The receiver determines the position of each SV based on ephemeris data in the navigation message. Know-

ing the positions of the tracked SVs, the apparent propagation times, and user time the receiver solves for its own position.

The control segment functions as the system administrator. It consists of a master control station at Falcon AFB, Colorado, and a few remote sites throughout the world. It is responsible for maintaining the satellites and the data they transmit. This includes updating the SV clock error coefficients, ephemeris data, health status, etc. A thorough discussion of the GPS segments can be found in the literature [29].

Although the GPS can act as a stand-alone navigation system, for dynamic applications it is most powerful when integrated with an INS. Specifically, GPS pseudo-range measurements are used to periodically improve the navigation system's accuracy. A free-running INS is subject to various errors which generally worsen with time. GPS range measurements, when processed by an appropriate Kalman filter, estimate the INS position error and can be used to improve position estimates. The Kalman filter combines satellite range measurements, with its own internal models of the system to derive an estimate of the user position error. This is then algebraically summed with the indicated position to yield the best estimate of user position [30].

3.6.1 GPS Measurements

The GPS error model consists of three model domains. First, is the deterministic dynamic behavior of the GPS errors. Second, the measurement model functionally relates the measured quantities, the pseudo-range and delta-range, to the quantities of interest, including the error variables. Third, the stochastic attributes of the model describe that behavior which is caused by the random nature of the measurement noise as well as the non-deterministic character of the error which drive the dynamic error variables.

3.6.1.1 True Range

The GPS receiver measures a range which is based on the time offset between the SV generated BPSK code and the user generated code. When this time offset is scaled by the speed of light c, a measure of range results. The true range, R_t, is simply the shortest distance between the two points in the three dimensional position space:

$$R_t = |r_{sv} - r_u| \qquad (33)$$

When the SV (r_{sv}) and user (r_u) coordinates are expressed in the same reference frame in R^3 space, this range is

$$R_t = \sqrt{(x_{sv} - x_u)^2 + (y_{sv} - y_u)^2 + (z_{sv} - z_u)^2} \qquad (34)$$

Note that the range is a nonlinear function of the whole-valued user r_u and SV r_{sv} position coordinate vector variables. These vectors are frequently calculated in an ECEF coordinate frame, e.g. r_{sv}^e, and r_u^e. The whole-valued position of the user is generally expressed as a latitude, longitude, altitude triplet, which can be converted to the Litton ECEF (or other desired reference) frame by a transformation such as:

$$r_u^e = \begin{bmatrix} x_u \\ y_u \\ z_u \end{bmatrix}^e = \begin{bmatrix} (R_N + h)\,cos\,\varphi\,sin\lambda \\ [R_N(1 - e_e^2) + h]\,sin\,\varphi \\ (R_N + h)\,cos\,\varphi\,cos\lambda \end{bmatrix}^e \qquad (35)$$

The symbols are defined in Section 9. Equations (33) and (34) are truth model representations of the range. In a simulation environment the values of $x_{sv}, y_{sv}, z_{sv}, x_u, y_u, z_u$, (or ϕ, λ, h) can be provided as true, whole-valued trajectory data.

3.6.1.2 Range Measurement Errors, δR

The measurement of true range R_t, called *raw pseudo-range*, is corrupted by several errors, the most significant being the user clock bias [31, 32]. The SV clock itself may be in error with

respect to the GPS time. The signal from the SV traverses the ionosphere which distorts the path and causes an apparent time delay. When the signal traverses the troposphere, the fluid retards the propagation. The accuracy with which the user generated code can be correlated to the received code is a function of noise, code loop bandwidth, and code chip length. The uncalibrated propagation delay through the receiver also adds to this error. The antenna-to-receiver cable propagation calibration error causes additional measurement error. However, this error is often indistinguishable from the user clock delay and is lumped therein. The raw pseudo-range model consists of the true range and various time delays (scaled by c)

$$R_{RP} = R_t + R_{clk_u} + R_{code} + R_{trop} + R_{iono} + R_{clk_{sv}} + v \qquad (36)$$

The range domain variables are defined in Section 9 and the following paragraphs. The white Gaussian additive measurement noise v statistical properties are in Equation (13).

3.6.1.3 Compensated Pseudo-range

Three components on the right-hand side of Equation (36), namely R_{trop}, R_{ion}, and $R_{clk_{sv}}$, can be estimated open-loop. The total tropospheric delay is a function of the GPS receiver altitude, elevation angle to the SV, and other environmental factors. The ionospheric delay can be open-loop compensated using dual-frequency reception or by a model involving total electron count (TEC) estimates, elevation angle to the SV, and other environmental factors. The SV clock advance can be open-loop estimated by using the SV clock correction coefficients which are transmitted in the navigation message. Even after these three open-loop corrections are subtracted from both sides of Equation (36), the left hand side can be called the compensated pseudo-range, or simply pseudo-range. The corresponding terms on the right-hand side are replaced by the residual error terms after open-loop compensation. This yields the pseudo-range error equation (with error variables defined in Section 9.)

$$R_P = R_t + \delta R_{clk_u} + \delta R_{code} + \delta R_{trop} + \delta R_{iono} + \delta R_{clk_{sv}} + v \qquad (37)$$

3.6.1.4 Coordinate Transformations

In a typical modelling, simulation, and analysis task, several different coordinate frames are necessary. Although the details of coordinate transformations are not presented in this paper, caution is necessary to correctly relate the required vector and matrix quantities in proper reference frames. The reader is referred to texts and other publications for explanations and examples of such applications [1, 22, 28, 24].

3.6.2 GPS Error Dynamics

The dynamic and statistical nature of the error variables is developed in a form suitable for Kalman filter implementation. First, we develop the range error dynamic relations as a system of first order linear differential equations in variable x driven by white Gaussian noises w. These equations follow the format discussed in Section 5.2.

3.6.2.1 User Clock Error δR_{clk_u}

Generally, user equipment sets use a quartz crystal oscillator which is less precise than the SV clocks. The user clock error is modeled as an offset and drift from GPS system time. Over a typical mission duration, the user clock error δR_{clk_u} can be modeled with two-states

$$\delta \dot{R}_{clk_u} = \dot{x}_{clk_u} = \begin{bmatrix} \dot{x}_1 \\ \dot{x}_2 \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \qquad (38)$$

where x_1 is the range equivalent of user clock bias, and x_2 is the velocity equivalent of user clock drift. The numerical subscripts on these and subsequent GPS states are specific to the development of the GPS truth model; the same states are num-

bered differently in the combined GPS/INS model (Appendix) The stochastic properties for these random variables (states) are [33]

$$\hat{x}_{clk_u}(t_0) = 0 \tag{39}$$

$$P_{clk_u}(t_0) = \begin{bmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_2^2 \end{bmatrix} \tag{40}$$

where $\sigma_1^2 = 9 \times 10^{14} ft^2$, $\sigma_2^2 = 9 \times 10^{10} ft^2/sec^2$ for a typical user clock bias and drift. Note that the Equation (38) is a stochastic differential equation describing the code loop error dynamics of the form: $\dot{x} = Ax + Bu + Gw$, with $u=w=0$. These random variables are assumed to be zero-mean Gaussian processes, where Equations (39) and (40) describe the mean $\hat{x}_{clk_u}(t_0)$ and covariance $P_{clk_u}(t_0)$ initial conditions.

3.6.2.2   Uncompensated Code Loop Error δR_{code}

The GPS phase lock loop can be modeled as a first order lag driven by white Gaussian noise, with a time constant of approximately *1 second* [34]. Assuming that the user set has a separate channel for monitoring each of four SVs, and assuming no correlation between channels, the dynamic equations for user set code loop are (with $\tau_c = 1 \ sec$)

$$\delta\dot{R}_{code} = \begin{bmatrix} \dot{x}_3 \\ \dot{x}_4 \\ \dot{x}_5 \\ \dot{x}_6 \end{bmatrix} = \begin{bmatrix} \delta\dot{R}_{code_1} \\ \delta\dot{R}_{code_2} \\ \delta\dot{R}_{code_3} \\ \delta\dot{R}_{code_4} \end{bmatrix} = \begin{bmatrix} \frac{-1}{\tau_c} & 0 & 0 & 0 \\ 0 & \frac{-1}{\tau_c} & 0 & 0 \\ 0 & 0 & \frac{-1}{\tau_c} & 0 \\ 0 & 0 & 0 & \frac{-1}{\tau_c} \end{bmatrix} \begin{bmatrix} x_3 \\ x_4 \\ x_5 \\ x_6 \end{bmatrix} + \begin{bmatrix} w_3 \\ w_4 \\ w_5 \\ w_6 \end{bmatrix} \tag{41}$$

$$\hat{x}_{code}(t_0) = 0 \tag{42}$$

$$P_{code}(t) = \begin{bmatrix} \sigma_3^2 & 0 & 0 & 0 \\ 0 & \sigma_4^2 & 0 & 0 \\ 0 & 0 & \sigma_5^2 & 0 \\ 0 & 0 & 0 & \sigma_6^2 \end{bmatrix} \tag{43}$$

where $\sigma_i^2 = 1.0 ft^2$, for $i=3,4,5,6$. The driving noise is

$$E\{w_{code}\} = 0 \tag{44}$$

$$E\{w_{code}(t) w_{code}^T(t+\tau)\} = \begin{bmatrix} q_c & 0 & 0 & 0 \\ 0 & q_c & 0 & 0 \\ 0 & 0 & q_c & 0 \\ 0 & 0 & 0 & q_c \end{bmatrix} \delta(\tau) \tag{45}$$

where $q_c = 0.5 ft^2/sec^2$. The code-loop error stochastic differential equation is in the form of Equation (38), but with $w_{clk_u}(t) \neq 0$. Equations (44) and (45) describe the first and second moments of the white Gaussian driving noise probability density function.

3.6.2.3   Uncompensated Tropospheric Range Error δR_{tr}

The range error introduced by tropospheric delay can be as large as 80 feet, and is a function of altitude, environmental conditions, and SV LOS elevation angle [31]. Much of this error can be open-loop compensated, with the residual error modeled as first order Gauss Markov with $\tau_{tr} = 500 \ sec$ [34]

$$\delta\dot{R}_{tr} = \begin{bmatrix} \dot{x}_7 \\ \dot{x}_8 \\ \dot{x}_9 \\ \dot{x}_{10} \end{bmatrix} = \begin{bmatrix} \delta\dot{R}_{tr_1} \\ \delta\dot{R}_{tr_2} \\ \delta\dot{R}_{tr_3} \\ \delta\dot{R}_{tr_4} \end{bmatrix} = \begin{bmatrix} \frac{-1}{\tau_{tr}} & 0 & 0 & 0 \\ 0 & \frac{-1}{\tau_{tr}} & 0 & 0 \\ 0 & 0 & \frac{-1}{\tau_{tr}} & 0 \\ 0 & 0 & 0 & \frac{-1}{\tau_{tr}} \end{bmatrix} \begin{bmatrix} x_7 \\ x_8 \\ x_9 \\ x_{10} \end{bmatrix} + \begin{bmatrix} w_7 \\ w_8 \\ w_9 \\ w_{10} \end{bmatrix} \tag{46}$$

$$\hat{x}_{tr}(t_0) = 0 \tag{47}$$

$$P_{tr}(t_0) = \begin{bmatrix} \sigma_{tr}^2 & 0 & 0 & 0 \\ 0 & \sigma_{tr}^2 & 0 & 0 \\ 0 & 0 & \sigma_{tr}^2 & 0 \\ 0 & 0 & 0 & \sigma_{tr}^2 \end{bmatrix} \tag{48}$$

where $\sigma_{tr}^2 = 1.0 ft^2$. The driving noise $q_t = 0.004 \ ft^2/sec^2$.

$$E\{w_{tr}\} = 0 \tag{49}$$

$$E\{w_{tr}(t) w_{tr}^T(t+\tau)\} = \begin{bmatrix} q_t & 0 & 0 & 0 \\ 0 & q_t & 0 & 0 \\ 0 & 0 & q_t & 0 \\ 0 & 0 & 0 & q_t \end{bmatrix} \delta(\tau) \tag{50}$$

3.6.2.4   Uncompensated Ionospheric Range Error δR_i

The ionospheric-induced range error can exceed 150 feet [31]. This error varies as a function of solar activity, time of day, etc. and is frequency dependent. Single *L*-band receivers can correct for some ionospheric error. GPS receivers with dual *L*-band capabilities can compensate for most of this error. However even after open loop compensation, the residual error can be about 3 meters [31]. This can be further reduced with *multiple measurement time averaging subject to temporal and spatial ionospheric delay gradients*. Whatever the uncompensated ionospheric error is, it is expected to behave dynamically in a manner similar to the tropospheric error, but with mutually uncorrelated statistics. The form of the ionospheric error model is again a stochastic differential equation driven by a white Gaussian noise with appropriate initial conditions and statistics

$$\delta\dot{R}_i = \begin{bmatrix} \dot{x}_{11} \\ \dot{x}_{12} \\ \dot{x}_{13} \\ \dot{x}_{14} \end{bmatrix} = \begin{bmatrix} \delta\dot{R}_{i_1} \\ \delta\dot{R}_{i_1} \\ \delta\dot{R}_{i_1} \\ \delta\dot{R}_{i_1} \end{bmatrix} = \begin{bmatrix} -\frac{1}{\tau_i} & 0 & 0 & 0 \\ 0 & -\frac{1}{\tau_i} & 0 & 0 \\ 0 & 0 & -\frac{1}{\tau_i} & 0 \\ 0 & 0 & 0 & -\frac{1}{\tau_i} \end{bmatrix} \begin{bmatrix} x_{11} \\ x_{12} \\ x_{13} \\ x_{14} \end{bmatrix} + \begin{bmatrix} w_{11} \\ w_{12} \\ w_{13} \\ w_{14} \end{bmatrix} \tag{51}$$

with $\tau_i = 1500 \ sec$. It should be noted that this ionospheric delay error time constant is based on the work of [31, 32, 35] and may require additional empirical validation and revision as necessary. However the basic structure and the chosen parameters should suffice for many applications.

$$\hat{x}_i(t_0) = 0 \tag{52}$$

$$P_i(t_0) = \begin{bmatrix} \sigma_i^2 & 0 & 0 & 0 \\ 0 & \sigma_i^2 & 0 & 0 \\ 0 & 0 & \sigma_i^2 & 0 \\ 0 & 0 & 0 & \sigma_i^2 \end{bmatrix} \tag{53}$$

$\sigma_i^2 = 1.0 ft^2$. The driving noise w with $q_i = 0.004 \ ft^2/sec^2$

$$E\{w_i\} = 0 \tag{54}$$

$$E\{w_i(t)\,w_i^T(t+\tau)\} = \begin{bmatrix} q_i & 0 & 0 & 0 \\ 0 & q_i & 0 & 0 \\ 0 & 0 & q_i & 0 \\ 0 & 0 & 0 & q_i \end{bmatrix} \delta(\tau) \tag{55}$$

3.6.2.5 Uncompensated SV Clock Range Error δR_{clk_s}:

The GPS navigation message contains correction coefficients for SV clock bias, drift, and drift rate. Open-loop corrections are accurate to about 1.0 meters [31]. This error is assumed to be a relatively constant over time, albeit random, thus it is modeled as a random bias. The dynamic equations for SV clock advance range error are:

$$\delta\dot{R}_{clk_s} = \begin{bmatrix} \dot{x}_{15} \\ \dot{x}_{16} \\ \dot{x}_{17} \\ \dot{x}_{18} \end{bmatrix} = \begin{bmatrix} \delta\dot{R}_{clk_{s1}} \\ \delta\dot{R}_{clk_{s2}} \\ \delta\dot{R}_{clk_{s3}} \\ \delta\dot{R}_{clk_{s4}} \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} x_{15} \\ x_{16} \\ x_{17} \\ x_{18} \end{bmatrix} + \begin{bmatrix} w_{15} \\ w_{16} \\ w_{17} \\ w_{18} \end{bmatrix} \tag{56}$$

$$\hat{x}_{clk_s}(t_0) = 0 \tag{57}$$

$$P_{clk_s}(t_0) = \begin{bmatrix} \sigma_s^2 & 0 & 0 & 0 \\ 0 & \sigma_s^2 & 0 & 0 \\ 0 & 0 & \sigma_s^2 & 0 \\ 0 & 0 & 0 & \sigma_s^2 \end{bmatrix} \tag{58}$$

where $\sigma^2_s = 25ft^2$. The driving noise statistics are

$$E\{w_{clk_s}\} = 0 \tag{59}$$

$$E\{w_{clk_s}(t)\,w_{clk_s}^T(t+\tau)\} = \begin{bmatrix} q_s & 0 & 0 & 0 \\ 0 & q_s & 0 & 0 \\ 0 & 0 & q_s & 0 \\ 0 & 0 & 0 & q_s \end{bmatrix} \delta(\tau) \tag{60}$$

where $q_s = 0.0\,ft^2/sec^2$ for the truth model, and $q_s = 1.0 \times 10^{-10}$ ft^2/sec^2 for the filter model. The non-zero driving noise in the Kalman filter is used to keep the filter gain, associated with this set of variables, from approaching zero. The non-zero, but small, value keeps the filter gain open to permit parameter variation, or other model inaccuracies from detracting from filter performance. This is one of various filter tuning artifices discussed in various texts, such as [7, 16].

3.6.2.6 Satellite Position Errors δr_{SV}

In addition to the above described 18 error states the satellite position and velocity errors also affect the navigation solution accuracy. The satellite position and velocity are known with an accuracy limited by the ground segment of GPS. The ground measurements are accumulated from the monitor stations, processed in a joint Kalman filter, and uplinked to the satellites. Even at the time of highest achievable accuracy, the satellite ephemeris errors are non-zero. More significantly, these errors grow in time due to various random orbital perturbations. The position and velocity errors increase with time, from the last update, at some rate. The time-tagged position and velocity vectors are transmitted to the user, however the transmitted quantities differ from the true ones. For most applications, the SV position errors alone are sufficient. Should an application require velocity error models, similar methodology can be used. The following model develops only the SV position errors.

The magnitude of these errors can be up to 2-4 meters, as reported in early literature on GPS. Actual satellite position and velocity error data, including the error statistics and dynamics, should be available from the GPS management. Each satellite position error is in three dimensions, thus 12 first-order Markov process states are required. In the actual Kalman filter, this three dimensional error does not entirely affect the user measurement. Only the scalar projection of the three-dimensional true error vector unto the line-of-sight (LOS) affects the measurement. Thus the filter requires only a single state for each satellite line-of-sight to account for this effect. With time, as the error vector changes its orientation with respect to the LOS, the projection magnitude also varies. The magnitude of this slow variation is a small contribution to the overall pseudo-range measurement error. Thus a useful truth model for GPS satellite position errors is

$$\delta\dot{r}_{sv} = \begin{bmatrix} \delta X_{sv_1} \\ \delta X_{sv_2} \\ \dots \\ \delta Z_{sv_3} \\ \delta Z_{sv_4} \end{bmatrix} = \begin{bmatrix} \dot{x}_{19} \\ \dot{x}_{20} \\ \dots \\ \dot{x}_{29} \\ \dot{x}_{30} \end{bmatrix} = \begin{bmatrix} \frac{-1}{\tau_{sv}} & 0 & \dots & 0 & 0 \\ 0 & \frac{-1}{\tau_{sv}} & \dots & 0 & 0 \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & \frac{-1}{\tau_{sv}} & 0 \\ 0 & 0 & \dots & 0 & \frac{-1}{\tau_{sv}} \end{bmatrix} \begin{bmatrix} x_{19} \\ x_{20} \\ \dots \\ x_{29} \\ x_{30} \end{bmatrix} + \begin{bmatrix} w_{19} \\ w_{20} \\ \dots \\ w_{29} \\ w_{30} \end{bmatrix} \tag{61}$$

with $\tau_{sv} = 1500\ sec$

$$\hat{x}_{sv}(t_0) = 0 \tag{62}$$

$$P_{sv}(t_0) = \begin{bmatrix} \sigma_{sv}^2 & 0 & \dots & 0 & 0 \\ 0 & \sigma_{sv}^2 & \dots & 0 & 0 \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & \sigma_{sv}^2 & 0 \\ 0 & 0 & \dots & 0 & \sigma_{sv}^2 \end{bmatrix}_{12 \times 12} \tag{63}$$

where $\sigma^2_{sv} = 25ft^2$. The driving noise statistics are described completely by the first and second Gaussian moments

$$E\{w_{sv}\} = 0 \tag{64}$$

$$E\{w_{sv}(t)\,w_{sv}^T(t+\tau)\} = \begin{bmatrix} q_{sv} & 0 & \dots & 0 & 0 \\ 0 & q_{sv} & \dots & 0 & 0 \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & q_{sv} & 0 \\ 0 & 0 & \dots & 0 & q_{sv} \end{bmatrix} \delta(\tau) \tag{65}$$

The GPS error truth model state vector consists of 30 elements (see Appendix for a complete listing of these states). In the actual Kalman filter only one state per LOS is required. In fact, attempting to estimate the other eight components of this error may lead to filter instabilities. This is due to the fundamental unobservability of the SV position error components which are orthogonal to the LOS.

3.6.3 Differential GPS

Differential GPS concepts are described in Section 6. The error models for differential GPS applications are derived from those of the absolute GPS. As is explained in that later section, the differential GPS error model, basically consists of variations of the model derived in this section.

3.7 Transponder Error Model

The Range/Range-Rate System (RRS) is a navigation aiding system which comprises a critical part of the *Completely Integrated Reference Instrumentation System (CIRIS)*. The RRS is used in the Navigation Reference System (NRS). It is an example of a system utilizing measurements from ground-based ranging transponders to obtain relative navigation information.

The RRS is quite similar to the orbiting transponder system, that is the GPS. Each provides relative range and range-rate measurements. In fact, during early stages of GPS testing, an inverted GPS system was used. The GPS transmitters that were placed on the Yuma Proving Grounds surface, performed a ranging function similar to that of the RRS. The few differences in the resulting range and range-rate measurements are in the character of the measurement errors. Six to ten transponders are typically used for Kalman filter updating.

Navigation information is obtained by *interrogating* ground transponders and subsequently processing the signals with which the transponders reply. The received information allows high quality range and range-rate measurements to be calculated by the RRS interrogating hardware [36]. Using these range and range-rate measurements, refinements to the NRS position and velocity estimates are then possible.

The RRS transponder measurement error model is in a form similar to that of the GPS [36, 37]. The lever-arm effect discussed in [36] is not included in this work. The lever-arm effect must be considered in hardware applications. Its omission in these studies does not change the validity of the conclusions. The RRS range measurement is computed from the time delay detected between the transmission of an interrogation request and the reception of a reply from the transponder. This temporal difference is scaled by the speed of light, $c/2$, to obtain an uncorrected range measurement. Correction factors are applied for tropospheric propagation delays and calibration residual [36]. The uncorrected range measurement R_{RRS} is (with variables defined in Section 9.)

$$R_{RRS} = R_t + \delta R_{atm} + \delta R_{br} + v \qquad (66)$$

A difference of two measurements is useful. The first is the RRS range measurement. The second is the range computed from the INS-indicated and surveyed-transponder positions. The INS and transponder positions are represented by vectors expressed in the Litton ECEF frame. The calculated range from the INS to the transponder is given by

$$R_{INS} = \sqrt{(x_u - x_t)^2 + (y_u - y_t)^2 + (z_u - z_t)^2} \qquad (67)$$

This equation is *linearized* about the error variables δx_u and δx_t in a Taylor series expansion truncated to first order [1, 7]

$$R_{INS} = R_t - \left[\frac{x_t - x_u}{|R_{INS}|}\right]\delta x_u - \left[\frac{y_t - y_u}{|R_{INS}|}\right]\delta y_u - \left[\frac{z_t - z_u}{|R_{INS}|}\right]\delta z_u$$

$$+ \left[\frac{x_t - x_u}{|R_{INS}|}\right]\delta x_t + \left[\frac{y_t - y_u}{|R_{INS}|}\right]\delta y_t + \left[\frac{z_t - z_u}{|R_{INS}|}\right]\delta z_t \qquad (68)$$

Note the unit-line-of-sight vector, u_{LOS} elements in the brackets. The difference measurement is formed

$$\delta z = R_{INS} - R_{RRS} \qquad (69)$$

The true range present in each range measurement is cancelled [1, 7]. Also note that the bracketed coefficients in the equation above determine the Kalman filter's H matrix.

The RRS error state vector consists of 26 elements. The first two states are zero mean random biases which model the airborne equipment range and range-rate calibration errors. The initial covariances for these states are

$$P(t_0) = \begin{bmatrix} 1ft^2 & 0 \\ 0 & 10^{-4}ft^2/sec^2 \end{bmatrix} \qquad (70)$$

These two states apply to all RRS measurements. There are two error sources unique to each transponder. First is the error due to R^3 transponder surveyed position uncertainty (x, y, z components in ECEF frame), and second is the error due to atmospheric propagation delays between the user and each individual transponder. The three position error sources are

modeled by zero mean random bias states, and the atmospheric error states are represented by first order Markov processes with a time constant of *300 seconds*. For each transponder, four states are defined [36]

$$P_{atm}(t_0) = \begin{bmatrix} 25ft^2 & 0 & 0 & 0 \\ 0 & 25ft^2 & 0 & 0 \\ 0 & 0 & 25ft^2 & 0 \\ 0 & 0 & 0 & 100(ppm)^2 \end{bmatrix} \qquad (71)$$

$$E\{w(t)w(t+\tau)\} = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \frac{2\sigma^2}{300} \end{bmatrix}\delta(\tau) \qquad (72)$$

with $\sigma^2 = 10^{-10}$. Equations (71) and (72) are for a single transponder. Six RRS transponders are used in this model.

The Kalman filter combines range measurements with its predicted measurement to calculate an optimal estimate of the state vector. The indicated positions of the transponder and user are modeled by

$$\tilde{x}_t = x_t + \delta x_t \qquad (73)$$

$$\tilde{x}_u = x_u + \delta x_u \qquad (74)$$

The truth model measurement equation is

$$z = u_{LOS}(\delta x_t - \delta x_u) - \delta R_{atm} - \delta R_{br} + v \qquad (75)$$

Caution is required in performing the computations. It is necessary that all of the vector operations be performed in the same coordinatized vector space. Typically the ECEF is chosen for these computations.

3.8 SAR Model.

Synthetic Aperture Radar (SAR) is an airborne radar capable of generating high resolution images of surface target areas and terrain. A fundamental limit in angular resolution a_{res} of an imaging system is due to the diffraction limit and is defined by $a_{res} = 1.22(D/\lambda)$, where λ is the electromagnetic wavelength, and D the system aperture (lens, mirror, or RF antenna). Thus the larger the aperture, the smaller minimum resolvable angle (higher resolution).

The term synthetic aperture is used because SAR radar utilizes the motion of the aircraft, or other platform, to synthesize a large aperture antenna from a smaller physical antenna aperture. It accumulates motion-induced, spatially-distributed information over a period of time and phased properly to within a fraction of the wavelength λ. As a result, high resolution images are obtainable using an lower instantaneous resolution antenna.

High resolution SAR radar images are employed in various applications. As with any physical device there exist measurement *uncertainties (errors)*. Combining the SAR measurements with other sensors provides an opportunity to reduce the measurement errors. This discussion, presented in [7, 39, 40, 41, 42, 43, 44] provides the SAR measurement model. The goal is to develop key ideas sufficient for implementation in a centralized Kalman filter.

3.8.1 *SAR Targeting Background*

The basic SAR geometry is shown in Figure 5. The SAR is characterized by a wide beam antenna which illuminates a large area on the ground. For an assumed instant in time, the SAR operates by transmitting a radar pulse and samples the magnitude and phase of the return signal. These samples are stored into a vector of data generally referred to as a *range bin*. As the aircraft travels along a given ground track, the process of illumi-

244

nating the ground with a radar signal and sampling the magnitude and phase of the return is repeated every T seconds for consecutively overlapping areas on the ground. The result is a two dimensional array of data which contains magnitude and phase information of radar returns in both range and azimuth directions.
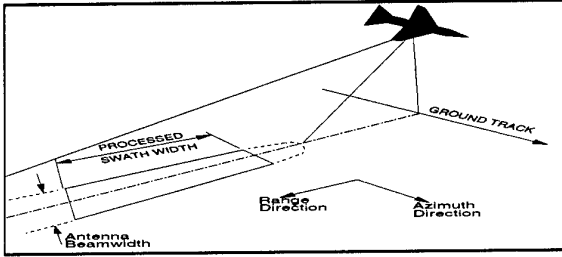

Figure 5. SAR Geometry

The SAR image alone cannot provide all the information needed to determine the target location, nor is it used for aircraft navigation information. The primary function of a SAR radar is to generate high resolution radar images of ground terrain. As a result, SAR images are employed to achieve the first task for target location, namely target designation. The operator designates the target by manually placing a cursor over the desired target image. SAR systems must employ other modes of operation to actually locate the target. These modes generally consist of conventional radar techniques which measure parameters associated with the target location. For instance, the range (R) and range-rate ($\dot{R}$) for each sampled cell in the image may be measured using conventional range-Doppler radar. These quantities are computed

$$R = \frac{cT}{2} \tag{76}$$

$$\dot{R} = \frac{f_d \lambda}{2} \tag{77}$$

where c is the speed of light, T is the round-trip transit time of the transmitted pulse, f_d is the Doppler frequency shift, and λ is the wavelength of the transmitted signal.

SAR target location measurements are often classified by two basic types: range-Doppler measurements and range-monopulse angle measurements. Both of these methods provide just enough information to compute the target location with only one set of measurements. When classical observer or Kalman filter is used with multiple-look measurements, it is possible to locate a target with a smaller set of measurements. Also, the Kalman filter provides a means to continually improve the estimate of the target location by utilizing the complete measurement history. In this report, we derive the Kalman filter measurement equations for both the range-Doppler and the range-monopulse measurements. Presented next is a brief explanation of both types of radar targeting methods.

In the range-Doppler method, range and range-rates are measured and stored for each cell in the SAR image. In addition to range and range-rate measurements, this method requires a radar altitude measurement for computing the location of a ground based target.

The range-monopulse method utilizes the range and precise angle measurements to compute the target position. In a monopulse radar, two spatially separated antenna form two similar beams. An RF signal receiver/combiner, obtains a sum and difference outputs. The antenna pattern for the difference port exhibits a null directly between the beams. Likewise, the pattern for the sum port exhibits a peak directly between the beams. This forms a precise pencil-beam tracking and/or angle measurement radar.

3.8.2 Coordinate Frames and Transformations

The coordinate frames and coordinate transformations matrices needed for radar measurements are the navigation n-, body b-, and path p-frames defined as a right handed orthogonal set. The path frame defines the attitude and direction of flight relative to the navigation frame via roll, pitch, and yaw angles α_x, α_y, and α_z. Its origin is at the aircraft center of mass with x^p pointing in the line-of flight (LOF) direction, y^p points out the right wing, and z^p pointing through the bottom of the aircraft[5]. Coordinate transformations between frames is required to perform vector additions and multiplications.

3.8.3 Transformation Matrix Perturbations

Error analysis utilizes perturbation methods which linearize the nonlinear differential equations [1]. For coordinate transformations matrices, this involves approximating the *indicated* transformation obtained from the INS as the sum of the *true* and *indicated* transformation matrix error

$$\bar{C}_p^n \approx C_p^n + \delta C_p^n \tag{78}$$

The indicated transformation matrix errors are δC_p^n and δC_a^p.

Orthogonalizing $\bar{C}_p^n$, about a small angle perturbation matrix Φ_{sk} provides the required linearized error model.

$$(\bar{C}_p^n)_o = (I - \Phi) C_p^n \tag{79}$$

where Φ_{sk} is given by a skew-symmetric matrix representation of the rotation error vector $\underline{\phi} = [\phi_x, \phi_y, \phi_z]$ [45]

$$\Phi_{sk} = \begin{bmatrix} 0 & -\phi_z & \phi_y \\ \phi_z & 0 & -\phi_x \\ -\phi_y & \phi_x & 0 \end{bmatrix} \tag{80}$$

This attitude error matrix Φ_{sk} (or rotation vector $\underline{\phi}$) is directly usable in a Kalman filter [45].

3.8.4 Range and Range-Rate Measurement Model

In this section the measurement equations for the radar range and range-rate measurements are presented. The basic vector geometry of an aircraft and a target in an earth centered earth fixed (ECEF) coordinatized frame is shown in Figure 6.
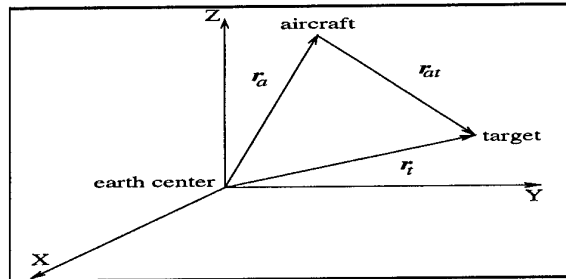
Figure 6. Target Location Geometry

The vector $r = Ru$, where R is the vector magnitude, and u is the unit vector indicating the vector direction, such that $|u| = 1$, and $|r| = R = u^T r$. The time-rate of change of vector r is:

$$\dot{r} = \frac{d}{dt} r = \frac{d}{dt} (Ru) = \dot{R}u + R\dot{u} \tag{81}$$

where $\dot{R}$ is the range-rate, and $\dot{u}$ is the line-of-sight (LOS) rotation rate. Range-rate can be expressed (using Equation (81), $u^T\dot{u} = 0$, and $u^T u = 1$) as

$$\dot{R} = u^T \dot{r} \tag{82}$$

---

5. This problem is simplified by the assumption (when valid) that the longitudinal axis of the aircraft is aligned with the velocity vector.

The ECEF coordinate frame is shown for clarity, however, we assume that all equations are solved in the navigation frame. Note from Figure 6 that the vector r_{at} is computed from r_a and r_t by the relationship $r_{at} = r_t - r_a$, where vectors r_{at}, r_a, and r_t are expressed in ECEF coordinates. Thus

$$R_{at}u_{at} = R_t u_t - R_a u_a \quad (83)$$

$$\dot{R}_{at} = u_{at}^T(\dot{R}_t u_t + R_t \dot{u}_t - \dot{R}_a u_a - R_a \dot{u}_a) \quad (84)$$

In special cases Equation (84) is simplified. For example when the target is stationary $\dot{R}_t = 0$ and $\dot{u}_t = 0$ resulting in

$$\dot{R}_{at} = -u_{at}^T \dot{r}_a \quad (85)$$

3.8.4.1 Radar Measurement

As previously stated, the radar is capable of obtaining an indirect measurement of R and $\dot{R}$. However, these measurement are not perfect and therefore must be modeled as the sum of true value and the measurement noise. $\tilde{R}$ and $\tilde{\dot{R}}$ denote the radar measurement, and may be modeled by

$$\tilde{R} = R + n_R + v_R \quad (86)$$

$$\tilde{\dot{R}} = \dot{R} + n_{\dot{R}} + v_{\dot{R}} \quad (87)$$

where n_R and $n_{\dot{R}}$ denote colored measurement noises, v_R and $v_{\dot{R}}$ denote white Gaussian measurement noises.

In general, the correlated measurement noises n_R and $n_{\dot{R}}$ are composed of many error sources. With an appropriate Kalman filter design it is possible to estimate the specific errors which make up the colored measurement noises n_R and $n_{\dot{R}}$, provided that these errors are individually observable. This requires that the errors comprising n_R and $n_{\dot{R}}$ be modeled as Gauss-Markov processes driven by white Gaussian noise. Likewise, the error sources comprising v_R and $v_{\dot{R}}$ must be modeled in a truth model. The truth model simulates with higher fidelity the true radar measurements.

The primary measurement noises observed in the range measurement include the target designation error and the clock error. Other errors which may also be included are multi-path and range glint errors. These are not modeled here because they are not always present in the measurements.

The target designation error is often modeled as a function of four parameters, namely clutter-to-noise ratio (CNR), signal-to-noise ratio (SNR), signal-to-clutter-plus-noise ratio (SCNR), and the resolution cell size. The CNR is computed to establish that the image quality is sufficient for individual fixed targets to be designated.

$$CNR = \rho_c \Delta G^2 L_a L_r T_i (\frac{r_0}{r}) \quad (88)$$

with

$$L_a = 10^{(2.16 \times 10^{-6})r} \quad \text{atmospheric loss}$$

$$L_r = 10^{(5.4 \times 10^{-6})\eta r} \quad \text{rain attenuation}$$

$$\rho_c \approx 0.1 (\Delta R)(\Delta az) \quad \text{clutter cross-section}$$

η is the rain rate in mm/hr, R is the range to the target, ΔR and Δaz are the range and azimuth resolutions, respectively, and r_0 is the range at which the SNR is unity for a $1m^2$ target, 1 sec array time, and undegraded main-lobe antenna gain. A typical value of r_0 is $1.5 \times 10^5 m$. If the CNR ratio is less than (3 dB), the resulting image is clutter. Consequently, when the CNR is less than 3dB, the variance of the target designation error, in both the

range and azimuth, is assumed equal to infinity. The SCNR is computed and compared to a threshold (of 10dB) to establish that the individual targets are recognizable by a human operator.

$$SCNR = \frac{SNR}{CNR + 0.1} \quad (89)$$

The SNR is given by

$$SNR = \rho_t \Delta G^2 L_a L_r T_i (\frac{r_0}{r})^4 \quad (90)$$

where ρ_t is the target cross-sectional area. For a side-looking SAR, range and azimuth resolutions are given by

$$\Delta R = \frac{c}{2B} = \frac{c\tau}{2} \quad (91)$$

$$\Delta\psi = \frac{D}{2} \quad (92)$$

respectively, where c is the speed of light, B is the bandwidth of the transmitted pulse, τ is the unmodulated pulse length, and D is the antenna diameter. For a spotlight mode SAR, the range resolution is the same as above and the azimuth resolution is expressed as

$$\Delta\psi = \frac{\lambda}{2\phi} \quad (93)$$

where λ is the wavelength of the transmitted pulse and ϕ is the angle between the aircraft and the target created by aircraft motion during SAR image data collection. The designation error in the range direction, denoted δR_D, and azimuth direction, denoted $\delta\psi_D$, have standard deviations given by

$$\sigma_{R_D} = K\Delta R \quad (94)$$

$$\sigma_{\psi_D} = K\Delta\psi \quad (95)$$

respectively, where typical values of K are on the interval $[0.5, 2.5]$. In [40] K is modeled by the following expression

$$K = \frac{1}{2(SCNR)^{1/2}} \quad (96)$$

This choice of K is given without discussion.

Two possible scenarios exist for SAR measurements each resulting in entirely different spectral characteristics for designation error. The first is a situation where SAR measurements are made only after target designation. In this scenario, the target designation errors in both range and azimuth directions are modeled as white Gaussian noises

$$\delta R_D = w_{R_D} \quad (97)$$

$$\delta\psi_D = w_{\psi_D} \quad (98)$$

where w_{R_D} and w_{ψ_D} denote white Gaussian noises. The second scenario consists of initially designating the target with a SAR then tracking it with monopulse radar. The errors present in the radar tracking mode are a direct result of the initial designation errors. Consequently, in this scenario a random bias model is appropriate for the designation error

$$\delta\dot{R}_D = 0 \quad (99)$$

$$\delta\dot{\psi}_D = 0 \quad (100)$$

In the Kalman filter design, a weak white Gaussian pseudo-noise is added to Equations (99) and (100) to compensate for possible hardware failures and model uncertainties. The clock error δT contribution to the range measurement δR_{clk} is expressed by

$$\delta R_{clk} = (\frac{c}{2})\delta T \quad (101)$$

where $c \approx 3.0 \times 10^8 (m/sec)$ is the speed of light. The clock error δT can also be modeled as a random bias

$$\delta \dot{T} = 0 \qquad (102)$$

Again a weak *pseudo-noise* is added in the Kalman filter to compensate for model inaccuracies.

The primary error in the *range-rate* measurement is the range-rate ambiguity error. Other errors which may also be included are the errors associated with surface wave motion when targeting is performed over bodies of water.

The range-rate ambiguity $\Delta \dot{R}_{amb}$ represents the smallest difference between two range-rates that could be detected by the radar and is characterized by

$$\Delta \dot{R}_{amb} = \frac{c}{2 f_0 \tau} \qquad (103)$$

where τ is the radar pulse length and f_0 is the frequency. As an approximation, we assume that the error in range-rate $\delta \dot{R}_{amb}$, resulting from the range-rate ambiguity, is white Gaussian noise with a standard deviation of $\Delta \dot{R}_{amb}$

$$\sigma_{\dot{R}_{amb}} \approx \left(\frac{c}{2 f_0 \tau} \right) \qquad (104)$$

where $\sigma_{\dot{R}_{amb}}$ is the standard deviation of $\delta \dot{R}_{amb}$. Use of this approximation results in an error that is slightly larger than the true errors that are normally expected from this source.

3.8.4.2 Error Measurement Model

The inputs to the Kalman filter are the differences between the indicated (e.g. INS) and the radar measured values. The range measurement model, for the general case where target and aircraft may both be moving with respect to navigation frame, differences the INS computed range $\bar{R}_I$ and the radar measured range $\tilde{R}_S$, and is expressed in terms of Kalman filter error states δr_t, δr_a, and n_R [45]

$$z_R = \bar{R}_I - \tilde{R}_S + n_R + v_R$$
$$\approx (u_{at}^n)^T (\delta r_t^n - \delta r_a^n) + n_R + v_R \qquad (105)$$

The range-rate measurement is similarly formed from the INS computed, $\bar{R}_I$, and the radar measured, $\tilde{R}_S$ range-rates. An example of such measurement for the case where the target is stationary with respect to earth's surface is [45]

$$z_{\dot{R}} = \bar{\dot{R}}_I - \tilde{\dot{R}}_S$$
$$\approx \frac{-\dot{r}_a^{nT}}{r} [I - u_{at}^n u_{at}^{n\,T}] [\delta r_t^n - \delta r_t^n] - u_{at}^{n\,T} \delta \dot{r}_a^n + n_{\dot{R}} + v_{\dot{R}} \qquad (106)$$

3.8.5 Azimuth and Elevation Angle Measurement Model

Next we present the measurement equations for the monopulse radar azimuth and elevation angle measurements. The basic vector geometry defining the azimuth angle ψ and elevation θ angle are shown below in Figure 7.
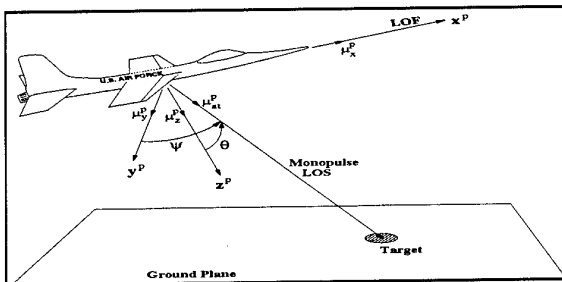
Figure 7. SAR Angular Geometry

To formulate the framework for incorporating the angular radar measurements, other-source derived LOS angles are again required. These are again generated from the INS provided information. The two error angles, azimuth $\delta \psi$ and elevation $\delta \theta$ are given by [45]

$$\delta \psi \approx - R \nabla \psi (u_{at}^n)_{sk} \underline{\phi} + \nabla \psi \delta r_t^n - \nabla \psi \delta r_a^n \qquad (107)$$

where $\nabla \psi$ is the gradient of ψ, and defined as a row vector

$$\nabla \psi = -\frac{(u_y^p)^{nT}}{R \sin (\psi)} [I - u_{at}^n (u_{at}^n)^T] \qquad (108)$$

$$\delta \theta \approx - R \nabla \theta (u_{at}^n)_{sk} \underline{\phi} + \nabla \theta \delta r_t^n - \nabla \theta \delta r_a^n \qquad (109)$$

where $\nabla \theta$ is the gradient of θ, and defined as

$$\nabla \theta = -\frac{(u_z^p)^{nT}}{R \sin (\theta)} [I - u_{at}^n (u_{at}^n)^T] \qquad (110)$$

3.8.5.1 Radar Measurement

A monopulse radar is capable of obtaining direct measurements of azimuth and elevation angles. However, these measurement are not perfect and therefore are modeled as the sum of the true value and the measurement noise. If $\tilde{\psi}$ and $\tilde{\theta}$ denote the radar measurement, then they may be modeled by the following equations

$$\tilde{\psi} = \psi - \eta_\psi - v_\psi \qquad (111)$$

$$\tilde{\theta} = \theta - \eta_\theta - v_\theta \qquad (112)$$

where η_ψ and η_θ denote colored measurement noises and where v_ψ and v_θ denote white Gaussian measurement noise. In reality, measurement noises η_ψ, η_θ, v_ψ, and v_θ consist of a number of physical error sources. Next follows a brief description of the primary error sources for use in the Kalman filter truth model design.

The primary error source for azimuth and elevation measurements include: target designation error, antenna boresight error, radar-INS harmonization error, antenna servo/pointing error, and radome correction error. First consider the effects of the designation error. In particular, consider the effects of an azimuth designation error on the azimuth angle. A similar development applies for the effects of range designation errors on the azimuth angle as well as the effects of both an azimuth and range designation errors on the elevation angle.

An azimuth designation error may be modeled as a vector which points in the direction of the aircraft velocity vector at the time that the image was formed. For detailed discussion the reader is referred to [45]. The radar azimuth designation errors are presented

$$\delta \tilde{\psi}_\psi = \nabla \psi u_a^n \delta \psi_D \qquad (113)$$

$$\delta \tilde{\theta}_\psi = \nabla \theta u_a^n \delta \psi_D \qquad (114)$$

The two radar line of sight (angular) errors are

$$\delta \tilde{\psi}_{R_D} = \nabla \psi u_{at_{xy}}^n \delta R_D \qquad (115)$$

$$\delta \tilde{\theta}_{R_D} = \nabla \theta u_{at_{xy}}^n \delta R_D \qquad (116)$$

The boresight, radar-INS harmonization, antenna servo/pointing, and radome refraction errors have both an azimuth and an elevation component which directly affect the azimuth and elevation angles, respectively. The azimuth and elevation errors consist of several physical components. With the exception of the radome errors, each of the error sources is characterized as random biases. In general, the radome errors are spatially

related to the azimuth angle ψ and elevation angle θ. However, often targeting is performed at large stand off distances with the aircraft in essentially a straight and level flight. Given this scenario, the azimuth and elevation angles change very little. In those cases they may also be modeled as a random biases.

3.8.5.2 Error Measurement Model

The inputs to the Kalman filter are the differences between the INS indicated, and the radar measured values

$$z_\psi = \overline{\psi} - \tilde{\psi} = -R\nabla\psi u^n_{at_{sk}}\phi + \nabla\psi\delta r^n_t - \nabla\psi\delta r^n_a + \eta_\psi + v_\psi \quad (117)$$

$$z_\theta = \overline{\theta} - \tilde{\theta} = -R\nabla\theta u^n_{at_{ss}}\phi + \nabla\theta\delta r^n_t - \nabla\theta\delta r^n_a + \eta_\theta + v_\theta \quad (118)$$

3.8.6 SAR Summary

The radar and SAR error model presented is neither complete, nor unique. It does present a framework for development of the required models. It is important to emphasize that *required* is relative to the task at hand. The truth model must always be of sufficient quality, such that the unmodelled errors would provide in the physical world a contribution which in effect is below the *noise level* of the system.

3.9 Other Sensors

As presented in the Section 1., the benefits of deep integration, go far beyond INS aiding. When a group of sensors, which share the same geometric vector, are processed jointly the natural result is that the stochastic cross-correlations are optimally computed. These cross-correlations become the agent for benefitting each of the sensor. The geometric errors of these geometrically related sensors are cross-calibrated, to the maximum degree possible (optimally).

A large variety of additional sensors, and other navigation update inputs have been demonstrated as useful in updating the INS. Some examples are: doppler radar, radar altimeter, map correlation, star/solar sensors, pilotage inputs, and others. The integration of those sensors for navigation updating has been studied, and reported on, by various organizations, corporations, and individuals. For an example, see *Models for Aided Inertial Navigation System Sensor Errors* by The Analytic Sciences Corporation (TASC) [38].

3.9.1 Angle (Azimuth, Elevation) and Range Sensors

One application of deep integration is the combining of the navigation sensors with other mission sensors, such as a laser radar (LADAR) or a synthetic aperture radar (SAR) which are capable of providing navigation update information as well as targeting information for a weapon system. The key advantage possible with deep integration is the calibration of the line-of-sight (LOS) and range bias and scale factor errors. Multiple "looks" by these sensors at calibrated and surveyed targets build up over time strong cross-correlations. These cross-correlations develop when the same common geometric vector reference is varied in time. This is analogous to the concept of inducing observability in the estimation theory [7]. The proper modeling, analysis, and simulation provides the baseline of performance and a foundation for implementation of trade-offs. The models for a LADAR or a SAR follow the framework developed earlier. Adequate models for the geometric vector component errors must be expressed in terms of their dynamic, stochastic, and measurement characteristics.

3.10 Geometric Sensor Model Summary

This section presented several models, of varying fidelities, which are deemed of sufficient quality for the analyses that had been performed in various studies and summarized in the following sections. The models presented in Section 3 were not all developed fully. Reader is referred to the referenced papers and reports for sufficient level of detail. The intent in this paper is to present the concepts of geometric vector commonality among all of the sensors modeled. This commonality, when represented in appropriate mathematical models suitable for Kalman filter implementation, offers opportunities to jointly estimate the errors that are geometrically related. This joint estimation is the key to substantial sensor error calibration and registration against some reference points, such as the illuminated target. When combined with the GPS, for referencing within the GPS-defined framework, tremendous opportunities are opened for relative registration on a global scale. This registration can be in relative GPS, differential GPS, or absolute GPS domains. Again the key to this precise, yet globally available, sensor calibration and registration is joint modeling and estimation with a single Kalman filter.

4. CIRIS: EXAMPLE OF GPS/INS INTEGRATION

CIRIS is a transponder aided INS test reference currently used by the Air Force for the development and testing of aircraft navigation systems [37, 46]. It is discussed in the transponder error model (Section 3.7). The office of primary responsibility (OPR) for CIRIS is the Central Inertial Guidance Test Facility (CIGTF), 6585th Test Group, Air Force Materiel Command (AFMC), Holloman AFB, NM. The operation of CIRIS involves flying the INS (or other system) to be tested, referred to as the test article, and the CIRIS through a pre-planned trajectory across a transponder range while data from both is recorded in-flight. The test article performance is compared (post-flight) to CIRIS.

CIRIS has served as the navigation system test standard since becoming operational in 1975. Until recently, CIRIS has been considered an accurate test baseline for determining the performance of aircraft navigation systems. CIRIS obtains an accurate navigation solution by combining information from three major subsystems: INS, barometric altitude (from CADC), and range/range-rate (RRS) transponder aiding [47, 48]. CIRIS determines the aircraft latitude and longitude with an accuracy of 14 ft horizontal and 40 ft vertical; the north and west velocity to 0.1 fps, and the vertical velocity to 0.4 fps (1-σ) [37, 46]. Improved navigation systems already exceed the CIRIS accuracy. CIRIS is being enhanced to provide a navigation solution an order of magnitude more accurate. By augmenting with DGPS measurements, it is possible to increase the CIRS navigation accuracy by an order of magnitude. The system presented in this section is called the Navigation Reference System (NRS). The higher CIRIS accuracy is possible by combining the strengths of the three systems.

An extended Kalman filter which incorporates an LN-93 INS model, transponder aiding (as implemented in CIRIS), and GPS aiding is designed and analyzed. Solomon's research specifically addresses GPS-CIRIS integration [37] while research performed by Snodgrass [36], and Cunningham [19] address the problem of *filter-driving-filter* instability which can occur when the integration is inadequately engineered. Solomon's GPS error model [37] and extensions developed at the Air Force institute of Technology (AFIT) form the GPS error model basis (Section 3.6.)

The following assumptions are made in this study. The truth models are of sufficient quality for these analyses and the conclusions drawn. The INS LN-93 error model, with the barometric altimeter error modifications [24], the RRS transponder error model by Snodgrass [36], and the GPS error model based on articles by Cox [30], Martin [31], and Milliken and Zoller [32] form the truth error model basis.

MSOFE is used to establish a performance baseline against which other full-order and reduced-order filters are compared. Time histories of variables such as aircraft position in R^3 space, attitude, and acceleration are created, and serve as the *true* aircraft trajectory in MSOFE simulations.

4.1 Extended Kalman Filter

A navigation *solution* is computed from a set of non-linear, coupled, deterministic, differential equations, such as illustrated in Equation (27). The accuracy of the solution depends on the INS error dynamics which are driven by several random sources whose statistics are known with varying degrees of accuracy. Given *a-priori* stochastic and dynamic models and initial conditions, the Kalman filter combines all measurements to produce optimal estimates. A system model must be linear in order to satisfy the optimality assumptions of a Kalman filter. Extended Kalman filter allows for relinearization about the best estimates at each computation cycle, thus enhancing the validity of the linearization process and improving filter performance [7].

4.2 INS and Altimeter Models

The Litton LN-93 strapdown INS is chosen for use in the proposed Navigation Reference System (NRS). The INS error model with the revisions made to the Litton baro-altimeter model, are presented in Sections 3.4 and 3.5. The INS subsystem, aided by a baro-altimeter, is combined with the RRS and GPS models to form the NRS. See Sections 3.7 and 3.6 for the respective error model descriptions.

4.3 RRS Transponder Error Model

For range/range-rate system description and error model, see Section 3.7. The six transponders in the NRS composite error model are located near Holloman AFB, NM. The locations of these transponders have been surveyed (typically to within 5 feet in each of the 3 axes), as given in Table 1.

Table 1. RRS Transponder Locations

Transponder ID	Latitude	Longitude	Altitude
005	33 01 36.14	-106 08 20.74	4339 ft
102	32 55 58.59	-106 08 50.33	4074
181	33 44 58.03	-106 22 14.63	7932
211	33 17 55.99	-106 31 44.31	8842
212	32 47 16.41	-105 49 15.47	9202
216	32 42 12.23	-106 07 38.90	4481

4.4 GPS Error Model

In a manner similar to the transponder system, GPS navigation information is obtained from signal propagation between the NRS and each of the SVs. As with RRS, GPS range measurements aid in estimating reference system errors. An *optimal* GPS constellation based on Green [49] is modeled. The model includes orbital calculations for all SVs, and simulates GPS receiver operation. The lever-arm effect is avoided, without the loss of generality, by assuming that the GPS and RRS antennae are collocated with the INS.

The GPS pseudo-range Equation (37) includes the true range along with the GPS measurement errors. Difference measurement is formulated in the GPS model in a manner similar to that of the RRS. The pseudo-range measurement from the GPS subsystem is subtracted from the range estimate constructed from the magnitude of a vector difference of INS-indicated position and SV broadcast positions.

4.5 Simulation

The system truth model is constructed from the LN-93 INS, the RRS, and the GPS subsystem models embedded in the MSOFE simulation routines. The truth model, along with PROFGEN generated trajectory and the SV orbit calculation software, generates measurements as well as reference variables which are used to test the performance of a full-order Kalman filter. This constitutes a simulation environment in which a variety of full-order and reduced-order Kalman filters may be tested against a common, high-fidelity standard.

Several system-level configurations are tested and analyzed using Monte Carlo simulations. Generally, analysis of a configuration consists of performing a series of 10 alignment runs, followed by a series of 10 flight simulations. The stochastic time history of error-state variables of interest is recorded. In the case of large dimensioned models with many measurement updates, fewer runs may be used, somewhat reducing confidence in the data sample statistics [7]. At the beginning of each alignment run, the truth model state vector is initialized. Each truth state is set to a random value based on the state's initial covariance and a random number to represent actual INS, RRS, and GPS error condition. The intent is to determine the effect of the randomization of truth model states on the Kalman filter's ability to perform its estimation task. Having initialized the truth model error states, the system is allowed to propagate its states for the period of the eight-minute alignment. During the alignment, the Kalman filter is provided with alignment measurements. At the end of the alignment runs, terminal conditions for the truth and filter state vectors and the final covariances are used as the initial conditions for flight simulations. Thus, a simulation is continuous for the entire analysis sequence.

4.6 Results

In the next step the RRS transponder aiding is added. The 26 RRS states are combined with the 72 INS states to produce a 98-state model whose performance establishes the baseline INS/RRS performance. The impact of providing RRS measurements to the Kalman filter is dramatic indeed. An improvement for the alignment simulations is evident in the horizontal position estimates. Latitude and longitude mean errors are closer to the zero-mean as predicted by estimation theory. Comparison of the vertical scales for horizontal channel states shows that the addition of RRS transponder measurements during alignment reduces the true- and filter-computed 1-σ values by more than 50 percent. A more substantial effect is evident in the flight runs. With RRS transponder measurements, horizontal position errors are reduced to less than 40 feet during all phases of the flight. Variations in the true and filter one sigma values are due to two sources. First, significant transitions occur in the altitude, velocity, and tilt states during periods of high dynamic maneuvering. Second, the aircraft range from the transponders has an obvious effect in horizontal position errors. Flight regimes in which the aircraft is at low altitude or is a long distance from the transponders result in increasing position uncertainty. When the aircraft is high overhead the transponder locations, much better estimates are possible. In all cases, the filter appears to be reasonably *tuned* in that the filter 1-σ values bound the mean error $\pm\sigma_{true}$ traces the majority of the time, without being overly conservative (Figure 8).
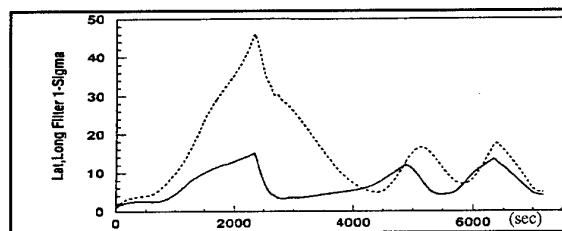
Figure 8. 98-state Model Filter Error (ft) 1-σ

In Figure 8, latitude error is indicated by the solid trace and longitude error is depicted by the dashed trace. Addition of the RRS measurements improved performance by roughly three orders of magnitude when compared to the INS performance aided by baro-altitude only. This is essentially the performance currently experienced by CIRIS. In the final phase, the GPS model is added to the 98-state INS/RRS model, bringing the

total number of NRS states to 128. Two modes of testing are performed with GPS addition. First the alignment and flight runs are performed with GPS measurements; no RRS measurements are included. Second, the alignment and flight series are re-run using both GPS and RRS measurements. Ten one-hour flight runs are performed for the 128-state NRS model. To facilitate a direct comparison of the horizontal positions states between this, previous, and subsequent configurations, Figure 9 is included.
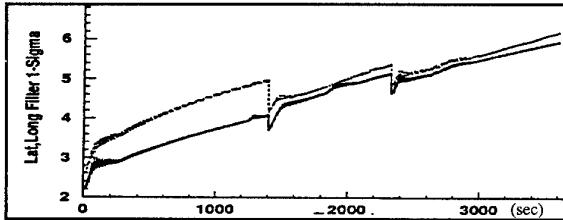
Figure 9. 128-state Model Filter Error (ft) 1-σ

In Figure 9, latitude error is indicated by the solid trace and longitude error by the dashed trace. Note the significantly reduced 1-σ magnitudes in this configuration compared to Figure 8. In some flight regimes (notably those far from the transponder sites, or those with depressed elevation angles between the user and transponders) performance appears to have improved by an order of magnitude compared to the INS performance achieved with RRS aiding alone.

In final configuration, both RRS and GPS measurements are used. For the alignment simulations, 10 Monte Carlo runs are performed. Alignment results are similar to those obtained with either GPS or RRS alone. The general trend is a slight reduction in the filter estimates of the overall error magnitudes during the alignment phase. In order to facilitate a direct comparison of the Kalman filter horizontal positions states between this, previous, and subsequent configurations, Figure 10 is included. In Figure 10, latitude error is indicated by the solid trace and longitude error is depicted by the dashed trace. The 1-σ magnitudes in this configuration are further reduced compared to Figures 8 and 9.
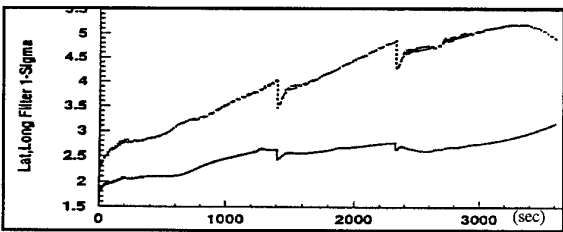
Figure 10. 128-state Model GPS & RRS Error (ft) 1-σ

The results achieved from the complete system simulation and analyses are presented in Figures 11-13. Key variables such as position, velocity, and platform tilt errors are plotted in the following figures. The true standard deviations of the indicated mean error variable is σ_{true}. The traces shown by dotted lines which bound and *track* the mean error time history represent the mean error plus and minus σ_{true}. The pair of traces shown by solid lines represents the filter-computed $\pm\sigma_{filter}$ for the error variables indicated. They are symmetrical about zero because the Kalman filter *assumes* that its errors are zero-mean [7].
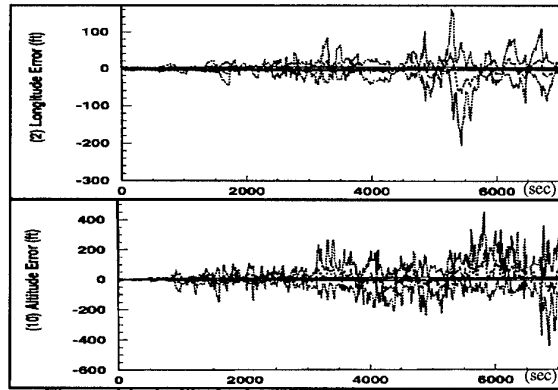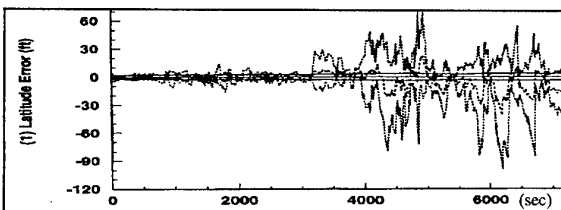
Figure 11. Latitude, Longitude, and Altitude Errors
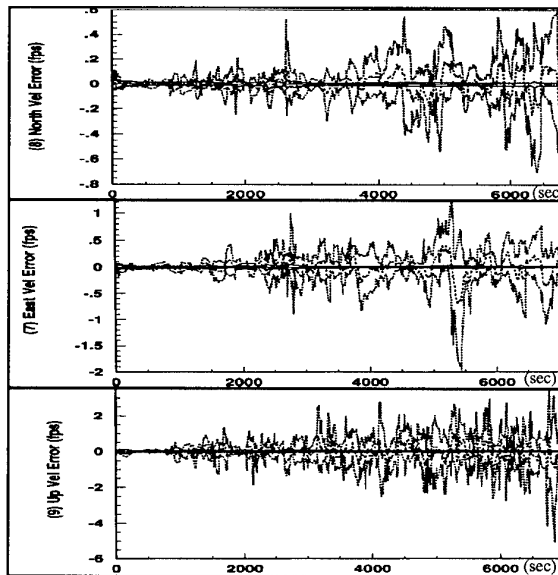
Figure 12. North, East, and Vertical (Up) Velocity Errors
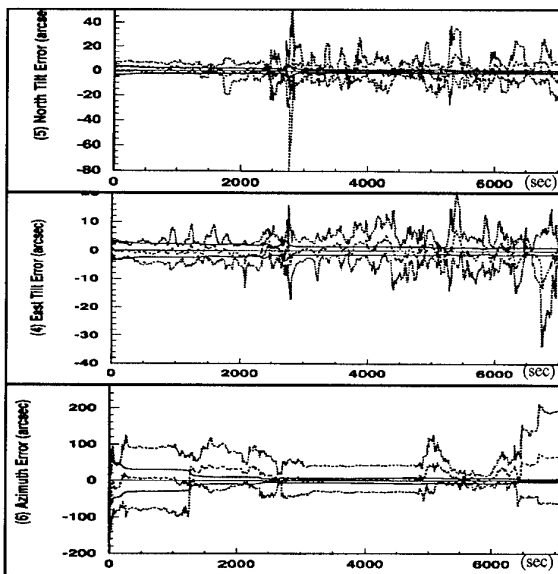
Figure 13. North, East, and Azimuth Tilt Error

After reviewing the different configurations, GPS aiding alone appears to offer a noticeable improvement over simple transponder-aided INS. In several flight regimes, the GPS-aided solution is significantly better than that of the RRS-aided INS. This fact is a direct result of two features which are inherent to

250

this simulation. The flight path extends far beyond the optimal coverage areas for the fixed transponders, and only six transponders are used; many actual flights use twenty or more transponders. The other important conclusion is that combining GPS and RRS measurements, in a joint Kalman filter, offers the best overall performance.

5. GRACEFUL DEGRADATION OF GPS/INS WITH FEWER THAN FOUR SATELLITES

This study examines the graceful degradation of a deep integration of GPS with INS and compared it to that of a federated filter architecture integration [50]. Precision first pass weapon delivery with exceptionally accurate navigation are vital to the operational survivability of attack aircraft. Systems such as the GPS are currently being fielded and integrated with INS to support such a capability. It offers the potential for less costly weapon targeting systems if target coordinates can be provided relative to the GPS reference. The current standard military GPS receivers requires simultaneous signals from 4 satellites (with good geometry) to calculate and output useful position and velocity information. However, signals from 4 satellites may not always be available, especially in and around heavily defended target areas protected by high power jammers. These high power jammers, even when recognized and appropriately nulled out by adaptive array GPS antenna, can result in large sectors of the sky surrounding these jammers no longer being monitored for any GPS satellite signals. This study examines Kalman filter algorithms which provide graceful degradation of aircraft navigation performance when only one, two, three, or intermittent satellite signals are available, and quantifies the level of relative performance expected for these conditions. This study also analyzes the relative performance of various Kalman filter mechanizations as a function of filter size.

The Standard GPS Receiver IIIA does not provide the pseudo-range and delta-range data from the GPS satellites on the MIL-STD 1553 data bus. The receiver only provides position and velocity data as an output of its own internal Kalman filter. This filter is based on a generic INS error model that is not optimized to any particular INS type or technology (it is crude in accuracy). If an aircraft requires a navigation solution optimized to a particular INS or if other sensors are required in the optimal navigation solution this GPS receiver output standard limits the aircraft Kalman filter designer to a cascaded filter approach. Use of GPS filtered position and velocity outputs to drive a separate aircraft Kalman filter can lead to filter instability, in part because the GPS and the INS position and velocity data are time correlated. This potential problem is minimized by processing GPS measurements in the aircraft filter at a much slower rate than available from the GPS receiver. This helps in reducing the GPS and INS measurement correlation. A spacing of 10-12 seconds is frequently considered sufficient when 4 satellites with good geometry are available.

When fewer than 4 satellites are available, the GPS receiver filter outputs degrade and closely track the rapidly growing INS errors. To avoid filter instability under these conditions the aircraft filter is designed to disregard inputs from the GPS receiver when fewer than 4 acceptable satellites are available. This cascaded filter approach therefore results in the binary on/off incorporation of GPS information. However, if pseudo-range and delta-range data were directly output, a single non-cascaded (joint) Kalman filter could be implemented. Such a filter would facilitate graceful position and velocity performance degradation when fewer than 4 satellites are available. One concern for GPS signal availability is during, or near, the weapon delivery time when jamming is likely to be present. Use of even momentary measurements from single random satellites by such a filter during a high jamming period will provide significantly improved performance over the cascaded filter approach. This discussion expands on the work presented in [50]. It describes the sensitivity analysis of the filter perfor-

mance as a function of the number of significantly contributing states. Subsequently, this performance is quantified against various GPS availability conditions.

5.1 Recommended Filter Mechanization

The GPS receiver should output pseudo-range and delta range, satellite position, and GPS time tag data. A joint Kalman filter that uses these GPS pseudo-range and delta-range measurements (instead of the filtered GPS position and velocity outputs) is recommended. The complexity of such a filter is not necessarily increased. For example a Kalman filter incorporating range and range-rate measurements from ground transponders has been "flying" since early 1970's as part of the CIRIS at the Central Inertial Guidance Test Facility (CIGTF), Holloman AFB, NM [51]. Modeling of GPS measurement error states in the Kalman filter results in higher estimation accuracy. When raw GPS data is available from 4 satellites with good geometry, three-dimensional position errors and user clock errors are directly computed by the aircraft filter. Furthermore, since "raw" GPS measurement errors are not time correlated with those of the INU, all GPS data can be used at the maximum GPS output rate. This enables INU errors, user clock bias and drift, and other errors to be rapidly and accurately estimated in the aircraft integrated (joint in the probability density function sense) Kalman filter. If, at a subsequent time, fewer than 4 satellites are available the now-calibrated user clock and other measurement errors (both GPS and barometric altimeter) will facilitate use of pseudo-range and delta-range data from the available satellites for up to 30 minutes and keep the navigation solution bounded [52]. Accurately estimated GPS and barometric measurement errors, when modelled in the filter, permit substantially improved navigation performance.

5.2 Filter Development

A 98-error state truth model is developed for the fully integrated GPS/INS navigation system. It combines 68-state INU truth model based on the LN-93, the Litton Inc. version of the Air Force Standard RLG inertial system [24] and a 30-state GPS error truth model developed by Solomon, in support of the advanced CIRIS for CIGTF [51]. The Litton's 93-state INU error model is expanded to include all of the inertial component misalignment states and the barometric altimeter measurement errors then reduced to 68 states by eliminating the trending, thermal transient, and compliance states (see the Appendix). From this truth model various reduced-order filters are created and evaluated using MSOFE against a PROFGEN-generated tactical mission trajectory.

The first reduced-order (51-state) filter analyzed is created by identifying and deleting those states whose error contributions are orders of magnitude less than the dominant error states. Additionally other error states, with similar coupling paths, are linearly combined. 15 INS states are an order of magnitude smaller in error contribution than other significant error states affecting either gyro platform tilt or velocity accuracy (e.g., the gyro scale factor asymmetry errors). Also 12 other "minor" INS error states are sufficiently similar in effect to other existing states that they are combined with respective dominant states (e.g., gyro misalignment errors with gyro scale factor error). Similarly the twelve GPS satellite position error states are not fully observable, except for the position error vector projection onto the respective line-of-sight, and can be deleted or combined. Additionally, the ionospheric and satellite clock errors can easily be combined with the tropospheric errors to reduce the GPS model by another eight states. Except for the user clock errors, the GPS measurement errors to each satellite are assumed to be statistically independent. (In fact some correlation exists, for example the low altitude atmospheric delay has some correlation as a function of angular proximity of the lines-of-sight and elevation angles.) Using this approach the 98-state truth model is reduced to a 51-state reduced-order fil-

ter. These 51 states are listed in the Appendix. The resulting 51-state reduced-order filter is tuned and its performance compared to that of the truth model. As expected, little or no tuning noise is added to selected states to insure filter stability (filter tuning). The 51-state reduced-order model filter performance is nearly identical to that of the 98-state truth filter.

The next, 32-state, reduced-order model combines the following 19 states into the remaining error terms (see the Appendix).

a. 3 gravity anomaly states
b. 3 accelerometer correlated bias states
c. 3 accelerometer scale factor asymmetry states
d. 6 accelerometer misalignment states
e. 4 GPS code loop errors

These 19 error states are primarily the least significant error contributors of the remaining 51 states. The resulting 32-state reduced-order filter is tuned and its performance compared to that of the truth model. The amount of tuning noise added to selected states to insure filter stability is relatively small. The performance of the 32-state reduced-order filter is discussed in the Filter Evaluation Section. Its performance is slightly degraded from that of the 51-state filter.

The subsequent, 23-state, reduced-order filter is generated by combining effects of the following 9 states into the remaining error terms (see the Appendix).

a. 3 accelerometer scale factor states
b. 3 gyro scale factor states
c. 3 vertical channel states

These 9 error states are the most significant error contributors of the remaining 32 states. The performance of the resultant 23-state filter is tuned. A moderate amount of noise strength is added to insure filter stability. The performance comparison of the 23-state filter is discussed in the Filter Evaluation Section. Performance of the tuned 23-state is somewhat degraded from that of the 32- or 51-state filters.

The final, 18-state, reduced-order filter error model is generated by deleting the following 5 states (see the Appendix).

a. 4 code loop error states
b. 1 user clock drift

These 5 error states are the least significant of the 23 primary error contributors. Deletion of these states should only be considered under severe computational limitations. A significant amount of tuning noise must be added to insure filter stability, especially for the conditions of reduced numbers of available satellite measurements.

The performance comparison of the 18-state, reduced-order filter is provided in the Section 5.3. The performance of the 18-state filter is significantly lower than that of the other three filters. The projected performance of the joint, 18-state filter promises to be better than the F-16 15-state cascaded filter performance demonstrated during recent F-16 GPS/INS flight test [53]. With four satellites available, the projected performance of the joint 18-state filter is 10 feet compared to the demonstrated performance of 27 feet (9 meters) of the cascaded filter.

5.3   Filter Evaluation

Before the simulated flights each filter is aligned using the flight filter. During alignment zero velocity, barometric altimeter, and the four GPS measurements are incorporated. This provides the filter with the opportunity to immediately begin estimating those error states which are observable. The significant advantage during the alignment is that the flight filter begins building up the cross-correlation terms in the covariance matrix between those observable states. Additionally only a single filter is required, rather than separate alignment and flight filters. During the alignment the estimated INU errors and GPS receiver user clock bias are reset at 10 second inter-

vals. During the first 2000 second flight segment these errors are reset at 30 second intervals

A representative weapon delivery mission for a modern attack aircraft is generated using PROFGEN. This mission flight trajectory is shown in Figure 14. The significant segments are the take-off, cruise, and descent during the 0 - 2000 second segment into the simulation. The second significant segment, from 2000 to 2800 seconds, is the ingress, pop-up maneuver, weapon delivery, evasive maneuvers, and escape, shown as an enlargement in the bottom of Figure 14 This segment is examined with various GPS satellite availabilities. The final segment, 2800 to 7295 seconds, is the return to home base. Single-run Monte Carlo simulation is performed for each case analyzed for relative filter performance comparisons
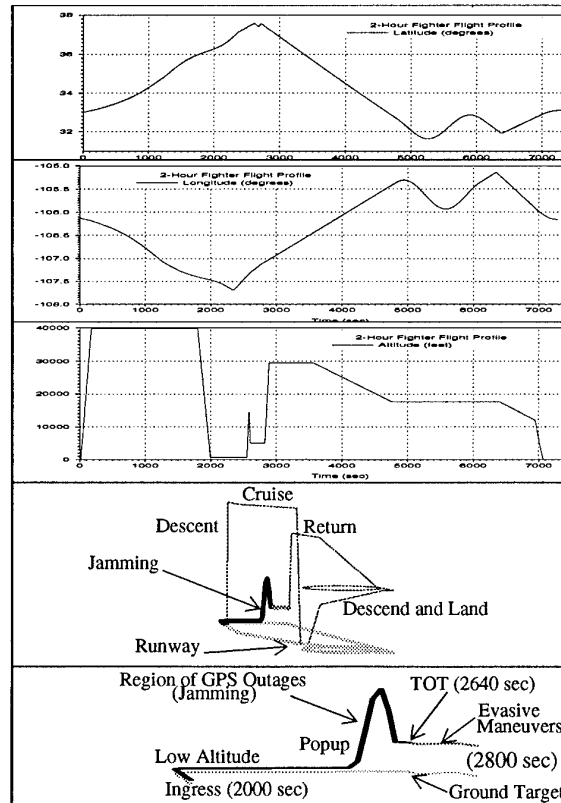
Figure 14. 2-Hour Fighter Flight Profile

These reduced-order filters are exercised using MSOFE simulation against this mission profile. During the takeoff, climb out, and cruise enroute to the target area four GPS satellites are simulated as continuously available until the 2000 second (33.33 minute) point. A single-run Monte Carlo simulation is performed from takeoff until the 2000 second point for each reduced-order filter and the truth model. This occurs immediately after the completion of descent to a low altitude. In this analysis only pseudo-range measurements are simulated.

Position and velocity error performance plots, for this initial 2000 second segment, are shown in Figures 15 and 16 The appropriately labelled performance plots provide the Kalman filter confidences expressed in 1-σ values for the tuned 18, 23, 32, and 51-state reduced-order models (gray line) as well as the estimation error (black line). These are presented as root sum squared (RSS) three-dimensional position and velocity errors (magnitudes of the position and velocity error vectors) versus time. It is important to point out that each filter was tuned until reasonably acceptable performance is obtained. No attempt is made to *locally optimize* the performance of any of the filters.
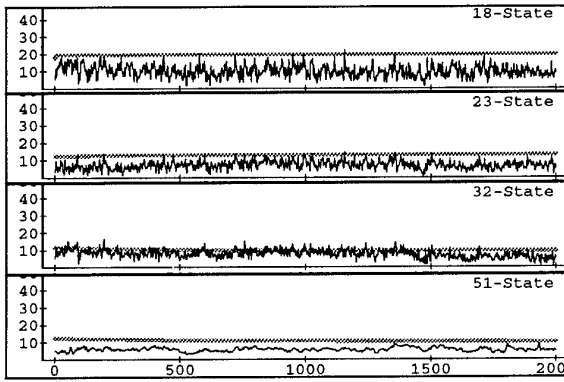
Figure 15. 3-D Position Error (ft), Take-off to 2000 Sec.
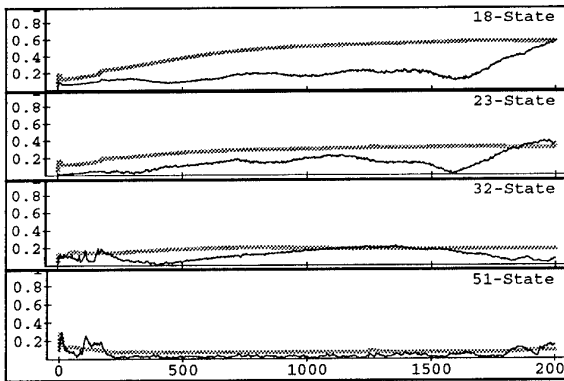
Figure 16. 3-D Velocity Error (fps), Take-off to 2000 Sec.

At approximately 10 minutes prior to time-on-target (TOT), (2000 seconds) varying qualities of GPS coverage are assumed. A separate Monte Carlo simulation run is performed for an additional 800 seconds for each reduced-order model against each assumed GPS coverage case. Three-dimensional RSS position and velocity error plots, for the reduced-order model filter cases, Figures 17-28, provide the Kalman filter confidence (1-σ value) and actual estimation errors.
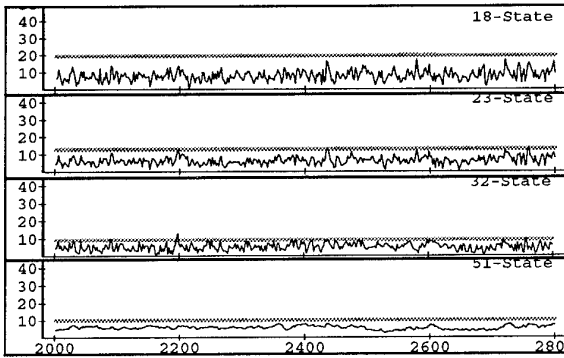
Figure 17. 3-D Position Error (ft), Four Satellites
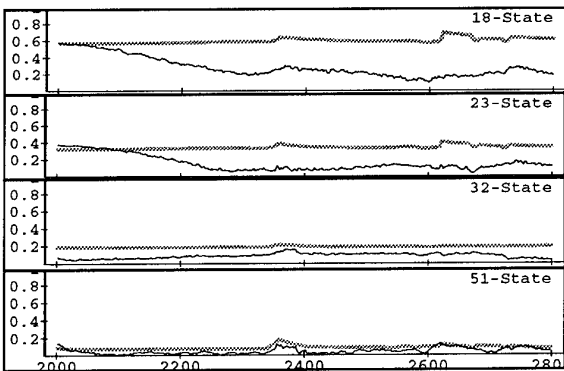
Figure 18. 3-D Velocity Error (fps), Four Satellites

The four-satellite performance is the baseline for comparison of the following three, two, one, zero, and occasional satellite measurement availability cases. The GPS satellite geometry assumed in this analysis is with a GDOP of approximately 3 with one satellite directly overhead. Note the relative performance of the significantly reduced-order filters against the 51-state filter, whose performance is very close to that of the truth model (98-state filter) performance (not shown). Also note the superior performance of the 51-state filter, especially in velocity. When the four satellite measurements are available the position performance differences are not significant, approximately 6-7 ft for the 51-state case compared to approximately 8-10 ft for the 18-state case, but more noisy and with less filter confidence for the 18-state case. The velocity performance is significantly better for 51-state versus the 18-state case.

When the overhead satellite is assumed not available (the three-satellite case), the 18-state filter position-performance in degrades by approximately a factor of two. The other filter performance remains essentially the same as the full satellite coverage case. Note that the 18-state filter confidence (1-σ value) is significantly larger than that of the other filters. Initially a higher confidence (smaller 1-σ value) was chosen. That worked well with four satellites. However when satellite availability was reduced, this particular filter became unstable. At that point the filter was detuned resulting in the larger 1-σ value as shown in the following performance plots.
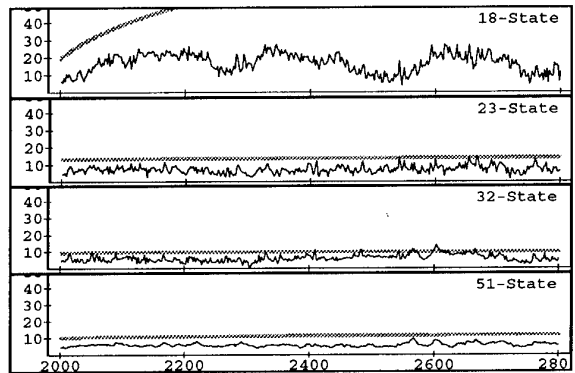
Figure 19. 3-D Position Error (ft), Three Satellites
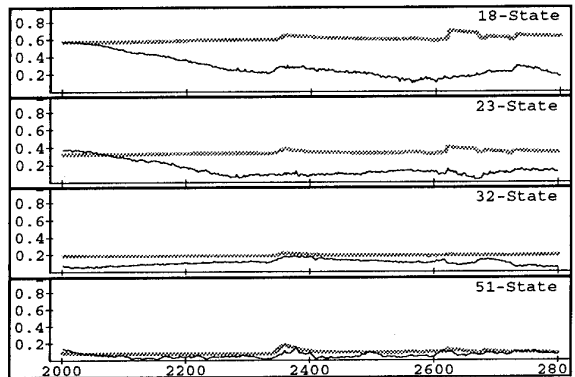
Figure 20. 3D Velocity Error (fps), Three Satellites

In the next case, only two satellite measurements are incorporated during this mission segment. Note that in all cases analyzed, the vertical channel model includes at least one state for the barometric altimeter, and the associated atmospheric pressure variation errors modeled as first-order Markov processes. The presence of this state appears to significantly enhance the performance of each filter, including the 19-state filter. In the two-satellite case the position error performance of the various filters varies substantially. The 32- and 51-state filters perform better than the 23- and 18-state filters. This is especially true during the first 400 seconds of this flight segment.
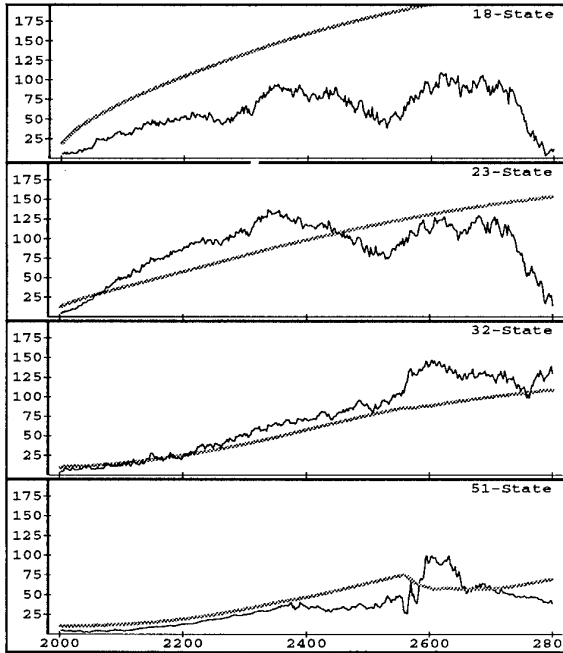
Figure 21. 3-D Position Error (ft), Two Satellites
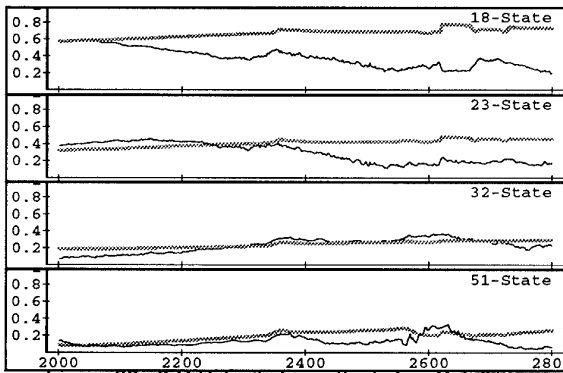
Figure 22. 3-D Velocity Error (fps), Two Satellites

The next case analyzed is that of a single satellite and barometric altimeter measurements. The position error performance is somewhat lower than that of the earlier cases. Noteworthy is the velocity performance of 51 and 32-state filters, where the error increases is minor. Also note the advantage of the additional states in the two higher order filters, especially during the first 400 seconds.

Although a particular single satellite is chosen it is believed that very similar performance will result with any other one. The exception would be if the single satellite used were the overhead one, in the direction that the barometric altimeter is already providing substantial information. The majority of the error growth rate observed in these position error plots is due to loss of estimation accuracy of the vertical channel (barometric) errors due to growing uncertainty of the atmospheric correlated Markov error growth.
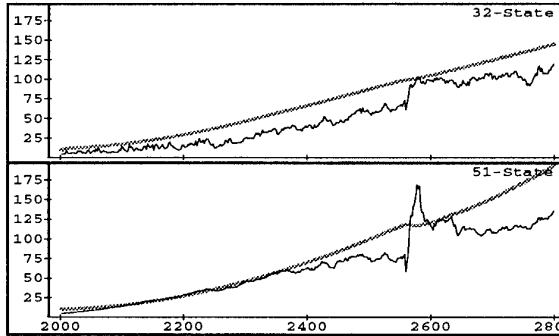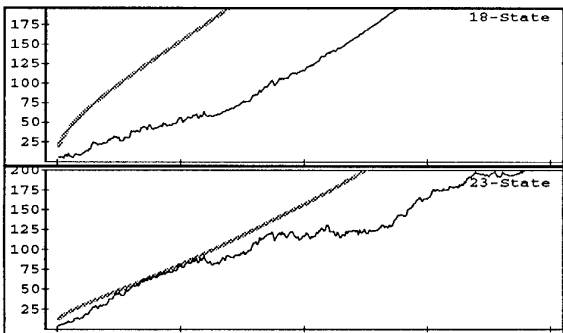
Figure 23. 3-D Position Error (ft), One Satellite
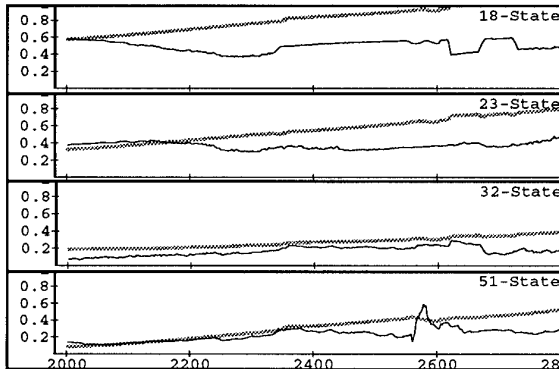
Figure 24. 3-D Velocity Error (fps), One Satellite
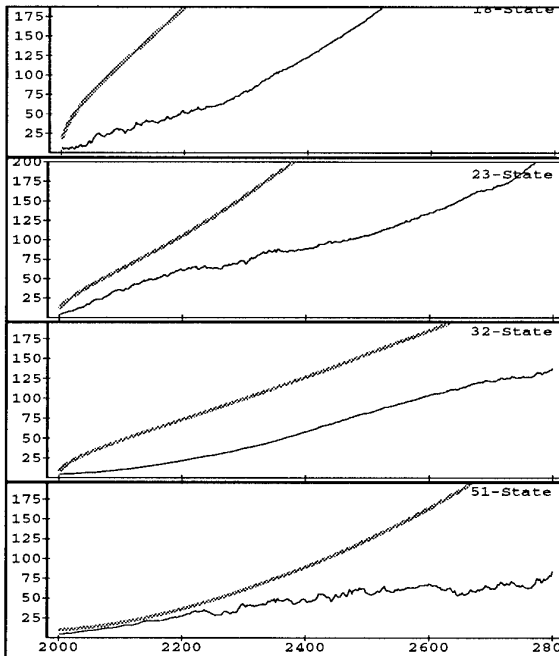
Figure 25. 3-D Position Error (ft), No Satellites
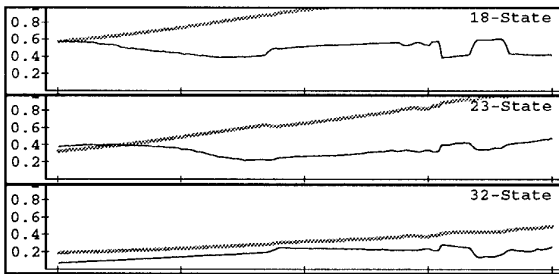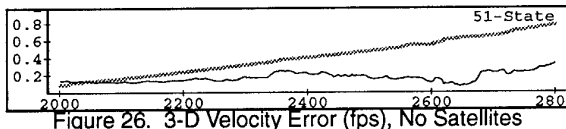
Figure 26. 3-D Velocity Error (fps), No Satellites

Figures 25 and 26 show the performance of each filter when only barometric altimeter measurements are available. In this case each filter relies on the quality of estimates of INS and barometric measurement errors. The estimation quality of these errors depends on the estimation quality of the GPS measurement errors during the time when GPS measurements are available. The position error performance is surprisingly good for the 51-state filter compared to that of other filters. The velocity error performance is also quite good, especially when compared to the velocity performance of cascaded filters [53].
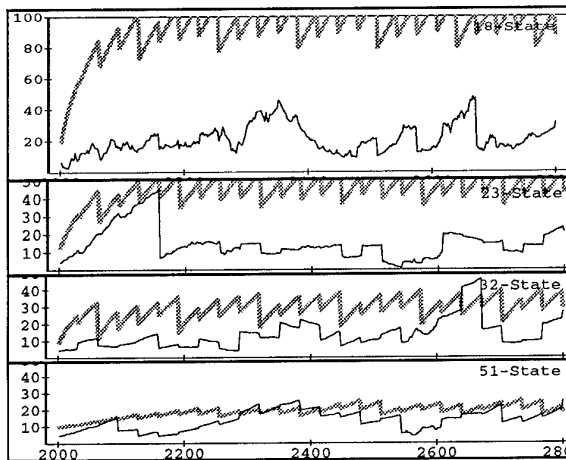
Figure 27. 3-D Position Error (ft), Occasional Satellites
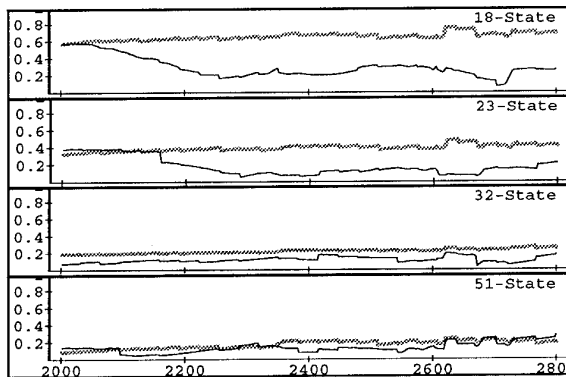
Figure 28. 3-D Velocity Error (fps), Occasional Satellites

This final case is the most interesting. It is a simulation of a situation where only occasional GPS measurements from various satellites are available. Although in reality the time distribution of the sporadic availability of GPS measurements is expected to be random, here it is approximated with a single satellite measurement available each 30 seconds on a rotating basis, that is each satellite is available only once every 120 seconds. The results are interesting, especially for the 32 and 51-state filters. The position error performance is near that of the four-satellite case, within a factor of two or so. *The velocity performance of each of these filters is nearly unchanged from the four satellite case!*

5.4 Simulation Results

The RSS position and velocity error results for each assumed GPS coverage case are shown in the previous Figures 17 through 28. These simulation results predict that a single Kalman filter receiving pseudo-range and delta-range measure-

ments and modelling significant GPS and barometric altimeter measurement errors can perform significantly better in actual flight than the current cascaded approach.

The results of the zero satellite case provide a *best case* approximation (upper bound) for a cascaded filter performance, with fewer than 4 satellites under these same conditions. This is because the cascaded filter does not use the filtered GPS position and velocity data when fewer than 4 satellites are available. Note that for all the reduced-order filters the RSS performance at the 10 minute point for the 3 satellite case is essentially the same as for the 4 satellite case, with the exception of the 18-state filter. The one, two and three satellite case 23-state RSS position error values at the 10 minute point are 130, 100, and 10 feet respectively. These values are significantly better than the *best case* value that could be expected for the cascaded filter approach using fewer than 4 satellites. Similarly the 23-state velocity errors at the ten minute point for the one, two and three satellite cases are 1.0, 0.8, and 0.6 feet per second. Note the position error growth rate for all Kalman filters is substantially less for the 1-satellite case versus the zero satellite case. Another interesting case is when only occasional single GPS measurements are simulated by a single measurement from each of four satellites, in turn, at 30 second intervals. It is clear for all the reduced-order Kalman filters that even occasional pseudo-range measurements provide sufficient information to significantly improve filter performance over that of the cascaded filter.

5.5 Centralized Filter Advantages

Precision first pass weapon delivery with exceptionally accurate navigation are vital to the operational survivability of attack aircraft. Systems such as the Global Positioning System (GPS) are currently being fielded and integrated with inertial navigation systems (INS) to support such a capability. It offers the potential for less costly weapon targeting systems if target coordinates relative to the GPS reference can be provided. The current standard GPS receiver requires simultaneous signals from 4 satellites (with good geometry) to calculate and output useful position and velocity information. However, signals from 4 satellites may not always be available, especially in and around heavily defended target areas protected by high power jammers. These high power jammers, even when recognized and appropriately nulled out by adaptive array GPS antenna, can result in large sectors of the sky surrounding these jammers no longer being monitored for any GPS satellite signals. It has been recommended [50] that changes be made to the existing GPS equipment and new aircraft Kalman filter algorithms to provide graceful degradation of aircraft navigation performance when only one, two, three, or intermittent satellite signals are available. In a paper by Lewantowicz and Keen [50], they quantify the level of relative performance expected for these conditions. They also analyzed the relative performance of various Kalman filter mechanizations as a function of filter size.

The analysis performed in [50] demonstrates that a significant level of navigational performance improvement can be realized for the case when fewer than four satellites are available by providing GPS pseudo-range and delta-range data on the MILSTD 1553 data bus and by implementing a single integrated aircraft Kalman filter. This same filter should be used for ground alignment as well. In fact, this filter configuration continues to estimate all of the observable INU and measurement errors including the misalignment errors. The quality of these error estimates improves significantly during any maneuver segments, be it horizontal or vertical. In fact the flight trajectory is rather sterile in the sense that all maneuvers are perfect in the sense that a straight and level flight is precisely straight and level. In reality pilots, and to a lesser degree autopilots, are not physically capable to fly these segments as perfectly. Thus any rolling, pitching, yawing, or linear accelerations, even relatively

small in magnitude should help the filter by making the less observable errors become more observable.

Additional significant performance improvement can be realized by incorporating as many significant error states into this single integrated Kalman filter as possible subject to the computational memory and thruput constraints. For the applications where the position, velocity, and attitude accuracy are the primary system contributors to weapon-on-target miss distances, this GPS/INS integration approach promises to significantly reduce those error components.

6. DIFFERENTIAL GPS CONCEPTS

This section discusses the application of differential GPS (DGPS) by way of an example to demonstrate how DGPS can be used to enhance performance of CIRIS. The enhanced system is the ENRS, which takes advantage of newer ring laser gyro strapdown INS (LN-93) technology, ground transponders from the current CIRIS, and DGPS measurements. Analysis conducted using MSOFE is presented. An enhanced navigation reference system (ENRS) truth model is developed. This truth model serves as a baseline for full- and reduced-order Kalman filter designs. Results suggest that the proposed ENRS with DGPS aiding can provide a navigation position solution at least one order of magnitude better than the current CIRIS.

This section presents the development of a 48 state post-processing Extended Kalman Filter (EKF) to augment the CIRIS navigation solution with DGPS pseudo-range measurements. An EKF is developed instead of a smoother due to the limited computer storage capacity available. The EKF is limited to 48 states to ensure 24 hour turn-around time for post-processing real measurements during INS testing at CIGTF. An in depth description of the design of this EKF can be found in [55]. Sections 3.4, 3.5, 3.6, 3.7 describe the truth models used in this research. The error-states for the test reference (INS and RRS) are described and then a short explanation of the DGPS implementation and error-states is given. Section 6.5 presents the DGPS measurement equation. Full- and reduced- order filter models are described in the Filter Models Section 6.6. Results of a 25-run Monte-Carlo simulation analysis are presented in Section 6.7. This section ends with a discussion of conclusions based on the performance of the ENRS.

6.1 Differential GPS Concepts

A GPS receiver determines its position by computing the range from the receiver to the satellite vehicles (SVs). This range is measured by determining the difference between the signal transmit time and the signal received time. It is in error due to several sources [32]. The DGPS concept improves the performance of the GPS receiver by eliminating errors that are common in a local vicinity between a user set and a GPS receiver that is at a known location. This latter receiver is referred to as the reference receiver. The placement of a receiver at a precisely surveyed location is the key to estimating DGPS corrections. After application of DGPS corrections, it is possible to achieve accuracies of 1 to 3 meters instead of the normal 10 meter accuracy one can achieve using P code without selective availability (SA). Use of DGPS allows one to mitigate much of the SA affect.

A good discussion of DGPS methods and concepts is found in a paper by Blackwell [56]. There are several methods of employing DGPS. Basically there are two methods to apply DGPS corrections; corrections can be applied in the pseudo-range measurement domain, or they can be applied in the solution domain. The key difference between the two approaches is in the specific data that must be transmitted between the two receivers involved, and whether an uplink versus downlink communication link is required.

Some DGPS implementations include:

1) Uplink of pseudo-range corrections for user vehicle processing

2) Uplink of position corrections for user vehicle processing

3) Downlink user receiver raw data for ground station DGPS processing

4) Downlink of uncorrected position data from user vehicle for ground station DGPS processing.

Methods 1 and 2 will provide corrected position on board the user vehicle. Method 1 is normally preferred when multiple users are to be accommodated. Methods 3 and 4 are used in applications where precise position information is required at a central ground facility. Method 3 is the most demanding to implement of these two.

There is also a concept called relative GPS where one achieves differential accuracies of the relative position between two users. This concept *is not* the same as DGPS. In relative GPS there is no reference receiver, but one simply takes the difference between two receivers in this same area. When this is done, the common biases seen by both receivers are essentially removed. In this technique, absolute position accuracy is not improved at all. However, relative position is often of greater interest in applications such as targeting or formation control.

6.2 CIRIS

CIRIS is described in Section 4. Up to this point in time, CIRIS has been considered more accurate than the test articles and has formed the baseline for determining the performance of aircraft INSs. Recently, state-of-the-art aircraft INSs have been developed (and many more in the design stage) approaching the accuracy of CIRIS. Interestingly, many of these new INSs use the Global Positioning System (GPS) to increase their accuracy. In order to use CIRIS as a baseline against these new INSs, CIRIS must be enhanced to provide an order of magnitude more accurate navigation solution. By using DGPS measurements to augment the navigation solution of CIRIS, it is possible to increase the accuracy of CIRIS to produce an order of magnitude better estimate of the navigation solution.

6.3 Truth Models

The 89-state truth model used in this research is divided into 3 sub-models based on the 3 subsystems forming the ENRS. The first sub-model contains a 41-state INS model consisting of 40 Litton LN-93 INS error-states and a single baro-altimeter error-state. The second 26-state sub-model defines the error-states associated with 6 transponders modeled in the RRS. The last sub-model contains the 22 error-states associated with DGPS measurements from 4 space vehicles (SVs).

The 41 error-state INS truth model is derived from the complete 93-state Litton LN-93 INS truth model [24]. The 41 error-states utilized in this research are the dominant error sources present in the 93 error-state truth model. This error-model is composed of 13 general error-states as well as 28 dominant gyro, accelerometer, and baro-altimeter error-states. The 52 gyro and accelerometer error-states not present in this truth model are not large in magnitude and have been combined with other states or eliminated according to the recommendations of Lewantowicz and Keen [50]. The first 13 states in the 41-state INS truth model represent the general error-states, states which are combinations of several other states in the error model. Three position (latitude, longitude, and altitude), three velocity (east, north, and vertical), and three platform tilt (east, north, and azimuth) errors along with four vertical channel errors comprise this group of error-states. Following the first 13 general error-states are 4 error- states modeled as first-order Markov processes. These states are the X, Y, and Z accelerometer noise states and the baro-altimeter state. The final 24 error-states are gyro and accelerometer error-states modeled as random biases. The gyro error-states include gravity, drift, and scale factor errors. The accelerometer error-states include bias, scale factor,

256

and misalignment errors. The 41-state truth model performance has been compared to the 93-state truth model performance with excellent correlation. A discussion of this comparison and a complete description of the 41-state INS truth model is found in [55].

The transponder based RRS provides the external measurements necessary to update the EKF for CIRIS as it is currently implemented. The RRS uses an airborne transceiver, referred to as the RRS interrogator, along with a computer to calculate position and velocity based on EM signals received from ground transponders. The transponders have been accurately surveyed, so that once the interrogator sends a signal to a particular transponder and after receiving the transponders reply, the computer can accurately calculate position and velocity information. Although only six transponders are modeled in this research, many more (or less) can actually be in range of the RRS interrogator depending on the location of the flight test. There are dozens of transponders located on the CIRIS flight test range and many more located across the country. The 26 error-state RRS truth model contains two types of error-states, two states common to all six transponders and four states which are transponder dependent. The first two states in the truth model are random bias states modeling the effects of user hardware (RRS interrogator) on range and range-rate calibration errors. These two states are modeled with initial variance values of $1\,ft^2$ and $10^{-4}\,fps^2$, respectively. The final 24 states, 4 states for each of six transponders, model transponder x, y, and z position errors and an atmospheric propagation error. The position errors, stemming from the fact that the transponders positions are not perfectly known, are modeled as random biases with initial variance values of $25\,ft^2$. The atmospheric error, induced by the propagation delay, is modeled as a first-order Markov process with a time constant of $300\,seconds$ and a white dynamics driving noise of $6.66\times10^{-3}\,ft^2$. Description of RRS states is found in [55].

6.4 DGPS Truth Model

GPS is designed to be an accurate, stand-alone navigation system. However, for this research, GPS is used as a subsystem to improve the navigation solution of the ENRS. GPS navigation information is obtained from EM signal propagation through the media (space and atmosphere) between the user (ENRS) and each of 4 SVs which the user *locks* into a reception channel of the GPS receiver. In a stand-alone GPS receiver, navigation information is obtained by receiving GPS SV ephemeris data broadcast continuously from each active (*locked-on*) SV, correlating the phase of the signal with a matching signal in the GPS receiver, and correcting for known error sources to produce an accurate range estimate between the user and each SV which is monitored.

In this research, uncorrected range measurements (known as pseudo-range measurements) are channeled to a Kalman filter which provides estimates of the error sources. Common GPS error sources which are considered dominant in this research include receiver clock bias and drift, ionospheric and tropospheric (atmospheric) propagation errors, SV clock, and SV position errors. Other non-dominant errors are also present in a true GPS signal, but are compensated for in this research in the GPS pseudo-range measurement noise. As in the RRS, GPS range measurements make refinements to the ENRS navigation solution possible.

Intermetrics, Inc. is the government sponsored contractor responsible for the DGPS reference station at CIGTF [57, 58].The following discussion of DGPS comes from interviews with Mr. Darwin Abbey and Mr. Scott Dance of Intermetrics, and the DGPS error model described is a combination of Intermetrics description and a course given by Navtech Seminars on DGPS error models [57, 58, 59].

In order to apply differential corrections to GPS measurements, a ground based reference receiver (GBR) is needed as well as an airborne GPS receiver (ABR). Figure 29 shows the basic DGPS system as it is being implemented at CIGTF.
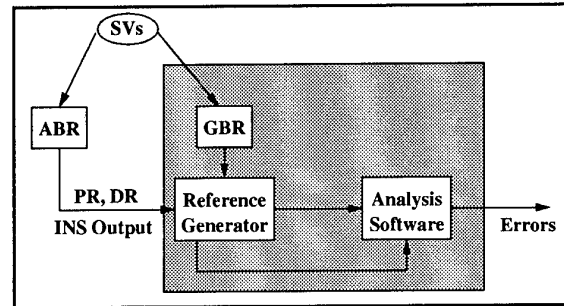
Figure 29. GPS Reference Station

The ground based receiver must be capable of tracking all SVs in view, possibly as many as eleven when the full GPS constellation is placed in orbit. The GBR's antenna position is permanently fixed and surveyed to within centimeter accuracy. A high accuracy rubidium clock is used in place of the GBR's normal clock, greatly decreasing the large clock errors common to GPS receivers. Using the transmitted SV ephemeris data, the GBR computes its position (different from the surveyed position) known as the ground truth. The data (pseudo-range measurements, SV transmitted corrections, GBR applied corrections, ground truth, clock errors, etc...) is fed to a 30386 personal computer which then estimates the SV position errors, SV clock errors, and atmospheric delays with great precision. Because the GBR's true position is accurately known and its clock errors are much smaller than the ABR's, the SV position and clock errors and atmospheric delays are highly observable. This is unlike a normal GPS model where the large ABR's clock errors and vehicle dynamics cause these states to be largely unobservable. These errors, now called differential corrections, are time tagged and stored on magnetic tape or disc. Note that the differential corrections could be immediately transmitted via a data link for real-time differential corrections if the need ever arises [57, 58].

Remembering that the ENRS is a post-processing filter, the raw airborne pseudo-range measurements (which are also stored magnetically and time tagged) now have the differential corrections applied before they are analyzed in the post-processing filter. Of course, this assumes the GBR is tracking the same four SVs the ABR was tracking (a good assumption if the ABR is within the CIRIS test range). Using differential corrections in this manner, the SV clock error is eliminated from each pseudo-range measurement and the SV position errors are nearly eliminated. Depending on the distance between the ABR and the GBR during the flight profile, the atmospheric propagation errors (ionospheric and tropospheric) can be almost totally eliminated for close trajectories or greatly reduced for flights within 200 miles of the GBR. Even with long-range DGPS (flights which extend more than 200 miles from the GBR), the post-processing in the 30386 personal computer eliminates some of the atmospheric propagation errors. The largest remaining errors in the pseudo-range measurements are the ABR clock errors. With this basic knowledge of DGPS, a DGPS error model is now developed assuming that differential corrections have previously been applied to the raw pseudo-range measurements from the ABR [57, 58, 59].

The DGPS error-model is composed of 22 states. The DGPS error-states are similar to any GPS error-state model except for elimination of the SV clock error and reduction of the atmospheric and SV position errors due to differential corrections. The first two error-states in the DGPS error model are the ABR clock error-states modeled by

$$\begin{bmatrix} \dot{x}_{Uclk_b} \\ \dot{x}_{Uclk_{dr}} \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} x_{Uclk_b} \\ x_{Uclk_{dr}} \end{bmatrix} \qquad (119)$$

symbols are defined in Section 9., and equivalencies applied:

$$x_{Uclk_b} = R_{clk_u} = range\ equivalent\ of\ ABR\ clock\ bias$$

$$x_{Uclk_{dr}} = \dot{R}_{clk_u} = velocity\ equivalent.of\ ABR\ clock\ drift$$

The initial state estimates and covariances are [55]:

$$\begin{bmatrix} \hat{x}_{Uclk_b}(t_0) \\ \hat{x}_{Uclk_{dr}}(t_0) \end{bmatrix} = \begin{bmatrix} 0.0 \\ 0.0 \end{bmatrix} \qquad (120)$$

$$P(t_0) = \begin{bmatrix} 9x10^{14}ft^2 & 0 \\ 0 & 9x10^{10}ft^2/sec^2 \end{bmatrix} \qquad (121)$$

Until the ABR clock error is determined, it is the single largest source of error in DGPS range measurements. While the two states discussed above apply to all DGPS measurements, the remaining 20 DGPS error-states are unique to each SV measurement. Each of the 4 EM signals originates from a different SV and travels through different amounts of atmosphere, hence the need for individual atmospheric and SV position errors for each pseudo-range measurement. Note that these errors have a much smaller contribution after differential corrections than they would in a GPS model [59]. The tropospheric and iono-spheric errors are modeled as first-order Markov processes such that:

$$\begin{bmatrix} \delta \dot{R}_{tr} \\ \delta \dot{R}_{ion} \end{bmatrix} = \begin{bmatrix} \dfrac{-1}{500} & 0 \\ 0 & \dfrac{-1}{500} \end{bmatrix} \begin{bmatrix} \delta R_{tr} \\ \delta R_{ion} \end{bmatrix} + \begin{bmatrix} w_{tr} \\ w_{ion} \end{bmatrix} \qquad (122)$$

$$P(t_0) = \begin{bmatrix} 1ft^2 & 0 \\ 0 & 1ft^2 \end{bmatrix} \qquad (123)$$

$$E\{w(t)\} = 0 \qquad (124)$$

$$E\{w(t)w(t+\tau)\} = \begin{bmatrix} 0.001 & 0 \\ 0 & 0.0004 \end{bmatrix} (ft^2/sec^2)\delta(\tau) \qquad (125)$$

The SV position errors are modeled as random bias errors:

$$\begin{bmatrix} \delta \dot{x}_{Si} \\ \delta \dot{y}_{Si} \\ \delta \dot{z}_{Si} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} \qquad (126)$$

$$P(t_0) = \begin{bmatrix} 0.35ft^2 & 0 & 0 \\ 0 & 0.35ft^2 & 0 \\ 0 & 0 & 0.35ft^2 \end{bmatrix} \qquad (127)$$

$$E\{w(t)\} = 0 \qquad (128)$$

$$E\{w(t)w^T(t+\tau)\} = \begin{bmatrix} xxx & 0 & 0 \\ 0 & xxx & 0 \\ 0 & 0 & xxx \end{bmatrix} (ft^2/sec^2)\delta(\tau) \qquad (129)$$

The set of equations above apply to a single SV. The initial covariance values in Equations (123) and (127) and atmospheric error dynamics noise variances in Equations (125) and (129) are obtained from [59] and are modeled slightly larger than Abbey and Dance [57, 58] recommend since a conservative model is developed. There are four such sets of matrix equations for DGPS SV errors modeled in this study. The error-state vector is completely specified in [55].

The sub-models described above for the INS, RRS, and DGPS subsystems compose the state dynamics equations (with initial conditions). These are entered into MSOFE simulation so that true error-state values can be computed and analyzed against the extended Kalman filter's estimated error-states. Now that all the error-states used in the models have been described, the 3 types of measurement equations can be overviewed.

6.5 Measurement Equation

In order for any Kalman filter to properly estimate state variables, external measurements encompassing some of the state variables must be provided to it. In this research, 3 types of measurements are provided to the post-processing EKF. The first type of measurement is the baro-altimeter measurement used to stabilize the vertical channel in the INS. This measurement occurs at 1-second intervals. The second measurement type is the RRS range measurement from each of six transponders, with measurements from all six occurring every six seconds. These measurements are well documented in [37] and [55], respectively, if the reader wishes to further research these areas. The final type of measurement, and the measurement this research focuses on, is the DGPS pseudo-range measurement.

The DGPS pseudo-range measurement is best described as a GPS pseudo-range measurement with differential corrections applied to it, as mentioned in the previous section. These measurements are obtained from 4 geometrically optimal SVs (out of 24 total and 11 possible, when the full GPS constellation is in orbit) at 10 second intervals. The full satellite constellation has been programmed into this simulation and geometric dilution of precision (GDOP) calculations are performed to obtain the 4 SVs used for measurement purposes, with satellite changes occurring when necessary for optimal GDOP considerations [46]. After applying differential corrections, the measurement equation is modeled as:

$$R_{DGPS} = R_t + \delta R_{tr} + \delta R_{iono} + \delta R_{clk_u} + v \qquad (130)$$

where the symbols are defined in Section 9.

The DGPS pseudo-range equation above includes the true range (which can never be known exactly) along with terms which reflect sources of error and uncertainty inherent to DGPS range measurements. With this equation in hand, it is desirable to formulate a difference measurement in the DGPS model. However, to accomplish this, two sources of range information must be obtained. The first source is the range measurement coming from the DGPS reference station and modeled by Equation (130). The second range estimate is constructed by differencing INS-indicated position and SV (broadcast) positions to calculate the range. This derivation is explained in detail in [55], and its result yields:

$$\delta z = R_{INS} - R_{DGPS}$$

$$= -\left[\frac{x_s - x_u}{|R_{INS}|}\right]\delta x_u - \left[\frac{y_s - y_u}{|R_{INS}|}\right]\delta y_u - \left[\frac{z_s - z_u}{|R_{INS}|}\right]\delta z_u$$

$$+ \left[\frac{x_s - x_u}{|R_{INS}|}\right]\delta x_s + \left[\frac{y_s - y_u}{|R_{INS}|}\right]\delta y_s + \left[\frac{z_s - z_u}{|R_{INS}|}\right]\delta z_s$$

$$- \delta R_{tr} - \delta R_{ion} - \delta R_{Uclk} - v \qquad (131)$$

The terms δx_u, δy_u, and δz_u directly relate to the INS position error terms (latitude, longitude, and altitude) while δx_s, δy_s, and δz_s are the SV position errors. The pseudo-range measurement noise variance is $9\ ft^2$ when DGPS pseudo-range measurements occur every 10 seconds. The true whole-valued range (R_t) formerly present is cancelled in the differencing operation. The bracketed coefficients in the equation above appear in the EKF update equations. The full derivation of this equation along with the EKF update and propagation equations is found in [55]. The DGPS error-state truth model and measurement

equations have now been shown, so now the two filter models used in this research are described.

6.6 Filter Models

Two filters are developed to implement DGPS pseudo-range measurements into the ENRS, a full-order filter of 89 error-states and a 48 error-state reduced order filter. The first filter is described as full-order since the 89 filter states are modeled exactly like the 89 truth model states. For the reduced-order filter, 41 states are eliminated after analysis determined their magnitudes are small when compared to other error-states being modeled.

The 89 error-state full-order filter model is used to establish a baseline for comparison to any reduced-order filters developed. As a reminder, this filter is composed of 41 INS error-states, 26 RRS error-states, and 22 DGPS error-states. Baro-altimeter, RRS range, and DGPS pseudo-range measurements are used in this EKF to provide accurate position estimates. At this time, no velocity aiding measurements are incorporated into this filter, although research is continuing in this area and eventually RRS range-rate and DGPS delta-range measurements will also be used to provide accurate velocity aiding as well [55].

In order to ensure 24-hour post-processing time of a typical flight profile used to test INSs, CIGTF requested that the final EKF be less than 70 error-states. This means that at least 19 error-states have to be eliminated. More could be eliminated if position (and eventually velocity) accuracy can be maintained. With these objectives in mind, filter order reduction was performed and a new 48 error-state EKF was developed. The first step in state reduction was taken when 20 DGPS error states were eliminated because their magnitudes were very small compared to the magnitudes of the ABR clock bias and drift states. A typical ABR clock bias error is on the order of 10^8 ft, while the differentially corrected atmospheric and SV position errors are 1 or 2 ft. The EKF often had trouble accurately estimating the small error-states because of this magnitude difference, so their elimination did not affect filter performance. Of course, the white Gaussian measurement noise variance was increased slightly to compensate for the eliminated states [55].

As a second step to eliminate more states from the EKF, the recommendations of Lewantowicz and Keen are again followed when 21 more INS error-states are eliminated. The eliminated states are small magnitude gyro and accelerometer errors, and small increases in the dynamics noise variance in some of the remaining states is necessary to compensate for their removal. Note that the removal of these 21 states significantly increased the attitude errors, and slightly increased the velocity errors, but the position errors remain relatively close to the values of the full-order filter. Again, for a complete description of the INS error-state filter reduction, consult [55]. At this point, enough states have been eliminated to ensure quick analysis while producing accurate estimates of the navigation solution. The EKF has been reduced to 48 error-states; 20 INS, 26 RRS, and 2 DGPS. It is important to note at this point that the RRS error-states could also be reduced. However, since this research is focused on the DGPS error model, no RRS filter state reduction is performed. The results of filter performance are now presented.

6.7 Results

This sub-section presents and discusses the results of the filter performance of the full-order and reduced-order filters. The current CIRIS 1-σ position and velocity accuracies are used as a baseline to judge the filters performance. Also, a 46 error-state filter called CIRIS-46 is included in this comparison. This filter is composed of the 20 error-state reduced-order INS model and the 26 error-state RRS model and uses baro-altimeter and RRS range measurements. This filter gives an indication of the per-

formance gained by using a higher-order post-processing filter, and will lend itself for a better comparison in the performance increase of DGPS alone. Each filter's performance is tested with MSOFE utilizing 25-run Monte-Carlo analysis of a 2-hour fighter flight profile. Figure 14 shows the latitude, longitude, and altitude information as well as a 3 dimensional representation for this fighter flight profile. As is readily seen, it incorporates several turns and dives to simulate a realistic flight profile.

Figure 30 shows typical latitude, longitude, and altitude plots utilizing the data provided by MSOFE. These particular plots show the relationship between the true 1σ errors and the filter's estimate of the errors. The single dashed line plots the mean error, or the difference between the true latitude error and the filter's estimate of that error. The two dotted lines are plots of the mean error ± the true 1-σ value. Finally, the solid outermost lines are the filter's estimate of the errors. Again these plots are an average over 25 Monte-Carlo runs and represents an adequately tuned filter.
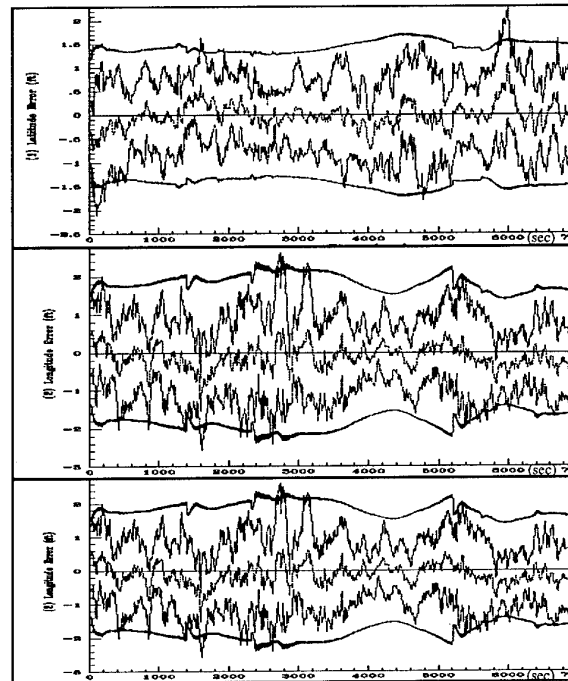
Figure 30. 89-State ENRS Filter, Latitude, Longitude, and Altitude Errors

The 89-state full-order ENRS filter's position and velocity estimation error performance is shown in Table 2 along with the current CIRIS and the CIRIS-46 filter. The north velocity error results are not presented due to space considerations, but their values are comparable to the east velocity error results. The temporal average of the ensemble average simply means that the true 1σ values for all 25 simulation runs are averaged together at ten second intervals and then this average is averaged to give the number shown. As seen, the full-order filter provides order of magnitude better position and velocity estimates than the current CIRIS filter and much better performance than CIRIS-46. The increase in velocity error estimation is attributed to the fact that if the filter's position error estimation is increased, it can also estimate the velocity errors with greater precision. It is easy to see the benefit that DGPS pseudo-range measurements have on increasing the accuracy of CIRIS.

The results of the performance analysis on the 48-state reduced-order ENRS EKF are also shown in Table 2. As a reminder, this filter is composed of 20 INS, 26 RRS, and 2 DGPS error-states. There is a slight decrease in position and velocity performance; however, this filter performs much better than either CIRIS or

CIRIS-46. Notice that the reduction in filter states (gyro and accelerometer bias error-states) did affect the filter's ability to estimate the velocity error (and attitude error, though not shown). It is again safe to assume that DGPS measurements dramatically increase the accuracy of the navigation solution. It is obvious from the results that DGPS pseudo-range measurements do indeed increase the accuracy of CIRIS and the post-processing 89- and 48-state ENRS filters increase the navigation solution position accuracy by one order of magnitude

Table 2. Temporal Average of True Error Ensemble

Filter	Lat (ft)	Long (ft)	Alt (ft)	East-Vel (fps)	Up-Vel (fps)
CIRIS	14.00	14.00	40.00	0.100	0.400
CIRIS-46	3.12	6.84	18.09	0.046	0.100
ENRS-89	0.84	1.04	3.80	0.013	0.042
ENRS-48	0.90	1.32	3.05	0.027	0.044

6.8 Summary on Differential GPS

The ENRS post-processing filters developed in this research increase the accuracy of the CIRIS system over one order of magnitude in position, so the objectives of this research are met. It is hoped that further research incorporating the velocity aiding measurements (both RRS and DGPS) will also increase the velocity estimates one order of magnitude. Even though it was not shown in this paper, the decrease in INS states from 41 to 20 decreased the attitude states estimation accuracy [55]. It is prudent to increase the INS states back up to 41 so that this accuracy can be maintained. This would increase the reduced-order filter to 69 error-states, still meeting the objective of having less than 70 filter error-states but increasing attitude estimation accuracy. Further work in this area will include incorporating the velocity-aiding measurement equations into the filter to aid in velocity estimation. Also, reducing the RRS filter error-states down from 26 will decrease processing time. The atmospheric error-state for each transponder has a small magnitude, so it should be possible to combine this error with the range bias error. The transponder position errors, now modeled as random biases in x, y, and z components, could be modeled as a single line-of-sight random bias in the filter. These steps reduce the 26 error-state RRS model to 8 error-states and produce a significant decrease in computer analysis time.

7. GPS/INS- CORNERSTONE OF SENSOR FUSION

The GPS contribution to navigational accuracy has been clearly established in military weapon systems. However the GPS signal provides enormously broader potential mission capability which has not yet been exploited. The GPS signal, when properly combined with information from INS and other sensors, provides position, velocity, attitude, and time of unprecedented accuracy and robustness. These ten elements, defined in a geometric vector in Section 3.1, are common, in various combinations, to most of the avionics functions. When viewed from a system perspective, this high precision information, can be thought of as the integration basis, or a reference set, which offers opportunities for reconfiguration of the offensive, defensive, communication, navigation, and other sensors. Various integration architectures for fusion of these sensors can inherently enhance, enable, or severely limit these potential mission capabilities. The choice of integration architecture, can directly and profoundly affect performance, cost of integration, cost of ownership, and exploitation of much greater mission capability.

7.1 Introduction

A multitude and variety of commercial and civil GPS applications are exploding, which in turn are fueling the explosion of commercial GPS receiver hardware development. The interest

in this tremendous capability is most clearly evident in the growth of attendance at the Institute of Navigation ION/GPS conferences and in the number and variety of commercial and civil GPS application products. Of significance to this discussion is the fact that the free enterprise spirit has produced a multitude of ingenious solutions to such a vast number and variety of commercial and civil problems.

There are various factors affecting the difference in utilization of the GPS by the military services, especially when compared to that of the commercial and civil sectors. Some of these factors are clearly identifiable and discussed in this paper while others are not known or not clearly understood. The more demanding military operating environment and mission requirements are among the more obvious differences. However there are other factors, some of which are systemic. Of the systemic factors, perhaps the most significant one is that of avionics integration architectures. Although GPS/INS integration is used as an illustration of a systemic factor, the discussion is presented from a broader perspective of system level avionics integration. Avionics architecture options are examined from the standpoint of the cost of integration, cost of ownership, and the yet unexploited mission capability.

7.2 Federated Architectures

The evolution and limitations presented by the federated architectures is described in the INTRODUCTION Section 1. This evolution, and federated architecture subsystem outputs resulted in unintended constraints on deep integration. However, despite some challenges in providing for multiple requirements, this limited integration concept was successful in that performance of some sensors was improved, to a lesser or greater degree, by providing information from other sensors. Most of the information provided from other sensors was at the post-processed output level or in the domain of output integration. This post-processing modifies the "raw" signal, available from the detector, in bandwidth, noise statistics, and other electronic compensation characteristics. Although theoretically it possible to "undo" the output processing if the internal algorithms are known, this is not practical due signal-to-noise losses, distortions, and typically high cost. The impact of this post-processing is that it limits the degree to which the information from other sensors can be utilized, where appropriate, to improve the performance of other sensors. This federated architecture, with its limitations, is the prevalent environment that the military GPS user equipment is required to interface to.

7.3 Military GPS Equipment

When the military user equipment segment of the GPS was fielded, the navigation function also was provided in a "black box" (really gray colored), Standard Rockwell Collins Receiver IIIA. This box was required to perform as a stand-alone instrument, while at the same time it output information for several other devices, such as the INS, the fire control computer, or fire control sensors. These outputs were defined under various standards, such as the MILSTD 1553, but other restrictions were also applied. For example, the GPS receiver did not include the pseudo-range and delta-range information on the MILSTD 1553 bus output. Security considerations prevent the output of the Selective Availability (SA) corrected precision pseudo- and delta-range information outside some boundaries of an avionics sensor container. Instead the Receiver IIIA outputs navigation position, velocity, and other data after significant processing by its own internal Kalman filter. This filter is based on a generic INS model that is not optimized to any particular INS type such as gimbaled or strapped-down system.

If an aircraft requires a navigation solution optimized to a particular INS or if other sensor information is required for an optimal navigation (or other variable) solution, this GPS receiver constraint limits the aircraft Kalman filter designer to a cas-

caded, and suboptimal, filter implementation. This results in significantly degraded performance and jam resistance, and requires a strongly disadvantageous GPS availability or visibility definition. Also use of GPS filtered position and velocity outputs to drive a separate aircraft Kalman filter (e.g. in a fire control computer) can lead to filter stability problems, since the GPS and the INS data are time correlated [60]. This potential problem is minimized by typically processing GPS measurements in the aircraft filter at a rate much slower than they are output from the GPS receiver. This helps in reducing the correlation between GPS and INS measurements, but not without penalties.

Addition of the GPS receiver to the air weapon systems increased the mission capability tremendously from that of INS alone, or non-GPS aided INS. The GPS Receiver IIIA was designed and installed in the aircraft (and other vehicles) essentially as a navigator. It provides an unprecedented improvement in navigation accuracy over that of the previously available from the INS, which was aided by barometric altimeters, doppler velocity sensors, and other navigation instrument outputs. Fundamentally this increase in navigation accuracy provided substantial increase in mission capability. Global scale navigation accuracies of 16m position and 0.1m velocity are two orders of magnitude better than possible before introduction of GPS. However this performance increase is limited when compared to the potential performance, robustness, jam-resistance, etc. of a properly integrated GPS/INS. Performance increase in the navigation function as well as other improvements and benefits remain largely unexploited.

7.4 Advanced Avionics Architectures

In contrast to the federated avionics architectures, the integrated avionics architectures offer numerous advantages. These advantages can be understood from a broader perspective of avionics system flexibility, redundancy, fault tolerance, real-time reconfigurability, decreased life cycle cost, adaptability, and other attributes. These architectures consist of multi-function, software programmable modules in common avionics enclosures. Many of the mission functions can be performed by individual or groups of modules. This concept replaces the discrete functions of the federated avionics architectures. The Integrated Communications, Navigation, Identification Avionics (ICNIA) and the INtegrated Electronic Warfare System (INEWS), are examples of such architectures, and are the basis for the U.S. Air Force Advanced Tactical Fighter (ATF), recently designated as the F-22, and the U.S. Army's next generation helicopter the LHX. The U.S. Navy is considering this avionics architecture for its AX system. Selection of modular avionics standard is a topic of high interest to the North Atlantic Treaty Organization (NATO), in particular the Four Power Group consisting of United States, United Kingdom, France, and Germany. The essence of the question before this group is not whether integrated avionics should be considered, but what form should the standards take for the avionics architectures and for one of its key elements, the modules.

Many advantages result from the integrated architecture concept. The essentially raw signals are brought into a single framework where real-time management of the available resources can be exercised. This immediately offers opportunities for real time reconfiguration for a spectrum of requirements, be it to account for failures, or reconfigure to adapt to a changing mission phase, or to respond to changing mission requirement, such as just identified enemy threat requiring a different and immediate response.

Perhaps the most significant advantage of integrated avionics is that most of the signals are available in various forms and stages of processing within the physical confines of this architecture. The communication, navigation, and identification signals in ICNIA, the defensive avionics sensors signals in INEWS, and the offensive avionics sensors signals are available in a variety of forms, from the "raw" to the fully processed for a variety of integration approaches. Signal level sensor fusion ranging from combining of two or three sensors to large scale fusion of a multitude of sensors is possible. In any of these cases, the availability of the "raw" pseudo- and delta-range GPS measurements provides unprecedented opportunities. When the GPS is properly combined with INS, this two-sensor subset produces a significantly more accurate, robust, jam-resistant, less costly navigation, attitude, and time reference set. This reference set is very important because not only does it provide superior navigation solution, but it also becomes an (geometric) integration reference for combining (fusing) of other sensor information.

When other sensor information, which contains either absolute or relative navigation, attitude, or time information, is processed jointly with that of GPS and INS, then additional benefits are available. For example the reference biases, and other errors, of these sensors are often observable and estimable, thus providing a cross-calibration function. For those sensors where this cross-calibration is significant, several opportunities immediately appear. For example, the absolute performance tolerances on those sensors perhaps could be relaxed, thus reducing the complexity, cost, and reliability of those sensors. In some cases more than one sensor performs a similar or an overlapping function, or measures information also measured by some other sensor. When the information from this ensemble of sensors is properly fused, some of those sensors are either redundant, or their omission could result in minor and acceptable mission performance degradation but at a significant reduction in cost. Another possibility may be the realization of a common and highly accurate time reference that could be derived from this joint fusion of signals which have the GPS time as a common reference. Also, when information from other sensors is jointly fused with that of GPS and INS, the reference set accuracy and robustness is further improved.

These are only few examples of what may be achievable by integrated architectures. It is essential to emphasize, that to realize this vision, system and component level modeling, analysis, and simulation are required to identify the system level sensitivities and perform various trade-offs.

7.5 GPS/INS Integration

The GPS/INS integration is a special subset of the larger set of advanced avionics architectures and sensor fusion discussion. It is treated separately because of the very synergistic properties that a properly integrated GPS and INS functions can provide. In many vehicles that do not, and may not have integrated avionics architectures, an opportunity exists for installing a properly integrated GPS/INS in a single box. The miniaturization of the GPS receiver processing is more than ready for imbedding into the INS. For those who may feel attachments to the GPS as the primary function, a possibility exists of imbedding the ring laser gyro (RLG), fiber optic gyro (FOG), or the hemispheric resonating gyro (HRG) INS into the GPS enclosure. The benefits derived from such an integrated system are increases in accuracy, jam resistance, reliability, fault tolerance, and decreases in cost of the required INS and certainly in life cycle cost.

Signal level fusion, as discussed earlier, produces better results. For an eloquent discussion of the synergistic benefits of proper integration of these two elegant and proven system, the GPS and INS, see the now classic, in this author's view, paper by Cox [30]. The vehicle dynamic maneuvering and the potential jamming of the GPS signals require the proper integration of the GPS with an INS. The significantly more demanding performance requirements of the military mission environment drive the implementation solutions in somewhat different directions than those of the civil and commercial applications.

Several analyses and experiments support the discussion by Dr. Cox and have pointed out significant mission capability advan-

tages of the proper integration. For example in the analysis in [54] and [50] the integration is performed at the raw pseudo-range signal level. That analysis demonstrates that a significant level of navigational performance improvement can be realized for the case when fewer than four satellites are available and pseudo-range and delta-range measurements are processed in a single integrated aircraft Kalman filter. Incidentally, this same filter should also be used for ground alignment as well. In fact, this filter configuration continues to estimate all of the observable INS, GPS, and barometric altimeter measurement errors including the gyro and accelerometer misalignment errors throughout all flight phases from power-up.

Thus a GPS/INS, properly integrated in a single box offers several advantages which are not possible with the federated GPS/INS integration. Those advantages range from significant reduction in life cycle cost to significantly better overall system performance. Placing the Kalman filter inside this box and on a "Kalman filter chip" should unconstrain this GPS/INS integration and unburden the fire control computer. The GPS jamming performance must be redefined, because even momentary lock-on to the GPS signals at relatively large (several minute) intervals provides performance of an essentially unjammed GPS! Of course the best solution is to integrate the GPS and INS with other avionics sensors in a modular integrated architecture.

7.6 Conclusions

The challenge for the military research and development community is to vigorously exploit the simultaneous arrival of the GPS, the explosion in computational capability, and availability of the integrated and modular avionics architectures for weapon systems. These factors offer unprecedented opportunities for much greater exploitation of avionics sensor fusion. With proper fusion of the multitude of information available from the variety of sensors aboard a weapon system, much greater benefits can be derived from the information contained in the GPS signal. This sensor fusion is a strong function of the avionics architecture and the variety of forms of information readily available to any of the fusion algorithms. In integrated and well-fused sensor avionics architectures, in addition to significant increases in performance potential, the concepts of standardization, system failure, and mission capability require redefinition. Entire sensor suites can then be revisited with the question: are all of the current sensors required for military mission performance?

8. ACKNOWLEDGMENTS

9. LIST OF SYMBOLS

a_E	earth's equatorial radius
a_{res}	angular resolution
A	system dynamics matrix
ABR	airborne receiver
$AFIT$	Air Force Institute of Technology
B	system control distribution matrix
bps	bits per second
$BPSK$	binary phase shift keying
$CADC$	Central Air Data Computer
$CIGTF$	Central Inertial Guidance and Test Facility
$CIRIS$	Completely Integrated Reference Instrumentation System
$C_e^n(t)$	direction cosine matrix relating e- to n-frames

D	direct control input to measurement
DCM	direction cosine matrix, e.g. C_e^n
$DGPS$	differential GPS
$ECEF$	earth-centered earth-fixed reference frame
EKF	extended Kalman filter
e_e	eccentricity of earth's ellipsoid
$E(\bullet)$	expectation operator
f	specific force vector: $[f_x\ f_y\ f_z]^T$
FOG	fiber optic gyroscope
G	geometric vector: $[r,\ v,\ \phi,\ t]^T$
G	driving noise w distribution matrix
GBR	ground-based receiver
GMT	Greenwich Meridian Time
GPS	Global Positioning System
h	true altitude, altitude above the reference ellipsoid
$\bar{h}$	altimeter output
H	measurement distribution matrix
HOT	higher order terms
HRG	hemispherical resonator gyroscope
I	identity matrix
ION	Institute of Navigation
INS	inertial navigation system
K_1	baro altimeter model gain
l	longitude
L	latitude
L_1	1575.42 MHz L-band signal
L_2	1227.6 MHz L-band signal
LOS	line-of-sight
$MSOFE$	Multimode Simulation for Optimal Filter Evaluation
NRS	Navigation Reference System
pdf	probability density function
p_i	time derivative operator (with respect to i-frame)
$PROFGEN$	profile generator
PSK	phase shift keying
r	position vector: $[r_x\ r_y\ r_z]^T = R\ u_{LOS}$
r_{sv}	$[x_{sv},\ y_{sv},\ z_{sv}]^T$ true SV position vector coordinates
r_t	$[x_t,\ y_t,\ z_t]^T$ transponder position vector coordinates
r_u	$[x_u,\ y_u,\ z_u]^T$ true user position vector coordinates
R	range; measurement noise covariance matrix
R_{clk_u}	user clock delay
$R_{clk_{sv}}$	SV clock advance
R_{code}	code loop delay
R_{DGPS}	DGPS pseudo-range measurement
R_{iono}	ionospheric delay
R_N	$= a_E / \sqrt{1 - e_e^2 sin^2 \varphi}$
R_P	compensated pseudo-range
R_{RP}	raw pseudo-range
R_{trop}	tropospheric delay
R_t	true range
RF	radio frequency
RLG	ring laser gyroscope
RRS	range/range-rate system
$\dot{R}$	range-rate
SA	selective availability
SAR	Synthetic Aperture Radar
SV	space vehicle (GPS)
s	Laplace transform operator
$(\bullet)_{sk}$	skew-symmetric matrix
t	time
t_0	initial time
$(\bullet)^T$	matrix transpose operator

TEC	total electron count
u	control input vector: $[u_x \; u_y \; u_z]^T$
u_{LOS}	unit-line-of-sight (ULOS) vector
v	zero-mean white Gaussian measurement noise
v	velocity vector: $[v_x \; v_y \; v_z]^T$
w	white, Gaussian driving noise vector: $[w_x \; w_y \; w_z]^T$
$WGS\text{-}84$	World Geodetic Survey 1984 datum
x	state vector: $[x_1 \; x_2 \; ... \; x_n]^T$
δh	baro pressure altitude variation error
δh_a	baro instrument output dynamic lag error
δh_b	baro instrument random bias error
δh_B	total baro altimeter error
δh_{sf}	baro instrument scale factor error
δR_{clk_u}	user clock bias (range equivalent) error
$\delta \dot{R}_{clk_u}$	user clock drift (range-rate equivalent) error
$\delta R_{clk_{SV}}$	SV clock advance bias range equivalent error
δR_{code}	tropospheric delay residual
δR_D	range designation error
δR_i	ionospheric delay residual
δR_{tr}	tropospheric delay residual
δr_t	$[\delta x_t, \; \delta y_t, \; \delta z_t]^T$ transponder survey error
δr_u	$[\delta x_u, \; \delta y_u, \; \delta z_u]^T$ user position error
$\delta \theta_D$	elevation designation error
$\delta \psi_D$	azimuth designation error
ΔG	differential geometric vector: $[\Delta r, \; \Delta v, \; \Delta \underline{\phi}, \; \Delta t]^T$
ζ	eigenvector
θ	elevation angle (LOS)
λ	eigenvalue, terrestrial longitude
$\underline{\lambda}$	vector of eigenvalues
σ_{R_D}	range designation error standard deviation
$\sigma_{\dot{R}_D}$	range-rate designation error standard deviation
σ_{ψ_D}	azimuth designation error standard deviation
σ	noise standard deviation
σ^2	noise variance
τ	integration time variable
τ_{dy}	instrument dynamic time constant
τ_{pco}	pressure altitude variation correlation time constant
Φ	state transition matrix
ϕ	attitude "vector," rotation vector: $[\phi_x \; \phi_y \; \phi_z]^T$
φ	geodetic latitude
ψ	azimuth angle (LOS)
$\underline{\omega}^n_{ie}$	earth angular velocity vector with respect to i-frame, coordinatized in n-frame: $[\omega^n_x \; \omega^n_y \; \omega^n_z]^T$
Ω^n_{ie}	skew symmetric matrix form of $\underline{\omega}^n_{ie}$
$\partial(\bullet)$	partial derivative operator
$\nabla\bullet$	gradient operator; defined as a row vector

References

[1] Britting, K.R., "Inertial Navigation Systems Analysis", New York, Wiley-Interscience, 1971.

[2] Broxmeyer, C., "Inertial Navigation Systems", New York, McGraw-Hill Book Company, 1964

[3] Farrell, J.L., "Integrated Aircraft Navigation", New York, Academic Press, 1976.

[4] Fernandez, M. and Macomber, G.R., "Inertial Guidance Engineering", Englewood Cliffs, NJ, Prentice-Hall Inc., 1962.

[5] Kayton, M., and Fried, W., "Avionics Navigation Systems", New York, Wiley, 1969.

[6] Markey, W.R., and Hovorka, J., "The Mechanics of Inertial Position and Heading Indication" London, Spottiswoode, Ballantyne & Co, 1961.

[7] Maybeck, P.S., "Stochastic Models, Estimation, and Control", Volumes 1-3, San Diego, Ca, Academic Press, 1982.

[8] O`Donnell, C.F., "Inertial Navigation Analysis and Design" New York, McGraw-Hill, 1964.

[9] Pitman, G.R., "Inertial Guidance", New York, Wiley, 1962.

[10] DeCarlo, R.A., "Linear Systems - A State Variable Approach with Numerical Implementation", Englewood Cliffs, New Jersey: Prentice Hall, 1989.

[11] Gaskill, J.D., "Linear Systems, Fourier Transforms, and Optics", New York, New York, John Wiley and Sons, 1978.

[12] Nering, E.D., "Linear Algebra and Matrix Theory", New York, Wiley, 1970.

[13] Reid, J.G., "Linear Systems Fundamentals", New York, New York, McGraw-Hill, 1983.

[14] Strang, G., "Linear Algebra and its Applications', New York, Academic Press, Second Edition, 1980.

[15] Kalman, R.E., "A New Approach to Linear Filtering and Prediction Problems", Transactions ASME, Series D: Journal of Basic Engineering, 1960

[16] Gelb, A., and others, "Applied Optimal Estimation", Cambridge, MA: MIT Press, 1984.

[17] Jazwinski, A.H, "Stochastic Precesses and Filtering Theory", New York, Academic Press, 1970.

[18] Meditch, J.S., "Stochastic Optimal Linear Estimation and Control", New York, McGraw-Hill, 1969.

[19] Cunningham, J.R., "Performance of GPS-aided INS During High-Dynamic Maneuvers", MS Thesis, AFIT/GE/ENG/87D-12, School of Engineering, Air Force Institute of Technology, W-P AFB, OH, December 1987.

[20] Carlson, N., and Musick, S., "User's Manual for a Multimode Simulation for Optimal Filter Evaluation (MSOFE)". AFWAL-TR-88-1138, W-P AFB OH, AFWAL/AAAN-2, Apr 1990.

[21] "PROFGEN - A Computer Program for Generating Flight Profiles", Avionics Laboratory, AFWAL/AARN-2, Wright-Patterson AFB, OH, November 1976.

[22] Vander Velde, W.E., "Strapped-Down Inertial Systems", Course 16.43 Notes, Department of Aeronautics and Astronautics, MIT, 1983.

[23] Wrigley, W., and others, "Gyroscopic Theory, Design and Instrumentation", Cambridge, MA, MIT Press, 1969.

[24] Litton Guidance and Control Systems, "Performance Accuracy (Truth Model/Error Budget) Analysis for the LN-93 Inertial Navigation Unit". DID No. DI-S-21433 B/T:CDRL No. 1002. Woodland Hills, CA, Jan 1985.

[25] Pearson, R.A., "Advanced Reference System", Proceedings of the Twelfth Biennial Guidance Test Symposium, October 1985.

[26] Windall, W.S., and Grundy, P.A., "Inertial Navigation System Error Models", Technical Report, AF-SWC TR 73-26, Intermetrics Inc., Cambridge, MA, 11 May 1973.

[27] Savage, P.G., "Strapdown Sensors", AGARD Lecture Series No. 95, Strapdown Inertial Systems - Theory and Applications, June, 1978.

[28] Savage, P.G., "Laser Gyros in Strapdown Inertial Navigation Systems", Proceedings: IEEE Position Location and Navigation Symposium, San Diego, CA, November, 1976.

[29] Spillaker, J.J.,"GPS Signal Structure and Performance Characteristics", Navigation, Journal of the Institute of Navigation, Summer 1978.

[30] Cox Jr., D.B., "Integration of GPS with Inertial Navigation Systems", Navigation: Journal of the Institute of Navigation, pp 144-153, 1980.

[31] Martin, E.H., "GPS User Equipment Error Models", Global Positioning System, Janiczek P.M., editor, Institute of Navigation, Washington, D.C., 1980.

[32] Milliken, R.J., and Zoller, C.J., "Principles of Operation of NAVSTAR and System Characteristics", Navigation: Journal of The Institute of Navigation, 25:3-14, Summer 1978.

[33] Bate, R.R., et al, "Fundamentals of Astrodynamics", New York, John Wiley and Sons, Inc., 1971.

[34] Solomon, J.K., "CIRIS Special Study, EENG699 Final Report", School of Engineering, Air Force Institute of Technology, W-P AFB OH, March 1989.

[35] "GPS Aided LN-93 INS," EENG735 Class Project Report, Department of Electrical and Computer Engineering, Air Force Institute of Technology (AU), W-P AFB, OH, Spring 1990.

[36] Snodgrass, F.B., "Continued Development and Analysis of a New Extended Kalman Filter for the Completely Integrated Reference Instrumentation System", MS Thesis, AFIT/GE/89M-8, School of Engineering, Air Force Institute of Technology, W-P AFB, OH, March 1990.

[37] Solomon, J.K., "Development of the Extended Kalman Filter for the Advanced Completely Integrated Reference Instrumentation System (CIRIS)". MS Thesis, AFIT/GE/89M-8. School of Engineering, Air Force Institute of Technology, W-P AFB OH, March 1989.

[38] The Analytic Sciences Corporation (TASC), "Models for Aided Inertial Navigation System Sensor Errors". TR-312-3, Prepared under Contract No. DMA700-74-C-0075, Reading, MA, Feb 1975.

[39]Musick, S.H., "Radar measurements and measurement matrices," Tech. Report AFAL-TM-76-48, Wright Laboratory (AFMC), W-P AFB, OH, May 1976.

[40]Grettenberg, T.L., and others, "SAR Targeting Accuracy Analysis Program (STAAP) - Hughes Aircraft Company, Final Report," Tech. Rep. AFWAL-TR-88-1074, Wright Laboratory (AFMC), W-P AFB, OH, October 1988.

[41]Jeffers, R.H.Biererman, and others, "SAR Targeting Accuracy Analysis Program (STAAP) - Emerson Electric Company, final report," Tech. Rep. AFWAL-TR-88-1030, Wright Laboratory, W-P AFB, OH, June 1988.

[42]Schmidt, D.J. "Error Analysis of an Air-to-Surface Missile with a Synthetic Aperture Radar Seeker," NAECON, pp. 363--370, Dayton, OH, May 1981.

[43]Nelander, J.C., and others, "Performance Sensitivities of Position and Velocity Determination Using SAR Vision Guidance Updates," Proceedings of the Southeast Conference on Energy and Information Technologies in the Southeast, pp. 90--94, Columbia, South Carolina, April 1989.

[44]Kenton, A.C., and others, "Vision Guidance Update· Synthetic Aperture Radar (SAR) Multiple Image Exploitation for Position and Velocity Determination," SPIE Millimeter Wave and Synthetic Aperture Radar, vol. 1101, pp. 158--169, 1989.

[45]Layne, J.R., "Integrated Synthetic Aperture Radar and navigation Systems for Targeting Applications", WL/AAAS Draft Report, W-P AFB, OH, Aug 1992

[46] Stacey, R.D., "A Navigation Reference System (NRS) Using Global Positioning System (GPS) and Transponder Aiding". MS Thesis, AFIT/GE/91M-04, School of Engineering, Air Force Institute of Technology, W-P AFB OH, March 1991.

[47] Guidance Test Division. "CIRIS II Test Summary Report," In-House Report, CIGTF, 6585th Test Group, Holloman AFB, NM, May 1988.

[48] Guidance Test Division. "The Completely Integrated Reference Instrumentation System (CIRIS)," Unpublished Contractor Informational Pamphlet, CIGTF, 6585th Test Group, Holloman AFB NM.

[49] Green, G.B., and others, "The GPS 21 Primary Satellite Constellation", Navigation, Journal of the Institute of Navigation, 25:9-24, Fall 1989.

[50] Lewantowicz, Z.H., and Keen, D.W., "Graceful Degradation of GPS/INS Performance With Fewer Than Four Satellites," The Institute of Navigation, National Technical Meeting, Jan 1991.

[51] Solomon, J.K., "CIRIS Special Study, EENG699 Final Report", Department of Electrical and Computer Engineering, Air Force Institute of Technology, W-P AFB, OH, March 1989.

[52] Hirt, E.H., and Yakos, M.D., "Time Dissemination Using NAVSTAR Global Positioning System (GPS) Phase IIB User Equipment", 35th Annual Frequency Control Symposium, USAERADCOM, Fort Monmouth, New Jersey, May 1981.

[53] "Flight Test Report for Performance Validation of the Global Positioning System on the F-16 Block 40 Aircraft," General Dynamics Test Report Number 16PR8816, 19 June 1990.

[54] Lewantowicz, Z.H., and Keen, D.W., "GPS Aided Navigation Performance for Weapon Release with Fewer Than Four Satellites", Joint Services Data Exchange, October 1990.

[55] Negast, W.J., "Incorporation of Differential Global Positioning Measurements for Improved Reference System Performance". MS Thesis, AFIT/GE/91D-41, School of Engineering, Air Force Institute of Technology, W-P AFB OH, Dec 1991.

[56] Blackwell, E.G., "Overview of Differential GPS Methods", Global Positioning System Papers, Institute of Navigation, Washington, D.C., Vol 3, pp 89-100, 1986.

[57] Abbey, D., and Dance, S.D., Intermetrics, Inc. GPS Data Analysis Reference Station, Briefing to AFIT Students, Holloman AFB, NM, June 1991.

[58] Dance, S.D., Personal Interviews., Intermetrics, Inc., Holloman AFB, NM, June 1991.

[59] Navtech Seminars, Inc., "Dynamic Differential GPS," Notebook from Course 311, Alexandria, VA, December, 1989.

[60] Johnson, G.B., and Lewantowicz Z.H., "Closed Loop Operation of GPS Aided INS," Proceedings, ION-GPS 90 Conference, Colorado Springs, CO, September 1990.

Appendix

State	Name	Truth Model	F_{51}	F_{32}	F_{23}	F_{18}
	INS Error States[a]					
$\delta\theta_x$	Angular Position	1	1	1	1	1
$\delta\theta_y$		2	2	2	2	2
$\delta\theta_z$		3	3	3	3	3
ϕ_x	Platform Tilts	4G	4	4	4	4
ϕ_y		5G	5	5	5	5
ϕ_z		6G	6	6	6	6
δV_x	Velocity	7A	7	7	7	7
δV_y		8A	8	8	8	8
δV_z		9A	9	9	9	9
δh	INS Altitude	10H	10	10	10	10
δh_l	INS Lagged Altitude	11H	11	11		
δS_3	Vert. Channel. Aiding K3	12H	12	12		
δS_4	K4	13H	13	13		
b_{xc}	Gyro Correlated Bias	14G				
b_{yc}		15G				
b_{zc}		16G				
Δ_{xc}	Accel. Correlated Bias	17A	14			
Δ_{yc}		18A	15			
Δ_{zc}		19A	16			
δg_x	Gravity Anomaly	20A	17			
δg_y		21A	18			
δg_z		22A	19			
δh_c	Baro Tropospheric	23H	20	14	11	11
δh_a	Baro Instrum. Dynamics	24H				
δh_{SF}	Baro Scale Factor	25H				
δh_b	Baro Fixed Bias	26H				
b_x	Gyro Bias	27G	21	15	12	12
b_y		28G	22	16	13	13
b_z		29G	23	17	14	14
S_{gx}	Gyro Scale Factor	30G	24	18		
S_{gy}		31G	25	19		
S_{gz}		32G	26	20		
χ_1	Gyro Misalignment (6)	33G				
χ_2		34G				
χ_3		35G				
ν_1		36G				
ν_2		37G				
ν_3		38G				
D_{xxx}	Gyro Nonlinearities	39G				
D_{yyy}		40G				
D_{zzz}		41G				
S_{Qbx}	Gyro S.F. Asymmetry	42G				
S_{Qby}		43G				
S_{Qbz}		44G				
Δ_{bx}	Accel. Bias	45A	27	21	15	15
Δ_{by}		46A	28	22	16	16
Δ_{bz}		47A	29	23	17	17
S_{ax}	Accel. Scale Factor	48A	30	24		
S_{ay}		49A	31	25		
S_{az}		50A	32	26		
S_{QAx}	Accel S.F. Asymmetry	51A	33			
S_{QAy}		52A	34			

State	Name	Truth Model	F_{51}	F_{32}	F_{23}	F_{18}
S_{QAz}		53A	35			
f_{xx}	Accel. Nonlinearities (9)	54A				
f_{yy}		55A				
f_{zz}		56A				
f_{xy}		57A				
f_{xz}		58A				
f_{yx}		59A				
f_{yz}		60A				
f_{zx}		61A				
f_{zy}		62A				
μ_1	Accel. Misalignment	63A	36			
μ_2		64A	37			
μ_3		65A	38			
σ_1		66A	39			
σ_2		67A	40			
σ_3		68A	41			
	GPS Error States[b]					
δR_{clk}	User Clock Bias	69	42	27	18	18
δD_{clk}	User Clock Drift	70	43	28	19	
δR_{cd1}	Code Loop Errors (4)	71	44			
δR_{cd2}		72	45			
δR_{cd3}		73	46			
δR_{cd4}		74	47			
δR_{trp1}	Tropospheric Delay (4)	75	48			
δR_{trp2}		76	49			
δR_{trp3}		77	50			
δR_{trp4}		78	51			
δR_{ion1}	Ionospheric Residual (4)	79				
δR_{ion2}		80				
δR_{ion3}		81				
δR_{ion4}		82				
δR_{clk1}	SV Clock Bias (4)	83				
δR_{clk2}		84				
δR_{clk3}		85				
δR_{clk4}		86				
δx_{sv1}	SV (x, y, z) Position Errors	87				
δx_{sv2}		88				
δx_{sv3}		89				
δx_{sv4}		90				
δy_{sv1}		91				
δy_{sv2}		92				
δy_{sv3}		93				
δy_{sv4}		94				
δz_{sv1}		95				
δz_{sv2}		96				
δz_{sv3}		97				
δz_{sv4}		98				

a. The LN-93 INS error state model actually contains 93 error states. 5 states were added to complete the accelerometer misalignments and 3 vertical channel states to account for the baro altimeter and pressure altitude errors. From this total of 98 states, 30 least significant states were deleted to form the error model in this table.

b. The GPS error model contains 30 error states, assuming five independent receivers, ionospheric correction having been made using dual L_1 and L_2 frequencies

FEDERATED FILTER FOR FAULT-TOLERANT INTEGRATED NAVIGATION

Neal A. Carlson
Integrity Systems, Inc.
31 Middlecot Street
Belmont, MA 02178 USA

SUMMARY

This paper describes federated filter applications to integrated, fault-tolerant navigation systems. The federated filter is an optimal or near-optimal estimator for decentralized, multisensor data fusion. Its decentralized estimation architecture is based on theoretically sound information-sharing principles. Federated filters consist of one or more sensor-dedicated local filters, generally operating in parallel, plus a master combining filter. The master filter periodically combines (fuses) the local filter solutions to form the best total solution. Fusion generally occurs at a reduced rate, relative to the local measurement rates. The method is well suited to real-time system implementation, and can provide significant improvements in data throughput, fault tolerance, and system modularity. This paper discusses federated filter applications to integrated navigation systems in terms of operating modes, accuracy, fault tolerance, computational efficiency (throughput), and real-time system features. Numerical simulation results are presented to demonstrate federated filter performance characteristics.

1. INTRODUCTION

Integrated multisensor navigation systems have the potential to provide high levels of accuracy and fault tolerance. The presence of multiple data sources provides functional redundancy as well as greater observability of the desired navigation states. However, that potential has not always been fully realizable, via the application of standard (centralized) Kalman filtering techniques. Applied to multi-sensor systems, centralized filters can result in severe computation loads when implemented in strictly optimal fashion. Worse, when used as the components of two-stage (cascaded) filter architectures, standard Kalman filters can exhibit poor accuracy and unpredictable behavior under certain conditions.

During the past fifteen years, the development of decentralized (or distributed) Kalman filtering methods has received increasing attention. Parallel processing technology, emphasis on fault tolerant system design, and availability of multiple specialized sensors strongly motivate the development of such methods. Potential applications include multisensor navigation systems, multisensor tracking systems, and other data fusion systems.

Early contributions to optimal decentralized filter theory were made by Speyer [1], Chang [2], Willsky et al [3], Levy et al [4], and Castanon et al [5]. While providing useful insights, these early methods did not appear practical for real-time navigation system applications, due to their computational burdens, inter-filter databus loads, and/or model restrictions.

Kerr proposed an attractive decentralized filtering structure [6], in which several parallel filters process data from separate navigation subsystems, and a master filter combines their outputs. Fault-tolerant aspects of this structure were

emphasized. Previous work (e.g., [1-4]) was cited as a general theoretical foundation, and no implementation equations for mechanizing the decentralized filter structure were presented.

A theoretically sound yet practical decentralized filtering method was developed by Bierman at JPL [6], for orbit estimation purposes. While this method was optimal and efficient on a per-cycle basis, it required the master filter to operate at the maximum local measurement rate, and the local filters to have infinite process noise (hence no memory). Bierman and Porter [7] extended that approach to decentralized parameter estimation. Belzer and Cho [8] also provided an extension of that approach.

Hashemipour et al [9] developed another unique decentralized filtering method during this same period. This method required inter-filter data communication at each measurement update cycle, which can lead to heavy databus loads.

More recently, Carlson [10-12] developed the "federated" filtering method based on rigorous information-sharing principles. This method provides globally optimal or near-optimal estimation accuracy, with a high degree of fault tolerance, and is practical for real-time, distributed navigation system applications. It was adopted as the basis of the Air Force's Common Kalman Filter for highly fault-tolerant, next-generation navigation systems [13]. The federated filter method is quite general, with several different information-sharing modes. Some modes require simultaneous updates from the LFs to the MF, while others permit independent updates.

Covino and Griffiths [14-15] subsequently developed the related "net information approach" to decentralized estimation. This globally optimal/near-optimal method is based on information-sharing principles similar to those of the federated method. It employs two local filters (one processes measurements, the other does not) and a differencing algorithm to determine the new information gained over each period. Local elements pass their new information to the master filter independently for data fusion.

Decentralized filters have also been examined in connection with multiple-model adaptive estimation. Watanabe [16] developed a decentralized multiple-model filter based on the Speyer-Chang approach [1, 2]. Martin and Carlson [17] extended the federated filter method to obtain a computationally efficient distributed model adaptive estimator. Both the number (multiplicity) and state-size of the adaptive filter banks are reduced, by implementing multiple sensor models at the local filter level, and combining the most likely local filter solutions in the master filter.

This paper focuses on practical applications of the federated filtering technique to integrated, fault-tolerant navigation systems. The remaining sections of this paper describe: (2) the distributed filtering problem, (3) limitations of standard Kalman filters, (4) the new federated filtering technique, (5) system architecture options, (6) real-time implementation features, (7) numerical simulation results, and (8) conclusions. Appendix A presents implementation equations, and Appendix B presents a mathematical derivation of the federated filter as a partitioned optimal estimator.

266

2. ESTIMATION PROBLEM STATEMENT

In concept, the federated filter is a partitioned estimation method. It employs a two-stage (cascaded) data processing architecture, in which the outputs of local, sensor-related filters are subsequently combined by a larger master filter, as illustrated in Figure 1. (This figure shows the major flow of information, but not all possible data exchanges.) As indicated, each LF is dedicated to a separate sensor subsystem, and also uses data from a common reference system, generally an inertial navigation system (INS).
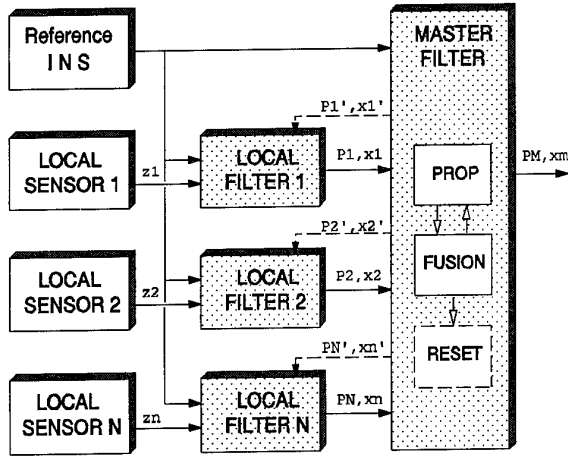

Figure 1: General Federated Filter Structure

The federated filter technique comprises a weighted least-squares solution to the following linear (or linearized) estimation problem. First, consider a system state vector $\underline{x}$ that propagates from time point t' to t according to the following dynamic model:

$$\underline{x} = \Phi \underline{x}' + G \underline{u} \qquad (1)$$

Here, Φ is the state transition matrix between time points t' and t, G is the process noise distribution matrix, and $\underline{u}$ is the additive uncertainty vector due to white process noise acting over the timestep. The error $\underline{e}_o$ in the initial state estimate $\hat{\underline{x}}_o$ and the sequential values $\underline{u}_j$ are uncorrelated, per the following error statistics (subscripts j and k refer to different time points t_j and t_k, and δ_{jk} is the Kronecker delta):

$$\hat{\underline{x}}_o = \underline{x}_o + \underline{e}_o; \quad E[\underline{e}_o] = \underline{0}; \quad E[\underline{e}_o \underline{e}_o^T] = P_o \quad (2)$$

$$E[\underline{u}_j] = \underline{0}; \quad E[\underline{u}_j \underline{u}_k^T] = Q_j \delta_{jk} \qquad (3)$$

$$E[\underline{u}_k \underline{e}_o^T] = 0 \implies E[\underline{u}_k \underline{e}_j^T] = 0, \quad k>j \quad (4)$$

Our system also has access to external measurements $\underline{\tilde{z}i}$ from $i = 1..n$ separate local sensor subsystems. Measurements from different local sensors are independent, and comprise disjoint data sets. The discrete measurements from sensor #i at time t_j are linearly related to the true state $\underline{x}_j$:

$$\underline{\tilde{z}i}_j = HI_j^T \underline{x}_j + \underline{vi}_j \qquad (5)$$

Here, HI_j^T is the sensor #i measurement observation matrix (often defined without the transpose), and $\underline{vi}_j$ is the additive, random measurement error. The sequential error val-

ues $\underline{vi}_j$ are uncorrelated, with these statistics:

$$E[\underline{vi}_j] = \underline{0}; \quad E[\underline{vi}_j \underline{vi}_k^T] = RI_j \delta_{jk} \qquad (6)$$

$$E[\underline{vi}_k \underline{e}_o^T] = 0; \quad E[\underline{vi}_k \underline{u}_j^T] = 0 \qquad (7)$$

$$E[\underline{vi}_k \underline{vm}_k^T] = E[\underline{vi}_k \underline{vm}_j^T] = 0 \quad (i \neq m) \quad (8)$$

This last condition is crucial: *measurement errors from different sensors i and m are statistically independent.* Disjoint sensor data sets permit the total estimation problem to be divided into sensor-related partitions with independent measurement processes, as described further in Section 4.

3. LIMITATIONS OF STANDARD KALMAN FILTERS

The primary limitations of standard Kalman filtering methods when applied to multisensor navigation systems and/or systems with embedded local filters are these:

a) heavy computation loads, and potential inability to keep up with high-rate sensor measurements;

b) poor fault-tolerance in terms of detecting gradual sensor faults, and of quickly regenerating a good solution after a failure; and

c) inability to correctly process pre-filtered data in a cascaded (two-stage) filter structure.

The first limitation of a standard, centralized Kalman filter (CF) is fairly obvious. In Figure 1, the presence of several sensors generally implies a relatively large number of filter states n, since each sensor typically introduces one to five measurement bias states. For a single Kalman filter, the per-cycle computation load grows roughly in proportion to $n^3 + \Sigma m \cdot n^2$, where Σm is the total number of measurements across all the sensors. The problem is especially severe when the multisensor suite requires a large, high-rate filter.

The second limitation of a CF relates to fault tolerance. Like any optimal filter, the CF attempts to make its data inputs agree in a weighted least-squares sense, thereby suppressing their differences. Hence an undetected failure in one sensor gets distributed into all of the navigation state and sensor bias estimates, so that they all tend toward agreement. Thus, while measurement residual tests can readily detect sudden "hard" failures, they may completely miss gradual "soft" failures. If the CF does incorporate faulty data from any sensor, its full solution becomes corrupted, and must be reinitialized.

The third limitation of a CF relates to cascaded filter processing, and can best be illustrated by means of an example. Figure 2 shows the major components of a cascaded filter designed for an integrated navigation system composed of an inertial navigation system (INS), Global Positioning System (GPS) receiver/navigator, and a radar subsystem. GPS receiver measurements and INS outputs are processed locally by an embedded GPS/inertial Kalman filter. Periodically, position and velocity outputs from the GPS local filter are incorporated as discrete "measurements" by a Kalman master filter (MF) in the central computer. The MF also uses the same INS reference solution as does the GPS local filter (LF).

Two aspects of this ad hoc cascaded filter design comprise potential causes of accuracy and/or stability problems, given a Kalman MF. First, errors in the position and velocity outputs from the local GPS filter are not sequentially random, but are time-correlated. Second, errors in the LF outputs are not independent of those in the MF state, but are cross-correlated due to common INS errors.
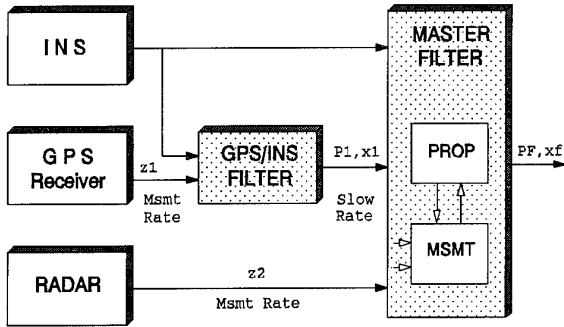
Figure 2: Ad Hoc GPS/Radar Integration Filter

Sequential correlations in the GPS filter outputs mean that each output contains some new information and some old information. Treating each GPS output as entirely new information causes the MF to become over-optimistic regarding its own accuracy (i.e, its covariance gets too small). The resulting problem is especially obvious if we suppose that GPS measurements become unavailable for several minutes. The MF will continue to incorporate successive GPS outputs as fresh "measurements", and to reduce its covariance accordingly, even though those outputs contain absolutely no new information. (During the outage, the GPS filter output covariance will grow, but not fast enough to prevent the MF from reusing the old GPS information.)

A common, ad hoc "fix" for these correlation problems is to limit the GPS incorporation rate in the Kalman MF to once every 10 to 20 seconds. Over these intervals, the GPS navigator output errors typically become decorrelated (particularly the velocity errors), so that they appear sequentially random to the MF. (Rapid decorrelation is caused by large process-noise terms in the GPS local filter; these suppress the time-correlated INS propagation errors, and emphasize the more random GPS measurement errors.)

This ad hoc cascaded filter design approach can yield satisfactory results in terms of estimation accuracy and stability, for many applications. However, questions remain as to whether the same approach will work with different sensors and local filters. For example, would performance remain satisfactory if the GPS local filter were tuned with less process noise, so that its outputs had longer correlation times? Or, if a less accurate sensor than GPS were used? Or, if other sensors were added to the navigation suite? Such questions provide further motivation to seek a theoretically correct foundation for designing and implementing cascaded navigation filters.

4. FEDERATED FILTER METHOD

The new federated filtering technique presented here avoids the theoretical and practical difficulties described in Section 3 by means of a simple yet effective information-sharing methodology. The advantages of information sharing, as implemented by the new federated filtering technique, are:

o increased measurement data throughput by parallel operation of local filters, and by data compression within local filters;

o enhanced system fault-tolerance by maintaining multiple component solutions to improve fault detection and recovery capabilities; and

o improved accuracy and stability of cascaded filter operations, via use of theoretically correct estimation algorithms.

The basic concept of the information-sharing approach

implemented by the federated filter is this:

o divide the total system information among several component (local) filters;

o perform local time propagation and measurement processing, adding local sensor information;

o recombine the updated local information into a new total sum.

The remainder of this section will illustrate how the federated filter applies information-sharing principles in its use of the n local filters (LFs) and one master filter (MF) in Figure 1. The presentation here will be somewhat heuristic, to emphasize the general concept. Appendices A and B provide implementation details and a more rigorous mathematical derivation of the federated method, respectively.

First, let the full, centralized filter solution be represented by the covariance matrix PF and state vector $\hat{xf}$; the local filter #i solution by PI and $\hat{xi}$; and the master filter solution by PM and $\hat{xm}$. We will use index $i = 1..n$ for the LFs alone, and $k = 1..n,m$ for the MF plus LFs, where $k = m$ represents the MF.

Now, if the LF and MF solutions are statistically independent, they can be optimally combined by the following additive information algorithm, where the inverse covariance matrix P^{-1} is known as the "information matrix":

$$PF^{-1} = PM^{-1} + P1^{-1} + .. + PN^{-1}$$

$$\tag{9}$$

$$PF^{-1}\hat{xf} = PM^{-1}\hat{xm} + P1^{-1}\hat{x1} + .. + PN^{-1}\hat{xn}$$

The key to the new federated filtering method is to construct individual LF and MF solutions so they can be combined or recombined at any time by the above simple algorithm. In particular, the construction avoids the need to maintain LF/LF or LF/MF cross-covariances. The procedure for doing so is the essence of the information-sharing approach. (See Appendix A for equivalent algorithms that simplify the P^{-1} operations indicated above.)

Suppose we start with a full solution $PF, \hat{xf}$. Now, divide that solution so that the $k = 1..n,m$ LFs plus MF each receive fractions β_k of the total information:

$$PF^{-1} = PM^{-1} + P1^{-1} + .. + PN^{-1} \tag{10a}$$

$$= PF^{-1}\beta_m + PF^{-1}\beta_1 + .. + PF^{-1}\beta_n \tag{10b}$$

$$PK^{-1} = PF^{-1}\beta_k \quad or \quad PK = PF \ \beta_k^{-1} \tag{10c}$$

$$\hat{xk} = \hat{xf} \qquad k = 1..n,m \tag{11}$$

To maintain constant total information across the sum in Eq. (10) (the *conservation of information principle*), the share-fraction values β_k must sum to unity:

$$\sum_{k=1}^{n,m} \beta_k = \beta_m + \sum_{i=1}^{n} \beta_i = 1 \tag{12}$$

It is clear that the LF and MF solutions given by Eqs. (10) and (11) can be recombined per Eqs.(9) to yield the correct total solution $PF, \hat{xf}$.

Next, consider the discrete time propagation process. This process can also be performed via independent, parallel operations of the LFs and MF, provided the common process noise information is divided in the same fashion as

were the fused solutions. The covariance propagation equations from time t' to t are:

$$PK = \Phi K\ PK'\Phi K^T + GK\ QK\ GK^T \qquad (13)$$

where $()K$ represents the filter #k matrix values. For the moment, let us assume that the LFs and MF are all full-sized, such that the transition matrices ΦK equal ΦF, and the noise distribution matrices GK equal GF. The process noise covariance matrices QK are governed by the information-sharing principle, however:

$$QF^{-1} = QM^{-1} + Q1^{-1} + .. + QN^{-1} \qquad (14a)$$

$$QK^{-1} = QF^{-1}\beta_k \quad or \quad QK = QF\ \beta_k^{-1} \qquad (14b)$$

Thus, if the PK' ($k = 1..n,m$) values have been initially obtained by Eqs.(10), and the QK values are obtained by (14), then the post-step values PK can again be summed by the simple fusion algorithm (9) to yield the correct total solution PF:

$$\sum_{k=1}^{n,m} PK^{-1} = \sum_{k} \left[\Phi F\ PF'\beta_k^{-1}\Phi F^T + GF\ QF\ \beta_k^{-1}GF^T \right]^{-1}$$
$$\qquad (15)$$

$$= \left[\sum_{k=1}^{n,m} \beta_k \right] \left[\Phi F\ PF'\Phi F^T + GF\ QF\ GF^T \right]^{-1} = PF^{-1}$$

Third, consider the measurement update process. Each LF #i incorporates discrete measurements $\underline{zi}$ from its own unique sensor #i. Measurement information is added to LF #i as follows, where RI^{-1} is the #i measurement information matrix:

$$PI_+^{-1} = PI^{-1} + HI\ RI^{-1}\ HI^T$$
$$\qquad (16)$$
$$PI_+^{-1}\ \hat{\underline{xi}}_+ = PI^{-1}\ \hat{\underline{xi}} + HI\ RI^{-1}\ \tilde{\underline{zi}}$$

where the subscript $+$ refers to post-measurement values. Again, combining the above results by the fusion algorithm (9) yields the correct total solution, i.e., the solution that would be achieved by a single centralized filter processing all of the $i = 1..n$ sensor measurement sets:

$$PM^{-1} + \sum_{i=1}^{n} \left[PI^{-1} + HI\ RI^{-1}\ HI^T \right]$$
$$\qquad (17)$$
$$= PF^{-1} + \sum_{i=1}^{n} HI\ RI^{-1}\ HI^T = PF_+^{-1}$$

A similar relation yields the correct total state vector $\hat{\underline{xf}}_+$.

Now, Eqs. (10) to (17) demonstrate that the federated filter solution is the same as that of a single, centralized Kalman filter, and hence is itself globally optimal, when certain implicit assumptions are satisfied: i) each filter employs a single β_k value for all of the full-system states and process noises, and ii) the information fusion and reset (dividing) operations are performed after every measurement cycle. However, much less restrictive conditions can be accommodated, with only a modest loss of optimality.

First, the federated filter can and should be implemented such that the LFs are of minimum size, each LF #i containing only the common INS states and its own unique

sensor-i bias states. Hence the matrices PK, ΦK, GK, and QK contain only the common and bias-k partitions of the full matrices. The β_k fraction values apply only to the common INS states, since only those states are shared among the LFs and MF. Appendix A provides representative implementation equations. (The MF can also be implemented in a global or large form containing all of the sensor bias states; then, each sensor bias-i partition is shared by the MF and one LF, with corresponding β_k values of 1/2.)

It can be shown (Appendix B) that this minimum-LF structure still produces a globally optimal solution in some cases. One case is when the MF retains all the fused information, and the LFs are reset to zero information after every fusion update (i.e., when $\beta_m = 1$ and $\beta_i = 0$). However, for general values of β_m and β_i, the minimum-LF structure introduces a slight loss of information with a conservatively suboptimal result. In practice, the resultant estimation accuracy is almost indistinguishable from that of the globally optimal filter.

Second, the federated filter can perform fusion updates at a reduced rate relative to the LF measurement rates, implying data compression in some or all of the LFs (multiple LF measurement sets are "compressed" into the latest LF state-vector estimate). LF data compression does introduce a small loss of global information, equivalent to neglecting a vector measurement of common process noise dimension at each interior step. This information loss is negligible when $Q \ll P$ over the fusion interval, as is the usual situation.

Last, the federated filter method has been described here in conventional covariance (P) and information (P^{-1}) terms. However, as shown in [18], the federated filter can be implemented in square root (factored U-D) form to maximize computational efficiency, numerical stability, and effective precision. One can choose covariance square root form, information square root form [7,8], or even a mix to best suit any particular application.

5. FEDERATED CONFIGURATION OPTIONS

The new federated filtering technique can be implemented in a variety of design configurations. Each design derives from different performance criteria that suggest a different information-sharing strategy, or "mode". This section describes four federated filter configurations. All four are suitable for "ideal" applications, in which the local and master filters can be freely designed to support whatever federated filter operations are required. Two of them are also suitable for "constrained" applications, in which the master filter may be freely designed, but the local filters are assumed to have been developed elsewhere for stand-alone filter operations. (We assume that pre-existing LFs can accommodate few if any software modifications.)

To make the discussion more concrete, we will consider as an example a hypothetical multisensor navigation system containing sensors typical of an advanced tactical aircraft. These sensors and their data output rates (Hz) or intervals (sec) are:

o   *Strapdown INS, baro-aided (50 Hz);*

o   *GPS receiver (1 - 2 sec);*

o   *Terrain-aided navigator (TAN) (1/4 - 1/2 sec);*

o   *Synthetic aperture radar (SAR) (100 - 300 sec).*

The strapdown INS is the common reference system, normally with a baro-stabilized vertical channel; it puts out indicated position, velocity, body attitude, and angular velocity. The GPS receiver typically puts out four pseudo-range and four pseudorange-rate (or delta-range) measurements per cycle. The TAN radar altimeter puts out one

height measurement per cycle. The SAR typically puts out two landmark imaging or "map" measurements (range and range-rate) and a set of 6-8 precision velocity measurements (multi-directional range-rate sets) per cycle. This example system serves to illustrate how various federated design options can be applied to advanced, integrated navigation systems.

Figures 3 to 6 illustrate four federated filter modes embodying different information-sharing strategies for this system. In each case, the federated filter structure consists of the MF and three LFs. Each LF processes measurements from one external sensor (GPS, TAN or SAR), while all of the filters use common INS data. Each LF provides a navigation solution of varying accuracy, depending on the inherent capabilities of its assigned sensor, and on environmental factors such as visibility, relative geometry, aircraft dynamics, and electronic interference. The MF performs generic fusion operations, and generates the best total solution from the three LF solutions.
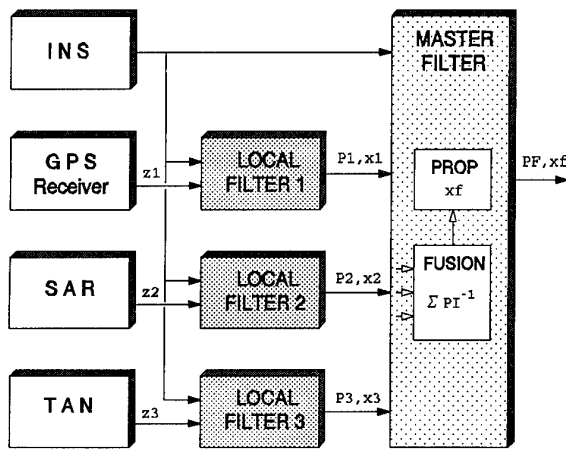
Figure 3: Federated No-Reset (NR) Mode

The primary difference among these four modes relates to information storage. In the first two modes (Figures 3 and 4), the LFs collectively store the long-term system information, while the MF acts as a short-term information combiner. Conversely, in the second two modes (Figures 5 and 6), the MF stores the long-term system information, while the LFs act as short-term data collectors, or data compression filters. Other important differences exist among these modes as well.

Figure 3 shows the federated *No-Reset (NR)* mode. Here, the LFs collectively maintain the system long-term memory, while the MF provides short-term propagation of the fused solution after combining (fusing) the LF outputs. There is no information feedback from the MF to the LFs; each LF retains its own, unique portion of the total system information. The NR mode comprises the least optimal but most fault-tolerant federated filter mode. This mode permits the LFs to operate independently, as standalone filters, with estimation accuracies essentially at their normal levels (except for the minor effects of process noise multipliers). The LFs all send solutions to the MF for fusion at the same time. The MF propagates the fused solution to intermediate time points, but does not use it in the next fusion update, since the next set of LF solutions contain all the accumulated system information.

Figure 4 shows the federated filter *Fusion-Reset (FR)* mode. Here, the LFs again collectively maintain the system long-term memory, each retaining a share of the total fused information while the MF retains none. The
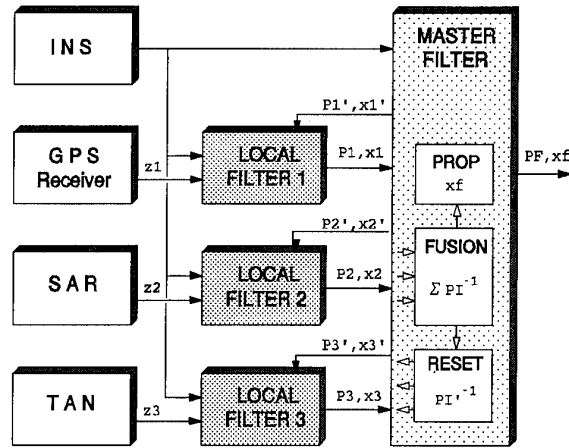
Figure 4: Federated Fusion-Reset (FR) Mode

MF provides temporary short-term memory, in propagating the full fused solution between fusion updates. The FR mode involves feedback of fusion-reset solutions from the MF to the LFs. Hence, the less accurate LFs operate at higher levels of accuracy than they can alone, since they gain information from the more accurate sensors via the MF resets. Conversely, the more accurate LFs operate at somewhat lower levels of accuracy, since they periodically give up more information than they receive back from the MF. The FR mode requires all the LFs to send solutions to the MF for fusion at the same time. Each LF then waits for a reset solution from the MF before proceeding.
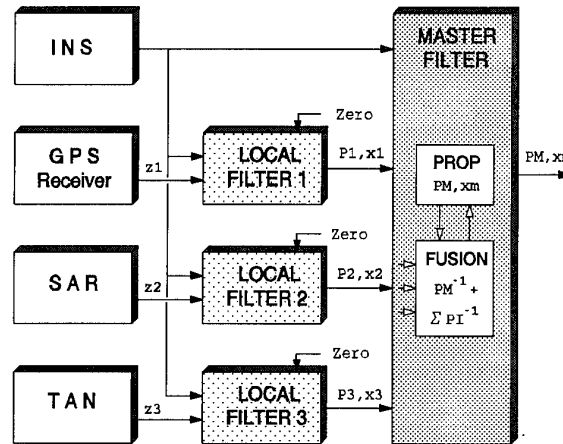
Figure 5: Federated Zero-Reset (ZR) Mode

Figure 5 shows the *Zero-Reset (ZR)* mode. Here, the MF retains all of the fused information ($\beta_m = 1$) or system long-term memory. The three LFs retain none of the fused information ($\beta_i = 0$), but act as data compression filters with short-term memory only. Each LF provides the MF with the new information it has gained since the last update. The MF adds each new LF solution input to its total information, in much the same way a standard Kalman filter adds new sensor measurement information. After the update, the LF resets itself to zero information. An attractive feature of the ZR mode is that the MF can process LF solutions at different times, according to how rapidly each LF gains information.

Figure 6 shows the federated filter *Rescale (RS)* mode. This mode can be considered a variation of the Zero-Reset
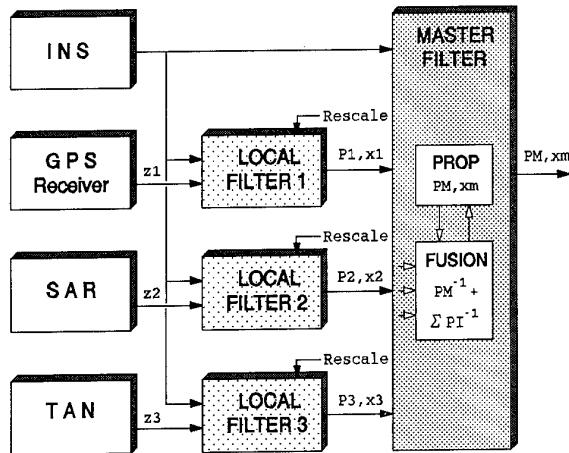
Figure 6: Federated Rescale (RS) Mode

mode. Here, each LF passes some but not all of its information to the MF at each fusion update (the fraction αm_i), and retains the remainder (the complementary faction $\alpha i_i = 1 - \alpha m_i$). With $\alpha m_i = 0.5$, the MF accumulates and retains over half of the total system information, while the LFs collectively retain the remainder. (The MF contains 1/2 of the LF's current-cycle information, 3/4 of the previous-cycle information, 7/8 of the next-previous, etc.) The LF resets are relatively simple to mechanize: the LF rescales its common covariance partition by a factor of $\gamma i_i = 1/\alpha i_i = 2$, and retains its state vector as-is.

A useful way to characterize these different information-sharing strategies is as follows. In the NR and FR modes, each LF tells the MF "what I know now," and the MF combines those results to obtain "what we all know now." Conversely, in the ZR Mode, each LF tells the MF "what I have learned since my last report," and the MF adds that new information to its total of "what we all know now." The RS mode is similar to the ZR mode, except that the LFs pass along what they have learned more gradually.

With regard to constrained filter applications, the NR mode is best suited for use with existing LFs, since it requires no additional software functions beyond standalone operations. (The common process-noise multipliers $1/\beta_i$ are often unnecessary, given conservative process noise models.) The RS mode may also be usable with existing LFs, if they can be modified to accommodate periodic rescaling of the covariance matrix. The ZR mode is not directly usable with existing LFs, but may be approximated by *restarting* each LF (at a large, initial covariance value) whenever a zero-reset is required. Last, the FR mode is not usable with existing LFs, since resetting the LF state and covariance to MF-supplied values is generally infeasible.

One final point should be noted here. These four federated filter modes involve exactly the same mechanization equations for all operations except the MF and LF resets (see Appendix A). They can all be implemented in a single software package, with a code branch for each reset option. Even a centralized filter can be implemented with the federated filter software [19], by assigning all the sensors to a single LF or MF and turning off the other filters.

6. REAL-TIME IMPLEMENTATION FEATURES

Four alternate federated filter designs were presented in the previous section, for the example integrated navigation system. This section examines those alternate designs more closely from the viewpoint of real-time implementation features, including processor and sensor fault tolerance.

With regard to *data processing rates*, the four federated filter designs of Section 5 all have the capability to adjust the MF fusion rate relative to the LF measurement update rates. When maximum accuracy is required, the MF can perform a fusion update after every measurement cycle of the most accurate LF (the GPS filter, for our example system).

When maximum MF accuracy is not required, and a reduction in computation burden is desirable, the MF fusion rate can be reduced to well below the LF measurement rates. Thus, the relatively large fusion computation burden can be reduced to, say, one-fifth or one-tenth of the fully optimal burden. For example, if the TAN LF performs measurement updates twice a second, and GPS does so once a second, the MF might perform fusion updates only once every ten seconds. In this way, the TAN LF compresses 20 measurements (20 x 1) into a single state vector, and the GPS LF compresses 80 measurements (10 x 8) into a single state vector, both used by the MF to obtain its fused solution.

The Zero-Reset (ZR) and Rescale (RS) modes are most flexible in this regard, since they permit fusion/reset processing to occur at different times for different LFs. For example, in the ZR mode, an LF can "dump" its information to the MF and reset itself to zero information at any time, regardless of when the other LFs do so. In contrast, the No-Reset (NR) and Fusion-Reset (FR) modes require all LF solutions to be propagated to the same time point (a fairly simple control process), and then to be passed to the MF where they are all combined to yield the total solution. Since the total system information is distributed across the LFs, all of the LF solutions must be collected at the same time, to obtain the best total solution.

Another real-time software issue relates to *resets* from the MF to the LFs. The NR mode is the simplest in this regard, for the obvious reason: there are no resets. The ZR and RS modes are relatively simple, since each LF can reset itself, after sending its solution to the MF, without waiting for a response from the MF. The FR mode is the most complex, since each LF has to wait for a reset solution from the MF, before it can continue its normal operations.

Next, consider *processor fault tolerance* -- the ability to detect and recover from filter dropouts caused by processor or bus failures. All of the federated filter modes provide enhanced processor fault tolerance in a way not possible with a single, centralized filter (provided the federated filter components are running on separate processors.) Each component sends out periodic status messages to the other components, and monitors their status messages in return. For example, if an LF disappears, due to a processor or bus failure, the federated controller (FC) detects its absence almost immediately, via the missing LF status message. The FC then modifies the information shares for the remaining filters, and tells the MF to use only those LFs in the fusion process. Thus, the federation gracefully reconfigures to the remaining set of available LFs. If the LF later reappears, the FC reconfigures the federation again, to re-include the LF.

Finally, consider *sensor fault tolerance* -- the ability to detect, isolate, and recover from sensor failures. Here, the federated filter designs support sensor fault detection, isolation and recovery (FDIR) at both the local and master filter levels. For purposes of comparison, it is helpful to begin by considering the sensor FDIR capabilities of a centralized filter (CF). The CF maintains a single, globally optimal solution, and incorporates measurements from all local sensors. It has good capability to detect and isolate some local sensor failures via measurement residual tests,

because each measurement is compared against all of the accumulated prior information.

One disadvantage of the CF is that, like any optimal filter, it attempts to make its data inputs agree in a weighted least-squares sense, thereby suppressing their differences. Hence, an undetected failure in one sensor will be distributed into all of the navigation state and sensor bias estimates, so that they all tend toward agreement. Thus, while measurement residual tests can readily detect sudden "hard" failures, they may completely miss gradual "soft" failures. A second disadvantage of the CF is that, if it does use bad data from any sensor, its full solution (accumulated information from the whole system) becomes irreversibly corrupted. When the faulty sensor is subsequently identified, the only safe means of recovery is to reinitialize the entire CF solution (state and covariance), and wait for it to reconverge to its normal accuracy level.

In comparison, the federated filter method provides several advantages with regard to sensor FDIR. Although specific advantages vary with the operating mode, there are a number of common features. First, the individual LFs accumulate sensor information over several cycles between MF fusion updates. Thus, a sensor soft failure has more time to reach detectable magnitude before it meets residual tests in the MF. Second, each LF maintains bias states for its local sensor alone. Other sensor bias states are maintained separately by other LFs, and are not drawn into agreement with the faulty sensor (except in the case of fusion resets). Third, once a sensor fault has been detected and isolated, uncorrupted LF solutions may still exist (depending upon the reset mode), so that recovery to a good MF solution can be quickly achieved.

More specifically, the federated No-Reset mode is highly fault-tolerant. Since there is no feedback of the MF fused solution to the LFs, there is no possibility of LF-to-LF cross-contamination. Each LF has good sensor FDI capability through measurement residual tests against all of the prior local information, except for certain types of soft failures. Most important, the MF has excellent LF:LF FDI capability, since each LF solution contains full accumulated information from its own sensor. Thus the effect of a sensor soft failure will be increasingly visible in one LF solution only, greatly increasing its probability of detection by the MF. The NR mode also provides a significant advantage for fault recovery. When the faulty LF is identified, a new MF solution can be generated immediately, from the remaining good LF solutions.

The other federated filter modes are less attractive with respect to sensor fault tolerance. The Fusion-Reset mode provides good measurement-residual screening at the LF level, since each LF contains a share of the full system information. However, the MF FDI capability is very weak, since the fusion-reset process causes all LF solutions to agree except for their most recent sensor measurements. Also, if faulty data from one LF is accepted by the MF, the fused solution is corrupted, and then the other LFs become corrupted via the MF resets. Thus, the MF and LF solutions all require reinitialization.

The Zero-Reset mode provides poor measurement-residual screening at the local level, since the LFs are reset to large covariance values after each fusion update. The MF FDIR capabilities are comparable to those of the CF. The MF has all of the prior system information for use in LF fusion-residual checks, but will still have difficulty detecting some types of soft failures. If faulty LF/sensor data is accepted by the MF, the total system solution held by the MF is corrupted, and must be reinitialized. The unfailed LFs do not hold any long-term information useful for quick MF recovery.

The Rescale mode provides fair measurement-residual screening in the LFs, since the LF residual tolerances increase by no more than $1/\alpha i_j^{\frac{1}{2}}$ (e.g., 1.414) in sigma units, after each fusion update. The MF FDI capabilities are similar to those of the ZR mode, and comparable to those of the CF; hence, certain types of soft failures will be difficult to detect. The recovery capability is slightly better than that of the ZR mode, since the unfailed LFs retain some accumulated information.

In summary, the No-Reset mode stands out from the others as having superior sensor FDIR capabilities. It also has the simplest real-time control requirements, since it requires no LF resets. Its only disadvantages are a slight degradation in estimation accuracy, and the need for all LF solutions to be combined by the MF at the same time (generally not a problem). The Zero-Reset and Rescale modes permit LF solutions to be processed at different times by the MF, but are not attractive from a sensor FDIR viewpoint. The Fusion-Reset mode provides the best estimation accuracy, but exhibits disadvantages with regard to sensor FDIR, computation and databus loads, and real-time software control requirements.

7. FEDERATED FILTER SIMULATION RESULTS

The federated filter and its several information-sharing modes have been implemented in computer software to support performance simulation testing. A general-purpose federated filter package has been built in FORTRAN-77, for non-realtime simulation testing on VAX and PC (IBM-compatible 80x86) computers [19]. In addition, a general-purpose federated filter has been built in Ada, with asynchronous multi-tasking capabilities, for real-time testing in an avionics simulation testbed [20].

The primary goals of the simulation tests reported in this section have been to:

a) demonstrate global optimality or near-optimality of federated filter configurations;

b) examine characteristics of master and local filter solutions relative to one another, and to that of the equivalent centralized filter;

c) demonstrate improved capability of federated filter to detect and isolate sensor faults, and to recover with minimal delay;

d) demonstrate reduced computation load per processor (improved throughput) of federated filter compared to centralized filter.

7.1 Performance Simulation Scenario

The federated filter has been tested in both non-real-time and near-real-time simulation environments. The non-real-time FORTRAN environment (the "DKF Simulator" [19]) contained a high-dynamic aircraft trajectory generator, a strapdown INS model, a baro altimeter model, a GPS satellite/receiver model, a SAR model (with landmark imaging and precision velocity updates), and a TAN model (radar altimeter plus synthetic terrain-map generator). The near-real-time FORTRAN/Ada environment (the "DKF-Ada Testbed" [20]) contained the same elements except for the SAR model. The sensor truth models were generally of medium-high fidelity, containing significantly more error sources than those in the corresponding filter models. Both environments permitted the filter navigation state estimates to be compared with the true states, to determine the estimation errors. Filter computation times were measured in the DKF Simulator.

In the simulation runs, the tactical aircraft trajectory consisted of a climb to altitude, a straight high-altitude seg-

272

Figure 7: Centralized Filter Estimation Errors

Figure 8: Federated No-Reset Estimation Errors

Figure 9: Federated Fusion-Reset Estimation Errors

The same levels of agreement -- very similar error and sigma plots -- occur for the velocity and attitude states as well. In each case, the errors and sigmas are relatively small while GPS data is available. (Plotting these segments on an expanded scale shows virtually identical results for the three filters.) The estimation errors grow during the GPS outages, with periodic reductions due to the SAR position and velocity measurements. In between SAR updates, the filter errors and sigmas increase somewhat, reflecting the reduced accuracy attainable from TAN radar/terrain measurements.

While we could discuss the relative performance attributes of the three navigation sensors at length here, the real point of interest is this: *the federated filter navigation performance is virtually identical to that of the globally optimal centralized filter.* Hence, these simulation results confirm the theoretical prediction that the federated filter is a near globally optimal filter formulation. The federated Fusion-Reset results are nearly identical to the CF results. The federated No-Reset results are slightly less accurate, as the theory (Appendix B) predicts.

7.3 Federated Filter Component Performance

The results in this subsection were generated with the real-time Ada version of the federated filter, running in the VAX-hosted DKF-Ada Testbed [20]. Each figure (10 to 14) shows the east position estimation error and sigma (feet) versus time (sec). The runs are for a relatively short 200-sec segment of the previous trajectory, including the final stage of descent, initial low-altitude penetration, and a few horizontal jinking maneuvers. Only the GPS and TAN sensors (and not the SAR) were employed for these runs; hence there were only two active LFs in the federated filter.

These real-time simulation results represent single-sample monte carlo runs, with the same random number sequences used for each case. The single-sample results are consistent with the more general results obtained from multi-sample monte carlo runs in the non-real-time DKF Simulator.

Figure 10 compares the globally optimal centralized filter (CF) results with those of the federated master filter operating in the Fusion-Reset (FR) mode, with equal (50/50) information-share fractions β_i for the two LFs. Figure 11 makes a similar comparison of the CF with the federated MF operating in the No-Reset (NR) mode, again with equal LF share fractions. In both cases, the federated MF results are nearly identical to those of the CF, and to each other. There are small differences in fine structure between the MFs and the CF, due to their different update rates. (The CF incorporates one GPS measurement set every 2 seconds, whereas the MFs combine two or three GPS measurement sets, via LF1, into one fusion update every 5 seconds.)

ment, a descent, and an extended low-altitude segment including several jinking maneuvers (zig-zags) and a large course change. GPS pseudorange and pseudorange-rate (or delta-range) measurements were processed by the filter at a rate of 1/5 Hz. Some runs included two significant GPS data outage periods, separated by a brief period of tracking. SAR position and velocity measurements were processed at a comparatively slow rate of once every 100 sec; the SAR was turned off near the end of the low-altitude segment, approaching the "target". TAN radar altimeter height measurements were processed during the entire low-altitude segment, at a relatively rapid rate of 1 Hz.

7.2 Federated Filter Optimality Results

The results in this subsection were obtained with the non-realtime DKF Simulator [19]. Figures 7, 8 and 9 demonstrate the global optimality of the federated filter. Each figure shows the east position estimation error (feet) for five separate monte carlo runs, plus the filter-computed one-sigma uncertainties, over a one-hour period from 1,000 to 4,600 sec. (The north and vertical errors and sigmas are qualitatively similar.) Figure 7 shows the performance of the globally optimal centralized filter, while Figures 8 and 9 show comparable results for two federated filters, operating in the No-Reset and Fusion-Reset modes, respectively. The federated filters performed fusion updates at the GPS measurement rate, once every 5 sec, for maximum MF accuracy. The same random number sequences were used to generate the truth data for each case.

The three sets of error and sigma plots are very similar to one another, although some minor differences are visible.

Figure 10: CF vs. Federated MF, Fusion-Reset 50/50

Figure 11: CF vs. Federated MF, No-Reset 50/50

Figure 12: Standalone LF1 (GPS) and LF2 (TAN)
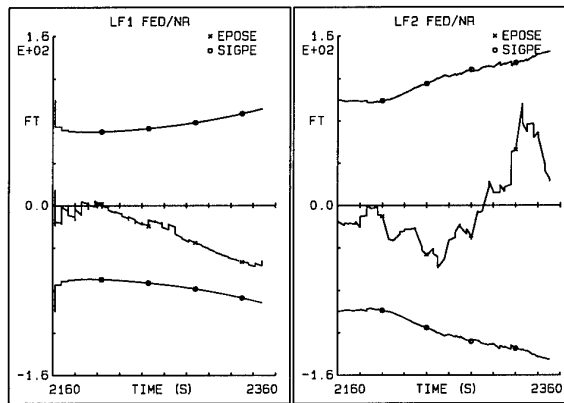
Figure 13: Federated LF1 and LF2 (NR 50/50)

Figure 14: Federated LF1 and LF2 (FR 50/50)

Note that the federated NR and FR modes again yield near-identical results (right-hand sides of Figures 10 and 11), just as they did in the earlier test case (Figures 8 and 9). We conclude that, even though the NR mode is theoretically less optimal than the FR mode, it is very nearly as accurate as the FR mode (and the CF) for typical navigation systems.

Figure 12 shows the two LFs (GPS, TAN) operating as *standalone* (locally optimal) filters. These standalone LF results are of interest primarily for comparison with federated LF results (later). However, comparing these results to those of the CF (Figure 10) also indicates the accuracy improvement in the CF obtained by using both sensors instead of just one.

Figure 13 shows similar results for LF1 and LF2 operating as components of the *No-Reset* federated filter with equal (50/50) information-share fractions. Here, the LF common process noise strengths are 1.414 (sqrt-2) times as large as those in Figure 12. However, virtually no difference in performance is evident in the position errors and sigmas of either LF. For the velocity errors (not shown), there is a very small increase in sigma growth rate between measurements, but the overall results are still virtually the same. Hence, we conclude that the federated process noise multipliers have very little effect on LF performance relative to the standalone case.

Likewise, Figure 14 shows results for LF1 and LF2 operating as components of the *Fusion-Reset* federated filter, again with 50/50 information shares. In contrast to the NR case, both LFs here show errors and sigmas significantly different from their standalone results. Except for the fine structure, the two sets of LF results are virtually identical to

each other. They are also virtually identical to the corresponding MF results (Figure 11), except that the LF sigmas are larger by a factor of 1.4, due to the $1/\beta_i = 2$ fusion-reset multipliers. We conclude that fusion resets dominate the performance of the component LFs in the FR mode, at least when the fusion interval is not much larger than the primary measurement interval.

7.4 Sensor Fault Detection Results

Federated and centralized filter performance results regarding sensor fault tolerance have been obtained in the non-real-time DKF Simulator environment. The results reported here are for a GPS satellite clock failure consisting of a frequency

274
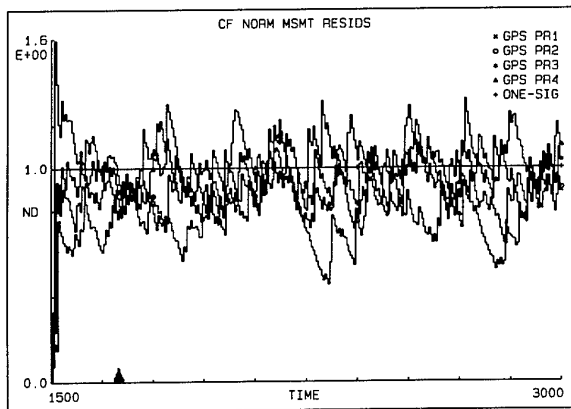
Figure 15: CF GPS PR Residuals after Failure

Figure 16: CF GPS PRR Residuals after Failure

Figure 17: MF Fusion Residuals after Failure
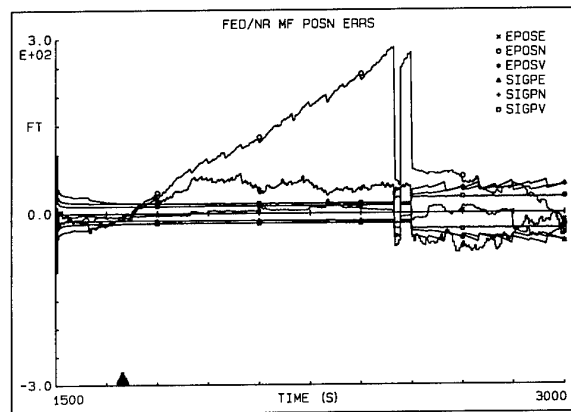
Figure 18: MF Position Errors after Failure

bias shift of 0.5 ft/sec, starting at time 1,700 sec (late in the high-altitude outbound cruise segment of the flight profile). The clock frequency failure introduced an 0.5 ft/sec bias into the subsequent pseudorange-rate measurements, and an 0.5 ft/sec ramp into the subsequent pseudorange measurements, from the faulty satellite. This scenario represents a "soft" (slowly-growing) failure in the most accurate navigation sensor -- generally the most difficult type of failure to detect.

Figures 15 and 16 show the centralized filter (CF) rms measurement residuals for the GPS pseudorange (PR) and pseudorange-rate (PRR) measurements, over the period from 1,500 to 3,000 sec; the failure occurs at the ▲ mark, 200 sec after the initial time shown. These rms residuals are normalized, with an expected value of unity, and are smoothed via a ten-point moving average, to make the trends more visible. The PR residuals show no effect of the satellite #3 failure, in spite of the ramp in PR #3. The smoothed PRR residuals for satellite #3 rise to about 3.5 sigma, while the unsmoothed PRR #3 residuals (not shown) rise to just over 5.0 sigma. The CF rejects the PRR #3 measurements while their residuals exceed the specified 3σ threshold. However, the sequence of PRR #3 residuals peaks, then falls back down to acceptable levels (due to filter covariance growth between measurements). Thus, *the CF loses its ability to detect the failure, and from that point on, uses the faulty data with no clear indication that it is bad.*

The CF measurement residuals from the less accurate SAR and TAN sensors likewise provide no clear indication of the GPS failure. (One of the SAR measurements does eventually show a ramp in its residual sequence; however, the CF then rejects those good SAR measurements.) After the

failure, the CF horizontal position estimates exhibit a steadily growing radial error that reaches 425 ft by time 3,000 sec (1,300 sec after the failure).

The enhanced FDI capability of the No-Reset federated filter is evident in Figure 17, which shows the MF fusion residuals. In each fusion cycle, the MF starts with the TAN LF solution, then adds the SAR LF solution, followed by the GPS LF solution. The SAR fusion residuals (SAR:TAN inconsistencies) are well behaved, whereas the GPS fusion residuals (GPS:SAR/TAN inconsistencies) ramp off and provide an unmistakable failure indication. (These are normalized rms fusion residuals, with an expected value of unity.)

Figure 18 shows the corresponding MF position errors. With a residual threshold set at 3σ, the MF rejected the GPS LF solution at about 2,500 sec (800 sec after the initial fault). At this time, the radial position error had grown to about 300 ft; this value is consistent with the SAR/TAN accuracy level of about 100 ft, and the 3σ threshold. (Note: the jumps in the MF errors near 2,500 sec were caused by noise in the GPS fusion residual as it crossed the 3σ threshold, such that the GPS LF solution was rejected, accepted, then rejected again.)

Note that, within the GPS LF of the No-Reset federated filter, the measurement residuals are qualitatively similar to those of the CF. The PR residuals give no indication of a failure. The PRR residuals give a temporary indication, then decay back to normal levels. This temporary indication is even weaker in the GPS LF than in the CF, since the LF covariance and computed residual variances are larger than those of the CF.
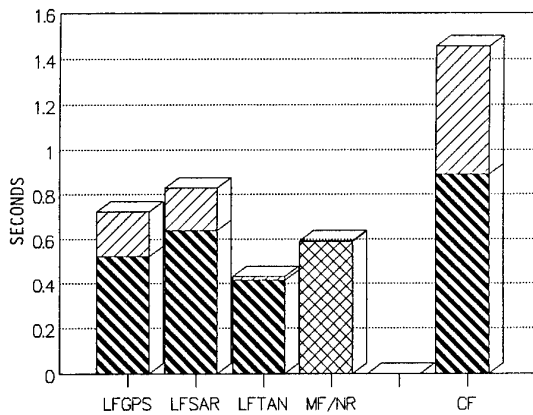
Figure 19: Peak Cycle Times, LFs and MF vs. CF
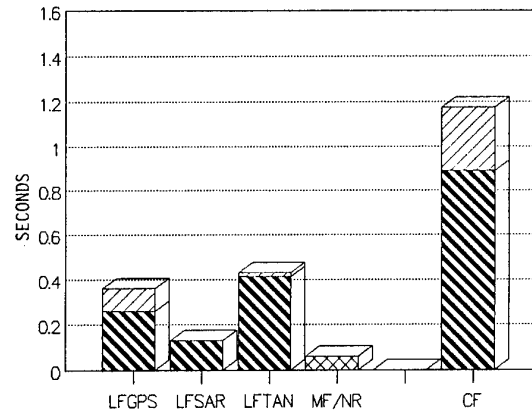
Figure 20: Average Cycle Times, LFs and MF vs. CF

7.5 Relative Computation Time Results

Computation times have been measured in the non-real-time DKF Simulator [19], for the federated filter No-Reset mode, and for the equivalent centralized filter. The various filter sizes and processing rates for these timing studies were:

Filter	States	Operation	Period
LF1	18	Propagation	2 sec
		8 GPS msmts	2 sec
LF2	17	Propagation	1 sec
		1 TAN msmt	1 sec
LF3	23	Propagation	5 sec
		7 SAR msmts	300 sec
MF	16	Propagation	10 sec
		Fusion (equiv to 32 msmts)	10 sec
CF	27	Propagation	1 sec
		1 TAN msmt	1 sec
		8 GPS msmts	2 sec
		7 SAR msmts	300 sec

In these cases, the filter INS model had 16 states (it had 13 in the previous tests), the GPS model had 2, the SAR model had 7, and the TAN model had 1. The MF implemented only the 16 common INS states, since only common information was shared. The MF propagated only its state vector, since the propagated MF covariance is not needed when the MF retains no information.

Figure 19 shows the *peak*-cycle execution times (sec) for each of the filters -- the three LFs and the MF from the federated filter, and the centralized filter (CF). Each total time is subdivided into the major functions: propagation, measurement processing, and fusion. Each of the federated filter components requires considerably less processing time for its peak cycle than the CF, ranging from approximately 29 to 56 percent of the CF peak cycle time. Thus, if implemented on separate, parallel processors, the federated filter would provide approximately a 2:1 advantage.

Figure 20 shows the *average*-cycle execution times (sec) for each of the filters. Here the federated filter components provide an even greater speed advantage, with average computation times ranging from 7 to 40 percent of the CF average time. Thus, the average per-processor computation loads of the federated filter are significantly less than that of the CF. Note that, even if the federated filter average times were added together (as if they all ran on the same processor), they would still sum to slightly less than the CF average time.

These timing results are meant to be illustrative. The speed advantage of the federated filter over a single centralized filter is greatest when the number of common INS states is relatively small, and the number of sensor bias states is relatively large. The illustrative case is not very favorable in this regard, even though the federated filter still exhibits a useful speed advantage. (The advantage would be greater, with fewer common INS states, and/or more sensor bias states.) These filter timing values were obtained on an 80386/20MHz processor with an 80387 math coprocessor. (With a different FORTRAN compiler on the same computer, we obtained computation times 50 percent less that those shown here, because of more efficient object code for the double-indexed matrix operations).

8. CONCLUSIONS

The new federated filter method presented in this paper forms a practical and straightforward basis for constructing integrated multisensor navigation filters with enhanced levels of fault tolerance. These filters exhibit a number of advantages over centralized Kalman filters, particularly for systems with embedded local filters and/or distributed components:

o theoretically correct partitioned estimation algorithms for cascaded systems with sensor-dedicated local filters (LFs);

o multilevel fault detection, isolation and recovery (FDIR) capability; highly fault tolerant in No-Reset mode with independent LF solutions;

o increased data throughput due both to parallel processing of LF operations, and to sensor data compression in LFs;

o potential reductions in system integration costs due to modular architecture with sensor-dedicated LFs and generic, sensor-independent MF.

Relative to previous decentralized filtering methods, the federated filter also provides one or more of these advantages:

o near globally optimal estimation accuracy of master filter fused (combined) solution;

o straightforward real-time implementation due to largely independent LF operations, particularly in No-Reset, Zero-Reset and Rescale modes;

o minimal data transfer requirements: filter states and covariances at fusion update times only.

The multisensor navigation system example illustrates the

variety of design configurations, or information-management strategies, permitted by the federated filter method. Of the several alternate configurations presented, each has attractive features for some but not all applications.

Numerical simulation results to date confirm the theoretical predictions regarding near global optimality of the federated filter method, reduced per-processor computation loads, and enhanced capability of the federated No-Reset mode to detect, isolate and recover from sensor failures. Additional simulation test results are reported in [18].

The practicality of implementing the federated filter method in real-time Ada software has also been demonstrated. The Ada federated filter has been tested in a near-real-time avionics simulation testbed, and again confirmed the theoretical performance predictions, including near globally optimal accuracy, reduced per-processor computation burden, and robustness in the presence of processor failures.

ACKNOWLEDGEMENTS

APPENDIX A: IMPLEMENTATION EQUATIONS

This appendix presents implementation equations for the federated filter. These equations are expressed in familiar covariance form, to facilitate understanding. However, the equations have been recast in factored U-D (unitary-diagonal) form [18,21] in software implementations where processing speed and numerical precision are important [19,20].

We will use subscript i to refer to LF #i, and subscript m for the MF. The set of LFs is denoted by $i = 1..n$, and the set of all filters (MF plus LFs), by $k = 1..n,m$. Each filter is partitioned into the common INS states and its own unique sensor bias states. These definitions apply to filter #k:

$$\underline{xk} = \begin{bmatrix} \underline{xk}_c \\ \underline{xk}_k \end{bmatrix} \quad \begin{array}{l} \textit{filter k common states} \\ \\ \textit{filter k bias-k states} \end{array} \quad (A-1)$$

$$PK = \begin{bmatrix} PK_{cc} & PK_{ck} \\ PK_{kc} & PK_{kk} \end{bmatrix} \quad (A-2)$$

where "bias-k" means the biases for sensor #k. The full or global state vector $\underline{xf}$ contains all of the filter common and sensor bias states. The filter #k state vector $\underline{xk}$ and the full state vector $\underline{xf}$ are related by a simple mapping matrix LK:

$$LK \doteq (\partial \underline{xf}/\partial \underline{xk}) = \begin{bmatrix} I_{cc} & 0_{ck} \\ : & : \\ 0_{kc} & I_{kk} \\ : & : \\ 0_{nc} & 0_{nk} \end{bmatrix} \quad (A-3)$$

In this appendix, we will present equations for the "small" form of the master filter, in which the MF contains only the common states and any MF-unique sensor biases, as indicated by Eq. (A-1) with $k = m$. (There is also a "large" or global form of the MF, in which the MF contains all of the full filter states $\underline{xf}$, i.e., the common states plus bias states

for all the sensors. When the MF retains information, the global form may provide somewhat better accuracy; however, it is much larger, slower, and more complex than the small form.)

The various information-share fractions and multipliers used by the federated filter satisfy these constraints:

$$\gamma_k = 1/\beta_k; \qquad \beta_m + \sum_{i=1}^{n} \beta_i = 1 \qquad (A-4)$$

$$\lambda k_i = 1/\alpha k_i; \qquad \alpha m_i + \alpha i_i = 1 \qquad (A-5)$$

The β_k fractions relate to initial information shares, process-noise shares, and fusion-reset shares (they can be in different ratios for each operation). The αk_i fractions relate only to the fusion-input and reset operations of the Rescale mode.

The federated filter algorithms will be presented in five steps: 1) the LF/MF initialization equations; 2) the LF/MF time propagation equations; 3) the LF/MF measurement update equations; 4) the MF fusion equations; and 5) the MF and LF reset equations for various reset modes. For each step, we will first indicate the basic operation (marked by ■) in unpartitioned covariance form, then provide the partitioned implementation equations and/or the relevant partitioned matrix factors. Note that, in Step 2, the prime refers to values at the beginning of the propagation; in Step 3, superscript + refers to post-measurement values; in Step 4, superscript + refers to values after each LF #i fusion; and in Step 5, the quote symbol " refers to values after the reset.

1) LF/MF Initialization:

$$\blacksquare \quad PF_o^{-1} = PO^{-1} => \sum_{k=1}^{n,m} LK \, PK_o^{-1} \, LK^T:$$

$$PK_o = \begin{bmatrix} PO_{cc}\gamma_k & PO_{kc}\gamma_k \\ PO_{kc}\gamma_k & PO_{kk} + PO_{kc}PO_{cc}^{-1}PO_{kc}^T(\gamma_k-1) \end{bmatrix} \quad (A-6)$$

$$\underline{\hat{xk}}_o = \begin{bmatrix} \underline{\hat{xo}}_c \\ \underline{\hat{xo}}_k \end{bmatrix} \quad (A-7)$$

2) LF/MF Propagation from time t' to t:

$$\blacksquare \quad PK = \Phi K \, PK' \, \Phi K^T + GK \, QK \, GK^T:$$

$$\underline{\hat{xk}} = \Phi K \, \underline{\hat{xk}}' \quad (A-8)$$

$$\Phi K = \begin{bmatrix} \Phi_{cc} & 0 \\ 0 & \Phi_{kk} \end{bmatrix} \quad (A-9)$$

$$\quad (A-10)$$

$$GK = \begin{bmatrix} G_{cc} & 0 \\ 0 & G_{kk} \end{bmatrix}; \quad QK = \begin{bmatrix} Q_{cc}\gamma_k & 0 \\ 0 & Q_{kk} \end{bmatrix}$$

3) LF/MF Measurement Update:

$$\blacksquare \quad (PK^+)^{-1} = PK^{-1} + HK \, RK^{-1} \, HK^T:$$

$$BK = PK \, HK; \quad AK = HK^T PK \, HK + RK \quad (A\text{-}11a)$$

$$PK^+ = PK - BK \, AK^{-1} BK^T \quad (A\text{-}11b)$$

$$\hat{\underline{xk}}^+ = \hat{\underline{xk}} + BK \, AK^{-1} (\tilde{\underline{zk}} - HK^T \hat{\underline{xk}}) \quad (A\text{-}12)$$

$$HK^T = [HK_c^T \quad HK_k^T] \quad (A\text{-}13)$$

4) MF Fusion Update (Common States):

$$\blacksquare \quad PF^{-1} = \sum_{k=1}^{n,m} LK \, PK^{-1} LK^T :$$

$$Start: \quad PF_{cc} = PM_{cc} \quad (A\text{-}14)$$

$$\hat{\underline{xf}}_c = \hat{\underline{xm}}_c \quad (A\text{-}15)$$

Do for LFs i = 1..n:

$$AI_{cc} = PF_{cc} + PI_{cc} \lambda m_i \quad (A\text{-}16a)$$

$$PF_{cc}^+ = PF_{cc} - PF_{cc} \, AI_{cc}^{-1} PF_{cc} \quad (A\text{-}16b)$$

$$\hat{\underline{xf}}_c^+ = \hat{\underline{xf}}_c + PF_{cc} \, AI_{cc}^{-1} (\hat{\underline{xi}}_c - \hat{\underline{xf}}_c) \quad (A\text{-}17)$$

5) MF/LF Fusion-Reset ("):

$$\blacksquare \quad PF^{-1} \Rightarrow \sum_{k=1}^{n,m} LK \, PK''^{-1} LK^T :$$

a) No-Reset Mode:

$$\left. \begin{array}{l} PM''_{cc} = PF_{cc} \\ \hat{\underline{xm}}''_c = \hat{\underline{xf}}_c \end{array} \right\} \begin{array}{l} \textit{Use between} \\ \textit{fusion times,} \\ \textit{then discard} \end{array} \quad \begin{array}{l} (A\text{-}18) \\ \\ (A\text{-}19) \end{array}$$

$$PI'' = PI \quad (no\ change) \quad (A\text{-}20)$$

$$\hat{\underline{xi}}'' = \hat{\underline{xi}} \quad (no\ change) \quad (A\text{-}21)$$

b) Fusion-Reset Mode:

$$PK'' = \begin{bmatrix} PF_{cc} \gamma_k & \gamma_k PF_{cc} PK_{cc}^{-1} PK_{kc}^T \\ PK_{kc} PK_{cc}^{-1} PF_{cc} \gamma_k & PK''_{kk} \end{bmatrix} \quad (A\text{-}22)$$

$$PK''_{kk} = PK_{kk} - PK_{kc} PK_{cc}^{-1} (PK_{cc} - PK''_{cc}) PK_{cc}^{-1} PK_{kc}^T$$

$$\hat{\underline{xk}}'' = \begin{bmatrix} \hat{\underline{xf}}_c \\ \hat{\underline{xk}}_k + PK_{kc} PK_{cc}^{-1} (\hat{\underline{xf}}_c - \hat{\underline{xk}}_c) \end{bmatrix} \quad (A\text{-}23)$$

c) Rescale Mode:

$$PM'' = \begin{bmatrix} PF_{cc} & PF_{cc} PM_{cc}^{-1} PM_{mc}^T \\ PM_{mc} PM_{cc}^{-1} PF_{cc} & PM''_{mm} \end{bmatrix} \quad (A\text{-}24)$$

$$PM''_{mm} = PM_{mm} - PM_{mc} PM_{cc}^{-1} (PM_{cc} - PF_{cc}) PM_{cc}^{-1} PM_{mc}^T$$

$$\hat{\underline{xm}}'' = \begin{bmatrix} \hat{\underline{xf}}_c \\ \hat{\underline{xm}}_m + PM_{mc} PM_{cc}^{-1} (\hat{\underline{xf}}_c - \hat{\underline{xm}}_c) \end{bmatrix} \quad (A\text{-}25)$$

$$PI'' = \begin{bmatrix} PI_{cc} \lambda i_i & PI_{ic}^T \lambda i_i \\ PI_{ic} \lambda i_i & PI''_{ii} \end{bmatrix} \quad (A\text{-}26)$$

$$PI''_{ii} = PI_{ii} + PI_{ic} PI_{cc}^{-1} PI_{ic}^T (\lambda i_i - 1)$$

$$\hat{\underline{xi}}'' = \hat{\underline{xi}} \quad (no\ change) \quad (A\text{-}27)$$

d) Zero-Reset Mode:

Use Rescale mode (c) with

$$\lambda i_i = 1/\epsilon_i \gg 1 \quad (A\text{-}28)$$

$$\lambda m_i = 1/(1-\epsilon_i) \approx 1 + \epsilon_i \quad (A\text{-}29)$$

Note that, in the Rescale and Zero-Reset modes, the MF can incorporate different LF solutions at different times, via Eqs. (A-14) to (A-17), with resets per Eqs.(A-24) to (A-27). Also, some care must be taken in resetting the LF covariance partitions to large values representing "zero" common-state information. For example, attitude error covariances may need to be limited in size, because of small-angle linearity assumptions, and weak observability of those states. (The reset multiplier λi_i can be limited in each LF, to prevent reset variances from exceeding specified maximums.)

The federated filter implementation equations presented in this appendix have been cast in equivalent U-D factored form, coded in both non-realtime FORTRAN software (all modes) and real-time Ada software (NR and FR modes), and validated by extensive simulation testing [18-20].

B. THEORETICAL FOUNDATION

This appendix presents a derivation of the federated filtering method outlined in Section 4. We will formulate the optimal estimation problem as a sequential, weighted least squares problem. The full or global least-squares problem will be partitioned into n local least-squares problems. Then, the solutions of the n local problems will be combined, via another least-squares operation, to yield the total fused solution. Under certain conditions, this fused solution is globally optimal -- identical to the full, unpartitioned solution. Under other, more practical conditions, it is near-optimal, using most but not all of the available information.

Consider the sequential system dynamics and discrete measurement processes described by Eqs. (1) to (8) of Section 2. First, define S, W and V as square roots of the state error, process noise, and measurement noise covariance matrices:

$$S = P^{\frac{1}{2}}; \quad W = Q^{\frac{1}{2}}; \quad V = R^{\frac{1}{2}} \quad (B\text{-}1)$$

Now, the globally optimal state estimate for this system at time t_k is the value of $\underline{x}_k$ that minimizes a weighted least squares cost function ψ_{k+} of all the random (unknown) error terms from step 0 through step k+. The initial estimate $\hat{\underline{x}}_o$ is known, with error covariance square root S_o. The unknown variables are the true initial state $\underline{x}_o$, the sequence of process noises $\underline{u}_j$, and the sequence of sensor #i measurement noises $\underline{vi}_j = \underline{zi}_j - HI_j^T \underline{x}_j$, where index j refers to

successive times t_j, adn index i, to the n independent sensors. Index $k+$ refers to the post-measurement value. The cost function ψ_{k+} can be expressed as follows [7]:

$$\psi_{k+} = \left\| S_o^{-1} (\hat{\underline{x}}_o - \underline{x}_o) \right\|^2 + \sum_{j=1}^{k} \left[\left\| W_j^{-1} \underline{u}_j \right\|^2 \right.$$

$$(B-2)$$

$$\left. + \sum_{i=1}^{n} \left\| VI_j^{-1} \underline{vi}_j \right\|^2 \right]$$

The sequence of true state vectors $\underline{x}_j$ is also a set of unknowns. These unknowns are linearly related to $\underline{x}_o$ and the process noises $\underline{u}_j$ by the state propagation equations:

$$\underline{x}_j = \Phi_j \underline{x}_{j-1} + G_j \underline{u}_j \qquad (j = 1..k) \ (B-3)$$

At step j, the $i = 1..n$ sensors provide measurements $\underline{\tilde{zi}}_j$ of the state $\underline{x}_j$, containing the independent measurement errors $\underline{vi}_j$ making up the last set of terms in ψ_{k+}:

$$\underline{\tilde{zi}}_j = HI_j^T \underline{x}_j + \underline{vi}_j \qquad (j = 1..k) \ (B-4)$$

The weighted least squares solution is the set of values for the unknowns $\underline{x}_o$, $\underline{u}_j$ (or equivalently $\underline{x}_j$), and $\underline{vi}_j$ that minimizes the value of the cost function ψ_{k+}, given the initial state estimate $\hat{\underline{x}}_o$ and the sequence of measurements $\underline{\tilde{zi}}_j$. The process of determining those optimal values (or estimates) can be performed sequentially. To do so, we rewrite the full cost function ψ_{k+} in terms of its accumulated residuals $\Delta\psi_{k-1}$ at the previous step, and the new terms introduced at step k:

$$\psi_{k+} = \Delta\psi_{k-1} + \left\| S_{k-1}^{-1} (\hat{\underline{x}}_{k-1} - \underline{x}_{k-1}) \right\|^2$$

$$(B-5)$$

$$+ \left\| W_k^{-1} \underline{u}_k \right\|^2 + \sum_{i=1}^{n} \left\| VI_k^{-1} \underline{vi}_k \right\|^2$$

The value of $\underline{x}_{k-1}$ that minimizes this cost function at step $k-1$ is $\hat{\underline{x}}_{k-1}$, the computed estimate after step $k-1$. S_{k-1} is the covariance square root of the error in that value. Both $\hat{\underline{x}}_{k-1}$ and S_{k-1} are assumed to have been propagated from their initial values by recursive step-wise operations, as will be outlined below for the step from $k-1$ to $k+$.

Now, we can determine the best estimate of $\underline{x}_k$ after the time propagation and before incorporating the measurement. We use Eq. (B-3) to eliminate $\underline{x}_{k-1}$ in the first term of (B-5), and then rearrange the elements of the first and second terms to produce a new first term in $\underline{x}_k$ alone:

$$\psi = \Delta\psi' + \left\| S^{-1} (\hat{\underline{x}} - \underline{x}) \right\|^2$$

$$(B-6)$$

$$+ \left\| B^{-1} [\underline{u} + E S^{-1} (\hat{\underline{x}} - \underline{x})] \right\|^2$$

In Eq. (B-6) we have dropped the subscript k for simplicity, using a prime (') to denote prior values at $k-1$. New terms on the right side of Eq. (B-6) have these values:

$$\hat{\underline{x}} = \Phi \hat{\underline{x}}'; \qquad \tilde{S} = \Phi S' \qquad (B-7)$$

$$S = [\tilde{S} \tilde{S}^T + F F^T]^{1/2}; \qquad F \doteq G W \qquad (B-8)$$

$$B = W [I - F^T (F F^T + \tilde{S} \tilde{S}^T)^{-1} F]^{1/2} \qquad (B-9)$$

$$E = W W^T G^T S^{-T} \qquad (B-10)$$

Thus, the optimal pre-measurement value of $\underline{x}$ is $\hat{\underline{x}}$, which zeroes the first new term in Eq. (B-6). The second term can be zeroed by choosing $\underline{u} = -E S^{-1} (\hat{\underline{x}} - \underline{x})$. The new S is the covariance square root of the propagated estimate $\hat{\underline{x}}$.

Next, the measurements at step k can be optimally incorporated. We first use Eq. (B-4) to rewrite the measurement error $\underline{vi}$ in terms of the computed measurement residual $\Delta\underline{\tilde{zi}}$ and the state error:

$$\underline{vi} = \underline{\tilde{zi}} - HI^T \underline{x} = \Delta\underline{\tilde{zi}} + HI^T (\hat{\underline{x}} - \underline{x}) \quad (B-11)$$

$$\Delta\underline{\tilde{zi}} \doteq \underline{\tilde{zi}} - HI^T \hat{\underline{x}} \qquad (B-12)$$

We then rearrange the elements of the squared state and measurement terms to obtain a new term in $\underline{x}$ alone. The final cost function at step $k+$ is this:

$$\psi_+ = \Delta\psi' + \left\| S_+^{-1} (\hat{\underline{x}}_+ - \underline{x}) \right\|^2$$

$$+ \left\| B^{-1} [\underline{u} + E S^{-1} (\hat{\underline{x}} - \underline{x})] \right\|^2 \quad (B-13)$$

$$+ \sum_{i=1}^{n} \left\| CI^{-1} \Delta\underline{\tilde{zi}} \right\|^2$$

where, for $i = 1..n$, the new terms in Eq. (B-13) can be computed in covariance square root form as follows:

$$FI = S^T HI \qquad (B-14)$$

$$S_+ = S[I - FI(FI^T FI + RI)^{-1} FI^T]^{1/2} \qquad (B-15)$$

$$CI = (FI^T FI + RI)^{1/2}; \qquad JI = S FI CI^{-T} \ (B-16)$$

$$\Delta\underline{\tilde{zi}} = \underline{\tilde{zi}} - HI^T \hat{\underline{x}} \qquad (B-17)$$

$$\hat{\underline{x}}_+ = \hat{\underline{x}} + JI CI^{-1} \Delta\underline{\tilde{zi}} \qquad (B-18)$$

Thus, the new optimal estimate of $\underline{x}$ is $\hat{\underline{x}}_+$, which zeroes the first new term in (B-13). S_+ is the new covariance square root. The process noise $\underline{u}$ can still be chosen to zero the second term, given the optimal value $\underline{x} = \hat{\underline{x}}_+$. The third term is an irreducible least-squares residual; it is composed entirely of measured and computed values, and contains no unknowns. In summary, Eqs. (B-7) to (B-10) and (B-14) to (B-18) represent the recursive solution of the globally optimal estimator.

With this foundation, we can proceed to develop the partitioned optimal estimator. First, we divide the original cost function ψ_+ in Eq. (B-5) into n partitions, each containing the measurement term from one unique sensor:

$$\psi_+ = \psi 1_+ + .. + \psi n_+ = \sum_{i=1}^{n} \psi i_+ \qquad (B-19)$$

$$\psi i_+ = \Delta\psi i' + \left\| \hat{SI}'^{-1} (\hat{\underline{xi}}' - \underline{x}') \right\|^2 \qquad (B-20)$$

$$+ \left\| WI^{-1} \underline{u} \right\|^2 + \left\| VI^{-1} \underline{vi} \right\|^2$$

where the SI, GI, WI, and $\hat{\underline{xi}}$ terms represent the ith parti-

tion. The original terms in Eq. (B-5) are exactly equal to the sums of the corresponding terms in Eq. (B-20) if these "conservation of information" conditions apply:

$$S'^{-T}S'^{-1} = \sum_{i=1}^{n} SI'^{-T}SI'^{-1} \qquad (B-21)$$

$$W^{-T}W^{-1} = \sum_{i=1}^{n} WI^{-T} WI^{-1} \qquad (B-22)$$

The underlying conservation of information principle can be stated as follows: *the total information in the partitioned estimator must sum to the same value as the total information in the unpartitioned global estimator.* This equality must hold true for a) the state information $P^{-1} = S^{-T}S^{-1}$, b) the process noise information $Q^{-1} = W^{-T}W^{-1}$, and c) the measurement information $R^{-1} = V^{-T}V^{-1}$. The measurement information is naturally divided according to the n independent sensors. The state and process noise information can be divided as follows, where Λ_i and Γ_i are diagonal share-matrices whose squares sum to the identity matrix:

$$SI^{-1} = \Lambda_i \ S^{-1}; \qquad \Sigma \ \Lambda_i^2 = I \qquad (B-23)$$

$$WI^{-1} = \Gamma_i \ W^{-1}; \qquad \Sigma \ \Gamma_i^2 = I \qquad (B-24)$$

Now, each partition of the global estimation problem represents a local estimator, or local filter. Each cost function ψi is minimized by the corresponding local filter operations. After local filter #i is propagated from time t' to t, and its unique sensor #i measurements are incorporated, its cost function ψi has the following form, similar to Eq. (B-13):

$$\psi i_+ = \Delta\psi i' + \left\| SI_+^{-1}(\hat{xi}_+ - \underline{x}) \right\|^2$$

$$+ \left\| BI^{-1}[\underline{u} + EI \ SI^{-1}(\hat{xi} - \underline{x})] \right\|^2 \quad (B-25)$$

$$+ \left\| CI^{-1} \ \Delta\tilde{\underline{zi}} \right\|^2$$

where the terms in (B-25) are defined analogously to those in the total cost function (B-13). Now, the globally optimal estimate of $\underline{x}$ is the value that minimizes the sum of the ψi partitions, i.e, the total cost. Each partition contains $\underline{x}$ in its first term, and both $\underline{u}$ and $\underline{x}$ in its second term. We can minimize the sum of the first terms in Eq. (B-25) as follows:

$$\sum_{i=1}^{n} \left\| SI^{-1}(\underline{x} - \underline{xi}) \right\|^2$$

$$\qquad\qquad (B-26)$$

$$= \left\| SF^{-1}(\underline{x} - \hat{\underline{xf}}) \right\|^2 + \left\| \hat{\Delta\underline{y}} \right\|^2$$

where $\hat{\Delta\underline{y}}$ is an irreducible fusion residual (a function only of known terms in $\hat{\underline{xi}}$). The full, combined state estimate $\hat{\underline{xf}}$ and its covariance square root SF can be computed either by orthogonal transformations on S^{-1} [7,21], or by the equivalent reduction operations on S similar to (B-15). The results can be expressed as the following information sum:

$$SF^{-T}SF^{-1} = \sum_{i=1}^{n} SI^{-T}SI^{-1} \qquad (B-27)$$

$$SF^{-T}SF^{-1}\hat{\underline{xf}} = \sum_{i=1}^{n} SI^{-T}SI^{-1}\hat{\underline{xi}} \qquad (B-28)$$

These operations represent the "fusion" process in the feder-

ated filter. Now, the question arises as to whether $\hat{\underline{xf}}$ and SF are equal to the globally optimal solution given by Eqs. (B-15) to (B-18). The answer is "yes," under certain limited conditions, and "almost," under other more practical conditions. The answer revolves around the second, process-noise term in Eq. (B-25). The globally optimal solution requires that the sum of these partition terms also be minimized. In principle, this sum could be rearranged such that there were only two squared terms in the unknowns $\underline{u}$ and $\underline{x}$:

$$JI \doteq BI^{-1}EI \ SI^{-1}; \qquad \hat{\underline{wi}} = JI \ \hat{\underline{xi}} \qquad (B-29)$$

$$\sum_{i=1}^{n} \left\| BI^{-1}\underline{u} - JI \ \underline{x} + \hat{\underline{wi}} \right\|^2$$

$$= \left\| BF^{-1}\underline{u} - JF \ \underline{x} + \hat{\underline{wf}} \right\|^2 \qquad (B-30)$$

$$+ \left\| JA \ \underline{x} - \hat{\underline{wa}} \right\|^2 + \left\| \hat{\Delta\underline{w}} \right\|^2$$

The first term in Eq. (B-30) can be zeroed by proper choice of $\underline{u}$, given the optimal value of $\underline{x}$. The third term is an irreducible residual. The second term has the same form as a vector measurement, i.e., the $\underline{vi}$ term in Eq. (B-11). Because this term involves $\underline{x}$, it could in principle be combined (via a final least-squares operation) with the earlier solution given by Eq. (B-28). However, doing so is generally not necessary, since the information content of this remainder term is generally negligible. Doing so is also quite undesirable, since the additional computations are very burdensome.

Now, this remaining information term is identically zero under some conditions, meaning that the federated filter solution is then globally optimal. For example, suppose that the total information is divided proportionately among the n local filters at time t', via the use of scalar share fractions in Eqs. (B-23) and (B-24), and with $\gamma_i = \lambda_i$:

$$SI^{-1} = \lambda_i \ S^{-1}; \qquad \Sigma\lambda_i^2 = 1 \quad (\Lambda_i = I\lambda_i) \quad (B-31)$$

$$WI^{-1} = \lambda_i \ W^{-1}; \qquad \Sigma\lambda_i^2 = 1 \quad (\Gamma_i = I\lambda_i) \quad (B-32)$$

This construction causes each i-term in Eq. (B-30) to be *exactly the same*, except for different scalar multipliers λ_i. Thus these terms readily reduce to a single term (having a scalar coefficient $\lambda_{rss} = 1$), with no residual. Hence, the remaining information term is identically zero for this case, and the fused solution (B-22) is globally optimal.

However, if each local filter takes several steps including measurement updates, then the i-terms in Eq. (B-30) are no longer simply related by scalar multipliers. In this case, the remaining information term is not zero, and the fused solution (B-26) is not globally optimal. If we examine the information content of this remainder term, we find that it has an equivalent measurement information (R^{-1}) value as follows:

$$\hat{\underline{wa}} - JA \ \underline{x} \doteq VA^{-1}(\underline{za} - HA^T\underline{x}) \qquad (B-33)$$

$$RA^{-1} \doteq VA^{-T}VA^{-1} \approx \Delta[(\tilde{P} + \tilde{P} \ Q^{-1}\tilde{P})^{-1}] \quad (B-34)$$

Eq. (B-34) represents the additional information "lost" in the multi-step process, where we have assumed $G = I$ for simplicity. The Δ implies a difference across the several local filters. The value $\tilde{P}$ equals $\Phi P \Phi^T$, before adding Q. Thus, when $Q << \tilde{P}$ over each local propagation step, as is generally the case for navigation filters, the information content of the unused remainder term is much less than the filter infor-

280

mation P^{-1} that *is* used in the fusion process:

$$Q \ll \tilde{P}: \quad RA^{-1} \approx \Delta[\tilde{P}^{-1}Q\,\tilde{P}^{-1}] \ll \tilde{P}^{-1} \quad (B-35)$$

(However, if $Q \geq \tilde{P}$, which is not usually the case, then the lost information could be significant.) Thus, operating the local filters independently over several steps is equivalent to ignoring an available but relatively weak (inaccurate) vector measurement at each such step. The resulting fused solution is quite valid, but somewhat suboptimal, in that not all of the available information has been incorporated. (This loss is minimal, compared to the information loss in a single large filter that can't keep up with all the measurements.)

A similar small information loss occurs for the case where the local filters contain only the common states plus their own sensor bias states. Here, the information-share matrices Λ_i and Γ_i for local filter #i contain only two non-zero partitions, where λ_{ci}^2 is the common-state share fraction:

$$\Lambda_i = \begin{bmatrix} I\lambda_{ci} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & I & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \begin{matrix} c \\ 1 \\ i \\ n \end{matrix} \qquad (B-36)$$

$$\sum_{i=1}^{n} \lambda_{ci}^2 = 1; \qquad \sum_{i=1}^{n} \Lambda_i^2 = I \qquad (B-37)$$

The previous cost-function partitioning and fusion operations still apply, given these values of Λ_i and Γ_i. The information remainder term for the fusion process -- the second term in Eq. (B-30) -- can also be expressed in partitioned form. This remainder is identically zero (meaning the result is globally optimal) in the hypothetical case of a system with no common process noises. When common process noises do exist (the usual case), the lost information is proportional to the ratio (B-35) for the common states. In navigation system applications, this value is generally quite small, meaning that the federated loss of information relative to the globally optimal filter is also quite small.

Note that Eq. (B-35) also implies that the one-step information loss is exactly zero when the LFs are reset to zero information ($\Lambda_i = 0$) after each fusion update, such that $\tilde{P}^{-1} = 0$, while the MF retains all the information ($\Lambda_m = I$). Hence, the federated filter *Zero-Reset* mode is globally optimal in the single-step fusion case, even when each LF implements only the common and local bias states.

REFERENCES

1. Speyer, J.L., "Computation and Transmission Requirements for a Decentralized Linear-Quadratic-Gaussian Control Problem," *IEEE Transactions on Automatic Control*, AC-24:2, Apr.1979.

2. Chang, T.S., "Comments on 'Computation and Transmission Requirements for a Decentralized Linear-Quadratic-Gaussian Control'," *IEEE Transactions on Automatic Control*, AC-25:3, June 1980.

3. Willsky, A.S., M.G. Bello, D.A. Castanon, B.C. Levy, and G.C. Verghese, "Combining and Updating of Local Estimates and Regional Maps Along Sets of One-Dimensional Tracks," *IEEE Transactions on Automatic Control*, AC-27:4, August 1982.

4. Levy, B.C., D.A. Castanon, G.C. Verghese, and A.S. Willsky, "A Scattering Framework for Decentralized Estimation Problems," *Automatica*, 19:4, April 1983.

5. Castanon, D.A., and D. Teneketzis, "Distributed Estimation Algorithms for Nonlinear Systems," *IEEE Transactions on Automatic Control*, AC-30, May 1985.

6. Kerr, T.H., "Decentralized Filtering and Redundancy Management for Multisensor Navigation," *IEEE Transactions on Aerospace and Electronic Systems*, AES-23:1, January 1987.

7. Bierman, G.J., and M.R. Belzer, "A Decentralized Square Root Information Filter/Smoother," *Proceedings of 24th IEEE Conference on Decision and Control*, Ft. Lauderdale, FL, Dec. 1985.

8. Bierman, G.J., and D.W. Porter, "Decentralized Tracking via New Square Root Information Filter (SRIF) Concepts," Business and Technological Systems, Inc. and Bierman and Associates, May 1987.

9. Belzer, M.R., and Y.M. Cho, "Micro-Computer Network Architecture for Range Instrumentation Applications," Mentor Technologies, Inc., March 1988.

10. Carlson, N.A. and Neily, C.M. Jr., "Distributed Kalman Filter Architectures," Phase I Final Report, AFWAL-TR-87-1181, Avionics Laboratory, Wright-Patterson AFB, OH, June 1987.

11. Carlson, N.A., "Federated Square Root Filter for Decentralized Parallel Processes," *IEEE Transactions on Aerospace and Electronic Systems*, AES-26:3, May 1990.

12. Carlson, N.A., "Federated Filter for Fault-Tolerant Integrated Navigation Systems," *Proceedings of Position Location and Navigation Symposium, PLANS*, Orlando, FL, November 1988.

13. Loomis, P.V.W., N.A. Carlson, and M.P. Berarducci, "Common Kalman Filter: Fault-Tolerant Navigation for Next Generation Aircraft," *Proceedings of the Institute of Navigation National Technical Meeting*, Santa Barbara, CA, January 1988.

14. Covino, J.M. and B.E. Griffiths, "A New Estimation Architecture for Multisensor Data Fusion," *Proceedings of the International Symposium and Exhibition on Optical Engineering and Photonics*, Orlando, FL, April 1991.

15. Griffiths, B.E. and J.M. Covino, "Net Information Approach Phase I Final Report," Synetics Corp., Wakefield, MA, May 1990.

16. Watanabe, K. "A Decentralized Multiple Model Adaptive Filtering for Discrete-Time Stochastic Systems," *Journal of Dynamic Systems, Measurement and Control*, Vol. 111, September 1989.

17. Martin, D.P. and N.A. Carlson, "Distributed Model Adaptive Estimation," Phase I Final Report, TR-91-003, Integrity Systems, Belmont, MA, August 1991.

18. Carlson, N.A., "Distributed Kalman Filter Architectures Phase II," Final Report, to be published as a WL technical report, Wright Laboratory, Wright-Patterson AFB, OH, January 1995.

19. Carlson, N.A., "User's Manual for the Distributed Kalman Filter Simulator (DKFSIM)", Integrity Systems, Belmont, MA, July 1994.

20. Carlson, N.A., D.P. Martin, and D. Kawa, "Distributed Kalman Filter (Ada) Real-Time Testbed," Final Report, WL-TR-93-1045, Wright Laboratory, Wright-Patterson AFB, OH, Mar 1993.

21. Bierman, G.J., *Factorization Methods for Discrete Sequential Processes,* Academic Press, New York, 1977.

GPS/INERTIAL INTEGRATION OVERVIEW

Richard L. Greenspan
Charles Stark Draper Laboratory, Inc.
555 Technology Square
Cambridge, MA 02139
USA

1. INTRODUCTION

GPS/Inertial integration is the process whereby a superior system navigation solution is produced by properly combining outputs from a GPS user equipment (UE) and from an inertial navigation system (INS). This process is receiving much attention because it is perceived to be a cost-effective means to satisfy navigation requirements that could not be met by either GPS or by an INS acting by itself. Some expectations levied on integrated systems are realistic; others, including hopes for mass-market commercial applications, will be delayed pending the development of suitable low-cost inertial technology.

The technical basis for considering GPS/INS integration is the complementary nature of the navigation errors for each system operating stand-alone. The GPS solution is relatively noisy; the noise-driven variance of GPS positioning errors is on the order of a meter per axis, per position determination. However, GPS errors are bounded, whereas inertial navigation errors are dominated by a low-frequency component that grows in proportion to the mission duration. (The high-frequency content of inertial errors is very small, amounting to a few centimeters (rms) over tens of seconds.) One expects that an integrated navigation solution would perform like an inertial navigator whose errors were bounded by the GPS solution. This performance is actually achieved using one of the least aggressive approaches to integration; further benefits achieved using more aggressive integration options are discussed in the following sections.

The first applications of GPS/INS integration have been in military aviation. This is not surprising because inertial navigation has been almost exclusively applied in military systems. One might object that commercial airliners have carried inertial navigation systems since the 1970's. Although this is true, civil aviation authorities have only recently granted approval to install GPS receivers on aircraft and their use will be limited to provide a "supplemental navigation aid" until at least the mid 1990's.

In contrast, military airborne applications emphasize the ability of an integrated system to meet requirements for precision delivery of weapons or materials on target, even during effective radio countermeasures that cause "outages" in GPS availability. This chapter will focus on the technical issues that arise in satisfying those requirements. The key issues for civilian applications are not yet as clear but the following are likely to be primary:

i) *System Integrity* - The use of internal or external signals and measurements to immediately detect and prevent "out-of-specification" measurements or data from corrupting the navigation solution.

ii) *Partial Outages* - Providing the means to use all available GPS measurements even when a "stand-alone" GPS solution cannot be formed.

iii) *Low Cost* - Developing technology that can be sold at commercially viable prices. This is especially a concern for inertial components.

Item (i) is not discussed in this chapter. However, there is already an extensive literature on integrity techniques. Ref (1) provides a good survey of receiver autonomous integrity monitoring (for GPS); numerous other techniques are being discussed that exploit the integration of several independent sensors (Ref. 2-4). Item (ii) is addressed in Section (3) in connection with the discussion of "deeply coupled integration." Item (iii) is raised in the discussion of error characteristics of inertial navigators also in Section 3.

This chapter is organized in three expository sections. These provide, 1) an overview of the benefits and trade-off issues of integration, 2) an overview of integration architectures and algorithmic concerns, and 3) sample results from performance evaluations of proposed integrations. In all of these discussions, the properties of GPS UE and inertial navigation systems cited here are generic rather than specific and are representative of technology circa 1992.

2. BENEFITS OF GPS/INERTIAL INTEGRATION

The design of any complex navigation system for civilian or military markets reflects the designer's judgment of the best trade-off among the following factors:

Cost
    a) development (non-recurring)
    b) life-cycle (recurring)

Installation Constraints
    a) volume, weight, power consumption
    b) interfaces

Performance
    a) mission requirements/mission environment
    b) reliability/graceful degradation
    c) options for improvement

The following remarks will emphasize the performance considerations because that is the area where the benefits of GPS/Inertial integration are most evident. However, cost and installation factors are often decisive. These are raised throughout the chapter wherever they may be a significant differentiator between alternative integration techniques. Ultimately, the system designer must justify his design as being the best way to satisfy the design

282

problem. It is of critical importance that the authorities who are managing the design team surface all requirements and constraints, both present and anticipated, so that informed and timely choices can be made among the alternatives.

The GPS system can provide a suitably equipped user with a Position Velocity, and Time Solution (PVT) whose errors are generally smaller than those of any alternative navigation system. This performance is achieved in all weather, at any time of the day, and under specified conditions of radio-frequency (RF), signal availability, and vehicle dynamics. Why then would one undertake the cost and complexity of integrating GPS User Equipment (UE) with any other navigation sensor, and in particular, with an inertial navigator?

The goal of integration is to provide more robust, and possibly more accurate, navigation service than is possible with stand-alone sensors. In particular, integration may be the only way to achieve the following (Ref. 5-10):

- Maintain a specified level of navigation performance during outages of GPS satellite reception.

- Provide a complete six degree-of-freedom navigation solution (translational and rotational motion) at a higher output rate than is conventionally available from GPS alone.

- Reduce the random component of errors in the GPS navigation solution.

- Maintain the availability of a GPS solution in the presence of vehicle dynamics and interference.

2.1 Operation During Outages
A stand-alone GPS receiver typically incorporates current measurements to four or more satellites to update its most recent PVT solution. Dead reckoning that may incorporate recent estimates of vehicle acceleration is used to propagate the current PVT solution in between measurement updates. A GPS outage occurs when fewer than four valid satellite measurements are available at each update. During a partial or complete outage, the software for a stand-alone receiver can continue to produce a navigation output if it mechanizes one of the following options, albeit with reduced accuracy:

a) Compute the "least-squares" solution with fewer measurements than there are "unknowns." Ref. (11)

b) Constrain one or more navigation outputs to be fixed, such as the UE clock bias or the vehicle altitude, or constrain the navigation solution to lie along a great-circle path. Ref. (12,13)

c) Incorporate measurements from an external sensor. A barometric indication of altitude is commonly available in military UE as are radar or pilot inserted position updates. Ref. (13,14)

During an outage, the navigation solution becomes less accurate the longer the outage and the greater the vehicle dynamics since the last full set of measurements. The key factor to be specified when deciding whether an auxiliary sensor is required is the maximum acceptable error growth during the outage. In a conser-vative design, maximum error growth is calculated under worst-case conditions of vehicle dynamics.

Outages may be a concern even for UE that track more than four satellites at a time. For example, a GPS antenna mounted on top of an aircraft will only "see" a limited number of satellites during a banked turn, and the "Dilution of Precision" (DOP) parameters for that visible constellation may be unacceptably high. In more extreme cases, a vehicle passing through a tunnel may see no satellites for an extended period, and a military UE may be jammed as it approaches its target.

Combining GPS with an independent navigation sensor (item c, above) is a particularly attractive means to maintain the quality of the navigation service during a GPS outage. In effect, the independent sensor can act like a "flywheel" to provide continuous high quality navigation outputs. Inertial navigators are commonly considered for this role because they are passive, self-contained, widely available, and are not subject to the causes of GPS outage. However, they are generally more expensive to buy and integrate than other radio-navigation sensors such as LORAN or Omega. Their use has generally been limited to military and commercial aircraft. However, low-cost, low-performance inertial sensors implemented using mass-production microelectronics technology are emerging from research laboratories. These may provide the technological basis for an economical solution to GPS outages in civil aviation and commercial applications such as trucks and automobiles. Ref.(15,16)

With respect to GPS/INS integration performance during outages, the key questions are:

(1) What quality INS is required?

(2) How complex is the integration required to exploit the inherent INS quality to achieve mission objectives?

The resulting performance must then be weighed against the cost to determine whether to implement the optimum integration or to accept a less expensive, lower performance solution.

2.2 Providing All Required Navigation Outputs
GPS UE's routinely estimate only the translational motion of a point referenced to the GPS antenna. Interferometric processing of GPS signals received at multiple antennas can also provide rotational (attitude) information. Ref. (17-19) However, we assert that inertial sensing is preferable to interferometry for terrestrial users whenever it is available. This preference is based on three considerations in addition to superior performance:

(a) The inertial system is self-contained and is not vulnerable to outages (except those caused by equipment failure).

(b) Installation of an inertial system on an aircraft is less demanding than an interferometer, and is probably less demanding on ships and vehicles too.

(c) The theoretical noise floor on the accuracy of a short-baseline (1-5m) interferometer has not yet been achieved. It appears that multipath is the culprit, and that is it premature to expect that anti-multipath techniques will be effective and practical. Ref. (20)

In addition to attitude indication, the inertial navigator is desirable because its accelerometers typically sense velocity changes at up to a 1.0 KHz rate, with a 200 Hz output rate being commonly available. Therefore the INS routinely outputs navigation solutions one to two orders of magnitude more often than a GPS UE. This high output rate allows the INS to provide accurate inputs to vehicle control subsystems, platform-stabilization systems, or to pilot displays, and velocity aiding inputs to GPS tracking loops.

It follows that an integration in which GPS is used to bound the error growth of an INS-based system navigation solution would be very effective whenever GPS was available, and the availability of a calibrated INS may be the only means to maintain nearly as good performance during an outage.

2.3 Reduced Noise in GPS Navigation Solutions
In a stand-alone GPS receiver, the navigation processor usually implements a linear filtering algorithm in which the previous navigation solution is propagated to the current measurement epoch. Because GPS does not directly sense acceleration, the propagated solution is sensitive to errors in the previous acceleration estimate or to jerk that changes the true acceleration during the propagation interval. In contrast, an inertial system measures position change very precisely in the interval between GPS updates. This property can be exploited by a well-tuned Kalman Filter, using GPS measurements to estimate *errors* in the INS output. Since these errors change slowly, the filter can smooth its update over many GPS measurements thereby reducing the effect of additive noise on any one update. The result is that the "integrated navigation" solution appears to be much "smoother" than the stand-alone GPS solution. Ref. (21,22)

2.4 Increased Tolerance to Dynamics and Interference
The INS velocity solution may be fed back to the GPS UE to reduce the apparent dynamics of the input to the GPS code and carrier loops. This has two effects; Ref. (5,8,10)

(1)   A fixed bandwidth aided tracking loop can maintain lock on GPS signals in the presence of dynamics that would cause the unaided receiver to break lock.

(2)   The tracking loop bandwidths can be reduced to the minimum amount required to track the *errors* in the INS aiding signals.* (INS position errors are mostly low-frequency).

The net result of these actions is that the INS aided GPS receiver can maintain lock and provide GPS measurements over a much wider range of vehicle dynamics and radio-frequency interference than the unaided, stand-alone receiver. In particular, there is provision to operate both RCVR 3A and the MAGR in the "INS" (aided mode), or the "DRS" (Doppler Aided Mode) when improvements in GPS availability at high levels of jamming or dynamics are required. In the following remarks, we consider only the INS mode because there is almost no experience with operating any GPS receiver in the Doppler mode. Operation in the INS mode provides three benefits to a GPS receiver.

(1) The INS velocity output can be used to "aid" the MAGR code-tracking loop when the carrier loop has lost lock. This in turn means that the code loop tracks only the *errors* in the INS aiding. The bandwidth of these errors is much less than the bandwidth of

---

* This feature breaks down for "unauthorized" users in the presence of selective availability clock dither.

the raw dynamics, hence the tracking loop bandwidth can be reduced. This bandwidth reduction gains additional jammer suppression. Thus the specification for State 3 (code loop) tracking is increased from J/S = 46dB to J/S = 54 dB in the INS aided mode.

(2) In the INS mode the INS position and velocity are treated by the MAGR as a reference navigation solution that is to be corrected by GPS measurements. A direct consequence of using the INS as a reference in the MAGR is that interpolation/ extrapolation/ synchronization errors in the receiver output are much less sensitive to acceleration and jerk than in the PVA mode. (i.e. the INS mode "foreground" solution is almost as good as the "background" solution.) Another consequence is smoother output. It is a property of the INS reference that the INS errors, though usually much larger than MAGR performance specifications, are smoothly growing functions of time: i.e., the short term quality of the INS-based reference (noise level, precision) is much better than the PVA solution maintained by an unaided receiver. Software in the receiver (a Kalman filter) is used to produce an output that combines the best features of the INS with the GPS: i.e., the receiver navigation solution output in the INS mode is smoother than in the PVA mode and (in the presence of valid measurements) has errors that satisfy the MAGR performance specifications over a wider range of dynamics and jamming.

(3) The combined INS/GPS solution is continuously available during any interruption of GPS signals (outage). It can be used to reduce the uncertainty in platform location *during* a GPS outage to a small enough uncertainty that certain rapid reacquisition strategies built into the MAGR can be exercised. This has the effect of reducing the duration of a GPS outage by hastening reacquisition.

3.   GPS INTEGRATION ARCHITECTURES AND ALGORITHMS

The degree of complexity of the integration should reflect the mission requirements; it may also be limited by the investment that can be made to obtain those objectives. Integration strategies and mechanisms may be very simple (for example: choose the GPS UE position and velocity as the "integrated" solution when GPS is available with a given precision, or else choose INS position and velocity as the "integrated" solution) or relatively complex (for example: optimally combine GPS UE measurements with INS outputs, Doppler radar outputs, baroaltimeter signals, true airspeed, and other sensor data). However, in the following remarks we limit our attention to alternatives involving only GPS integrated with an inertial system.

3.1 Integration Architectures
Figure 3-1 illustrates three generic functional architectures for GPS INS integration. The GPS receiver and the INS are treated as navigation systems in architectures a and b, with GPS supplying a Position, Velocity, and Time solution (P,V,T) and the INS supplying a Position, Velocity, and Attitude solution (P,V,θ), respectively. In architecture c, the GPS and INS are treated as sensors producing line-of-sight measurements $\left(\rho, \dot{\rho}\right)$ and accelerations and angular rates ($\Delta V, \Delta \theta$) respectively. In addition to the GPS and INS units, each one includes various data paths and a processor unit that mechanizes the integration algorithm. These alternatives are distinguished by the data that passes between the sub-system components. The proper interfacing and control of

these components may incur the largest part of the cost of an integration project, but those concerns are not within the scope of architectural considerations.

3.1.1 Stand-Alone Mode

Figure 3-1a illustrates the configuration in which GPS UE and an INS produce independent navigation solutions with no influence of one on the other. The integrated navigation solution is mechanized by an external integration processor that may be as simple as a selector or as complex as a multi-mode Kalman Filter. All data busses are "simplex" (unidirectional). The characterization of Figure 3-1a as a "stand-alone" mode is based on the independence of the GPS and INS navigation functions. Note that, in principle, the *hardware* could all be packaged in one physically integrated ("embedded") unit yet the functionality would be that of a stand-alone architecture.

The potential benefits of a stand-alone integration are:

a) It is the easiest, fastest, and potentially the cheapest approach when an INS and GPS are both available.

b) It provides some tolerance to failures of subsystem components (except in the embedded configuration, see 3.3).

c) Using an integration processor as simple as a selection algorithm can provide en route navigation at least as accurate as available from an INS.

A) STAND-ALONE MODE

B) LOOSELY COUPLED MODE

C) TIGHTLY COUPLED MODE
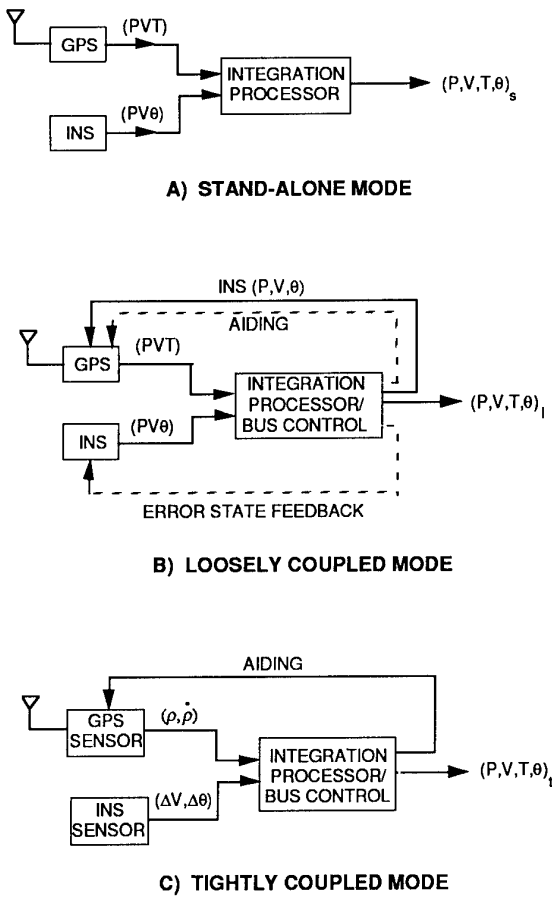
Figure 3-1. Generic GPS/INS Integration Architectures.

3.1.2 Loosely Coupled Mode

Figure 3-1b illustrates a configuration in which there are several data paths between the integration processor and the GPS and the INS equipment. Among these, the provision of the system navigation solution to the GPS UE is the most important for getting the maximum benefit from the integration filter. The inertial aiding of GPS tracking loops is of next greatest benefit, and feedback of error states to the INS is of second-order benefit. There may also be some benefit for system reliability to the extent that individual components are mature and have been well tested compared to emerging technology.

3.1.2.1 Reference Navigation Solution

GPS UE generally employ a Kalman filter mechanization to compute (P,V,T) updates based on current tracking loop measurements. A stand-alone UE does not directly sense acceleration; it must use relatively "noisy" acceleration estimates based on recent velocity measurements for a "dead reckoning" propagation of the previous navigation solution forward to the epoch of the current tracking loop outputs. The situation changes dramatically when the System Navigation solution is fed back to perform that propagation. In effect, the GPS measurements can now be used (within the UE navigation filter) to correct the system navigation solution. Over short periods of time, that solution is very accurate because it incorporates INS data based on acceleration sensing. The UE filter is then mechanized to estimate INS (or System) error states having relatively low dynamics and low uncertainty (process noise). The filter can be "tuned" to have a longer time constant (filter memory) thereby increasing the effective averaging of each noisy GPS measurement. The result is a great reduction in noiseness of the GPS P,V,T solution.

3.1.2.2 Inertial Aiding of GPS Tracking Loops

As mentioned in Section 2, the availability of a GPS navigation solution can be significantly increased when inertial aiding is used to reduce the vehicle dynamics that are tracked by the UE code and carrier loops. In principle, this aiding could be applied directly from the INS to the GPS UE, but it is shown as an output of the integration processor in Figure 3-1b because:

• GPS tracking loops must be aided by the projection of vehicle velocity along the line-of-sight (LOS) to each satellite being tracked. The conversion from inertial coordinates to GPS LOS coordinates is most appropriately done in the integration processor or in the GPS UE itself. In either case, INS velocity information is available within the processor hence aiding can be part of the data flow to the UE. This avoids the expense and risk of developing a custom interface from the INS to the GPS UE.

• Executing the coordinate transformation external to the INS retains flexibility in the selection of INS equipment and avoids the need to develop custom GPS/INS interfaces for each application. However, this raises a concern for "data latency" (i.e. feeding delayed data to the tracking loops) as discussed below in Section 3.3.

3.1.2.3 Error-State Feedback to the INS

Most inertial navigation systems have the means to accept external inputs to reset their position and velocity solutions and to adjust the alignment of their stable platform. The adjustment may be executed by a mathematical correction in a "strap-down" inertial system, or it may be realized by torquing a gimballed platform. In

either case the use of feedback can maintain inertial navigation errors at a level for which their dynamics are accurately modeled by the linearized error state propagation equations usually embodied in the integration filter. However, this is a minor consideration because position and velocity errors would have to exceed many km and many m/s before non-linear effects become important.

3.1.3 Tightly Coupled Mode

Figure 3-1c illustrates the so-called tightly coupled integration mode. It differs from the loosely coupled mode in that both the GPS receiver and the inertial components are limited to their sensor functions. They are treated as sources of GPS code and carrier measurements and inertial indications of acceleration (velocity change) and angular rate respectively. These sensor outputs are then combined in one navigation processor that may mechanize an appropriately high-order integration filter. Ref. (23-25)

In the tightly coupled mode, there is only one feedback from the navigation processor. Figure 3-1c illustrates the use of velocity aiding to the GPS tracking loops. Acceleration aiding could also be effectively used, but we are not aware of any particular mechanization using other than velocity aiding. The other paths used in loosely coupled architectures are not needed here because all computations involved in navigation processing are now internal to one processor.

The concept of tightly coupled integration is often raised in connection with "embedded" GPS receivers. These are not necessarily synonymous. However, it is reasonable that one would choose to mechanize a tightly coupled integration algorithm if one had already taken the effort to design a GPS receiver that is physically and electrically integrated with an inertial sensor or with a powerful navigation processor. We return to this point in Section 3.3.

As of January 1993, it was uncertain whether the option to implement deeply-coupled integraton would be denied to authorized GPS users. Plans to develop tightly-coupled systems such as the GPS Guidance Package and the Honeywell/Texas Instruments Model 764-C3 were well underway when the GPS Joint Program Office introduced a new GPS security implementation called the "SAASM" (Selective Availability, Anti-spoofing Module). Comments from industry were due in January; there was a widespread concern that the amount of integration processing (e.g. Kalman Filtering) that could be supported by a prospective SAASM mechanization would not be adequate to achieve the full benefits attributable to deep integration using tightly coupled sensor outputs. The jury is still out on this issue.

3.2 Integration Algorithms

The basic choices for GPS integration algorithms are:
- (I) Selection, with or without INS resets
- (II) Fixed Gain Filter
- (III) Time-varying filter

These are listed in order of increasing complexity and optimality. Each one can be used with any one of the architectures listed in 3.1 but the incremental pay-off of a more complex filter is directly related to the quality of the input information.

3.2.1 Selection

A selection algorithm chooses the GPS indicated (P,V,T) as the system navigation solution whenever the GPS UE indicates that this solution is within acceptable bounds on its accuracy (via the GPS "Figure-of-merit (FOM)). INS data may be used to interpolate between successive GPS updates when a higher output rate is needed than can be provided by the UE. During GPS outages the INS solution extrapolates from the last valid GPS solution. (The process of forcing the INS solution to equal the current GPS indicated velocity and/or position is known as a "reset" if that correction is actually fed back to the INS).

3.2.2 Filtering

The general filtering problem involves trying to estimate time-varying "states" whose evolution is characterized by known laws of propagation, which are usually taken to be a coupled system of linear differential equations driven by white noise.

States cannot usually be measured directly but are inferred from measurable quantities to which they are related. These measurements may be made simultaneously, or sequentially at a series of distinct points in time. The filter will generally incorporate knowledge of the statistics of the measurements.

Knowledge of the way the states change (propagate) in time, knowledge of the way the measurements are related to the states, measurement statistics and measurement data are all used in each state update. The most common update algorithms use linear filters, e.g., ones in which the updated state is a linearly weighted sum of the measurements and the previous state value.

Position and velocity of an aircraft are examples of quantities that may be chosen as *states* in a filter (these are referred to as whole-value filter states). For whole-value position and velocity states, the propagation equations are simply the equations of motion of the aircraft. To make the whole-value filter propagation equations a better reflection of the real world, acceleration states could be added (otherwise, by its omission, acceleration must be treated as "noise," driving the derivative of velocity). GPS indicated position and velocity are examples of *measurements* that might be processed by an integration filter with whole-value states. At one extreme, the integration "filter" could ignore everything except the GPS receiver position and use this as the integrated position. This degenerate case is the selection mode cited above in which the state propagation equations and any other available measurements would be ignored. In general, some rule must be used in order to determine how much weight should be put on a measurement and how much weight should be put on the propagated states. For the above degenerate case, the weight on the GPS UE position is one and the weight on the propagated state is zero. The weight on the measurement is referred to as the filter *gain*.

Another choice of states are the *errors* in position and velocity indicated by the INS (these are referred to as error states). For a filter whose states are INS errors, accurate representations and linear approximations of the propagation equations are well known. As in the case of whole-value states, additional INS error states (for example, states for azimuth and tilt errors, accelerometer bias, and gyro drift) could be added to the filter in order to make the propagation equations a better model of the real world. Of course, the degree to which the filter must reflect the real world is a function of the estimation accuracy required, and that is a reflection of the mission requirements.

For a GPS/INS integration filter with INS error states, the measurements would actually be the differences between GPS position and INS position and the differences between GPS velocity and INS velocity. As with the case of whole-value states, some

286

rule must be used in order to determine how much gain should be put on the measurements and how much weight should be put on the propagated states when computing state updates.

We mention in passing that the "optimum" filter may require an impractically large number of states. Separation of the 3-dimensional filtering problem into a "vertical" estimator (possibly mechanized as a fixed-gain filter) and a "horizontal" estimator is a common strategy in GPS inertial integration. Options to decompose a high dimensional estimator into combinations of lower dimensional filters have also been described in the literature. Distributed filtering and federated filtering are the terms under which these options are usually cited as in References (24,25). Both are believed to be more robust than the optimum filter when the design must be tolerant to imperfect information about the estimation problem. It is claimed that federated filters are more fault-tolerant. These details are beyond the scope of this survey.

3.2.2.1 Fixed Gain Filters
In a fixed filter, the propagated estimates are combined with new measurement data using predetermined gains. The gains are fixed in the sense that they have been loaded into computer memory a priori, so that the filter selects from a short list of gains, rather than computing them. Different gains may be used for different sensor status and operational status, reflecting the uncertainties in the propagated solution and in the measurements. In general, the gains in a fixed-gain filter can have any value (they should at least properly reflect the relationships among the measurements and the states).

If the state dynamics and their uncertainty are limited, and there is negligible variation of measurement noise during the interval of interest, it may turn out that the optimum filter gains will not vary very much during the mission. In that case the performance penalty of mechanizing one fixed set of gains (or a few selectable sets of gains) compared to optimum time-varying gains may be acceptably small. The benefit to the integrator is a vast decrease in computational burden and memory required to implement the filter. It may even be effective to pre-calculate and store time-varying gains for use during standard missions.

3.2.2.2 Time Varying Gain (Kalman Filter)
In the Kalman filter, new gains are computed every time measurements are available. The Kalman Filter is a recursive implementation of the optimum "least-squares error" estimation algorithm. It is optimum in the sense that it strikes the correct balance between uncertainly in the presumed dynamics of the states being estimated (process noise), uncertainly in the measurements (measurement noise) and the observability of individual states (sensitivity), required to minimize the figure-of-merit. See References (26-28) for a detailed discussion of Kalman Filtering. In the present context, we note that the updating of N states by M measurements involves substantial matrix manipulations, propagation of difference equations, and memory to store the matrices. Current technology can handle updates of around 20 "states" at up to a few times per second in a reasonably cost-effective processor. Since upwards of 100 error sources may influence an integrated GPS INS solution, the brute force approach to real-time integration is not yet computationally feasible. Each designer must complete detailed design studies to determine the minimum number of states and the update rate that will result in an acceptable navigation error using the available processor resources and with sufficient design margin. Given the rapidly changing computational

capabilities available to avionics integrators, questions of computational feasibility should be reconsidered every few years.

3.2.2.3 Custom Tuning the GPS Internal Navigation Filter
Custom tuning is an action that can be taken internal to the GPS UE. It does not require an external integration filter to derive some degree of more robust performance in the presence of high vehicle dynamics or jamming.

For example, the standard MAGR or RCVR 3A internal Kalman filter is a complex computer program which embodies a mathematical algorithm for combining the INS reference with GPS measurements in a manner that is statistically optimal. About two dozen parameters in the program characterize the dynamics, timing, and statistics of the INS and GPS. Adjusting these parameters is called tuning. The standard tuning in the MAGR assumes a very low quality INS (large errors, large error growth, very noisy, and very unstable calibration coefficients) a poor clock in the UE, and very high noise in the carrier-tracking (delta-range) measurements. Using these assumptions, Rockwell Inc. produced a navigation solution that is impervious to bad UE clocks, carrier loop tracking cycle slips, and substandard INS quality.

Although the standard tuning of the MAGR produces a very robust solution in the presence of GPS measurements, this solution may deteriorate within a few minutes when no GPS measurements are available. Custom tuning tries to take advantage of a navigation-grade INS by matching the tuning parameters to the expected INS performance. One effect achieved is similar to what happens when an external integration filter is used: The INS is better calibrated using the nine INS error states in the retuned MAGR internal Kalman filter so that when a GPS outage occurs (e.g. from jamming) the MAGR's navigation solution error growth is much slower than with the standard tuning. Thus, the uncertainty in vehicle position that must be searched to reacquire GPS signals after an outage is reduced. This is the mechanism whereby custom retuning can result in faster reacquisition after jamming. This result was observed in a recent comparison between a MAGR integrated with an RLG and a tightly-coupled embedded navigator.

Another advantage of custom tuning, not available to an external integration, is that the effective code tracking bandwidths in state 3 can be made more narrow. The adjustment is made using certain tuning parameters. The procedure is successful because the INS-aided loop is known to be tracking the errors of a higher quality INS than is assumed by the standard filter tuning parameter selection. Taking full advantage of the INS aiding means that the loops are more resistant to jamming and high dynamics and stay in lock longer.

Under no circumstances will custom tuning improve the positioning accuracy of the MAGR when GPS code measurements are available from four satellites. When satellite measurements are available, accuracy is limited by the 3 to 5 meter bias errors in the measurements and by the satellite geometry (GDOP). In summary, the benefit of custom tuning is not to improve accuracy in the benign case, but to increase the receiver's resistance to jamming and high dynamics and to obtain faster reacquisition of specified performance after code loops lose lock for any reason.

Negative factors of retuning include the cost of requalifying the set, and the on-going logistics burdens/costs of stocking replace-

ments for many different varieties of sets at depots. The use of a few predefined sets of strapable coefficients might reduce this burden at the cost of adding safety procedures to ensure that the wrong coefficients could never be loaded. Some additional risks are also incurred because the custom tuned MAGR assumes a high quality INS, but unlike an external integration program, no explicit checks to monitor INS performance can be made to identify substandard INS performance. Thus, the retuned receiver is more sensitive to INS quality than the standard tuned receiver.

3.2.3 Discussion

The stand-alone mode is inferior (in performance) to the loosely coupled integration mode. The stand-alone and loosely coupled integration models are inferior (in performance) to a tightly coupled mode because information inherent in the sensor measurements is lost in the receivers mechanization of its Kalman filtered (P,V,T) solution, i.e., it is not always possible to backtrack from a (P,V,T) solution to the raw GPS measurements with sufficient bandwidth and precision to support a tightly coupled integration..

The feedback of the system navigation solution to the INS (via resets) is a "second order" improvement in the loosely coupled mode (and is inherent in the tightly coupled mode).

The current generation of military high dynamics GPS UE (Receiver 3A, MAGR) is often criticized because it must be integrated in the loosely coupled mode. This criticism is valid, but it really addresses a cost tradeoff in which the integration filter in RCVR 3A was limited to twelve states, of which nine represent very generic INS errors (P,V,θ). The decision for RCVR 3A was based on unit cost and the desire to produce a generic standard equipment that did not burden any user with features that were not justifiable in his application. In principal, a modified RCVR 3A operating in the INS mode with an expanded internal filter and appropriate software could perform nearly as well as a "tightly coupled" integration.

In design studies of elaborate GPS/INS integration filters as many as eighty inertial error states are modeled, in addition to GPS error states related to delay measurement bias, tracking loop errors, propagation errors, and user clock errors. In some ultra-precise systems, it may even be useful to incorporate additional states that model multipath effects. Nevertheless, many studies have shown that most of the benefit of expanded error state formulations is gained with 25-30 states, and that adequate performance can usually be obtained from 14-17 states. Ref. (29,30)

The tightly coupled mechanization does avoid one problem commonly attributed to loosely coupled integration, namely the possibility of instability (in state estimates) arising when the GPS navigation errors become highly correlated with INS navigation errors. This situation may occur at low input signal to noise ratios when GPS code loops remain in lock only because inertial aiding allows the loop bandwidth to be reduced thereby reducing the effective levels of noise and interference. Now, the narrower the loop bandwidth, the more the loop error approximates the error of the aiding signals so that the correlation cited above becomes significant. See References 31 and 32 for further discussion of this point.

3.3 Embedded Systems

As GPS approaches its operational status, there has been a massive increase in investment in civil GPS technology which has led to smaller, lower power consuming, higher performance UE than were dreamed of as recently as the late 1980's. One consequence of this trend is that GPS UE can be packaged on a single card that can be embedded in other systems. As noted in Section 3.1, the concept of GPS embedded in an INS is one such application that is being prominently discussed at present, with several efforts underway to demonstrate the concept. Ref. (33-36)

Setting aside the valid claims of savings in size, power weight that accrue from embedding, it is reasonable to ask whether there is any functional or performance payoff directly attributable to embedding. The answer is a qualified yes. There are potential performance improvements, but the system may be vulnerable to a single-point failure, such as a power supply or the processor.

3.3.1 Tight Coupling

There is no inherent reason to claim that embedding implies tight-coupled integration. An embedded receiver could be stand-alone, loosely-coupled or tightly coupled. However, developers of embedded systems have tended to mechanize tight coupling as a performance feature.

3.3.2 Carrier Loop Aiding

Standard military UE use inertial aiding only for code loops, and only after carrier loops have lost lock. The decision to limit the INS aiding goes back to the late 1970's when it was argued that the latency (time delay) between the sensing of inertial velocity and its receipt at a GPS receiver could be as large as tens of milliseconds, even with the high speed data busses that were available. With this much delay, it was argued that errors in the aiding signal during accelerations or turns could be large enough to drive the carrier loop out of lock.

There are at least two ways to mitigate this concern in an embedded system. The most common approach is to customize the data link between the INS and the GPS carrier loop to reduce the latency to a few tens of microseconds and to minimize the uncertainty in the latency. With that small a delay, the maximum error of the aiding signal is negligible even for an aircraft rolling as fast as 1 rev/s, and moving towards a satellite with a relative velocity of 2000 ft/second. Under those conditions the error caused by a 20 μs delay in attitude indication would be approximately V_ε, where

$$V_\varepsilon < 2\pi\,(20 \times 10^{-6})\ \text{radian} \times 2000\ \text{f/s} < .25\ \text{f/s}$$

which is well within the acceptable range for GPS receivers. An alternative that has not yet been mechanized is to delay the GPS signals by an amount that matches the latency, before the aiding signal is applied. For modem pre-correlation digital GPS receivers one could store tens of milliseconds of GPS samples in a buffer mechanized by a single memory chip.

3.3.3 Tracking Fewer Than Four GPS Satellites

The loose-coupling approach (see Figure 3-1b) integrates a GPS (P,V,T) solution with an inertial (P,V,θ) solution. When the GPS solution is unavailable, the integrated solution "flywheels" using the inertial solution as corrected at the start of the GPS outage. In contrast, the tightly coupled solution uses raw GPS measure-

ments, which are available as long as one or more satellites are being tracked. Thus it is a more robust solution vs. outages that could prevent a GPS navigation solution from being formed in the loosely coupled configuration. Reference (29) gives a good insight into the potential performance improvement.

However, there is one caveat to consider. Conventional GPS UE, such as GPS RCVR 3A can continue to provide a navigation solution (albeit degraded) with only two or three satellites. Thus it is inappropriate to claim that a loosely coupled integration *must* convert to a free-running inertial solution in the presence of one or more satellite outages. The performance will depend on the details of the GPS UE mechanization.

3.3.4 Quantization
All calculations within an embedded system are more likely to be executed as "full-precision" quantities than in a system wherein the GPS and INS and navigation processor are connected by data busses. These busses are usually so heavily used that data must be coarsely quantized for data transmission (compared to their internal precision) in order to satisfy communication bandwidth constraints and to conform to data transmission protocols. An alternative that has not been explored in GPS navigation data communications is to use data-compression to increase the information content of the message structure. This would require reworking interfaces and message protocols, but the effort might be cost-effective in high precision applications.

4. INTEGRATION CASE STUDIES

We shall consider two case studies that illustrate performance benefits of GPS/INS integration. The properties to be addressed include:

a)   In flight INS alignment

b)   Reduced error growth of a "calibrated" INS during a GPS outage

4.1 Using GPS for Inflight Alignment
Alignment is the process that ties inertial platform coordinates to the geographic frame in which the host vehicle navigates. Alignment establishes the conditions necessary for the proper integration of accelerometer outputs to accurately estimate the changes in user position and velocity.

Gyro compassing alignment typically extends over a 10-15 minute period prior to a standard aircraft takeoff. We need to look at the factors that contribute to alignment time in order to understand the opportunities for using GPS to speed up the process, or to allow it to proceed in-air in addition to on the ground.

4.1.1 Conventional Alignment Procedures
Alignment consists of platform leveling and establishing a reference bearing. The following discussion is based on the alignment of a gimballed platform INS, but the general conclusions and timing estimates are good approximations to the performance of a "strapdown" INS.

Leveling
Leveling is the process of establishing one plane of the IMU instrument package perpendicular to the local gravity field. This is typically accomplished by mounting two accelerometers on that plane so that their input axes are not colinear (they are usually

mounted at right angles), and tilting the plane to null the accelerometer outputs.[*] The perpendicular to this plane defines "UP" in an East, North, Up coordinate system.

The error signal that drives the leveling loop is proportional to the instantaneous value of the tilt error, i.e., to approximately 1.0 g times the alignment error (in radians). In mathematical notation, the output for each horizontal accelerometer is $\hat{g}$, where

$$\hat{g} = \left(1 + s_f\right) g \, e_\phi + b$$

where

g   is the magnitude of the local gravity vector

e_ϕ is the angular error of the platform normal to the accelerometer axis (i.e., the tilt error)

b   is the accelerometer bias

s_f is the accelerometer scale factor error

For an aircraft at rest $\hat{g}$ should be zero when the platform is level, however, the accelerometer bias causes the null to occur for a non-zero value of tilt error. For accelerometers used in contemporary fighter aircraft b is about 150 μg and s_f is about 500 ppm. For g=1.0, the steady state tilt error e_ϕ is then approximately

$$\left|e_\phi\right| = \frac{b}{1 + s_f} \approx 150 \mu radian = 30 arc \sec \quad (4.1)$$

where we assume that angle quantization effects in the leveling feedback loop are negligible. This 30 arcsec is a bias error; the random component of leveling errors is on the order of a few arcseconds.

North Seeking
North Seeking is the process of establishing a reference direction (azimuth) in the leveled plane containing the "east" and "north", accelerometers. The most widely used scheme for self-alignment of the platform is gyrocompassing.

Gyrocompassing exploits the following properties of gyroscopes, for a local level north-oriented system at latitude λ:

(I)   If the gyro input axes are physically aligned with the geographic axes (North, East, Up), then the system will remain aligned if each gyro is individually torqued at a rate equal to the projection of the earth rate on its input axis.

(II)   If the platform is level, but not aligned to north it will rotate around the east axis, causing the north axis to rotate from horizontal. This rotation produces a level error than can be sensed by the north accelerometer, and used to drive the azimuth error to zero.

---

[*] Alignment for "strap-down" inertial systems is a mathematical operation that does not involve physical motion of the sensor platform.

The rate of rotation of the north axis in response to an azimuth misalignment of radians is given by $\dot{\theta}_N$

$$\dot{\theta}_N = \theta_A \Omega_e \cos \lambda$$

where Ω_e is the earth rate. If this rate acts for t seconds it produces net level error (tilt) of θ_N (radians)

$$\theta_N = \theta_A \Omega_e t \cos \lambda \qquad (4.2)$$

From Eq. 4.2 we see that for gyrocompassing to convert a misalignment error θ_A into an equally large observable tilt error (i.e., for $\theta_N = \theta_A$), then the effect of the error must be integrated for at least t_0 seconds, where

$$t_o = (\Omega_e \cos \lambda)^{-1}$$

At mid latitudes, (say $\lambda = 45°$) and for

$$\Omega_e = 2\Pi radians\ /\ day$$

we have

$$t_o = \frac{1}{\sqrt{2\Pi}} days = 5.4 hours \qquad (4.3)$$

In practice, the requirement for alignment accuracy is substantially less than the requirement on leveling accuracy, and this allows the alignment to proceed faster than indicated by Eq. (4.3). We can estimate this speedup using the approximation that the standard deviation of leveling errors is inversely proportional to the time spent in leveling. For medium accuracy inertial sets that are found in contemporary aircraft the random noise in leveling is on the order of a few arcseconds, and the uncertainty in azimuth alignment after gyrocompassing is on the order of 160-200 arcseconds. The ratio of these standard deviations is about 50. Substituting $\theta_A = 50\,\theta_N$ in Eq. (4.2) yields $t_0 = t_0 /50$, which is approximately 7 minutes. This estimate is consistent with Air Force specifications for the time to achieve standard accuracy alignment by gyrocompassing under favorable conditions. Ref. (37)

4.1.2 GPS Aided Alignment
Leveling is essentially an instantaneous process (especially in a strap-down system) because it seeks to drive two directly sensed gravity indications to be equal. This gives the clue as to the benefits of in-flight alignment. Velocity changes sensed by the inertial system are compared to velocity changes sensed by GPS and the alignment parameters are adjusted to drive the residuals to zero. Sensing INS alignment errors from velocity residuals is significant because it eliminates the delay required to integrate a velocity error into a position error of detectable magnitude for gyrocompassing alignment. Now, combine this observability with the opportunity to use aircraft maneuvers (turns, climbs, dives) to apply acceleration in the horizontal navigation plane (North, East), in addition to the vertical plane (up), and the potential use of GPS for in-flight alignment is clear.

The issues for in-flight INS alignment using GPS are related to the "noisiness" of GPS velocity measurements compared to inertial measurements. Unless the INS misalignment is so large that, at the vehicle velocity, it produces a velocity error that exceeds the GPS noise level, the alignment will require several minutes, or more, of observations. Thus the time to align is a function of the accuracy goals, the magnitude of acceleration that can be applied during maneuvers, and the noise level of GPS measurements.

Figures 4-1 to 4-3 were generated in a linear covariance simulation of INS alignment for a Fighter Aircraft using a standard (1.0 nm/h) inertial navigator. The integration filter mechanized eleven states (3 position, 3 velocity, 3 misalignment, 2 clock), and the inputs are GPS pseudo range and delta range and velocity changes output by the INS.

Two takeoff profiles were investigated:

Standard Takeoff
(1)   Accelerate at 0.3 g to 160 knots indicated airspeed (KIAS) using MIL (standard) acceleration.

(2)   Begin rotation to 5° pitch angle and accelerate to 170 KIAS takeoff speed.

(3)   Climb at 5°, and accelerate to 385 KIAS, on 45° heading (northeast).

(4)   Continue climbing at 5°, and constant velocity to cruise altitude at 20,000 ft, and 385 KIAS.

Alert (Scramble) Takeoff
(1)   Accelerate at 0.4 g to 165 KIAS using maximum afterburner thrust.

(2)   Rotate to 12° pitch angle and climb at 0.4 g to 415 kt at 2500 ft altitude.

(3)   Climb from 2500 ft to 20,000 ft altitude at constant speed and 12° pitch, arriving at 20,000 ft approximately 180 seconds after takeoff.

(4)   Level flight at 20,000 feet then accelerate to Mach 0.85 (533 kt).

Navigation for "standard" takeoff assumes that the INS has been aligned by gyrocompassing and that GPS is continually available. Navigation for "alert" takeoff assumes that the INS is aligned by a stored heading reference, and that GPS is not available until five minutes after takeoff. In both cases, different turn, climb, and dive maneuvers after takeoff were considered for their effectiveness in reducing alignment errors after takeoff.

The values cited in Table 4-1 were used to characterize the standard errors (1 - s levels) remaining after gyrocompassing and stored heading alignment respectively. These are the initial conditions for subsequent Kalman filtering of GPS plus inertial sensor outputs.

Table 4-1. Alignment Errors at Takeoff (arcseconds)

	Level	Azimuth
Stored Heading	40	450
Gyrocompassing	37	225

Three individual test cases are presented here. These are listed in Table 4-2.

Table 4-2. Test Cases for Inflight Alignment Studies

Case	Takeoff Mode	Alignment Mode	GPS On At	Alignment Maneuver
1	standard	gyrocompass	0 s	2/3 g turn
2	alert	stored heading	300 s	1 g turn
3	alert	stored heading	300 s	2 g S-turn

Case 1

Case 1 illustrates the potential benefits of using GPS aided alignment in a standard mission, where the INS is gyrocompassed prior to takeoff, and GPS is available throughout takeoff. A 2/3-g turn that produces a 45° heading change beginning at t = 495 seconds is included to illustrate the additional improvement available from in-flight maneuvers.

Figure 4-1 illustrates the history of RMS alignment errors. As expected, the availability of GPS during the period of acceleration at takeoff reduces the alignment error; it appears to decrease asymptotically towards about 50 arcsec. The additional acceleration after t = 495 seconds produces a modest further improvement. The final azimuth error is approximately the same as the level error, which signifies that accelerometer bias is the limiting factor in improving alignment by this procedure. An analysis of the error sensitivity of the tilt estimates confirms this intuition, with nonorthogonality of the horizontal accelerometers with respect to the "up" direction being the only other significant error source.

Case 2

Case 2 illustrates potential benefits of GPS for a "scramble" takeoff, when GPS is available only after 300 seconds. The aircraft takes off with a stored heading alignment having a nominal azimuth uncertainty of 450 arcseconds. The aircraft executes a 1-g turn that produces a 45° heading change starting at t = 495 seconds. (The heading angle for Case 2 is essentially the same as for Case 1, but the turn is completed faster.)

Figure 4-2 illustrates the history of RMS alignment errors. The error is constant until t = 300 seconds when GPS becomes available. The availability of GPS makes a dramatic improvement in alignment even without maneuvers because the position and velocity errors that have accumulated in the INS navigation results are highly correlated with misalignments. We see that, following an initial decrease to 280 arcseconds, the alignment errors fall to an asymptote of 200 arcseconds which is comparable to gyrocompass alignment accuracy. The execution of 1-g lateral maneuver drives the alignment errors to the level predicted by accelerometer bias.

Case 3

Case 3 involves an "alert" takeoff with nominal stored heading azimuth errors and a 2-g "s" turn maneuver at t=460 seconds. GPS is not available until t=300 seconds (corresponding to a five-minute warm-up and acquisition time).

Figure 4-3 illustrates the rms alignment error. As expected, Figures 4-2 and 4-3 are identical until the maneuver begins, and then the higher g's involved in Case 3 reduce the alignment error below the level for Case 2.

4.2.3 Discussion

The simulations reported above support the following performance conclusions:

(1) Inflight GPS measurements during periods in which lateral acceleration is on the order of 1g can be used to reduce azimuthal alignment errors to the same magnitude as level errors. Larger accelerations lead to smaller errors.

(2) Even in the absence of lateral acceleration the availability of GPS measurements can be used to reduce the errors of a stored-heading alignment to the level of gyrocompassing. This reduction exploits the correlation between INS alignment errors and the INS navigation errors that are uncovered by comparison with GPS observations.

(3) Simple maneuvers such as 1 g coordinated turns over as little as 45°, and lasting for no more than 30 seconds are adequate for realizing the benefits of inflight alignment.
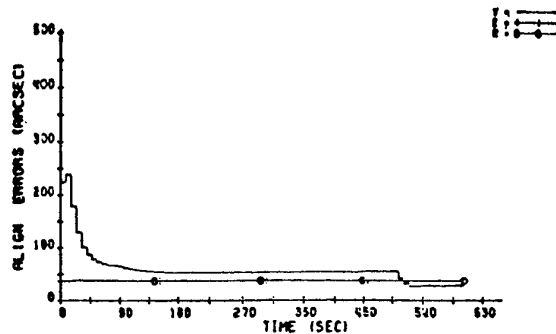
Figure 4-1. INS alignment errors with GPS available at takeoff and gyrocompassed alignment. (standard takeoff)
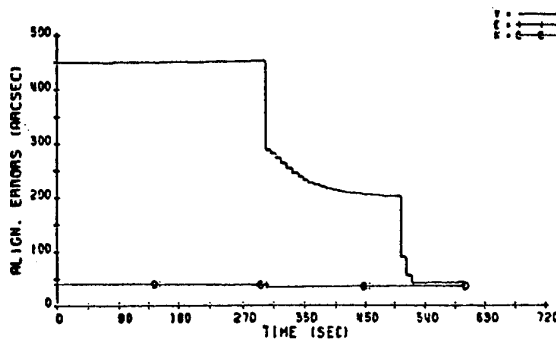
Figure 4-2. INS alignment errors with GPS available 5 minutes after an ALERT takeoff.
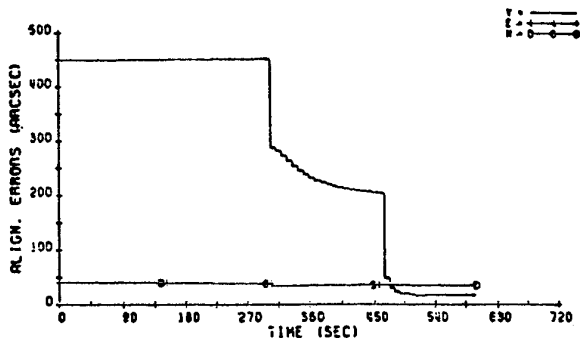
Figure 4-3. INS alignment errors with GPS available at 300 seconds after an alert takeoff with stored heading alignment.

4.2 Integrated Navigation Solutions During a GPS Outage

The previous examples illustrated certain benefits of GPS/INS integration in a benign environment, with low to moderate vehicle dynamics. The use of integration to lengthen the interval that GPS data is available, and to ride-out the period of outages caused by the interference or high dynamics is the focus of the following remarks.

4.2.1 Integration Case Study Overview

In a recent military aircraft avionics upgrade, GPS was to be integrated with a "navigation-grade" gimbaled-platform INS. The principal performance objective was to limit the error growth after five minutes of GPS outage, providing that the INS had been calibrated by at least seven minutes of GPS measurements prior to the outage. Based on extensive simulations, it was determined that:

1.   The horizontal CEP after five minutes of GPS outage could be held to 30 meters. (Recall that the free-running INS is specified as a 1.0 nmi/hr system.)

2.   A 21-state filter mechanized as two independent "horizontal" and "vertical" filters would provide close enough to optimum performance to be a cost-effective, computationally effective approach.

These results were derived from a covariance analysis that incorporated a 73-state truth model for the INS, plus additional error states for GPS.

Table 4-3 lists these truth states. Table 4-4 characterizes the mission segments of the standard aircraft flight path used in the simulation.

4.2.2 Navigation Performance of the Integrated System

Figure 4-4a illustrates the growth in rms "east" positioning navigation error versus time after the loss of GPS at approximately 2550 seconds. In this case, the GPS and INS position and velocity measurements were processed by a full optimal 73-state filter with dynamics, plant noise parameters and measurement noise parameters that exactly match those in the truth model. In other words, this represents the best that any filter can do given this truth model

and trajectory. Five minutes after the loss of GPS, the "east" component of horizontal position errors has grown to 80 feet (24.4 meters rms), exclusive of the low-frequency GPS bias. Figure 4-5 shows the contribution of various error sources to the net navigation error. For at least ten minutes, the accelerometer and gravity disturbance terms are dominant. Gyro errors do not become dominant until more than a quarter of a Schuler period (22 minutes) after the outage.

Figure 4-4b illustrates the growth in "east" errors for a sub-optimum 21-state filter, whose states are listed in Table 4-5. These results are almost indistinguishable from Figure 4-4a, for the "optimum filter." Finally, Figure 4-4c gives the "east" error for a 15-state "horizontal" filter whose states are listed in Table 4-4. These results too are nearly indistinguishable from "optimum" for about the first 420 seconds of GPS outage.

Table 4-3. Truth Model States

State #	Description	Comment
1,2	Horizontal position error	
3	Wander angle error	
4,5,6	Platform misalignment	100 arcsec initial horizontal; 6 degree initial vertical
7,8,9	Velocity error	
10	Vertical position error	
11-13	Auxiliary baro-inertial states	Used in describing baro-inertial loop
14-16	Markov gyro bias	.002 degree/hr rms horizontal .005 degree/hr rms vertical; Correlation time = 5 minutes
17-18	Markov accelerometer bias	3 micro-g rms; 10 minute correlation time
19	Markov baro bias	100 feet rms
21-23	Gravity disturbance	35 micro-g rms; 20 nautical mile; correlation distance
24-26	Gyro bias	.01 degree/hr rms horizontal; .022 degree/hr rms vertical
27-29	Gyro scale factor error	0.0002 rms
30-32	Gyro misalignments about spin axes	3.3 arcsec rms
33-35	Remaining gyro misalignments	20 arcsec rms
36-41	Gyro g-sensitivity	.015 degree/hr/g rms
42-44	Gyro g-squared-sensitivity	.02 degree/hr/g^{2}rms
45-47	Accelerometer bias	150 micro-g rms
48-50	Acc. scale factor error	0.0002 rms
51-53	Acc. scale factor asymmetry	0.0001 rms
54-56	Acc. nonlinear scale factor asymmetry	10 micro-g/g^{2}rms
57-59	Accelerometer non linearity	10 micro-g/g^{2}rms
60-65	Accelerometer orthogonal g-squared sensitivity	30 micro-g/g^{2}rms
66-71	Accelerometer misalignments	20 arcsecond 4ms
72	Baroaltimeter bias	300 feet rms
73	Baroaltimeter scale factor error	.04 rms

Table 4-4. Test Trajectory

Time (sec)	Description
0-300	Gyrocompass at true heading = 120 degrees.
300-490	Taxi and turn left approximately 75 degrees to prepare for takeoff at a heading of approximately 45 degrees.
490-540	Takeoff. Speed = 434 f/s, altitude approximately 1000 feet.
540-840	Climb, accelerate to 655 f/s, altitude approximately 9660 feet.
840-1020	Level off at approximately 9660 ft. Cruise at 655 f/s.
1020-1100	Descend to approximately 200 feet.
1100-2720	Level off at 200 feet, cruise at 655 f/s.
2720-2780	90 degree right turn.
2780-2806	Pop-up, climb to approximately 12,000 feet, accelerate to approximately 820 f/s.
2806-2838	Dive to 635 feet.
2838-2858	Level out at 635 feet, decelerate to 655 f/s. Weapons delivery.
2858-2918	90 degree right turn.
2918-3226	Low altitude combat, 300 to 2700 feet, 485 to 655 f/s.
3226-3276	Turning climb to 9900 feet, speed = 655 f/s.
3276-3346	High altitude combat.
3346-5086	Turn towards home. High altitude cruise at 9900 feet, 655 f/s.
5086-5446	Descend to 3000 feet, accelerate to 820 f/s.
5446-5876	Loiter, decelerate to 434 f/s.
5876-6126	Land.
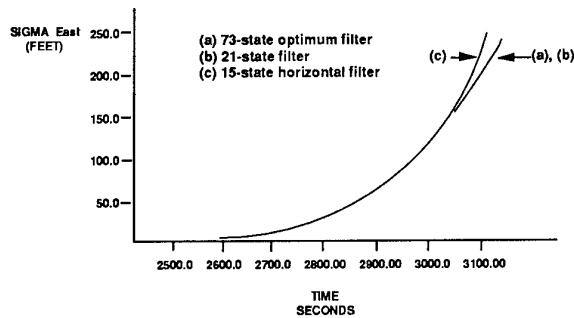
Figure 4-4. RMS East Position Error vs. Time: GPS measurements Stop at 2550 seconds.
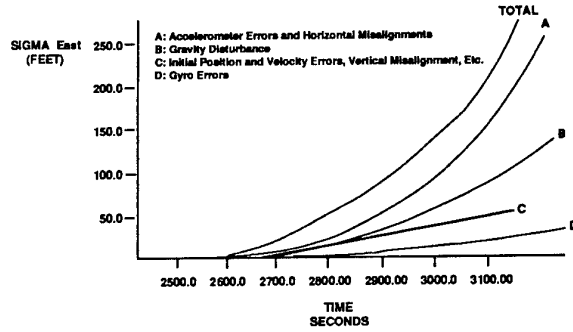
Figure 4-5. Contribution to INS Error Growth During a GPS Outage.

Table 4-5. States of the 15-state Horizontal Filter

State #	Description	Comment
1,2	Horizontal position error	
3	Wander angle error*	
4,5,6	Platform misalignment	Plant noise of 9.4E-17 rad^2/sec (horizontal) and 5.9E-16 rad^2/sec (vertical) used to account for unmodelled gyro errors
7,8	Horizontal velocity error	Plant noise of 9.4E-12 ft^2/sec^3 used to account for unmodelled accelerometer errors
9-11	Gyro bias	
12,13	Horizontal gravity disturbance (also absorbs other accelerometer errors)	
14,15	Accelerometer scale factor error	

* In an actual implementation, wander angle error would be combined into a single state with vertical misalignment.

5.0 CONCLUSIONS

We have presented a description of the processes whereby the combination of GPS with an inertial navigator can produce a system performance that is superior to either one acting alone. These benefits include:

1. Smaller random errors than seen in stand-alone GPS navigation solutions.

2. Improved availability of GPS operations during maneuvers and in the presence of radio frequency interference (RFI).

3. A navigation solution whose position and velocity errors are bounded by the errors in the GPS navigation solution.

4. A calibrated navigation solution whose errors grow slower than those of a free-running uncalibrated INS during GPS outages.

They are available because the long-term (low frequency) content of INS errors is negligible.

All four benefits are available to GPS users who are authorized to obtain the GPS Precise Positioning Service (PPS). Only the first and third are guaranteed to GPS users who are vulnerable to the selective availability clock dither that corrupts the GPS Standard Positioning Service (SPS). Clock dither prevents the tracking loop bandwidths from being reduced to the bandwidth of the INS errors, and corrupts the use of GPS measurements to calibrate the INS error states.

GPS receiver technology is evolving rapidly in response to pressure from the civilian market. This trend is evident in the miniaturization of full-capability receivers that can be physically embedded in a host system such as an INS or a mission computer. This in turn makes it practical to obtain even higher performance levels by treating both GPS and the INS as sensors that produce measurements to be optimally combined by a navigation filter into an "optimum" navigation solution.

For many users, the primary benefit of embedded configurations is not improved performance. These users focus on savings in volume, weight, power consumption, and cost that are predicted for the embedded system. Some cost savings are non-recurring (e.g. the one-time investment in developing integration software may be borne by the vendor of the embedded system rather than by the integrator), whereas others are recurring (e.g. reduced production cost of the system hardware).

In addition to these cost and performance benefits, the integrated system may be able to support functional capabilities that were previously not available to the user. For example, the use of in-flight INS alignment can make it possible for an aircraft to take off without waiting for the 5-10 minutes routinely reserved for INS alignment. This may be a life-saving capability for military aircraft that have to take off immediately after an alert that hostile forces are incoming. Another example is the inclusion of error states for gravity anomalies in the integration filter. This is the basis for balloon-borne GPS/INS instrumentation that will allow for more extensive gravity mapping than has ever been possible. (Ref. 38)

The design of a Kalman integration filter that mechanizes the navigation solution must also address certain standard questions such as:

- How many states to estimate?

- How often should the filter be updated?

- How should correlated measurements be treated?

- By how much can the computational burden of the filter be reduced by exploiting sparseness of the state's dynamics matrix or the measurement covariance matrix?

- How should the filter accommodate transient events such as changes in the constellation of satellites being tracked?

- How should the filter be tuned to provide robust performance versus unknown aspects of the design problem?

- How can the filter be made robust against variations in the error characteristics of individual INS or GPS units?

- How can the filter be used to detect the onset of anomalous conditions that may indicate a failure in the GPS or the INS sub-systems?

These design questions are addressed in more detail in other chapters of this book.

REFERENCES

1. Brown, R. Grover, "A Baseline GPS RAIM Scheme," NAVIGATION, Vol. 39, No. 3, Fall 1992, pp. 301-314.

2. Van Graas, Frank, "Sole Means Navigation and Integrity Through Hybrid Loran-C and NAVSTAR GPS," Proc. ION National Technical Meeting, Santa Barbara, CA, January 26-29, 1988, pp. 294-301.

3. Brown, A.K., Schmid, T., "Integrity Monitoring of GPS Using a Barometric Altimeter," Proc. ION National Technical Meeting, Santa Barbara, CA, January 26-29, 1988, pp. 245-253.

4. Tsang, Wai L., and Stein, Barry A., "A Robust RAIM Scheme Using GPS/GLONASS Systems," Proc. ION GPS-89, Colorado Springs, CO, September 27-29, 1989, pp. 373-378.

5. Cox, D. B., Integration of GPS with Inertial Navigation Systems, Reprinted in Collected GPS Papers Volume I, Institute of Navigation, Alexandria, VA.

6. Johannessen, R., Asbury M. J. A., Towards A Quantitative Assessment of Benefits INS/GPS Integration can offer to Civil Aviation, NAVIGATION, Vol. 37, No. 4, Winter 1990-1991, pp. 329-346.

7. Greenspan, R. L., et al, The GPS Users Integration Guide, Proceedings, ION National Technical Meeting, Santa Barbara, CA, 26-29 January 1988, pp. 104-112.

8. Wiederholt, L., Klein, D., Phase III GPS Options for Aircraft Platforms, NAVIGATION, Vol. 31, No. 2, Summer 1984, pp. 129-151.

9. Brown, A. K., Bowles, W. M., Interferometric Attitude Determination Using GPS, Proceedings, Third Geodetic Symposium on Satellite Doppler Positioning, Las Cruces, NM, February 1982, pp. 1289-1302.

10. Eller, D., GPS/IMU Navigation in a High Dynamics Environment, Proceedings, First International Symposium on Precise Positioning with GPS, April 15-19, 1985, Rockville, MD, pp. 773-782.

11. Bridges, P. D., Influence of Satellite Geometry Range, Clock and Altimeter Errors, on Two-Satellite GPS Navigation, Proceedings, ION GPS '88, Colorado Springs, CO, September 19-23, 1988, pp. 139-146.

294

12. Kalafus, R. M., and Knable, N., Clock Coasting and Altimeter Error Analysis for GPS, NAVIGATION, Vol. 31, No. 4, Winter 1984-85, pp. 289-302.

13. Bartholomew, R. G., et al, "Software Architecture of the Family of DoD Standard GPS Receivers." Proceedings First Technical Meeting of ION Satellite Division, Colorado Springs, September 1987, pp. 23-24.

14. Dellicker, S. H. and Henckel, D., "F-16/GPS Integration Test," Proceedings of ION GPS-89, Colorado Springs, September, 1989, pp. 295-303.

15. Barbour, Neil, et al, "Inertial Instruments - Where to Now 7," Proceedings AIAA GN&C Conference, Hilton Head, S.C., 10-12 August 1992, AIAA-92-4414-CP.

16. Boxenhorn, B., et al, "The Micromechanical Inertial Guidance System and it's Application," Proceedings Fourteenth Biennial Guidance Test Symposium, Holloman AFB, New Mexico, 3-5 October 1989, pp. 97-112.

17. Ward, P., Rath, J., Attitude Estimation Using GPS, Proceedings, ION National Technical Conference, San Mateo, CA, 23-26 January 1984, pp. 169-178.

18. Kruczynski, L. R., and Li, P. C., Using GPS to Determine Vehicle Attitude, Proceedings, ION GPS '88, Colorado Springs, CO, September 19-23, 1988, pp. 139-146.

19. Jurgens, Richard, Real Time GPS Azimuth Determining System, Proceedings, ION National Technical Conference, San Diego, CA, 23-25 January 1990, pp. 105-110.

20. van Graas, F., and Braasch, M., GPS Interferometric Attitude and Heading Determination; Flight Test Results, Proceedings, 47th Annual Meeting (ION) Williamsburg, VA, June 10-12, 1991, pp. 183-191.

21. Graham, W. R. and Johnston, G. R., Standard Integration Filter State Specification and Accuracy Predictions, NAVIGATION, Vol. 33, No. 4, Winter 1986-1987, pp. 295-313.

22. Cunningham, J., and Lewantowicz, Zdzislaw, H., Dynamic Integration of Separate INS/GPS Kalman Filters, Proceedings, ION GPS '88, Colorado Springs, CO, September 19-23, 1988, pp. 139-146.

23. Diesel, John W., Integration of GPS/INS for Maximum Velocity Accuracy, NAVIGATION, Vol. 34, No. 4, Fall 1987.

24. Berman, G. J., and Belzer, M., "A Decentralized Square Root Information Filter/Smoother," Proceedings 24th IEEE Conference on Decision and Control, Ft. Lauderdale, Florida, December 1985.

25. Carlson, N. A., "Federated Square Filter for Decentralized Parallel Processes, IEEE Trans. Aerospace and Electronic Systems, Vol. 26, No. 3, May 1990, pp. 517-525.

26. Gelb, A., (ed) Applied Optimal Estimation, MIT Press, Cambridge, MA (1974).

27. Maybeck, P. S., Stochastic Models, Estimation and Control, (Volume 1) Academic Press, Orlando, FL, 1979.

28. Bletzacker, F., et al., Kalman Filter Design for Integration of Phase III GPS with an Inertial Navigation System, Proceedings, ION National Technical Meeting, Long Beach, CA, 21-23 January 1986, pp. 113-129.

29. Lewantowicz, Z. H., and Keen, D. W., Graceful Degradation of GPS/INS Performance with Fewer than Four Satellites, Proceedings, ION National Technical Meeting, Phoenix, AZ, January 21-24, 1991, pp. 269-276.

30. Widnall, W. S., Stability of Alternate Designs for Rate Aiding of a Non-coherent Mode of a GPS Receiver, Intermetrics Corp., Cambridge, MA, 25 September 1978.

31. Buechler, D., and Foss, M., Integration of GPS and Strapdown Inertial Subsystems into a Single Unit, NAVIGATION, Summer 1987, Vol. 34, No. 2, pp. 140-159.

32. Franklin, M., Pagnucco, S., Development of Small Embedded GPS/INS RLG and FOG systems, Proceedings, ION National Technical Conference, San Diego, CA, 27-29 January 1992, pp. 3-12.

33. Tazartes, D. A., Mark, J. G., Integration of GPS Receivers into Existing Inertial Navigation Systems, NAVIGATION, Spring 1988, Vol. 35, No. 1, pp. 105-120.

34. Homer, W. C., An Introduction to the GPS Guidance Package (GGP) Proceedings, 15th Biennial Guidance Test Symposium, CIGTF Guidance Test Division, Vol. 1, pp. 12-15, Holloman AFB, New Mexico, September 24-26, 1991.

35. Donna, J., and Greenspan, R., Measurement Errors in GPS Observable, NAVIGATION, Vol. 33, No. 4, Winter 1986-87, pp. 319-334.

36. Soltz, J. A., et al., An Option for Mechanizing Integrated GPS/INS Solutions, NAVIGATION, Vol. 35, No. 4, Winter 1988-89, pp. 443-458.

37. "Specification for USAF Standard Form, Fit and Function (F^3) Medium Accuracy Inertial Navigation Unit," SNU-84-1, Rev. 3, Amendment, 13 February 1987, USAF Aeronautical Systems Division (now Aeronautical Systems Center).

38. Jekeli, C., et al., Instrumentation Design and Analysis for Balloon-Borne Gravimetry and Attitude Determination Using GPS and INS, Proceedings ION National Technical Meeting, San Francisco, CA, 20-22 January 1993 (To Be Published).

Design Considerations for a Suboptimal Kalman Filter

D.J.DiFilippo
Defence Research Establishment Ottawa
Ottawa, Canada
K1A 0Z4

1. INTRODUCTION

In designing a suboptimal Kalman filter, the designer must decide how to simplify the system error model without causing the filter estimation errors to increase to unacceptable levels. Deletion of certain error states and decoupling of error state dynamics are the two principal model simplifications that are commonly used in suboptimal filter design. For the most part, the decisions as to which error states can be deleted or decoupled are based on the designer's understanding of the physics of the particular system. Consequently, the details of a suboptimal design are usually unique to the specific application.

In this paper, the process of designing a suboptimal Kalman filter is illustrated for the case of an airborne transfer-of-alignment (TOA) system used for synthetic aperture radar (SAR) motion compensation. In this application, the filter must continuously transfer the alignment of an onboard Doppler-damped master inertial navigation system (INS) to a strapdown navigator that processes information from a less accurate inertial measurement unit (IMU) mounted on the radar antenna. The IMU is used to measure spurious antenna motion during the SAR imaging interval, so that compensating phase corrections can be computed and applied to the radar returns, thereby preventing image degradation that would otherwise result from such motions. The principles of SAR are described in many references, for instance [1], [2].

The primary function of the TOA Kalman filter in a SAR motion compensation system is to control strapdown navigator attitude errors, and to a lesser degree, velocity and heading errors. Unlike a classical navigation application, absolute positional accuracy is not important. The motion compensation requirements for SAR imaging are discussed in some detail in [3]. This TOA application is particularly appropriate as a vehicle for discussing suboptimal filter design, because the system contains features that can be exploited to allow both deletion and decoupling of error states.

In Section 2, a high-level background description of a SAR motion compensation system that incorporates a TOA Kalman filter is given. The optimal TOA filter design is presented in Section 3 with some simulation results to indicate potential filter performance. In Section 4, the suboptimal Kalman filter configuration is derived. Simulation results are also shown in this section to allow comparison between suboptimal and optimal filter performances. Conclusions are contained in Section 5.

Throughout this paper, it is assumed that the reader is familiar with basic Kalman filter theory. If not, a treatment of this subject can be found in [4], [5] or [6].

2. SAR MOTION COMPENSATION SYSTEM DESCRIPTION

2.1 Sensors

A functional block diagram of a SAR motion compensation system (SARMCS) that utilizes a TOA Kalman filter is shown in Figure 1.
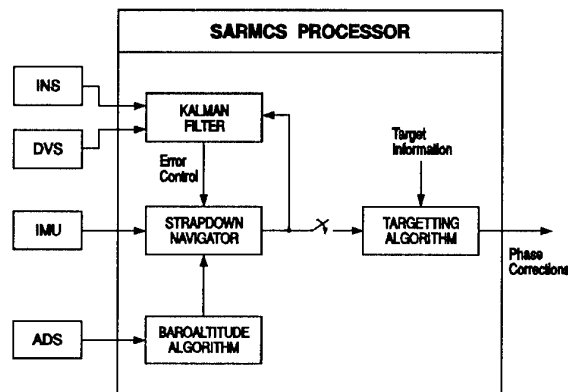
Figure 1. Functional Block Diagram of a SARMCS

The principal motion compensation sensor is a small inertial measurement unit (IMU) mounted directly on the antenna structure. It provides raw measurements of the angular rates and linear accelerations of the antenna in the standard form of angular and velocity increments. The antenna structure is assumed to be physically stabilized in roll and pitch. Consequently, the strapdown IMU is exposed to only residual angular motion about the aircraft's roll and pitch axes resulting from imperfect antenna stabilization, while it is fully exposed to angular motion caused by aircraft heading changes. It is assumed that the IMU contains two dry tuned-rotor gyroscopes and three accelerometers with the specifications shown in Table 1.

The onboard master inertial navigation system (INS) is remotely located from the IMU. It is assumed to be a medium-accuracy system (1 nm/hr, 1σ) with a baroaltitude-damped vertical channel. It employs three ring laser gyroscopes and three accelerometers, which have the characteristics indicated in Table 2.

The Doppler velocity sensor (DVS) employs a standard strapdown lambda 3-beam configuration [7] to measure aircraft velocity in aircraft body coordinates relative to the terrain surface. The primary time-correlated error in the measurement of the forward velocity component is a scale factor type error in the order of 1-2% (1σ), depending upon the terrain. The error in measuring the lateral velocity component principally results from an antenna boresight azimuth misalignment which is typically less than 1° (1σ). In addition to these time-correlated errors, there is also a

significant noise component to the DVS measurements of velocity; the level of noise is typically 1 m/s (1σ) for the forward velocity measurement, and 1.5 m/s (1σ) for the lateral velocity measurement. It is assumed that the DVS is only operated over land, so that additional errors introduced by the effects of sea currents and waves can be ignored.

The *air data sensors* (ADS) include a pressure transducer to measure static air pressure, and an air temperature probe to measure outside air temperature. This information is used to compute baroaltitude.

IMU Gyro Performance Characteristics	
Characteristic	Performance
Scale factor repeatability Bias repeatability G-sensitive drift repeatability	150 ppm (1σ) 0.2 °/hr (1σ) 0.2 °/hr/g (1σ)
IMU Accelerometer Performance Characteristics	
Characteristic	Performance
Scale factor repeatability Bias repeatability	200 ppm (1σ) 100 μg (1σ)

Table 1. IMU Gyro and Accelerometer Performance Characteristics.

INS Gyro Performance Characteristics	
Characteristic	Performance
Scale factor repeatability Bias repeatability Random drift	5 ppm (1σ) 0.01 °/hr (1σ) 0.003°/√ hr (1σ)
INS Accelerometer Performance Characteristics	
Characteristic	Performance
Scale factor repeatability Bias repeatability	50 ppm (1σ) 50 μg (1σ)

Table 2. INS Gyro and Accelerometer Performance Characteristics.

2.2 Processing Overview

Velocity and angular increments from the IMU are processed in a strapdown navigator algorithm which mechanizes the navigation equations in a wander azimuth frame. Since the IMU is mounted directly on the antenna, the strapdown algorithm fundamentally computes antenna attitude and heading, and velocity and position of a particular point at the IMU location. A lever arm correction is applied within the navigator to obtain estimates of the velocity and position of the antenna phase centre which is

the point of interest for SAR motion compensation. The computed baroaltitude is used in a classical third-order damping loop to stabilize the strapdown vertical channel.

The strapdown navigator estimates of antenna phase centre motion are provided to a targetting algorithm during intervals of SAR imaging. The targetting algorithm initially computes a line-of-sight vector from the antenna phase centre to the target, using target information supplied from other parts of the SAR system. This line-of-sight vector is then updated throughout the imaging interval. At each point in time, the targetting algorithm calculates the component of the antenna phase centre velocity lying along the line-of-sight vector, and integrates this velocity component in time to obtain phase centre displacement along the line-of-sight. From this displacement, an appropriate value of phase correction is computed for each point in time, and applied to the radar signal returns to compensate for the spurious antenna motion.

The outputs of the strapdown navigator, along with information from the master INS and DVS, are supplied to a Kalman filter which estimates various system and sensor errors associated with these devices. The filter-estimated strapdown navigation errors and IMU instrument errors are fed back to the strapdown navigator algorithm and used there to correct the relevant quantities in a closed loop fashion. The net result of this Kalman filter implementation is that alignment of a Doppler-damped INS is continually transferred to the strapdown navigator.

3. OPTIMAL TOA KALMAN FILTER DESIGN

3.1 State Vector

The state vector in the optimal TOA Kalman filter design contains 26 states, representing all significant time-correlated errors. The state vector can be expressed as

$$x = \begin{bmatrix} x_M \\ x_{INS} \\ x_{DVS} \\ x_S \\ x_{IMU} \end{bmatrix} , \tag{1}$$

where the subvectors x_M and x_S contain master INS system error states and strapdown navigator system error states respectively, x_{INS} and x_{IMU} contain states that represent INS and IMU instrument errors respectively, and x_{DVS} contains augmenting states that represent errors in the DVS. These 26 states are listed in Table 3. The error states modelled in x_M and x_S are defined as system-indicated values minus true values, while the instrument error quantities modelled in x_{INS}, x_{IMU} and x_{DVS} are defined as factory-calibrated values minus true values.

3.2 Initial Covariance Matrix

The initial covariance matrix P_0 for the state vector is expressed in terms of the initial estimation error $\tilde{x}_0$ as

$$P_0 = E[\tilde{x}_0 \tilde{x}_0^T] , \tag{2}$$

where "E" denotes expectation value, the superscript "T" indicates matrix transpose, and the initial estimation error is defined by

$$\tilde{x}_0 \equiv \hat{x}_0 - x_0 \quad , \tag{3}$$

with $\hat{x}_0$ denoting the estimated value of x_0, the true value of the initial state vector. The initial covariance matrix for this TOA system is expressed in terms of submatrices as

$$P_0 = \begin{bmatrix} P_{M_0} & 0 & 0 & P_{M/S_0} & 0 \\ 0 & P_{INS_0} & 0 & 0 & 0 \\ 0 & 0 & P_{DVS_0} & 0 & 0 \\ P_{M/S_0}^T & 0 & 0 & P_{S_0} & 0 \\ 0 & 0 & 0 & 0 & P_{IMU_0} \end{bmatrix} \quad , \tag{4}$$

where

$$P_{M_0} = E[\tilde{x}_{M_0} \tilde{x}_{M_0}^T] \quad , \tag{5}$$

$$P_{INS_0} = E[\tilde{x}_{INS_0} \tilde{x}_{INS_0}^T] \quad , \tag{6}$$

$$P_{S_0} = E[\tilde{x}_{S_0} \tilde{x}_{S_0}^T] \quad , \tag{7}$$

$$P_{IMU_0} = E[\tilde{x}_{IMU_0} \tilde{x}_{IMU_0}^T] \quad , \tag{8}$$

$$P_{DVS_0} = E[\tilde{x}_{DVS_0} \tilde{x}_{DVS_0}^T] \quad , \tag{9}$$

$$P_{M/S_0} = E[\tilde{x}_{M_0} \tilde{x}_{S_0}^T] \quad . \tag{10}$$

Subvector	State	Description	Coordinate Frame
x_M	δR_{M_X} δR_{M_Y} δV_{M_X} δV_{M_Y} ϕ_{M_X} ϕ_{M_Y} ϕ_{M_Z}	master INS position error along X axis master INS position error along Y axis master INS velocity error along X axis master INS velocity error along Y axis master INS platform misalignment about X axis master INS platform misalignment about Y axis master INS platform misalignment about Z axis	Wander Azimuth[1]
x_{INS}	A_{M_x} A_{M_y} G_{M_x} G_{M_y} G_{M_z}	master INS x accelerometer bias master INS y accelerometer bias master INS x gyro bias master INS y gyro bias master INS z gyro bias	Aircraft Body[2]
x_{DVS}	δS γ	DVS forward scale factor error DVS azimuth boresight error	Aircraft Body
x_S	δR_{S_X} δR_{S_Y} δV_{S_X} δV_{S_Y} ϕ_{S_X} ϕ_{S_Y} ϕ_{S_Z}	strapdown navigator position error along X axis strapdown navigator position error along Y axis strapdown navigator velocity error along X axis strapdown navigator velocity error along Y axis strapdown navigator platform misalignment about X axis strapdown navigator platform misalignment about Y axis strapdown navigator platform misalignment about Z axis	Wander Azimuth
x_{IMU}	A_{S_x} A_{S_y} G_{S_x} G_{S_y} G_{S_z}	IMU x accelerometer bias IMU y accelerometer bias IMU x gyro bias IMU y gyro bias IMU z gyro bias	IMU Body[3]
[1] Wander azimuth frame: X,Y axes level, Z axis up [2] Aircraft body frame: x axis forward, y axis out right wing, z axis down [3] IMU body frame: x axis forward, z axis down, y axis oriented to form right-handed coordinate system			

Table 3. Description of States for the Optimal TOA Kalman Filter.

The submatrix $\mathbf{P}_{\mathbf{M}_0}$ is given by

$$
\mathbf{P}_{\mathbf{M}_0} =
\begin{bmatrix}
\sigma^2_{MR} & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & \sigma^2_{MR} & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & \sigma^2_{MV} & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & \sigma^2_{MV} & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & \sigma^2_{M\phi} & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & \sigma^2_{M\phi} & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & \sigma^2_{M\phi_z}
\end{bmatrix} , \quad (11)
$$

where the variables are defined as follows:

σ^2_{MR} - the variance of the error in the initial estimates of δR_{M_X} and δR_{M_Y},

σ^2_{MV} - the variance of the error in the initial estimates of δV_{M_X} and δV_{M_Y},

$\sigma^2_{M\phi}$ - the variance of the error in the initial estimates of ϕ_{M_X} and ϕ_{M_Y},

$\sigma^2_{M\phi_Z}$ - the variance of the error in the initial estimate of ϕ_{M_Z}.

In this system, position, velocity, wander angle and heading in the IMU strapdown navigator are initialized to master INS values for these quantities. This implies that at filter initialization, the following relationships between estimation errors are true:

$$
\begin{aligned}
\delta \tilde{R}_{S_i} &= \delta \tilde{R}_{M_i} , \quad i = X,Y \\
\delta \tilde{V}_{S_i} &= \delta \tilde{V}_{M_i} , \quad i = X,Y \\
\tilde{\phi}_{S_Z} &= \tilde{\phi}_{M_Z} .
\end{aligned}
\quad (12)
$$

Strapdown navigator roll and pitch are initialized independently from the master INS values because the IMU is mounted on the stabilized antenna while the master INS is mounted on the airframe. This last point, together with the equations in Eq.(12), results in the following expressions for $\mathbf{P}_{\mathbf{S}_0}$ and $\mathbf{P}_{\mathbf{M/S}_0}$:

$$
\mathbf{P}_{\mathbf{S}_0} =
\begin{bmatrix}
\sigma^2_{MR} & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & \sigma^2_{MR} & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & \sigma^2_{MV} & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & \sigma^2_{MV} & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & \sigma^2_{S\phi} & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & \sigma^2_{S\phi} & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & \sigma^2_{M\phi_z}
\end{bmatrix} , \quad (13)
$$

$$
\mathbf{P}_{\mathbf{M/S}_0} =
\begin{bmatrix}
\sigma^2_{MR} & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & \sigma^2_{MR} & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & \sigma^2_{MV} & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & \sigma^2_{MV} & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & \sigma^2_{M\phi_z}
\end{bmatrix} . \quad (14)
$$

In Eq.(13), $\sigma^2_{S\phi}$ is defined as the variance of the error in the initial estimates of ϕ_{S_X} and ϕ_{S_Y}.

The submatrices $\mathbf{P}_{\mathbf{INS}_0}$, $\mathbf{P}_{\mathbf{IMU}_0}$ and $\mathbf{P}_{\mathbf{DVS}_0}$ are expressed as

$$
\mathbf{P}_{\mathbf{INS}_0} =
\begin{bmatrix}
\sigma^2_{MAB} & 0 & 0 & 0 & 0 \\
0 & \sigma^2_{MAB} & 0 & 0 & 0 \\
0 & 0 & \sigma^2_{MGB} & 0 & 0 \\
0 & 0 & 0 & \sigma^2_{MGB} & 0 \\
0 & 0 & 0 & 0 & \sigma^2_{MGB}
\end{bmatrix} , \quad (15)
$$

$$
\mathbf{P}_{\mathbf{IMU}_0} =
\begin{bmatrix}
\sigma^2_{SAB} & 0 & 0 & 0 & 0 \\
0 & \sigma^2_{SAB} & 0 & 0 & 0 \\
0 & 0 & \sigma^2_{SGB} & 0 & 0 \\
0 & 0 & 0 & \sigma^2_{SGB} & 0 \\
0 & 0 & 0 & 0 & \sigma^2_{SGB}
\end{bmatrix} , \quad (16)
$$

$$
\mathbf{P}_{\mathbf{DVS}_0} =
\begin{bmatrix}
\sigma^2_{DSF} & 0 \\
0 & \sigma^2_{DB}
\end{bmatrix} . \quad (17)
$$

In the above expressions, the variables are defined as:

σ^2_{MAB} - variance of the master INS accelerometer bias, as specified in Table 2,

σ^2_{MGB} - variance of the master INS gyro bias, as specified in Table 2,

σ^2_{SAB} - variance of the strapdown IMU accelerometer bias, as specified in Table 1,

σ^2_{SGB} - variance of the strapdown IMU gyro bias, as specified in Table 1,

σ^2_{DSF} - variance of the Doppler forward scale factor error, as specified in Section 2.1,

σ^2_{DB} - variance of the Doppler antenna azimuth boresight misalignment, as specified in Section 2.1.

3.3 State Dynamics

The continuous model for the error state dynamics has the standard form applicable to linear systems:

$$\dot{x}(t) = F(t)x(t) + w(t) \quad . \tag{18}$$

In the above equation, $w(t)$ is a 26 element vector of zero-mean white noise processes. This noise vector has the same structure as the state vector, namely

$$w = \begin{bmatrix} w_M \\ w_{INS} \\ w_{DVS} \\ w_S \\ w_{IMU} \end{bmatrix}, \tag{19}$$

and $F(t)$ is a 26 by 26 element matrix with the following structure:

$$F = \begin{bmatrix} F_M & F_{M/INS} & 0 & 0 & 0 \\ 0 & F_{INS} & 0 & 0 & 0 \\ 0 & 0 & F_{DVS} & 0 & 0 \\ 0 & 0 & 0 & F_S & F_{S/IMU} \\ 0 & 0 & 0 & 0 & F_{IMU} \end{bmatrix}, \tag{20}$$

where F_M and F_S are 7×7 matrices, $F_{M/INS}$ is a 7×5 matrix, $F_{S/IMU}$ is a 7×5 matrix, F_{INS} is a 5×5 matrix, F_{IMU} is a 5×5 matrix, and F_{DVS} is a 2×2 matrix. Using the expressions from Eqs.(1), (19) and (20) in Eq.(18), the system dynamics can be expressed in terms of the state subvectors as

$$\dot{x}_M = F_M x_M + F_{M/INS} x_{INS} + w_M \quad , \tag{21}$$

$$\dot{x}_{INS} = F_{INS} x_{INS} + w_{INS} \quad , \tag{22}$$

$$\dot{x}_{DVS} = F_{DVS} x_{DVS} + w_{DVS} \quad , \tag{23}$$

$$\dot{x}_S = F_S x_S + F_{S/IMU} x_{IMU} + w_S \quad , \tag{24}$$

$$\dot{x}_{IMU} = F_{IMU} x_{IMU} + w_{IMU} \quad . \tag{25}$$

The elements of F_M correspond to a true frame inertial error model [8] mechanized for a wander azimuth coordinate system that has its origin at the master INS, and that has X, Y, Z axes and wander angle α as depicted in Figure 2.
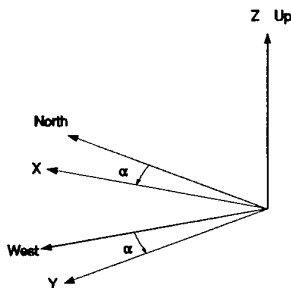
Figure 2. Orientation of Wander Azimuth Frame

This yields

$$F_M = \begin{bmatrix} \dfrac{V_Z}{R_X} & 0 & 1 & 0 & 0 & 0 & 0 \\[2mm] 0 & \dfrac{V_Z}{R_Y} & 0 & 1 & 0 & 0 & 0 \\[2mm] F_{31} & F_{32} & -\dfrac{V_Z}{R_X} & 2\omega_{IE_Z} & 0 & -f_Z & f_Y \\[2mm] F_{41} & F_{42} & -2\omega_{IE_Z} & -\dfrac{V_Z}{R_Y} & f_Z & 0 & -f_X \\[2mm] -\dfrac{\omega_{IE_Z}}{R_X} & 0 & 0 & -\dfrac{1}{R_Y} & 0 & \omega_{IE_Z} & -\omega_{IW_Y} \\[2mm] 0 & -\dfrac{\omega_{IE_Z}}{R_Y} & \dfrac{1}{R_X} & 0 & -\omega_{IE_Z} & 0 & \omega_{IW_X} \\[2mm] \dfrac{\omega_{IW_X}}{R_X} & \dfrac{\omega_{IW_Y}}{R_Y} & 0 & 0 & \omega_{IW_Y} & -\omega_{IW_X} & 0 \end{bmatrix} \tag{26}$$

with

$$F_{31} = \frac{V_Y}{R_X}(\omega_{IE_Y} + \omega_{IW_Y}) \quad ,$$

$$F_{32} = \frac{V_Y}{R_Y}(\omega_{IE_Y} + \omega_{IW_Y}) + 2\frac{V_Z}{R_y}\omega_{IE_Z} \quad ,$$

$$F_{41} = -\frac{V_X}{R_X}(\omega_{IE_Y} + \omega_{IW_X}) - 2\frac{V_Z}{R_X}\omega_{IE_Z} \quad ,$$

$$F_{42} = -\frac{V_X}{R_Y}(\omega_{IE_Y} + \omega_{IW_Y}) \quad ,$$

$$R_X = r_X + h \quad ,$$

$$R_Y = r_Y + h \quad .$$

In the above equation, the variables are defined as follows:

V_X, V_Y, V_Z - the X, Y, and Z components of V, the master INS velocity vector,

f_X, f_Y, f_Z - the X, Y, and Z components of f, the master INS specific force vector,

$\omega_{IE_X}, \omega_{IE_Y}, \omega_{IE_Z}$ - the X, Y, and Z components of ω_{IE}, the angular velocity vector of the earth with respect to inertial space,

$\omega_{IW_X}, \omega_{IW_Y}, \omega_{IW_Z}$ - the X, Y, and Z components of ω_{IW}, the angular velocity vector of the wander azimuth coordinate frame with respect to inertial space,

h - altitude of the master INS above the earth reference ellipsoid,

r_X, r_Y - local radii of curvature of the earth reference ellipsoid along the X and Y axes respectively.

Vertical channel error states for the master INS are not modelled in the Kalman filter because vertical channel errors are controlled with a third order fixed-gain damping loop implemented internally in the INS.

The elements of F_S have exactly the same form as that expressed in Eq.(26) for F_M, except that the wander azimuth frame of interest has its origin at the strapdown IMU, and h, V, and f apply to the strapdown IMU rather than the master INS.

The submatrix $F_{M/INS}$ has the standard expression for a strapdown mechanization:

$$F_{M/INS} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ C_{11} & C_{12} & 0 & 0 & 0 \\ C_{21} & C_{22} & 0 & 0 & 0 \\ 0 & 0 & -C_{11} & -C_{12} & -C_{13} \\ 0 & 0 & -C_{21} & -C_{22} & -C_{23} \\ 0 & 0 & -C_{31} & -C_{32} & -C_{33} \end{bmatrix}, \quad (27)$$

where C_{ij} is the element in the ith row and jth column of the direction cosine matrix C_A^{MW} which expresses the transformation from the aircraft body frame to the master INS wander azimuth frame. The aircraft body frame is defined in Table 3. C_A^{MW} is computed as

$$C_A^{MW} = \begin{bmatrix} \cos A \cos P & \begin{matrix} -\sin A \cos R + \\ \cos A \sin P \sin R \end{matrix} & \begin{matrix} \sin A \sin R + \\ \cos A \sin P \cos R \end{matrix} \\ -\sin A \cos P & \begin{matrix} -\cos A \cos R - \\ \sin A \sin P \sin R \end{matrix} & \begin{matrix} \cos A \sin R - \\ \sin A \sin P \cos R \end{matrix} \\ \sin P & -\cos P \sin R & -\cos P \cos R \end{bmatrix} \quad (28)$$

with

$A = \alpha + H$,

and the variables defined as:

α - wander angle for the master INS wander angle frame,

R - aircraft roll,

P - aircraft pitch,

H - aircraft heading from North.

The submatrix $F_{S/IMU}$ has exactly the same form as $F_{M/INS}$ in Eq.(27) except that in the case of $F_{S/IMU}$, the C_{ij} are elements of the direction cosine matrix C_B^{SW} which describes the transformation from the strapdown IMU body frame to the wander azimuth frame centered at the IMU. The strapdown IMU body frame is described in Table 3. C_B^{SW} is given by the same expression as in Eq.(28) with roll, pitch and heading interpreted as being that of the IMU. In general, the roll and pitch of the strapdown IMU is close to

zero, since the IMU is mounted on the antenna which is physically stabilized in roll and pitch.

The dynamics of the error states in x_{INS}, x_{IMU}, and x_{DVS} are modelled as first order Markov processes, resulting in the following submatrices:

$$F_{INS} = \begin{bmatrix} -\beta_{MAB} & 0 & 0 & 0 & 0 \\ 0 & -\beta_{MAB} & 0 & 0 & 0 \\ 0 & 0 & -\beta_{MGB} & 0 & 0 \\ 0 & 0 & 0 & -\beta_{MGB} & 0 \\ 0 & 0 & 0 & 0 & -\beta_{MGB} \end{bmatrix}, \quad (29)$$

$$F_{IMU} = \begin{bmatrix} -\beta_{SAB} & 0 & 0 & 0 & 0 \\ 0 & -\beta_{SAB} & 0 & 0 & 0 \\ 0 & 0 & -\beta_{SGB} & 0 & 0 \\ 0 & 0 & 0 & -\beta_{SGB} & 0 \\ 0 & 0 & 0 & 0 & -\beta_{SGB} \end{bmatrix}, \quad (30)$$

$$F_{DVS} = \begin{bmatrix} -\beta_{DSF} & 0 \\ 0 & -\beta_{DB} \end{bmatrix} . \quad (31)$$

In Eqs.(29) to (31), the variables are defined as:

β_{MAB} - inverse of the correlation time for the master INS accelerometer bias error,

β_{MGB} - inverse of the correlation time for the master INS gyro bias error,

β_{SAB} - inverse of the correlation time for the strapdown IMU accelerometer bias error,

β_{SGB} - inverse of the correlation time for the strapdown IMU gyro bias error,

β_{DSF} - inverse of the correlation time for the DVS forward scale factor error,

β_{DB} - inverse of the correlation time for the DVS azimuth boresight error.

The vector of random forcing functions in Eq.(18), denoted by w, is described in terms of its covariance matrix which is given in the continuous formulation by

$$P_w = E[w(t)w(\tau)^T] = Q(t)\delta (t - \tau) , \quad (32)$$

where $\delta(t-\tau)$ is the Dirac delta function and $Q(t)$ is a spectral density matrix. For this system, Q is expressed in terms of submatrices as

$$Q = \begin{bmatrix} Q_M & 0 & 0 & Q_{M/S} & 0 \\ 0 & Q_{INS} & 0 & 0 & 0 \\ 0 & 0 & Q_{DVS} & 0 & 0 \\ Q_{M/S}^T & 0 & 0 & Q_S & 0 \\ 0 & 0 & 0 & 0 & Q_{IMU} \end{bmatrix} . \quad (33)$$

In the above equation, the submatrix $\mathbf{Q_M}$ is given by:

$$\mathbf{Q_M} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & q_{VD} + q_{MASF} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & q_{VD} + q_{MASF} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & q_{MGN} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & q_{MGN} & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & q_{MGN} \end{bmatrix}, \quad (34)$$

where q_{VD} and q_{MASF} are the spectral densities of white noise models to account for the effects of vertical deflections and INS accelerometer scale factor errors respectively, and q_{MGN} is the spectral density of the white noise associated with the random drift of the INS ring laser gyros, as specified in Table 2. The spectral densities q_{VD} and q_{MASF} are given by the expressions

$$q_{VD} = \frac{2d_{VD}\,g^2}{V_g}\sigma_{VD}^2 \quad , \quad (35)$$

$$q_{MASF} = \frac{4V_g\,a}{\pi}\sigma_{MASF}^2 \quad , \quad (36)$$

where the variables have the following definitions:

g - the nominal value of gravity,

d_{VD} - the correlation distance of the random deflections of the vertical,

σ_{VD}^2 - the variance of the deflections of the vertical,

V_g - aircraft ground speed,

a - magnitude of the aircraft acceleration vector,

σ_{MASF}^2 - the variance of the master INS accelerometer scale factor error, as indicated in Table 2.

A derivation for the expressions in Eqs.(35) and (36) is given in [9].

The submatrix $\mathbf{Q_S}$ is written as:

$$\mathbf{Q_S} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & q_{VD} + q_{SASF} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & q_{VD} + q_{SASF} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & q_{SGMU} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & q_{SGMU} & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & q_{SGMU} + q_{SGSF} \end{bmatrix}, \quad (37)$$

where q_{VD} is as defined previously, and q_{SASF}, q_{SGMU} and q_{SGSF} are the spectral densities of white noise models that account for the effects of strapdown IMU accelerometer scale factor errors, IMU gyro mass unbalances excited by aircraft accelerations, and the z gyro scale factor error respectively. Of course, gyro mass unbalances are also excited by the constant gravitational field, but this component of the mass unbalance effect is indistinguishable from a gyro bias effect and so is lumped in with the modelled gyro bias states. It should also be noted that the effects of the IMU x and y gyro scale factor errors are not modelled because in this case, these errors do not contribute to inertial platform misalignments; since the x and y gyro input axes are physically stabilized in the horizontal plane, they do not experience any significant angular velocities that would excite scale factor error effects. On the other hand, the z gyro input axis, which essentially points along the local vertical, does experience a significant angular velocity during an aircraft heading change, so that during a turn, the z gyro scale factor error causes a platform misalignment about the Z wander azimuth axis.

The spectral densities q_{SASF}, q_{SGMU} and q_{SGSF} are given by the expressions:

$$q_{SASF} = \frac{4V_g\,a}{\pi}\sigma_{SASF}^2 \quad , \quad (38)$$

$$q_{SGMU} = \frac{4V_g\,a}{\pi}\sigma_{SGMU}^2 \quad , \quad (39)$$

$$q_{SGSF} = \pi\,|\omega_z|\sigma_{SGSF}^2 \quad , \quad (40)$$

where the variables have the following definitions:

σ_{SASF}^2 - the variance of the IMU accelerometer scale factor error, as specified in Table 1,

σ_{SGMU}^2 - the variance of the strapdown IMU gyro mass unbalance, as specified in Table 1,

σ_{SGSF}^2 - the variance of the strapdown IMU gyro scale factor error, as specified in Table 1,

$|\omega_z|$ - the magnitude of the z IMU body axis component of the IMU angular rate vector.

Derivations for the expressions in Eqs.(38) to (40) are found in [9].

The submatrix $\mathbf{Q}_{M/S}$ is defined by the equation

$$\mathbf{Q}_{M/S}(t)\delta\,(t-\tau) = E[\mathbf{w}_M(t)\mathbf{w}_S(\tau)^T] \qquad (41)$$

and is given by

$$\mathbf{Q}_{M/S} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & q_{VD} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & q_{VD} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} . \qquad (42)$$

In the above expression, the two non-zero terms along the diagonal represent the correlation that exists between elements of $\mathbf{w}_M$ and $\mathbf{w}_S$ as a result of modelling vertical deflection effects as a white noise process that drives both the master INS and strapdown navigator velocity error states.

The submatrices $\mathbf{Q}_{INS}$, $\mathbf{Q}_{IMU}$, and $\mathbf{Q}_{DVS}$ are

$$\mathbf{Q}_{INS} = \begin{bmatrix} q_{MAB} & 0 & 0 & 0 & 0 \\ 0 & q_{MAB} & 0 & 0 & 0 \\ 0 & 0 & q_{MGB} & 0 & 0 \\ 0 & 0 & 0 & q_{MGB} & 0 \\ 0 & 0 & 0 & 0 & q_{MGB} \end{bmatrix} , \qquad (43)$$

$$\mathbf{Q}_{IMU} = \begin{bmatrix} q_{SAB} & 0 & 0 & 0 & 0 \\ 0 & q_{SAB} & 0 & 0 & 0 \\ 0 & 0 & q_{SGB} & 0 & 0 \\ 0 & 0 & 0 & q_{SGB} & 0 \\ 0 & 0 & 0 & 0 & q_{SGB} \end{bmatrix} , \qquad (44)$$

$$\mathbf{Q}_{DVS} = \begin{bmatrix} q_{DSF} & 0 \\ 0 & q_{DB} \end{bmatrix} . \qquad (45)$$

The spectral densities in the expressions above are computed from the following expressions:

$$q_{MAB} = 2\sigma_{MAB}^2\,\beta_{MAB} \quad , \qquad (46)$$

$$q_{MGB} = 2\sigma_{MGB}^2\,\beta_{MGB} \quad , \qquad (47)$$

$$q_{SAB} = 2\sigma_{SAB}^2\,\beta_{SAB} \quad , \qquad (48)$$

$$q_{SGB} = 2\sigma_{SGB}^2\,\beta_{SGB} \quad , \qquad (49)$$

$$q_{DSF} = 2\sigma_{DSF}^2\,\beta_{DSF} \quad , \qquad (50)$$

$$q_{DB} = 2\sigma_{DB}^2\,\beta_{DB} \quad , \qquad (51)$$

where all variables have been previously defined. The spectral density expressions in Eqs.(46) through to (51) represent the level of driving white noise for each Markov process model that results in the steady-state variances for the models being equal to the initial estimation error variances described in Section 3.2.

From the description of the state vector and state dynamics model given in this section, it is clear that even this so-called "optimal" design is not truly optimal, in the sense that certain simplifying assumptions have already been made to restrict the number of modelled states. The model simplifications include the following:
1) Vertical position error, vertical velocity error and vertical accelerometer bias states have not been modelled in the Kalman filter for either the master INS or the IMU strapdown navigator because these errors are kept small by separate fixed-gain baroaltitude damping loops.
2) The effect of the scale factor error in the master INS ring laser gyros has not been modelled because it is very small (5 ppm) and its effect is expected to be insignificant compared to the effect of other master INS gyro errors.
3) Certain time-correlated errors in the strapdown IMU sensors, such as gyro mass unbalances and gyro scale factor errors, have only been modelled, if at all, as white noise processes driving the inertial system error states, because for this configuration where the IMU is physically stabilized in roll and pitch, their effects are too small to allow these errors to be observable as separately modelled states in the Kalman filter.
4) Vertical deflection effects have been modelled as a white noise process driving velocity error states because these effects are expected to be too small to warrant modelling as filter states.

Whether or not a filter designer feels confident in making such model simplifications right at the outset of the design process depends largely on the designer's experience and understanding of the system at hand. Typically, it is desireable to make some initial model simplifications simply to reduce the development time required to arrive at the final Kalman filter configuration. However, if there is any doubt as to the significance of certain errors, it is better to err on the side of caution and include them as modelled states in the initial filter design. Later, through further analysis, detailed simulations and field trials, the designer gains additional insights into the system error behaviour, which may then allow him to reduce the model complexity with a higher degree of confidence. The initial filter formulation then serves as the designer's "optimal" configuration against which the performance of other more suboptimal designs is compared. In this respect, the state dynamics model presented in this section is considered the optimal Kalman filter model for the TOA system, despite containing the model simplifications described.

3.4 Measurements
Two different sets of measurements are constructed for use in the Kalman filter. One set is a velocity matching measurement $\mathbf{z}_D$ between the DVS and the master INS. This

measurement type is used to estimate velocity and attitude errors in the master INS. The other set is a velocity matching measurement z_V between the master INS and the strapdown navigator, which effectively accomplishes the transfer of alignment.

The measurement vector z_D is computed as

$$z_D^A = \begin{bmatrix} z_{D_x} & z_{D_y} & z_{D_z} \end{bmatrix}^T$$
$$= (C_A^{MW^T} V_M^{MW} - V_D^A) + (\omega_A^A \times d_D^A) \quad , \tag{52}$$

where the variables are defined as follows:

z_D^A - the master INS/Doppler measurement vector resolved in the aircraft body frame, denoted by the superscript "A",

C_A^{MW} - the direction cosine matrix computed from Eq.(28) by using the roll, pitch, heading and wander angle indicated by the master INS,

V_M^{MW} - the velocity vector computed by the master INS and resolved in the master wander azimuth frame, as denoted by the superscript "MW",

V_D^A - the aircraft velocity vector indicated by the DVS and resolved in the aircraft body frame,

ω_A^A - the angular velocity vector of the aircraft body frame with respect to an earth-fixed frame, as computed by the master INS, resolved in the aircraft body frame,

d_D^A - the lever arm vector from the master INS to the DVS, resolved in the aircraft body frame.

The measurement vector z_V is calculated by the expression

$$z_V^{MW} = \begin{bmatrix} z_{V_x} & z_{V_y} & z_{V_z} \end{bmatrix}^T$$
$$= (C_{SW}^{MW} V_S^{SW} - V_M^{MW}) - (C_A^{MW} [\omega_A^A \times d_S^A]) \quad , \tag{53}$$

where the variables not previously defined are :

z_V^{MW} - the master INS/strapdown measurement vector, resolved in the master INS wander azimuth frame,

V_S^{SW} - velocity vector computed by the IMU strapdown navigator, and resolved in the strapdown IMU wander azimuth frame, denoted by the superscript "SW",

d_S^A - the lever arm vector from the master INS to the strapdown IMU, resolved in the aircraft body frame,

C_{SW}^{MW} - the direction cosine matrix describing the transformation from the strapdown IMU wander azimuth frame to the master INS wander azimuth frame.

The matrix C_{SW}^{MW} is computed by

$$C_{SW}^{MW} = C_{MW}^{E^T} C_{SW}^{E} \quad , \tag{54}$$

where C_{SW}^E and C_{MW}^E are the direction cosine matrices describing respectively the transformation from the strapdown IMU wander azimuth frame to a specified earth-fixed frame, and the transformation from the master INS wander azimuth frame to the earth-fixed frame. The matrix C_{SW}^E is calculated as

$$C_{SW}^E = \begin{bmatrix} \cos\alpha \cos L & -\sin\alpha \cos L & \sin L \\ \begin{matrix}\sin\alpha \cos l + \\ \cos\alpha \sin L \sin l\end{matrix} & \begin{matrix}\cos\alpha \cos l - \\ \sin\alpha \sin L \sin l\end{matrix} & -\cos L \sin l \\ \begin{matrix}\sin\alpha \sin l - \\ \cos\alpha \sin L \cos l\end{matrix} & \begin{matrix}\cos\alpha \sin l + \\ \sin\alpha \sin L \cos l\end{matrix} & \cos L \cos l \end{bmatrix} , \tag{55}$$

where α, L and l are the wander angle, latitude and longitude computed by the IMU strapdown navigator. The matrix C_{MW}^E is calculated with exactly the same expression as in Eq.(55) except that the wander angle, latitude and longitude used are those indicated by the master INS.

The Kalman filter actually processes only z_{V_x} and z_{V_y}, the horizontal components of z_V^{MW}, and z_{D_x} and z_{D_y}, the x and y components of z_D^A. Also, the master INS/Doppler measurements are processed only during nominally straight and level flight to avoid additional inaccuracies in the DVS that result when the aircraft is steeply rolled or pitched. The measurement vector z that is processed by the filter is then written as

$$z = \begin{bmatrix} z_{D_x} \\ z_{D_y} \\ z_{V_x} \\ z_{V_y} \end{bmatrix} \quad . \tag{56}$$

3.5 Measurement Models

The linear model for discrete measurements is given by

$$z = Hx + v \quad , \tag{57}$$

where H is the measurement matrix with dimension 4×26 and v is a 4 element vector of measurement errors that are assumed to be zero-mean sequences uncorrelated in time. The measurement matrix is derived by perturbing the right hand side of Eqs.(52) and (53) and expressing the result in terms of the modelled error states. The following assumptions are made in this derivation: 1) the effect of errors in computing the $\omega \times d$ terms and the transformation matrix C_{SW}^{MW} can be neglected (these small effects are included in the measurement noise model), 2) the x and y axes of the aircraft body frame are nominally in the horizontal plane when master INS/Doppler measurements are constructed, which allows the measurement error terms containing vertical velocity errors to be neglected, and 3) vertical velocities are small enough (i.e. < 10 m/s) to allow measurement error terms containing vertical velocity to be neglected. These last two assumptions are valid since, as

pointed out earlier, master INS/Doppler measurements are processed only during nominally straight and level flight. The resulting measurement matrix is expressed in terms of submatrices as:

$$H = \begin{bmatrix} H_{D_M} & 0 & H_{D_{DVS}} & 0 & 0 \\ H_{V_M} & 0 & 0 & H_{V_S} & 0 \end{bmatrix} \quad . \tag{58}$$

The submatrices H_{D_M} and $H_{D_{DVS}}$, which have dimensions of 2×7 and 2×2 respectively, are given by

$$H_{D_M} = \begin{bmatrix} 0 & 0 & C_{11} & C_{21} & C_{31}V_{M_Y} & -C_{31}V_{M_X} & (C_{21}V_{M_X} - C_{11}V_{M_Y}) \\ 0 & 0 & C_{12} & C_{22} & C_{32}V_{M_Y} & -C_{32}V_{M_X} & (C_{22}V_{M_X} - C_{12}V_{M_Y}) \end{bmatrix} \tag{59}$$

$$H_{D_{DVS}} = \begin{bmatrix} -V_{D_x} & 0 \\ 0 & V_{D_x} \end{bmatrix} \quad , \tag{60}$$

where the variables are defined as:

C_{ij} - the elements of C_A^{MW},

V_{D_x} - the x component of V_D^A

V_{M_X}, V_{M_Y} - the X and Y components of V_M^{MW},

V_{M_x}, V_{M_y} - the x and y components of V_M^A which is computed as

$$V_M^A = C_A^{MW^T} V_M^{MW} \quad . \tag{61}$$

The submatrices H_{V_M} and H_{V_S}, both of dimension 2×7, are given below:

$$H_{V_M} = \begin{bmatrix} 0 & 0 & -1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 & 0 & 0 & 0 \end{bmatrix} \quad , \tag{62}$$

$$H_{V_S} = \begin{bmatrix} 0 & 0 & C_{11}' & C_{12}' & 0 & 0 & 0 \\ 0 & 0 & C_{21}' & C_{21}' & 0 & 0 & 0 \end{bmatrix} \quad , \tag{63}$$

where the C_{ij}' are the elements of C_{SW}^{MW}.

The measurement noise vector v is characterized by its covariance matrix R which is given by

$$R = E[v \, v^T] = \begin{bmatrix} \sigma_{D_x}^2 & 0 & 0 & 0 \\ 0 & \sigma_{D_y}^2 & 0 & 0 \\ 0 & 0 & \sigma_V^2 & 0 \\ 0 & 0 & 0 & \sigma_V^2 \end{bmatrix} \quad . \tag{64}$$

In the above equation, the variables are defined as

$\sigma_{D_x}^2$ - the variance of the measurement noise for z_{D_x},

$\sigma_{D_y}^2$ - the variance of the measurement noise for z_{D_y},

σ_V^2 - the variance of the measurement noise for z_{V_X} and z_{V_Y}.

The main contributor to the measurement noise for z_{D_x} and z_{D_y} is the noise on the DVS velocity measurements, which is specified in Section 2.1. The measurement noise modelled for z_{V_X} and z_{V_Y} includes the effects of quantization of master INS velocity data, and the effects of airframe flexing which causes errors in computing relative velocity between the master INS and strapdown IMU.

The fact that the matrix R is diagonal implies that the elements of v are uncorrelated with each other. Strictly speaking, this is not true. There is a correlation between the measurement noise for z_{D_x} and z_{D_y} due to the fact that the noise associated with the DVS forward and lateral velocity measurements is somewhat correlated. However, by assuming that the noise elements are uncorrelated, there is only a small loss in optimality, while there is a significant advantage gained in that the scalar components of the measurement vector can be processed sequentially by the filter at a given update time, rather than processing the z vector as a whole. This allows matrix inversion to be avoided, thereby increasing computational efficiency.

3.6 Simulation Results

A high-fidelity computer simulation is an extremely useful tool in evaluating a Kalman filter design and predicting filter performance. Such a simulation was performed for the optimal TOA filter described in this section.

The simulation package that was used generates simulated master INS, strapdown IMU, DVS and air data, which are then used as inputs to a processing package. The simulated sensor data corresponds to a particular user-defined flight profile, and includes the effects of all the sensor errors discussed in Section 2.1, as well as other errors like quantization of the sensor output data. The simulation package also provides files containing true position, velocity, attitude and heading as a function of time for both the master INS and strapdown IMU. These files serve as the reference against which the processing package outputs are compared in order to evaluate Kalman filter estimation errors. The processing package consists of the Kalman filter, strapdown navigator and baroaltitude algorithm shown in the block diagram of Figure 1.

For the Kalman filter processor, the Kalman filter equations are implemented in their discrete form using Bierman's U-D factorization algorithm [10] to efficiently propagate and update the error covariance matrix . Other features that are used to enhance computational efficiency include sequential processing of measurements, sparse matrix multiplication techniques and exploitation of the block structure of certain matrices to increase the execution speed of matrix operations. In the discrete Kalman filter formulation, the model for the error state dynamics as expressed in Eq.(18) is converted to a discrete-time model of the form:

$$\mathbf{X}_{k+1} = \Phi_k \mathbf{X}_k + \mathbf{W}_k \quad , \tag{65}$$

where $\mathbf{w}_k$ is a vector of zero-mean white noise sequences with covariance matrix $\mathbf{Q}_k$, and $\mathbf{x}_k$ and $\mathbf{x}_{k+1}$ refer to the values of the state vector at times t_k and t_{k+1} respectively. The interval $\Delta T_k \equiv (t_{k+1} - t_k)$ is the interval between Kalman filter measurement updates. The update interval for this simulation is 10 seconds. To obtain the transition matrix Φ_k, a transition matrix Φ_j is first computed over a smaller subinterval, referred to as the propagation interval, by retaining the first three terms of the matrix exponential expression below:

$$\begin{aligned}\Phi_j &= \exp(\overline{\mathbf{F}}_j \Delta t_j) \\ &= \mathbf{I} + \overline{\mathbf{F}}_j \Delta t_j + \frac{1}{2}\overline{\mathbf{F}}_j^2 (\Delta t_j)^2 + \frac{1}{6}\overline{\mathbf{F}}_j^3 (\Delta t_j)^3 + \dots \quad ,\end{aligned} \tag{66}$$

where Δt_j is the propagation interval defined by $\Delta t_j \equiv (t_{j+1} - t_j)$, and $\overline{\mathbf{F}}_j$ is the average value of $\mathbf{F}(t)$ over Δt_j, with $\mathbf{F}(t)$ given by the expression in Eq.(20). In the simulation, the basic propagation interval is 2 seconds. The transition matrix Φ_j is recalculated every propagation interval, and Φ_k is then calculated as

$$\Phi_k = \Phi_{j+4}\Phi_{j+3}\Phi_{j+2}\Phi_{j+1}\Phi_j \quad . \tag{67}$$

The covariance matrix $\mathbf{Q}_k$ is computed from the expression

$$\mathbf{Q}_k = \int_0^{\Delta T_k} e^{\mathbf{F}_{k+1}\tau}\mathbf{Q}(\tau)(e^{\mathbf{F}_{k+1}\tau})^T d\tau \tag{68}$$

by expanding the matrix exponential terms, multiplying out and integrating the products with the assumption that $\mathbf{Q}(t)$, the spectral density matrix given in Eq.(33), is constant over the interval ΔT_k. In the above expression, $\mathbf{F}_{k+1}$ is the value of $\mathbf{F}(t)$ at t_{k+1}.

For this simulation, the elements of the submatrices $\mathbf{P}_{M_o}$ and $\mathbf{P}_{S_o}$ in the initial error covariance matrix $\mathbf{P}_0$ were assigned values consistent with expected system errors in the INS following a stationary ground alignment. The flight profile used for the simulation is depicted in Figure 3. A description of the various manoeuvres included in the simulated flight profile is given in Table 4. It should be noted that the DVS data is not used to construct Kalman filter measurements until cruising altitude is reached. The velocity matching measurements between the strapdown navigator and master INS are constructed and processed right from time zero.

The performance of the optimal TOA Kalman filter is indicated in Figures 4 to 8. The solid traces in the plots indicate the strapdown navigator velocity, attitude and heading estimation errors over the course of the flight, assuming closed loop correction of strapdown navigator quantities with the filter-estimated error states. The estimation errors are obtained by subtracting the "true" velocity, attitude and heading values, as generated by the simulation package, from the corresponding values

computed by the strapdown navigator in the processing package. The dashed lines show the filter-predicted standard deviation bounds for these estimation errors. The sharp decreases in the standard deviation bounds generally occur right after turns, when DVS error states and inertial platform misalignments about the Z axis become separately observable from the other error states. It can be seen that the time histories of the estimation errors are fairly consistent with the predicted standard deviation bounds.
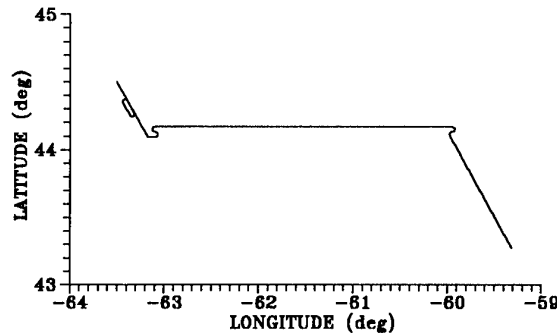
Figure 3. Simulated Flight Profile

TIME (s)	DESCRIPTION OF MANOEUVRE
0	Stationary; heading is 150 degrees
3	Start accelerating down runway
75	Takeoff; aircraft pitch increases to 6 degrees
116	Achieve and maintain cruising speed of 108 m/s
173	Reach and maintain cruising altitude of 1000 m; aircraft pitch returns to zero
330	Enter racetrack; aircraft roll increases to 30 degrees
392	Complete first 180 degree turn of racetrack
500	Enter second turn of racetrack
562	Complete second 180 degree turn of racetrack; aircraft heading returns to 150 degrees
870	Start turn to heading of 90 degrees
891	Complete turn to heading of 90 degrees
941	Start S-manoeuvre
1067	Complete S-manoeuvre; aircraft heading returns to 90 degrees
3400	Start S-manoeuvre
3506	Complete S-manoeuvre; heading held at 150 degrees
4500	Complete profile

Table 4. Description of Manoeuvres in Simulated Flight Profile.
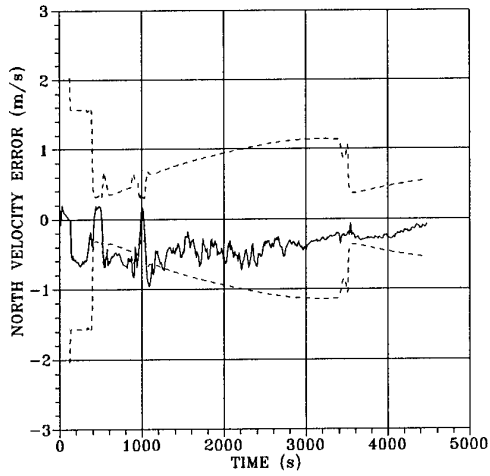
Figure 4. Strapdown Navigator North Velocity Error and Predicted Error Standard Deviation (Optimal Filter).
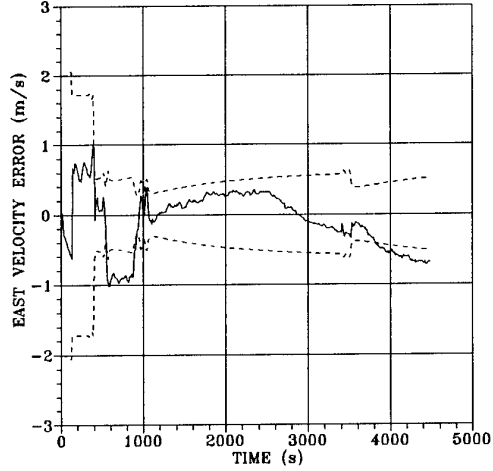
Figure 5. Strapdown Navigator East Velocity Error and Predicted Error Standard Deviation (Optimal Filter).
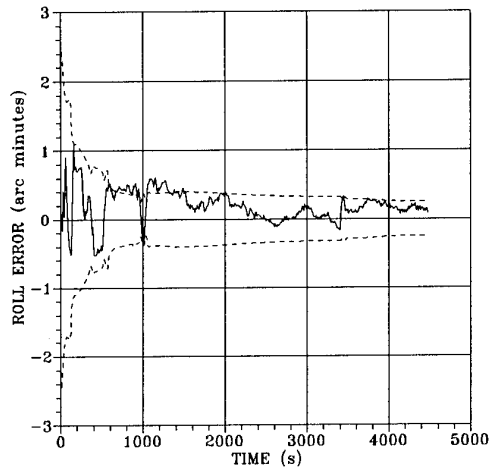
Figure 6. Strapdown Navigator Roll Error and Predicted Error Standard Deviation (Optimal Filter).
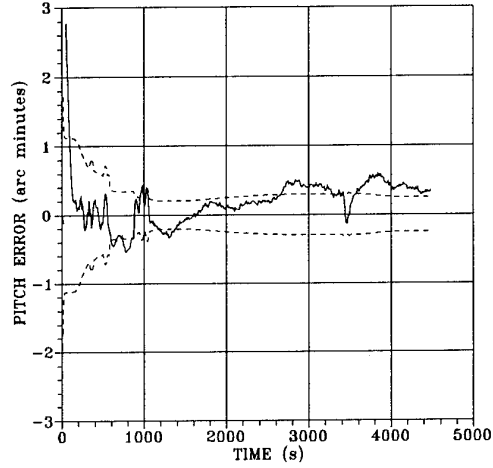
Figure 7. Strapdown Navigator Pitch Error and Predicted Error Standard Deviation (Optimal Filter).
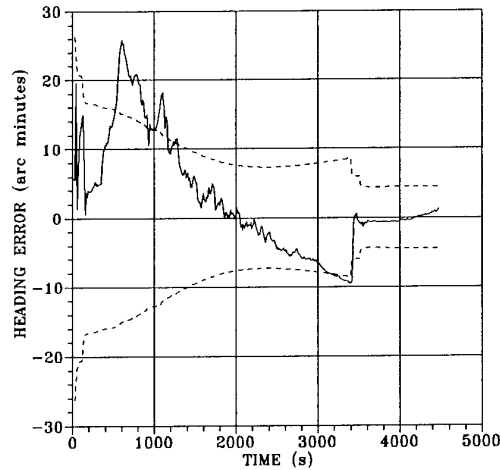
Figure 8. Strapdown Navigator Heading Error and Predicted Error Standard Deviation (Optimal Filter).

4. SUBOPTIMAL TOA KALMAN FILTER DESIGN

4.1 State Decoupling

If certain parts of a system state vector are only weakly coupled, it is often possible to split the one Kalman filter into two or more lower order filters. This can reduce the computational burden by a significant amount since the number of multiplications required to perform the most intensive filter task - the solution of the error covariance equations - varies roughly as the third power of the number of states.

In this section, it is shown that the optimal TOA Kalman filter described in Section 3 can effectively be split into two filters. The Doppler damping of the master INS is performed in one filter, while the transfer of alignment from the master INS to the strapdown navigator is performed in a second filter which is mathematically decoupled from the first.

The means by which the original Kalman filter can be decoupled becomes more apparent if one considers an alternate filter formulation for which the *differences* between the absolute errors in the master INS and strapdown navigator are modelled, instead of the absolute strapdown navigator errors themselves. Such a filter formulation is based on the fundamental insight that measurements constructed by comparing equivalent information from two systems with the same error dynamics will only allow differences between the two systems to be observable, rather than the absolute errors in each system. Thus the velocity matching measurement between the master INS and strapdown navigator provides direct observability only for x'_s where

$$x'_s \equiv x_s - x_M = \begin{bmatrix} \delta R'_{S_X} \\ \delta R'_{S_Y} \\ \delta V'_{S_X} \\ \delta V'_{S_Y} \\ \phi'_{S_X} \\ \phi'_{S_Y} \\ \phi'_{S_Z} \end{bmatrix} , \qquad (69)$$

and not for x_S or x_M themselves, whereas the DVS measurements, as defined in Eq.(52) do allow elements of x_M to be observable. From the viewpoint of observability then, an alternative selection of inertial system error states for the original fully coupled filter could just as well be x_M and x'_s instead of x_M and x_S. In that case, the dynamics of the state vector x'_s are derived by subtracting Eq.(21) from Eq.(24) to give

$$\dot{x}'_s = F_S x'_s + (F_{S/IMU} x_{IMU} - F_{M/INS} x_{INS}) + w'_s , \qquad (70)$$

where

$$w'_s = w_s - w_M , \qquad (71)$$

and it has been assumed that $F_M \approx F_S$. This assumption is valid as long as the axes of the master INS and strapdown navigator wander azimuth frames are nearly coincident (i.e. $C_{SW}^{MW} \approx I$, the identity matrix) and the relative motion between the master INS and strapdown IMU is small compared to the nominal values of velocity and specific force appearing in the F matrices. Both these conditions are true in this case.

The measurement model for the master INS/strapdown velocity matching measurements is obtained in terms of x'_s by first writing out the following expression, using Eqs.(56), (57) and (58):

$$\begin{bmatrix} z_{V_X} \\ z_{V_Y} \end{bmatrix} = H_{V_M} x_M + H_{V_S} x_S , \qquad (72)$$

where H_{V_M} and H_{V_S} are given by Eqs.(62) and (63). Substituting $x_S = x'_s + x_M$ into the equation above and assuming again that $C_{SW}^{MW} \approx I$ yields

$$\begin{bmatrix} z_{V_X} \\ z_{V_Y} \end{bmatrix} = H_{V_S} x'_s . \qquad (73)$$

The form of the above equation clearly confirms that only x'_s is fundamentally observable from the z_V measurement.

In order for the optimal TOA filter to be decoupled, the estimation of one set of states x_1 must be mathematically decoupled from the estimation of the second set x_2, where in this case, it is desired that

$$x_1 = \begin{bmatrix} x_M \\ x_{INS} \\ x_{DVS} \end{bmatrix} , \qquad x_2 = \begin{bmatrix} x'_s \\ x_{IMU} \end{bmatrix} . \qquad (74)$$

For x_1 to be decoupled from x_2, there must be no modelled correlation between x_1 and x_2. In other words, the cross-covariance matrix of x_1 and x_2 must be zero at all times. There are generally three ways that correlation between x_1 and x_2 can be introduced into the Kalman filter:
1) through the measurement model,
2) through the model of the system dynamics,
3) through the initialization of the covariance matrix.

The measurement model for z_{V_X} and z_{V_Y}, as expressed in Eq.(73), is only a function of x_2, and the measurement model for z_{D_X} and z_{D_Y} discussed in Section 3.5 is only a function of x_1. So these measurement models are already in a form that avoids correlation between x_1 and x_2.

Examination of the error dynamics models in Eq.(70) and (21) indicates that correlation will develop between x_1 and x_2 because both x_M and x'_s are driven by a common error state term $F_{INS} x_{INS}$. However, the error covariance for the states

in x_{IMU} is modelled as being significantly larger than the error covariance for x_{INS}, so that the effect of $F_{INS}x_{INS}$ on the estimation of x_S' is expected to be much less than the effect of the $F_{IMU}x_{IMU}$ term, and can therefore be neglected. There is also a certain amount of correlation between w_M and w_S'. The cross spectral density matrix for these two process noise vectors, using Eqs.(34) and (42), is found to be

$$Q_{M/S}' = Q_{M/S} - Q_M$$

$$= \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & -q_{MASF} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -q_{MASF} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & -q_{MGN} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & -q_{MGN} & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & -q_{MGN} \end{bmatrix} \quad (75)$$

However, consider the spectral density matrix for w_S'. This matrix Q_S' is expressed in terms of previously defined spectral density matrices as

$$Q_S' = Q_M + Q_S - 2Q_{M/S} \quad , \quad (76)$$

which yields

$$Q_S' = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & q_{SASF}+q_{MASF} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & q_{SASF}+q_{MASF} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & q_{SGMU}+q_{MGN} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & q_{SGMU}+q_{MGN} & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & q_{SGMU}+q_{SGSF}+q_{MGN} \end{bmatrix} \quad (77)$$

The principal terms in the above expression that affect the estimation of x_S' are q_{SGMU} and q_{SGSF}, which, in general, are significantly larger than the q_{MGN} terms that appear in the same locations in the matrix $Q_{M/S}'$. Consequently, it can be assumed that $Q_{M/S}' = 0$ with little effect on the estimation of x_S'. It may be noticed that q_{VD} does not appear in Eq.(77). This is consistent with the expectation that vertical deflection errors, which affect both x_S and x_M in the same way, should not affect the difference states. Based on the above discussion, the state dynamics models for the two decoupled filters can be written as

$$\begin{aligned} \dot{x}_1 &= F_1 x_1 + w_1 \quad , \\ \dot{x}_2 &= F_2 x_2 + w_2 \quad , \end{aligned} \quad (78)$$

where

$$F_1 = \begin{bmatrix} F_M & F_{M/INS} & 0 \\ 0 & F_{INS} & 0 \\ 0 & 0 & F_{DVS} \end{bmatrix} , \quad w_1 = \begin{bmatrix} w_M \\ w_{INS} \\ w_{DVS} \end{bmatrix} , \quad (79)$$

and

$$F_2 = \begin{bmatrix} F_S & F_{S/IMU} \\ 0 & F_{IMU} \end{bmatrix} , \quad w_2 = \begin{bmatrix} w_S' \\ w_{IMU} \end{bmatrix} . \quad (80)$$

The initial cross-covariance matrix between x_1 and x_2 is expressed as

$$P_{12_0} = E[\tilde{x}_{1_0} \tilde{x}_{2_0}^T] = \begin{bmatrix} E[\tilde{x}_{M_0} \tilde{x}_{S_0}'^T] & E[\tilde{x}_{M_0} \tilde{x}_{IMU_0}^T] \\ E[\tilde{x}_{INS_0} \tilde{x}_{S_0}'^T] & E[\tilde{x}_{INS_0} \tilde{x}_{IMU_0}^T] \\ E[\tilde{x}_{DVS_0} \tilde{x}_{S_0}'^T] & E[\tilde{x}_{DVS_0} \tilde{x}_{IMU_0}^T] \end{bmatrix} . \quad (81)$$

All of the terms in the above expression are found to be zero except for the initial cross-covariance matrix between x_M and x_S'. This term is evaluated, with the help of Eqs.(11) and (14), as

$$E[\tilde{x}_{M_0} \tilde{x}_{S_0}'^T] = P_{M/S_0} - P_{M_0}$$

$$= \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & -\sigma_{M\phi}^2 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & -\sigma_{M\phi}^2 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} . \quad (82)$$

The magnitude of the non-zero terms in the above expression is much smaller than the initial variance $\sigma_{S\phi}^2$ assigned to the estimation error for ϕ_{S_X}' and ϕ_{S_Y}'. This variance is large to reflect the large uncertainty in the knowledge of the initial roll and pitch of the IMU. Under these conditions then, it can be assumed that $P_{12_0} = 0$, with little consequence for the estimation of x_2.

In summary then, it can safely be assumed that the three conditions to allow state decoupling, as stated previously in this section, are satisfied for the two vectors x_1 and x_2, and consequently, the optimal TOA filter can be mechanized as two decoupled filters.

It should be noted that for the purposes of closed loop correction of the strapdown navigator parameters, it is still an estimated value of x_S that is required, not x_S'. Consequently, for this decoupled filter configuration where x_S is not explicitly estimated, it must be constructed when it is needed as

$$\hat{x}_S = \hat{x}_M + \hat{x}_S' \quad , \tag{83}$$

where the symbol $^\wedge$ denotes estimated quantities.

4.2 State Deletion

If certain error states that are being modelled in the filter are too small to be observable with the given measurement models, then they can often be eliminated with little impact on the filter performance. For example, consider the plots in Figures 9 to 11 obtained from the simulation run with the optimal filter. The solid traces in these plots show the time histories of the master INS gyro bias state estimation errors, and the dashed lines show the filter-computed standard deviations for these estimation errors. It is obvious that the predicted standard deviations do not change significantly from their initial values over the course of the run. This is a clear indication that these states are too small to be observable by the filter.

When states are eliminated, it is usual to attempt to compensate for their loss by adding a component to the process noise. For example, from the expression for $F_{M/INS}$ in Eq.(27), the first-order effect of a gyro bias state is to cause a linear growth in inertial platform misalignments. Thus, the variance growth for the platform misalignment estimation error as a result of a modelled gyro bias state is given approximately by

$$\sigma_{M\phi}^2(t) = \sigma_{MGB}^2 t^2 \quad . \tag{84}$$

If the effects of a gyro bias were modelled by a white noise process with spectral density q_{MGB}, then the variance growth for the platform misalignment estimation errors is expressed as

$$\sigma_{M\phi}^2(t) = q_{MGB} t \quad . \tag{85}$$

One criterion that the filter designer may use to assign an appropriate value to q_{MGB} is to select a value of q_{MGB} that results in the $\sigma_{M\phi}^2$ buildup being equal for both cases at a chosen time T. This condition is expressed as

$$\sigma_{MGB}^2 T^2 = q_{MGB} T \quad , \tag{86}$$

which yields

$$q_{MGB} = \sigma_{MGB}^2 T \quad . \tag{87}$$

The choice of T is itself fairly arbitrary, but a suitable choice is the filter settling time, which is in the order of about 1000 seconds. The quantity q_{MGB} computed from the above expression is then added to the q_{MGN} terms in the spectral density matrix Q_M to compensate for the deletion of the gyro bias states.

Another case for deleting states can be made when it is not important to know the state estimate, and the state does not
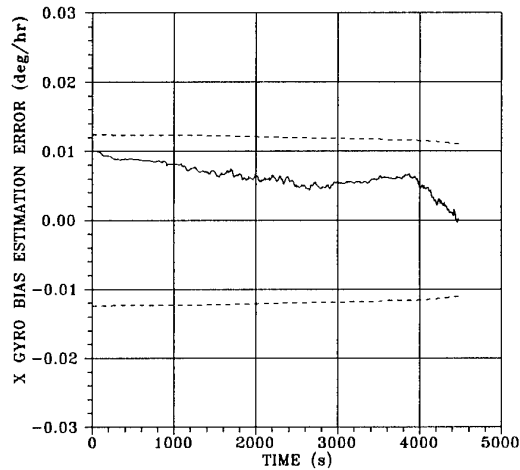
Figure 9. Master INS X Gyro Bias Estimation Error and Predicted Error Standard Deviation (Optimal Filter).
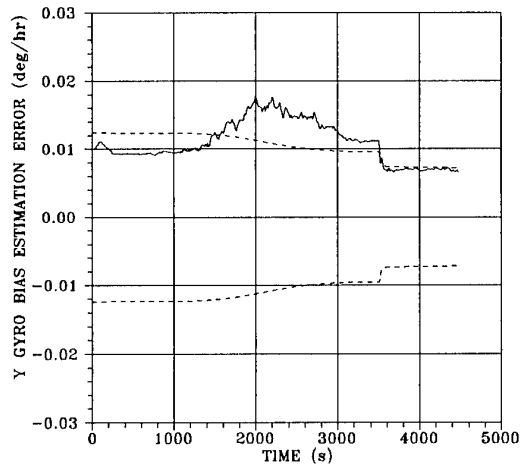
Figure 10. Master INS Y Gyro Bias Estimation Error and Predicted Error Standard Deviation (Optimal Filter).
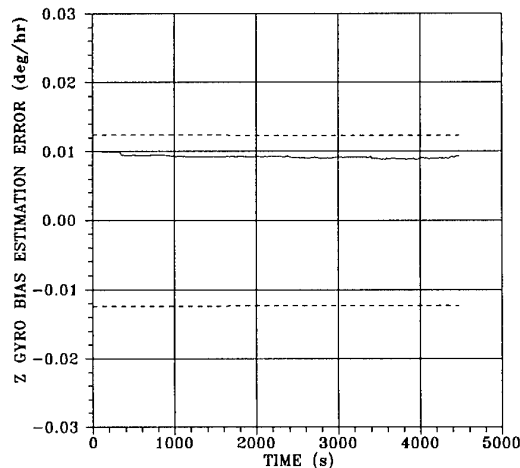
Figure 11. Master INS Z Gyro Bias Estimation Error and Predicted Error Standard Deviation (Optimal Filter).

have a significant effect on the estimation of other states of interest. A good example of this is the position error states modelled for the master INS and the strapdown navigator. The primary requirement for the TOA filter is to control attitude errors and velocity errors in the strapdown navigator. Absolute positional accuracy is not a requirement. Furthermore, deletion of the position error states does not impact on estimation of the other states. This can be seen by considering the measurement model and error dynamics model for the TOA filter. Since the measurement model does not include any terms that contain the position error states, then obviously it is not affected by the absence of these states. With regard to the error dynamics, from examination of the expression for $\mathbf{F_M}$ in Eq.(26), it can be seen that position error terms which drive the velocity errors and platform misalignments are weak compared to other driving terms for these states. The dominant term that would be omitted as a result of deleting

the position error states is of the form $\dfrac{\omega_{IEz}}{R}\delta R$. To first order, this term causes a linear growth of platform tilts. Following a similar approach to that taken for the gyro bias states, an appropriate selection for the spectral density of an equivalent white noise process is

$$q_{MR} = \left(\frac{\omega_{IEz}}{R}\right)^2 \sigma_{MR}^2 T \quad , \qquad (88)$$

where σ_{MR}^2 is chosen to be the expected variance of the position error over the entire flight, and T , as before, is selected to be the filter settling time. This new spectral density term is included in $\mathbf{Q_M}$ by summing it with the q_{MGN} terms that affect the estimation of platform tilts. A similar spectral density term q_{SR} can be incorporated into $\mathbf{Q'_s}$ to compensate for the deletion of position states from the $\mathbf{x'_s}$ vector.

4.3 Simulation Results
An identical simulation run to that performed for the optimal TOA filter was conducted with the suboptimal filter design. For this run, the modifications to the measurement and error dynamics model discussed in this section were implemented to realize two decoupled filters without position error states or master INS gyro bias states.

The results of the run are shown in Figures 12 to 16 which depict the velocity, attitude and heading estimation errors of the strapdown navigator, along with associated filter-predicted standard deviations, after closed loop correction with the filter-estimated error quantities. From comparison to the corresponding plots in Figures 4 to 8, it is clear that the suboptimal filter performance is very nearly the same as that of the optimal configuration. These results confirm the validity of the model simplifications discussed in this section.

5. CONCLUSIONS
The process of designing a suboptimal *transfer-of-alignment* (TOA) Kalman filter for a SAR motion compensation application has been detailed at some length. While a

certain amount of the information presented is fairly unique to this particular application, the criteria used in making the filter model simplifications provide useful guidelines for general suboptimal filter design. Also, much of the discussion regarding inertial system behaviour is of course relevant to aerospace navigation applications involving inertial technology. Further information on general principles of suboptimal filter design and implementation can be found in [6].

Specifically, two principal techniques of Kalman filter model simplification, namely state deletion and state decoupling, have been demonstrated for the TOA filter example. Starting with a description of the optimal Kalman filter, the process of designing modifications to accommodate state deletion and decoupling has been discussed, along with the rationale for these modifications. Simulation results for the optimal and suboptimal filter configurations were presented to confirm that the suboptimal filter design provided similar performance to the optimal filter.

6. REFERENCES
1. J.J.Kovaly, *Synthetic Aperture Radar*, Artech House Inc., Dedham, Massachusetts, 1976.

2. M.I.Skolnik, ed., *Radar Handbook*, Chapter 23, McGraw-Hill Book Company, New York, 1970.

3. D.J.DiFilippo, G.E.Haslam, and W.S.Widnall, "Evaluation of a Kalman Filter for SAR Motion Compensation", proceedings of PLANS 88, Orlando, Florida, Nov.29-Dec.2, 1988.

4. R.E.Kalman, "A New Approach to Linear Filtering and Prediction Problems", J. Basic Eng., March 1960, pp.35-46.

5. R.E.Kalman and R.S.Bucy, "New Results in Linear Filtering and Prediction Theory", J. Basic Eng., March 1961, pp.95-108.

6. A.Gelb, ed., *Applied Optimal Estimation*, The M.I.T. Press, Cambridge Massachusetts, 1974.

7. M.Kayton and W.R.Fried, eds., *Avionics Navigation Systems*, Chapter 6, John Wiley and Sons, Inc., New York, 1969.

8. D.O.Benson, "A Comparison of Two Approaches to Pure Inertial and Doppler-Inertial Error Analysis", IEEE Transactions on Aerospace and Electronic Systems, Vol. AES-11, No.4, July 1975, pp.447-455.

9. D.B.Reid and J.Hepburn, "Final Report for SAR Motion Compensation Study", Report H8402-03/SARMC/DBR, Huntec (70) Ltd., February, 1984.

10. G.J.Bierman, *Factorization Methods for Discrete Sequential Estimation*, Academic Press, New York, 1977.
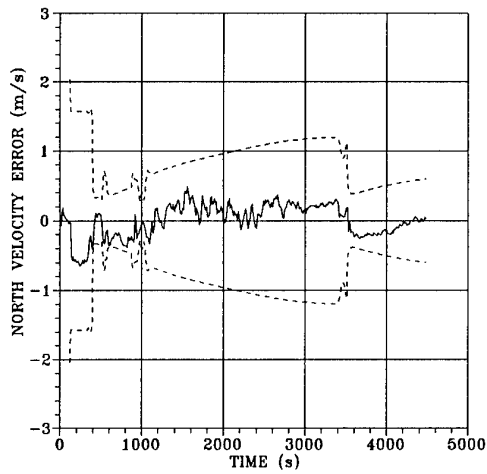
Figure 12. Strapdown Navigator North Velocity Error and Predicted Error Standard Deviation (Suboptimal Filter).
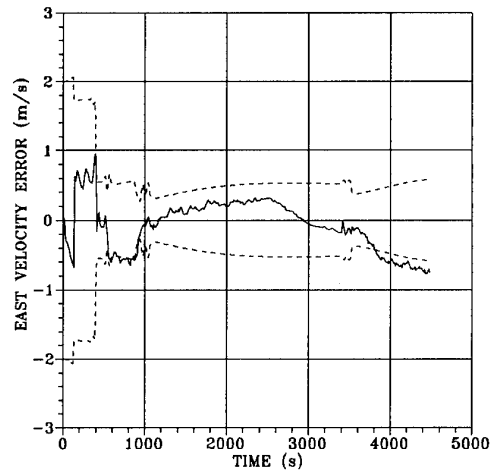
Figure 13. Strapdown Navigator East Velocity Error and Predicted Error Standard Deviation (Suboptimal Filter).
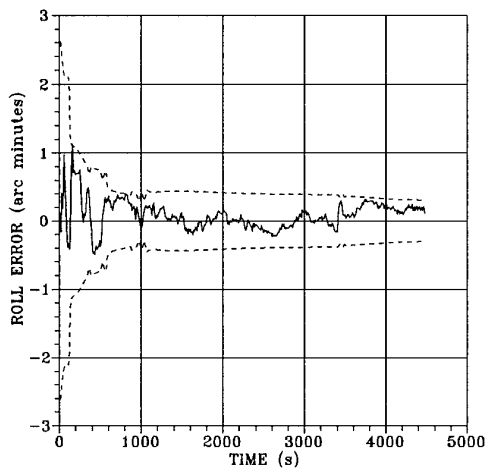
Figure 14. Strapdown Navigator Roll Error and Predicted Error Standard Deviation (Suboptimal Filter).
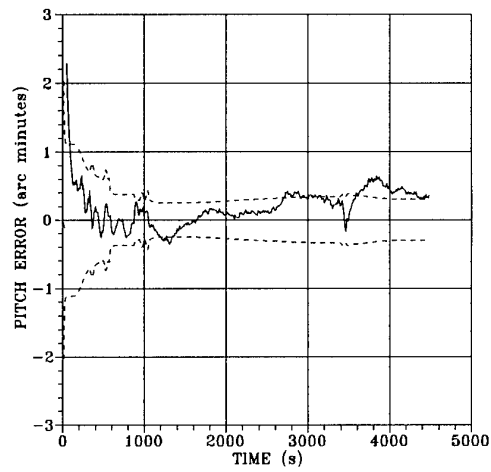
Figure 15. Strapdown Navigator Pitch Error and Predicted Error Standard Deviation (Suboptimal Filter).
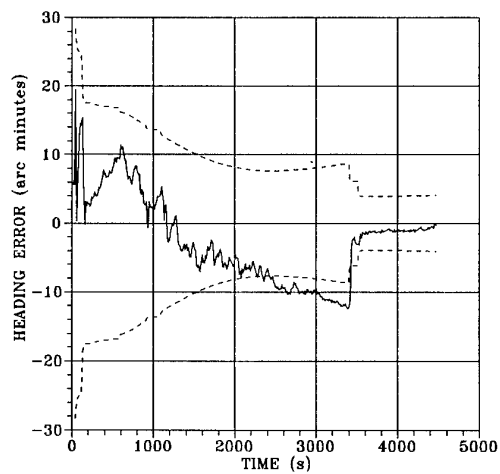
Figure 16. Strapdown Navigator Heading Error and Predicted Error Standard Deviation (Suboptimal Filter).

VERTICAL CHANNEL DESIGN CONSIDERATIONS

J. Stanley Ausman
Litton Guidance and Control Systems Division
5500 Canoga Avenue
Woodland Hills, CA 91367-6698
USA

1. Introduction

The vertical channel of an inertial system is unstable. This instability is caused by the gravity compensation fed back to the vertical accelerometer output. The gravity compensation, computed as a function of altitude after doubly integrating the output of that accelerometer, creates an unstable, positive feedback loop. The time constant of this instability is about 560 seconds near the earth's surface. For ballistic missiles and rockets this does not pose a problem because the guidance is completed before the instability becomes serious. For aircraft systems, however, one must augment the inertial measurements, typically with barometric altimeter information, in order to stabilize the vertical inertial channel.

Earliest mechanizations of the baro-inertial loop employed second-order feedback with constant gains (References 1 and 2). The next step was to add integral feedback in order to bias the vertical accel-erometer, thus creating a third-order system. Widnall and Sinha (Ref 3) investigated the third-order loop to find the optimum set of fixed gains. Not surprisingly, they found that the optimum set of gains depended on the values assumed for the noise characteristics of the accelerometer and the barometric altimeter. Because the noise magnitudes will vary as a function of the aircraft's flight regime, the baro-inertial feedback gains should not be constant, but should also vary.

Litton first mechanized a third-order variable gain baro-inertial loop in CLASS, an all-weather close air support system, successfully demonstrated in 1972 (Ref. 4). They gradually improved upon that basic design over the years as successive systems, principally ARIS (Refs. 5 and 6); LW-33 (Ref. 7); and LN-39 (Ref. 8), uncovered more and more barometric altitude error characteristics which had to be accommodated. The culmination of this evolutionary development is the baro-inertial loop currently mechanized in the LN-93 and LN-94 systems for the USAF Standard RLG INU (Ref 9).

Following a discussion of barometric and inertial errors, we will take a detailed look at the LN-93/94 conventional vertical channel mechanization, the reasons behind the loop design, and some simulation results illustrating the loop performance when subjected to certain flight maneuvers and baro-metric errors.

These simulations show that a slow descent (approximately 60 fps) is the most difficult situation for the baroinertial loop to handle because it is hard to discriminate between an inertial velocity error and a barometric scale factor error. In the final section we see that a Kalman filter mechanization can alleviate this difficulty by including barometric scale factor as one of the filter states. The resulting mechanization is a 5-state Kalman filter (inertial altitude, vertical velocity, accelerometer bias, baro scale factor, and baro bias) which uses barometric altitude updating. One can also augment the updates with GPS altitude or radar altimeter measurements, if available. The performance of the Kalman filter mechanization is supe-rior to that of the conventional mechanization of the vertical channel.

In order to properly design the vertical channel, one must first understand the various error - inducing mechanisms. The following paragraphs discuss the various error

sources that may occur in the barometric altimeter and in the vertical channel of the inertial system.

Barometric Error.

There are two general categories of barometric errors. These are (1) modeling errors in mathematically characterizing the real-world atmosphere as a function of altitude, and (2) sensor errors in measuring the static pressure and temperature of the atmosphere. Typical air data computers model pressure and temperature as a function of altitude according to the U.S. Standard Atmosphere (Ref 10). Nonstandard day conditions will cause the air data computer to compute the wrong altitude even if the sensor measurements are correct.

Nonstandard day atmospheric pressure generates a bias-like altitude error. Near sea level, a 10 mm Hg change in pressure corresponds approximately to a 120m change in altitude. Hence, a 10 mm Hg pressure deviation from standard day pressure (7560 mm Hg) would cause a 120m altitude error.

Because the weather and its associated parameters, pressure and temperature, are subject to change, the error due to nonstandard day pressure may also vary with time and geographic location. The correlation time associated with diurnal variations in atmospheric pressure would be $24/2 = 3.82$ hr. The movement of weather systems over the earth adds more uncertainty and tends to decrease the correlation time. Based on these considerations, a correlation time of 2 to 3 hours seems reasonable for atmospheric pressure variations.

Blanchard (Ref 11) analyzed deviations in mean heights of pressure surfaces listed in (Ref 12) and concluded that the RMS day-to-day variability is 380 ft. This data covered latitudes from 20 deg N to 80 deg N and altitudes from 10,000 ft to 50,000 ft.

In addition to temporal variations in pressure at one location, changes may also occur due to spatial changes as the aircraft moves from one location to another. Barham and Manville (Ref 13) give a value of 1 ft per nmi as the RMS slope of a constant pressure surface. This value is in general agreement with the conclusions of Asbury et. al. (Ref 14) in their study of the same problem.

Nonstandard day atmospheric density can cause an apparent altimeter scale factor error as the aircraft changes altitude. Deviations in atmospheric density from standard atmosphere are altitude dependent (Ref 10). Nonstandard density at low altitudes (below 8000m) is negatively correlated with non-standard density, at high altitude (above 8000m). Nonstandard day temperatures have a similar negative correlation between low and high altitudes. At the "isopycnic" level of 8000m, the atmospheric density stays constant to within 2 percent year round and over all latitudes (Ref 10). At sea level, air density can reach extreme deviations from standard day values by as much as $\pm$ 15 percent. For an RMS value, 5 percent deviation is appropriate.

Because air density is subject to the same diurnal and weather-changing patterns as air pressure, the correlation times should be the same. Hence, a 2- to 3- hour autocorrelation time for air density variations appears reasonable.

The principal errors encountered in air data sensing systems arise from the static pressure ports, which may either be on an air data probe or flush-mounted on the aircraft skin. Errors in the air data pressure sensing device and computer are generally small in comparison with the errors in the porting and tubing system, which transmits air pressure to the sensing device.

Static pressure ports attempt to pick up the pressure of undisturbed (static) air. However, on an aircraft moving through the air, the pressure picked up may be contaminated with dynamic pressure $(q = 0.5 \ V^2)$. One refers to the amount of contamination as "static pressure defect". Static pressure defect is principally a function of Mach number, angle-of-attack, and

314

side-slip. All new aircraft undergo
an elaborate calibration procedure
to measure the static pressure
defect, which can then be
compensated for in the air data
computer. Nevertheless, the
calibration is imperfect, and
residual static pressure defect
errors remain. These can be
especially large as the aircraft
approaches Mach 1 or actually
transits the transonic region.
False, instantaneous altitude jumps
of several hundred feet are not
uncommon in these instances.

Time lag is another type of pressure
sensing error. It is caused by the
volume of air contained in the input
end of the pressure sensor and the
tubing which connects it to the
static pressure port. As the port
pressure changes, air must flow in
or out of the tubing and pressure
sensor input chamber in order to
reach a new equilibrium. The larger
the volume of air and the smaller
the tubing (which restricts air
flow), the greater the time lag.
Typical air data installations have
time lags of a few tenths of a
second. The effect of this time lag
is to cause the barometric altimeter
to read low during a climb and high
during a dive.

Inertial Errors.

Inertial system errors in the
vertical channel are primarily
caused by the vertical
accelerometer. Gyro errors tend to
be small and temporary by
comparison, their predominant effect
being to couple a small amount of
any horizontal acceleration into the
vertical channel. Medium-accuracy
inertial navigation systems have
accelerometers with bias
repeatability of 50 to 100 ug and
scale factor accuracies of 50 to 100
ppm. These may vary somewhat with
time due to temperature changes.
Typical autocorrelation times for
such changes are of the order of
1000 seconds in ring laser gyro
(RLG) systems.

Gravity anomalies can also cause
errors which look like accelerometer
bias errors and can change values as
the aircraft moves over the earth.
A typical RMS value for gravity-
induced bias noise is 25 ug with a

correlation distance of 20 nmi.

2. Conventional Baro-Inertial
   Loops.

Figure 1 is a software block diagram
of the LN-93/94 vertical channel.
Note that it is divided into two
sets of computations: a fast loop
and a slow loop. The iteration rate
for the fast loop computations is 64
Hz; the slow loop computation rate
is 1 Hz. The function of the fast
loop is to integrate vertical
acceleration to generate vertical
velocity and short-term dynamic
changes in altitude. The slow-loop
computations contain the barometric
stabilizing feedback paths and their
gain-setting logic.

Fast Loop.

The fast-loop (64 Hz) computations
include the integration of inertial
acceleration and velocity to compute
vertical velocity and altitude
changes, respectively. In these
fast-loop mechanization equations
the terms ABZ and AFB (see Figure 1)
are treated as constants, but they
actually change at a once-per-second
rate determined by the slow-loop
computations.

After the double integration, which
generates inertial altitude H, there
is a first-order lag filter with a
time constant 1/KF equal to 1
second. This matches the 1-second
low-pass filter on the raw baro-
altitude input, which is needed to
accommodate applications providing
only very coarse (100 ft) resolution
of the baro-altitude data.

The last item computed in the fast
loop is the difference between the
filtered inertial and baro-
altitudes, HL and HB. The slow loop
computations use this difference as
an input only once per second, but
computing it at 64 Hz ensures that
the HL and HB used to form the
difference are synchronized to
within 1/64 second.

Slow Loop.

The slow-loop (1 Hz) computations
contain; (1) the gravity
compensation as a function of
altitude, (2) the third-order
feedback stabilization loops, (3)

the gain variation logic for setting the feedback gains, (4) a baro-bias compensation differentiator-integrator combination, and (5) a 16-second bleedoff of the baro-bias compensation integrator. The following paragraphs discuss each of these functions in detail.

Gravity Compensation and Stabilization Loops.

The H part of the feedback term $(K1/a)$ $(H-S2)$ to the acceleration node mechanizes the variation in gravity with altitude. This is the positive feedback which necessitates the stabilizing influence of the barometric altimeter. The $-S2$ part of that term is the negative feedback necessary to just bring the loop to a neutrally stable condition. The other three paths for S2 through KB2, KB3, and KB4 provide a stabilizing third-order feedback control system. The quantity S2 is the error signal for the third-order feedback loops. The gains KB2, KB3, and KB4 are chosen so that there is one real solution with a root at $-K/T_v$ and a pair of complex roots at $-K/T_v \pm jK/T_v$. Thus, the sinusoidal solution has a 0.707 damping ratio and decays at the same rate as does the real solution to provide good setting characteristics. The quantity Tv is the nominal loop time constant, while K is a parameter that can vary between 0 and 1 in order to vary the loop time constant from infinity to Tv, respectively.

Gain Variation Logic.

The original gain changing logic (Ref.4) for setting K was as follows:

$$K = 1 - KB5 \tag{1}$$

where

$$KB5 = Vz^2/(Vz^2 + DO^2) \tag{2}$$

and

$$DO = 30 \text{ fps}$$

The above logic makes K approach zero, effectively opening the feedback loops, whenever the aircraft is in a steep climb or dive. Under these conditions, the barometric altimeter input is suspect, and one should rely more, heavily on the inertial data. Subsequently, several problems arose

(Ref.5) which caused us to alter the logic somewhat. The first of these was the observation that sudden step changes of several hundred feet could occur in the barometric altimeter input when the aircraft approaches Mach 1 even though the aircraft is in straight and level flight. These steps can result from shock waves sweeping across the static pressure ports of the air data system. As the aircraft is in level flight, Vz=0, and the feedback loops are closed (K=1). Hence, these steps introduced large transients into the vertical loop.

To circumvent this problem, we first tried to replace Vz in Eq.(2) with HBdot, the rate of change of filtered barometric altitude. This effectively opened the loops when the steps occurred. However, this mechanization, employed in the LN-39, ran into difficulty during long, slow climbs or letdowns in the presence of nonstandard day lapse rate (barometric scale factor error) and large quantization (100 ft) steps in the barometric altimeter input. The final solution, therefore, was to return to the Vz formulation of Eq.(2), but to make a separate test comparing HBdot with Vz in order to reject step changes in barometric altitude. This test takes the following form.

IF

$$|\dot{HB} - Vz| > 2DO, \text{ then } KB5 = 1, \text{ else}$$
$$\text{use eq(2).} \tag{3}$$

Originally, the parameter DO (Eq.3) was fixed at 30 fps. This has the effect of opening the feedback loops whenever HBdot differs from Vz by more than 60 fps, Eq. (3), or whenever Vz is much larger than 30 fps, Eq.(2). During air combat maneuvers or repeated bombing runs, however, the aircraft may be climbing and diving almost continuously for long periods of time. Because of the unstable nature of the vertical inertial channel, we cannot leave the barometric feedback loop open for an appreciable length of time relative to the 560-second time constant of the instability. Otherwise, time variations in accelerometer bias and spatial variations in gravity anomalies would be sufficient to

cause unacceptable divergence in the vertical channel.

Consequently, in the LW-33 and LN-39 vertical channels, we caused D0 to increase gradually whenever Vz exceeded D0. The rate of increase in those systems was linear with time and at such a rate that D0 doubled in either 4 seconds (LW-33) or in 30 seconds (LN-39). These systems worked well for either severe vertical maneuvers or level flight, but long slow climbs and letdown in the presence of barometric scale factor errors still caused problems because the loop tended to track the barometric error rate too closely.

To alleviate this behavior in the LN-93/94 vertical channel, we delay the increase in D0 so that it doubles in 300 seconds. Furthermore, the increase occurs as a parabolic function of time, essentially flat to begin with and increasing rapidly after 300 seconds to ensure loop closure. Figure 2 displays this nonlinear increase in the parameter D0, which occurs whenever the magnitude of Vz exceeds D0.

Once loop closure is accomplished, as recognized by the magnitude of Vz being less than D0, then D0 decreases back to its minimum value of 30 fps at a rate of 0.5 ft/sec. Hence, if it took 300 seconds to close the loop by increasing D0 to 60 fps, it would take only 60 seconds to bring D0 back to 30 fps once the magnitude of Vz diminished and stayed below D0.

Baro-Bias Compensation.

Reducing the feedback gains during climbs and dives protects the feedback loops by blocking out, or at least greatly attenuating, changes in barometric altimeter error, but only during the climb or dive. When the aircraft levels out, the feedback loops will reclose, and any change in barometric error which still persists, as caused for example by a scale factor error in the altitude measurement, would now introduce a new loop transient into S2, which would take several minutes to damp out. The baro-bias compensation path through KB5

(Figure 1) overcomes this difficulty by accumulating the changes in barometric error when the feedback paths are open and subtracting them from the inertial-baro difference signal DHB0 before forming the feedback error signal S2. The tacit assumption here is that during climbs and dives the inertial data is more reliable than the barometric data and most of the difference between the two represents barometric error.

Observed changes in DHB0 are multiplied by KB5 and then summed in an integrator. The gain KB5 is the complement of K, the feedback gain parameter. When K is low, KB5 is high and vice versa. The output of the integrator represents the accumulated changes in baro-bias error and is subtracted from the inertial-baro difference signal, thus canceling out the change in baro-bias.

To see how this works, let us consider the case where the barometric altitude error HB suddenly changes by 400 ft in level flight. This causes KB5 to be set to 1, because $|HBdot - Vz|$ exceeds 60 fps. With KB5 = 1 and K = 0, the feedback loops open up and the full amount of the 400 ft step change in the baro input becomes stored in the summing integrator following KB5 (see Figure 1). This 400 ft, when subtracted from DHB0 to form S2, exactly cancels the 400 ft step in DHB0 so that S2 sees no change at all. Now when the feedback loops reclose, there is no transient to ring through the system. In the paragraph under Response to large Step in Barometric Altitude Error we will see simulation results of a similar example complete with time histories of several of the loop variables.

Bias Compensation Bleedoff.

When we first mechanized the baro-bias compensation scheme described in the previous section (Ref.4), we soon found that it tended to rectify. That is, during repeated climbs and dives the integrator following KB5 would build up an ever increasing compensation value because air-craft climb rates tend to be different than their descent

rates. As a result, KB5 would typically have higher values during dives than it would during climbs. The net effect was to cause the inertial altitude H to gradually diverge from the barometric altitude. This is not a serious problem from a system standpoint because the vertical velocity Vz is still valid, and the sole purpose of the baro-inertial channel is to produce an accurate vertical velocity. One should never use the altitude output from a baro-inertial loop when precision is required. In weapon delivery systems, altitude change should be obtained by integrating Vz directly outside the baro-inertial loop.

Nevertheless, for display purposes the vertical channel does output "inertial altitude," and the users are disturbed if this "inertial altitude" diverges from the barometric altitude. Hence, to prevent this divergent behavior we added a bleedoff path around the baro-bias compensation integrator with a 16-second time constant. This is the K/16 path from DHHB (Figure 1) back to the integrator's input. The value K/16 is really (1 - KB5)/16 wherein the 1/16 portion is the direct 16-second bleedoff. We shall soon see the reason for the KB5/16 portion.

When we bleed off DHHB, we have to do it in such a way that S2, the feedback error signal, remains undisturbed. We accomplish this by also feeding -1/16 of DHHB into AFB from whence it travels through the altitude integrator back to S2, thus canceling at the S2 node the direct bleedoff of DHHB back through its own integrator. The effect of this bleedoff cancellation is to drive the filtered inertial altitude HL back to the filtered barometric altitude HB. However, this bleedoff cancellation path can also go up through "s" and KB5 to the input of the DHHB integrator. To cancel this effect, an additional bleedoff - KB5/16 is added to the DHHB direct bleedoff, making its total (1 - KB5)/16, which is the same as K/16.

There is still another secondary bleedoff path from the altitude output, H - S2, back to the gravity compensation which enters the K1/a

block (Figure 1). We cancel this path by feeding DHHB/16 through an integrator (to match the vertical velocity to altitude integrator in the other path) to the K1/a block. This forestalls any change in gravity compensation which would otherwise occur when the inertial altitude is forced to track a possibly erroneous barometric altitude.

Vertical Velocity and Altitude Outputs.

As previously stated, the primary purpose of the baro-inertial loop is to provide accurate vertical velocity. To this end we output Vz and not Hdot (see Figure 1) for vertical velocity. The quantity Hdot can be corrupted by barometric altimeter noise coming directly through KB2 in 1-second steps, whereas Vz is isolated by the smoothing effect of at least one stage of integration from any baro noise. Hdot can exhibit discontinuities (steps) at 1-second intervals while Vz at most may have slope discontinuities at 1-second intervals.

The "altitude output" H-S2, tracks the barometric altitude and should not be used where precision altitude changes are required as in weapon delivery systems. For one thing, H-S2 is not necessarily the integral of Vz. Weapon delivery systems should only use the Vz output and integrate it externally to the baro-inertial loop. This open loop integration scheme is the only way to ensure that changes in altitude indeed represent the integral of vertical velocity.

The reason for listing H-S2 as the altitude output is as follows: Dynamically, H is a better measure of altitude, whereas statically the raw baro altitude HBR is preferable. The error signal S2 indicates the low frequency difference between H and HBR. Hence, the quantity H - S2 provides an altitude measure both with good dynamic response as well as good static accuracy. Figure 3 provides a good example. This figure shows the loop response to an initial S2 error of -400 ft created by initializing HBR and HB to 400 ft but initializing H and HL

to zero. Note that H approaches 400 ft with the response characteristics of the third-order feedback loop. However, H - S2 immediately achieves a value only slightly above 400 ft and then slowly approaches 400 ft in an exponential fashion.

One undesirable feature of the H - S2 output is that S2 is computed only once per second. This can introduce step changes into H - S2 at one-second intervals. For display purposes this is not serious, and as stated earlier H - S2 should never be used for precision altitude applications anyway.

In spite of our protestations that the "altitude output" is unsuitable for weapon delivery, some users insist on using it anyway. To try to accommodate these people, we added a pseudo open-loop Vz integrator in the LN-93. The LN-94 still retains H - S2 altitude output. Figure 4 is a block diagram of the altitude output computation in the LN-93. To prevent a divergent output due to integration errors, we wrap a first-order feedback loop around the integrator with a very long time constant. This keeps the altitude output very loosely coupled to the barometric altitude so that for short time intervals the altitude output is very nearly the integral of the vertical velocity Vz. This pseudo open-loop integration process is entirely outside the baro-inertial loop and does not affect the vertical velocity accuracy.

Vertical Loop Corrections.

USAF specification SNU 84-1 requires that the baro-inertial loop accept corrections to (1) vertical velocity, (2) inertial altitude, and (3) barometric altitude. Such corrections may be made as a result of GPS updates, for example. Figure 1 shows where these intermittent corrections will enter the baro-inertial loop. In order not to disturb the loop, it is important that the inertial altitude correction and the baro-altitude correction be equal and applied simultaneously.

Simulations of Vertical Loop Performance.

This section contains simulation results showing the performance of the LN-93/94 baro-inertial loop when subjected to various barometric disturbances. The first simulation is the response to a large step change in barometric altimeter error. This simulation demonstrates the effectiveness of the baro-bias compensation and bleedoff network in squelching this type of barometric altimeter error. The second simulation shows the effect of a 30-second burst of barometric altimeter noise as might be encountered when an aircraft traverses the transonic region.

The third simulation displays the loop behavior during and following a steep dive and high-g pullout. The final set of simulations explores the most difficult flight regime for the baro-inertial loop, namely, long, slow descents in the presence of barometric scale factor errors.

In all of these simulations, we presume that the barometric altimeter input is quantized in 100-ft increments and has a 0.3-second time lag. Furthermore, the altimeter scale factor error is modeled as being -5 percent at sea level and varying linearly to +5 percent at 52,000 ft. This is fairly representative of density variations in the real atmosphere under nonstandard day conditions (Ref.10).

Response to Large Step In Barometric Altitude Error.

To test the effectiveness of the baro-bias compensation and bleedoff networks, let us introduce a step change in altimeter bias error of 400 ft. This differs from the earlier simulation (Figure 3) in that, instead of initializing the feedback error signal S2 to - 400 ft, this time S2 is initially zero, the loop is completely settled, and then we introduce a step into the altimeter bias. In this simulation, the aircraft is in level flight (Vz = 0) and the baro bias step enters at time t = 10 seconds. We then watch the loop response for another 190 seconds. Figure 5 shows

the Vz and S2 responses as well as the DHHB (output of the baro-bias compensation integrator) and H-S2 responses.

Note that neither VZ nor S2 are significantly affected by the 400-ft step in baro bias and that DHHB jumps immediately to -335 ft to cancel most of the step input before it can reach S2. Thereafter it bleeds off exponentially with a 16-second time constant. During the bleedoff, H-S2 builds up to a value of 400 ft to match the new baro input value. The reason that DHHB does not jump all the way to -400 ft is that the 1-second filter on the baro input reduces the peak in HBdot somewhat.

This example serves to emphasize the point that the "altitude output," H-S2, is not necessarily the integral of vertical velocity Vz. In this case, H-S2 changed by 400 ft whereas Vz remained essentially zero.

Response to 30-Second Burst of Altimeter Noise.

When an aircraft penetrates the transonic regime, large variations in the static pressure sensed by the altimeter can occur. Figure 6 shows the response of the LN-93/94 baro-inertial loop when subjected to a 30-second burst of altimeter white noise of magnitude 800 ft peak-to-peak. For simulation purposes, the altimeter error is randomized between -400 ft and +400 ft with any value within that range being equally likely. Figure 6a shows a sample of one such 30-second burst. After 30 seconds, the altimeter error is zeroed, but we continue the simulation for another 170 seconds to observe the vertical velocity transient effects caused by the noise burst.

Figure 6b displays a set of 10 Monte Carlo runs each with a different set of random numbers to generate the 30-second burst of simulated white noise. In the worst case of these 10 runs, the vertical velocity error reaches a peak value of 1.8 fps. The RMS peak value over these 10 runs is 0.93 fps.

Response to Steep Dive and Pullout.

This type of maneuver frequently appears in the specification for a vertical channel, and for this reason we include it. However, we shall see shortly that a shallow climb or dive is a more severe test of vertical channel accuracy. The maneuver simulated here starts with the aircraft flying level at 30,000 ft. It then pulls down for 5 seconds at 4.7 g to a 750 fps rate of descent.

After 35 seconds the aircraft pulls up at 4.7 g for 5 seconds to level off at sea level. Elapsed time from level flight at 30,000 ft to level flight at sea level is 45 seconds. We then continue the simulation for another 155 seconds to observe any overshoot which might occur.

Figure 7d shows the vertical trajectory as a function of time. Figures 7a and 7b display the vertical velocity error and the feedback error signal S2. In Figure 7c we see the gain cutoff parameter HBdot-Vz in comparison with its cutoff value 2(D0). We see from this last figure that |HBdot-Vz |exceeds 2(D0) in both the high-g pull down and pullout, so that K=0 and the feedback loops are wide open. During the dive itself the high rate of descent makes K very close to zero, and the feedback paths are still essentially open. Hence both S2 and the vertical velocity error remain very small until the loops reclose at time T = 45 seconds. Then we see a small transient in both S2 and in Vz, but the vertical velocity error over-shoot is still only 0.25 fps.

Response to Shallow Dives.

These simulations also start with the aircraft flying level at 30,000 ft. We simulated 6 different rates of descent - 30 fps, 40 fps, 50 fps, 60 fps, 75 fps, and 100 fps - in order to find the worst case which yields the largest vertical velocity error. In each case the aircraft descends from 30,000 ft to sea level using 1 second to establish the rate of descent and 1 second to pull out at sea level. Because these rates are of the same order of magnitude as the gain variation parameter D0, the feedback loops are still closed, albeit with a somewhat longer that

320

nominal loop time constant. Consequently, the vertical velocity error will tend to track the rate of change of barometric altimeter error. If the baro scale factor error is 5 percent and the rate of descent is 60 fps for example the baro error rate would be 3 fps. For slower rates of descent the baro error rate would be slower, but would persist for a longer time because the aircraft takes longer to reach sea level. Conversely, for faster descents the baro error rate is higher, but for a shorter time. We expect the worst-case descent rate, i.e. the one that causes the largest vertical velocity error, to be between 30 fps and 100 fps, and that is the reason for simulating six different rates covering this range.

Figures 8a through 8d show the vertical velocity error during and shortly after the 30,000-ft descent. Note that the worst-case peak value is almost 3 fps. This is for the 60-fps case wherein the baro error rate at sea level is indeed 3 fps (5 percent of 60 fps). Figure 9 is a plot of the peak vertical velocity errors taken from Figure 8 as a function of the descent rate. We see from this graph that a 60-fps rate of descent represents the worst-case situation for this vertical channel mechanization.

Because the real culprit here is the baro scale factor error and the descent times are of the same order of magnitude as the 560-second time constant of the basic instability of the vertical inertial channel, there is not much more that we can do to the baro-inertial loop itself to reduce these errors. One way of improving the vertical velocity performance is to eliminate the baro scale factor error in the air data computer. This can be accomplished through use of a Blanchard algorithm (Ref.15) which uses air temperature as well as static pressure to determine air density and then integrates the reciprocal of air density (scaled by gravity) with respect to pressure to calculate altitude.

One might also ask if slow climb rates produce the same errors as those seen for slow descents in

Figure 8. The answer is no, because the scale factor error is largest at sea level, diminishes to zero at 26,000 ft. and reverses sign above that altitude. Hence, in slow climb starting from sea level, the parameter D0 is smallest (30 fps) during the greatest baro error rate, and the baro-inertial loop can do a better job of combating the error. In the case of descent, the aircraft encounters the highest baro error rate at the end of the long, slow descent and D0 has built up to the same value as the descent rate itself, making K = 0.5. Hence, the vertical velocity error after a 60-fps descent from 30,000 ft to sea level is greater than it is after a 60-fps ascent from sea level to 30,000 ft. Figure 10 verifies this fact by showing the vertical velocity error during and after a 60-fps ascent from sea level to 30,000 ft. The peak vertical velocity in this case is -1.9 fps as compared with the 2.9-fps value observed in the descent case (Figure 8d).

3. **Kalman Filter Baro-inertial Loop.**

The increasing availability of GPS for altitude updating makes it attractive to consider a Kalman filter mechanization of the vertical channel. The mechanization presented here is the same as that published in Reference 16.

Reference 13 provides an approach to designing a vertical channel Kalman filter. It uses 5 states: inertial altitude error, vertical velocity error, accelerometer bias error, baro-scale factor error and pressure altitude error. In this filter, baro-scale factor is modeled as a Markov process with a 10,000-second time constant, which allows for some slow variation with time.

Reference 10 indicates that air density deviations from the standard atmosphere tend to reverse sign at about 26,000 ft. This suggests that a better model of baro-scale factor error would be:

$$k=k_0 (1-h_t/h_0) \qquad (4)$$

where k is the scale factor error at true altitude h_t, k_0 is the scale

factor error at sea level, and

$h_0 = 26,000$ ft.

With this model, the error in pressure altitude, hp is

$$\int h_P = \int h_b + \int_0^{hi} k\, dh_t = \int h_b + k_0 h_t (1 - h_t/2h_0) \quad (5)$$

where $\int h_b$ is the error in pressure altitude at sea level. We see from Eq (5) that this model requires two pressure altimeter error states: a bias state, $\int h_b$ and a state we shall call the sea level scale factor error, k_0.

In this next section we describe the resulting Kalman filter mechanization using Eq (5) as the model for pressure altimeter error. Table 1 summarizes the mechanization.

4. Analysis

State Equations.

As suggested in Ref 13, we will use 5 states for the vertical channel error model. Three of these are the same as in Ref 13: inertial altitude, vertical velocity, and accelerometer bias. However, we will replace the pressure altitude state with an altimeter bias state, and instead of a scale factor error independent of altitude, we will use Eq (4). That state variable will be k_0 the sea level value of the scale factor error

The state equations are:

$$\begin{aligned}
\int \dot{h}_i &= \int Vz \\
\int \dot{V}z &= \int a_z + 2(g/a_c)hi \\
\int \dot{a}_z &= -10^{-3} \int az \quad (6) \\
\dot{k}_0 &= -10^{-4}(0.5 + v_h/v_0)k_0 \\
\int \dot{h}_b &= -10^{-4}(0.5 + v_h/v_0)\int h_b
\end{aligned}$$

where

a_c = Earth's Astronomic Radius in feet
g = Gravity in ft/sec^2

v_h = Magnitude of aircraft horizontal velocity in fps
V_0 = 500 fps
In Eq (6), the term $2(g/a_e)hi$ is the positive feedback term due to the altitude-dependent portion of the gravity compensation. Eq (6) models

the accelerometer bias error as a Markov process with a time constant of 1000 seconds. The $-10^{-4}(0.5 + V_b/V_o)k_o$ term is similar to the Markov process with a 10^4-second time constant used in Ref 3, but modified here to be a function of horizontal velocity. This effectively makes the change in k_o somewhat spatially dependent as well as time dependent. The state equation for h_b is similar and uses the same correlation time as that for the ko state.

Equation (6) leads us to the dynamics matrix (F-Matrix) given in Table 1.

Covariance Matrix.

Table 1 lists the initial values for the covariance matrix. Note that the off-diagonal element which represents the correlation between pressure altitude and inertial altitude is initialized with a unity correlation coefficient. This presumes that inertial altitude is initially set equal to the pressure altitude, so that their errors are equal.

Table 1 enumerates the Q-Matrix elements which augment the covariance diagonals at each update cycle. Small amounts of "plant noise" or Q are added to the inertial altitude and velocity states to model computer resolution. Additionally, an amount of 0.125×10^{-6} (ΔVz^2) Δt_k is added to the inertial velocity error state to model vertical accelerometer scale factor error, where V_z is the change in vertical velocity from one Kalman filter update to the next, and t_k is the time between updates.

To model accelerometer bias shifts, g^2 errors, and gravity anomalies, we set Q_{33} equal to $(2+V_h/V_o) \times 10^{-9} ft^2/sec^5$. The constant part of Q_{33} models the accelerometer errors while the $V_h/V_o \times 10^{-9}$ ft^2/sec^5 portion models gravity anomalies which may change as the aircraft moves over the earth.

For baro scale factor error plant noise, we use a value Q_{44} equal to $0.2 \times 10^{-6} + (Vx/V_o)^2 \times 10^{-5}$ sec^{-1}. The first term is comparable to the

value of $0.18 \times 10^{-6} sec^{-1}$ used in Ref 13 and the second term allows for modeling errors in case the real scale factor deviations from the standard atmosphere do not conform exactly to our model with its linear dependence with altitude.

For pressure altitude error variance plant noise, Q_{55}, Ref 13 recommends 4×10^{-6} ft^2 per each 10-second update cycle. This seems unrealistically small. The same reference states that the isobaric surfaces typically are sloped relative to the horizontal plane with a standard deviation of 1 ft per nautical mile. At 500 fps, an aircraft would travel almost a mile in 10 seconds, and one should expect to add approximately $(1 \ ft)^2$ to C_{55} each 10-second period. The actual value for Q_{55} listed in Table 1 is $(1 + 0.5 \ Vh /Vo) ft^2/sec$. The 1 ft^2/sec portion of this term accounts for a time variation in pressure at a fixed place, while the $0.5 \ V_h/V_o$ portion accounts for the spatial variation in pressure as the aircraft moves to another place.

Measurement Equations.

Table 1 shows the measurement matrix (H-Matrix) for altimeter updating. Its derivation is as follows: If h_t represents the true altitude, then the inertial altitude is:

$$h_i = h_t + \delta h_i \qquad (7)$$

Similarly, the pressure altitude is:

$$h_p = h_t + h_t (1-h_t/2h_o) \ k_o + \delta h_b \qquad (8)$$

Differencing h_t and h_p to form the error signal, we have

$$E = h_i - h_p = \delta h_i - h_t(1-h_t/2h_o)k_o - \delta h_b \qquad (9)$$

Because we do not know the true altitude in flight, we approximate it with the inertial altitude h_i to form

$$E = \delta h_i - h_i (1-h_i/2h_o)k_o - \delta h_b \qquad (10)$$

The H-matrix given in Table 1 follows directly from Eq (10) by inspection.

Measurement Noise.

The R-matrix listed in Table 1 consists of two parts. The first part, $(0.12)^2 (0.6 \ h_i^2 \times 10^{-7} + 27.8)^2$, comes from Ref (13) and represents 0.12 millibars of pressure noise converted to altitude noise by the conversion factor $(0.6 \ h_i^2 \times 10^{-7} + 27.8)$. Actually, Ref (13) recommended using 0.12 millibars of noise for vertical velocity magnitude less than 25 ft/sec and 0.32 millibars of noise for vertical velocity magnitude in excess of 25 ft/sec. The second part of the R-matrix, $(0.3 \ V_z,)^2$, adds additional measurement noise during high rates of climb or dive. The 0.3 V_z value models a 0.3-sec altimeter time lag. At $|Vz| = 25$ fps, the amount of equivalent pressure noise added by this term is 0.23 millibar. When added to the 0.12 millibar figure for the first term, the total is 0.35 millibar, which is close to the Ref 13 recommended value of 0.32 millibar for vertical speeds in excess of 25 fps.

Validity Test.

One problem that arises in baro-inertial loops is the possibility of large (~500 ft) step changes in the barometric altimeter input. These changes can occur near Mach 1 as shock waves sweep across the static pressure ports of the air data system. Conventional vertical loops (Ref 9) handle this situation by drastically reducing the feedback gains and by computing a baro-bias compensation. In effect, when the baro-inertial error signal is excessively large, one assumes that the inertial signal is correct and that the problem is in the barometric signal.

We can use the same philosophy to devise a strategy in case of failure of the validity test in our Kalman filter vertical channel. For example, consider the following strategy: If $E^2 > HCH^T + R$ then set $HCH^T + R = E^2$ and after the update, add $E^2 - R$ to C_{55} (except during absolute altitude updates on the ground).

The reasoning behind the above approach is as follows: Because the

reason for failing the validity test is presumed to be a sudden change in the altimeter bias, we acknowledge this fact by increasing the covariance of the altimeter bias state (after the update) by the amount of the observed error squared less the measurement noise. During the update itself, we reduce the feedback gains by setting the denominator ($HCH^T + R$) in the gain equation equal to E^2, the square of the observed error. This is equivalent to adding additional measurement noise, R, to account for the observed error difference.

5. Simulations

Description of Simulation Program.

Figure 11 shows an overall block diagram of the simulation program used to simulate the Kalman filter vertical channel. A desktop computer was used to perform the simulations. The Kalman filter update interval was one update every 8 seconds, although some runs were made with 4-second and 2-second update intervals. Figure 12 shows little difference in vertical velocity accuracy as a function of update rate, and for that reason we chose the slowest of the three rates for the remaining simulations.

To initialize the filter, we had it perform absolute altitude updates for the first 98 sec and then switch to barometric altitude updates. The absolute altitude updates employ the same Kalman filter except that the H-Matrix, observation error signal, and measurement noise are changed to:

$$H = [1\ 0\ 0\ 0\ 0\ ] \qquad (11)$$
$$E = h_i - h_t \qquad (12)$$
$$R = (10\ ft)^2 \qquad (13)$$

This procedure simulates turning the system on when the aircraft is on the ground at a known elevation above sea level

Truth Model.

The simulation truth models included the following:

    a.   An initial baro-bias error of 500 ft

    b.   Sea level baro scale factor error of 5 percent

    c.   Initial vertical velocity error of 1 fps

    d.   Accelerometer bias error of 0.001 ft/sec^2

    e.   Accelerometer scale factor error of 400 ppm

    f.   Baro-altimeter resolution of 2.5 ft

    g.   Baro-altimeter time lag of 0.3 sec

    h.   Baro-scale factor nonlinearity with altitude per Figure 13.

If we simulated the real-time barometric scale factor as a linear function of altitude, the truth model would match the Kalman model exactly, and we might achieve unrealistically good results. To avoid over-optimism, we introduced a modeling error by assuming a nonlinear variation in barometric scale factor error with altitude. The simulated nonlinear truth model for pressure altitude is

$$h_p = (k_o t + .04\ \beta/h_o)\beta + dh_b$$

$$\text{where } \beta = h_t\ (1 - h_t/2h_o) \qquad (14)$$

and k_{ot} is true barometric scale factor error at sea level.

With this formulation for the baro error model, the baro scale factor error variation with altitude is

$$\frac{d h_p}{d h_t} = (k_o t + .8\ \beta/h_o)\frac{d\beta}{d h_t}$$

$$\text{and } \frac{d\beta}{d h_t} = 1 - h_t/h_o \qquad (15)$$

Combining the above equations, we have,

$$\frac{d h_p}{d h_t} = (Ko\ t + .8\beta/h_o)(1 - h_t/h_o) \qquad (16)$$

A plot of Eq (16) vs altitude appears in Figure 13.

Simulation Results.

The above error model was subjected to several different vertical maneuvers and barometric anomalies.

These were: (1) a series of fast climb and dive maneuvers, (2) a fast climb followed by a slow descent, and (3) a series of sudden steps in baro-bias. A discussion of these simulations and their results follows.

Fast Climbs and Dives.

In this simulation, after ground alignment, the aircraft takes off and pulls up at 150 ft/sec^2 (4.7g) to a 750 fps rate of climb. At t = 140 seconds it pulls down at 150 ft/qec^2 (4.7g) to level flight at 30,000 ft altitude. At t = 200 seconds it pulls down in 5 seconds to a 750 fps dive. At t = 240 seconds it begins a 4.7g pullout to level flight at zero altitude. This climb and dive maneuver is repeated every 200 seconds as shown in Figure 14a.

Figure 14b shows both the vertical velocity error and the covariance estimate of the vertical velocity error. The dashed, symmetric curves are the plus and minus values of the square root of the vertical velocity covariance, C_{22}. The solid third curve is the vertical velocity error itself, which should and does stay between the $\pm$ 1-sigma error values calculated by the filter.

The inertial altitude error appears in Figure 14c along with the plus and minus values of the filter estimates of the RMS altitude accuracy. The initial absolute altitude updates bring the altitude error down to about 10 ft. After 700 seconds of climbing and diving maneuvers, the altitude error gradually increases to about 50 ft and remains within the $\pm$1-sigma values calculated by the Kalman filter.

Figure 14d is a plot of the sea level baro scale factor error, i.e. the difference between S_4, the value estimated by the Kalman filter, and the true value, 0.05 in this example. Also shown (dashed curves) is the 1-sigma range of the error in S_4 as determined by the square root of the covariance element C_{44}. After the first climb to altitude, the filter correctly estimates baro scale factor to about one-percent accuracy, and the error is in general agreement with the filter

covariance estimates. In the subsequent maneuvers the error increases somewhat to about 2.5 percent after 700 seconds of climbing and diving.

At times, especially during the level flight legs at sea level, the error in the filter estimate of k_o exceeds its one-sigma value as calculated from SQRT (C_{44}), and the whole curve is biased to the positive side of zero. The discrepancy is due to the mismodeling of the baro scale factor as shown in Figure 13. The filter attempts to fit a straight line through the actual nonlinear function of altitude. As readily seen from Figure 13, a best straight-line fit through the truth model will intercept the zero altitude axis at a value somewhat higher than 0.05. This causes the bias offset observed in Figure 14d.

In Figure 14e we see the error in the filter's estimate of baro bias in comparison with its expected RMS value as calculated by the filter. The spike-like increases in SQRT (C_{55}) result from failing the validity test during some of the high-g pullup or pulldown maneuvers. These increases help to maintain the $\pm$1-sigma range outside of the actual error. This imparts an adaptive behavior to the filter, which helps prevent divergence.

Finally, in Figure 14f we see that the filter reduces accelerometer bias error from about 100 ug to about 50 ug, and the actual error stays comfortably within the $\pm$1-sigma limits $\pm$SQRT (C_{33}) calculated by the filter.

Fast Climb and Slow Descent.

This simulation starts out the same as the last one, but after reaching 30,000 ft, the aircraft then descends slowly at 60 fps back down to sea level (Figure 15a). This slow descent represents the worst case for conventional baro-inertial loops (Ref 9) and also for the Kalman filter vertical channel of Ref 13. In these mechanizations, the slow descent case created peak

velocity errors of 2.9 fps (Ref 9) and 1.8 fps (Ref 13). These errors peaked out near the end of the descent.

In contrast, the mechanization proposed herein limits the peak vertical velocity error to about 0.5 fps, and this occurs early in the descent (see Figure 15b). The reason for this exceptional performance is that the Kalman filter estimates the baro scale factor error during the climbout and can then compensate for the baro scale factor error during the descent.

Figure 15c shows the inertial altitude error. It also peaks out early in the descent phase at about 80 ft and thereafter diminishes as the filter compensates for the baro scale factor error. In Ref 13, a similar gradual descent generated ten times as much altitude error.

Figure 15d illustrates the ability of the filter to estimate baro scale factor error. Again, it shows both the actual error as well as the $\pm$1-sigma values for the error as calculated by the filter. This figure shows that the filter estimates sea level baro bias error to about one-percent accuracy, and the filter 1-sigma estimates are in agreement with the actual error itself. Once again however, the actual error is biased somewhat to the positive side because of the baro scale factor modeling error (Figure 13).

Figure 15e shows the baro bias error relative to the true value of 500 ft and compares it with the filter estimated value of the expected RMS error, $\pm$SQRT(C_{55}). The baro bias error in this example stays within the $\pm$SQRT(C_{55}) range throughout the maneuver. The spike in SQRT(C_{55}) at T=140 seconds is caused by failing the validity test during the high-g pulldown to level flight at 30,000 ft.

In Figure 15e we see the ability of the filter to estimate accelerometer bias. It shows a gradually improving estimate as judged by the narrowing range between the $\pm$1-sigma values. The accelerometer bias error itself remains between the

filter $\pm$1-sigma values and diminishes from almost 100 ug at the start of alignment to about 30ug at the end of the fast climb and slow descent maneuver.

Steps in Baro-Bias.

In this simulation, we start off with zero baro-bias, but at t=100 sec we suddenly change it to 500 ft. Then at t=200 sec we reset the baro-bias to zero. We keep repeating this pattern every 200 sec as indicated in Figure 16a.

Figure 16b shows the resulting vertical velocity error and compares it with the filter estimated $\pm$1-sigma values of the velocity error. The actual velocity error is less than 0.2 fps throughout this simulation (after alignment) and stays within the filter estimated 1-sigma range.

The inertial altitude error caused by the baro bias steps appears in Figure 16c along with the filter estimated $\pm$1- sigma values for the inertial altitude error. Note that the altitude error gradually increases to about 30 ft after the eighth bias step. This error is consistent with the $\pm$40 ft 1-sigma range calculated by the filter.

Filter 16d is a plot of the plus and minus RMS values for the baro bias error as computed by the filter. The actual baro bias error itself appears in Figure 15e. We separated these two graphs because it would be difficult to distinguish one set of lines from the other if we had plotted them together as in the previous figures. Note that the 500-ft error introduced at each bias step is almost completely corrected at the very next Kalman filter update. The 500-ft stepup in the baro altitude RMS magnitude at the first update following each bias step is the result of failing the validity test. This stepup in the filter 1-sigma value persists only until the next update, at which time it returns to nearly the same low value it had before the step occurred. This shows that the filter is able to follow and correct for the sudden jumps in baro-bias.

The error in sea level baro scale

factor error, k_o, appears in Figure 16f along with the filter RMS estimates, $\pm$SQRT (C_{44}). Because the actual altitude remains at sea level throughout this simulation, the filter cannot estimate k_o. However, because the filter computes H_4 from the inertial altitude instead of true altitude, and the inertial altitude deviates slightly from zero (Figure 16c), the filter computes a nonzero H_4 and thinks it cal update k_o. Hence, the $\pm$SQRT (C_{33}) curves tend to converge slightly instead of staying flat at their initial values of ± 0.05.

The last figure in this set, Figure 16g, shows the accelerometer bias error plotted along with its filter computed 1-sigma values, $\pm$SQRT (C_{33}). The baro bias steps have no appreciable effect, and the filter is able to update the accelerometer bias to within 25ug of its correct value. At the same time the filter's RMS estimate of the baro bias accuracy is consistent with the actual error.

6. Conclusion.

The Kalman filter vertical channel design summarized in Table 1 exhibits much superior performance in comparison with either the LN-93/94 conventional baro-inertial loop (Ref 9) or the Kalman filter design proposed in Ref 13. It obtains this superior performance by modeling the baro scale factor error as a linear function of altitude, a relationship which the real atmosphere tends to follow.

7. Nomenclature.

ABZ term	Acceleration feedback
AFB term	Altitude rate feedback
a or a_e	Earth's radius
a_z	Vertical accelerometer bias error
C	Covariance matrix
DVZ	Vertical Velocity increments from vertical acceleromleter
DHBO	Difference between filtered baro and filtered inertial altitudes
E	Error signal, $h_i - h_p$
F	Dynamics matrix
g	Gravity
g_o	Gravity at sea level on the equator
H	Inertial altitude in conventional filter. Observation matrix in Kalman filter.
Hdot or $\dot{H}$	Time rate of change of H
HB	Filtered baro altitude
Hbdot or $\dot{HB}$	Time rate of change of HB
HBR or h_b	Raw baro altitude input
h_i	Inertial altitude in Kalman filter
HL	Filtered inertial altitude
h_o	Isopycnic level, 26,000 ft
h_p	Pressure altitude ft
h_t	True altitude ft
k	Baro scale factor error
k_o	Value of k at sea level
K	Gain changing parameter = 1 - KB5
K1	$2\ g_o$
K2	0.005302
KB2	Velocity feedback gain
KB3	Acceleration feedback gain
KB4	Integral acceleration feedback gain
KB5	Gain change parameter computed as $= V_{z2}/(V_{z2}+DO^2)$
KF	1 sec^{-1}
Q	Plant noise matrix
R	Measurement noise
S	Laplace operator. S signifies differentiation; 1/S signifies integration. Represents state vector in Kalman filter.
S2	Feedback error signal
t	Time
	Time between Kalman updates
T_v	Nominal loop time constant = 20 sec during alignment and 100 sec thereafter
V_h	Magnitude of aircraft horizontal velocity
V_o	500 fps
V_z	Inertial vertical

	velocity
ΔV_z	Change in vertical velocity between Kalman updates
β	$h_t (1 - h_t/2\, h_o)$

8. References.

1. K.L. Perkins, R.R. Palmer and J.S. Ausman, Vertical Velocity Measuring System, U.S. Patent 3,005,348, October 24, 1961

2. M. Kayton and W.R. Fried, Avionics Navigation Systems, John Wiley & Sons, New York, 1969, pp 317-319

3. W.S. Widnall and P.K Sinha, Optimizing the Gains of the Baro-Inertial Vertical Channel, AIAA Guidance and Control Conference, August 1978, Palo Alto, CA

4. J.S. Ausman, et al., Close Air Support System (CLASS) F-4D Flight Test,, Vol I Summary, Air Force Avionics Laboratory, TR-73-363, WPAFB, Ohio, September 1973, pp 96-102

5. J.S. Ausman and F.J. Hellings, Development of a System for Scoring Simulated Bombing Runs, NATO AGARD-CP-198, October 1975, pp 25-1 to 25-19

6. C.H. Leatherbury, J.S. Ausman and H.E. Thompson, Development and Test of an Airborne Range Instrumentation System (ARIS), Proceedings of the 23rd International Instrumentation Symposium, May 1977, pp 269-278

7. M. Weintraub, LW-33 Advanced Inertial Nav/Attack System for Improving the Effectiveness of Combat Aircraft, Section 5 International and foreign Development, 12th Joint Services Data Exchange for Inertial Systems, October 1978, Norfolk, VA

8. LN-39 Standard Inertial Navigation system AN/ASN-141 Technical Description, Litton Document No. 15556H, September 1987

9. J.S. Ausman, Baro-Inertial Loop for the Standard RLG INU, Proceedings of the Institute of Navigation, National Technical Meeting, Phoenix, AZ, January 22-24 1991, pp 303-308

10. U.S. Standard Atmosphere, 1976, NOAA-S/T 76-1562, October 1976

11. R.L. Blanchard, An Improvement to an Algorithm for Computing Aircraft Reference Altitude, IEEE Transactions on Aerospace and Electronic Systems, September 1972, pp 685-687

12. S.L. Valley, Scientific Ed., Handbook of Geophysics and Space Environments, New York; McGraw-Hill, April 1965

13. P.M. Barham and P. Manville, Application of Kalman Filtering to a Baro/Inertial Height Systems, Royal Aircraft Establishment, Technical Report 69131, June 1969.

14. M.J.A. Ashbury, D.A. Forrester, C.S. Dixon and R. Johannessen, Probability Distributions that are Important when Assessing Barometric Aiding to GPS, Proceedings of Institute of Navigation GPS-90 Meeting, Colorado Springs, CO, September 19-21, 1990, pp 445-453

15. R.L. Blanchard, A New Alogorithm for Computing Inertial Altitude and Vertical Velocity, IEEE Transactions on Aerospace and Electronic Systems, Vol AES-7, No. 6, November 1971, pp 1143-1146

16. J.S. Ausman, A Kalman Filter Mechanization for the Baro-Inertial Vertical Channel, Proceedings of the Institute of Navigation Forty-Seventh Annual Meeting, Williamsburg, VA, June 10-12, 1991, pp 153-159.

SECTION V
REPRESENTATIVE VEHICLE IMPLEMENTATIONS
INTRODUCTION
DR. JOHN NIEMELA

This section describes the navigation suites of a diverse though representative spectrum of aerospace vehicles - fixed and rotary wing aircraft as well as spacecraft. Each aerospace vehicle is configured to meet dissimilar performance requirements, environment constraints and mission profiles.

However, there is a common methodology in the requirement analysis, sensor selection and system synthesis used in the development of the navigation system for each aerospace vehicle type. It is the intent of this section to describe generic classes of navigation systems with emphasis on the technical and operation factors that influence their design.

Aerospace Navigation Systems Requirements for Fixed Wing Aircraft

by
Carlos A. Bedoya
McDonnell Douglas Aerospace – East
St. Louis, MO 63034
USA

Introduction

The purpose of this paper is fourfold, (1) to provide the reader with a status report on the state-of-the-art of navigation system technology as applied to fixed wing aircraft, (2) recommend a disciplined systems engineering process to be used in determining navigation system requirements as they relate to navigation, sensor cueing and targeting as well as weapon requirements (3) examples of how to functionally decompose each aircraft and analyze its mission requirements into the navigation system requirements and (4) A typical set of requirements for a fixed wing aircraft. The results will allow the navigation systems designer to have a process he can follow in determining or verifying the requirements of a particular application as well as a baseline set of requirements..

1.0 Navigation Systems Current Status

Aerospace navigation systems have undergone revolutionary changes in the past fifteen years (References 1, 3, 5, 10, 12, 13, 14, 16, 18, 19. Major developments have occurred in the inertial navigation Systems (INS) and satellite based navigation systems with the advent of the ring laser gyroscope and the development of the Global Positioning System (GPS) as well as the explosion in the field of microprocessors and integrated circuits. Electromechanical gimbaled inertial navigation systems using mechanical gyroscopes have been replaced by Strapdown systems using ring laser gyros and powerful microprocessors providing equivalent accuracy and twenty fold increases in reliability. GPS has begun to provide world wide satellite based navigation to military and civilian users with unprecedented accuracy in position, velocity and time. It has the potential to replace all current radio navigation aids by the turn of the century. In addition to in-creases in performance, these developments have significantly reduced the weight and the cost as well as increased the reliability and maintainability of state-of-the-art aerospace navigation systems (Fig 1.0-1).

The increased demands in performance, reliability, maintainability and decreased weight and cost of fixed wing aircraft have also brought significant changes to the architecture of avionics and flight control. Federated, independent, dedicated systems have evolved to more and more integrated systems with common interfaces such as standard multiplexed buses, multipurpose displays and controls, integrated functions and integrated controls. Standard navigation systems such as the United States Air Force Standard Navigation Unit (SNU-84-1) the Fighter Navigation Unit (FNU-85), the United States Navy Carrier Align Inertial Navigation System (CAINS II), and the multiservice Standard GPS Receivers, the 3A and the Miniature Advanced GPS Receiver (MAGR). The miniaturization of GPS receivers to the size of one electronic circuit card such as the Rockwell Collins GEM III, has allowed the incorporation of the GPS in the same electronic enclosure or box as the INS allowing easy retrofit of GPS into existing aircraft, designs such as the Embedded GPS Inertial (EGI) and the GPS Integrated Navigation Assembly (GINA). This design is finding more and more advocates as it allows a low cost method of incorporating the upgrade to include the GPS function by reducing the fielded aircraft structural and wiring modifications while also providing a more tightly coupled INS/GPS for better performance (Reference 17). The new GPS receiver shares the power supply and existing wiring of the INS while only needing the addition of an antenna wire to the INS box.

Parameter		Supplier A	Supplier B
Bias	(deg/hr)	0.005	0.003
Scale Factor	(PPM)	5	5
Random Walk	deg/$\sqrt{hr}$	0.003	0.0015
Path Length	(cm)	15.24	18.4
System Wieght	(lb)	16	19
System Power	(w)	45	22
Type		Dithered	Multioscillator
Nav Performance	(nm/hr)	0.8	0.8
Velocity	(ft/secrms	2.5	2.5
Cooling Air		None	None
Size	(in.)	7x7x11	7x7x10
Cost	(1990 Dollars)	$75k	$75k
MTBF	Hours	4000+	4000+
Global Positioning System		Yes	Yes
Air Data Capability		Yes	Yes
Radar Altimeter Capability		Yes	Yes
Common Module Maintainability		Yes	Yes

Figure 1.0-1. Current State-of-the-Art Inertial Navigation Systems
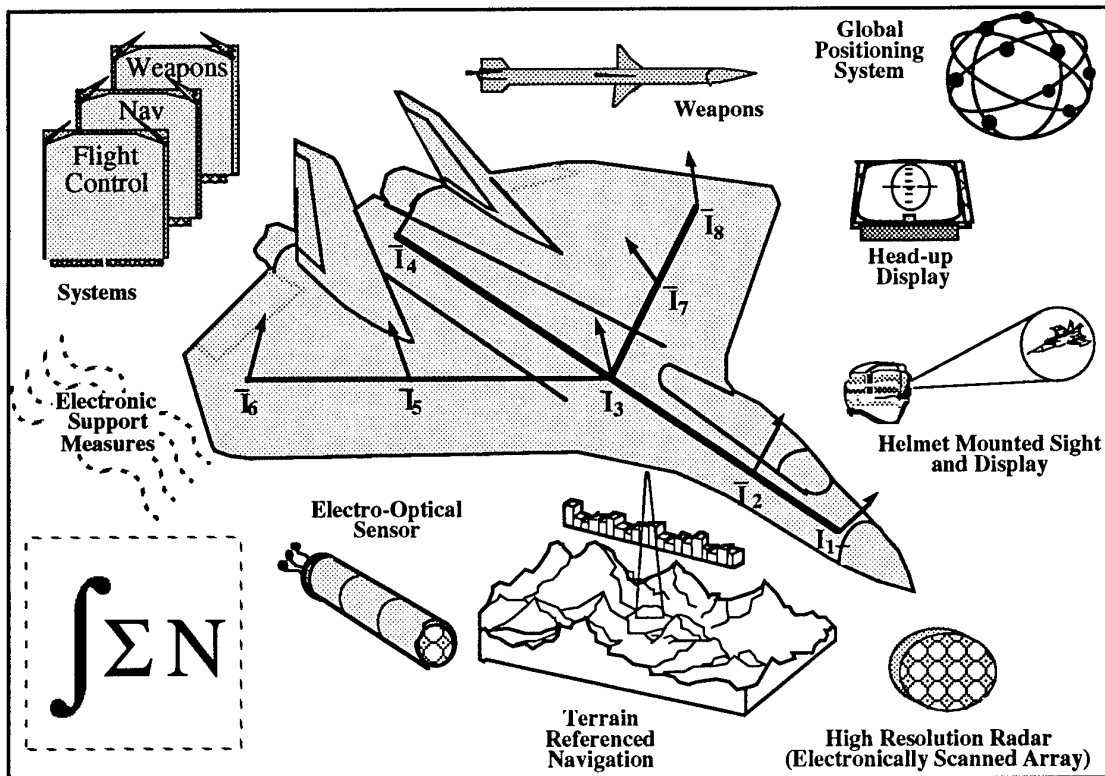
Figure 1.0-2. Integrated Navigation / Inertial Network Concept

CORE PROCESSOR #2

INS #1

	40 Hz	
• Closed Loop KF#1 - GC, CV, IFA, **NAV Mode** **Performance 0.8 nm/hr**	Position, Velocity Attitude, Altitude, ...	

CORE PROCESSOR #1

Update Rates

• Preprocess External Sensors - Measurement Noise and Residuals

INS #2

	40 Hz	
• Closed Loop KF#2 - GC, CV, IFA, **NAV Mode** **Performance 0.8 nm/hr**	Position, Velocity Attitude, Altitude, ...	

Open Loop Kalman Filter #1
Navigation Aiding
Correction Algorithms

Corrected Nav Data

Navigation Data Source
Selection Logic

Maintenance Diagnostics / **Validity Cross Check**

Dual INS Averaging Backup Mode

Open Loop Kalman Filter #2
Navigation Aiding
Correction Algorithms

Corrected Nav Data

Best Available Navigation Outputs Offer Improved Performance

$$\frac{0.8}{1.4} = 0.57 \text{ nm/hr}$$

- **Need separate Open Loop Kalman Filters for each INS to Properly Model Correlations Built up during Independent Alignment Functions**

Figure 1.0-3 Dual INS Performance is Improved Over Single System

The current generation of military aircraft now under development will be using common modules as building blocks of even more integrated and interdependent navigation system architectures as shown in Figure 1.0-2. The design and development of fixed wing aerospace navigation systems is based on operational requirements of the vehicle, mission scenarios, sensor suites, weapons suites, reliability and maintainability requirements. Because of the high degree of integration of current and future systems, it is imperative that this design and development follow a rigorous analytic process in order that the establishment of the required accuracy, signal characteristics, interface requirements, noise, bandwidth, reliability, physical characteristics and cost of the system be well defined.

The system engineering process allows the designer to perform trade studies in order to optimize system design (References 2, 5 and 6). For example, it may be more cost effective to achieve higher system performance by using lower cost inertial navigation systems aided by aircraft sensors performing velocity and position updates. Another example could be a configuration using averaged dual medium accuracy INS's to achieve high accuracy and higher mission reliability instead of single high accuracy fault tolerant INS as illustrated in Figure 1.0-3. A dual INS installation would also allow placement of one system at the Radar antenna for better motion compensation and the other system at a structurally benign position for flight control/autopilot reference or located at another sensor. Section 2.0 will describe a recommended approach for this process.

1.0.1 Definitions

From the previous discussion it can be seen that significant changes have taken place in the field of aerospace navigation systems. This paper uses the definition of navigation as defined in Reference 4, "the process of directing the movements of a craft from one point to another", and will deal primarily with fixed wing military aircraft such as fighters, attack, strategic and transports. It will deal with three main missions, the tactical, the strategic and the transport. At this point it is necessary to define some new concepts and terms used in today's aerospace vehicles. As shown in Figure 1.0.1-1

(Reference 5), aerospace vehicle functions can be divided into three main systems, (1) the Pilot Vehicle Interface (PVI), (2) the Mission Management System (MMS) and (3) the Vehicle Management System (VMS). The first deals with the aircrew/machine interfaces, the second consists of all the mission critical systems, while the third is made up of all flight critical systems. Figure 1.0.1-1 also indicates that there exists some overlapping functions between the mission management and the vehicle management functions such as the trajectory control of the vehicle when performing functions such as terrain following and automatic landings or other integrated control functions.

loss of the capability of the vehicle to successfully complete its designated mission.

Aerospace navigation system functions are critical to both these systems and must address dual requirements in some cases which in turn will affect system reliability, redundancy, survivability and cost. Current navigation system functions in fixed wing vehicles must provide much more than the classical outputs of position, velocity and time. Vehicle attitude, attitude rate, acceleration, acceleration rate in body and geodesic coordinates with time correlation for all outputs are also required. In addition to geographic
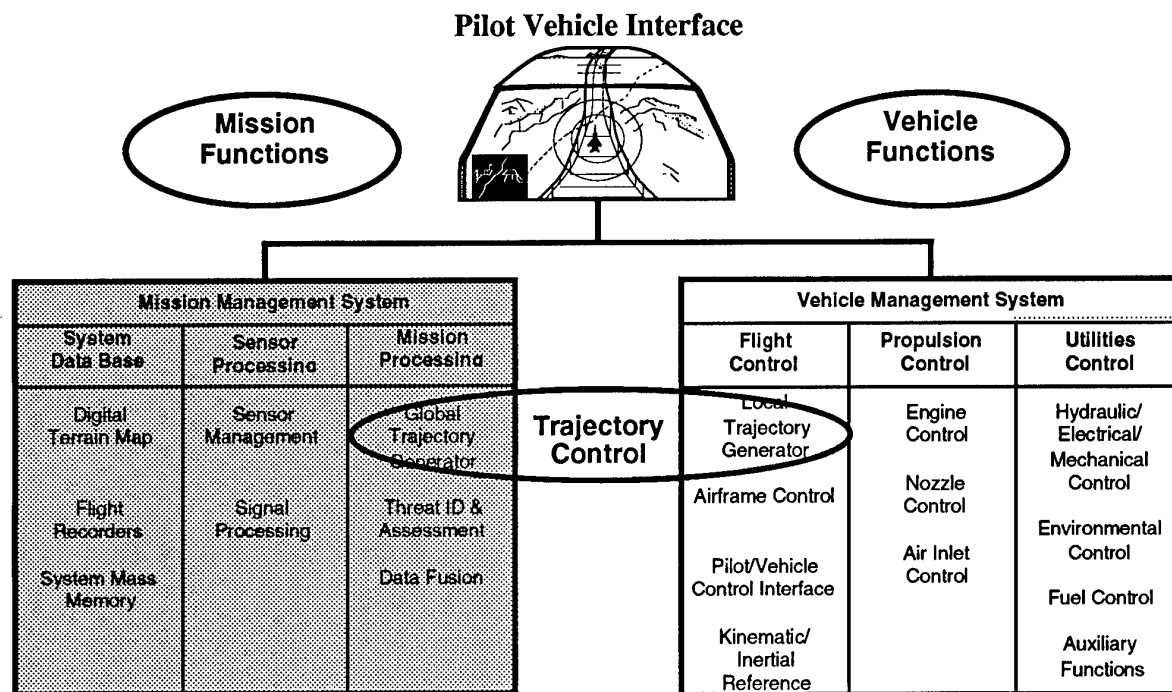
Figure 1.0.1-1. Aerospace Vehicle Functions

In this paper we will define a VMS as an Integrated System that manages flight and vehicle critical systems and the interfaces with those systems that support the flight and vehicle critical systems. Flight critical systems being defined as those elements whose failures cause loss of flight path control. Vehicle critical are those elements whose failure or degraded performance cause loss of the vehicle (Reference 2 and 5). Mission critical systems are defined as those elements whose failures will cause the

navigation, relative navigation to a fixed point or one or more aircraft is also needed. With the above major functional definitions in place, it is now appropriate to examine the purpose of aerospace navigation systems in fixed wing aircraft.

**System
Engineering Process
Air Vehicle Level**

**System
Engineering Process
VMS Level**

Hardware
Design → Fabrication

System
Design

Building
Blocks

System
Integration
and Test

Software
Design → Coding

Hardware/Software Development Level

CC24-0032-11-D/js
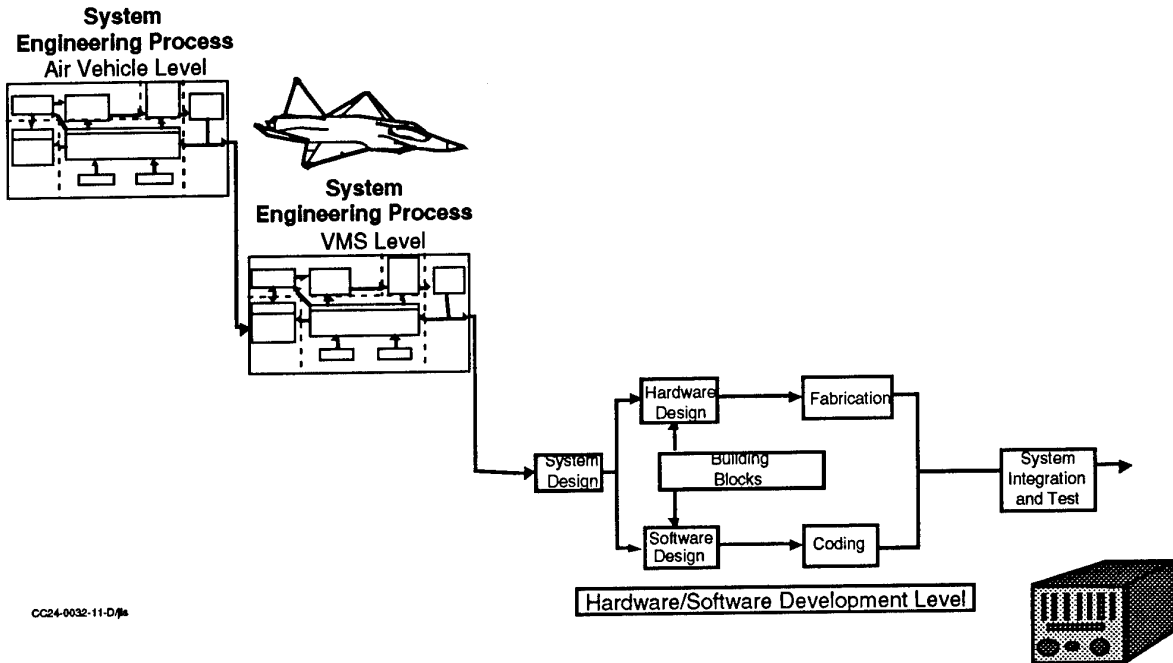
Figure 2.0-1. Navigation System Development Process

2.0 The System Engineering Process

In order to determine the requirements of modern fixed wing aircraft navigation systems we need a disciplined process. An disciplined engineering development process would have the following characteristics:

1. Top down, requirements driven.
2. Disciplined and repeatable.
3. Incorporate concurrent engineering principles.
4. Incorporate effectiveness and balance trade-off analyses.
5. Allow for design iteration and refinement.

Figure 2.0-1 provides an overview of the recommended navigation system development process/methodology which incorporates the needed characteristics (Reference 5). The process is a top down approach wherein air vehicle level requirements are flowed down to the navigation system level. System engineering (Reference 6) is a management process which controls total system development. Its objective is to achieve the optimum balance of all system elements.

MIL-STD-499A defines Systems Engineering as, "the application of scientific and engineering efforts to (a) transform an operational need into a description of system performance parameters and a system configuration, through the use of an iterative process of definition, synthesis, analysis, design, test, and evaluation; (b) integrate related technical parameters and ensure compatibility of all physical, functional, and program interfaces in a manner that optimizes the total system definition and design; and (c) integrate reliability, maintainability, safety, survivability, human engineering, and other such factors into the total engineering effort to meet cost, schedule, supportability, and technical performance objectives".

The requirement to use systems engineering is invoked by DoD DIR 5000.1 Part I, Paragraph C.1.b dated 23 February 1991. It specifies that, "Program plans must provide for a systems engineering approach to the simultaneous design of the product and its manufacturing, test, and support processes.
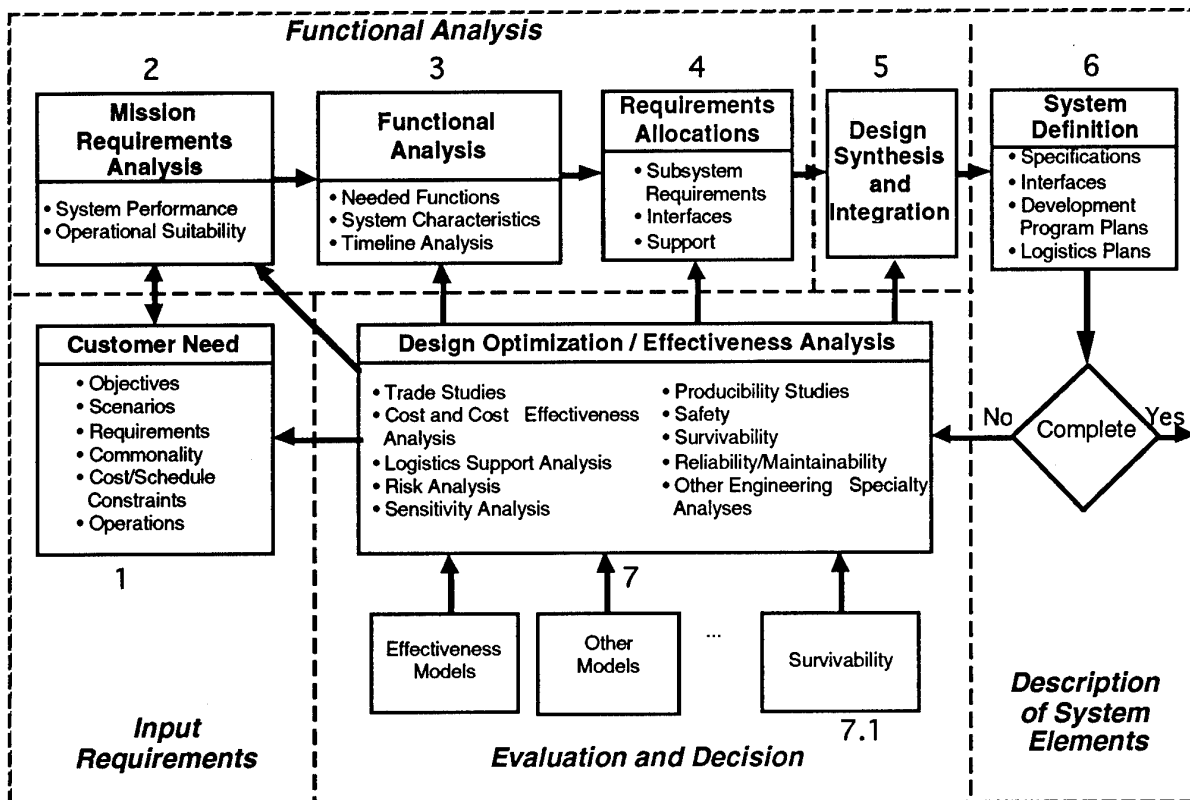
Figure 2.0-2. System Engineering Process

2.0.1 Systems Engineering Process

As shown in Figure 2.0-2, systems engineering comprises five major activities, some of which have been decomposed into smaller tasks. The output of the system engineering process is documentation including specifications, plans, functional flow block diagrams, and trade study reports. The approach is to apply the process of Figure 2.0-2 at each level as outline in Figure 2.0-1.

The first step at each level is to determine the customer needs as shown in step 1, convert those to mission requirements of step 2, perform a functional analysis as shown in step 3 and then allocate the requirements as shown in step 4. Steps 2, 3, and 4 constitute the system functional analysis which is then used in step 5 to perform the design synthesis and system integration. The next step is to define the system for specifications, interfaces, development plans and logistics as shown in step 6 which results in a description of the system elements. If the design is still not complete, step 7 is performed in order to optimize the design by performing trade studies and effectiveness analysis, as well as other related analyses. This final step is the evaluation and decision step. This step may result in further iterations through steps 1 to 5 until the decision path in 6 is that the design is complete. Section 3.0 will perform some examples of Step 1, input requirements, step 2, mission requirements analysis and step 3, functional analysis and Step 4, requirements allocations in order to give the reader a methodology to follow and some results that may be useful in his applications. Section 3.2 will present a table of typical requirements allocations for a modern fixed wing tactical aircraft that can be used as a basline as it would be difficult to define all possible requirements without performing an analysis for all aerospace vehicles. To define all possible requirements could also result in overspecifing the system and increased cost, weight and maintenance.

3.0 Analysis of Fixed Wing Aircraft (steps 1 through 4)

This section will examine a top level functional decomposition of possible areas of coverage for navigation systems in fixed wing aircraft in order to establish major baseline functions that may be required which in turn can be decomposed into sub functions. The purpose of this section will be to define the possible functions of the navigation system which will then be later used to relate them to aircraft requirements, vehicle functions, mission functions, mission tasks, sensor suites, weapons suites, survivability, reliability and maintainability. Figure 3.0-1 shows the first step in the decomposition of aircraft requirements which simply shows how mission and aircraft requirements can be mapped into the three major system requirements.

The navigation system in modern aerospace vehicles serves for more than directing the vehicle from point to point. As described in the previous section, the navigation system has functions dedicated to support traditional navigation position and velocity but is also required to provide aircraft attitude for the flight instruments used by the pilot, aircraft references for flight control and autopilot functions, references for stabilizing and aiding aircraft sensors, parameters for weapon delivery, references for filtering and improving the air data parameters, references for integrated control functions such as trajectory control and fire control.

The increased functional requirements have in great part resulted from increased operational requirements of the vehicle. Improvements in aircraft performance, weight
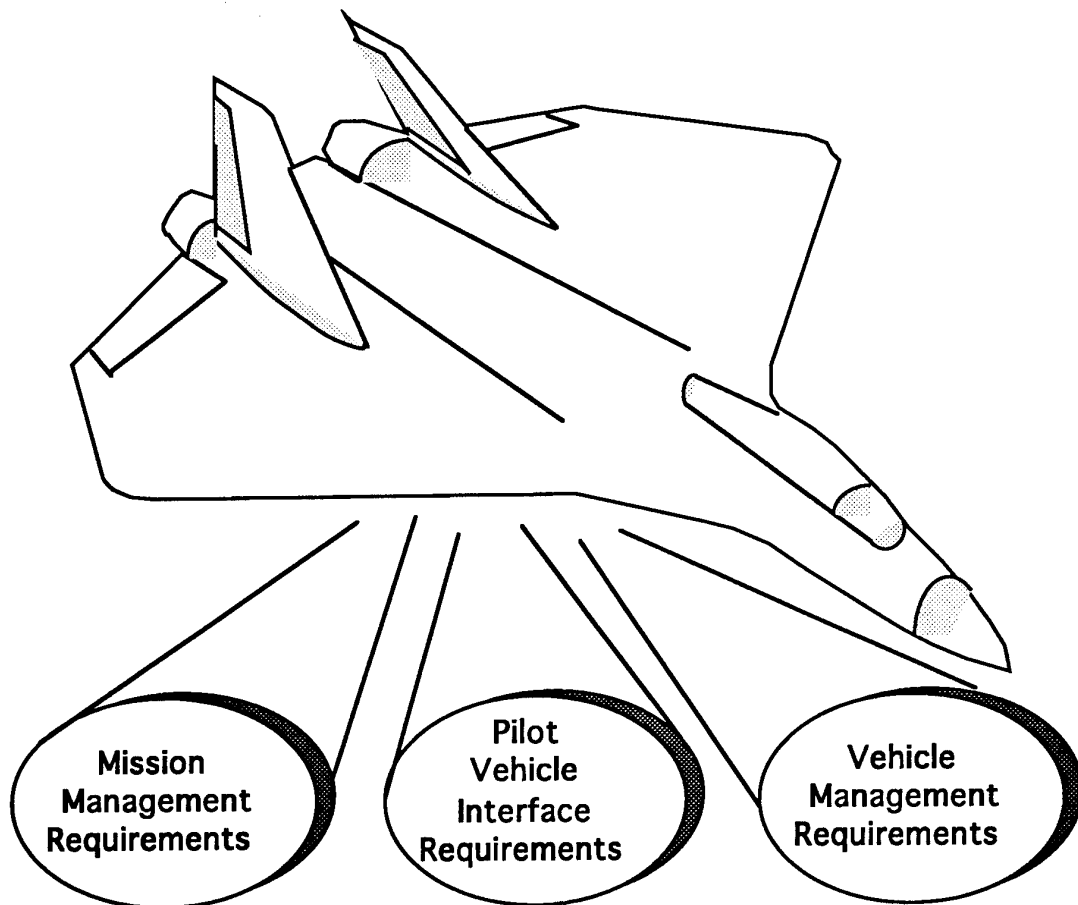
Figure 3.0-1. Aircraft Functional Decomposition

337

and cost have been achieved by reducing the size of aerodynamic surfaces which in turn reduce aircraft stability. Stability is then recovered by the use of more accurate inertial data by the flight control system. Fuel and time savings can be achieved by optimal path control during climb out and cruise mechanized by coupling the flight control system, the propulsion system, navigation and inertial data. Weapon systems have also been improved by the integration of the flight control, the fire control and the navigation system where the fire control system is coupled in order to aid the pilot in air-to-air combat or air-to-ground weapon delivery.

3.0.1 Military Fixed Wing Aircraft Missions

Due to the previously stated operating modes, it has become more complex to define the navigation system that meets its operational requirements. In order to define all functions and ensure that they map into navigation system requirements and then that these requirements in turn are mapped into a correct technical specification, a well defined process must be used. For military aircraft this process is very complex as these aircraft have wide ranging aircraft and

divided into multiple aircraft. These multirole aircraft must be able to perform their mission while based at full service facilities, austere bases and aircraft carriers.

Military fixed wing aircraft can be divided into their tactical, strategic and transport missions. As seen if Figure 3.0.1-1 the tactical mission can then be subdivided into air-to-ground and air-to-air, these in turn, for example air-to-ground can be further subdivided into (1) close air support, (2) battle field air interdiction, (3) deep interdiction, (4) reconnaissance. The air-to-air mission can be subdivided into (1) offensive counter air and (2) defensive counter air. The strategic mission can be subdivided into (1) hi / lo extremely long range interdiction and (2) hi / hi extremely long range interdiction. The transport aircraft mission can be divided into two main divisions (1) rapid deployment of air and ground forces and (2) long term supply of air and ground forces.

3.0.2 Typical Tactical Aircraft Customer Requirements

A tactical aircraft must be capable of rapid deployment on both air-to-air and air-to-ground missions and offer affordability, availability, survivability, and lethality.. Typical aircraft performance requirements

Strategic		Tactical						Transport	
		Air-to-Ground				Air-to-Air			
Hi / Lo Long Range Interdiction	Hi / Hi Long Range Interdiction	Close Air Support	Battlefiel Air Interdictio n	Deep Interdiction	Reconnaissance	Offensive Counter Air	Defensive Counter Air	Rapid Deployment	Long Term Supply

Figure 3.0.1-1 Military Aircraft Missions

mission requirements. Military aircraft typically require that the aircraft flight envelope range from low subsonic to high supersonic (mach 2 or 3), from low altitude (50 to 100 feet) to extremely high altitudes (70,000 feet). Environmentally they must operate from temperatures of -40 degrees to 70+ degrees C with reaction times of 30 seconds to five minutes. In addition, the trend is to multimission / multirole aircraft that must meet the requirements traditionally

are summarized in the paragraphs below:

Altitude and speed. The equipment shall operate with the accuracy state herein at altitudes form -1000 feet to 75,000 feet and over a speed range of 0 to 300 feet/second horizontal and minus to plus 1500 feet/second vertical.

Aircraft capability. The equipment shall be capable of operation in all aircraft attitudes.

Air Interdiction Mission

(4) MK-84 IAMs
(2) AIM-9M
20 mm (1,000 Rounds)

Combat
0.8 - 0.9 Mach
500 ft

Loiter
20 min
@ S.L.

Optimum Cruise

550 KCAS Dash

200 ft Altitude

150 NM

275 NM

Defensive Counter Air Mission

Combat
0.9 Mach
25,000 ft

(4) AIM-120 (AMRAAM)
(2) AIM-9M
20 mm (1,000 Rounds)
(1) DEW

1 hr Loiter
@ 0.7M/25k

Loiter
20 min
@ S.L.

Optimum Cruise

1.4 Mach Dash

25 NM

100 NM

Close Air Support

(4) MK-82 GPBs
(2) AIM-9M
20 mm (1,000 Rounds)

Combat
0.8 - 0.9 Mach
500 ft

Loiter
20 min
@ S.L.

Optimum Cruise

550 KCAS Dash

200 ft Altitude

15 NM

100 NM

Offensive Counter Air Mission

Combat
0.9 Mach
25,000 ft

(4) AIM-120 (AMRAAM)
(2) AIM-9M
20 mm (1,000 Rounds)
(1) DEW

Loiter
20 min
@ S.L.

Optimum Cruise

1.4 Mach Dash

50 NM

100 NM

CC14-0032-292-D/vc

Figure 3.0.2-1 Tactical Aircraft Missions

Battlefield Air Interdiction	1. Pre-mission Planning 2. Pre-flight and Pre-takeoff 3. Takeoff 4. Departure 5. Cruise (haven)	Takeoff / Ingress
	6. Let Down toTerrain Following 7. Ground Target Acquisition and Attack 8. Post Attack	Air-to-Ground Attack
	9. Return to Base 10. Penetration / Approach 11. Land	Egress / Landing
	12. Turnaround	Service and Maintenance
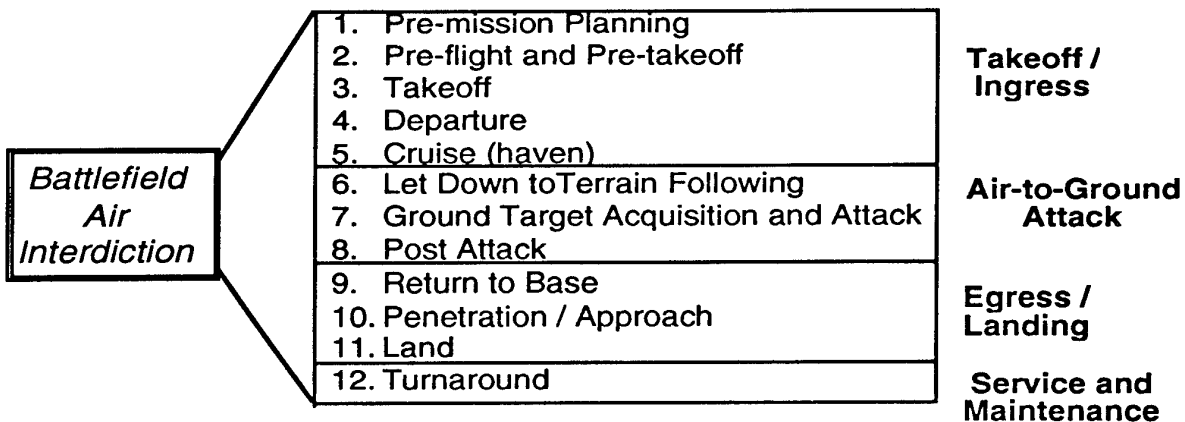
Figure 3.0.2-2 Battlefield Air Interdiction Mission Phases and Segments

Aircraft dynamic response. The equipment shall be capable of functioning and maintaining alignment such that overall system performance will not be degraded by aircraft maneuvers up to the following limits:

(a)   Aircraft maximum angular rate

Pitch Rate = 2.0 radians/second
Roll Rate = 5 radians/second
Yaw Rate = 3.5 radians/second

(b)   Aircraft maximum accelerations
Pitch Acceleration = 3.5 radians/second^2
Roll Acceleration = 18 radians/second^2
Yaw Acceleration = 3.5 radians/second^2

Tactical Aircraft Mission Requirements
The tactical aircraft will be required to perform multiple missions. The primary missions are air-to-ground missions and air-to-air. A pictorial representation of some tactical mission profiles is shown in Figure 3.0..2-1.

To further define a typical mission, it is necessary to decompose the mission into detailed tasks. First, the mission is divided into mission phases. Within each mission phase there are mission segments, and within each mission segment there are the detailed tasks. The Air Interdiction mission phases and their segments are shown in Figure 3.0.2-2 as an example of decomposition.

porating a low-observable phased-array antenna, an extremely accurate inertial navigation system, a strategic Doppler radar altimeter, and a star tracker. Its defensive avionics are built around the AN/ALQ-161 electronic counter measures system, with extended frequency coverage, and includes a tail warning radar and expendable decoys such as chaff and flares.

It has three internal weapons bays with the capacity to carry up to eight AGM-86B air-launched cruise missiles (ALCMs) and twenty-four short-range attack missiles (SRAMs). In a conventional, non-nuclear role, it can carry up to eighty-four 500
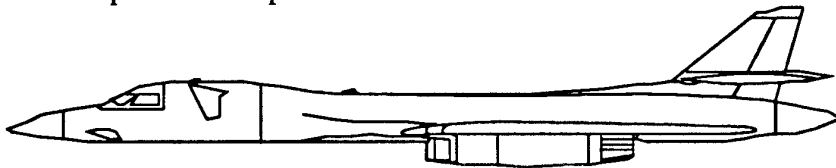
Figure 3.0.3-1 Representative Strategic Aircraft

3.0.3 Strategic Aircraft Customer Requirements

The strategic aircraft selected as an example is representative of the B-1B and B-2 bombers. The strategic aircraft is a long range, multi-role strategic bomber and the mission/sensor/weapons are considered to be representative of advanced bombers. It is capable of carrying, in three weapons bays, varying combinations of nuclear air-to-ground missiles, conventional or nuclear free-fall bombs, and auxiliary fuel. In addition it has electronic jamming equipment, infrared countermeasures, radar location and warning systems, other advanced avionics and low observable technology.

The strategic aircraft, shown in Figure 3.0.3-1, has a maximum gross weight of 477,000 pounds and is capable of supersonic flight at speeds approaching Mach 2.0. Its primary role will be high subsonic, low-altitude penetration to the target area. Its variable-geometry wing permits operation from shorter runways in the upswept position and supersonic flight in the fully swept position. The avionics systems include advanced forward-looking and terrain following radars incor-

pound MK-82 or twenty-four 2,000 pound MK-84 bombs. It also has the capability to carry, on external stations, an additional forty-four MK-82s or fourteen MK-84s.

Strategic Aircraft Mission Requirements
The typical strategic mission is normally an Air Interdiction type of mission entailing much longer ranges and greater weapon capacity than the tactical aircraft missions. In a wartime scenario, the tactical aircraft would operate out of Forward Operating Bases where the strategic aircraft would be operating from Main Operating Bases much farther removed from the area of conflict. Strategic missions are longer duration, more autonomous, and have different target assignments and associated weapon loads than a tactical mission. Reference 3 provides additional information on strategic missions.

The nuclear mission profile includes two different profiles. One mission, carrying ALCMs, involves a high altitude launch of the ALCMs outside of enemy airspace and a return to base. The other mission profile, involving delivery of nuclear bombs, is executed at a low altitude and high subsonic airspeed during penetration into enemy airspace. This mission profile would re-

semble the tactical air interdiction profile in Figure 2.2.2-1 except with longer cruise-to-letdown distances. The weapons would be delivered in a level delivery (one g-loading) so the g-loading on the aircraft during the entire mission would likely range between one and three g's. Delivery of its nuclear cruise missiles (one g-loading) would be preceded with a transfer alignment of the cruise missile inertial reference system. The terrain correlation matching system in the missile is used for enroute and terminal position updates. Terrain reference navigation (TRN) compares real-time sensed terrain variations with onboard reference maps to determine the position of the missile.

cruise missile is aligned to the reference INS prior to launch. Alignment of the cruise missile is allocated sufficient time for convergence because of the longer mission duration, as compared to tactical missions. After launch, the cruise missile descends to 200 feet, flies over water for 500-1000 NM, and ingress's to the target at 0.7 Mach which is 2350 NM from launch point (a total of about 5 hours flying time from launch). Approaching the target, a single three-fixed update is made to improve probability of kill.

The non-nuclear mission profile is a high altitude profile for the entire mission thus sur-

Figure 3.0.3-2 Strategic Aircraft/Cruise Missile Mission Profile
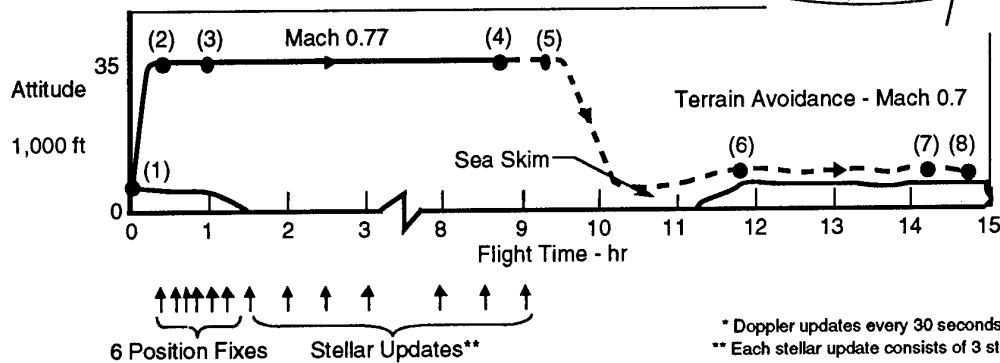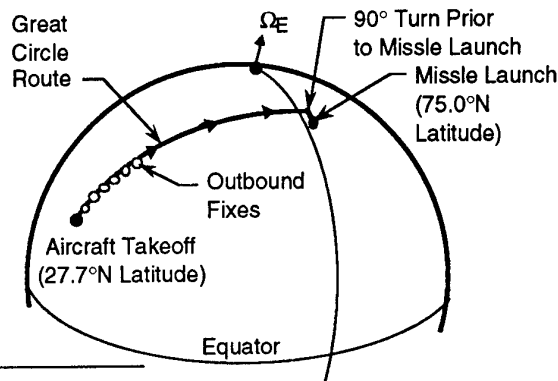
The profile for the cruise missile mission is depicted in Figure 3.0.3-2. The events table shows tasks of both the strategic aircraft and cruise missile. The reference INS is started in flight using Doppler updates every 30 seconds and radar position fixes to align the INS. During flight over water (up to eight hours), stellar updates (three star fixes) are performed every 30 minutes to better calibrate the INS. During captive flight, the

face-to-air and enemy air threats are not significant factors. Carpet bombing is accomplished by taking advantage of the large numbers of conventional weapons, (MK-82 or MK-84 low drag general purpose bombs), that can be carried on the strategic aircraft.

3.0.4 Transport Aircraft Customer Requirements

Examples of this type of aircraft are: C-17 and C-130 Advanced Theater Transport, and Special Operations Aircraft. The mission requirements for a transport aircraft are outlined as follows:

```
Covert operation
Combat zone
Austere basing
Low level (below 200 ft.) flight
All weather / night operation
Long / extended range
Cooperative tactics
Aerial delivery
Medical evacuation
```

Transport Aircraft Mission Requirements

Typical mission segments for a transport aircraft are shown in Figure 3.0.4-1. The most critical segment for a VMS from a performance point of view is the low altitude parachute extraction segment. This requires holding altitude at 4 to 6 feet above the ground while off-loading cargo of up to 60,000 pounds. Although this altitude-hold requirement is more stringent than typical, it does not impact the VMS architecture. The most significant architectural impact is due to the length of the mission. The long duration, medical evacuation long range requirement drives the safety and redundancy requirements for its' VMS architecture.

```
1. Pre-mission Planning
2. Preflight and Pre-takeoff
3. Takeoff
4. Departure/Cruise
5. Medium Altitude Airdrop
6. Outbound Flight
7. Rendezvous
8. Returning Flight
9. Low Altitude Parachute Extraction
10. Return To Base
11. Descent And Land
12. Turnaround
```

Figure 3.0.4-1 Mission Segments For Transport Aircraft

3.1 Functional Analysis

A functional analysis must be performed in order to examine the navigation system performance requirements of the aircraft. Requirements addressed should include those for initial alignment, basic navigation, sensor cueing and designation, Inertial Navigation System (INS) drift requirements, and Vehicle Management System (VMS) flight critical functions. Weapon delivery systems impose additional requirement on the navigation system to provide updating to improve weapon delivery functions. "Aided" inertial sensing performance requirements should also be addressed. VMS functions to be examined are inner-loop flight control (basic flight) and integrated flight modes.

Initial Alignment Analysis Inertial navigation performance during flight is greatly dependent on the quality of alignment received before takeoff or during an in-air alignment. Performance parameters are usually based on a full-up, ground, gyrocompass alignment. Shortened reaction time alignments can be mechanized resulting in reduced accuracy or in special parking and handling of the aircraft prior to starting the alignment. Several alignment modes are required in order to provide flexibility in reaction time and resulting accuracy.

Other alignment requirements are (1) automatic interrupted ground alignment detection using measurements of vehicle velocities, and (2) alignment quality indication (approximate CEP drift rate in nm/hr).

Normal Ground Alignment Analysis During normal alignment the INS will autonomously determine the correct reference frame. The vertical reference is established by sensing the earth's gravitational acceleration. The true north heading reference is determined by sensing earth rates and equating those to the known earth rates for the latitude at the aircraft position. The alignment time required for full performance is typically four minutes.

Alignment requirements usually specified for a navigation system are alignment time and accuracy. These two requirements primarily affect the quality of gyros required. Although detailed gyro specifications are not usually included in a procurement specification, knowledge of operation is essential to evaluate the feasibility of candidate systems and alignment requirements.

• Typical Allocation of Error Sources During Alignment

Error Source	Allocated Error	spec.
1. Gyro WRN	0.36	0.001°/√hr
2. Gyro Bias	0.13	0.0015°/hr
3. Vel Disturbance	0.13	0.02 ft/sec
4. All Other Errors	0.03	
	0.41 mrads (about 0.4 NMI/hr)	

Figure 3.1-1 Typical Allocation of Error Sources During Alignment

Figure 3.1-1 shows a typical error budget for a Strapdown INS alignment (Reference 7). The largest single contributor to the residual alignment error is the gyro wide band random noise (WRN) or random walk which typically dominates for alignment times greater than 2-3 minutes. The relationship for wide band random noise and residual error is shown in Figure 3.1-2 (Reference 7).

$$\varepsilon = \frac{\eta}{\Omega \times Sqrt(T) \times Cos(\lambda)}$$

where,

ε = heading error allocated to wide band random noise

η = gyro wide band random noise specification

Ω = earth rate (15.04°/hr) (sidereal time)

λ = latitude

T = alignment time

Figure 3.1-2 Random Wide Band Noise and Alignment Error

Stored Heading Alignment Analysis A stored heading ground alignment is provided to minimize INS alignment time. During this mode the INS heading is set to the value that was stored during the previous ground alignment or at the end of the last flight. This mode is usable only if the aircraft has not been moved since the last alignment or flight. Reaction time is typically 30 seconds.

Attitude And Heading Reference (AHRS) Ground Alignment Analysis The AHRS ground alignment mode provides a 30 second alignment without stored heading available. In the AHRS mode the INS provides heading, attitude, angular rates, and acceleration with a lower accuracy. Subsequent in-flight aiding or in-flight alignment results in full navigation performance.

Inflight Alignment Analysis The inflight alignment mode enables the INS to perform an inflight start up or improve system performance if less than a normal full alignment was performed. The INS will use the best present position, velocity, and heading data available from the non-inertial navigation aids.

Basic Navigation Analysis The basic navigation functions for all of the tactical and the strategic mission are similar. The strategic mission places a different requirement on the basic navigation function due to extended mission duration. Inertial navigation is the process of calculating position and velocity based solely on inputs from self-contained acceleration sensing instruments. Accelerometers provide the acceleration magnitude sensing function. Gyros provide the acceleration direction sensing function (i.e. define the direction of the accelerometer sensing axes). The basic navigation concept is to integrate the sensed acceleration vector data (corrected for gravity) to determine vehicle velocity, and then integrate the velocity data to determine vehicle position.

Basic navigation parameter outputs required are latitude, longitude, altitude, vertical velocity, horizontal velocities, true heading, roll angle, pitch angle, and load factors (linear acceleration). Accuracy's needed for basic navigation are indicated in the following paragraphs.

Latitude And Longitude During normal navigation, aircraft latitude and longitude are required for on-course flight to a way point or rendezvous point. Autonomous position accuracy available with current, standard inertial navigation technology (0.6 to 1 nm/hr CEP drift rate) have been and are still adequate for basic navigation present position. Worldwide operation ($\pm$ 90∞) latitude, $\pm$ 180∞ longitude) of the inertial present position function is required.

Present position requires fault tolerance and survivability for safety at night, in weather, and for return to friendly territory/base. To increase system availability, a fail-op continuous inertial navigation capability is provided. Survivability is provided by use of radio navigation aids in case battle damage completely disables the fault tolerant inertial navigation function.

Heading And Attitude True heading, roll angle, and pitch angle are required for attitude control, coordinated turns, landing and takeoff, autopilot and proper display orientation. Autonomous true heading and attitude accuracy available with current inertial navigation sets (better than 0.1 deg, 1 s) have been and are still adequate for safe navigation. Range required is all possible heading and attitudes, usually expressed as 0 to 360 degrees heading, $\pm$ 90 degrees in pitch and $\pm$ 180 degrees in roll.

In previous aircraft, the INS, Attitude Heading Reference System (AHRS), Magnetic Azimuth Detector (MAD), and backup instruments provided the aircraft true heading, magnetic heading, pitch-roll attitudes, and rate of turn for navigation and safe flight at night and in weather. Consolidation of these equipment sets requires at least a fault tolerant INS or two sources of attitude data (2 INSs or INS plus AHRS). In addition, a survivable source is required for attitude when coupled directly into the flight control system for low altitude operations such as short take off and landing (STOL) and Terrain Following/Terrain Avoidance (TF/TA). Survivable heading and attitude requires at

least two (2) physically separated sources (such as IIRA, 2 INSs, INS plus AHRS, or INS plus backup in flight control system). The configuration which meets these requirements in a most cost effective manner should be part of a trade study in the design process.

Horizontal Velocities Analysis Navigation frame (North/South and East/West) velocities are required for on-course/on-time flight to a way point or rendezvous. Body frame (nose, right wing, belly) are required for the display of the aircraft velocity vector during low level maneuvers, landing, and weapon delivery. Autonomous velocity accuracy available with current inertial navigation sets (2.0 to 2.5 ft/sec (1 sigma) navigation or body frame) have been and are still adequate for navigation. Range required is -3000 ft/sec to +3000 ft/sec in either frame.

Velocity fault tolerance (fail-op), and survivability requirements are similar to those for present position. Velocity survivability will be provided by air data computations and radio navigation aids in case battle damage completely disables the inertial navigation function.

Altitude And Vertical Velocity Analysis Inertial altitude and vertical velocity are required to supplement the altimeter during large variations in altitude. Pressure-altitude can be invalid during and after steep dives or climbs. Short-term inertial vertical velocity is required to be accurate to 2 ft/sec (1 sigma) during this period primarily to provide accurate velocity vector calculations. The long-term inertial altitude is required to track the pressure-altitude within 2 feet (1 sigma) during steady state conditions and to track true altitude within 150 feet during maneuvers. Range required is -1000 feet to +75,0000 feet for altitude and + 1500 ft/sec for vertical velocity.

Altitude and vertical velocity fault tolerance and survivability are required due to the requirement for valid velocity vector display during landing. Fail-op inertial outputs will be provided to increase availability and safety during large altitude variations.

Linear Acceleration Analysis Longitudinal, lateral, and normal linear accelerations (expressed as load factors) are required for aircraft load warning, recording, and limiting. Normal accuracy and ranges for a tactical fighter are 0.05 g's (1 sigma) accuracy, + 3g's range for longitudinal and lateral accelerations, and -4g's to +12g's range for normal accelerations.

Cooperative Functions Reference Analysis Relative position between cooperating aircraft, and each aircraft's heading are required for triangulation during some cooperative operations. INS heading accuracy of 0.05 deg (1 sigma) is required to achieve a ranging accuracy of 3% under nominal conditions. Relative position can be derived using INS data or by several radio navigation aids. The required accuracy of a few hundred feet is not obtainable with an INS in autonomous operation, but requires frequent position updates or navigation fusion.

During lengthy missions, the aircraft will require inflight refueling. Rendezvous with the tanker aircraft is required prior to refueling. Based on F-15 and F-4 experience a position accuracy of 1.5 nm has been demonstrated to be adequate in the existing cooperative refueling environment.

Sensor Management Analysis The navigation system must provide present position and attitude on data with sufficient accuracy to ensure that the way point or target is within the sensor field of view (FOV) or mapped area when expected.

The INS position and position drift requirements to support sensor cueing in this example will be based on the air interdiction mission discussed in section 3.0. The position error allowed for successful cueing is based on a cueing probability of 0.99. The cueing probability is a function of the way point or target location within the sensor FOV or mapped area is calculated as shown in Figure 3.1-3. The major contributors to way point or target position error within the sensor FOV or mapped area are the target location uncertainty and the aircraft position error.

(1) Position Uncertainties

Half Field Along LOS (Fa)

Line of Sight (LOS)

Half Field Cross LOS (Fc)

Target Location With Position Error

Ideal Patch Location

Patch Location With Position Error

(2) P (Target in Patch) = P (Target in Along Track) * P (Target in Cross Track)

= P (Fa/Along Track Error) * P (Fc/Cross Track Error)

f (x)

-3 -2 -1 0 1 2 3
X

f (x)

-3 -2 -1 0 1 2 3
X
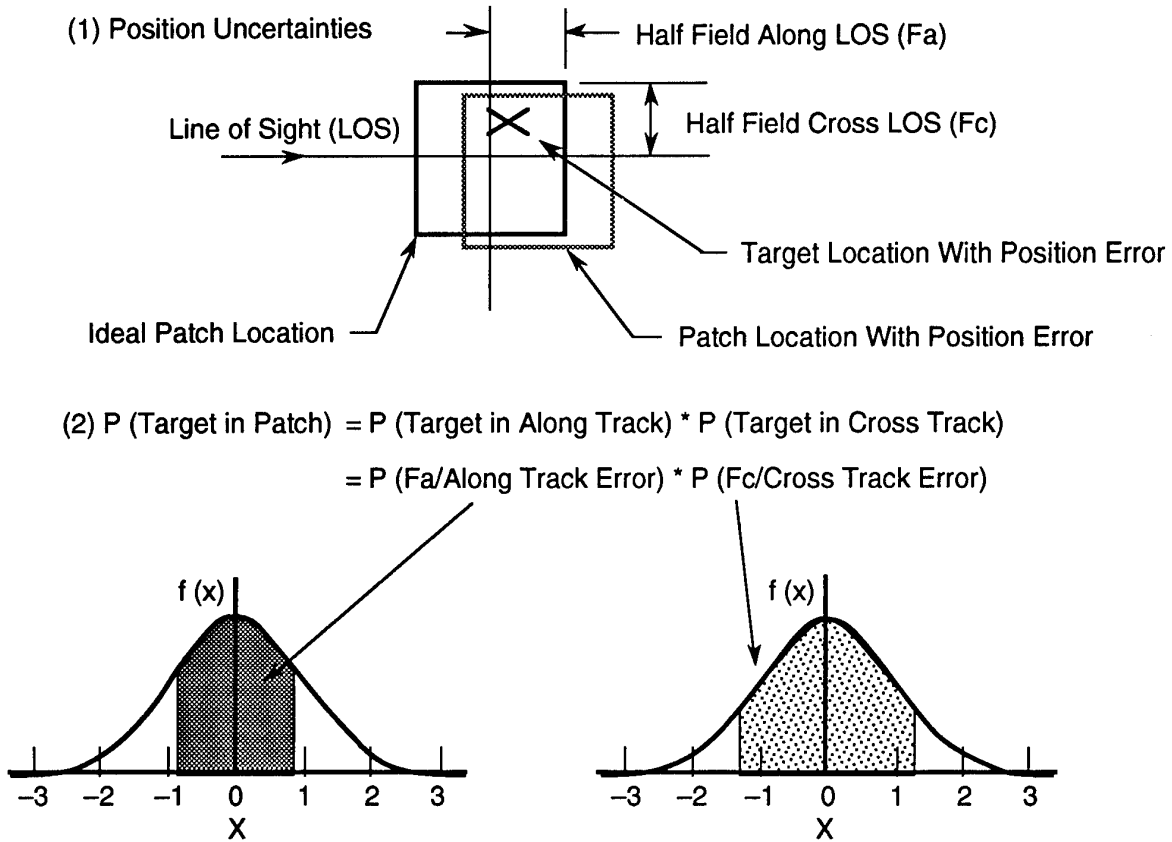
CC24-0157-87-D

Figure 3.1-3 Calculating the Cueing Probability

LOS

1. Perform
Velocity
Update

2. Create
SAR
Map

3. Designate
Waypoint
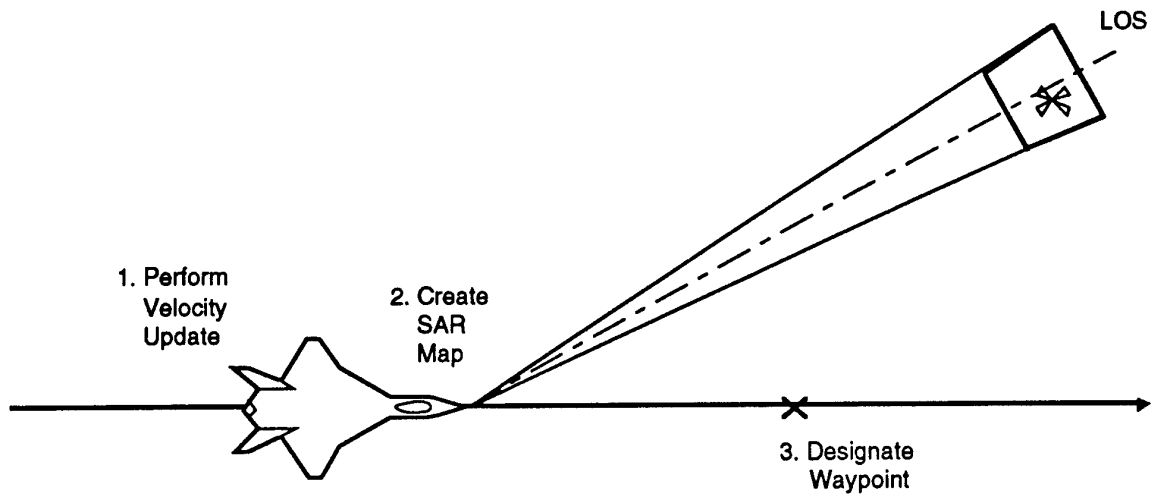
CC24-0157-63-D

Figure 3.1-4   Cueing and Way point Detection with Radar

Patch Size

3.34 x 3.34 NM

1.34 x 1.34 NM

0.67 x 0.67 NM

Probability
Waypoint in
SAR Patch

Patch Error - ft, CEP
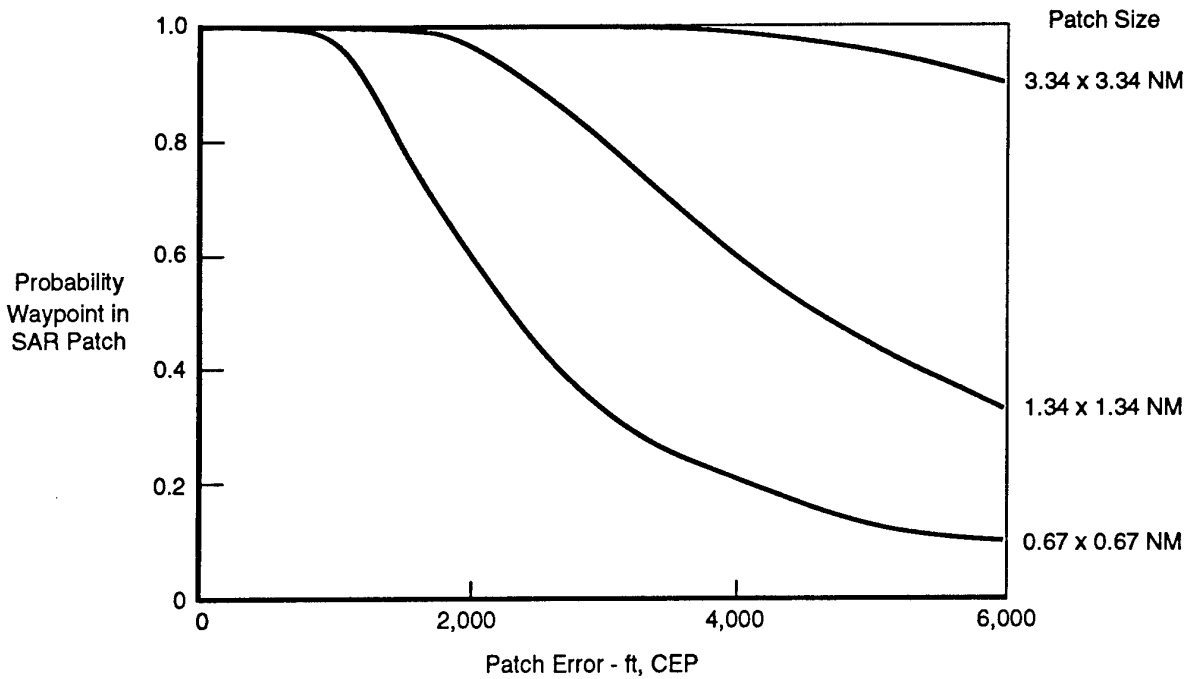
CC24-0157-64-D

Figure 3.1-5   SAR Map Cueing Probability

347

Map Size (NM x NM)	Total Allowed (ft, CEP)	Target Location Uncertainty (Note) (ft, CEP)	Aircraft Present Position Uncertainty (ft, CEP)
0.67 x 0.67	850	295	800
1.34 x 1.34	1705	295	1680
3.34 x 3.34	4250	295	4240
4.67 x 4.67	5945	295	5935
10 x 10	12.7k	295	12.7k
20 x 20	25.5k	295	25.5k
40 x 40	50.9k	295	50.9k
80 x 80	102k	295	102k

CC24-0157-65-D

Figure 3.1-6. Position Error Allowed for Cueing Radar with P=0.99

Wide FOV 5° x 5° or
Narrow FOV 1.25° x 1.25°

Target Location Uncertainty = 295 ft CEP

Pop-Up to 5,000 ft Altitude

Cross Track

Along Track

Detect Range
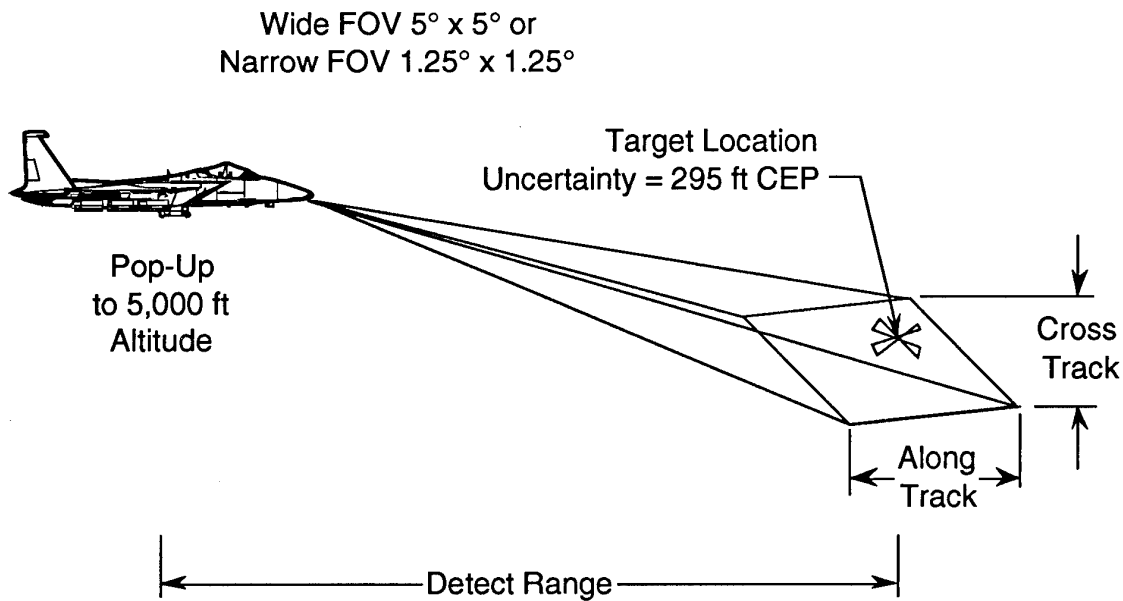
CC24-0157-66-D

Figure 3.1-7 Cueing and Way point Detection with FLIR

348
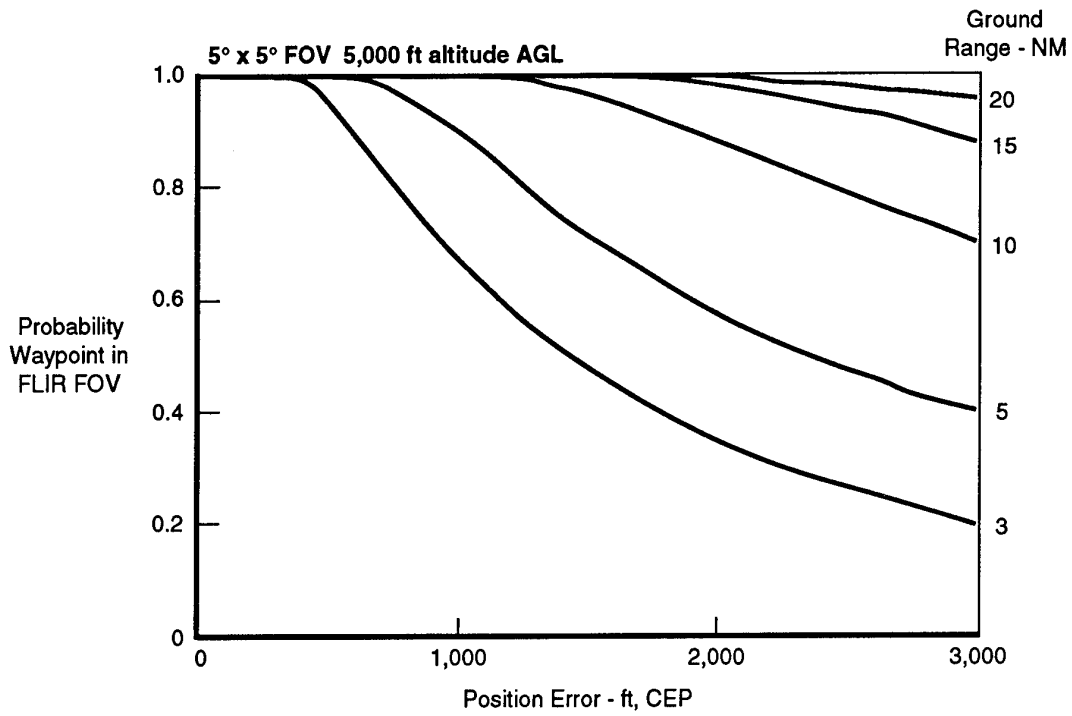
Figure 3.1-8. Position Accuracy for Cueing Wide FOV FLIR
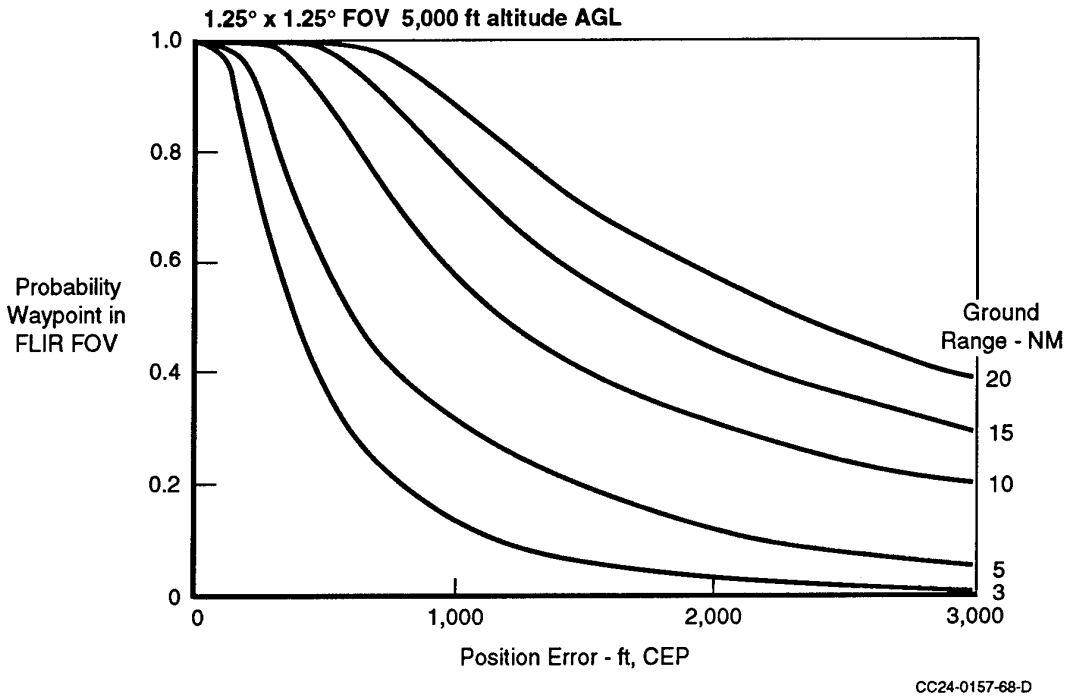
Figure 3.1-9. Position Accuracy for Cueing Narrow FOV FLIR

	Field-of-View (degs x degs)	Maximum Allowed (ft, CEP)	Target Location Uncertainty (ft, CEP)	Aircraft Present Position Uncertainty (ft, CEP)
20	Wide	2427	295	2409
15	5 x 5	1822	295	1797
10		1217	295	1180
5		614	295	539
3		377	295	235
20	Narrow	606	295	530
15	1.25 x 1.25	455	295	347
10		304	295	73
5		154	295	——
3		94	295	——

Conditions: 1. PFOV = 0.99
2. 5,000 ft altitude
3. INS to sensor Boresight better than 4 mrads

CC24-0157-69-D

Figure 3.1-10 Position Accuracy for Cueing FLIR with P=0.99
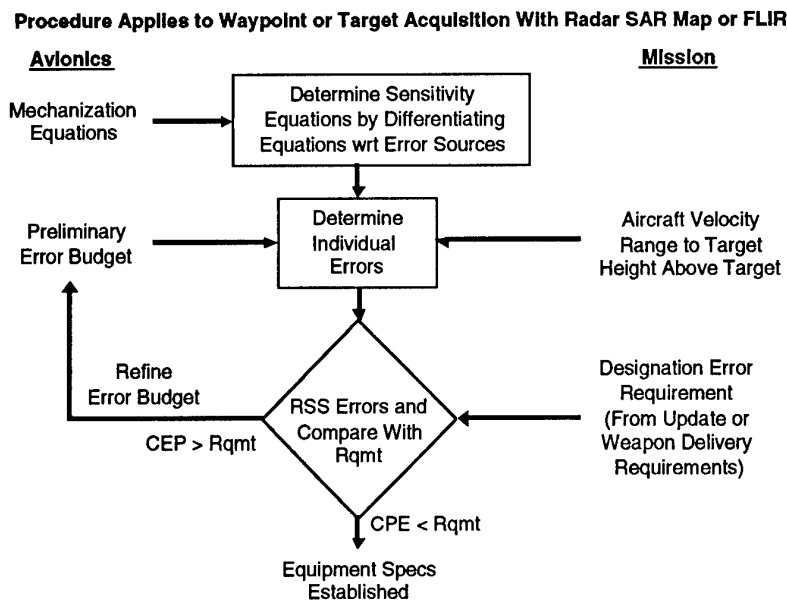
Figure 3.1-11 Equipment Specifications From Designation Error Analysis

Radar Sensor Cueing Figure 3.1-4 shows a typical Synthetic Aperture Radar (SAR) mapping procedure. First a precision velocity update is performed to limit designation errors. Next the map is created for the land area assumed to contain the way point or target. The probability that the way point is in the map is based on map size and aircraft position error as shown in Figure 3.1-5 for several example map sizes. The maximum position accuracy allowed for a given probability are shown in Figure 3.1-6 for a probability of 0.99. The maximum position accuracy allowed consists of aircraft position error and target position uncertainty. The allowable aircraft position error is based on a 295 ft, CEP target (map) uncertainty which has been subtracted in a Root-Sum-Square (rss) fashion.

FLIR Sensor Cueing Analysis The methodology for determining the position error allowed for FLIR cueing is the same as that used for radar cueing once the FLIR along and cross LOS fields are found as illustrated in Figure 3.1-7. This is done by projecting the FLIR FOV onto the earth surface. The cross line of sight (LOS) half field is simply equal to:

$$Rc = Rs \times Tan(q/2)$$

where:

Rc = Cross LOS Field

Rs = Slant Range

q = Sensor Field of View

The probability that the way point is in the along LOS FOV is assumed to be unity due to the geometric enlargement along the ground track as shown.

The probability that the way point (or target) is in the projected field is shown in Figures 3.1-8 and 3.1-9 for wide and narrow FOV's respectively. As the FOV gets smaller and the distance to way point gets smaller, the projected area reduces thus reducing the probability as shown in the figures.

The computed allowable position errors are shown in Figure 3.1-10 for a 0.99 probability of cueing. Thus, from the figure the aircraft position error allowed for a 5 degree by 5 degree FOV FLIR at 20 nm is 2409 ft, CEP (P=O.99). The requirement for the narrow FOV FLIR is only 530 ft, CEP (P=0.99). Some conditions shown in the figures are not realizable even with perfect aircraft position accuracy due to the way point position uncertainty. In these cases the target must be found first using the wide FOV. Once the target is detected and stabilized in the wide FOV, the FLIR can be switched to narrow FOV for tracking and subsequent target state estimation.

The INS position accuracy requirement for sensor cueing is defined by the most stringent position accuracy for the Radar or FLIR. This is found to be 235 ft in the case of FLIR cueing at 3 nautical miles in the wide FOV. Attitude and heading accuracy required for sensor cueing were defined in Reference 10 and 15 and are specified as 0.1 degrees (1 sigma).

Designation Management Analysis The inertial navigation system must provide the necessary platform references to the appropriate sensor to establish relative location of the platform with respect to target or way point. The relative data is necessary to compute ballistic solutions for bombing or to provide an accurate platform position fix relative to a known map feature. The methodology used to determine the contribution of inertial reference parameters to designation errors is outlined in Figure 3.1-11. The procedure applies to way point or target designation with radar SAR map or FLIR.

Sensitivity equations are determined by differentiating the mechanization equations with respect to the error sources of interest. Error values and aircraft parameters are then entered into the sensitivity equations. The resulting error contributions from each sensor source are RSS'ed together and compared with the required designation error. The error budgets are iterated until the requirement is met and a suitable compromise is made between equipment specifications and equipment producibility.
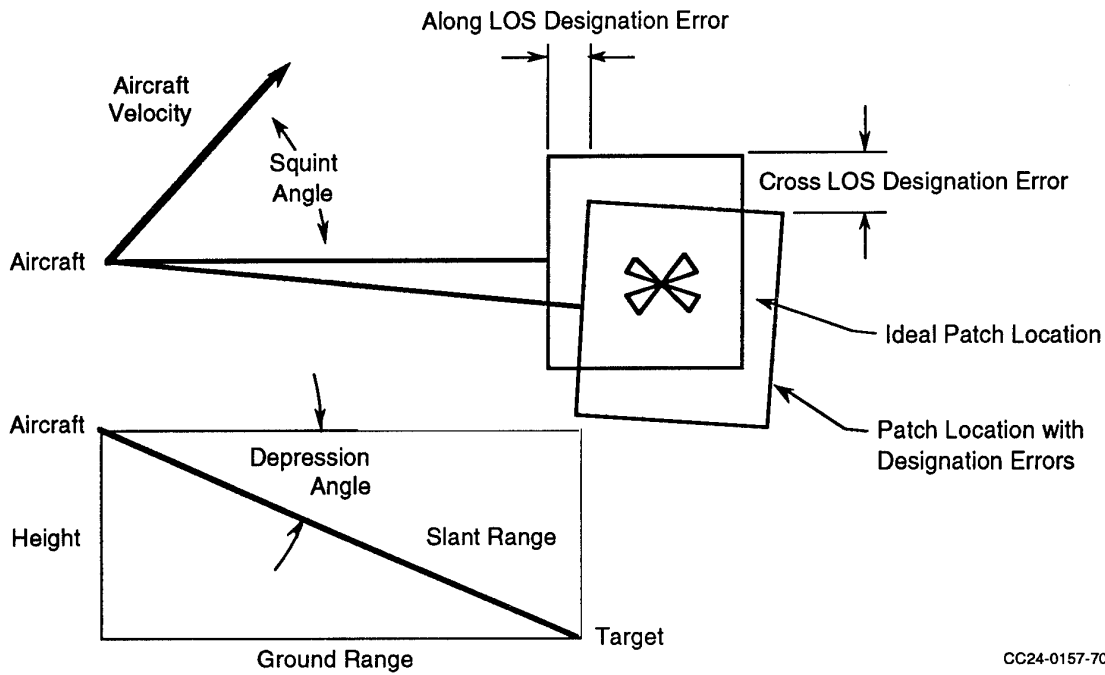
CC24-0157-70-D

Figure 3.1-12. Radar SAR Map Geometry for Designation Error Determination

Error Source	Sensitivity Equations	
	Along LOS	**Cross LOS**
Operator Designation Cursor Related Ranging - Doppler Process	$d_R \bullet N_{pxl}$	$d_A \bullet N_{pxl}$
Aircraft Altitude	$\dfrac{h}{R_s \cos\phi} \bullet \Delta h$	$\dfrac{h}{R_s \tan\theta \cos^2\phi} \bullet \Delta h$
Velocity - x	$T_n \cos\theta \bullet \Delta V_x$	$\left[\dfrac{R_s}{V \tan\theta} + T_n \sin\theta \right] \bullet \Delta V_x$
Velocity - y	$T_n \sin\theta \bullet \Delta V_y$	$\left[\dfrac{R_s}{V} - T_n \cos\theta \right] \bullet \Delta V_y$
Velocity - z	0	$\dfrac{h}{V \sin\theta \cos\phi} \bullet \Delta V_z$
Heading	0	$R_s \bullet \Delta\psi$

d_A = Azimuth Distance Resolution R_s = Slant Range ψ = Azimuth Angle

d_R = Range Resolution h = Height T_n = Nav Time

N_{pxl} = Number of Pixels θ = Squint Angle V = Aircraft Velocity

ϕ = Depression Angle

CC24-0157-71-D

Figure 3.1-13. Radar SAR Sensitivity Equations

Error Source	Units	Standard Deviation	Along LOS ft	Cross LOS ft
Operator Designation	pxl	1.50	12.8	12.8
Cursor Related	pxl	2.10	18.0	18.0
Ranging - Doppler Process	pxl	0.30	2.5	2.5
Aircraft Altitude	ft	100	5.2	5.2
Velocity - x	fps	0.80	8.5	94.8
Velocity - y	fpx	0.80	8.5	77.8
Velocity - z	fpx	2.00	0.0	16.0
Heading	mr	0.96	0.0	87.5
REP			17.0	
DEP				104
Circular Error Probable			**105 ft**	

Conditions: 15 nmi grd range, 4,800 ft alt, 500 kts,
45° squint, 15 sec nav, 8.5 ft resolution

CC24-0157-72-D

Figure 3.1-14. Radar SAR Designation Error for 8.5 feet Resolution

Error Source	Units	Standard Deviation	Along LOS ft	Cross LOS ft
Operator Designation	pxl	1.50	190.5	190.5
Cursor Related	pxl	2.10	266.7	266.7
Ranging - Doppler Process	pxl	0.30	38.1	38.1
Aircraft Altitude	ft	100	5.2	5.2
Velocity - x	fps	0.80	8.5	94.8
Velocity - y	fpx	0.80	8.5	77.8
Velocity - z	fpx	2.00	0.0	16.0
Heading	mr	0.96	0.0	87.5
REP			330	
DEP				363
Circular Error Probable			**408 ft**	

Conditions: 15 nmi grd range, 4,800 ft alt, 500 kts,
45° squint, 15 sec nav, 127 ft resolution

CC24-0157-73-D

Figure 3.1-15. Radar SAR Designation Error for 127 feet Resolution
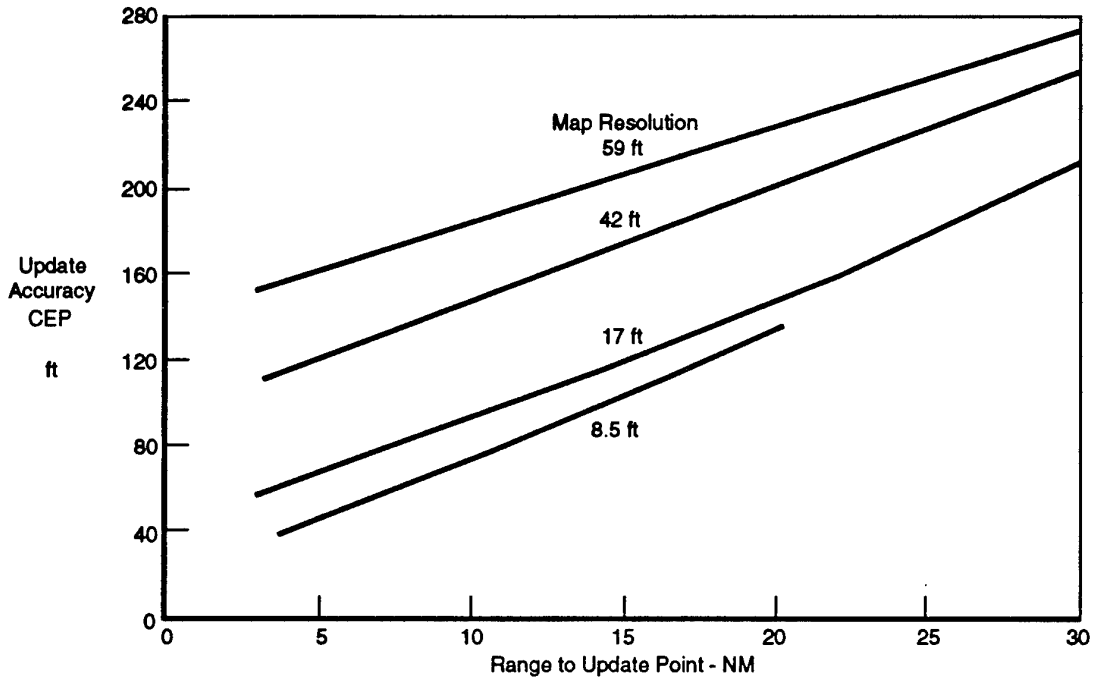
Figure 3.1-16. Position Update Accuracy with Radar SAR Map
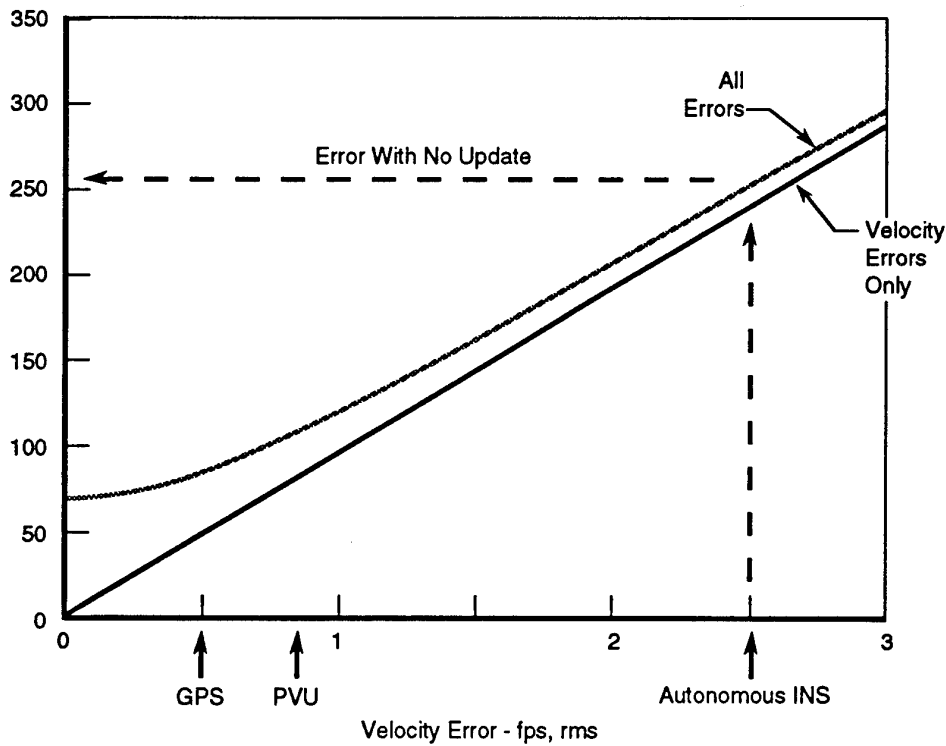
Figure 3.1-17. Designation Error Sensitivity to INS Velocity Error

354

V

Depression
Angle

Az Angle

Height

Slant
Range

Ground
Range

Cross Error

Along Error

Height Error

CC24-0157-76-D
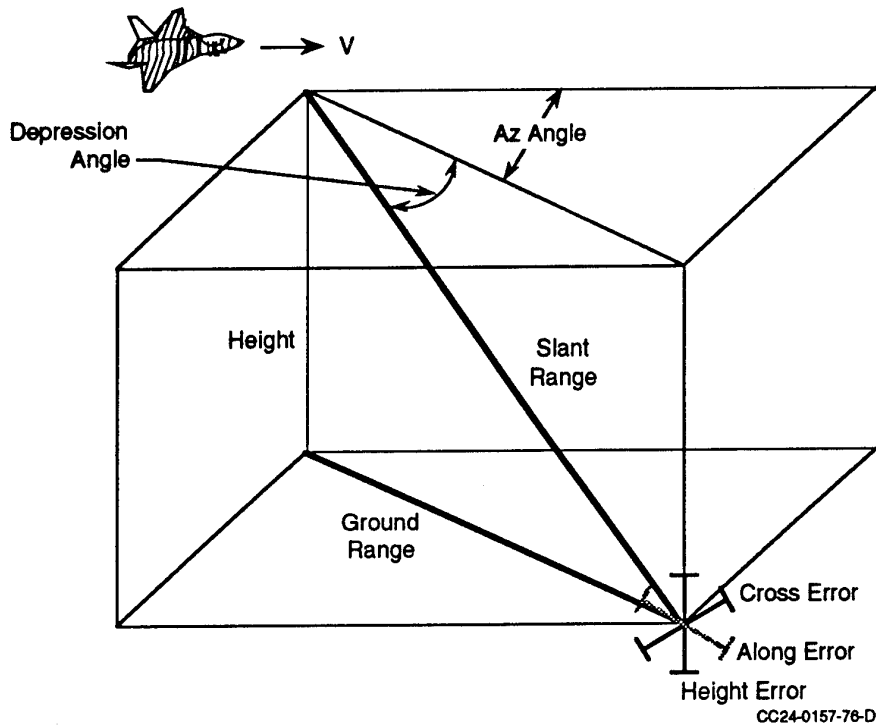
Figure 3.1-18. FLIR Geometry for Designation Error Determination

Error Source	Sensitivity Equations		
	Along LOS	Cross LOS	Height
Laser Range	$\cos\phi * \Delta Rs$	0	$\sin\phi * \Delta Rs$
Laser Range Timing	$V\cos\psi\cos^2\phi * \Delta t$	0	$V\cos\psi\cos\phi\sin\phi * \Delta t$
Laser Alignment	$Rs\cos\phi\cot\phi * \Delta\phi$	0	$Rs\cos\phi * \Delta\phi$
Gimbal Readout Timing	$V\sin^2\phi * \Delta t$	0	$V\sin\phi\cos\phi * \Delta t$
Gimbal Readout Tracking Loop Gimbal Boresight FLIR Boresight Flexure Uncertainty	$Rs\sin\phi * \Delta\phi$	$Rs * \Delta\psi$	$Rs\cos\phi * \Delta\phi$
Azimuth Reference	0	$Rs\cos\phi * \Delta\psi$	0
Elevation Reference	$Rs\sin\phi * \Delta\theta p$	0	$Rs\cos\phi * \Delta\theta p$

Rs = Slant range
ψ = Azimuth angle
ϕ = Depression angle
$\Delta\theta p$ = Pitch angle uncertainty
$\Delta\psi$ = Azimuth angle uncertainty
V = Aircraft velocity
Δt = Timing uncertainty

CC24-0157-77-D

Figure 3.1-19. FLIR Sensitivity Equations

Level, 1,000 ft

Error Source	Units	Standard Deviation	Along Track ft	Cross Track ft	Height ft
Laser Range	ft	15	14.9	0.0	1.3
Laser Range Timing	ms	33.7	28.3	0.0	2.4
Laser Alignment	mr	0.15	22.1	0.0	1.8
Gimbal Readout Timing	ms	9.6	0.1	0.0	0.7
Gimbal Readout	mr	0.985	1.0	11.8	11.8
Tracking Loop	mr	0.141	0.1	1.7	1.7
Gimbal Boresight	mr	0.1	0.1	1.2	1.2
FLIR Boresight	mr	0.37	0.4	4.5	4.5
Flexure Uncertainty	mr	0.33	0.3	4.0	4.0
Azimuth Reference	mr	0.96	0.0	11.5	0.0
Elevation Reference	mr	0.71	0.7	0.0	8.4
Subtotals (ft, 1 Sigma)			38.9	17.7	16.2
Horizontal Circular Error Probable (ft)			33.4		

Conditions: 12k ft Slant Range, 1,000 ft Altitude, 0° Squint, 500 kts

CC24-0157-78-D

Figure 3.1-20. FLIR Designation Error
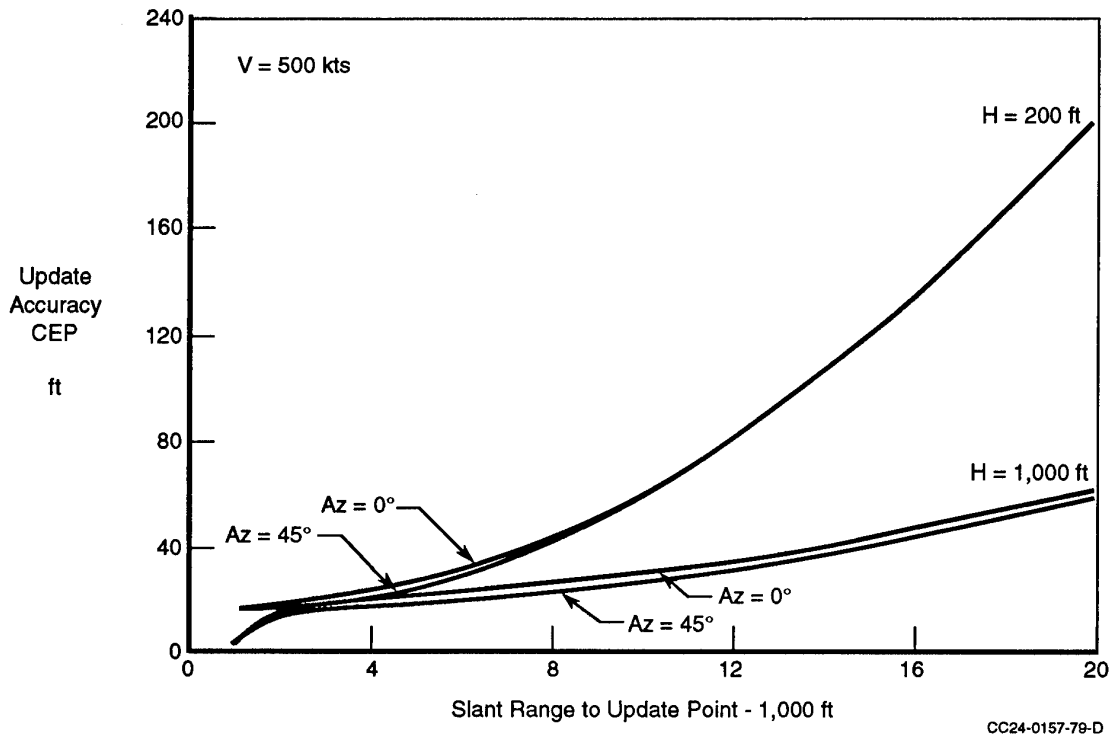
Figure 3.1-21. Position Update Accuracy with FLIR

Radar SAR Designation Analysis In the SAR mode the radar constructs a map of the cued area. Position updates or target designation is accomplished by identifying a known landmark and indicating its relative position to the map coordinates. Figure 3.1-12 shows the radar SAR map geometry for determining the designation error. The designation error consists of along LOS and cross LOS errors. The sensitivity equations used to find these errors are shown in Figure 3.1-13 for the significant contributors. Derivations of these equations can be found in References 10, 15 and 17 .

A typical SAR error budget is shown in Figures 3.1-14 and 3.1-15 for two different map resolutions. (All error contributors discussed in this section are 1 sigma values.) The dependence of position update accuracy on map resolution and update range shown in Figure 3.1-16. The parametric curve shown in Figure 3.1-17 shows the dependence of SAR map designation errors on INS velocity error. Even at 0.8 ft/s velocity error the designation error is increased significantly over what it would be if no horizontal velocity error was present. A typical INS velocity specification of 2.5 ft/s would more than double the designation error with 0.8 ft/s velocity errors. This dependence explains the need for performing a precision velocity update (typically a radar mode) immediately prior to construction of a SAR map or the use of GPS velocities to update the INS.

FLIR Designation Analysis The FLIR provides a relative position update or target designation by acquiring a landmark and indicating its relative position to the aircraft. The landmark must be recognized and designated by the pilot unless automatic target recognition is provided. If the latitude and longitude of the landmark is known, then a new aircraft absolute position can be calculated. Depending on the aircraft and electro-optical system one of two techniques is typically used for performing navigation updates and target designation. One technique uses a FLIR in combination with a laser ranger/designator. In this system the relative pointing angles to the target, in conjunction with slant range data from the laser ranger, are used to derive the aircraft

position. The other system uses track angles combined with measured angle rates and known aircraft velocity to derive aircraft position. This technique is often referred to as an angle rate system. Figure 3.1-18 shows the FLIR geometry for determining the designation errors for either FLIR system.

The sensitivity equations for the laser designator system for along LOS, cross LOS, and height are shown in Figure 3.1-19. A typical error budget for this system is shown in Figure 3.1-20. The parametric curve in Figure 3.1-21 shows the dependence of position update accuracy on slant range to target and altitude. (These update accuracy do not include any way point location uncertainties.)

① 300 NM to Target
   Last Update
   150 ft, CEP

③ Need Update ──────────
   GPS Still Inoperable
   TRN Data Base not Available
   for Region
   Find Prebriefed Waypoint
      (a) In Weather, Use Radar
      (b) Waypoint Selected can B
          With 127 ft Resolution SA
      (c) 127 ft Resolution Availab
          a 10 x 10 NM Map Size

⑤ 125 NM to Target
   15 min. @ 500 kts
   GPS, TRN, Radio
   Nav Still not Availal

② No Updates for 0.5 hr
   GPS not Available
   TRN not Functional Ove
   No Intermediate Waypo
   No Radio Nav Available
   No Star Tracker on Airc

④ Update
   Resulted
   408 ft, (

⑥ Blind Bomb Target
   • Cue Radar to T
   • Use 0.67 x 0.6
     Map Size for B
     Bombing Accu

CC24-0157-80-D

Figure 3.1-22. Scenario I: Radar Cueing

Cueing Requirements
@ (99%) Prob

Radar 20 x 20 NM Patch
(255 ft Resolution)

Radar 10 x 10 NM Patch
(127 ft Resolution)

Radar 4.67 x 4.67 NM Patch →
Radar 3.34 x 3.34 NM Patch →
FLIR 5 x 5° @ 20 NM →

Position Error
(NM CEP)

INS Drift
Rate (NM/hr, CEP)

1.0

0.8

0.6

0.42 NM/hr @ 5 hr

0.4

0.2

0.1

Time Since Last Update - hr
(Last Update Accuracy = 150 ft, CEP)
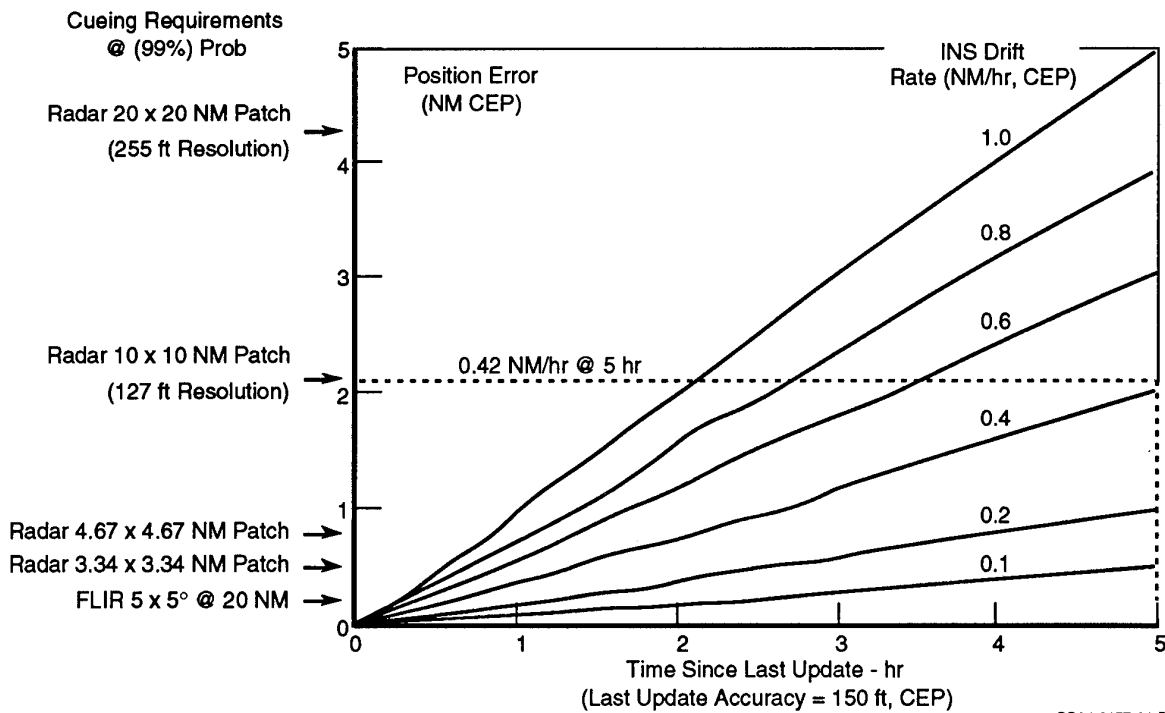
CC24-0157-81-D

Figure 3.1-23 INS Drift After 5 Hours Autonomous Flight

358
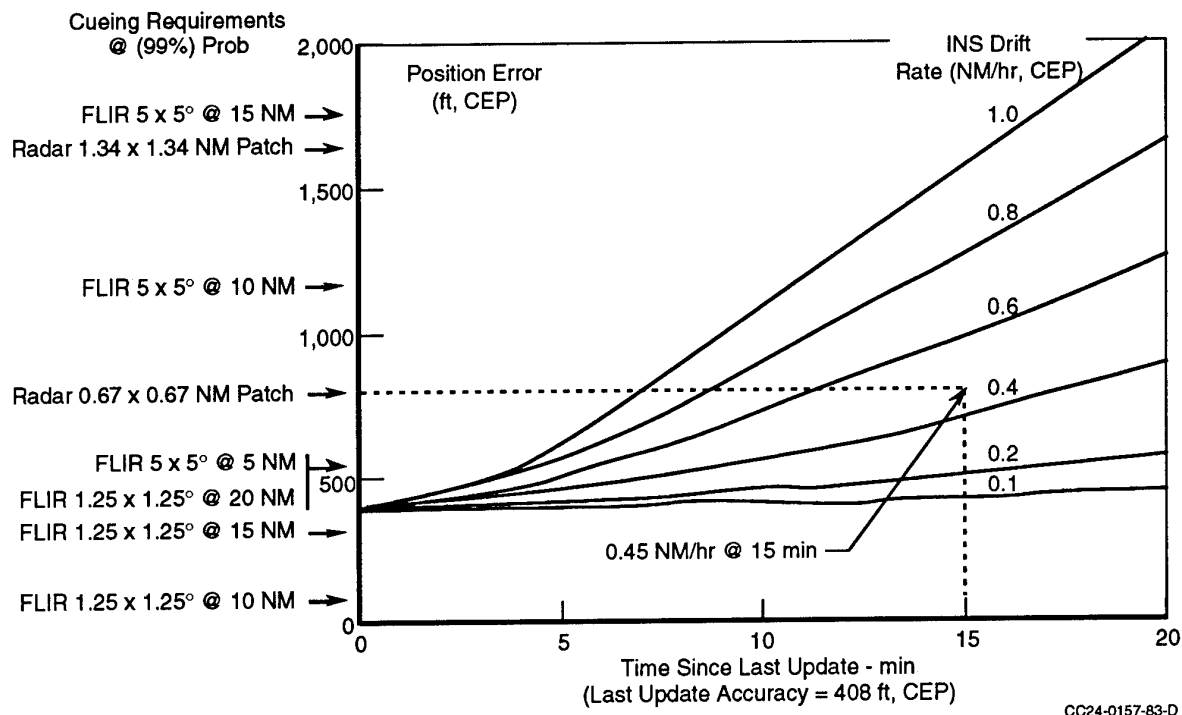
Figure 3.1-24. INS Drift After Autonomous Flight (Scenario I)

INS Drift Analysis Example The INS drift requirements for sensor cueing are defined relative to the air interdiction mission defined earlier. Two scenarios are used as examples for analysis, one using the Radar and one using the FLIR.

Example scenario I The first scenario, Figure 3.1-22 uses the radar for both way point and target acquisition. The assumptions for this scenario are that the navigation system will cue a 3.34 nm x 3.34 nm SAR radar map after a .5 hour flight with no navigation aids available. With an update accuracy of 170 ft (consistent with the SAR map size), the navigation system must cue a 0.67 nm x 0.67 nm radar map to a target after an additional 15 minutes of unaided flight.

The graphs in Figure 3.1-23 show aircraft position error propagation after an initial position update accuracy of 150 ft. CEP (an initial position fix from GPS or terrain refer-

ence navigation has been assumed). The initial error (150 ft, CEP) is combined in an rss manner with the product of INS drift and time. Shown are drifts from 0.1 nm/hour to 1.0 nm/hour, CEP.

The INS drift requirement for the first update using the 3.34 nm x 3.34 nm map is found by using the position accuracy requirement for the 3.34 nm x 3.34 nm from Figure 3.1-6 above (4240 ft) and the curves of Figure 3.1-23. The required INS drift is found on Figure 3.1-23 at the intersection of the allowable position error for the 3.34 nm x 3.34 nm map and the time of the last update. The first update requirement is found to be in excess of 1 nm/hr. The drift requirement for cueing the .67 nm x .67 nm SAR map for target acquisition is found in a similar manner. From Figure 3.1-24 the allowable drift rate is found to be 0.45 nm/hr CEP for .99 probability.

① **Last Update with TRN, 150 ft, CEP**
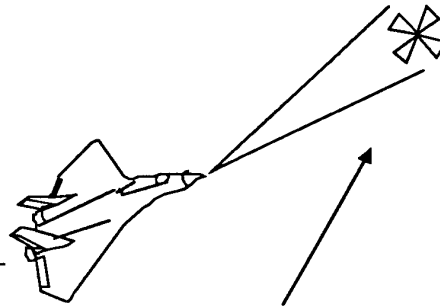
② **No Update for 10 min**
   – GPS Not Available
   – Radar Alt Turned Off for Last 80 nmi
   – Radar Not Used for Acquisition

③ **Acquire Target**
   – Cue FLIR to Target at 5 nmi
   – Use 5° x 5 FOV
   – See Weapon Delivery
     Analysis for CEP Achieved

CC24-0157-82-D

Figure 3.1-25. Scenario II: FLIR Cueing

Cueing Requirements @ (99%) Prob

FLIR 5 x 5° @ 15 NM →
Radar 1.34 x 1.34 NM Patch →

FLIR 5 x 5° @ 10 NM →

Radar 0.67 x 0.67 NM Patch →

FLIR 5 x 5° @ 5 NM →
FLIR 1.25 x 1.25° @ 20 NM
FLIR 1.25 x 1.25° @ 15 NM →

FLIR 1.25 x 1.25° @ 10 NM →

Position Error (ft, CEP)

INS Drift Rate (NM/hr, CEP)

0.51 NM/hr @ 10 min

Time Since Last Update - min
(Last Update Accuracy = 150 ft, CEP)
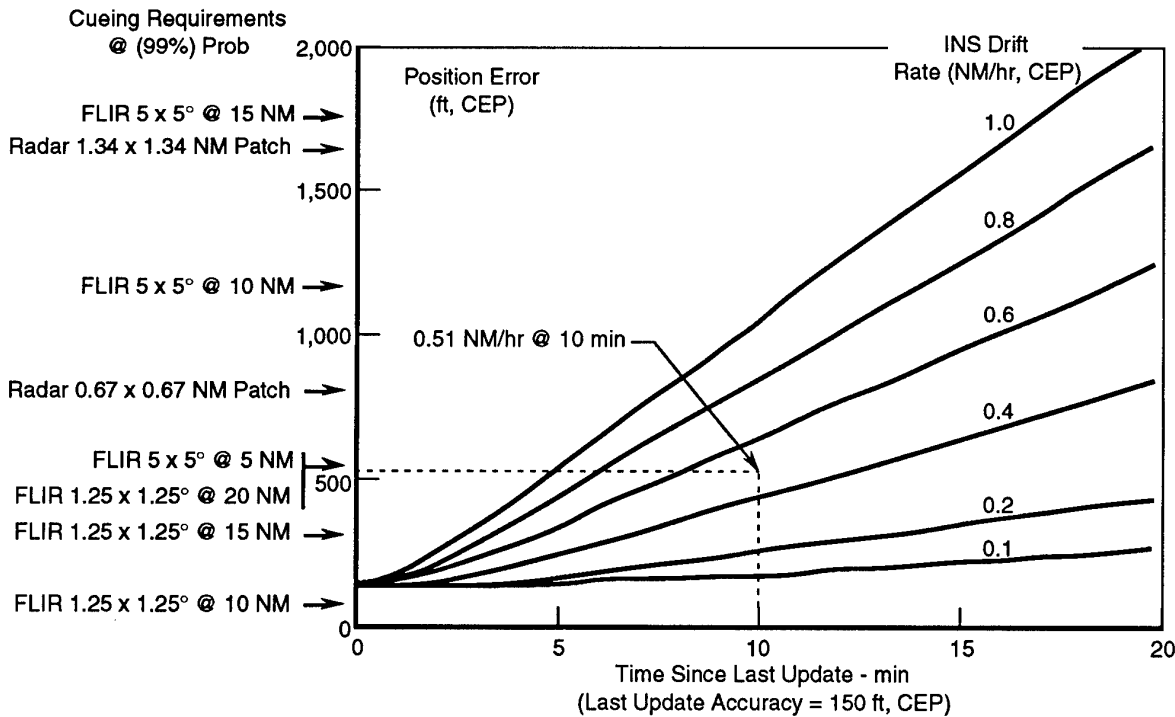
CC24-0157-84-D

Figure 3.1-26. INS Drift After Autonomous Flight (Scenario II)

Example scenario II The second scenario (Figure 3.1-25) is the same as the first except in this case the FLIR is used to acquire and track a ground target to support a bombing mission. This scenario requires cueing the FLIR to a target after 10 minutes of unaided flight. Figure 3.1-26 shows the requirement for cueing the 5 degree x 5 degree FLIR wide FOV. From this illustration the INS drift allowed is 0.51 (P=0.99) nm/hour, CEP.

Map; 8.5 ft, 10 NM 45° Squint, Release; 540 kts, 1,000 ft Level

	Error Source	Units	Standard Deviation	Along Track ft	Cross Track ft
1	Operator	pix	1.5	12.75	12.75
2	Cursor	pix	2.1	18.00	18.00
3	Range/Doppler	pix	0.29	2.45	2.45
4	Altitude	ft	100	5.23	5.23
5	Radar Velocity - x	f/s	0.8	39.17	92.51
6	Radar Velocity - y	f/s	0.8	39.17	14.16
7	INS Velocity - x	f/s	2.5	9.04	12.79
	INS Velocity - y	f/s	2.5	9.04	5.28
8	INS Velocity - Vertical	f/s	2.0	53.91	9.87
9	Attitude Angle Reference (El)	mr	0.713	17.39	–
	(Az)	mr	0.96	–	0.18
10	Radar Slant Range	X	3.3	16.27	–
11	Beam LOS Position	mr	1.78	43.46	–
12	True Airspeed	f/s	1.69	0.38	–
13	Pilot Steering	mr	2.50	–	17.63
14	Bomb - Ejection Velocity	f/s	1.15	30.76	–
15	Bomb - Release Time Delay	ms	3.0	2.73	–
16	Bomb - Dispersion	mr	2.0	51.27	14.24
17	Ballistic Fit	mr	0.33	3.47	–
		ft		REP = 78	
		ft			DEP = 66
		ft	CEP = 126		

CC24-0157-89-D

Figure 3.1-27 Radar SAR Blind Bombing Error Budget

Level, 1,000 ft

Error Source	Units	Standard Deviation	Along Track ft	Cross Track ft
Aircraft Position at Release				
Slant Range	ft	21	10.01	0.0
Ranger Boresight	mr	0.2	14.00	0.0
FLIR El Angle	mr	1.2	27.90	0.0
FLIR Az Angle	mr	1.3	0.00	8.7
Pilot Az Steering	mr	3.0	0.00	20.4
Weapon Range				
Bomb Ejt Velocity	fps	2.0	52.50	0.0
Release Time Delay	ms	10.0	9.10	0.0
Bomb Dispersion	mr	3.0	75.40	20.6
INS Velocity Reference	fps	3.0/3.0	22.80	22.8
INS Vert Velocity	fps	2.0	52.90	0.0
INS Elevation Ref	mr	1.5	38.20	0.0
INS Boresight (Elevation)	mr	1.5	38.20	0.0
True Airspeed	kts	2.7	0.47	0.0
Side-Slip Angle	mr	1.5	0.00	0.1
Ballistic Fit	mr	0.33	15.50	0.0
Total 1 Sigma Error	ft		126.80	37.9
Circular Error Probable			**94.7 ft**	

CC24-0157-90-D

Figure 3.1-28. FLIR with Laser Ranger Bombing Error Budget

$$\vec{V} = \vec{\omega} \times \vec{R}_{sl}$$

$$R_{sl} = \frac{-H}{\sin(\sigma)}$$

$$H = \frac{-V_{tas}\sin(\lambda)}{\omega}\sin(\sigma) - \frac{\left(V_{wind} - V_{tar}\right)\sin^2(\sigma)}{\omega}$$

$$R_{tar} = R_{sl}\cos(\sigma)$$

CC24-0157-91-D
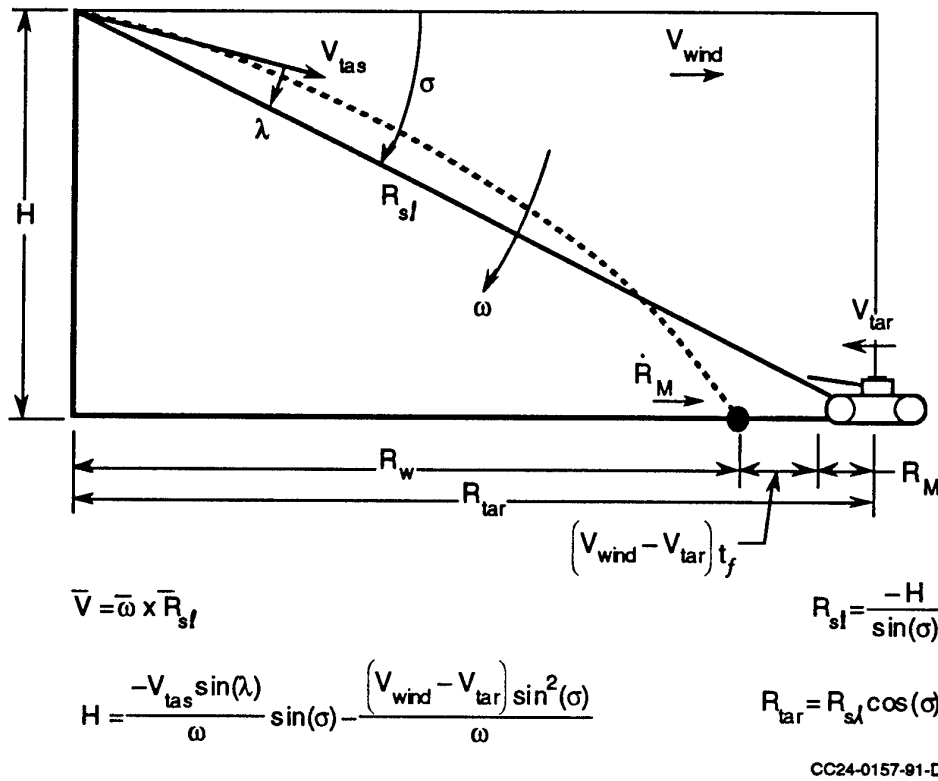
Figure 3.1-29. Angle Rate Bombing Geometry

Level, 1,000 ft

Error Source	Units	Standard Deviation	Along Track ft	Cross Track ft
Aircraft Position at Release				
FLIR Az/El	mr	1.5/1.5	7.7	0.07
FLIR Boresight	mr	7.5/1.5/1.5	7.7	7.8
El Angle Rate	mr/s	0.2	30.9	0.0
Az Angle Rate	mr/s	0.1	0.0	5.0
FCS Roll Ref	mr	10.0	0.0	10.4
INS Angle Ref	mr	1.45/1.45/1.45	3.2	0.2
Pilot Az Steering	mr	3.0	0.0	20.5
Weapon Range				
Bomb Ejt Velocity	fps	2.0	50.7	0.1
Release Time Delay	ms	10.0	9.4	0..0
Bomb Dispersion	mr	3.0	71.1	20.5
INS Velocity Reference	fps	3.0/3.0	12.3	0.6
INS Vert Velocity	fps	2.0	0.0	0.0
INS Boresight	mr	3.0/1.5/1.5	3.3	2.8
True Airspeed/AOA	kts/mr	2.7/4.2	11.8	0.8
AOA Boresight	mr	2.1	5.9	0.0
Ballistics Fit	mr	0.33	15.5	0.0
Total 1 Sigma Error	ft		96.8	32.3
Circular Error Probable		74.3 ft		

CC24-0157-92-D

Figure 3.1-30 Angle Rate Bombing Error Budget

On the basis of the scenarios and sensors analyzed the requirements placed on the navigation system to support sensor cueing and way point/target designation are drift rate and velocity accuracy. For the specific cases presented the constraining condition for drift rate is the long duration flight without position updates. For this case hand-off to a 3.34 nm x 3.34 nm SAR map with a 0.99 probability after a 15 minute flight requires an INS with a drift rate of less than 0.45 nm/hr, CEP. Obviously, changes in the particular aspects of the scenario will change the limiting conditions.

Weapon Delivery Bombing Modes Analysis The navigation system provides the velocity and attitude reference required to compute weapon ballistic trajectory. Errors in the navigation parameters propagate into bomb impact errors.

The methodology used to determine the contribution of error sources to bomb impact error is analogous to that used for navigation and designation updates. First, sensitivity equations are determined by differentiating the mechanization equations with respect to the error sources of interest. Weapon release conditions (mode, altitude, dive angle, and velocity), and error source values (target sensor errors, inertial reference errors, pilot steering error, bomb dispersion, etc.) are entered into the sensitivity equations. The resulting contributions from each error source are combined together to compute the bomb impact error. The impact error is then compared to the requirement. Requirements are iterated until a suitable compromise is made between equipment specifications and equipment producibility.

Blind Bombing Analysis The most stringent velocity reference requirements come from blind bombing. The target is acquired using the radar SAR mode as discussed above. The bomb impact error is equal to the rss of (1) the target designation error, (2) relative navigation error incurred from target designation to weapon release, (3) the release error due to uncertainty in the release conditions, and (4) bomb dispersion after release.

Figure 3.1-27 shows a blind bombing error budget. The velocity reference uncertainties dominate. Most of the level velocity contribution is due to the designation error. The velocity accuracy required to limit designation error to a reasonable value, and thus limit bomb impact error, is 0.8 fps, 1 sigma.

Conventional Bombing Analysis Conventional bombing with a head up display or FLIR sensor requires less velocity accuracy. Designation with the FLIR is similar to way point designation as previously discussed except that the target is locked on and tracked until weapon release. Figure 3.1-28 shows an error budget for bombing with a FLIR and laser ranger. The error contribution due to velocity is significant but not dominant.

Angle Rate Bombing Analysis In angle rate bombing (ARB), a FLIR tracks the target, and provides line-of-sight angles and angle rates to the target relative to the aircraft. These angles and angle rates are an implicit measure of range to the target and target relative velocity normal to the line-of-sight. The ARB geometry and relevant equations are shown in Figure 3.1-29. The advantage of angle rate bombing is that the first order effects of aircraft INS velocity errors cancel. The nullifying effect results from the fact that both target range and weapon range are nearly linear functions of aircraft velocity. A positive error in velocity would indicate both an increased target range and a compensating increase in weapon range therefore nullifying the impact of the error. Figure 3.1-30 shows an error budget for angle rate bombing. The inertial velocity and attitude contributions are small compared to the total error.

From examination of the budgets the navigation system requirements to support weapon delivery are primarily for aircraft velocity. Of the three modes reviewed "blind bombing" is the only mode that delivery accuracy is directly dependent on velocity accuracy. It is for this reason that blind bombing is usually supported by a precision velocity update either from radar or GPS. In the other modes of automatic weapon delivery the sensor (FLIR or radar) provides most of the precision targeting data with the navi-

gation system supporting the sensor. The error budgets in Figures 3.1-28 and 3.1-30 show that the relative impact of navigation system velocity error on the ARB system is less than for conventional bombing.

Vehicle Management System (VMS) Analysis - Basic Flight Overall, the basic flight functions do not require highly accurate inertial state vector information. The inertial sensing requirements can be readily met with an Attitude Heading Reference System (AHRS). Pilot relief modes also require velocity information which can be supplied by the reference INS. However, inertial data must be provided with enough fault tolerance and redundancy to must meet safety-of-flight issues (Reference 5).

Air Data Backup Analysis If the Air Data System fails, inertial data can be used to provide an air data backup capability. Inertially derived velocity data can be combined with aircraft attitude to estimate angle of attack and side slip angle needed for flight control. These estimates ignore the effects of winds, and therefore would support a degraded mode of operation (Reference 5).

Pilot Relief Modes Analysis Autopilot functions are provided through the VMS computer using inputs from the aircraft motion sensors and air data sensors. Autopilot modes provided are navigation steering mode for great circle course guidance, ground track steering mode, attitude/heading hold, and barometric or radar altitude hold. The pilot relief modes are non-critical both in terms of flight safety and mission success (i.e. the mission can still be completed with this mode of operation although a greater workload may be placed on the crew) and are readily met with a medium accuracy inertial navigation system or Attitude Heading Reference System (AHRS) (Reference 5).

Vehicle Management System (VMS) Analysis - Integrated Modes Overall, the integrated flight modes of the VMS do not require highly accurate inertial state vector information. The inertial sensing requirements can be easily met with an aided medium accuracy INS. Inertial data must be provided with enough fault tolerance and re-

dundancy to must meet safety-of-flight issues (Reference 5).

Terrain Following/Terrain Avoidance (TF/TA) Analysis TF/TA mode requires both absolute navigation data and relative navigation data. Absolute position accuracy of 0.5 nm CEP is required for initial way point acquisition. Velocity accuracy of 2.0 ft/sec (1 sigma) is required for use in computation of commands for control of the aircraft velocity vector. Relative navigation data will be provided by use of radar and radar altimeter to achieve low level TF/TA. Terrain following systems that blend radar and DTED or rely solely on DTED for terrain definition require navigation position accuracy better than 300 ft and a heading accuracy of 0.1 deg (References 5, 10, 15).
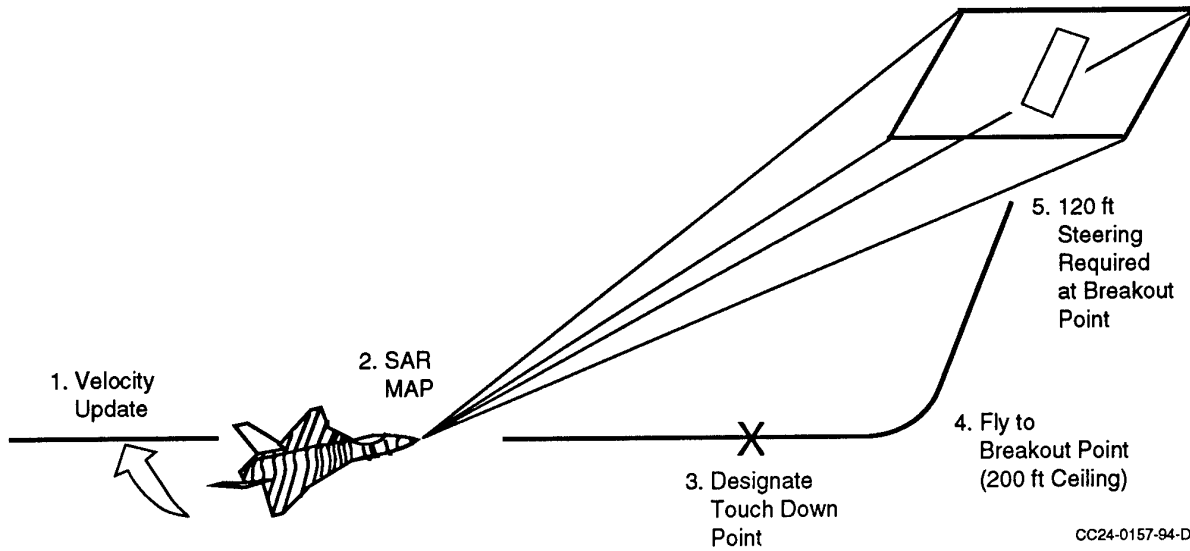
Figure 3.1-31. Autonomous Precision Landing

	INS Velocity Error = 0.5 fps		INS Velocity Error = 1.0 fps		INS Velocity Error = 2.0 fps	
	Along Track (ft)	Cross Track (ft)	Along Track (ft)	Cross Track (ft)	Along Track (ft)	Cross Track (ft)
Operator Designator Error	13	13				
Cursor Related Errors	18	18				
Ranging - Doppler Filter Process Errors	9	9				
Output Lag Error	5	5	Same	Same	Same	Same
Aircraft Altitude at MAP	5	5				
Radar Velocity - Along-Track	44	78				
Radar Velocity - Cross-Track	98	22				
INS Velocity - Along Track	11	11	22	22	44	44
INS Velocity - Cross Track	11	11	22	22	44	44
INS Velocity - Vertical	—	9	—	18	—	36
CEP - ft	117		122		141	

Note 1: Pilot makes steering corrections upon breakout

CC24-0157-93-D

Figure 3.1-32. Autonomous Precision Landing Touchdown Point Accuracy

Autonomous Precision Landing Analysis
Navigation parameters required to support an autonomous precision landing capability are present position, velocity, attitude, and heading as illustrated in Figure 3.1-31. A present position accuracy of 2 nm CEP is required for navigation to the landing area. On board target sensors then locate and designate the touchdown point. A position error subsequently propagates due to several error sources. This position error must be corrected by the pilot prior to touchdown. The integration of INS velocity errors over time after sensor designation contribute to

3.2 Tactical Fixed Wing Aircraft Requirements Allocation (Reference 10)

Vehicle Management System Requirements Allocation

The basic flight functions do not require highly accurate inertial state vector information. The inertial sensing requirements can be readily met with an Attitude Heading Reference System (AHRS). Based on the results from the Advanced Vehicle Management System Architecture Study (AVMS), a triplex fly-by-wire flight control

Error Source	Units	Refresh Rate
Body Axis Rate	deg/sec	200 Hz
Body Axis Acceleration	ft/sec/sec	200 Hz
Aircraft Attitude	deg	200 Hz
True Heading	deg	200 Hz
Ground Speed	knots	50 Hz
Ground Track Angle	deg	50 Hz
Velocity (NED)	ft/sec	200 Hz
Velocity (Body Axis)	ft/sec	50 Hz
Acceleration, Vertical	ft/sec/sec	50 Hz

Figure 3.2-1: Typical INs Outputs to Flight Control Systems

position errors as shown in Figure 3.1-32. Based on the total touchdown errors presented in Figure 3.1-32, an INS velocity accuracy requirement of + 1 ft/sec is established. Attitude and heading accuracy of 0.1 degrees (1 sigma) is required for final approach and landing (Reference 10).

system was assumed to the most cost-effective means for satisfying flight safety requirements. The VMS must provide aircraft stability and performance which satisfy the requirements of MIL-F-8785 and MIL-F-9490D for Class IV aircraft with Level I flying qualities in the A, B, and C flight phase categories. Aircraft body rates and acceleration sensing must be fail-op/fail-op, survivable, and have a mission reliability of 1.0 E-10. Typical flight control data are shown in Figure 3.2-1.

366

Parameter [1]	Units	Range	Accuracy	Mean Accuracy	Jitter	Quantization/ Resolution	Refresh Rate	Data Latency (Max)
Latitude	deg	+/- 90	0.8 nm/hr CEP 1 hr mission		2.14x10E-5	8.38x10E-8	25 Hz	not specified
Longitude	deg	+/- 180	1.7 nm/hr 95% Probability		2.14x10E-5	8.38x10E-8	25 Hz	not specified
Baro-Inertial Altitude (2)	ft	-1k to 75k	150	149.5	12	2.0	50 Hz	
True Heading	deg	+/- 180	0.05	0.048	0.012	0.005	50 Hz	10.5 msec
Roll	deg	+/- 180	0.033	0.03	0.012	0.005	50 Hz	10.5 msec
Pitch	deg	+/- 90	0.033	0.03	0.012	0.005	50 Hz	10.5 msec
N/S Velocity	ft/sec	+/- 3000	2.5	2.5	0.125	0.125	50 Hz	55 msec (3,4)
E/W Velocity	ft/sec	+/- 3000	2.5	2.5	0.125	0.125	50 Hz	55 msec (3,4)
Vertical Velocity (2)	ft/sec	+/- 1500	2.0	2.0	0.0625	0.0625	50 Hz	55 msec (3,4)
N/S Acceleration	ft/sec/sec	+/- 9 G	0.5	0.064 (6)	0.49	0.0078	50 Hz	35 msec
E/W Acceleration	ft/sec/sec	+/- 9 G	0.5	0.064 (6)	0.49	0.0078	50 Hz	35 msec
Vertical Acceleration	ft/sec/sec	- 4 to +12 G	0.5	0.064 (6)	0.49	0.0156	50 Hz	35 msec
Roll Rate	deg/sec	+/- 300	0.054	0.04	0.036	0.0156	200 Hz	14.5 msec
Pitch Rate	deg/sec	+/- 200	0.054	0.04	0.036	0.0078	200 Hz	14.5 msec
Yaw Rate	deg/sec	+/- 200	0.054	0.04	0.036	0.0078	200 Hz	14.5 msec
Long Acceleration	ft/sec/sec	+/- 12 G	0.5	0.064	0.49	0.0156	200 Hz	23 msec
Lateral Acceleration	ft/sec/sec	+/- 4 G	0.5	0.064	0.49	0.0156	200 Hz	23 msec
Normal Acceleration	ft/sec/sec	+/- 9 G	0.5	0.064	0.49	0.0156	200 Hz	23 msec
Wander Angle (Alpha)	deg	+/- 180	Note (5)	Note (5)	0.012	0.0055	25 Hz	not specified
Roll Angle Accel	deg/sec/sec	+/- 1000	2.65	0.5	2.6	0.17	200 Hz	23 msec
Pitch Angle Accel	deg/sec/sec	+/- 200	2.65	0.5	2.6	0.17	200 Hz	23 msec
Yaw Angle Accel	deg/sec/sec	+/- 200	2.65	0.5	2.6	0.17	200 Hz	23 msec
Ground Speed	ft/sec	0 to 3000	3.4 ft/sec	3.4	0.125	0.125	50 Hz	not specified
Gorund Track Angle	deg	+/- 180				0.0055	50 Hz	not specified
Dir Cosines (Cyx)	none	+/- 1	Note (7)		Note (7)	2.38x10E-7	200 Hz	not specified
Dir Cosines (Cyy)	none	+/- 1	Note (7)		Note (7)	2.38x10E-7	200 Hz	not specified
Dir Cosines (Cyz)	none	+/- 1	Note (7)		Note (7)	2.38x10E-7	200 Hz	not specified
Time Tag (1553B)	μs		50			50	(8),(9)	
Nav Boresight	mrad		0.25					

CC24-0157-105-C

Notes: All values are RMS except as noted

(1) All data shown is based upon a full performance gyrocompass alignment
(2) Requires IIN Navigation Updating - pressure altitude aiding is used to stabilize the vertical channel
(3) Includes digital "body bending" filter to remove airframe cycle effects
(4) Velocities can be lead compensated using acceleration to minimize delay (< 2.5 msec)
(5) True Heading, Platform Azimuth, and Wander Angle accuracies shall comply with the following equation

$$\text{True Heading Accuracy} >= \sqrt{[(\text{Wander Angle Accuracy})^2 + (\text{Platform Azimuth Accuracy})^2]}$$

(6) X,Y,Z Acceleration accuracy = 0.5 ft/sec/sec = $\sqrt{[(0.064 \text{ ft/sec/sec})^2 + (0.49 \text{ ft/sec/sec})^2]}$
(7) Accuracy shall be consistent with source data (Wander Azimuth (Alpha) and Latitude)
(8) The refresh rate shall be a multiple of 25 Hz
(9) The refresh rate of the time tag is the rate of the associated data
(10) The Aircraft c.g. State Estimation (Reference Nav) Vector Defines the Aircraft Rigid Body Axis

CC24-0157-106-D

Figure 3.2-2: Aircraft c.g. State Estimation Performance Allocations

(0.8 Nm/Hr INS)

Mission Related		O Aided Inertial			
Top Level Functions	Events &Tasks	Position	Altitude	Horizontal Velocity	Vertical Velocity[1]
Steering / Penetration	TF/TA - Passive (GPS, DTED)	300 ft [2]	30 ft [3]	2.0 ft/s	2.0 ft/s
	TF/TA - Active (TF Radar)		[3],[4]	2.0 ft/s	2.0 ft/s
	Threat Avoidance	300 ft [2]			
Sensor Management	FLIR Sensor Cueing	235 ft CEP [5]			
	Radar Sensor Cueing	800 ft CEP [6]			
Designation Management	Visual Target Designation		100 ft		
	Radar Target Designation		100 ft	0.8 ft/s	
Air-to-Surface	Guided Weapons	1177 ft CEP			
Landing / Approach	Autonomous Landing	129 ft CEP	5 ft	1.0 ft/s	

Flight Related					
Vehicle Management Functions		Position	Altitude	Horizontal Velocity	Vertical Velocity
Integrated Modes	IFPC - Flight Path Optimization	2000 ft CEP			

Note: All values RMS except as noted
(1) Requires pressure altitude aiding
(2) DTED position accuracy 300 ft / cell
(3) Relative altitude (AGL - Above Ground Level)
(4) Driven by radar altimeter accuracy requirements
(5) Cue 5 deg x 5 deg (WFOV) FLIR at 3 NM from target with P=0.99
(6) Cue 0.67 NM x 0.67 NM SAR map with P=0.99

CC24-0157-108-D

Figure 3.2-3: "Aided" Aircraft c.g. State Estimation Performance Allocations

Table 3.2-1: Flight Control Pitch and Yaw Gyro Performance Allocations

Range	±200 deg/sec
Linearity*	±0.5% Full Scale (0-50% Full Scale)
	±3.0% Full Scale (50-100% Full Scale)
Zero Offset	0.10 deg/sec
Threshold	0.02 deg/sec
Resolution	0.02 deg/sec
Hysteresis	0.05 deg/sec
Natural Frequency	>30 Hz
Damping	0.5 to 1.0 (−20 to +160 °F)
Cross-Axis Sensitivity	0.05 deg/sec/deg/sec
Residual Oscillation and Noise	0.05 deg/sec (0-5 Hz)
	0.10 deg/sec (5-20 Hz)
Scaling Tolerances	±5%

* Linearity is defined as deviation of actual output at any point from the least square straight line fit of measured input/out relationship

CC24-0157-58-D

Table 3.2-2: Flight Control Roll Gyro Performance Allocations

Range	±300 deg/sec
Linearity*	±0.5% Full Scale (0-50% Full Scale)
	±2.0% Full Scale (50-100% Full Scale)
Zero Offset	0.80 deg/sec
Threshold	0.02 deg/sec
Resolution	0.02 deg/sec
Hysteresis	0.30 deg/sec
Natural Frequency	>30 Hz
Damping	0.5 to 1.0 (−20 to +160 °F)
Cross-Axis Sensitivity	0.05 deg/sec/deg/sec
Residual Oscillation and Noise	0.05 deg/sec (0-5 Hz)
	0.10 deg/sec (5-20 Hz)
Scaling Tolerances	±5%

* Linearity is defined as deviation of actual output at any point from the least square straight line fit of measured input/out relationship

CC24-0157-59-D

Table 3.2-3: Flight Control Lateral and Longitudinal Accelerometer Performance Allocations

Range	±3G
Linearity*	+1% Full Scale
Zero Offset	0.01G
Threshold	0.001G
Resolution	0.001G
Hysteresis	0.001G
Natural Frequency	>30 Hz
Damping	0.5 to 1.0
Cross-Axis Sensitivity	0.002 G/G
Residual Oscillation and Noise	0.004G
Scaling Tolerances	±2.5%

* Linearity is defined as deviation of actual output at any point from the least square straight line fit of measured input/out relationship

CC24-0157-60-D

Table 3.2-4: Flight Control Normal Accelerometer Performance Allocations

Range	±12G
Linearity*	+1% Full Scale
Zero Offset	0.05G
Threshold	0.005G
Resolution	0.005G
Hysteresis	0.005G
Natural Frequency	>30 Hz
Damping	0.5 to 1.0
Cross-Axis Sensitivity	0.002 G/G
Residual Oscillation and Noise	0.002G
Scaling Tolerances	±2.5%

* Linearity is defined as deviation of actual output at any point from the least square straight line fit of measured input/out relationship

CC24-0157-61-D

Table 3.2-5: AHRS Accuracy Allocations

Parameter	Units	Requirement with Aiding *	Requirement without Aiding **
Heading	degs, RMS	0.25	0.25 Initial
Heading Drift	degs/hour, RMS	n/a	1
Attitude	degs, RMS	1	1.5
Horizontal Velocities	fps, RMS	3	16

* INS and air data aiding
** AHRS must meet requirement for 15 minutes after loss of aiding.

Aircraft cg State Estimation Performance Allocations

The aircraft c.g. state estimation performance allocations are shown in Figure 3.2-2, which is representative of medium accuracy 0.8 nm/hr inertial navigation system. Aided requirements are shown in Figure 3.2-3 which require navigation updating with an external navigation aiding sensor.

The VMS must provide aircraft stability and performance which satisfy the requirements of MIL-F-8785 and MIL-F-9490D for Class IV aircraft with Level I flying qualities in the A, B, and C flight phase categories. Aircraft body rates and acceleration sensing requirements to provide this performance are listed in Tables 3.2-1 through 3.2-4. Inertial sensing must be fail-op/fail-op, survivable, and have a mission reliability of 1.0 E-10. Equivalent AHRS allocations needed for the flight control function is shown in Table 3.2-5.

Figure 3.2-4: Local State Estimation Functional Allocations

Mission Related — Top Level Functions	Functions & Tasks	Present Position	Attitude	Attitude	True Heading	Velocity	Acceleration	Body Rates	Angular Acceleration	Manufacture Tolerance-7.5mr	Static Boresight (2)	Quasi-Steady State Flex (3)	Vibration (3)
Stores Management	Weapon Transfer Alignment (4)	●	●	●	●	●				(7)			
Sensor Management	FLIR Stabilization		●	●	●	●	●						
Sensor Management	Radar Motion Compensation		●	●	●			●	●				
Air-to-Surface	Ejection Velocity (Stick Bomb.)					●							
Weapon / Sensor Alignment	Weapon/Sensor Alignment (5)												
Weapon / Sensor Alignment	- Nav FLIR, IRST (6)		●							●	●	●	(8)
Weapon / Sensor Alignment	- LTD/R, Targeting FLIR (6)		●							●	●	●	(8)
Weapon / Sensor Alignment	- Radar (SAR)		●							●	●	●	●
Weapon / Sensor Alignment	- HUD		●							●	●	●	
Weapon / Sensor Alignment	- HMD		●							●	●		
Weapon / Sensor Alignment	- Flush Air Data Ports		●							●	●		
Weapon / Sensor Alignment	- MWR, LWR (6)		●							●	●	●	
Weapon / Sensor Alignment	- RWR		●							●	●	●	
Weapon / Sensor Alignment	- MK 84 GPB		●							●			
Weapon / Sensor Alignment	- GBU-24 LGB		●							●			
Weapon / Sensor Alignment	- AGM-65 EO/IR Maverick		●							●	●	●	
Weapon / Sensor Alignment	- AIM-9M Sidewinder		●							●	●	●	
Weapon / Sensor Alignment	- M61A1 20 mm Gun		●							●	●	●	
Weapon / Sensor Alignment	- DEW		●							●	●	●	(8)
Weapon / Sensor Alignment	- Basic Navigation Sensors		●							●	●	(9)	
Weapon / Sensor Alignment	- Flight Control Sensors		●							●			

CC24-0157-109-D

Notes:

(1) Flexure Results from Changes in the Aircraft's Reference Frame (Rigid Body) Caused by Factors such as Maneuvers, Structural Bending Modes and High Frequency Vibration Effects

(2) Static Boresight Improves Manufacturing Boresight Tolerances

(3) Quasi-Steady State Flexure Bandwidth (0.01 - 1 Hz), Vibration Flexure Bandwidth (> 1 Hz)

(4) Transfer Alignment Requires an Inertial Reference Source, Alignment is with Respect to Inertial Navigation Frame (Weapons Include: AGM-65 MMW Maverick, JDAM, JSOW, SLAM, AGM-86, ALCM, AIM-120 AMRAAM)

(5) Misalignment is Taken with Respect to Aircraft Rigid Body Axis (.i.e. Local Attitude = A/C Attitude + Δ Attitude)

(6) Collocated on Airframe

(7) Weapon Attachment Tolerances are 0.5 Deg in Pitch and Yaw, 1.0 deg in Roll

(8) Vibration Stabilized via Local Gimbal Platform

(9) A/C c.g. State Estimation (Reference Nav) Vector Defines the Aircraft Rigid Body Axis Reference Frame

CC24-0157-110-D

Figure 3.2-4: Local State Estimation Functional Allocations

Parameter [1]	Units	Range	Accuracy [4]	Jitter	Quantization/ Resolution	Refresh Rate	Data Latency (Max)	Notes
Latitude	deg	+/- 180	800 ft	2.14x10E-5	8x10E-8	25 Hz	not specified	6
Longitude	deg	+/- 180	(CEP)	2.14x10E-5	8x10E-8	25 Hz	not specified	6
Baro-Inertial Altitude	ft	-1k to 75k	150	12	2	50 Hz	not specified	6
True Heading	deg	+/- 180	0.05 - 0.1	0.0057 - 0.012	0.005 - 0.022	50 Hz	8 msec	6,7,8
Roll	deg	+/- 180	0.033 - 0.05	0.0057 - 0.012	0.005 - 0.022	50 Hz	8 msec	6,7,8
Pitch	deg	+/- 90	0.033 - 0.05	0.0057 - 0.012	0.005 - 0.022	50 Hz	8 msec	6,7,8
N/S Velocity (3)	ft/sec	+/- 3000	2.5 - 4.0	0.002 - 0.008	0.001	50 Hz	8 msec	8
E/W Velocity (3)	ft/sec	+/- 3000	2.5 - 4.0	0.002 - 0.008	0.001	50 Hz	8 msec	8
Vertical Velocity (3)	ft/sec	+/- 1500	2.0 - 2.5	0.002 - 0.008	0.001	50 Hz	8 msec	8
Roll Rate	deg/sec	+/- 300	0.054	0.036	0.0156	200 Hz	14.5 msec	8
Pitch Rate	deg/sec	+/- 200	0.054	0.036	0.0078	200 Hz	14.5 msec	8
Yaw Rate	deg/sec	+/- 200	0.054	0.036	0.0078	200 Hz	14.5 msec	8
Roll Angle Accel	deg/sec/sec	+/- 1000	2.65	2.6	0.17	200 Hz	23 msec	8
Pitch Angle Accel	deg/sec/sec	+/- 200	2.65	2.6	0.17	200 Hz	23 msec	8
Yaw Angle Accel	deg/sec/sec	+/- 200	2.65	2.6	0.17	200 Hz	23 msec	8
N/S Acceleration	ft/sec/sec	+/- 9 G	0.5	0.49	0.0078	50 Hz	23 msec	8
E/W Accleration	ft/sec/sec	+/- 9 G	0.5	0.49	0.0078	50 Hz	23 msec	8
Vertical Acceleration	ft/sec/sec	- 4 to +12 G	0.5	0.49	0.0156	50 Hz	23 msec	8
Time Tag (Radar)	μs	0 to 1.64 s	50		50	200 Hz	not specified	8
Radar (SAR) Alignment	mrad		0.5 (9)					8

Notes:

(1) Different vendors and motion compensation techniques require different inertial requirements, therefore some parameters show a range of values rather than values for a particular mechanization

(2) All values are RMS except as noted

(3) Precision Velocity Updates (PVU) improves horizontal velocity accuracy to 0.5 ft/sec, 1.0 ft/sec needed for High Resolution Mapping (HRM). GPS improves velocity to 0.3 ft/sec. Accurate velocity reference improves SAR resolution by reducing pointing errors under acceleration and the along track error.

(4) Accuracy = square root (mean accuracy^2 + jitter^2)

(5) Requirements translate into accelerometer accuracy: 100 μg bias, 200 ppm scale factor

(6) Required for cueing

(7) Required for designation

(8) Required for motion compensation

(9) Also known as radar harmonization

CC24-0157-115-D

Figure 3.2-5: SAR Inertial Sensing Allocations

Local State Estimation (LSE) Allocations

Local State Estimation provides the local motion/attitude/position inertial state vector at sensor and weapon locations which supports static boresight, weapon/sensor dynamic alignment, motion compensation, stabilization, and the transfer alignment of weapon Inertial Measurement Units (IMUs). Figure 3.2-4 shows these allocations.

SAR Motion Compensation

One of the basic radar motion compensation functions is the control and stabilization of the antenna beam pointing. This control function is required in all radar modes to provide sufficient antenna illumination gains for target detection and terrain imaging purposes. Figure 3.2-5 shows this requirement.

Electro-Optical Sensor Stabilization

The targeting FLIR provides a high resolution IR display with provides accurate pointing information. Using other onboard sensors, the targeting FLIR can be cued to a designated point, or it can be manually positioned. After a specific target is identified, the targeting FLIR can track the target by processing LOS commands to ground stabilize the electro-optics via a gimbaled platform and using the LTD/R laser designate if required. The Electro-Optical (EO) sensor suite has the capability to hand-off a single, tracked target to an IR MAVERICK, requiring only aircrew consent for launch. It can also be used to update the reference INS by supplying accurate position information. Figure 3.2-6 summarizes these allocations.

372

Parameter [1]	Units	Range	Accuracy [3]	Jitter	Quantization/ Resolution	Refresh Rate	Data Latency (Max)	Notes
Latitude	deg	+/- 90	235 ft	2.14x10E-5	0.000021	25 Hz	not specified	5
Longitude	deg	+/- 180	(CEP)	2.14x10E-5	0.000021	25 Hz	not specified	5
Baro-Inertial Altitude	ft	-1k to 75k	150	12	2	50 Hz	not specified	5
True Heading	deg	+/- 180	0.05 - 0.1	0.012	0.005 - 0.022	50 Hz	10.5 msec	5,6,7
Roll	deg	+/- 180	0.033 - 0.1	0.012	0.005 - 0.022	50 Hz	10.5 msec	5,6,7
Pitch	deg	+/- 90	0.033 - 0.1	0.012	0.005 - 0.022	50 Hz	10.5 msec	5,6,7
N/S Velocity	ft/sec	+/- 3000	2.5 - 4.0	0.002	0.001 - 0.25	50 Hz	20 msec	7
E/W Velocity	ft/sec	+/- 3000	2.5 - 4.0	0.002	0.001 - 0.25	50 Hz	20 msec	7
Vertical Velocity	ft/sec	+/- 1500	2.0 - 2.5	0.002	0.001 - 0.25	50 Hz	20 msec	7
N/S Acceleration	ft/sec/sec	+/- 9 G	0.1 - 0.5	0.49	0.0078 - 0.25	50 Hz	23 msec	7
E/W Acceleration	ft/sec/sec	+/- 9 G	0.1 - 0.5	0.49	0.0078 - 0.25	50 Hz	23 msec	7
Vertical Acceleration	ft/sec/sec	- 4 to +12 G	0.1 - 0.5	0.49	0.0156 - 0.25	50 Hz	23 msec	7
Roll Rate	deg/sec	+/- 300	0.054	0.036	0.0156	200 Hz	14.5 msec	7
Pitch Rate	deg/sec	+/- 200	0.054	0.036	0.0078	200 Hz	14.5 msec	7
Yaw Rate	deg/sec	+/- 200	0.054	0.036	0.0078	200 Hz	14.5 msec	7
Direction Cosines (4)	none	+/- 1	Note (4)	Note (4)	2.38x10E-7	200 Hz	not specified	5
Time Tag (1553B)	µs	0 to 1.64 s	50		50		not specified	7
Nav FLIR,IRST Alignment to INS		0.5 mrad						
Target FLIR,LTD/R Alignment to INS		0.25 mrad						

Notes:
(1) Different vendors and stabilization techniques require different inertial requirements, therefore some parameters show a range of values rather than values for a particular mechanization
(2) All values are RMS except as noted
(3) Accuracy = square root (mean accuracy ^{2} + jitter ^{2})
(4) Accuracy shall be consistent with source data (wander azimuth (alpha) and latitude)
(5) Required for cueing
(6) Required for designation
(7) Required for motion stabilization
(8) Requirements for gimballed platform gyro translates into 0.004 °/hr (North,East), 0.015°/hr (Down), 500 ppm Scale Factor

CC24-0157-120-D

Figure 3.2-6: EO Inertial Sensing Allocations

Weapon/Sensor	Δ Pitch Attitude (mrad) (3)	Δ Roll Attitude (mrad) (3)	Δ Yaw Attitude (mrad) (3)	Refresh Rate (Hz) (4)	Bandwidth (Hz) (4)
Nav FLIR, IRST (2)	0.5	0.5	0.5	100	>10
LTD/R, Targeting FLIR (2)	0.25	0.25	0.25	100	>10
Radar (SAR)	0.5	0.5	0.5	200	>10
HUD	0.7	0.7	0.7	25	0.01
HMD	1.0	1.0	1.0	Note (5)	-
MWR, LWR	1.5	1.5	1.5	50	>10
RWR (2)	1.5	1.5	1.5	50	>10
MK 84 GPB	None	None	None	-	-
GBU-24 LGB (6)	None	None	None	-	-
AGM-65 EO/IR Maverick (6)	2.0	2.0	2.0	25	< 1
AIM-9M Sidewinder	2.0	2.0	2.0	25	< 1
M61A1 20 mm Gun	1.0	6.0	1.0	25	< 1
DEW	0.25	0.25	0.25	100	>10
Basic Nav Sensors	0.25	0.25	0.25	50	< 1
Flight Control Sensors (6)	None	None	None	-	-

Notes: All values are RMS except as noted

(1) Compensation is required to meet alignment accuracies shown (static and/or dynamic flexure)

(2) Collocated on airframe

(3) Misalignment is represented as a Δ pitch, Δ roll and Δ yaw taken with respect to aircraft rigid body axis

(.i.e. local attitude = A/C attitude + Δ attitude)

(4) Data rates, data latency, bandwidth are driven by weapon/sensor location, compensation technique and overall system implementation. Simulation/analysis will update parameters.

(5) Static boresight is a constant by definition and only needs to be estimated once at power on

(6) Does not require static boresight compensation

CC24-0157-112-D

Figure 3.2-7: Weapon/Sensor Alignment Allocations

Weapon/Sensor Alignment (Pointing) Allocations

Mission sensor performance is dependent on accurate inertial information and the precise alignment (boresight) of the system sensitive axis. Manufacturing installation tolerances are cited at 7.5 milliradians, but advanced mission sensor alignment requirements can be as stringent as 0.25 milliradians. Manual static boresighting techniques are currently used to align mission sensors. These techniques utilize complex optical-mechanical procedures (expensive) and ground support personnel. This approach is only partially effective and does not address dynamic alignment as the aircraft structure flexes in flight, especially during severe maneuvers.

Weapon/store manufacturing attachment tolerances are typically 1/2 degree in pitch and yaw and 1 degree in roll. The overall misalignment can easily reach 10 milliradians (mils) or more. The values above are all static, that is, they do not include the effects of flexure (wing bending and twisting relative to a zero g condition). Wing flexure is most pronounced at the outboard stations and at higher g levels. Flexure can displace the weapon by over 10 inches, change the relative pitch angle by over 0.75 degrees, and change the relative roll angle by over 5 degrees. The angular displacement levels can be larger than the weapon/store attachment tolerances themselves.

Figure 3.2-7 shows the sensors and weapons requiring angular alignment or local attitude. The local attitude accuracy allocations are based on experience on F-15, F-18, AV-8B, A-12.

Weapon (1)	GPS (2)	Transfer	Weapon Launch Errors (3)												
			A/C Posn Handoff (ft CEP)	Weapon Position (ft CEP)	RSS Position (ft CEP)	A/C Vel Handoff (ft/s)	Weapon Velocity (ft/s)	RSS Velocity (ft/s)	A/C Azimuth Handoff (mr)	Weapon Azimuth (mr)	RSS Azimuth (mr)	Weapon Nav Tilt (mr)	A/C to Weapon Matrix Accuracy (mr)	Data Latency (msec)	DL Uncertainty (msec) (8)
AGM-65 MMW Maverick	No	Pre-Launch Alignment, Mil-Std 1760	500	-	500	2.5	0.1	2.5	0.87	3.4	3.5	1.0	-	200	10
JDAM (See Note 4)	Yes	Pre-Launch Alignment, Mil-Std 1760	500	-	500	2.5	0.1	2.5	0.87	1.0	1.3	1.0	-	200	<1
JSOW (See Note 4)	Yes	Pre-Launch Alignment, Mil-Std 1760	500	-	500	2.5	0.1	2.5	0.87	1.0	1.3	1.0	-	200	<1
SLAM (See Note 5)	Yes	A/C to Weapon Matrix Handoff	-	-	(6)	-	-	(6)	-	-	-	-	(6)		
AGM-86B ALCM	No	Pre-Launch Alignment	-	-	(7)	-	-	(7)	-	(7)	(7)	(7)	(7)	180	7
AIM-120 AMRAAM	No	1.5 sec to Launch, A/C to Missile Matrix Handoff	800	-	800	2.5	-	2.5	-	-	-	-	(9)		

CC24-0157-111-D

Notes: All values RMS except as noted
(1) Weapon Transfer Alignment is with Respect to the Inertial Navigation Frame (Local Level)
(2) Accuracies Shown Assumes GPS Not Available (e.g. jammed), otherwise; Weapon Performance Approaches GPS Accuracy (0.3 ft/s, 30 ft).
(3) Weapon Accuracies Shown are Relative to the Inertial Reference Source Data,
Total Weapon Alignment Error = Square Root (A/C Handoff Error**2 + Transfer Alignment (Relative) Accuracy**2)
(4) Advanced Weapons, Requirements Shown are Projected, Fast GPS Acquisition of Satellites Requires Precise Timing Accuacies
(5) Transfer Alignment Only Requires Handoff Data to Initialize Weapon - No Prelaunch Alignment
(6) Numbers Classified - No Impact on IIN
(7) Numbers Classified - Impacts IIN Design
(8) Transformation Matrix Relating the Aircraft Inertial Nav Reference Frame to the Missile Body Frame
(9) Consists of Weapon Attachment Tolerances and A/C c.g. State Attitude/Heading Accuracies

Figure 3.2-8 Weapon IMU Transfer Alignment Allocations

Weapon IMU Transfer Alignment Allocations

Weapons with Inertial Measurement Units (IMUs) for midcourse and/or terminal guidance usually require prealignment before launch. Transfer alignment procedures have been developed which prealign the weapon IMU by dynamically matching time histories of the aircraft reference INS data with the weapon slave IMU data. Such procedures

The challenge of transfer alignment is to utilize the physical error models of the INS and IMU to separate and identify the various alignment and instrument error terms so that they can be compensated prior to weapon launch. Certain error parameters (notably the alignment errors) enter the error model as terms multiplied by the aircraft acceleration or rotation rate. This dictates the need to provide an aircraft maneuver with sufficient acceleration amplitude and time varia-

Table 3.2-6 Error Contributors to Weapon Position Error

Initialization
- Final Checkpoint   – Position Error at Launch Due to Designation Errors
- Master System (Position Drift)   – Position Error at Launch Due to Aircraft Navigator Errors

Transfer Alignment
- Measurement Noise   – Position Uncertainty Between Aircraft and Missile IMU Due to Flexure
- Master System (Attitude and Velocity Errors)   – Attitude and Velocity Errors Transferred From Aircraft to Missile IMU

Missile System   – Quality of Missile Inertial Measurement Unit (IMU)

External
- Gravity Anomaly
- Mapping   – Vertical Deflection
      – Target to Identification Point Uncertainty

CC24-0157-156-D

are mature with extensive research and successful applications. Exceptions to aligning the IMU prior to launch include weapon systems with Global Positioning System (GPS) which allow an inflight alignment capability. However, weapons with GPS still require coarse initialization to ensure quick acquisition of satellites. Table 3.2-6 shows other error contributors to weapon IMU position errors.

tion so that terms can be separated. Higher acceleration and/or rotation rates can lead to better estimation accuracy and faster estimate times but the higher accelerations and rates also increases aircraft flexure. Aircraft flexure can limit transfer alignment accuracies and convergence times since the uncertainty of the "true" weapon position increases. Weapon IMU transfer alignment performance allocations are shown in Figure 3.2-8.

Supports Aircraft c.g. State Estimation "Aided" Performance Requirements
Supports Inflight Alignment Function
Supports Navigation Fault Detection, Isolation & Recovery • Provide Measurement Residuals • Provide Kalman Filter Covariance Values
Update Types • GPS Velocity • Radar Velocity • Generalized Velocity • GPS Position • Radar Position • Generalized Position
Other Capabilities • Reasonableness Checks (i.e. Chi-Squared) • Adaptable/Flexible to Allow Cost-Effective Integration of New Sensors or Upgrades

CC24-0157-127-D

Figure 3.2-9 Navigation Updating Functional Allocations

Updating Modes	Update Accuracy	Advantages	Disadvantages
Radar SAR Map [1]	105 ft CEP [6] 0.5 ft/sec [7]	Good Range, Relatively Unaffected by Weather, Velocity Update Available, Autonomous	Emissions Required
FLIR Update [2]	33 ft CEP [6]	Good Accuracy, No RF Emissions, Autonomous	Limited Range, Affected by Weather
Global Positioning System (GPS) [3]	30 ft CEP 0.3 ft/sec	Best Accuracy, No Emissions, Jam-Resistant, All Weather, World-Wide	Not Jam-Proof, Satellites and Ground Stations Subject to Attack
Terrain Referenced Navigation	150 to 300 ft	Low Power Emissions, Jam-Resistant, All Weather, Autonomous	Limited Map Availability, Not Operational Over Water, Cities, or Very Smooth Terrain
TACAN [4]	1.3K ft CEP @ 10 NM 11K ft CEP @ 100 NM	Widely Deployed, Available for Rendezvous	Emissions Required, Poor Accuracy at Range, Not Available in Enemy Territory
HUD Visual Update [5]	97 ft CEP [6]	No Emissions, Limited Avionics Required, Autonomous	Affected by Weather, Limited Range

Notes: All values RMS except as noted
(1) Based on 15 NM ground range, 4,800 ft altitude, 500 kts, 45 deg squint, 15 sec navigation time, 8.5 ft resolution
(2) Based on 12K slant range, 1,000 ft altitude, 500 kts, laser ranging
(3) Based on good satellite geometry
(4) Based on accuracy of 1.5 deg one sigma in bearing, 600 ft one sigma range
(5) Based on 6K ft slant range, 100 ft altitude, level
(6) Waypoint mapping errors not included
(7) Accuracy of Precision Velocity Updates (PVU)

CC24-0157-128-D

Figure 3.2-10 Navigation Updating Accuracies

Navigation Updating Allocations Navigation aiding sensors include Global Positioning System (GPS), pressure altimeter, SAR, Targeting FLIR, Tactical Air Navigation (TACAN), Head-Up Display (HUD), and a terrain referenced navigation with a digital map such as SITAN (Sandia Inertial Terrain-Aided Navigation) or TERPROM (Terrain Profile Matching).

Navigation "fusion" updating consists of integrating information from multiple navigation sensors in a statistically optimum way. This is accomplished by a Kalman filter. Although it is a filter in the digital processing sense, it may more fundamentally be considered a statistical algorithm which provides an optimum estimate of the values of certain parameters associated with a dynamic process. In a navigation system, these parameters (or "states" of the system) consist of various navigation errors, so that the filter provides an optimum, ongoing estimate of such errors, which are then used to correct the navigation system outputs. Historically, the software and processor for accomplishing this has resided in the INS and has been provided by the supplier as part of the system. As navigation becomes more dependent on information from multiple sensors, and as other systems become more integrated with the INS, it is not obvious that this location will continue to be appropriate for the Kalman filter.

The navigation system updating functional allocations are shown in Figure 3.2-9. The ability to obtain periodic navigation system updates during the course of the mission is extremely important to overall mission effectiveness. However, mission scenario and crew workload considerations may place constraints on the frequency and quality of the navigation update. Several update modes have been quantified in terms of accuracies, advantages and disadvantages as shown in Figure 3.2-10.

4.0 Conclusions

This paper has reviewed the current state-of-the-art of navigation system technology for fixed wing aircraft, suggested a methodology to decompose the mission of the particular aircraft in order to determine requirements and a methodology to apply systems engineering to determine the detailed requirements of the system to the particular set of mission segments, sensors and functions required of the system. Some examples of the analysis required for several functions such as sensor queing, targeting and weapon delivery have been shown. The results of the example analysis have been shown in tables throughout section 3.1 and may be of some use to the reader for application to his or her particular need. It would be difficult to establish one set of requirements for all applications, therefore the goal is to provide the reader a basis for a methodology to follow in his or her analysis. Section 3.2 provides a typical set of requirements allocations for fixed wing aircraft as a baseline system that can be used as a reference by the reader. The references called out in the text and outlined in the reference section should serve as a more detailed basis for the reader to use in his application.

<u>Definitions</u>

Several error measurement terms are used in this paper for specifying accuracy requirements and/or bombing errors. Definitions of each are given below. The error distribution is assumed to be linear and normal (References 1, 4, 9, 10 and 17).

Standard Deviation (Sigma): The square root of the mean of the squares of the deviations (68.27% of all errors occur within the standard deviation limits of + or - 1 sigma for a normal distribution).

Mean Deviation or Probable Error (PE): The sum of the absolute deviations divided by their number. There is equal probability that the error will be either larger or smaller than this value(50% of all errors will not exceed this value). Probable error is commonly used to express weapon impact error.

Range Error Probable: The probable error measured along the range axis (along the weapons flight path).

Deflection Error Probable: The probable error measured perpendicular to the weapons flight path (cross axis).

Circular Error Probable: A two dimensional measurement of error defined as the radius of the circle for which 50% of the errors fall within its boundary. In practice, the circle is centered at the true position of the target, which is equal to the mean in the absence of systematic errors.

References

1. Kayton, Myron and Fried, Walton, "Avionics Navigation Systems", John Wiley and Sons Inc., New York, NY. 1969

2. Thompson, Dan and Schrank Michael, "Advanced Vehicle Management System Architectures Study", Digital Avionics Conference, October 1991, IEEE Proceedings.

3. Kelley, Ronald, "Integrated Inertial Networks", Joint Services Data Exchange Conference, October 1992, JSDE Proceedings.

4. US Naval Oceanographic Office, "Navigation Dictionary", Pub. No 220, Washington, DC., 1956.

5. Gault, Kenneth, et al, "Advanced Vehicle Management System Architecture Studies Final Report", TR-WL-92-3106, September 1992.

6. Defense Systems Management College, "Systems Engineering Management Guide", January 1990.

7. Savage, P. G., "Laser Gyros In Strapdown Inertial Navigation Systems", IEEE Position Location and Navigation Symposium, San Diego, California, Nov. 1-3, 1976.

8. Bevington, J. E. and Marttila, C. A., "Precision Aided Inertial Using SAR and Digital Map Data", IEEE Position Location and Navigation Symposium, IEEE Proceedings, 1990.

9. Greenwalt, C. R., "Principles of Error Theory and Cartographic Applications", Aeronautical Chart and Information Center Technical Report No. 96, Feb. 1962.

10. Kelley, Ronald et al, "Integrated Inertial Network", WL-TR-94-1012, Volume 1, Final Report, January 1994.

11. Martin, E. H., "GPS User Equipment Error Models," Journal of the Institute of Navigation, Vol. 25, No. 2, 1978.

12. SS-GPS-300, "System Specification for the NAVSTAR Global Positioning System"

13. SNU 84-1, "Specification for USAF Standard Form, Fit and Function Medium Accuracy Inertial Navigation Unit", Aeronautical Systems Division, Air Force Systems Command, Wright-Patterson AFB, OH, 28 February 1986.

14. "Procurement Specification For Inertial Navigation Set", PS 68-870204, McDonnell Aircraft Company, Revision B, 30 November, 1987.

15. Burns, R. C., et al, "Strapdown Performance Study," AFWAL-TR-82-1074, Volume 1, July 1982.

16. Interface Control Document (ICD-GPS-200), "NAVSTAR GPS Space Segment/Navigation User Interfaces, December 1986.

17. Beck, G. R., et al, "GPS Exploitation for Precision Targeting Study Final Report'", WL-TR-94-1068.

18. "Specification for Global Positioning System Integrated Navigation Assembly (GINA)", Naval Command, Control and Ocean Surveillance Center, Warminster, PA.

19. SME-EGI-92, "Systems Requirements Document for an Embedded Global Positioning System Receiver in an Inertial Navigation System (EGI)", Rev-2, 20 January, 1994.

REPRESENTATIVE VEHICLE IMPLEMENTATION - ROTARY WING AIRCRAFT

JOHN NIEMELA

U.S. ARMY COMMUNICATION-ELECTRONICS COMMAND

RESEARCH & DEVELOPMENT CENTER

FORT MONMOUTH, NJ

USA

DAVID F. LIANG

NATIONAL DEFENCE HEADQUARTERS

OTTAWA, ONTARIO

CANADA

1. INTRODUCTION

Rotary wing aircraft have seen an ever expanding role in support of land and sea military operations including: search and rescue, scout, attack, troop transport, anti-submarine warfare, anti-surface ship targeting, cargo and electronic warfare. They are ideally suited for operation in confined and unprepared areas where no other form of aerial transport is suitable.

For a helicopter to survive on the modern battlefield, flight profiles are mandated that impose unique requirements on the navigation system. These include contour and nap-of-the-earth flight as well as hover in defilade in close proximity to obstructions. Critical mission segments are conducted in nap-of-the-earth flight during which the pilot conceals the helicopter with terrain, foliage and buildings. These missions must be accomplished in all weather and visibility conditions. In the course of these mission, the rotary wing aircraft crew must maintain accurate self-location to maintain geographic orientation and situational awareness relative to friendly and hostile forces. A typical tactical mission profile overlaid on a contour map is shown in Figure 1, illustrating the non-linear flight path which takes advantage of terrain masking [1].

For the anti-submarine warfare missions the helicopter navigation system must maintain stable and accurate tactical plots over long periods of time. In the anti-surface ship targeting role, a high degree of absolute and relative navigational accuracy are vital to rapid and successful action. There are further complicating factors as well. Operations must often take place under radio silence and shore-based or satellite navigation aids may be destroyed or jammed during wartime. The small crew of the helicopter must not be burdened with monitoring the functioning of, or updating, the navigation system.

2. TACTICAL HELICOPTER NAVIGATION

In general, two accuracy related operational aspects drive the tactical helicopter configuration. The first of these is for the helicopter aircrew to maintain geographic orientation during low altitude flight. Geographic orientation assures point-to-point navigation, rendezvous with friendly forces, and avoidance of known enemy threats. To maintain geographic orientation the helicopter aircrew must be able to identify natural or cultural features and correlate them with a hand-held or electronic map. Failure to do so may jeopardize the mission, in that the helicopter may not arrive at its destination in a timely manner, inadvertently overfly a hostile threat or impact unexpected terrain or obstacles. A hundred meter navigation error, when displayed on a paper or electronic map at an appropriate scale (nominally 1:50,000), could direct the helicopter aircrew up the wrong stream-bed or draw when operating at nap-of-the-earth altitudes.

The metric, height above terrain divided by forward velocity, has been applied to aircrew geographic orientation and is proportional to the time available for the aircrew to fixate, identify and correlate a ground feature. Considering the limited number of features that the aircrew sees at any one instance when operating at nap-of-the-earth height (5-10 m AGL) and their nominal velocities (20-50 kn) the metric is comparable to a low level tactical fighter at near supersonic velocity.

The second accuracy driver for the tactical helicopter is for target location. Many scout and attack helicopters have weapon direction systems that locate targets relative to the aircraft. It is the function of the navigation system to provide data to register these locations to target coordinates on a geographic grid. The navigation position error of the helicopter adds to the target location error. The system error must meet ordinance related miss distances after being handed over to weapon system elements. Twenty-five meter position accuracy meets a large class of such requirements when considered in the context of system error budgets.

The nominal position accuracy requirements that arise from consideration of the above discussed requirements for tactical helicopter are 25 meters when external aids are available or 0.5% to 1.0% of the distance traveled for the self-contained system. The accuracies are typically specified at the ninety-five percent (95%) Probable Error Level. Note that if the external navigation reference is not available, the aircrew must

manually update the self-contained navigation system fairly frequently (every 5-10 km) to maintain geographic orientation as well as maintain position accuracy sufficient for weapon direction.

3. MARITIME HELICOPTER NAVIGATION

For maritime military helicopter applications, flight duration tends to be rather extensive, therefore the performance accuracy of the navigation system depends on the flight profile specified and the type of radio equipment available. In general, maritime helicopter operations can be represented by the following two flight profiles:

3.1 CONVOY SCREENING MISSION PROFILE

The helicopter searches a moving sector ahead of a convoy of ships. The profile is normally characterized by alternative short cruise phases and hover phases. The hover phases represent periods in which the helicopter is using a dipping sonar to detect threats in the vicinity of the convoy. Figure 2 shows the plan view for the convoy screening mission.

3.2 SONOBUOY MISSION PROFILE

This mission includes a constant heading cruise from the takeoff point to the vicinity of a target. The mission profile undergoes a sonobuoy deployment phase, followed by a Magnetic Anomaly Detection (MAD) phase and a weapon delivery maneuver. Return to the mother ship consists of another extended cruise phase. Total mission duration can exceed several hours. Figure 3 shows the plan view and figure 4 shows the altitude profile for the sonobuoy mission.

With these two mission profiles in mind, the operational accuracy requirements can be specified as follows:

Radial Position Error (95%):

- with external aids* 2.0 nautical miles (nm)

- without external aids 1.5 nm/hr

Radial Velocity Error (95%):

- with external aids** 3.0 ft/sec

- without external aids 4.0 ft/sec

Attitude Error (95%):   0.5 deg

Heading Error (95%):   0.5 deg

* External aids are those systems such as Omega, Loran and the Global Positioning System (GPS) which rely upon transmitters which are located external to the

aircraft and may be unavailable during wartime.

** It is recognized that these performance levels can be improved on by a large margin if, as expected, a GPS is part of the navigation system configuration.

3.3 ALIGNMENT CONSIDERATIONS

For maritime military helicopter applications, one of the key navigation requirements is the appropriate alignment of the INS, which requires initialization of position, velocity and attitude data. When GPS is available as an alignment reference, both the fixed base and moving base alignment can be readily accomplished, since GPS will provide an excellent position and velocity reference.

When GPS is not available, during the initialization process, the initial position can be inserted manually or automatically. For a fixed base initialization, the initial velocity can be assumed to be zero; but the same cannot be assumed for a moving base initialization. For a moving base, a velocity reference is required or else it must be computed from a sequence of reference positions. Attitude alignment can be inferred from the accelerometer measured specific force. However, the most difficult task is the initial heading alignment, which can be accomplished by several alternative methods.

Heading alignment can normally be divided between the so-called coarse and fine alignment. For a fixed base, the coarse alignment can be accomplished with wide angle gyrocompassing, which attempts to calculate body roll, pitch and heading from gyroscope and accelerometer measurements in body coordinates. For an in-motion coarse heading alignment, the true heading can be derived from a magnetic compass or magnetometer.

In gyrocompassing alignment, reference velocity or position measurement noise error is a fundamental source of error that determines the time required for alignment. Therefore for a rapid alignment, the reference measurement noise must be very small, which is the condition normally achieved with a fixed base alignment.

For in-motion alignment, the most rapid alignment is achieved with transfer of alignment from the ship's inertial navigation system. For this, very small short correlation measurement error must be achieved, which is difficult to accomplish for a smaller size ship. The next rapid alignment method requires the ship to maintain steady course and speed and uses inserted course and speed as the velocity vector reference. However, because of the surge and sway of the ship, the inserted constant velocity has significant short correlation error. The accuracy of this method is limited by the gyro drifts and by the bias error of the inserted velocity.

383

When doppler aiding is available, the most prominent source of error is the doppler velocity long term correlation error due to waves and current effects. However, when shipboard coarse alignment is used in conjunction with inflight doppler aiding, it is not necessary to have long duration shipboard alignment to achieve excellent integrated navigation performance.

4. PILOT VEHICLE INTERFACE

The means by which the navigation data, generated by the sensors and processed by computer configuration to the required accuracy, is displayed to the aircrew is subject to many constraints. Typically navigation displays provide the following capabilities:

(1) Position in geographic reference (latitude,longitude)/UTM/tactical grid coordinates.

(2) Steering, bearing, track display, distance and time to go.

(3) Look ahead to destination, and drift angle.

(4) Groundspeed and track.

(5) Windspeed and direction.

(6) Position updating including "on-top", manual and offset.

(7) Manual slew of map image

(8) Display freeze.

(9) Map marking.

(10) Ability to permit the aircrew to view a continuous tactical map display of various scales (typically 1:50,000) over which the mission is being flown.

(11) Portray three dimensional topographic perspective.

(12) 100 waypoint entry and storage.

(13) Allow for map orientation as desired by the aircrew.

(14) Alternative displays (e.g., heads-up, etc.) providing required tactical information shall also be acceptable.

(15) Displayed navigation information should be compatible with the aircrew's night vision and target acquisition systems.

5. OTHER MILITARY REQUIREMENTS

Operation off the deck or in the proximity of NATO surface vessels requires that the navigation system must be compatible with the electromagnetic interference experienced in that environment.

For rapid response missions, the tactical helicopter, including its navigation system, must be operational within two minutes of power application.

The system must be capable of providing a compatible navigation input to an instrument letdown in Instrument Meteorological Conditions (IMC) acceptable to civil authorities for operation at civil airports.

In addition to the above, as with all military avionics systems it is desired that the helicopter's navigation capability be realized with minimum impact on aircraft system cost and gross weight as well as be modular in construction to facilitate system growth and maintenance.

6. ARCHITECTURE

A typical rotary wing requirement definition considers the several factors discussed in the introduction and includes extensive analysis and simulation to determine the class of sensors and means of integration. With a fair degree of consistency, a set of navigation sensor types and integration concepts are widely used to meet helicopter mission requirements. Aircrews require a position location and navigation system that will permit accurate navigation worldwide when operating at terrain flight altitudes under all visibility and meteorological conditions. The system must be accurate, low-cost, lightweight, and provide anti-jam protection as well as positional information in Latitude/Longitude and in the Military Grid Reference System (MGRS) or the Universal Transverse Mercator (UTM) coordinates. The system shall have the capability to store multiple waypoints/destinations and display azimuth and distance to these locations.

In general, the desire to have at least one mode that is not dependent on external radio navigation aids has led to the requirement for a self-contained navigation mode - a mode in which the navigation system is not dependent on external radio navigation aids. The need for a self-contained mode is driven by the operational and technical concerns:

° Externally referenced systems may be susceptible to enemy electronic countermeasures and physical destruction. External sensor input may be susceptible to jamming or spoofing and in times of crisis or war, may not be fully usable or continuously available.

° Externally referenced systems may be affected by natural phenomena and man-made obstructions. Ground based navigation aids that require line-of-sight reception do not provide the required capability at terrain flight altitude.

Hence, helicopter navigation requirements are typically met by the integration of self contained and externally referenced navigation sensors and modes. The combination of these two navigation sensors create a hybrid navigation system (e.g. GPS/inertial, GPS/doppler/AHRS, GPS/doppler/inertial/terrain referenced navigation (TRN). The sensors which comprise such hybrid navigation systems have complementary error spectrums and/or data availability characteristics which yield a more accurate and reliable navigation system.

7. NAVIGATION MODES AND SENSORS

7.1 HYBRID NAVIGATION MODES

Typical hybrid navigation systems aboard military helicopters are comprised of a set of self-contained (doppler and inertial) and externally referenced navigation sensors (GPS and TRN). These sensors are configured via standard data busses and Kalman filters to yield system performance and integrity appropriate to the particular helicopter mission.

The benefits of hybrid navigation systems are numerous. The sensor data integration software can perform the following tasks.

a. Integrate the available navigation input data from the sensor subsystems. Appropriate blendings of sensor data can lead to rapid alignment sensor calibration. Optimum navigation accuracy and enhanced reliability through redundancy and navigation.

b. Continuously calculate the appropriate gains to weight the information provided from the different sensor subsystems to generate real-time gain calculation.

c. Calculate the measurement of uncertainty (covariance) in the estimate of the system navigation states and the individual subsystem states.

d. Perform system failure detection and isolation as well as "outlier" editing to detect and isolate both hard and soft failures in the sensor and/or computer subsystems.

e. Perform automatic system reconfiguration to remove the failed subsystem while integrating the remaining navigation sensors.

When the on-board navigation systems include both the GPS and INS, the integrated navigation solution can provide excellent position, velocity and attitude accuracy. As soon as GPS has achieved its position fix, the integrated position and velocity solution will be excellent, and the well calibrated inertial system accuracy will remain sufficiently accurate, should the GPS reception be interrupted or lost.

In the complete absence of GPS, the INS performance can be aided using Doppler velocity subsystem. However, the INS/Doppler performance over the sea depends on initial alignment; sea bias effects calibration; as well as on the surface wind effects compensation.

For maritime military helicopter applications, the hybrid navigation systems can provide the following additional alignment capabilities:

(1) Fixed-base ground alignment mode,

(2) Shipboard Transfer alignment mode,

(3) Shipboard GPS-aided alignment mode,

(4) Airborne GPS-aided alignment mode, and

(5) Airborne Doppler-aided alignment mode.

7.2 SELF-CONTAINED MODES

The self-contained mode is typically obtained with a doppler navigation radar coupled with a accurate heading reference. In instances where a more accurate heading reference is needed, such as for weapon direction, a doppler/inertial hybrid configuration is employed. The doppler navigation system operated with an accurate heading reference, is particularly suitable for low altitude, slow flying air vehicles as the tactical helicopter, as compared to a stand-alone inertial navigation system. This is because a doppler navigation error propagates in proportion to distance traveled whereas inertial propagates error in proportion to time. Considering the respective state-of-the-arts of doppler and inertial technology, and the relatively low speed that a tactical helicopter achieves over the battlefield, leads to the conclusion that doppler navigation is a more accurate means of providing self-contained navigation for this category of aircraft. For example, a doppler properly integrated with a good quality heading reference can yield one-half percent of distance traveled navigation accuracy independent of the helicopter's velocity. To achieve an equivalent level of navigation accuracy with a stand-alone inertial would require one-quarter of a nautical mile per hour performance - a more costly alternative. This trade-off is portrayed in Figure 5 which shows that for the tactical helicopter critical mission segment of 20 - 50 kn at nap-of-the-earth flight, one percent of distance traveled performance of a doppler/heading reference is equivalent to a 0.3 NM/HR inertial system. As respective advancements have been made in doppler, heading reference and inertial system technology the performance of all have improved. However, the general observation that a doppler with a good heading reference can outperform a stand-alone inertial system in the helicopter environment is as valid

today as it was when noted by T. J. Penfold in 1975 [2].

Additional characteristics for helicopter doppler navigation technology that have made it an attractive means of navigation is that the velocity data is very useful for air to ground fire control and for hover stabilization. This feature is of particular importance for the anti-submarine warfare mission where the helicopter must hover over a field of sonabouys. Also its rapid reaction time, typically less than two minutes, is a useful characteristic for tactical operations.

Extensive experience with doppler navigation systems have shown them to be a reliable and cost effective means of providing a self-contained navigation capability for tactical aircraft, particularly helicopters.
Vulnerability to ECM is minimal due to the frequency of operation (13.3 Ghz), low radiated power, and very low side lobe radiation resulting in a low RF signature relative to other signatures such as infrared and acoustic.

A typical configuration is shown in Figure 6 which portrays the principle sensing, processing, and control & display functional elements as well as primary data interfaces. The functions of each of these elements are as follows:

7.3 ATTITUDE & HEADING REFERENCE.

This element senses the roll, pitch and heading of the helicopter which when combined with aircraft body axis velocity from the doppler provides the self-contained navigation solution. A secondary function is to provide data to aircraft flight instruments, flight and fire control systems. Ring laser gyros have been found to be superior to spinning mass gyros, particularly for strapdown AHRS application due to greatly reduced scale factor and gyro axis alignment stability errors [3].

For utility and cargo helicopters this function is typically realized with a magnetically slaved heading reference or pendulously erected AHRS. Nominal accuracies are 1° RMS - heading; 0.5° RMS - pitch & roll. For scout and attack helicopters in which heading accuracy can be directly related to targeting accuracy, a strapdown AHRS, which employs gyrocompassing techniques for alignment, is typically employed. Nominal accuracies are 0.3° - 0.5° RMS heading; 0.15° - 0.25° RMS pitch and roll.

7.4 DOPPLER VELOCITY SENSOR.

This element senses the helicopter translation velocity over land or water for use by the navigation processor in generating the navigation solution. A secondary function is to provide ground reference velocity data for air to ground fire control, and hover stabilization. Nominal accuracies are 0.25% of velocity plus a bias of 0.1 kn.

7.5 GPS RECEIVER.

This element senses the helicopters position, velocity and time in earth centered - earth fixed coordinated and provides a statistically stationary navigation accuracy (not time or distance dependent) to bound the error of the self-contained portion of the navigation system. Trade-offs and consideration on GPS integration for tactical helicopter is given in Ref 2.

7.6 RADAR ALTIMETER.

This element senses height above terrain for display to the aircrew and for use by on-board TRN equipment. Nominal accuracies of 0.5 - 1.0 m RMS bias plus a 2 - 3% altitude dependent scale factor error.

7.7 NAVIGATION PROCESSOR.

This element combines the nav sensor data through use of a recursive statistical filter. In addition, it computes bearing and distance to waypoints, waypoint storage, magnetic variation and navigation moding. For a hybrid navigation system, the functions of the navigation processor expand considerably to include sensor detection isolation and system reconfiguration.

8. CONCLUSION:

Typical navigation suites aboard tactical helicopters are significantly different than those of fixed wing fighter or transport aircraft. This is due primarily to the difference in flight profile between these categories of aircraft. The lower and slower flight profile of the helicopter favors a low cost inertial - doppler ensemble for the self-contained navigation function whereas a medium to high accuracy inertial system is in general favored for fixed wing aircraft. In addition, the translational and rotational degrees of freedom of the helicopter are typically an octave more benign then that of high performance fighter aircraft. This allows adjustments to be made in integration of the Kalman filter to achieve improved performance in the helicopter environment.

REFERENCES

[1] Coffee, J. N. "GPS - INS Integration Requirements for Robust NOE Helicopter Navigation: american Helicopter Society 47th Annual Forum, 6 - 8 May 1991.

[2] Penfold, T. J. "Medium Accuracy Low cost Navigation System for Helicopters", AGARD Conference Proceeding No 176 on Medium Accuracy Low Cost Navigation, September 1975.

[3] Niemela, R. J., et al "Helicopter Flight Test of a Ring Laser Gyro Attitude and Heading Reference System". Proceeding of the Institute of Navigation National Technical Meeting, San Diego, CA 17 -19 Jan 1984.
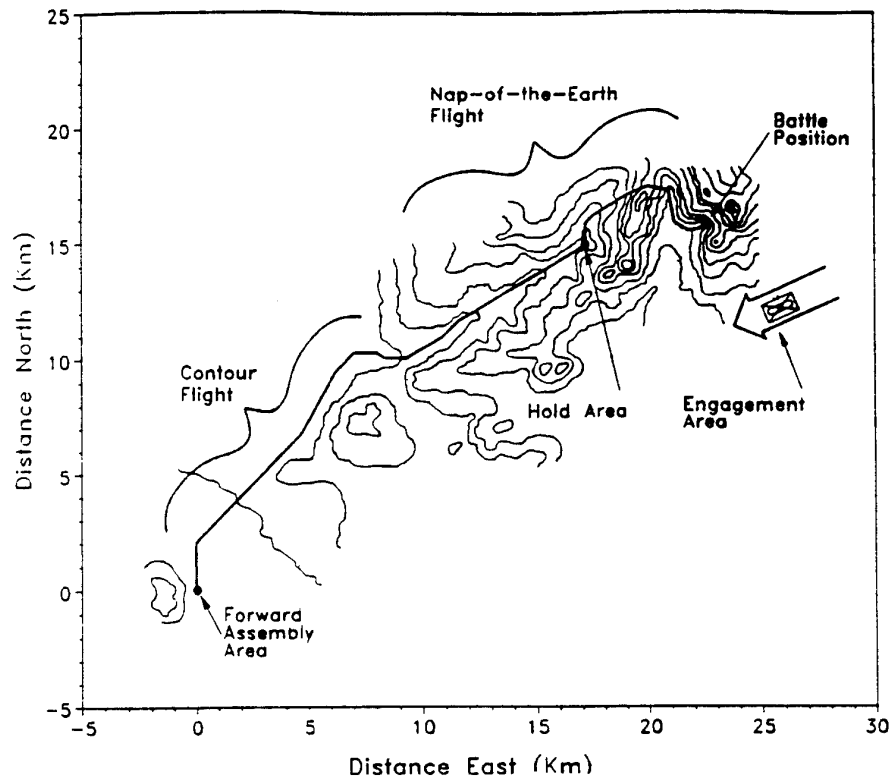
FIGURE 1: TYPICAL TACTICAL HELICOPTER MISSION PROFILE
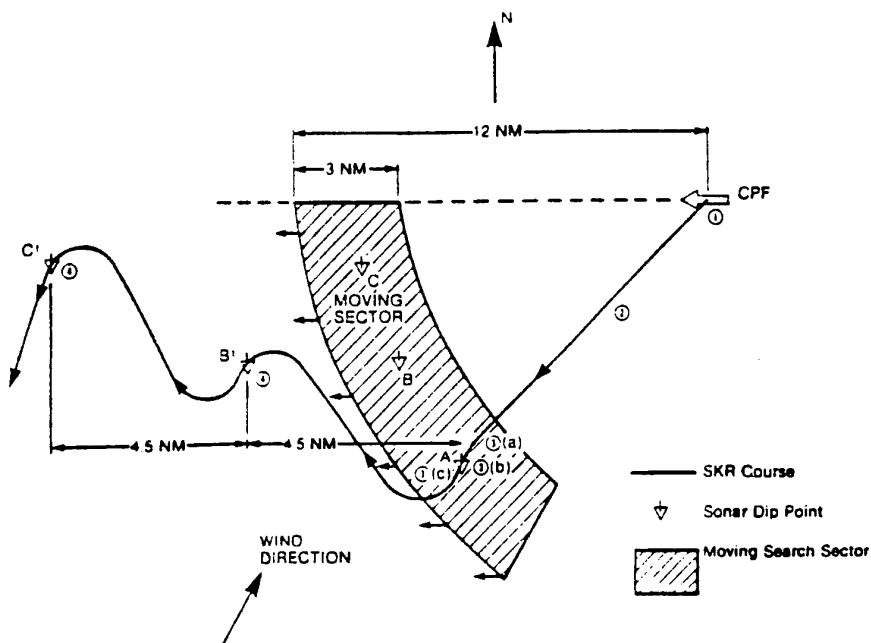
FIGURE 2: PLAN VIEW OF CONVOY SCREENING MISSION

Sea King Replacement (SKR) Course

Canadian Patrol Frigate (CPF) Course

Sub-surface Contact Course

Sonobouy

Torpedo Attack Zone

NM   Nautical Mile

FIGURE 3:   PLAN VIEW OF SONOBUOY MISSION

SKR Altitude Profile

Canadian Patrol Frigate

FIGURE 4:   ALTITUDE PROFILE FOR SONOBUOY MISSION

FIGURE 5:   COMPARISON OF DOPPLER WITH INERTIAL NAVIGATION PERFORMANCE

FIGURE 6:   TYPICAL TACTICAL HELICOPTER NAVIGATION SUITE

Spacecraft Navigation Requirements

Judy L. Miller
Charles Stark Draper Laboratory, Inc.
555 Technology Square
Cambridge, MA 02139
USA

1. SUMMARY
Spacecraft operation depends upon knowledge of vehicular position and, consequently, navigational support has been required for all such systems. Technical requirements for different missio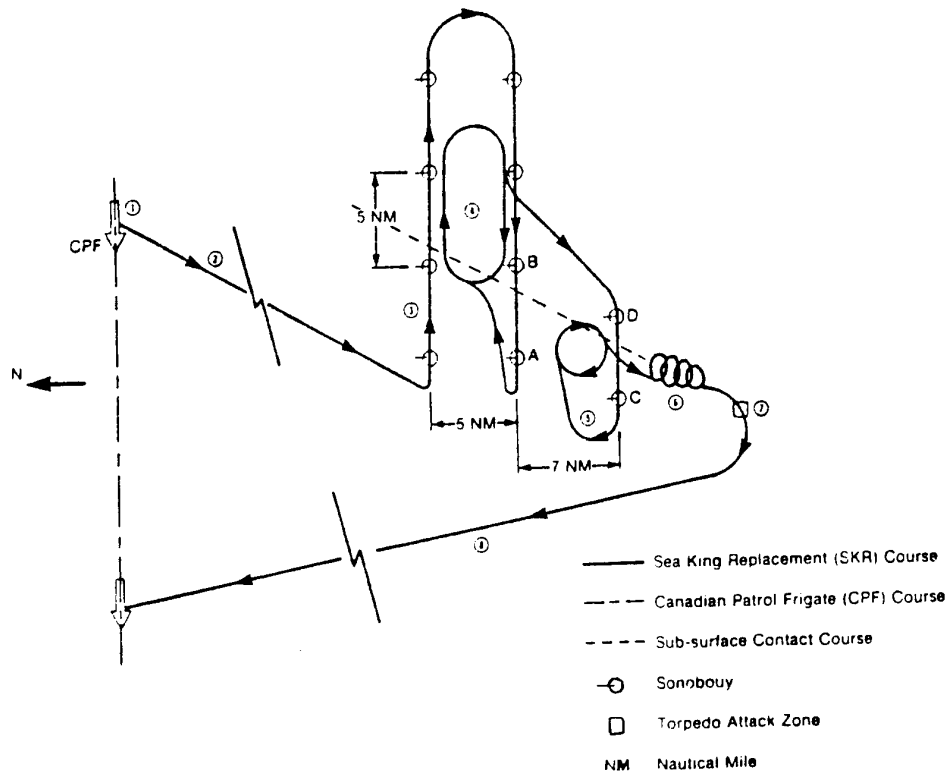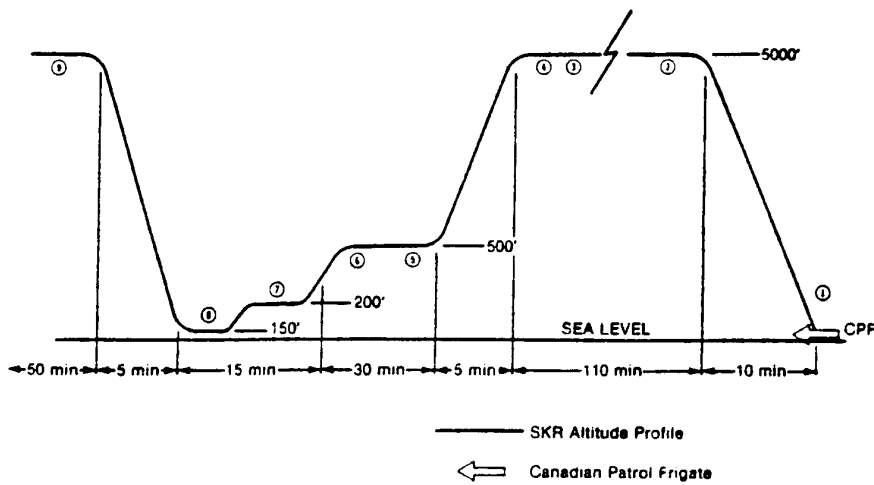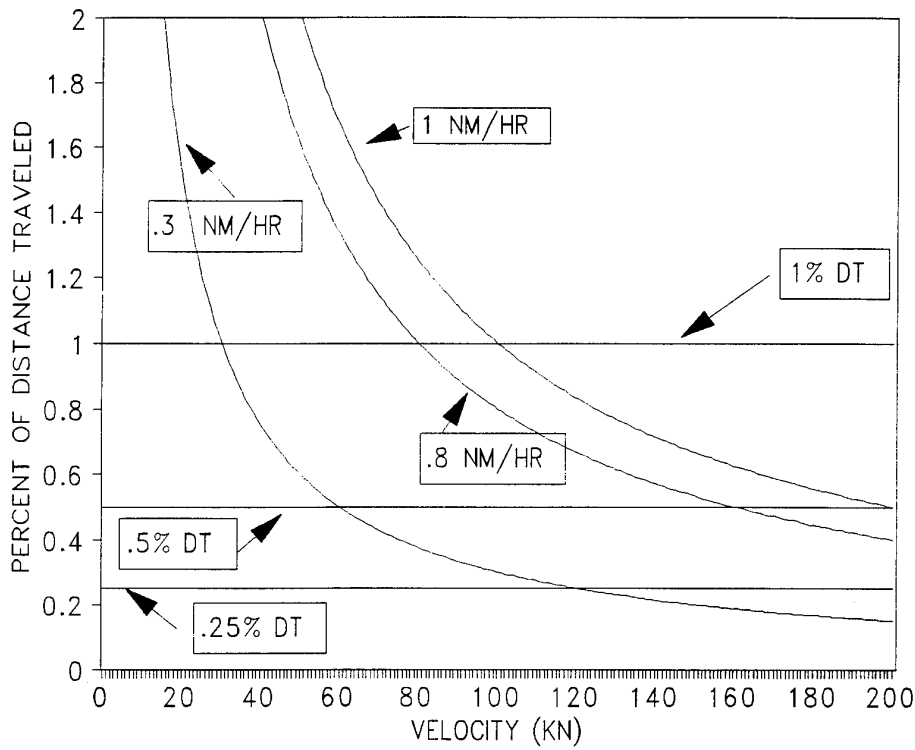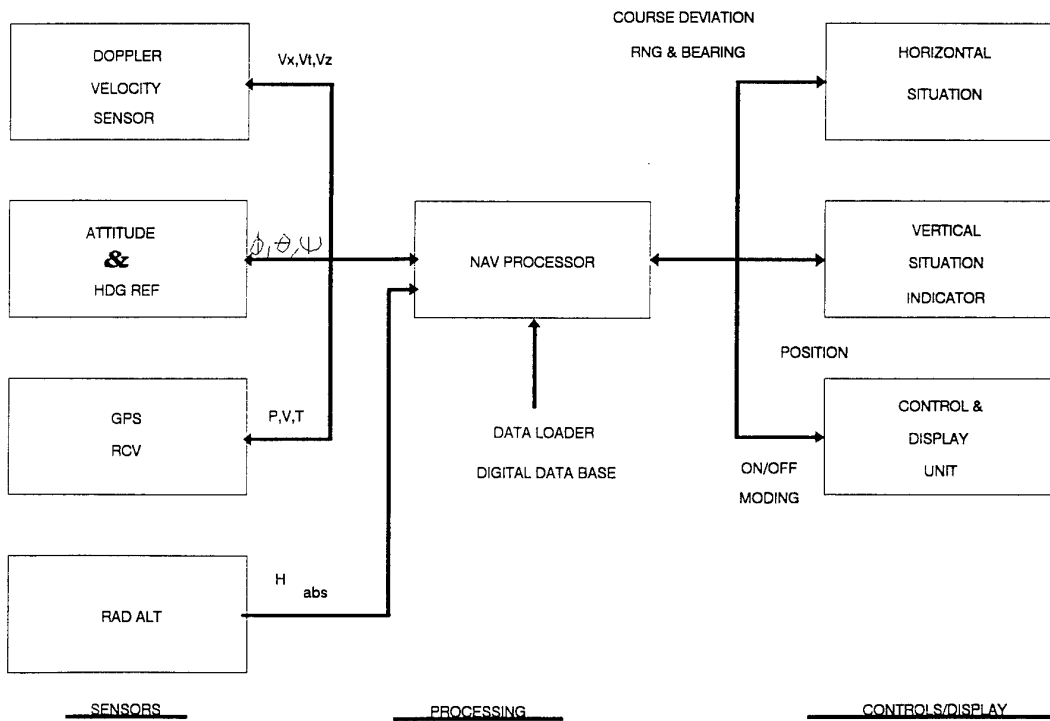n trajectories and orbits are addressed with consideration given to the various tradeoffs which may need to be considered. The broad spectrum of spacecraft are considered with emphasis upon those of greater military significance (i.e., near earth orbiting satellites.) Technical requirements include, but are not limited to, accuracy; physical characteristics such as weight and volume; support requirements such as electrical power and ground support; and system integrity. Generic navigation suites for spacecraft applications are described. It is shown that operational spacecraft rely primarily upon ground-based tracking and computational centers with little or no navigational function allocated to the vehicle, while technology development efforts have been and continue to be directed primarily toward onboard navigation suites. The military significance of onboard navigators is shown to both improve spacecraft survivability and performance (accuracy).

2. INTRODUCTION
The diversity of space missions and spacecraft presents a wide range of technical requirements, and it is useful to consider requirements in terms of spacecraft types. After a brief description of the space environment, spacecraft will be categorized by mission characteristics and flight regime. Flight trajectories are typically comprised of some combination of powered flight, coast, and aerodynamic flight segments. Technical requirements for different mission orbits will be presented and applicable navigation suites identified. A brief discussion of orbital mechanics is included to introduce concepts and nomenclature related to orbital navigation.

Following a description of spacecraft classes and missions, technical requirements and figures of merit for requirements specification are presented. With an understanding of the space environment, vehicle classes designed to operate within that environment, and technical requirements, we will then be in a position to describe state-of-the-art generic navigation suites for spacecraft. Finally, the relationship between technical requirements and generic navigation suites is characterized with tradeoffs considered.

3. SPACECRAFT
For most space missions, the spacecraft physical configuration changes at discrete time intervals during the mission as hardware is separated and jettisoned. Using an example of boosting a surveillance sensor (the payload) into earth orbit aboard a satellite, the final spacecraft configuration consists of sensor and spacecraft bus where the spacecraft bus provides support such as attitude control and electrical power for the payload. Initially, however, the spacecraft configuration as erected at the launch pad is a stack consisting of launch vehicle, upper stage, spacecraft bus, and payload. While the launch vehicle is active, its effective payload consists of the upper stage, spacecraft bus, and sensor. Similarly, after launch vehicle separation and upper stage activation, the effective payload is spacecraft bus and sensor. Therefore, in this example there are three vehicle configurations involved with support of this single payload and, as we shall see, navigation requirements and suites are considerably different among these vehicles. Prior to considering technical requirements and navigation suites for spacecraft, it is useful to characterize the space environment and vehicles designed for space operation.

3.1 The Space Environment
To describe the space environment, we consider first the boundary of space. The question of where space begins does not have a unique answer. Propulsion engineers may consider space to begin at 45 kilometers since this is the altitude at which a vehicle must provide its own oxidizer along with fuel. Aeronautical engineers may consider space as the altitude above which aerodynamic forces acting on a vehicle are considered negligible, which occurs above 100 kilometers. At approximately 160 kilometers altitude, there is insufficient air to transmit sound waves and scatter light so the region remains in silence and darkness. For navigational purposes, an altitude of 160 kilometers may be selected.

The abundance of space enables a wide range of trajectories which may be categorized according to: (a) surface-to-surface, (b) surface-to-space, (c) space-to-space, and (d) space-to-surface. These mission trajectories are sufficiently different to have resulted in development of several space vehicle classes. These classes, in turn, possess distinct technical requirements and employ differing navigation suites. We initially consider a broad spectrum of space vehicles and subsequently focus primarily upon those with greater military significance.

3.2 Space Vehicles
As indicated previously, the term *spacecraft* refers to that collection of unmanned and manned vehicles designed for operation within the space environment. Sounding rockets have surface-to-surface trajectories. Surface-to-space trajectory spacecraft consist of launch vehicles for boosting satellites from the earth surface into orbit, although booster components themselves fall back to the earth surface. Trajectories remaining within the space environment include upper stages for transferring satellites from parking orbit to final mission orbit, earth orbit satellites which remain in stable orbit around the earth, and interplanetary spacecraft. Vehicles with space-to-surface trajectories include descent, landing, and recovery vehicles.

Surface-to-surface trajectory vehicles are characterized by booster rocket stages to provide high thrust over a short time interval (near-impulsive) for payload delivery into a high altitude ballistic trajectory. *Sounding rockets* provide a short duration space environment and are typically used for environmental data acquisition while in the upper atmosphere or short-term (<15 minutes) microgravity environmental conditions for scientific experiments. The trajectory objective is primarily to establish a controlled vehicle attitude so sounding rockets are typically stabilized but usually unguided. These vehicles can be tracked by ground stations and additional terrestrial-based assets such as aircraft, and do not typically require onboard navigational data.

Launch vehicles provide propulsion and trajectory control to boost a payload into a specified orbit. These vehicles impart nearly all (but usually not quite all) the impulse required and are usually ballistic. Although typically ignited from a vertical orientation at a fixed site launch pad, one recently developed vehicle (Pegasus) for smaller payloads is released horizontally from an aircraft undercarriage. Several missions flown to date have released the launch vehicle from a B-52 aircraft at approximately 40,000 feet. Following release, Pegasus employs a two stage rocket to propel the payload into low earth orbit.

Boosters are either expendable or reusable. Current U. S. expendables are derivatives of ballistic missiles originally dating back to the 1950's, and this trend has been continued with recent refurbishment of Titan II missiles as launch vehicles for military satellites. Consistent with their ballistic missile heritage, most expendable launch vehicles use inertial guidance systems. Radio guidance is still in use by some older Atlas vehicles (Atlas E/F vintage), but these vehicles are no longer in production. The sole reusable vehicle in the U.S. inventory is the

NASA Space Transportation System (Shuttle) which is also the only U. S. manned spacecraft in operation. The Shuttle is launched into orbit in the traditional manner but atmospheric reentry is as a winged glider using aerodynamic surface control for cross range maneuver capability of approximately 2000 km during unpowered descent to runway landing. The Shuttle mission profile is shown by Figure 1.

During ascent, three redundant gimballed inertial measurement units (IMU) are used for navigation. There are two orbital navigation methods depending upon whether vehicle operation is powered flight or orbital coast. During coast, a ground uplinked state vector is propagated forward in time with force models for gravity, aerodynamic drag, reaction control system jet firings, and waste vents. In the presence of thrust, the navigation method is switched to propagate the trajectory using accelerometer data and a less sophisticated gravity model. Since all state vector data are derived from the IMU, disturbing acceleration models, and propagating the initial conditions, position accuracy degrades over time and ground-derived state vector data must be uplinked periodically. Scheduling of these uplinks depends upon the accuracy needed to support the mission timeline and flight operations.

The reentry navigation suite consists of IMU, tactical air navigation (TACAN), microwave scan beam landing system (MSBLS), and radar altimeter. Navigation aid usage during descent and landing is shown by Figure 2. Shuttle flight test has demonstrated an ability to detect TACAN from low altitude orbit, and on-orbit navigation software could be modified to incorporate this measurement type onboard. There is also interest in integrating a Global Positioning System (GPS) receiver onboard Shuttle which could be used during both on-orbit and descent mission phases.
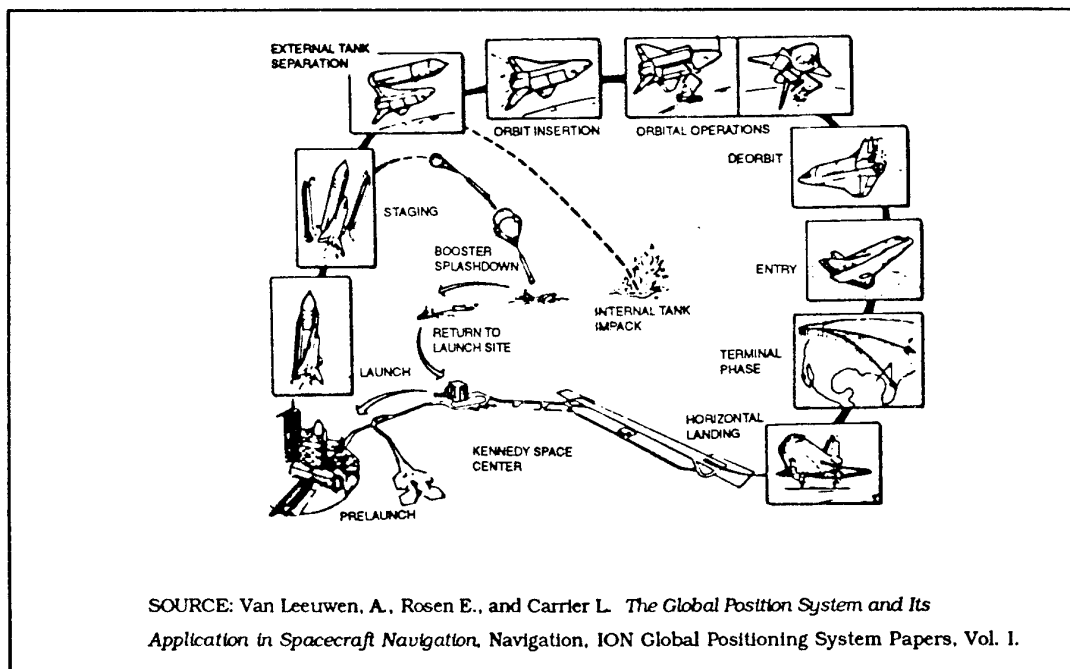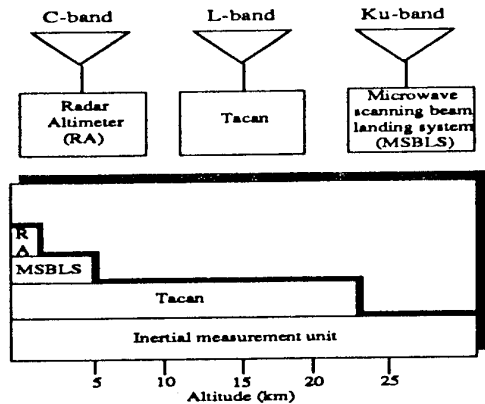

SOURCE: Van Leeuwen, A., Rosen E., and Carrier L. *The Global Position System and Its Application in Spacecraft Navigation.* Navigation, ION Global Positioning System Papers, Vol. 1.

Figure 1. Shuttle mission profile.

Figure 2. Shuttle descent and landing navigation aids.

Vehicles used to transfer satellites from one orbit to another are referred to as *upper stages* or, equivalently, *orbit transfer vehicles (OTV)*. Upper stages provide transfer from the launch vehicle established orbit to final mission orbit. Most upper stages are inertially navigated and guided. An exception is the Payload Assist Module (PAM-A and -D) which is unguided. The PAM provides payload transfer from parking orbit to geosynchronous transfer orbit (GTO). The absense of radio guided upper stages may be attributable to poor geometry between fixed, ground-based control centers and upper stages during the burn. Geostationary satellites are a primary user of upper stages as are interplanetary spacecraft which rely upon the transfer vehicle to impart sufficient velocity for escape from the earth sphere of influence.

The Inertial Upper Stage (IUS) is compatible with Shuttle and the Titan III and IV launch vehicles. For Shuttle configuration, the IUS inertial system propagates pre-launch alignment through launch and during on-orbit

phases of Shuttle operation. Deployment from the Shuttle payload bay is typically scheduled for 10 to 35 hours after launch so the IUS inertial navigation system is subject to accuracy degradation due to instrument errors and disturbing accelerations. The navigation state is typically updated prior to deployment by the Shuttle to eliminate error resulting from uncompensated atmospheric drag effects on the Shuttle vehicle. The IUS attitude can be updated pre-deployment with a star scan technique. The Titan trajectory is direct orbit insertion of IUS so the coast period is short and inertial navigation is adequate.

Earth orbital spacecraft operate at altitudes sufficiently above the sensible atmosphere to maintain stable orbit. We shall use the term *satellite* to refer to earth orbital spacecraft. Satellites have historically performed support roles for strategic military systems and more recently have emerged as a force multiplier for tactical operations [Ref. 1]. Military satellite missions include surveillance, meteorology, navigation, and communication. Mission scenarios for typical low earth orbit (LEO) and geosynchronous earth orbit (GEO) satellites launched by an expendable launch vehicle are shown by Figures 3 and 4, respectively.

Satellite navigational state information requirements are derived from two sources. The first requirement is the need for predictive navigational state information to support operations such as scheduling of ground contacts, radiofrequency (RF) signal acquisition and tracking, and command and control. These operational requirements can be satisfied with moderate navigational state information accuracy. The other requirement is for support of payload data processing such as data location and analysis. Payload users may require high accuracy such as for image data registration in geographic coordinates, but calculation of satellite position at the time of data collection may be accomplished long after-the-fact using precision orbit re-determination techniques as described later.
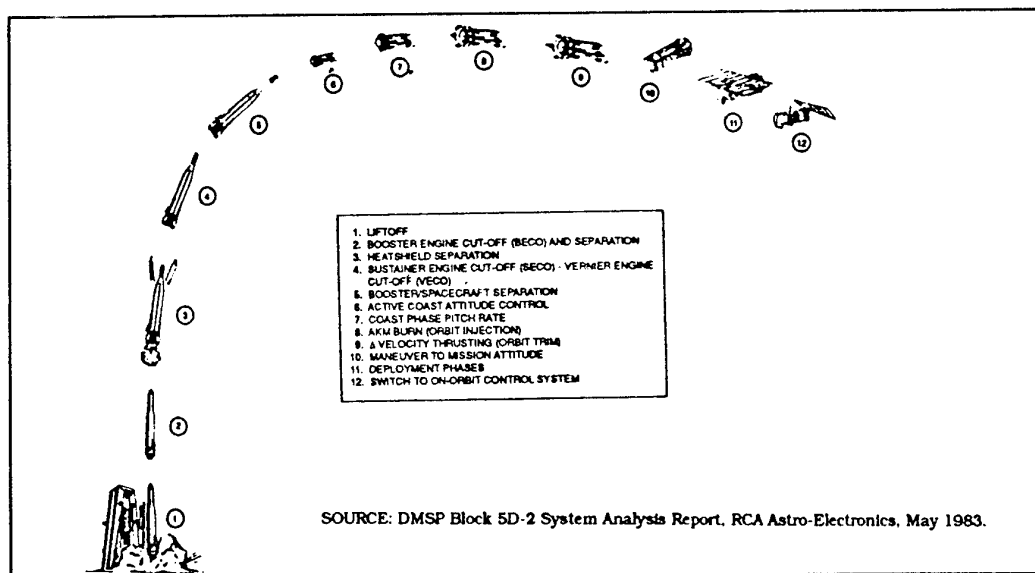
Figure 3. Mission scenarioes for typical low earth orbit launched on an ELV
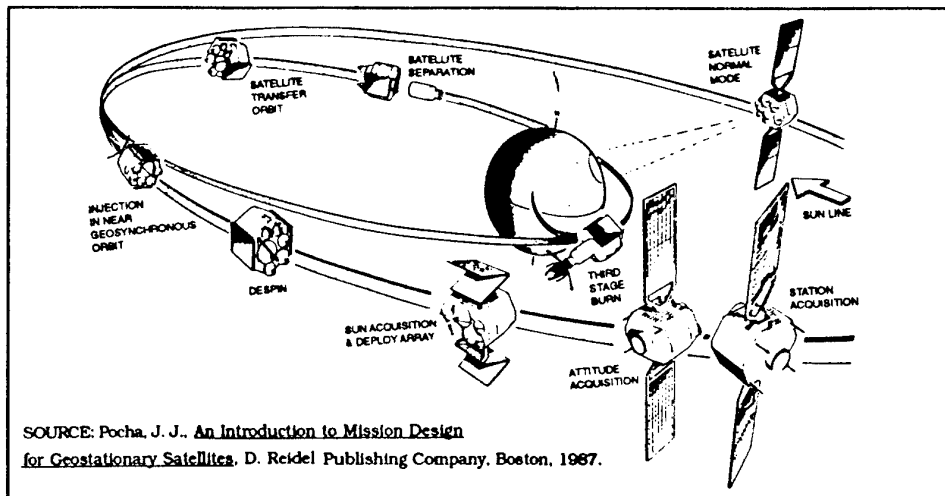
SOURCE: Pocha, J. J., An Introduction to Mission Design for Geostationary Satellites, D. Reidel Publishing Company, Boston, 1987.

Figure 4. Typical mission for GEO satellite launched on an ELV.

Onboard navigation for satellite missions is generally limited to the transportation vehicles (launch and orbit transfer). Once on-orbit, the satellite is tracked by the ground segment using multi-program and/or program-dedicated ground stations with support from computational centers. Several features of tracking and computational facilities which have been developed by North American Aerospace Defense Command (NORAD), U.S. Air Force (USAF), U.S. Navy (USN), and NASA are shown in Table 1. Tracking data are used in conjunction with orbital mechanics to determine the satellite orbit. The determined orbit parameters correspond to an epoch within the tracking data arc and provide the basis for prediction to future time using orbital mechanics and dynamics models. The tracking arc length may be as short as several minutes (for a LEO satellite with limited ground station visibility) or as long as several days (for GEO satellites with continuous line-of-sight visibility).

This solution to the satellite navigation problem has proved practical because in contrast with terrestrial-based systems, the nature of orbital kinematics and dynamics permits adequate prediction of vehicle navigation state (e.g., inertial position and velocity) over meaningful time intervals from knowledge of current vehicle navigation state. There are two methods by which satellites make onboard use of the predicted orbit parameters. First, the orbit prediction can be applied over a desired time interval (e.g., a day) in the form of an ephemeris which is essentially a catalog of position and velocity vectors at discrete times (e.g., every 12 minutes). For spacecraft requiring higher accuracy than can be provided with the ephemeris, interpolation between the onboard vectors may be performed. The other method is prediction of the determined orbit to a desired time with uplink of this time-tagged state vector as initial conditions for onboard extrapolation. With the ephemeris technique, error remains bounded over the applicable time interval, but a new ephemeris must be generated and uplinked by the ground segment for accuracy is to be maintained. With the initial condition extrapolation approach, accuracy degrades with time due to unmodeled errors in the disturbing accelerations. The ground-based tracking, orbit determination, and orbit pre-

diction technique is shown by Figure 5.

There are numerous reasons for this dichotomy between onboard inertial navigation for transportation spacecraft and ground-based tracking, orbit determination/prediction approach for satellites. One major reason is because most satellites have not required *real-time* navigation state knowledge. For those systems in need of onboard

ORGANIZATION	FACILITIES	FEATURES	MEASURABLES
North American Aerospace Defense Command (NORAD)	Space Sensor Network COBRA DANE COBRA JUDY (Ship) FPS-85 PAVEPAWS Ground-based Electro-Optical Deep Space Surveillance (GEODSS)	Cheyenne Mountain computational facilities; FPS-85 has range of ~8000 km Dish radars, phased array radars, electro-optical sensors (20 sensors total as of 1985) including Ballistic Missile Early Warning System (BMEWS) GEODSS range 5500-40,000 km altitude	Object data report (distance, azimuth, altitude, time)
U.S. Air Force	Satellite Control Network (SCN)		
U.S. Navy	Navy Space Surveillance (NAVSPASUR)	CONUS radar fence at ~33 deg N. latitude with 3 transmitters (1 primary and 2 backup) and 6 receivers. Can detect satellites with inclinations > 30 deg and altitude > 12,000 km	
NASA	Space Tracking and Data Network (STDN); Deep Space Network (DSN)	Ground tracking stations (C - band, S-band, and laser) in and outside CONUS; Tracking and Data Relay Satellite System (TDRSS)	Range, Doppler, Angles

Table 1. Tracking and computational featues.

position and velocity data, accuracy requirements have been satisfied with ground-based techniques. Another major reason is that technology to support on-orbit navigation (primarily significant digital computational capability) has not been available to satellites. This does not imply, however, that onboard navigation suites for satellites have not been of interest. Since the 1960's there has been continuing interest in onboard navigation to enhance the survivability of military satellites and reduce operating costs for civilian satellites. Consequently, numerous studies of autonomous navigation concepts have been conducted and reported in the literature with sporadic technology development and demonstration efforts interspersed. These concepts and the resulting technology

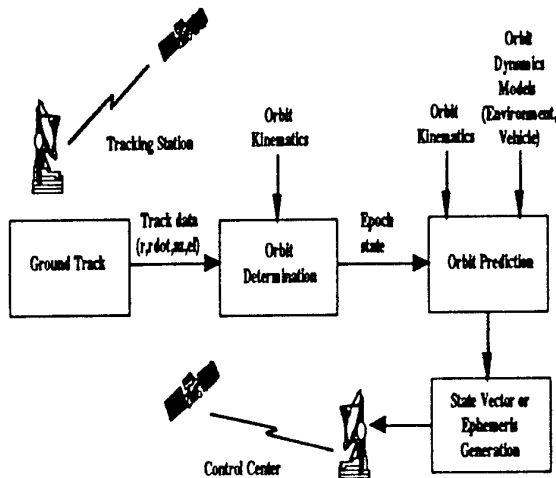efforts are considered in detail in a subsequent section.

Figure 5. Ground track, orbit determination/prediction technique.

The enormous technological evolution in inertial navigation for aircraft, missile, and naval systems over the past twenty-five years has had limited impact upon satellite navigation because the orbital navigation problem is inherently different from these other navigation problems. The basic goal of all non-orbital navigation problems is to measure specific force (the non-gravitational accelerations) which is compensated for gravity effects and then integrated to obtain velocity and integrated again to obtain position. In the orbital navigation problem there are essentially no specific forces acting on the vehicle (atmospheric drag, solar radiation pressure, and spacecraft-generated effects are typically five orders of magnitude smaller than gravitational effects), and the primary requirement is to characterize the complex gravitational field including the effects of a non-spherical earth and multi-body (e.g., earth, sun, and moon) interactions. In their simplest forms, non-orbital navigation is a problem of measurement whereas orbital navigation is a problem of modeling. Hence, the need for significant digital computational resources is driven by the need for complex onboard models.

Interplanetary spacecraft escape the earth gravity field and travel into deep space for encounter with other planets in the solar system. These vehicles are characterized by extremely long mission durations which may exceed ten years of continuous, non-serviceable operation at distances from the earth which can result in round trip communication time delays measured in hours. During the long interplanetary coast period, spacecraft orbital dynamics are extremely low and ground-based tracking, orbit determination/prediction techniques are employed using the Deep Space Network (DSN). From Figure 6, the DSN consists of three complexes of tracking sites located approximately 120 degrees of longitude apart to provide worldwide tracking coverage. Spacecraft designs to date have employed celestial, fixed attitude hold during interplanetary coast and inertial equipment used during other mission phases such as descent and landing has been turned off to conserve electrical power and extend operating life times. The purpose of the interplanetary missions which have flown to date has been to acquire basic

scientific knowledge of the universe and these vehicles are not directly relevant for military missions.

Descent and landing vehicles perform controlled de-orbit from space and land on a planet or moon surface. These vehicles use propulsion and aerodynamic surfaces to control the descent trajectory during a combination of maneuvering and coasting periods. Typically, these vehicles navigate with inertial navigation systems and radar altimeters. Probably the most well known of this class of vehicle is the Apollo Lunar Module. The mission profile for the more recent landing of Viking on the surface of Mars is shown by Figure 7.

Spacecraft return into the earth atmosphere at a known time is sometimes referred to as *reentry*. The reentry time is noted and the vehicle will either impact the earth or burn up in the atmosphere. Physical retrieval of a spacecraft which has reentered and landed on the earth surface is sometimes termed *recovery*.
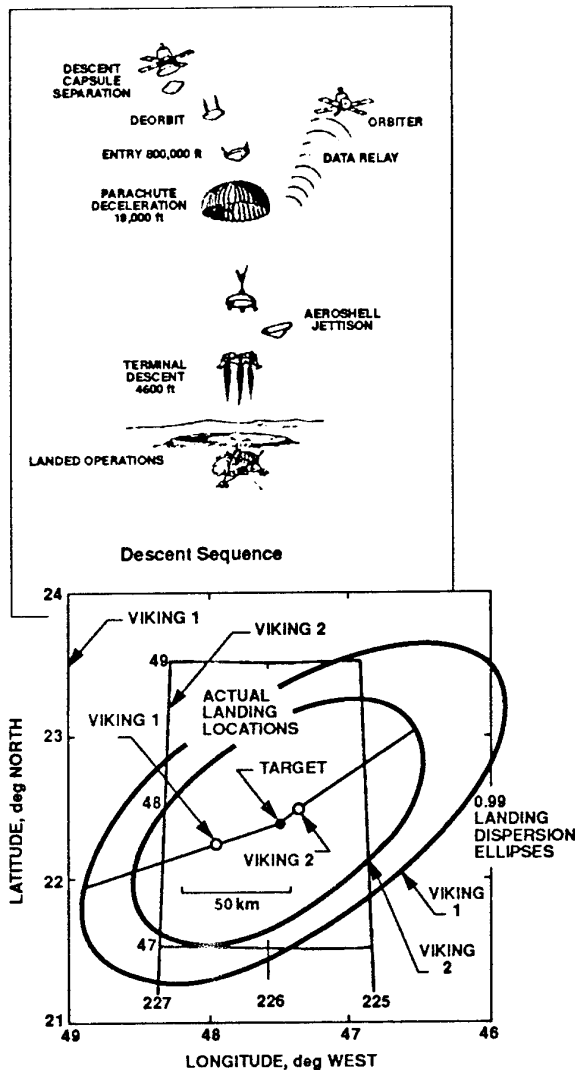
Descent Sequence

Figure 7. Profile of Viking Lander

3.3 Mission Orbits

Satellites operate in mission orbits selected to establish and maintain an acceptable environment for the payload. Orbital payloads perform data collection or data relay missions. Meteorology, surveillance, and scientific mission payloads observe and collect data then transmit the data to the ground segment for processing and analysis. Payload data analysis and post-processing may require precise orbit reconstruction using tracking data, extensive force models, and more sophisticated data reduction techniques than those applied during the original orbit determination process used for prediction. For example, The Ocean Topography Experiment (TOPEX) satellite mission (1334 km altitude and 63.1 degree inclination) will carry a radar altimeter with subdecimeter accuracy. As a goal, the precision orbit determination process will incorporate data from global GPS ground terminals and an onboard receiver which will be edited, calibrated, archived, and processed to generate precision orbits using techniques such as tuning the gravity model to fit the observation data.

Communication mission payloads receive data from a transmitter and retransmit the data to the receiver. Navigation satellites act as a data relay by broadcasting navigational information received from the ground. Data relay missions capitalize upon geometrical advantages offered by high altitudes of space to overcome the line-of-sight limitations introduced by curvature of the earth and, therefore, are typically earth-viewing payloads. Data collection missions are a mixture of earth- and space-viewing payloads. Store and forward payloads blend data collection with relay to collect transmitted data and store it onboard until properly positioned in space for retransmission to a cooperative but remotely located earth-based receiver.

Figure 8 indicates orbits of interest for space missions. Low earth orbit (LEO) is typically of low eccentricity, high inclination, and altitudes less than 1000 km. Mid Earth Orbit (MEO) is usually with low eccentricity, mid inclination, and altitudes between 10,000 and 20,000 km. Geosynchronous Earth Orbit (GEO) is nearly circular ($e<$ 0.1, typically), low inclination ($i<10$ degree, usually and often zero), approximately 40,000 km altitude, and an orbital period approximating one mean sidereal day (0.9 revolution/day $\leq n \leq 1.1$ revolution/day usually.) Molniya orbits are typically highly elliptical orbits at mid-inclinations. Super Synchronous Orbits (SSO) are those above 40,000 km and typically to 200,000 km. Earth escape requires velocity in excess of 11.2 km/s. Of these, orbits of military interest are primarily LEO, MEO, GEO, and SSO. We introduce these orbits because both navigation requirements and generic navigation suites are sensitive to orbit characteristics.

CURRENT SYSTEMS SUPPORTED BY AIR FORCE SATELLITE CONTROL

70 SATELLITES CURRENTLY SUPPORTED
90+ SATELLITES PROJECTED BY FY 97

SOURCE: U. S. Air Force

SPACE MISSIONS

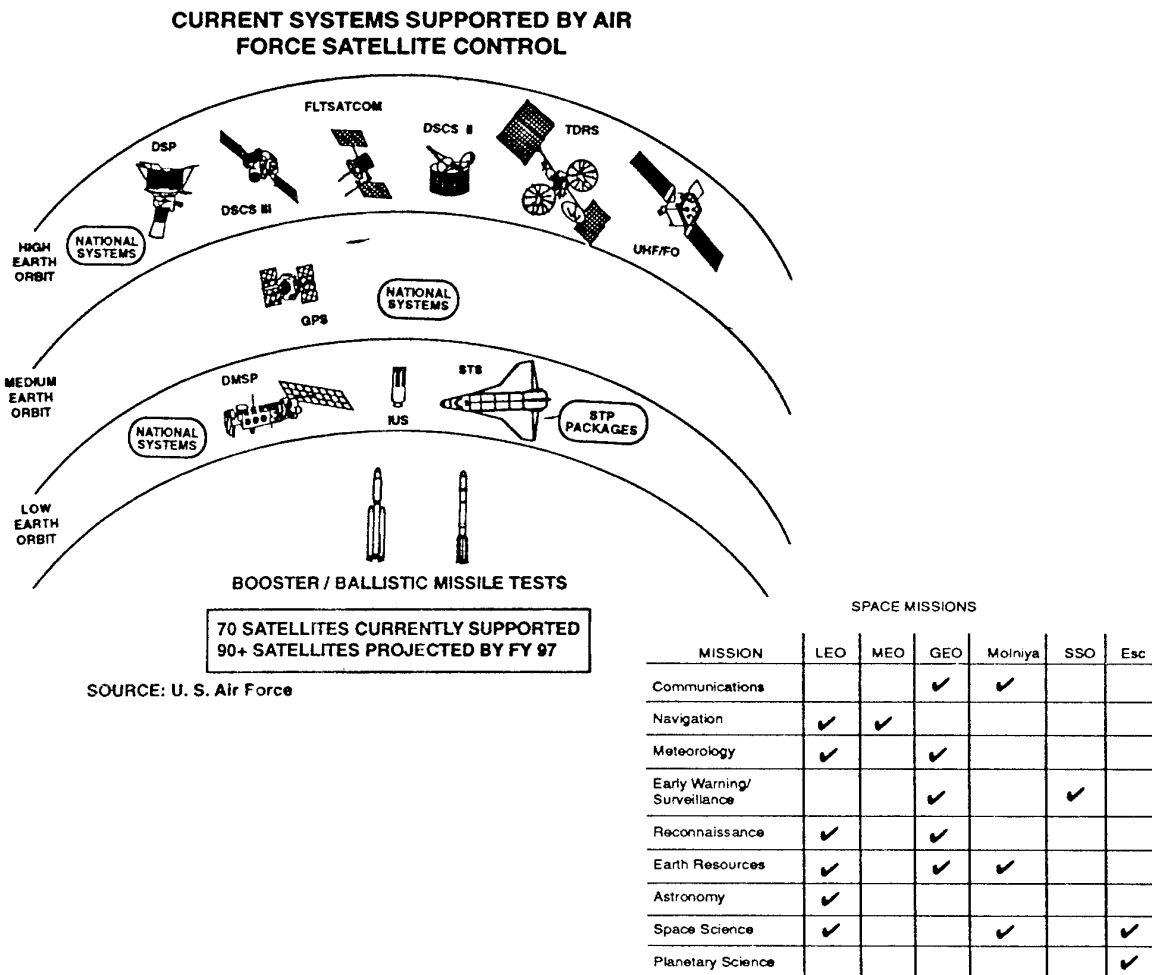| MISSION	LEO	MEO	GEO	Molniya	SSO	Esc
Communications			✔	✔		
Navigation	✔	✔	✔			
Meteorology	✔		✔			
Early Warning/ Surveillance			✔		✔	
Reconnaissance	✔		✔			
Earth Resources	✔		✔	✔		
Astronomy	✔					
Space Science	✔			✔		✔
Planetary Science						✔

Figure 8. Mission orbits of interest.

3. ORBIT MECHANICS

Orbital mechanics is the study of trajectories and orbits of space vehicles. Numerous books on the subject of orbital mechanics have been published [e.g., Refs. 2-5] and should be consulted for comprehensive development of the topic including the derivation of equations. This section is intended to introduce the basic kinematics of elliptical trajectories as applied to satellites in Earth orbit.

The ellipse orientation relative to an earth-centered reference system can be specified with three angles (Ω, ω, i) as shown in Figure 9. These angles are the longitude of the ascending node (Ω), argument of perigee (ω), and inclination (i). Specification of three additional constants related to the size and shape of the ellipse and time provide a set of six elements which completely specify the Keplerian motion. These constants are semimajor axis (a, orbit size), eccentricity (e, shape), and time of perigee passage (τ).
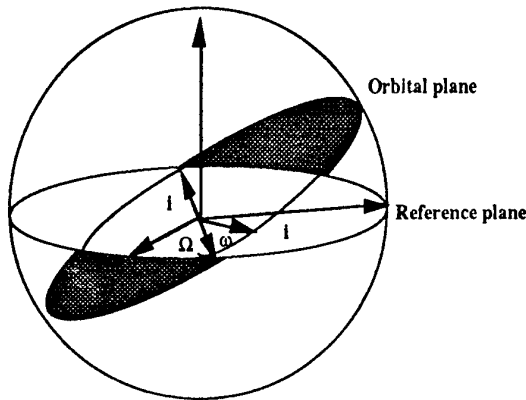

Figure 9. Orbital coordinates.

The six elements are uniquely related to position and velocity in Cartesian coordinates, and one useful form of this relationship is given by:

$$x = r (\cos \Omega \cos (\omega + f) - \sin \Omega \sin (\omega + f) \cos i)$$

$$y = r (\sin \Omega \cos (\omega + f) + \cos \Omega \sin (\omega + f) \cos i)$$

$$z = r (\sin i \sin (\omega + f))$$

where

$$r = \frac{a (1 - e^2)}{1 + e \cos f}$$

and angle f is the true anomaly measured from perigee. The relationship between τ and f is through mean anomaly (M), eccentric anomaly (E), and Kepler's equation according to:

$$M = n (t - \tau) = E - e \sin E$$

where n is mean motion. True anomaly is related to eccentric anomaly as:

$$\tan \left(\frac{f}{2} \right) = \sqrt{\frac{1 + e}{1 - e}} \tan \left(\frac{E}{2} \right)$$

In practice, forces other than the central force which act upon satellites cause deviations from a Keplerian orbit and variations in the orbital elements. Several approaches have been developed to describe the perturbed motion of satellites, and the reader may pursue development of these approaches in the cited references. The equations shown below are known as the Gaussian form of Lagrange's planetary perturbation equations. Several recommended references which derive these equations are [Refs. 2-4]. The nomenclature shown below is most closely attributed to Jerardi. The method involves resolving the perturbing force into three orthogonal components. The height component (H) is along the radius vector, the along track component (L) is in the orbital plane along the direction of satellite motion and perpendicular to H, and the cross track component (C) is normal to the orbital plane. First order perturbations to the orbital elements can then be formulated in terms of velocity changes (ΔV) along this local vertical local horizontal (LVLH) coordinate system centered at the satellite. The chosen set of orbital elements is comprised of the semi-major axis (a), eccentricity (e), inclination angle (i), longitude of the ascending node (Ω_N), argument of perigee (ω_p), and mean anomaly at epoch (χ) such that $M = n t + \chi$.

$$\Delta a = \frac{2 e \sin f}{n \sqrt{1 - e^2}} \Delta V_H + \frac{2 a \sqrt{1 - e^2}}{n r} \Delta V_L$$

$$\Delta e = \frac{\sqrt{1 - e^2} \sin f}{n a} \Delta V_H + \frac{\sqrt{1 - e^2}}{n a} \left(\cos f + \frac{e + \cos f}{1 + e \cos f} \right) \Delta V_L$$

$$\Delta \chi = \frac{a (1 - e^2) \cos f - 2 e r}{n a^2 e} \Delta V_H + \frac{r + a (1 - e^2) \sin f}{n a^2 e} \Delta V_L$$

$$\Delta \Omega_N = \frac{r \sin (\omega_p + f)}{n a^2 \sqrt{1 - e^2} \sin i} \Delta V_C$$

$$\Delta \omega_p = \frac{- \sqrt{1 - e^2} \cos f}{n a e} \Delta V_H + \frac{\sin f}{n a^2 e} \left(\frac{a (1 - e^2) + r}{\sqrt{1 - e^2}} \right) \Delta V_L - (\cos i) \Delta \Omega_N$$

$$\Delta i = \frac{r \cos (\omega_p + f)}{n a^2 \sqrt{1 - e^2}} \Delta V_C$$

The ΔV_H, ΔV_C, and ΔV_L correspond to velocity increments in H, C, and L directions, respectively. These velocity increments may represent disturbance sources or control inputs. Note that Ω_N and i are dependent only upon C while a, e, and χ are independent of C, and the argument of perigee is influenced by all three components.

The astute reader has by now observed the potential numerical problems associated with small eccentricity (e ≈ 0) and low inclination (i ≈ 0 degree) orbits. These cases can be solved through a change of variables [Ref. 5].

4. TECHNICAL REQUIREMENTS

Technical requirements imposed on spacecraft navigation systems may be derived from the vehicle physical characteristics, natural physical environment, natural operating environment, and mission orbit. Additionally, there may be requirements tailored to specific mission objectives. Although military and civilian weather satellites share many requirements, military systems will likely have survivability and tactical data distribution requirements which are unnecessary for civilian systems. Technical requirements for earth satellites will be emphasized as these vehicles currently possess military roles, with expectations for expanded roles in the future, and yet the navigation capability of these vehicles remains extremely limited.

We define performance parameters as accuracy, operating life, and computational loading. Accuracy is dependent upon the functions which use the navigation state data. For satellites, navigation state data generally supports functions which do not impose stringent accuracy requirements such as magnetic momentum management, solar array pointing, and mission data annotation. Table 2 illustrates functions which utilize satellite navigational data in terms of both current functions performed by the ground and space segments as well as future functions which could be performed onboard. A wide range of operating life exists for spacecraft. Launch vehicles may operate for 10 minutes while satellites with 5-year operating life must function continuously for 43,830 hours, and interplanetary spacecraft with a 10-year mission operate for 87,660 hours.

CURRENT		FUTURE
GROUND SEGMENT	SPACE SEGMENT	
•Plans for satellite commands both daily and long range	• Geographic pointing (0.01 deg)	• Space sequencing
		• Mission data analysis
• Orbit adjustment maneuvers for stationkeeping	• Magnetic momentum management	• Autonomous orbit control
	• Solar array pointing	– Stationkeeping
• Monitor satellite health and status	• Mission data annotation	– Co-orbit flying – Evasive maneuvers
	• Antenna pointing for command and data relay	– Rendezvous
• Satellite collision avoidance		• Long-term formation-flying with platforms
• Predict satellite reentry location		– Cooperative – Non-cooperative
• Detect newly-launched space objects		• Extensive interactions with spacebased data relay assets

Table 2. Functions using navigation data.

Finally, computational loading depends upon whether spacecraft translational kinematics are driven primarily by measurable, specific forces or complex gravitational effects which must be modeled, since models tend to be more demanding of computational resources. As indicated previously, acceleration is dominated by specific forces for sounding rockets, launch vehicles, and upper stages so the gravity model need not be highly complex. For non-maneuvering satellites, the gravitational effects and, to a much lesser extent, other non-contact forces dominate acceleration. Therefore, it is necessary to include higher order terms in the gravity model. During the process of a posteriori precision orbit determination, the gravity model may be "tuned" to the specific satellite orbit in such a manner as to provide a better fit of the computed orbit with the tracking data. Interplanetary spacecraft must account for gravity fields of other bodies such as sun, moon, and other planets depending upon location within the solar system.

The vehicle physical characteristics place constraints on the navigation suite. Primary parameters of interest are mass, power, and volume. Total electrical power available to satellites is limited by onboard power sources as shown by Figure 10. Satellite mass is limited by the launch vehicle/upper stage mass-to-orbit capability as shown by Figure 11. An upper limit on satellite physical dimensions will be introduced by a number of factors such as launch vehicle, upper stage, and fairing (payload covering to reduce drag during ascent) dimensions.

POWER SOURCE	CAPABILITY
Solar arrays	
Specific power	36 W/kg
Design life	10 years (GEO)

Primary batteries
Energy density

(Ag/Zn)	160 W-hr/kg
(LiSOCL2)	200 W-hr/kg

Design life

(Ag/Zn)	2 years
(LiSOCL2)	3 years

Secondary batteries
Energy density

(LiX, NAX)	80 W-hr/kg (GEO)
(NiCd)	15 W-hr/kg (GEO)
	5 W-hr/kg (LEO)
(NiH2)	30 W-hr/kg (GEO)
	25 W-hr/kg (LEO)

Design life

(NiCd)	5 years (LEO)
(NiH2)	2 years (LEO)

Fuel Cells
Primary
Specific power   110 W/kg
Radiothermal generators (RTG)
Specific power   5 W/kg

SOURCE: CRC Handbook of Space Technology: Status and Projections

Figure 10. Capabilities of satellite power sources.

	SCOUT	PEGASUS	ATLAS E	TITAN I	DELTA I
LEO	155-255	700	2090	5100	11,100
GTO				1043	1447
ESC			1500		1000 lunar
POLAR		600		4200	8300

	ATLAS I	TITAN IV	SPACE SHUTTLE
LEO	14,600	39,100	51,000
GTO	2770		1250-5910
ESC	1940-5890		2270
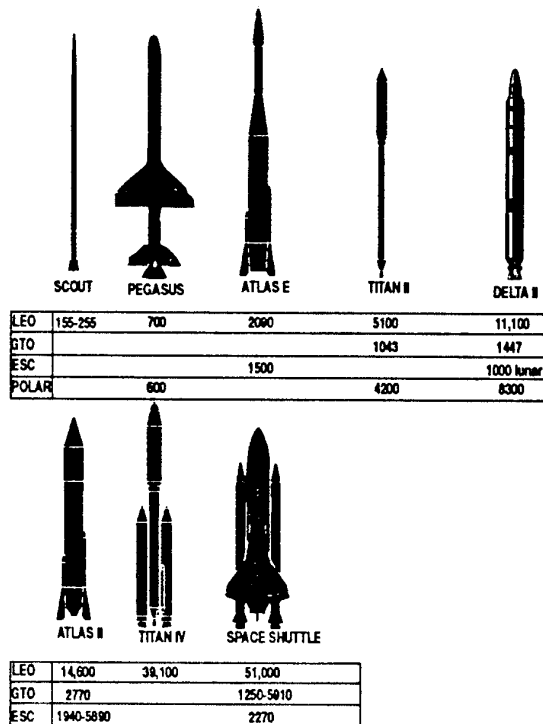
Notes:

1. LEO Low Earth Orbit
   GTO Geosynchronous Transfer Orbit
   Esc Earth Escape

2. Mass in kilograms

3. Variations due to orbit altitude and inclination

Figure 11. Mass to orbit capabilities of U.S. launch vehicles.

The natural environment imposes requirements on electronics and surfaces to withstand space radiation effects, thermal extremes, and the vacuum of space. Space radiation imposes hardening (i.e, protection of electronics against radiation effects which can result in temporary or permanent device failure) requirements against total and maximum dose exposures. Repeating sunlit and eclipse conditions to which the satellite is exposed result in significant temperature variations which must be accommodated. Materials and coatings are effected by space vacuum conditions.

The navigation suite must withstand the operating environment from launch through orbital operations. Launch operation introduces substantial forces and vibration while steady-state orbital operations may impose field-of-view and angular rate constraints on strapdown optical sensors thus influencing device performance characteristics, mounting location, and shading devices. Lastly, remote satellite operation precludes repair and maintenance opportunity for all but a fraction of satellite orbits limited by the "reach-ability" of the Shuttle.

Although requirements vary widely among missions and vehicles, a set of *ideal* technical requirements for an advanced, earth-viewing satellite are postulated by Figure 12.

POSITION ACCURACY	< 100 meters
POWER	< 20 watts
MASS	< 10 kilograms
VOLUME	<1000 cu. in.
LIFE EXPECTANCY	> 7 years
TEMPERATURE	-5 to 45 deg C
RADIATION	> 10^6 rad-Si
GROUND SUPPORT	≤ 1 update/mo.

Figure 12. *Ideal* technical requirements for satellite navigation.

4.1 Military Requirements

Since the early 1960's there has been concern regarding military satellite dependency upon overseas tracking stations which are perceived as vulnerable to a regime of threats including political climate, sabotage, and direct attack as well as natural disasters such as volcanic eruptions and hurricanes. There has also been concern regarding vulnerability of space-to-ground data links to denial from intentional jamming, nuclear blackout, and scintillation. Satellite autonomy has been regarded as a means to improve military asset survivability, but the belief that autonomy adds complexity and, therefore, reduces reliability has been an obstacle to autonomous satellite development. Typical life expectancy of military satellites ranges from 3-4 years at lower altitudes to 7-10 years at higher altitudes. At extremely low altitudes where atmospheric effects are more severe, operating life expectancy may be substantially shorter.

Many satellite requirements and designs reflect the fact that military satellite development and operation have only recently been transferred from the research and development community to military operations through establishment of U.S. Space Command. Satellites have had low production rate (often less than one per year) and volume (typically, a buy of six or less for a given design) with negligible technology transfer across program boundaries, resulting in custom-designed satellite systems. The GPS Navstar satellites represent the largest production base to date with procurement of 28 Block II vehicles and planned procurement of 20 Block IIR replenishment vehicles. Even within a given program, satellites are operated as isolated assets which may relay data from another satellite to the ground in "bent pipe" fashion to enhance data link survivability, but which do not make onboard use of the data. In the next century, the GPS Block IIR system is intended to actually exchange data among the constellation satellites and perform onboard data processing to provide autonomous navigation capability in the event the ground station becomes unable to provide ephemeris updates.

There are indications that future generations of spacecraft will have requirements reflecting enhanced operational capabilities and direct support for tactical as well as strategic operations [Ref. 1]. Emerging requirements levied upon satellites may include survivability against hostile threats and direct (and timely) reporting and/or relay of mission data to tactical users. In turn, implications for the navigation system might include:

(a) Direct distribution of collected data to tactical users without going through ground stations will increase the need for accurate mission data annotation and location so tactical users can receive and interpret data locally. Onboard navigation may need to support the location of tactical users near hostile territory such as via geographic latitude and longitude coordinates. As transmitted signal beamwidths become narrower to mitigate the potential for intentional jamming and unintentional self-jamming among multiple satellites and users, precision pointing requirements for the satellite transmitting antenna will become more demanding.

(b) Performance of autonomous orbit maneuvers ranging from routine orbit adjustment maneuvers (e.g., drag compensation, stationkeeping) to time-critical, responsive maneuvers for threat evasion and collision avoidance may be required. Autonomous return to normal operating conditions following maneuver completion has implications on response time requirements to damp out the maneuver effects on satellite attitude and rate.

(c) Intra-constellation data-sharing may be required. As indicated earlier, this trend has been initiated by the U.S. DoD GPS Block IIR program which intends to crosslink navigation data among the satellites to enhance survivability.

(d) Navigation accuracy may have to be maintained through conflict. This has historically been of concern for satellites supporting strategic missions but as the boundary between strategic and tactical becomes less distinguishable, satellites supporting tactical missions may expect to encounter sophisticated threats during regional conflicts.

4.2 Civilian Requirements

Civilian satellites provide science data such as stellar information and earth observation data such as land resources. An important civilian system is the Geostationary Operational Environmental Satellite (GOES) for detection of damaging weather systems such as hurricanes. Science and earth observation data post-processing may require precise orbit determination after the fact. Generally, the approach has been to reconstruct the orbit, analyze the residual error to identify improvements to the models, and re-process the orbit in an iterative manner. Future requirements for improved accuracy and shorter response time may impose new requirements which can only be satisfied with an onboard navigator.

4.3 Crew Requirements

Manned vehicles such as Shuttle are concerned with crew safety requirements. Safety-critical systems such as guidance, navigation, and control must provide an independent backup capability during flight-critical operations. Abort scenarios are needed to enable safe crew return throughout all mission phases. Navigation data must interface with crew command and display systems. Manned spacecraft have stringent reliability and redundancy requirements for navigation during flight-critical operations. With the exception of Skylab, manned US missions flown to date have been relatively short duration (≤ 2 weeks) in comparison with unmanned missions.

4.4 Interplanetary Requirements

Interplanetary spacecraft must be capable of autonomous operation during the long signal delay caused by the round trip light time at long ranges between ground control and spacecraft. Other special requirements are exceptionally long life expectancy and planet-relative navigation during planet encounters.

4.5 Commercial Requirements

Commercial space applications are primarily geostationary communication satellites which may be continuously tracked and controlled from a single earth station. Since service interruption equates to revenue loss, satellite design is frequently kept as simple as possible. The combination of satellite simplicity, favorable orbit characteristics, and strong relationship between payload mass and revenue are all drivers for minimizing satellite onboard requirements.

5. A SPECIAL CASE: NAVIGATION OF NAVIGATION SATELLITES

Kepler's laws of planetary motion are inadequate for precise orbit determination as needed for navigation satellites, and a two-step procedure is used. First, numerical integration of force models for external and internal disturbing accelerations provides a "computed orbit". Then a least squares fit of the computed orbit to a series of observations from ground tracking stations is performed [Ref. 6].

The ephemeris determination approach developed for GPS Navstar satellites is described in Reference 7. Major components of the determination process are: (a) pseudoranging measurement data from ground monitor stations, (b) batch estimator, and (c) on-line Kalman filter. Briefly, the batch estimator includes detailed force models to generate a batch-weighted least-squares estimate of the ephemerides, which are then predicted forward in time to generate an ephemeris reference for the Kalman filter. Both the estimator and filter operate on identical measurement data obtained from ground monitor stations. The filter fine tunes the reference ephemeris and provides a best estimate ephemeris as well as estimates for clock errors. The batch estimator models effects of the geopotential, atmospheric drag, solar radiation pressure, satellite thrusting, tidal forces, and multi-body force effects of the sun and moon.

6. REFERENCE SYSTEMS

Navigation state data requires a coordinate frame and time reference. Two coordinate systems of interest for spacecraft navigation are earth centered inertial (ECI) and earth centered earth fixed (ECEF) frames. The most recent ECEF frame is DoD World Geodetic System 1984 (WGS 84), and most recent ECI frame is the Conventional Inertial System (CIS) known as the FK5 System of Epoch J2000.0. The WGS 84 provides a reference frame, geometric earth figure, gravitational field model, and a means for relating positions on various local geodetic systems to an ECEF coordinate system. The J2000.0 defines an ECI frame at the new standard epoch J2000.0 (2000 January 1.5). Associated with this system are new theories for precession and astronomic nutation and new definition of universal time as adopted by the International Astronomical Union. Transformation between these frames is well-defined and real-time transform requires earth orientation prediction (i.e., earth rotation and polar motion).

The WGS 84 system is described in detail by Reference 8. The origin is located at the center of mass of the earth, the Z-axis is parallel to the direction of the pole, the X-axis is the intersection of the reference meridian plane and the plane of the equator, and the Y-axis completes a right-handed, earth-fixed orthogonal system. The coordinate system definition is shown by Figure 13.

Two time systems are universal and sidereal time - both of which are based upon diurnal rotation of the earth. Sidereal time is derived from Earth rotation with respect to the stars while universal time is related to diurnal motion of the sun. These time systems include effects of non-uniform rate of earth rotation and thus are non-uniform time systems. One mean sidereal day is equivalent to 23h 56m 04s of mean universal (solar) time. When corrections are applied to universal time to account for variations due to observed polar motion, the time scale is called UT1. A time scale whose epoch is periodically adjusted by integral seconds to maintain synchronization within 0.9 seconds of UT1, but whose rate is based upon atomic time (which is continuous) is known as Coordinated Universal Time (UTC). This time is provided by the U.S. Naval Observatory and is appropriate for annotation of events, observations, and other occurrences requiring a time tag, and for the independent variable in the equations of motion. Sidereal time is used in the transformation between WGS 84 and J2000.0.
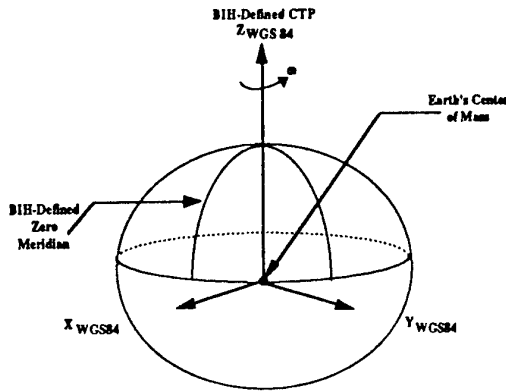
Figure 13. WGS coordinate frame axes.

Ephemeris time is related to the orbital motions of the earth, moon, and planets in the solar system. The ephemeris second is fixed by definition so ephemeris time is theoretically uniform. The relationship between ephemeris and universal time is given as:

$$E.T. = U.T. + \Delta T$$

where ΔT is a yearly increment provided in Reference 9. Thus, the relationship between ephemeris time and sidereal/universal time must be determined empirically. Reduction formulae for conversions among these time systems are included in detail in Reference 10.

7. NAVIGATION SUITES

The two primary components of navigation suites are the sensor complement and computational processor. Generally, for ground-based navigation, both components are located on the ground and similarly, for onboard navigation both are physically located onboard the satellite. One exception is ground-based processing of optical data from interplanetary spacecraft. As interplanetary spacecraft approach a planet, optical sensors can image the planet against the stellar background. This data is processed on the ground to estimate spacecraft flight path relative to the planet.

The type of observation measured by the sensor is sometimes used to classify an onboard system as either autonomous or aided. Autonomous navigation relies upon measurement of naturally occurring phenomena such as line-of-sight to astronomical bodies. Aided navigators process measurements from artificial sources such as GPS and ground beacons. This concept has evolved primarily from concerns of measurement denial and corruption which could adversely effect navigation system integrity. A formal definition for autonomous navigation is given by Figure 14. Conceptually, the space to ground interfaces among these techniques are shown by Figure 15. The aided navigator shown in the figure assumes a navigation satellite.

AUTONOMOUS NAVIGATION

An orbital navigator is considered autonomous under the following conditions:

- Its long-term error growth is bounded
- The navigation measurements do not require any (cooperative or un-cooperative) artificial support external to the system containing the navigator
- It is tolerant of moderate initial condition errors

Figure 14. Definition of autonomous navigation.

Ground-based systems may be thought of as tracking stations and computer programs which perform orbit analysis. Therefore, in a strict sense of the terminology these systems are not truly *navigation* systems. However, this technique is so pervasive for satellite position/velocity vector determination that it cannot practically be omitted from discussion of spacecraft navigation.
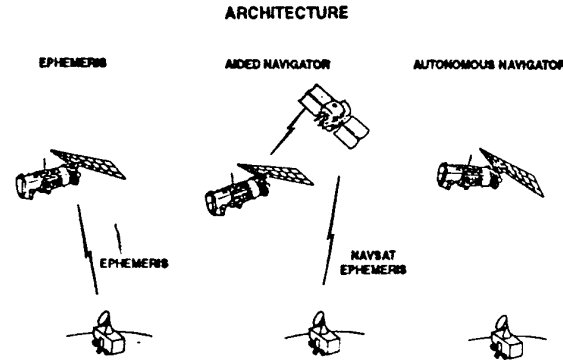
Figure 15. Ground-based, aided, and autonomous navigation conceptual interfaces.

7.1 Ground Systems

The ground-based approach to navigational state information currently used to support orbital satellites can be divided into four major functional areas: (a) data collection, (b) orbit determination, (c) orbit prediction, and (d) ephemeris generation. The former two areas are usually accomplished by shared facilities such as the NORAD Space Sensor Network, USAF SCN, USN NAVSPASUR, and NASA STDN systems. The latter two areas are typically handled by program-dedicated facilities. There are examples of systems, however, such as GPS where the end-to-end process relies primarily upon program-dedicated facilities.

7.1.1 Measurement Techniques

Common observables include line-of-sight (LOS) angles, range, and range-rate. Each observable may be collected cooperatively or non-cooperatively (to varying degrees of accuracy) with existing sensors. The DoD Space-Ground Link System (SGLS) provides accurate range and range-rate measurements via two-way S-band radio links. The inherent accuracy of sensors in use varies but is sufficient to preclude measurement errors from becoming a dominant error source during steady state operation. Opportunities for observation vary from approximately 10% for LEO satellites (<1000 km altitude) using only ground

stations to 100% for satellites in geosynchronous orbit. Non-cooperative observations can be obtained from radar (skin tracking) and optical sensors. Radar measurements often consist of time-tagged azimuth and elevation angles, range, and range-rate observables. Passive optical sensors are typically limited to angular observations.

Two-way ranging uses a coherent forward and return link to measure the elapsed time between transmission of the forward pseudorandom noise (PRN) code epoch (or an equivalent technique) and reception of the same epoch by the receiver channel. Delays introduced by signal propagation through ground station equipment and cables are subtracted from elapsed time by computer to calculate actual time delay. This measurement type requires the satellite to synchronize the return to the forward link code and requires a stable clock. If a satellite possesses a transceiver with proper ratio between forward and return carrier frequencies, then the ground system can measure two-way Doppler. One-way Doppler measurements can be made for any satellite.

For interplanetary spacecraft, telemetered optical data from celestial sensors onboard the satellite can be ground processed for planet-relative navigation data.

7.1.2 *Models*
Orbit determination techniques are well-established for stable orbits. Epoch state accuracies of less than 100 meters are achievable for a wide range of orbits and cooperative tracking can be used to reduce these errors by at least an order of magnitude. Since ground-based models need not execute in real-time, complex environmental and vehicle disturbances can be incorporated. As a minimum, higher order gravity terms are usually included. Other disturbances which may be modeled include aerodynamic drag, solar radiation pressure, lunar-solar attraction, and disturbances resulting from thruster firings. The magnitude of these disturbances is shown by Figure 16 as a function of satellite altitude. For satellites demanding higher accuracy navigational state information, the disturbance models and, in particular, model parameters, may be derived from empirical data acquired over long periods of time.
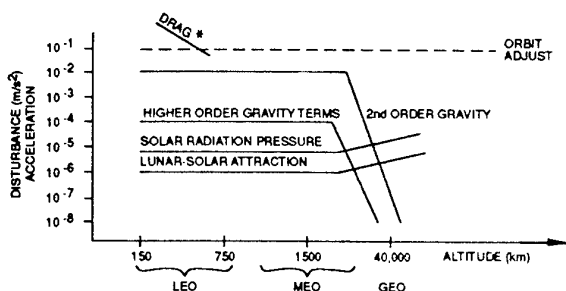
Figure 16. Disturbances as a function of orbital altitude.

7.1.3 *Clocks*
Ground-based systems employ precision oscillators of cesium beam and hydrogen maser atomic oscillator types. These oscillators exhibit high short-term (10^{-12}, $\tau = 1$ second) and long-term (10^{-14}, $\tau = 1$ day) stability. Clock drift is of the order of 10^{-15} over one day and 10^{-13} over one year [Ref. 11].

7.1.4 *Uplink/Downlink*
The ground system uplinks the predicted satellite navigational state as a time-tagged state vector or ephemeris. The frequency with which this must be performed is on-demand and depends upon the accuracy required so may occur as often as several times per day or as infrequently as once per month. For satellites with short passes over transmitting ground stations, it may be necessary to limit the amount of uplink data and ephemeris compression techniques may be employed. Additionally, commands to fire thrusters may be uplinked to effect orbit maneuvers such as drag adjustment and stationkeeping.

Downlink of computed position and velocity vectors enables ground operators to monitor the difference between onboard estimates and ground-based calculations. This allows unscheduled state updates to be uplinked to the satellite should the error become excessively large.

7.2 Flight Systems
Onboard navigators estimate satellite position and velocity vectors in real-time based upon a set of initial conditions, an *a priori* system model, and a sequence of external observations of the navigation state. Conceptually, the navigator consists of two basic elements: state propagator and measurement processor. Navigator inputs are a set of initial conditions and sequence of measurements which relate to the navigation state. The current estimate of the state vector is the primary navigator output. The basic structure for the onboard satellite navigator is depicted by Figure 17 [Ref. 12].
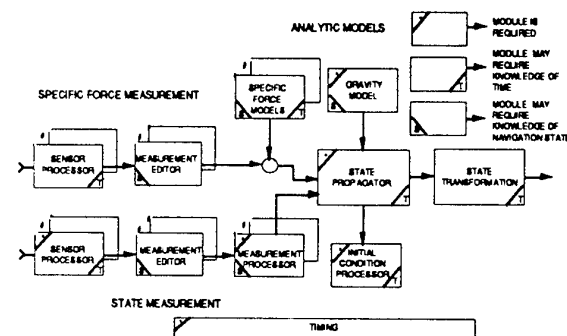
Figure 17. Conceptual diagram of onboard navigator.

The function of the state propagator is to extrapolate the initial state vector forward in time by numerically integrating a non-linear, vector differential equation which describes the motion of the satellite in a perturbed orbit. The state propagator is characterized by a state vector, a set of disturbing acceleration models and a numerical integration algorithm.

When specific force measurements are available, the state propagator may incorporate these as well. For example, if performance characteristics of inertial instruments are capable of measuring the drag force, these measurements could be incorporated to improve accuracy.

7.2.1 *Measurement Techniques*
Onboard measurements of navigation state may be optical or radiofrequency (RF). Astronomical optical sources include earth horizon, sun, moon, and stars. The measured

phenomenology is the angle between two optical sources (typically, a star and near body such as star/earth horizon). An optical sensor oriented toward the earth surface may use natural surface features for landmarks. Optical navigation aids include artificial ground landmarks and laser passive satellites. Radiofrequency aiding sources include GPS, TRANSIT, TDRSS, and RF ground beacons. Note that the measurements described are useful for attitude determination or communications systems and, therefore, such equipment may be shared with the navigator.

7.2.2 Measurement Sources

There are three existing satellite systems with navigational missions which provide radiometric measurements. GPS provides one-way range and range-rate data and broadcasts ephemeris data. TRANSIT provides range-rate data and broadcasts orbital elements. These space-based navigation systems were developed for terrestrial-based users, and transmitting antennas point toward the earth. Therefore, these systems provide measurement coverage for lower altitude satellites (h <20,000 km for GPS and <1000 km for TRANSIT). Past studies have considered GPS for satellites at altitudes above 20,000 km by receiving signals from satellites on the far side of the earth as shown by Figure 18. The GPS signal structure was developed to be tolerant of intentional and non-intentional interference, and signal features appear in the literature [Ref. 13]. The DoD policy to phase in GPS as the primary navigation aid and phase out TRANSIT and TACAN systems discourage development of satellite navigators dependent upon these systems.
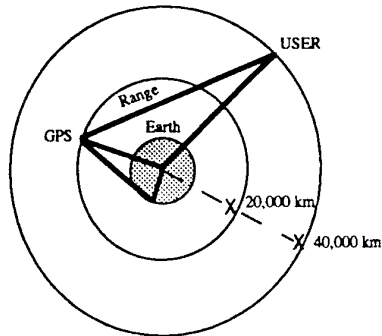

Figure 18. GPS reception at GEO.

As currently used, TDRSS provides a bent pipe between user satellite and ground terminal. Measurement types provided are one-way Doppler, and two-way range and Doppler. Past studies have examined one-way range and Doppler from the ground station through TDRSS with measurement processing onboard the user satellite [Ref. 14]. There are two active TDRSS satellites in geostationary orbit at 41 and 171 degrees west longitude which require ground support for ephemeris generation. The TDRSS ephemeris error is larger than GPS, and accuracy achievable with TDRSS cannot match that of GPS. Moreover, TDRSS satellites do not broadcast ephemeris information. The satellites possess steerable antennas so they may be capable of supporting higher altitude users, although relatively low geometrical dynamics between TDRSS and higher altitude users may limit achievable accuracy. A major limitation of TDRSS for military applications is the lack of survivability of the TDRSS

system. Use of TDRSS for onboard navigation would be most appropriate for satellites already requiring a TDRSS antenna for the data relay capability.

Ground beacons at known locations can provide range and/or range-rate measurements. Transmitters provide one-way range and/or range-rate while transponders provide two-way range and/or range-rate. For survivability purposes, ground beacons for U.S. satellites should be located within the continental U. S., which limits measurement coverage.

Inertially-referenced line-of-sight measurements to ground landmarks can be used for satellite navigation. Two types of tracking algorithms have been reported in the literature [Ref. 15]. Known landmarks are earth-fixed at a known location. Unknown landmark tracking also assumes earth-fixed landmarks but rather than require *a priori* location knowledge, landmark coordinates are included in the state vector and estimated by the navigation algorithm. Optical sighting to ground landmarks possess several disadvantages. At visible wavelengths, landmarks are obscured by clouds and depend upon the availability of natural lighting (i.e., daylight only). Longer wavelengths such as infrared (IR) penetrate clouds, but have lower resolution. The image must be recognizable and uniquely identifiable and at some altitude (which will vary with tracker design) detection and discrimination will become infeasible. Some limitations of passive optical measurements may be overcome with active (cooperative) systems. For example, an onboard IR laser could operate with ground beacons (retroreflectors) or an onboard laser could operate with orbiting satellite retroreflectors. Again, U. S. satellites using ground beacons should rely only upon those within the continental U.S. for reasons of survivability.

Astronomical measurements of the angle between a star and near body can provide navigational updates. For near-earth orbiting satellites, the body is typically the earth horizon or moon. A related technique is observation of stellar refraction or dispersion as stars set behind the atmospheric horizon of the earth as seen from the satellite. Optical sightings from the satellite are highly coupled with satellite attitude, and attitude knowledge errors contribute to navigation error for these systems. Radiometric measurements made by satellites can provide higher accuracy by at least an order of magnitude and are essentially decoupled from vehicle attitude.

Passive satellites in highly stable orbits which are predictable over long time periods can also provide a navigation aiding source. These laser retroreflector satellites are simple in design and highly reliable. The laser geodynamic satellite (LAGEOS) launched in May 1976 to an altitude of 5900 km is such a satellite. This measurement source has the potential for high accuracy and immunity to jamming since narrow beamwidths can be used and only space-to-space links are involved. Small, mobile ground tracking stations are sufficient for tracking the laser passive network.

With a constellation containing an adequate number and distribution of satellites, the member satellites can make inter-satellite measurements using communications equipment to provide relative navigation information. This

type of navigation has been considered for the Strategic Defense Initiative space-based assets [Ref. 16] based upon direct vehicle-to-vehicle ranging techniques. The GPS Block IIR program plans to use a UHF time slot broadcast technique [Ref. 17] for crosslink measurements.

Studies have also examined the ejection of a navigation aid satellite from the primary user satellite into a near orbit for tracking. [Ref. 18] Radar measurements of range, range-rate, and angles were considered as well as laser measurements of range and angles. The method is predicated upon reference of the relative state measurement to an inertial coordinate system as established with an IMU and/or star trackers.

7.2.3 Models

To achieve the accuracy required, onboard navigator models provide compensation for dominant disturbing accelerations acting upon the satellite. Since modeling errors are bounded by incorporation of measurements of the navigation state, reduced order gravity models should prove adequate for many satellite applications. Low altitude satellites may also need to incorporate an atmospheric drag model. High altitude systems may also need to model the effects of solar radiation pressure and lunar-solar attraction.

7.2.4 State Vector

Satellite navigation state is typically defined as inertially-reference Cartesian position and velocity vectors. Although this particular state vector has a convenient interpretation in terms of the physical quantities of interest, the state vector elements suffer from a wide dynamic range due primarily to the central force gravity term. A satellite at a 1000 km orbital altitude around the equator of a spherical earth will experience a variation of ± 7378 km in position over the 100 minute orbital period and, correspondingly approximately ± 7.4 km/s in velocity. This large dynamic range imposes a significant computational burden on the state propagation technique by requiring extremely small integration step sizes and a high level of numerical accuracy.

The constrained nature of the on-orbit navigation problem suggests the use of a state vector which is time-invariant for undisturbed two-body motion. A set of six orbital parameters defines the motion of a satellite in its orbit. For the special case of a true Keplerian orbit, these parameters will be constants. In general, these time-varying orbital parameters are functions of the perturbing accelerations.

7.2.5 Flight Computers

State-of-the-art military satellite computers in the early 1980's were examined by Reference 19. Since then, commercial advances in digital data processing hardware capability have repeatedly increased performance such as throughput and memory size while simultaneously decreasing size and weight. Although the full extent of these advances has not been realized for satellite computers, some advances have been made here as well. Processor development using the MIL-STD-1750A standard instruction set architecture (ISA) has been actively pursued by several vendors, and was selected by the MILSTAR program. A 1750A Ada™ compiler has been validated

and personal computer (PC) support tools for 1750A software development have become available. More recently, the Generic VHSIC Spaceborne Computer (GVSC) program is developing radiation-hardened 1750A computers using VHSIC technology. These computers are expected to provide three to four mega-instructions per second throughput capability.

7.2.6 Clocks

Quartz crystal, rubidium, and cesium oscillators are candidates for onboard clocks. Both quartz crystal and rubidium oscillators exhibit advantages with respect to size, weight, power, and cost. Rubidium clocks provide good performance in the presence of a hostile environment. Cesium clocks provide excellent long-term performance at a higher cost, weight, power, and volume penalty.

8. TRADE-OFFS

The first trade-off involves determining whether navigational state data is required onboard the satellite, and, if so, whether the data should be provided to the vehicle by ground-based orbit determination techniques or onboard navigation. Ground-based techniques offer advantages for vehicle weight, power, volume, and complexity. Onboard approaches offer advantages of autonomy, survivability, and accuracy. Figure 19 [Ref. 1, Paper No. 24] shows the potential accuracy improvement with onboard GPS navigation as compared with ground-based orbit determination. This figure illustrates that ground-based techniques achieve best performance in the 2,000 to 20,000 km altitude region. At lower altitudes the coverage achievable by ground stations is limited and atmospheric drag effects are most severe. At higher altitudes, there is little change in geometry between the satellite and ground stations.
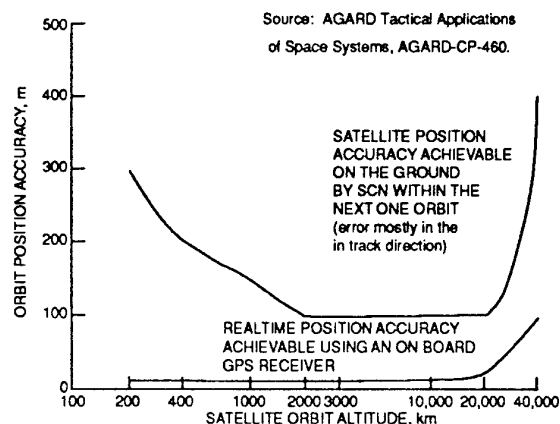
Figure 19. Onboard navigation accuracy with GPS as compared with ground-based orbit determination techniques.

Passive optical measurements from satellite to ground are subject to cloud coverage, lighting conditions, seasonal effects, and sensor pointing errors. For survivability reasons, the landmarks should be located within the continental U. S. (CONUS) which may significantly limit measurement opportunities. These observation types are more applicable to lower altitude satellites since accuracy depends upon ground resolution. Some problems associ-

ated with natural landmarks (e.g., seasonal and weather effects, correlation with reference landmarks) can be ameliorated with artificial landmarks such as search beacons. These landmarks then require maintenance to ensure availability.

Cooperative techniques from space-to-ground would have lasers onboard the satellites and ground-located optical beacons (retroreflectors). Laser beams are subject to atmospheric effects and may require two frequencies for refraction compensation. Survivability against blinding and beacon denial requires beacons be CONUS located which will constrain measurement opportunities. Cooperative space-to-space systems with laser rangers onboard the user satellites and cornercube retroreflecting satellites are not subject to atmospheric effects, but require ephemeris maintenance of the retroreflector satellites although highly stable and, therefore, predictable orbits can be used. This technique appears to have the potential for very high survivability. The narrow beamwidth of lasers may require stringent pointing capability for the user.

Sightings to earth horizon, sun, and moon from the user satellite also offer high survivability as it is difficult to deny these phenomenologies. The accuracy achievable with horizon sensors is limited by knowledge of the horizon and is subject to seasonal variations as well as local variations. The measurement is an angle between the local vertical and another celestial body such as sun, moon, or star. As an angle measurement, the position accuracy degrades with orbit altitude so at geosynchronous altitude the variations in horizon are less detectable but achievable accuracy is decreased. The sun and moon sightings are limited by the ability to determine the center of mass of these bodies. The use of star sightings with an earth horizon offers the best measurement availability since sun and moon sightings are highly geometry specific.

Radio frequency measurements emanating from the ground and received by the satellite would enable a relatively simple receiver since transmitter location could be known and fixed. Anti-jamming and anti-spoofing techniques need to be applied and the transmitter should be CONUS located for systems with survivability requirements. This technique is essentially independent of satellite pointing errors. The transmitters could be sufficiently powerful so only a low power omni-directional antenna would be required by the satellite. If the signal were modulated with position and time information, this system would be the inversion of GPS. For accuracy, two frequencies are required to correct for ionospheric effects on signal propagation. Reception from GPS satellites directly provides greater measurement opportunity for low altitude satellites and does not involve signal propagation through the atmosphere. Ground beacons could still be used to supplement GPS if so desired. The survivability of satellites using GPS is dependent upon the survivability of the GPS system. The GPS signal is specially designed for protection against electronic jamming and spoofing. This is not the case for TDRSS which is a "soft" system. GPS offers the potential for high accuracy without demanding onboard resources in terms of power, weight, and computational processing.

Before concluding, we consider the use of radar for satellite navigation from space to ground. Radar can measure range and range rate to the surface through uncooperative radar return from the surface or cooperative return from microwave reflectors. Radar provides an all-weather, day/night measurement capability. Real aperture radars consume substantial power and are terrain model sensitive. The support requirements for real aperture radar (primarily size, weight, and power) with range capability in excess of several hundreds of kilometers are not compatible with satellite support capabilities. Additionally, since power requirements increase substantially with range, onboard radar for anything but the lowest altitude satellites is not practical. The need for large antennas also is a problem. Generally, radars are inappropriate for satellite navigation. The role of radars to date has been for rendezvous with other spacecraft and as altimeters during surface landing. Synthetic aperture radar (SAR) offers antenna size advantages over real apertures but use of the data onboard would be extremely demanding of signal and data processing resources to obtain a robust correlation of SAR images with onboard reference images.

9. SUMMARY
In this section we have considered technical requirements and generic navigation suites for spacecraft with emphasis upon military satellites. Both ground-based and satellite-based navigation techniques were addressed. Navigation suites were considered in terms of the phenomenology providing the measurement source. We also considered the other major components of the navigation system such as algorithms, computers, and clocks.

DISCLAIMER
Review of this material does not imply United States Department of Defense indorsement of factual accuracy or opinion.

References
1. AGARD Tactical Applications of Space Systems, AGARD CP-460, October 1989.

2. Jerardi, T. W., Orbital Mechanics, Anthology, (undated).

3. Taff, L. G., Celestial Mechanics, A Computational Guide for the Practitioner, John Wiley & Sons, Inc., New York, 1985.

4. Moulton, F. R., Celestial Mechanics, Macmillan, New York, 1902.

5. Battin, R. H., Astronautical Guidance, McGraw-Hill, New York, 1964.

6. Ashkenazi, V. and Moore, T., The Navigation of Navigation Satellites, J. Navigation (GB), Vol. 39, No. 3, September 1986, pp. 377-93.

7. Van Dierendonck, A. J., Melton, W. C., Birnbaum, M., and Harkens, M. D., *The Approach to Satellite Ephemeris Determination for the NAVSTAR Global Positioning System*, Navigation, Vol. 23, No. 1, Spring 1976, pp. 76-86.

8. Department of Defense World Geodetic System 1984, *Its Definition and Relationship with Local Geodetic Systems*, DMA TR 8350.2, 30 September 1987 and Supplements DMA TR 8350.2-A Part I, *Methods, Techniques, and Data Used in WGS 84 Development*, DMA TR 8350.2-B Part II, *Parameters, Formulas, and Graphics for the Practical Application of WGS 84*.

9. *The Astronomical Almanac*, (USA/GB), US Government Printing Office, Washington.

10. *Explanatory Supplement to the Astronautical Ephemeris and American Ephemeris and Nautical Almanac*, Her Majesty's Stationery Office, London, 1961.

11. Fruejauf, H., *Precision Oscillators and Their Role and Performance in Navigation Systems*, PLANS 1982.

12. Miller, J. L. and Daly, K. C., *Autonomous Navigation for Orbital Satellites*, Journal of Defense Research, Vol. 17, No. 3, Fall 1985, pp. 242-268.

13. Spilker, Jr., J. J., *Signal Structure and Performance Characteristics*, ION GPS Papers, Vol. I, pp. 29-54.

14. Elrod, B. D., *Low Earth Orbit Navigation in the Tracking and Data Acquisition System Era*, AIAA J. Guidance, Vol. 8, No. 4, July-August 1985.

15. Kau, S. P., *Autonomous Satellite Orbital Navigation Using Known and Unknown Landmarks*, AIAA Guidance and Control Conference, AIAA 75-1096, 1975.

16. Treder, A. J., *Autonomous Navigation - When Will We Have It?*, Navigation, ION, Vol. 34, No. 2, Summer 1987, pp. 93-114.

17. Ananda, M. P., Bernstein, H., Bruce, R. W., Cunningham, K. E., Feess, W. A., Jorgensen, P. S., Menn, M., and Price C. M., *Autonomous Navigation of the Global Positioning System*, AIAA 84-1863-CP, 1984.

18. Muller, E. and Kachmar, P., *A New Approach to On-Board Orbit Navigation*, CSDL-E-2566, February 1971.

19. Gilley, G., *Digital Hardware for Use in Spacecraft Control Applications*, AAS 80-031, Guidance and Control 1980, Vol. 42, pp. 631-50.

SECTION VI
TEST METHODOLOGY

Coy L. Hunt
46th Test Squadron/TGGTN
1644 Vandergrift Road
Holloman AFB NM 88330-7850
USA

6. TEST METHODOLOGY

6.1 Introduction

This section of the document describes the various methods employed to evaluate the performance of aerospace navigation systems. Included are discussions regarding test equipment, test data, reference systems, environmental conditions, laboratory tests, flight test profiles and procedures, and statistical methods used to measure system performance. The test methods detailed in this document are somewhat specific and may need to be adapted to particular systems and their specifications.

6.1.1 *Test System*

The basic component of the test system is an inertial navigation unit (INU) and may be integrated into a system utilizing a multi-sensor configuration. All sensor components should include their own monitoring units to facilitate real-time validation of operational performance.

6.1.2 *Data Acquisition System*

The data acquisition equipment should be capable of recording the necessary parameters output by the integrated system as well as the navigation parameters from the individual sensors. The required data parameters, sample rates, and timing requirements are specified in Section 6.2.

6.1.3 *Reference System*

The requirements for reference system data acquisition equipment are quite similar to those in Section 6.2.2 above. Additional discussion related to the reference system is presented in Section 6.3.

6.1.4 *Laboratory Test Equipment*

The required laboratory test equipment includes test tables capable of positioning the test system in any number of specific orientations. Equipment capable of subjecting the test system to various environmental conditions (vibration, heat, cold, humidity, etc.) will also be required. Additionally, equipment for testing Electromagnetic Interference (EMI) should be available.

6.2 Test Data

Test log data should include complete identification of the test article and all test equipment and accessories as well as basic data accuracies. The log data should include the actual test sequence used, ambient test conditions recorded periodically during the test, and performance data recorded periodically during the test. Laboratory data should include as a minimum: ambient temperature, humidity, latitude, longitude, altitude, cooling air temperature, cooling air flow rates, and power. It is recommended that aircraft data on temperature, humidity, vibration, cooling air temperature, cooling air flow rates, and power be included to help isolate system performance problems or anomalies. Performance data should include, as a minimum, indicated position and velocity components. As a minimum, the data recording rate for static tests should be one-per-minute. For dynamic tests, the recording rate should be one-per-second.

6.3 Reference System for Flight Tests

6.3.1 *Description*

An independent reference system is required for use as a truth model against which the test system performance will be compared. The accuracy and reliability of the reference system must have been previously established in order to reliably establish the performance of the test system.

6.3.2 *Accuracy*

It is highly desirable that the accuracy of the reference system exceed the estimated accuracy of the test system by a factor of ten. This will insure that the significant portion of the error attributed to the test system is in fact generated by the test system and not by the reference system.

6.3.3 *Data Rate*

Data from the reference system should be available at least once-per-second. If the test aircraft is performing maneuvers that involve high dynamics, then the sample rate should be higher than one-per-second to prevent the introduction of interpolation errors.

6.3.4 *Restrictions*

Flight test profiles should be designed to fly in the area where the reference system can provide the most reliable tracking. If this is not possible, then it should be noted that the reference system is likely producing data with degraded accuracy when flying in areas where reference sensor coverage is marginal.

6.4 Environmental Conditions

6.4.1 *Restrictions*

Discretion must be exercised in connection with the testing of any navigation system to insure that the environmental limitations of the test system are not exceeded. These limits are defined by the manufacturer and should be noted prior to the initiation of any type of testing.

6.5 Laboratory Tests

6.5.1 *Assumption*

This test procedure assumes that a series of tests has previously been conducted on the individual components of which the integrated system is comprised. Therefore, the purpose of all laboratory tests will be to determine integrated system performance for a variety of static and dynamic modes while experiencing diverse environmental conditions. These tests are primarily designed for the unit containing the inertial navigation system whether or not other systems are integrated into the same unit.

6.5.2 *Static Tests*

These tests should be conducted to assure that the system functions within designated limits for the desired operational

intervals. Two tests should be conducted for each type of ground alignment which the system will be expected to perform (Gyrocompass, Stored Heading, Enhanced-Interrupted, Best Available True Heading, etc.). Data from the static tests will be used to establish baseline performance for the navigation system.

6.5.3 *Random Vibration Tests*

The objectives of these tests are to evaluate the effects of endurance level vibration and to determine if performance is satisfactory following the prescribed vibration. The system should be mounted with a north heading on a vertical exciter, at ambient temperature. A normal gyrocompass alignment will be performed after which the system is switched to the navigate mode. After five minutes of static navigation, random vibration will be applied to the Z-axis for 1 hour. The acceleration spectral density will be 0.001 g-squared/Hz from 10 to 1000 Hz, with a 6 dB/octave roll off to 2000 Hz, 1.22 g rms (see figure 1). When the vibration input is removed, static navigation should continue for an additional hour. One test should be conducted.

6.5.4 *Electrical Power Variation Tests*

These tests should be conducted to assure that the navigation system will function under power conditions expected in a flight environment. These tests consist of six parts for a system powered by AC or two for a system powered by DC. One test should be conducted for each part. The voltage and frequency levels specified in these tests are typical of levels used to test navigation systems. However, these levels should be interpreted with respect to the specifications of the particular system to be tested.

a. Normal and Emergency AC Power, Power Factor, and Power Consumption. With the system in the navigate mode, the three power phases are simultaneously varied to 113, 104, 113, 108, 113, 119, 113, 122, and 113 volts AC, remaining at each voltage level for 10 minutes. The individual phase power factor and power consumption are to be recorded.

b. AC Frequency Variation – Navigation Mode. While operating in the navigation mode, the power frequency is varied to 400, 420, 400, 380, and 400 Hz, remaining at each frequency for 10 minutes.

c. Three Phase Voltage Transients – Navigation Mode. The system is operated in the navigation mode for 42 minutes after which simultaneous three-phase voltage transients are applied at 10-minute intervals as follows:

Volts (RMS)			Duration (Seconds)
122	132	122	2.0
122	152	122	0.8
122	180	122	0.1
104	0	104	2.0

The system is operated in the navigation mode for an additional 30 minutes following the transients.

d. AC Power Variation – Alignment Mode. The voltage is changed to 420 Hz, 118 VAC during the alignment mode. After switching to the navigation mode, the system is allowed to navigate on normal power for 90 minutes.

e. AC Power Variation – Alignment Mode. The voltage is changed to 380 Hz, 108 VAC during the alignment mode. After switching to the navigation mode, the system is allowed to navigate for 90 minutes.

f. AC Voltage Transients – Alignment Mode. With the AC voltage and frequency set to 115 VAC, 400 Hz, the transients of part "c" are imposed during the alignment mode at 1-minute intervals. After switching to the navigation mode, the system is allowed to navigate for 90 minutes.

g. DC Power Variation – Alignment Mode. The system should be mounted with a north heading and gyrocompass aligned with the DC power level at 10 percent below nominal. Following alignment, the system should be allowed to navigate for 1 hour. This test should be repeated with the DC power level at 10 per cent above nominal.

h. DC Power Variation – Navigation Mode. The system should be mounted with a north heading and gyrocompass aligned. The following input voltages with specified durations should be applied to the system during the navigation segment, starting 1 minute after the completion of alignment. One test should be performed with a total navigation time of 67 minutes. Nominal power is 28 volts.

Volts	Duration (minutes)
28	1
24	10
28	1
25	10
28	1
26	10
28	1
27	10
28	1
29	10
28	1
30	10
28	1

For testing systems whose nominal power differs from 28 volts, the table above should be used as a guideline to vary the voltages by similar percentages.

6.5.5 *Electromagnetic Interference (EMI) Tests*

Electromagnetic Interference testing will be conducted to measure the effects of conducted or radiated radio frequency waves on system performance. All tests should be conducted in an anechoic chamber. Four different types of EMI tests should be conducted. Additional details regarding these tests are contained in MIL–STD–461A, Notice 3 and MIL–STD–462, Notice 2, test methods CS06, RS02, and RS03. The voltage and frequency levels specified for the EMI tests are typical of levels used to test navigation systems. However, these levels should be interpreted

with respect to the specifications of the particular system to be tested.

a. Radiated Susceptibility-Magnetic Induction Field Test. These tests consist of two parts. In part one, a wire is wrapped around the cables connecting the Inertial Navigation Unit (INU), the data acquisition system, and the control-display unit. This wire is wrapped around the cables spiraling at two turns per meter and running the length of the cables to within 15 centimeters of each end connector. The wire is powered by a 400 Hz generator and terminated with a 20 amp load. The 400 Hz is applied to the wire-wrapped harness during both the alignment and navigation segments. With the same configuration, 400 volts AC at 10 pulses-per-second are applied; positive pulses starting at the beginning of alignment, then switching to negative pulses for the first 45 minutes of the navigation segment. The system is then allowed to navigate for an additional 45 minutes. For the second part of the test, the wire is wrapped around the system itself. The 400 Hz and 20 amp load and the 400 volt pulses are again applied as stated above. A total of four tests should be conducted.

b. Conducted Susceptibility, Spike, Power Test. Two tests should be conducted. In the first test, positive 100 volt, 1 kHz pulse modulated pulses of 10 microsecond duration should be applied to the three input phase voltages by way of pigtails installed parallel to the three power lines. Positive pulses are applied during the alignment segment, and negative pulses are applied during the first 15 minutes of the navigation segment. The system is then allowed to navigate for an additional 75 minutes. For the second test, pulses are applied to the phase A line first, then to phases B and C.

c. Radiated Susceptibility, Electric Field, 14kHz to 10 GHz. The system is installed on the ground plane in the anechoic chamber with the connector end facing the radiating antenna which is placed one meter from the system in a vertical orientation. After entering the navigation mode, the electric field intensity is brought up to 20 volts/meter. The frequency sweep is then initiated from low to high in the given bandwidth. The antenna is then rotated 90 degrees to the horizontal plane and the sweep is repeated. The sweep is computer controlled to produce discrete frequency steps. As frequency is increased, the step width is also increased. Time at each step is dependent upon the radio frequency generator stabilizing at that frequency as well as allowing for the power level to adjust. The nominal duration is two seconds per step. The nine tests which should be conducted are detailed in the following table.

Frequency	Field	Modulation
10 kHz–25 MHz	Vertical	AM
25 – 200 MHz	Vertical	Pulse
25 – 200 MHz	Horizontal	Pulse
1 – 2 GHz	Horizontal–Vertical	Pulse
2 – 4 GHz	Vertical–Horizontal	Pulse
4 – 8 GHz	Horizontal–Vertical	Pulse
8 – 10 GHz	Vertical–Horizontal	Pulse
500 MHz – 1 GHz	Horizontal	Pulse
200 – 500 GHz	Vertical–Horizontal	Pulse

d. Transient Impulse (Relay Chatter) Test. A wire from the relay test box is located parallel to the system cable bundle, then attached to three sides of the INU before returning to the relay. Two double throw switches are used to provide four combinations of current flow conditions. One set of conditions provides forward and reverse current to the relay coil, and the second set provides forward and reverse outputs of the relay. Each combination is used in sequence for 5 minutes and is initiated 1 minute after entering the navigation mode. The system is allowed to navigate for an additional 70 minutes. The four combinations provided by the switches are:

Fore	Fore
Aft	Fore
Aft	Aft
Fore	Aft

6.5.6 Combined Environments Test

The purpose of the combined environments test is to determine if the system performs satisfactorily while operating in a simulated service environment. Three types of environments will be simulated; arctic, desert, and tropic. Altitude, temperature, relative humidity, and vibration will be varied while the system is operating in the navigational mode following a precision alignment. Only the tropic tests require precise humidity control. Each test cycle should consist of a 65-minute temperature soak, a simulated flight, a 15-minute system-off period, and another simulated flight for a total of 480 minutes. These tests yield 40

hours of simulated flight (10 simulated flights) as indicated in the table below.

Cycle	Temperature Soak (deg C)	Type Cycle	Relative Humidity
1	71	Desert	0 – 50%
2	50	Desert	0 – 50%
3	50	Tropic	91–100%
4	–40	Arctic	0 – 50%
5	4	Arctic	0 – 50%

For each simulated flight, the system will be precision aligned and placed in the navigation mode. The profiles for temperature, altitude, relative humidity (tropic only), and vibration are shown in figures 2 and 3. A tabulated description of the profiles is provided in table 1. For each test, the system shall be provided with cooling air at laboratory room ambient temperature (23 plus/minus 10 degrees C). Cooling air will be turned on just prior to system turn-on and turned off immediately after the system is turned off.

6.5.7 *Moving Base Alignment*
This test will be performed to measure the effects of base motion during gyrocompass alignment and the resulting effects on navigation accuracy. The system will be mounted on a single axis vibration table, temperature stabilized for 2 hours, and gyrocompass aligned with a north heading. While in the alignment mode, the table will be driven in the east-west direction at a 1 Hz, 2.54 cm peak-to-peak (equivalent to 0.05 g (zero-to-peak) rate. An additional motion, a 2 cm step in 0.5 seconds, will be added 5 seconds prior to the completion of alignment. At the completion of the alignment, the system will be placed in the navigation mode for a period of 60 minutes. This test will be repeated with an initial alignment heading of 180 degrees.
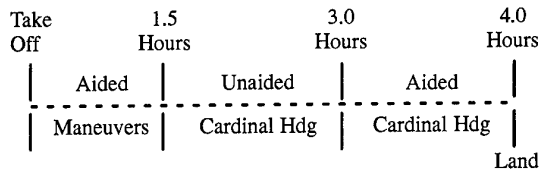
6.6 Flight Tests

6.6.1 *Discussion*
The goal of flight testing is to subject the test navigation system to the type of flight environment which will be experienced when the system becomes operational. This involves testing the system in the various aircraft (cargo, fighter, helicopter) in which the system is designed to be flown and in a manner which will provide a meaningful indication of the system's performance capabilities in both benign and dynamic flight environments. The flight profiles and procedures detailed in the next four sections are designed to determine the system's performance while being subjected to realistic operational conditions.

6.6.2 *Cargo Aircraft Flight Tests*
The purpose of these tests is to determine navigation system performance in both the aided and unaided modes of operation during benign and dynamic flights. To satisfy the objectives, a flight profile composed of medium (less than 5 g's) and low dynamic (less than 2 g's) maneuvers will be used. This profile will include maneuvers such as tight turns with changes in altitude and benign straight-and-level flying. This flight test profile will begin with a low dynamic racetrack type pattern, establishing a baseline for maneuvers. This will be followed by a medium dynamic profile designed to test the performance of the integrated system. Next will be a cardinal heading profile with the system operated in the unaided mode. Finally, a return cardinal heading will be flown in the fully aided mode, testing the recovery capabilities of the system. The chart below illustrates the timeline for this cargo flight profile. Table 2 along with figure 4 are provided as a description of a sample maneuver profile to be used with these tests.

```
Take          1.5              3.0              4.0
Off          Hours            Hours            Hours
 |    Aided    |   Unaided    |    Aided    |
 - - - - - - - - - - - - - - - - - - - - - - - -
 |  Maneuvers  | Cardinal Hdg | Cardinal Hdg |
                                           Land
```

A minimum of six flights should be flown to achieve the desired level of statistical confidence in the estimate of the system's performance. References 3 and 4 present detailed discussions regarding the required number of flights.

Since the position and velocity errors may be non-zero at the beginning of the unaided segment, it would be necessary to compensate these parameters during the entire unaided interval by the amount of the error present at the beginning of the interval. Only by performing this compensation will a true picture of the system's actual performance during this segment be obtained.

The system's ability to perform in-flight alignments (IFA) should also be tested. A total of four flights should be flown; two benign flights and two maneuver flights. Each flight will consist of three IFA's, each followed by 30 minutes of navigation to verify the validity of the IFA. The decision regarding the operational mode (aided or unaided) of the system should be determined by the test objectives. The IFA's for the benign flights and for the maneuver flights should be divided into separate ensembles.

6.6.3 *Helicopter Flight Tests*
The purpose of this series of tests is to determine the navigation system performance in the fully aided mode of operation during benign and maneuver flights with the system subjected to the vibration environment of the helicopter. To satisfy this objective, two different flight profiles will be flown.

The first profile is a cardinal heading cruise profile. This profile consists of a cardinal heading flight at 3,000 feet above-ground-level (AGL) for a duration of 90 minutes.

The second profile is the tactical maneuver flight profile and consists of a clockwise, 360 degree turn at takeoff followed by a 20-minute cardinal heading cruise leg, 45 minutes of maneuvering, and a 25-minute return leg. Specific tactical maneuvers performed are a rapid spiraling descent (3,000 feet to 500 feet AGL with two 360 degree turns), a counterclockwise pinwheel turn, a series of four S-turns (45-degree bank with 60 degree heading changes), a series of two shallow S-turns (30-degree bank with 30-degree heading changes) followed by two vertical S-turns (100 feet up, 200 feet down, 100 feet up), terrain following, and a high performance climb-out to 3,000 feet AGL. Six flights of each profile should be flown.

6.6.4 *Fighter Flight Tests*

The purpose of fighter aircraft flight testing is to determine navigation system performance in the fully aided mode during benign and high dynamic flights. The fighter test bed may vary with test requirements, but in most cases, a high performance fighter aircraft should be used as the test bed.

Three basic types of flight profiles should be flown during fighter testing: benign flight, ordnance delivery maneuvers, and air combat maneuvers. Each flight will be 90 minutes in duration, with any maneuvers conducted at the end of the initial benign leg. Six flights will consist of straight-and-level, out-and-back benign profiles, six will include air combat maneuvers, and six will include ordnance delivery maneuvers. The maneuver flights will consist of a 40-minute benign leg followed by 10 minutes of maneuvers, then a 40-minute benign return leg. The profiles, with the exception of the maneuver segments, should be flown at an altitude of 35,000 feet MSL. Throughout the fighter testing, the pilot should attempt to conduct the flights in a repeatable manner, noting deviations from the intended profiles as they occur. Entry headings, altitudes, and air speed should be consistent for all maneuvers whenever possible.

6.6.5 *Terrain Referenced Navigation Flight Tests*

The purpose of these tests is to evaluate the accuracy of the terrain referenced navigation (TRN) system while flying over various terrain types at different altitudes. The TRN system produces trajectory data by comparing the measured terrain profile from the radar altimeter with the digital terrain data base. The performance accuracy of the TRN system is determined by comparing its trajectory data with time-correlated reference data.

In order to adequately characterize the performance of a TRN system, it should be flown over flat terrain, gently rolling hills, moderately rough terrain, and rough terrain. Additionally, the test flights should be flown at both 1,000 and 2,000 feet above ground level (AGL) due to the altitude-sensitive nature of radar altimeters. Six flights for each ensemble should be flown. The duration of these flights will be determined by the type of terrain in the immediate area of the testing headquarters.

In addition to determining the navigation accuracy of the TRN system, other items which need to be determined are as follows:

a. How long does it take the system to acquire the proper track?

b. How vulnerable is the system to acquiring a false track?

c. How effectively does the system recover from tracking errors?

d. How well does the system perform in the coasting mode over flat terrain?

6.7 Data Reduction Procedures

6.7.1 *Generation of Error Outputs*

Time-correlated system test data and reference data should be recorded during each test. Data rates are prescribed in sections B and C. Errors are calculated as a function of time by comparing system data with reference data. Error outputs are generated by linearly interpolating the reference data to system times and comparing the interpolated reference data to system data. Latitude, longitude, and altitude position errors are output as well as north, east, and vertical velocity errors. The horizontal position and velocity errors are root-sum-squared (RSS) to produce radial position and velocity errors. Provisions should be made for suspending the comparison process during intervals where the reference system data are of marginal value. In these cases, gaps in position and velocity error data will occur.

6.7.2 *Plots to be Presented*

It is useful for comparison purposes to represent the error plots of several tests with a single plot. This is accomplished in one of two ways. The first type is obtained by presenting the radial error curves from an ensemble of tests on a composite plot. These ensembles contain only those tests which come from similar test profiles. The statistical analysis of these ensembles will reflect the system's overall performance and will establish certain confidence limits with respect to that performance. This analysis should output time histories of the mean, standard deviation, median, 50th, and 95th percentiles of the radial position and velocity errors. Plots of any or all of these parameters may be presented as an indication of system performance. For details regarding statistical methods, refer to section 6.9.

6.7.3 *Tabular Data to be Presented*

For the purpose of comparing one system against another, it is useful to specify a system's performance with a single number. For inertial navigation systems, where level-axis gyro drifts generating ramps in position error have been the dominant error source, the straight-line/zero-intercept fit to the radial position error has been the most popular performance parameter. However, with the advent of hybrid navigation systems, which produce errors that are largely time-invariant, this is no longer the case. It is much more meaningful to present the mean and standard deviation of the 50th and 95th percentiles of radial position and velocity errors for an ensemble of flights. These values should be presented as indications of the position accuracy of a hybrid navigation system. To obtain a measure of velocity accuracy, the radial velocity error data for all flights in an ensemble should be merged, then the composite RMS computed for the radial velocity error.

6.8 Statistical Methods

6.8.1 *Introduction*

For the purpose of comparing one navigation system with another, it is highly desirable to specify a system's performance in a simple, straightforward manner, preferably with a single quantity stated as a function of time or distance. However, there are a variety of methods for describing the performance accuracy of a navigation system as well as numerous assumptions which can be made regarding a system's error characteristics. These dissimilarities may introduce inconsistencies in comparing the performance of one navigation system with another. The goal of this document is to standardize the method of presenting navigation system accuracies.

6.8.2 Definitions

The statistical parameters most frequently used to quantify a system's radial position and velocity errors are mean, standard deviation, root-mean-square (RMS), geometric mean (GM), 50th percentile (Circular Error Probable, CEP), and the 95th percentile. The mean, standard deviation, RMS, and GM are computed using simple algorithms. These are defined as follows:

$$Mean \equiv \overline{R} = \frac{\displaystyle\sum_{i=1}^{N} R_i}{N} \tag{1}$$

$$Std\ Dev \equiv \sqrt{\frac{\displaystyle\sum_{i=1}^{N} (R_i - \overline{R})^2}{N}} \tag{2}$$

$$RMS \equiv \frac{\sqrt{\displaystyle\sum_{i=1}^{N} R_i^2}}{N} \tag{3}$$

$$GM \equiv \sqrt[N]{\prod_{i=1}^{N} R_i} \tag{4}$$

where R_i is the radial error at time t_i and N is the number of data points.

The 50th percentile (R50) and 95th percentile (R95) are defined respectively as being 50 or 95 percent probable that the error in a randomly selected flight from an ensemble of similar flights will fall below this level, or that the error will be below this level in 50 or 95 percent of flights if the number of flights in the ensemble is very large.

6.8.3 Discussion

The process of determining the accuracy of a navigation system involves the conducting of flight tests. However, due to fiscal and time constraints, it is rarely ever possible to conduct a large enough sample of tests to completely characterize a system's performance. Because of this limitation, a method has been developed which makes it possible to extrapolate from a limited number of samples of system performance to an estimation of system performance for the parent population. This method involves the ratio of the GM and the RMS values of the radial position error for an ensemble of tests. The GM/RMS algorithm was developed at Holloman Air Force Base, New Mexico, and is widely used by facilities involved in the testing of navigation systems. Derivation of this method is detailed in reference 1, and further explanation can be found in references 2 and 3.

From the radial errors computed at each time point in an ensemble of valid flights, calculate the RMS value and the GM of the errors as defined in section 2 above. Form the ratio GM/RMS. Also, form the ratio R(P)/RMS, where R(P) is the desired percentile level, usually 50 or 95. The relationship between these two ratios is a group of functions, one for each percentile level P. These functions are independent of the actual test data and can be precomputed. These functions are presented in figure 5. Using the GM/RMS value along the ordinate axis, determine the multiplier corresponding to the desired percentile P, and multiply this value by the RMS sample value. The results will be the Pth percentile level of error at time t for this ensemble of flights. The quantity t represents the time elapsed since switching to the navigate mode or from the time of initiation of "aiding."

Once the Pth percentile has been computed for all available data points, this curve is used to determine system performance. For systems whose radial position errors are functions of time or distance, the most commonly used method of determining system performance has been to perform a least-squares fit (constrained to pass through the origin) to the Pth percentile curve of the radial position error. This produces the radial position error growth rate. For systems whose radial position errors are not functions of time or distance, compute the Mean and Standard Deviation of the Pth percentile curve. These values will then be used to characterize the performance of a hybrid navigation system.

413
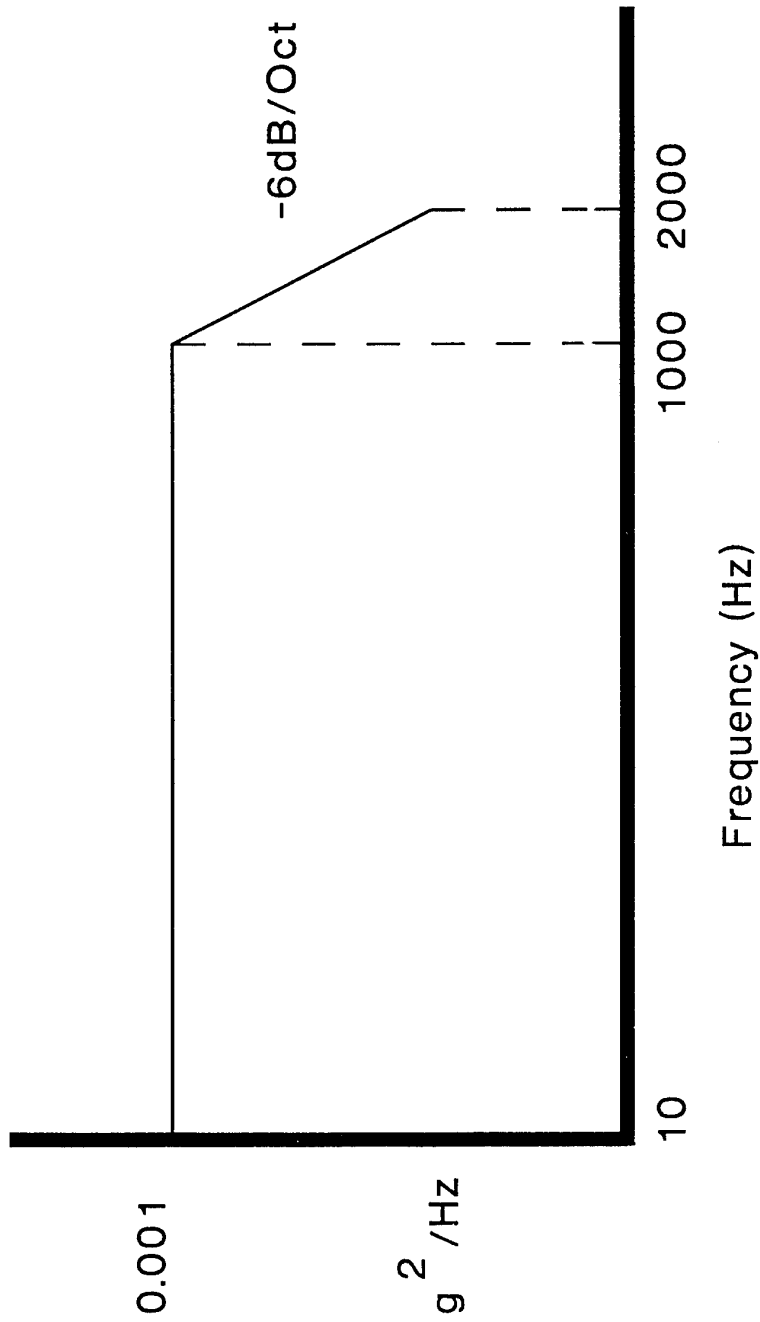
Figure 1. Random Vibration Levels.
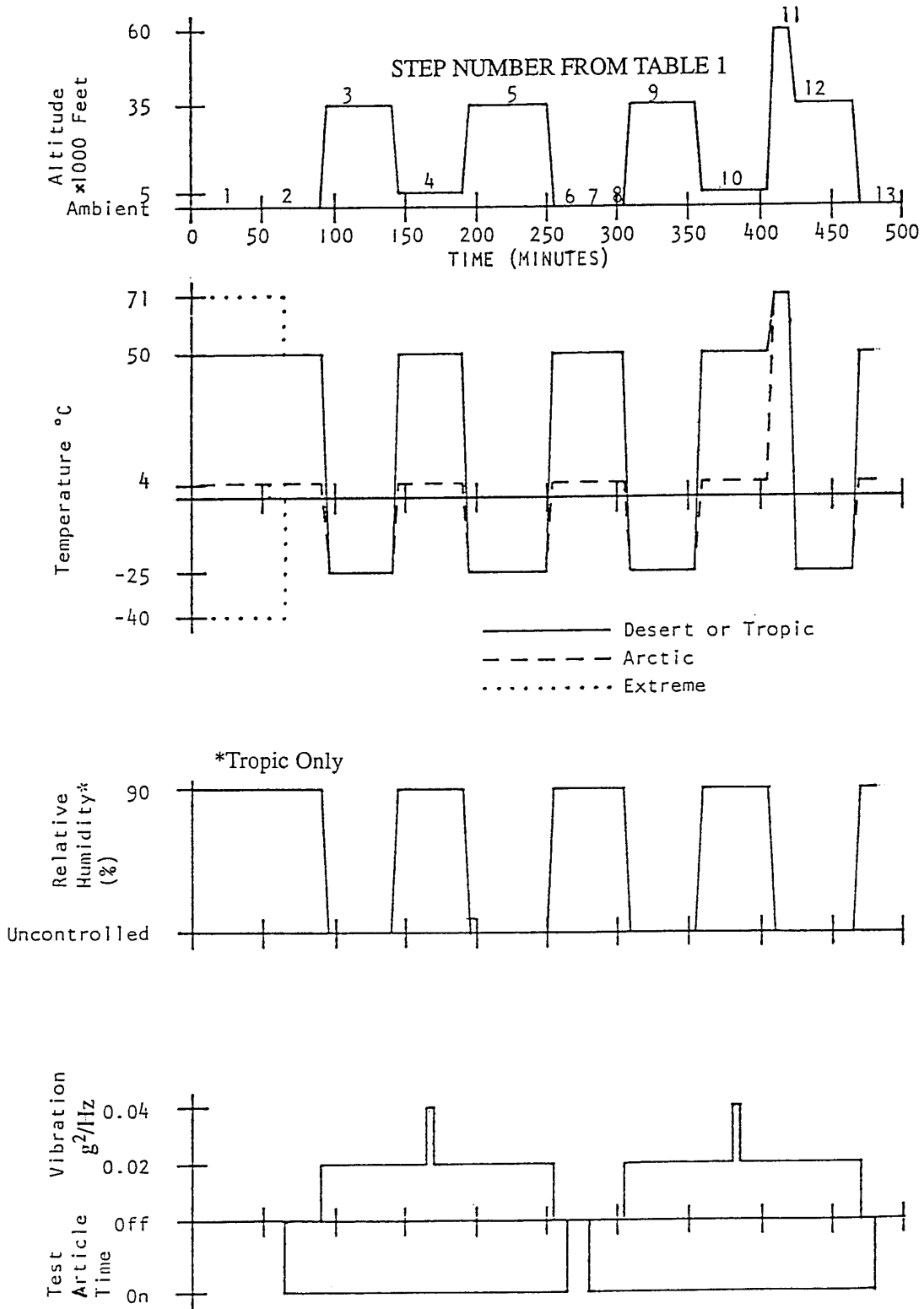
Figure 2. Combined Environments Test Profile.

POWER SPECTRAL DENSITY g^2/Hz

.04

.02

②

①

-6 dB/OCT

-6 dB/OCT

15   1000   2000

Frequency (Hz)
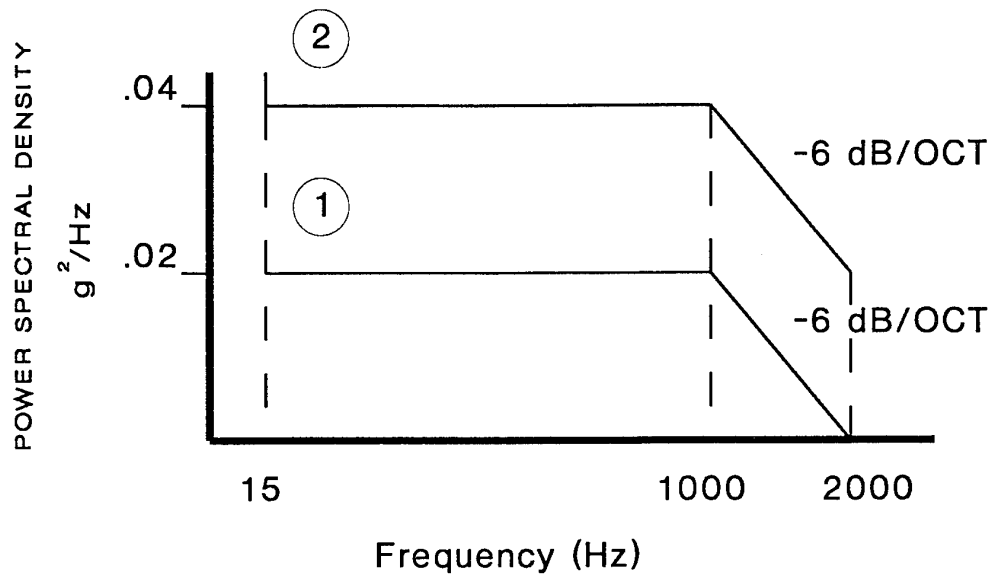
① This level is to be run during all vibration-on times except for five minutes out of each 5,000 ft. altitude segment of the flight profile.

② This level is to be run for five minutes during each 5,000 ft. altitude segment of the flight profile (total test time is ten minutes for each 8-hour cycle.

Figure 3. Random Vibration Levels for CET.

Maneuver Checkpoint Locations

Checkpoint	Longitude	Latitude
1	-106 15.8	33 13.8
2	-106 05.6	33 13.8
3	-106 05.6	33 00.0
4	-106 15.8	33 00.0
5	-106 09.8	33 13.8
6	-106 09.8	33 00.0
7	-106 12.0	33 12.3
8	-106 12.3	33 08.5
9	-106 12.0	33 05.2
10	-106 11.5	33 02.0
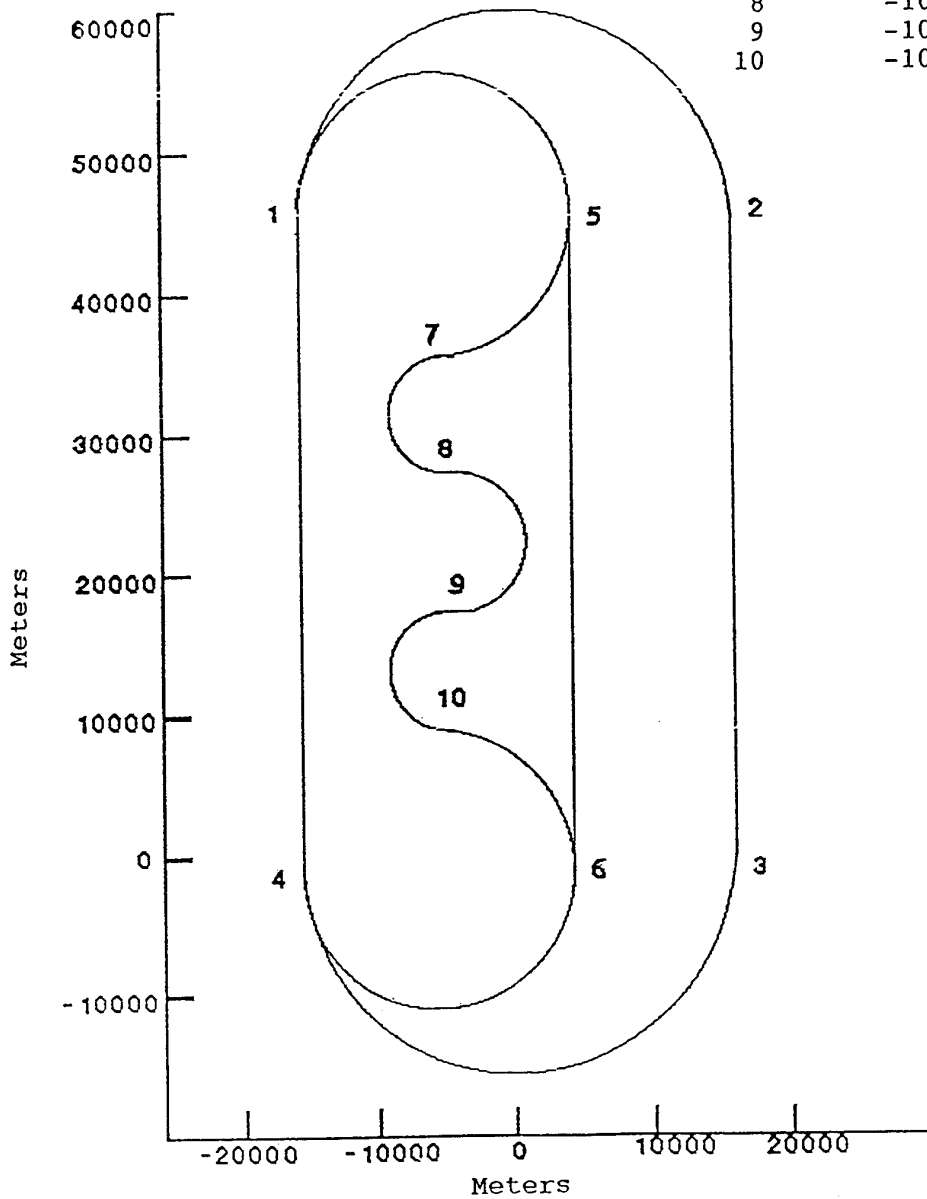
Figure 4.   Cargo Maneuver Profile
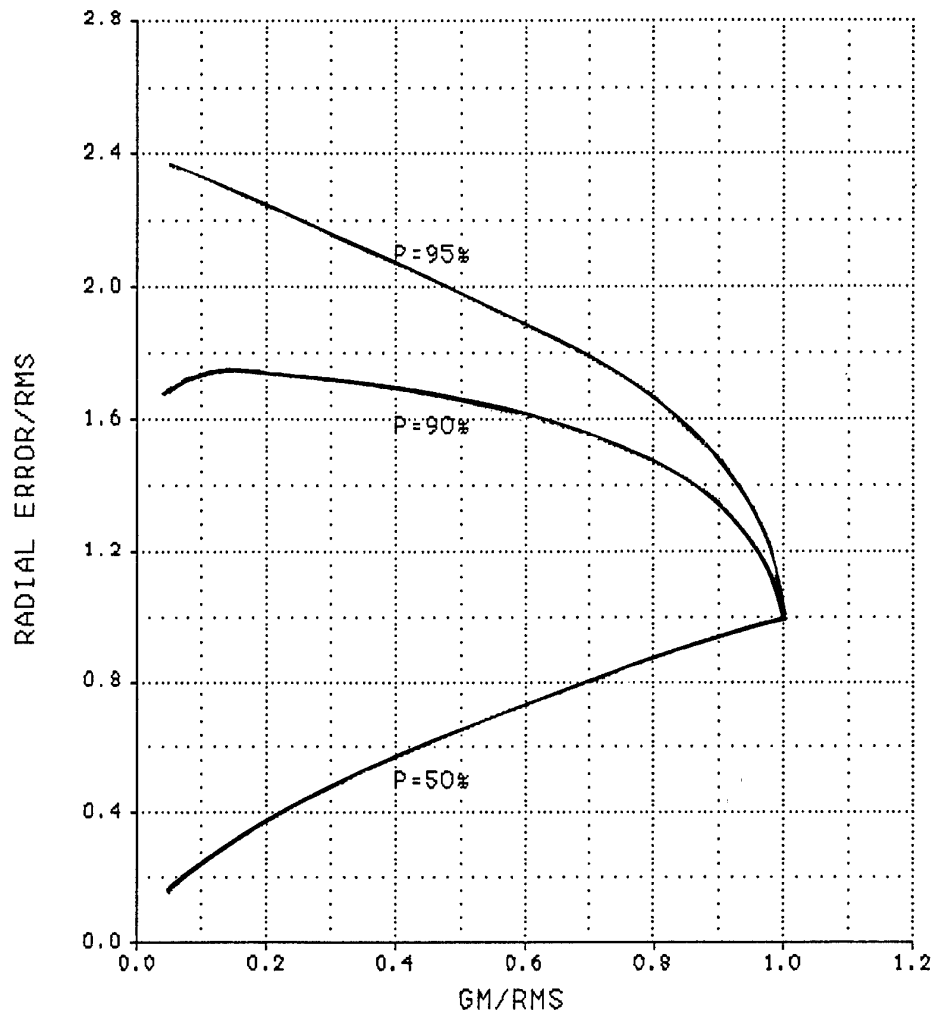
PERCENTILES OF THE MAXIMUM LIKELIHOOD
ESTIMATE OF THE DISTRIBUTION OF
RADIAL ERROR

Table 1. Combined Environments Test Profiles.

	Step Number	Time Interval (min)	Altitude (x1000 ft.)	Desert Temp (deg C)	Desert Humidity (% RH)	Arctic Temp (deg C)	Arctic Humidity (% RH)	Tropic Temp (deg C)	Tropic Humidity (% RH)	Vibration (g²/Hz)
Soak Temp	1A	0 to 65	ambient	71	*	-40	*	—	—	0
	1B	0 to 65	ambient	50	*	4	*	50	>90	0
Simulated Flight	2	65 to 90	ambient	50	*	4	*	50	>90	0
	3	95 to 140	35	-25	*	-25	*	-25	*	0.02
	4	145 to 190	5	50	*	4	*	50	>90	.02 to .04¹
	5	195 to 250	35	-25	*	-25	*	-25	*	0.02
	6	255 to 265	ambient	50	*	4	*	50	>90	0
System off	7	265 to 280	ambient	50	*	4	*	50	>90	0
Simulated Flight	8	280 to 305	ambient	50	*	4	*	50	>90	0
	9	310 to 355	35	-25	*	-25	*	-25	*	0.02
	10	360 to 405	5	50	*	4	*	50	>90	0.02 to 0.04¹
	11	410 to 420	60	70	*	70	*	70	*	0.02
	12	425 to 465	35	-25	*	-25	*	-25	*	0.02
	13	470 to 480	ambient	50	*	4	*	50	*	0

* Uncontrolled

¹ See footnotes on Figure 3.

Table 2. Aircraft Controller Data.

TRK POS	TRK DEG	ALT MSL	LONGITUDE	LATITUDE	KTAS	MANEUVER
1A	0	150	−106 15.8	33 13.8	350	20 DEG BANK RIGHT
2A	180	150	−106 05.6	33 13.8	350	END TURN
3A	180	150	−106 05.6	33 00.0	350	20 DEG BANK RIGHT
4A	0	150	−106 15.8	33 00.0	350	END TURN
1B	0	150	−106 15.8	33 13.8	350	32 DEG BANK RIGHT
5A	180	150	−106 09.8	33 13.8	350	END TURN
6A	180	150	−106 09.8	33 00.0	350	32 DEG BANK RIGHT
4B	0	150	−106 15.8	33 00.0	350	END TURN, SLOW TO 325 KTAS
1C	0	150	−106 15.8	33 13.8	325	28 DEG BANK RIGHT
5B	180	150	−106 09.8	33 13.8	325	40 DEG BANK RIGHT BEGIN CLIMB TO FL 170 BEGIN SLOW TO 275 KTAS
7A	270	150 '	−106 12.0	33 12.3	325	*40 DEG BANK LEFT
8A	90	160@	−106 12.3	33 08.5	300	*40 DEG BANK RIGHT
9A	270	170	−106 12.0	33 05.2	275	40 DEG BANK LEFT BEGIN DIVE TO FL 150 ACCELERATE TO 350 KTAS
10A	90	160@	−106 11.5	33 02.0	300	*40 DEG BANK RIGHT
6B	180	150	−106 09.8	33 00.0	325	*28 DEG BANK RIGHT
4C	0	150	−106 15.8	33 13.2	350	END TURN BEGIN CLIMB TO FL 200
1D	0	160@	−106 15.8	33 13.8	350	20 DEG BANK RIGHT
2B	180	170@	−106 05.6	33 13.8	350	END TURN
3B	180	180@	−106 05.6	33 00.0	350	20 DEG BANK RIGHT
4D	0	190@	−106 15.8	33 00.0	350	END TURN
1E	0	200	−106 15.8	33 13.8	350	32 DEG BANK RIGHT
5C	180	200	−106 09.8	33 13.8	350	END TURN
6C	180	200	−106 09.8	33 00.0	350	32 DEG BANK RIGHT
4E	0	200	−106 15.8	33 00.0	350	END TURN, SLOW TO 325 KTAS
1F	0	200	−106 15.8	33 13.8	325	28 DEG BANK RIGHT
5D	180	200	−106 09.8	33 13.8	325	40 DEG BANK RIGHT BEGIN CLIMB TO FL 220 BEGIN SLOW TO 275 KTAS
7B	270	200	−106 12.0	33 12.3	325	*40 DEG BANK LEFT
8B	90	210@	−106 12.3	33 08.5	300	*40 DEG BANK RIGHT
9B	270	220	−106 12.0	33 05.2	275	40 DEG BANK LEFT BEGIN DIVE TO FL 200 ACCELERATE TO 350 KTAS

Table 2. Aircraft Controller Data (Continued).

TRK POS	TRK DEG	ALT MSL	LONGITUDE	LATITUDE	KTAS	MANEUVER
10B	90	210@	−106 11.5	33 02.0	300	*40 DEG BANK RIGHT
6D	180	200	−106 09.8	33 00.0	325	*28 DEG BANK·RIGHT
4E	0	200	−106 15.8	33 13.2	350	END TURN BEGIN CLIMB TO FL 250
1G	0	210@	−106 15.8	33 13.8	350	20 DEG BANK RIGHT
2C	180	220@	−106 05.6	33 13.8	350	END TURN
3C	180	230@	−106 05.6	33 00.0	350	20 DEG BANK RIGHT
4F	0	240@	−106 15.8	33 00.0	350	END TURN
1H	0	250	−106 15.8	33 13.8	350	32 DEG BANK RIGHT
5E	180	250	−106 09.8	33 13.8	350	END TURN
6E	180	250	−106 09.8	33 00.0	350	32 DEG BANK RIGHT
4G	0	250	−106 15.8	33 00.0	350	END TURN, SLOW TO 325 KTAS
1I	0	250	−106 15.8	33 13.8	325	28 DEG BANK RIGHT
5F	180	250	−106 09.8	33 13.8	325	40 DEG BANK RIGHT BEGIN CLIMB TO FL 270 BEGIN SLOW TO 275 KTAS
7C	270	250	−106 12.0	33 12.3	325	*40 DEG BANK LEFT
8C	90	260@	−106 12.3	33 08.5	300	*40 DEG BANK RIGHT
9C	270	270	−106 12.0	33 05.2	275	40 DEG BANK LEFT BEGIN DIVE TO FL 250 ACCELERATE TO 350 KTAS
10C	90	260@	−106 11.5	33 C2.0	300	*40 DEG BANK RIGHT
6F	180	250	−106 09.8	33 00.0	325	*28 DEG BANK RIGHT
4H	0	250	−106 15.8	33 13.2	350	END TURN

@ Indicates approximate altitude MSL

* Indicates approximate true air speed

6.9 References

1. Mason, F.J. and Bodwell, C.A. "Maximum Likelihood Estimation of the Distribution of Radial Errors," Directorate of Guidance Test, Air Force Missile Development Center, Holloman Air Force Base, New Mexico, Working Paper WP–MDSGA–65–4, Rev 1, 1 Oct 1969, DTIC AD–A950 316.

2. Air STD 53/16A, "The Specification for Evaluation of the Accuracy of Hybrid Navigation Systems," Air Standardization Coordinating Committee, 6 April 1988.

3. Sakran, F.C. Jr., "Considerations in Use of the GM/RMS Estimation Method For Navigation System Position Error," Force Warfare Aircraft Test Directorate, Patuxent River, Maryland, Technical Memorandum TM 91–03 FW, 17 July 1991.

4. "Calculation of Confidence Intervals for Percentile Levels," Memo for The Record dtd 20 May 1983, Holloman Air Force Base, New Mexico.

REPORT DOCUMENTATION PAGE

1. Recipient's Reference	2. Originator's Reference	3. Further Reference	4. Security Classification of Document
	AGARD-AG-331	ISBN 92-836-1018-0	UNCLASSIFIED/ UNLIMITED

5. Originator Advisory Group for Aerospace Research and Development
North Atlantic Treaty Organization
7 rue Ancelle, 92200 Neuilly-sur-Seine, France

6. Title
Aerospace Navigation Systems

7. Presented at/sponsored by
the Guidance and Control Panel of AGARD

8. Author(s)/Editor(s)	9. Date
Various	June 1995

10. Author's/Editor's Address	11. Pages
Various	432

12. Distribution Statement There are no restrictions on the distribution of this document.
Information about the availability of this and other AGARD
unclassified publications is given on the back cover.

13. Keywords/Descriptors

Air navigation	Space navigation
Inertial navigation	Doppler navigation
Satellite navigation	Magnetic navigation
Radar navigation	Global positioning system
Multisensors	Integrated systems
Rotary wing aircraft	Fixed wing aircraft
Spacecraft	Kalman filtering

14. Abstract

The need for an up to date, comprehensive treatise on aerospace navigation systems has been recognized by the navigation community. Such a document will serve as a lasting reference as did previously published material prepared in the late 60's and early 70's. These earlier documents, though remarkably prescient in their technical forecasts, have been overtaken by very rapidly advancing technology.

It is anticipated that the target reader of this AGARDograph will be an individual who has responsibility for the integration of navigation equipment aboard an aerospace vehicle. He may not have, but is desirous of obtaining, an expert's perspective on the capabilities and limitations of the various navigation sensor and integration techniques.

The AGARDograph is organized into six sections with the initial section providing the motivation for establishing the requirements to assure that the development of an aerospace navigation system will meet its operational requirement. The second section of this document reviews navigation coordinate frames with a discussion of inertial, terrestrial and geodetic coordinate systems. The third section of the document describes the navigation sensor technologies that are employed in modern aerospace navigation systems.

The fourth section of this document addresses system analysis and synthesis methods. Examination and trade-off of the technology, for each sensing, processing and control and display element, generates alternative preliminary designs. The fifth section of the document describes representative state-of-the-art navigation system implementations in fixed wing, rotary wing aircraft and spacecraft. The sixth and concluding section of the document describes the various test methods employed to verify the performance of aerospace navigation systems used by NATO countries.

AGARD

NATO ⊕ OTAN

7 RUE ANCELLE • 92200 NEUILLY-SUR-SEINE

FRANCE

Télécopie (1)47.38.57.99 • Télex 610 176

DIFFUSION DES PUBLICATIONS

AGARD NON CLASSIFIEES

Aucun stock de publications n'a existé à AGARD. A partir de 1993, AGARD détiendra un stock limité des publications associées aux cycles de conférences et cours spéciaux ainsi que les AGARDographies et les rapports des groupes de travail, organisés et publiés à partir de 1993 inclus. Les demandes de renseignements doivent être adressées à AGARD par lettre ou par fax à l'adresse indiquée ci-dessus. *Veuillez ne pas téléphoner.* La diffusion initiale de toutes les publications de l'AGARD est effectuée auprès des pays membres de l'OTAN par l'intermédiaire des centres de distribution nationaux indiqués ci-dessous. Des exemplaires supplémentaires peuvent parfois être obtenus auprès de ces centres (à l'exception des Etats-Unis). Si vous souhaitez recevoir toutes les publications de l'AGARD, ou simplement celles qui concernent certains Panels, vous pouvez demander à être inclu sur la liste d'envoi de l'un de ces centres. Les publications de l'AGARD sont en vente auprès des agences indiquées ci-dessous, sous forme de photocopie ou de microfiche.

CENTRES DE DIFFUSION NATIONAUX

ALLEMAGNE
    Fachinformationszentrum,
    Karlsruhe
    D-76344 Eggenstein-Leopoldshafen 2
BELGIQUE
    Coordonnateur AGARD-VSL
    Etat-major de la Force aérienne
    Quartier Reine Elisabeth
    Rue d'Evere, 1140 Bruxelles
CANADA
    Directeur, Services d'information scientifique
    Ministère de la Défense nationale
    Ottawa, Ontario K1A 0K2
DANEMARK
    Danish Defence Research Establishment
    Ryvangs Allé 1
    P.O. Box 2715
    DK-2100 Copenhagen Ø
ESPAGNE
    INTA (AGARD Publications)
    Pintor Rosales 34
    28008 Madrid
ETATS-UNIS
    NASA Headquarters
    Code JOB-1
    Washington, D.C. 20546
FRANCE
    O.N.E.R.A. (Direction)
    29, Avenue de la Division Leclerc
    92322 Châtillon Cedex
GRECE
    Hellenic Air Force
    Air War College
    Scientific and Technical Library
    Dekelia Air Force Base
    Dekelia, Athens TGA 1010

ISLANDE
    Director of Aviation
    c/o Flugrad
    Reykjavik
ITALIE
    Aeronautica Militare
    Ufficio del Delegato Nazionale all'AGARD
    Aeroporto Pratica di Mare
    00040 Pomezia (Roma)
LUXEMBOURG
    Voir Belgique
NORVEGE
    Norwegian Defence Research Establishment
    Attn: Biblioteket
    P.O. Box 25
    N-2007 Kjeller
PAYS-BAS
    Netherlands Delegation to AGARD
    National Aerospace Laboratory NLR
    P.O. Box 90502
    1006 BM Amsterdam
PORTUGAL
    Força Aérea Portuguesa
    Centro de Documentação e Informação
    Alfragide
    2700 Amadora
ROYAUME-UNI
    Defence Research Information Centre
    Kentigern House
    65 Brown Street
    Glasgow G2 8EX
TURQUIE
    Millî Savunma Başkanliği (MSB)
    ARGE Dairesi Başkanliği (MSB)
    06650 Bakanliklar-Ankara

Le centre de distribution national des Etats-Unis ne détient PAS de stocks des publications de l'AGARD.

D'éventuelles demandes de photocopies doivent être formulées directement auprès du NASA Center for AeroSpace Information (CASI) à l'adresse ci-dessous. Toute notification de changement d'adresse doit être fait également auprès de CASI.

AGENCES DE VENTE

NASA Center for AeroSpace Information (CASI) 800 Elkridge Landing Road Linthicum Heights, MD 21090-2934 Etats-Unis	ESA/Information Retrieval Service European Space Agency 10, rue Mario Nikis 75015 Paris France	The British Library Document Supply Division Boston Spa, Wetherby West Yorkshire LS23 7BQ Royaume-Uni

Les demandes de microfiches ou de photocopies de documents AGARD (y compris les demandes faites auprès du CASI) doivent comporter la dénomination AGARD, ainsi que le numéro de série d'AGARD (par exemple AGARD-AG-315). Des informations analogues, telles que le titre et la date de publication sont souhaitables. Veuiller noter qu'il y a lieu de spécifier AGARD-R-nnn et AGARD-AR-nnn lors de la commande des rapports AGARD et des rapports consultatifs AGARD respectivement. Des références bibliographiques complètes ainsi que des résumés des publications AGARD figurent dans les journaux suivants:

Scientific and Technical Aerospace Reports (STAR)
publié par la NASA Scientific and Technical
Information Division
NASA Headquarters (JTT)
Washington D.C. 20546
Etats-Unis

Government Reports Announcements and Index (GRA&I)
publié par le National Technical Information Service
Springfield
Virginia 22161
Etats-Unis
(accessible également en mode interactif dans la base de
données bibliographiques en ligne du NTIS, et sur CD-ROM)

AGARD

NATO ⊕ OTAN

7 RUE ANCELLE • 92200 NEUILLY-SUR-SEINE

FRANCE

Telefax (1)47.38.57.99 • Telex 610 176

DISTRIBUTION OF UNCLASSIFIED

AGARD PUBLICATIONS

AGARD holds limited quantities of the publications that accompanied Lecture Series and Special Courses held in 1993 or later, and of AGARDographs and Working Group reports published from 1993 onward. For details, write or send a telefax to the address given above.. *Please do not telephone.*

AGARD does not hold stocks of publications that accompanied earlier Lecture Series or Courses or of any other publications. Initial distribution of all AGARD publications is made to NATO nations through the National Distribution Centres listed below. Further copies are sometimes available from these centres (except in the United States). If you have a need to receive all AGARD publications, or just those relating to one or more specific AGARD Panels, they may be willing to include you (or your organisation) on their distribution list. AGARD publications may be purchased from the Sales Agencies listed below, in photocopy or microfiche form.

NATIONAL DISTRIBUTION CENTRES

BELGIUM
Coordonnateur AGARD — VSL
Etat-major de la Force aérienne
Quartier Reine Elisabeth
Rue d'Evere, 1140 Bruxelles

CANADA
Director Scientific Information Services
Dept of National Defence
Ottawa, Ontario K1A 0K2

DENMARK
Danish Defence Research Establishment
Ryvangs Allé 1
P.O. Box 2715
DK-2100 Copenhagen Ø

FRANCE
O.N.E.R.A. (Direction)
29 Avenue de la Division Leclerc
92322 Châtillon Cedex

GERMANY
Fachinformationszentrum
Karlsruhe
D-76344 Eggenstein-Leopoldshafen 2

GREECE
Hellenic Air Force
Air War College
Scientific and Technical Library
Dekelia Air Force Base
Dekelia, Athens TGA 1010

ICELAND
Director of Aviation
c/o Flugrad
Reykjavik

ITALY
Aeronautica Militare
Ufficio del Delegato Nazionale all'AGARD
Aeroporto Pratica di Mare
00040 Pomezia (Roma)

LUXEMBOURG
See Belgium

NETHERLANDS
Netherlands Delegation to AGARD
National Aerospace Laboratory, NLR
P.O. Box 90502
1006 BM Amsterdam

NORWAY
Norwegian Defence Research Establishment
Attn: Biblioteket
P.O. Box 25
N-2007 Kjeller

PORTUGAL
Força Aérea Portuguesa
Centro de Documentação e Informação
Alfragide
2700 Amadora

SPAIN
INTA (AGARD Publications)
Pintor Rosales 34
28008 Madrid

TURKEY
Millî Savunma Başkanliği (MSB)
ARGE Dairesi Başkanliği (MSB)
06650 Bakanliklar-Ankara

UNITED KINGDOM
Defence Research Information Centre
Kentigern House
65 Brown Street
Glasgow G2 8EX

UNITED STATES
NASA Headquarters
Code JOB-1
Washington, D.C. 20546

The United States National Distribution Centre does NOT hold stocks of AGARD publications.
Applications for copies should be made direct to the NASA Center for AeroSpace Information (CASI) at the address below.
Change of address requests should also go to CASI.

SALES AGENCIES

NASA Center for
AeroSpace Information (CASI)
800 Elkridge Landing Road
Linthicum Heights, MD 21090-2934
United States

ESA/Information Retrieval Service
European Space Agency
10, rue Mario Nikis
75015 Paris
France

The British Library
Document Supply Centre
Boston Spa, Wetherby
West Yorkshire LS23 7BQ
United Kingdom

Requests for microfiches or photocopies of AGARD documents (including requests to CASI) should include the word 'AGARD' and the AGARD serial number (for example AGARD-AG-315). Collateral information such as title and publication date is desirable. Note that AGARD Reports and Advisory Reports should be specified as AGARD-R-nnn and AGARD-AR-nnn, respectively. Full bibliographical references and abstracts of AGARD publications are given in the following journals:

Scientific and Technical Aerospace Reports (STAR)
published by NASA Scientific and Technical
Information Division
NASA Headquarters (JTT)
Washington D.C. 20546
United States

Government Reports Announcements and Index (GRA&I)
published by the National Technical Information Service
Springfield
Virginia 22161
United States
(also available online in the NTIS Bibliographic
Database or on CD-ROM)