RTO-MP-43
AC/323(SCI)TP/13

RTO-MP-43

NORTH ATLANTIC TREATY ORGANIZATION

RESEARCH AND TECHNOLOGY ORGANIZATION
BP 25, 7 RUE ANCELLE,
F-92201 NEUILLY-SUR-SEINE CEDEX, FRANCE

CENTRAL SCIENTIFIC AND RESEARCH INSTITUTE "ELEKTROPRIBOR"
30 MALAYA POSADSKAYA STREET
ST PETERSBURG 197046, RUSSIA

RTO MEETING PROCEEDINGS 43

6th Saint Petersburg International Conference on Integrated Navigation Systems

(6eme conférence internationale de Saint Petersbourg sur les systèmes de navigation intégrée)

Papers presented at the 6th Saint Petersburg International Conference organised by the Scientific Council of the Academy of Sciences on the Problems of Motion Control and Navigation, the State Research Center of the Russian Federation — Central Scientific and Research Institute "Elektropribor", and the Systems Concepts and Integration Panel (SCI) of RTO-NATO, held at "Elektropribor", St. Petersburg, Russia on 24-26 May 1999.

Published October 1999

Distribution and Availability on Back Cover

The Research and Technology Organization (RTO) of NATO

RTO is the single focus in NATO for Defence Research and Technology activities. Its mission is to conduct and promote cooperative research and information exchange. The objective is to support the development and effective use of national defence research and technology and to meet the military needs of the Alliance, to maintain a technological lead, and to provide advice to NATO and national decision makers. The RTO performs its mission with the support of an extensive network of national experts. It also ensures effective coordination with other NATO bodies involved in R&T activities.

RTO reports both to the Military Committee of NATO and to the Conference of National Armament Directors. It comprises a Research and Technology Board (RTB) as the highest level of national representation and the Research and Technology Agency (RTA), a dedicated staff with its headquarters in Neuilly, near Paris, France. In order to facilitate contacts with the military users and other NATO activities, a small part of the RTA staff is located in NATO Headquarters in Brussels. The Brussels staff also coordinates RTO's cooperation with nations in Middle and Eastern Europe, to which RTO attaches particular importance especially as working together in the field of research is one of the more promising areas of initial cooperation.

The total spectrum of R&T activities is covered by 7 Panels, dealing with:

- SAS Studies, Analysis and Simulation
- SCI Systems Concepts and Integration
- SET Sensors and Electronics Technology
- IST Information Systems Technology
- AVT Applied Vehicle Technology
- HFM Human Factors and Medicine
- MSG Modelling and Simulation

These Panels are made up of national representatives as well as generally recognised `world class' scientists. The Panels also provide a communication link to military users and other NATO bodies. RTO's scientific and technological work is carried out by Technical Teams, created for specific activities and with a specific duration. Such Technical Teams can organise workshops, symposia, field trials, lecture series and training courses. An important function of these Technical Teams is to ensure the continuity of the expert networks.

RTO builds upon earlier cooperation in defence research and technology as set-up under the Advisory Group for Aerospace Research and Development (AGARD) and the Defence Research Group (DRG). AGARD and the DRG share common roots in that they were both established at the initiative of Dr Theodore von Kármán, a leading aerospace scientist, who early on recognised the importance of scientific support for the Allied Armed Forces. RTO is capitalising on these common roots in order to provide the Alliance and the NATO nations with a strong scientific and technological basis that will guarantee a solid base for the future.

Printed on recycled paper

Published October 1999

ISBN 92-837-1018-5

Printed by Canada Communication Group Inc.
(A St. Joseph Corporation Company)
45 Sacré-Cœur Blvd., Hull (Québec), Canada K1A 0S7

6th Saint Petersburg International Conference on Integrated Navigation Systems

(RTO MP-43)

Executive Summary

This International Conference, which is held annually in Saint Petersburg, brought together this year some 140 engineers and scientists from 17 different countries. It was organised into 4 sessions, covering the main fields involved in integrated navigation systems. The first session, "Inertial Sensors and Systems", emphasised the current importance of laser gyroscopes and the hopes placed in fibre optics gyros. Three interesting papers were presented during the session on the theory and technology of Hemispherical Resonator Gyroscopes (HRG). A triad cluster of accelerometers with common magnet and current source was proposed for inclinometry. The algorithms of dynamics integration in strapdown systems were discussed in two complementary and original presentations. Two generations of redundant strapdown attitude references used in the Alpha International Space Station were described. Theoretical results on fault detection sensitivity in redundant strapdown systems were presented for known and new algorithms.

The second session was entitled "Satellite Navigation". A new technique was described for increasing the accuracy and robustness of GPS receivers by reducing acquisition and tracking thresholds. The comparative performance of a GPS code loop for several suggested new waveforms were analysed but better results were given by the current P(Y) and C/A signals. A presentation was made of the initial results concerning a GPS/GLONASS receiver, tested on SOYUZ in orbital flight, and intended for the station ALPHA. These results show the benefits of using the two constellations for better accuracy and availability as well as the importance of ionospheric errors at negative elevations. An experiment carried out on a GPS receiver in highly elliptical orbital flight demonstrated that it was possible to make measurements at an altitude much greater than that of the GPS constellation. An experimental GPS/GLONASS reference station, used to monitor the integrity of measurements with detection and correction of errors, was described.

The third session, "SatNav/INS Technology", proposed a brief review of existing navigation systems, with emphasis on the exceptional complementarity between inertial and satellite navigations and the benefit of their integration. Some interesting results were presented on the use of an INS for terrestrial geodesy and in-flight gravimetry. A low cost, integrated inertia/DGPS system, in development for flight testing, was described. Simulation results were presented using real flight data. Flight test on an agile aircraft of a GPS receiver in differential mode were commented on. A low cost inertia/GPS/GLONASS preliminary project for civil aircraft was described. A neural technique was proposed for rugged adaptive monitoring of the integrity of an inertia/satellite navigation/barometric altitude system with Kalman filter.

The fourth and final session was given over to "Integrated Navigation" proper. A non-linear Markovian filter was suggested as a solution to the non-linear problem posed by turns in navigation systems with map matching for land vehicles. An experimental integrated inertia/GPS/altitude correlation system, for low altitude flight and CAT2 landing was described and flight test results presented. Integration uses a bank of Kalman filters for fault detection and post-fault reconfiguration. An original presentation was given on federated filters by analogy with electrical networks. This approach casts an interesting light on the loss of optimality of federated filters. Theoretical results concerning stereoscopic image observation equations for navigation and observation were presented. A preliminary study of an integrated system for supervision of the position of an offshore platform was described, with a comparison made between several types of adaptive Kalman filters used to separate low frequency and high frequency movements.

6eme conférence internationale de Saint Petersbourg sur les systèmes de navigation intégrée

(RTO MP-43)

Synthèse

Cette conférence internationale, qui se déroule tous les ans à Saint Pétersbourg, a réuni cette année environ 140 ingénieurs et scientifiques venant de 17 pays. Elle était structurée en quatre sessions qui couvraient les principaux domaines concernés par les systèmes de navigation. La première session, "Senseurs et systèmes inertiels", a rappelé l'importance actuelle des gyroscopes Laser et les espoirs mis dans les gyroscopes à fibre optique. Des travaux intéressants ont été présentés sur la théorie et la technologie des gyroscopes vibrants à résonateur hémisphérique (HRG). Un bloc de trois accéléromètres avec aimant et source de courant communs a été proposé pour l'inclinomètrie. Les algorithmes d'intégration de la dynamique dans les systèmes à composants liés ont fait l'objet de deux présentations complémentaires originales. Deux générations de références d'attitude à composants liés redondantes utilisées dans la Station Spatiale Internationale Alpha ont été décrites. Des résultats théoriques sur la sensibilité de la détection des pannes dans les systèmes à composants liés redondants ont été présentés pour des algorithmes classiques et d'autres nouveaux.

La deuxième session était intitulée "Navigation par satellites". Une nouvelle technique a été décrite pour l'augmentation de la précision et de la résistance aux interférences des récepteurs GPS par réduction des seuils d'acquisition et de poursuite. Les performances comparées d'une boucle de code GPS pour plusieurs nouvelles modulations de sous porteuses envisagées ont été analysées et montrent de meilleurs résultats pour la modulation actuelle. Les premiers résultats en vol orbital sur Soyuz d'un récepteur GPS/GLONASS destiné à la station Alpha ont été présentés. Ces résultats montrent qu'avec deux constellations on gagne en précision et en disponibilité et que les erreurs ionosphériques sont importantes aux élévations négatives. Une expérimentation de récepteur GPS en vol orbital fortement elliptique a vérifié la possibilité de mesures à une altitude très supérieure à celle de la constellation GPS. Une station de référence expérimentale GPS/GLONASS a été décrite qui permet de surveiller l'intégrité des mesures avec détection et correction des erreurs.

Au cours de la troisième session, "Technologie de navigation inertielle et par satellites", une revue rapide des avantages et inconvénients des systèmes de navigation existants a été présentée. La complémentarité parfaite de l'inertie et de la navigation par satellite a été rappelée. Des résultats intéressants de système inertiel pour la géodésie terrestre et la gravimétrie en vol ont été présentés. Un système intégré inertie/DGPS à bas prix pour essais en vol, en cours de développement, a été décrit et des résultats de simulation de performance ont été présentés. Des résultats d'essais sur avion très manoeuvrant d'un récepteur GPS en mode différentiel ont été commentés. Un projet préliminaire d'intégration inertie/GPS/GLONASS à bas prix, pour aviation civile, a été décrit. Une technique neuronale a été proposée pour la surveillance robuste et adaptative de l'intégrité d'un système inertie/navigation par satellite/altitude barométrique avec filtre de Kalman.

La quatrième et dernière session était consacrée à la "Navigation intégrée" proprement dite. Un filtre Markovian non linéaire a été proposé pour résoudre le problème non linéaire posé par les virages à la navigation intégrée avec corrélation topographique pour véhicules terrestres. Un système intégré expérimental inertie/GPS/corrélation d'altitude, pour vol basse altitude et atterrissage CAT 2, a été décrit et des résultats d'essais en vol ont été présentés. L'intégration utilise une banque de filtres de Kalman pour la détection de panne et la reconfiguration après panne. Une présentation originale de la technique des filtres fédérés a été faite par analogie avec les réseaux électriques. Cette approche apporte un éclairage intéressant sur la perte d'optimalité des filtres fédérés. Des résultats théoriques ont été présentés sur les équations d'observations d'images stéréoscopiques utilisables par des systèmes de navigation intégrée. Une étude préliminaire de système intégré pour la supervision de la position d'une plate-forme de forage a été décrite avec comparaison de plusieurs types de filtres de Kalman adaptatifs utilisés pour séparer les mouvements basse fréquence et haute fréquence.

Contents

SESSION II: SATELLITE NAVIGATION

SESSION III: SATNAV/INS TECHNOLOGY

SESSION IV: INTEGRATED NAVIGATION

Preface

The Saint Petersburg International Conferences on Integrated Navigation Systems, which has been held annually since 1994, has become a very significant international scientific event in the field of navigation, guidance and control. This conference is distinguished among international conferences and symposiums on navigation as it provides the opportunity for scientists and engineers to meet, who for many decades had been separated from each other by the constraints of the Cold War. Because of the economic considerations of Russian participants, it has been found efficient to hold the conference within the territory of Russia. Holding the conference in Russia is also rather convenient for the citizens of other countries within the Commonwealth of Independent States (CIS).

Experience of the last decade indicated St. Petersburg to be the optimal place in Russia for the conference site. Traditions initiated by the emperor Peter the Great, when he founded the city as a "window to Europe", are still preserved. He also founded, 275 years ago, the Saint Petersburg Academy of Sciences (now the Russian Academy of Sciences). Among the most important problems set before it were research challenges in the field of navigation. Well known to the scientists all over the world were members of the Russian academy; from the great mathematician and mechanician of the XVIII century academician Leonard Euler to the patriarch of modern Russian gyroscopy, academician Alexander Yu. Ishlinsky.

Continuing this tradition, the Scientific Council on the Problems of Motion Control and Navigation of the Russian Academy of Sciences established the conference and assigned responsibility for it to the Saint Petersburg section of the Council. The section works at the institute "Elektropribor", which considers it an honor to conduct and ensure the conference's success.

The previous five conferences drew 120-130 participants representing many leading firms and universities from 12 countries. The presented papers were mainly devoted to the following problems:
- gyro sensors;
- inertial navigation systems;
- satellite navigation systems;
- integrated navigation systems GPS/INS.

This same set of topics was considered at the 6th Conference. However, there were some important differences. The Research and Technology Organization of NATO participated in the organization of the 6th Conference. The RTO has some experience in holding scientific events in Russia. Specifically the Lecture Series "System Implications and Innovative Application of Satellite Navigation" was held in St. Petersburg in 1996 with great success. The host organization accepted the RTO conference standards for the format, providing only the plenary sessions.

The extended Program Committee consisted of the scientists from France, Italy, Russia, Turkey, the UK, the USA which examined about 60 paper abstracts. The conference format provided for presentation of no more than 27 papers, so the competition was very serious. The final program includes 25 papers presented by 61 authors working for 28 organizations from 9 countries. The papers were divided into the following subjects:
- gyro sensors - 7 papers;
- gyro systems - 2 papers;
- satellite navigation - 5 papers;
- integrated satellite navigation systems - 6 papers;
- integrated navigation systems - 5 papers.

The papers presented at the conference reflected the principal trends in modern navigation development.

Any scientific conference offers several opportunities for the exchange of technical information. The primary means is via the presentations and content of papers; in this respect the conference had a wealth of valuable technical information. Personal contact with the colleagues from the other organizations and other countries is another very important aspect. Each of the 140 participants representing 70 organizations from 17 countries participated had a excellent opportunity for face to face technical interchange during the plenary sessions and the cultural events.

The number of the conference participants was about evenly divided between Russian and foreign attendees. There was a full representation of the leading scientific organizations and leading scientists of Russia (27 organizations from 6 cities: St. Petersburg, Moscow, Yekaterinburg, Perm, Miass, Orenburg). This was the first conference using simultaneous interpretation in both Russian and English, which gave an opportunity to every specialist from Russia and other CIS countries, as well as those who did not have a command of the Russian language, to participate fully in the conference.

Great contribution was made by the scientists from the USA, France, Germany and Ukraine who participated in our conferences from the very beginning. It should be emphasized that we have never seen at the conferences such a remarkable representation of the leading firms and universities from Germany. In addition to these countries the conference attracted many new participants from Bulgaria, Canada, Holland, Hungary, Iran, Italy, Poland, Romania, South Africa, Spain, Turkey and the United Kingdom.

On the whole, this conference came up the participants' expectations and its experience will be used when preparing the 7th Conference in 2000. New scientific and technical results contained in the papers presented at the conference, show that the integrated systems development is the most important trend for development of navigation technology both today and for the foreseeable future.

V. Peshekhonov
Co-chairman (Russia)

J. Niemela
Co-chairman (USA)

Systems Concepts and Integration Panel

Dr Murat EREN
TUBITAK-SAGE
Gudum Kontrol Lab.
O.D.T.U. Kampusu
16531 Ankara
Turkey

Mr Loic CAMBERLEIN
32, Avenue de la Bourdonnais
75007 Paris
France

Dr-ING Luigi CROVELLA
Alenia
10072 Caselle Torinese
Italy

RTA/SCI Panel Executive (Pro-Tem)

Lt Col Richard VANTINE
Deputy, Operations & Coordination
RTA
BP 25
7, rue Ancelle
92201 Neuilly-sur-Seine Cedex
France

TECHNICAL EVALUATION REPORT

Loïc Camberlein
Consultant
32, Avenue de La Bourdonnais, 75007 Paris
Tel/Fax : 33- (0)1 45 51 47 59
e-mail : Lcamberlin@aol.com

INTRODUCTION

The 6th Saint Petersburg International Conference on Integrated Navigation Systems was held on the 24 – 26 of May. It was organized this year by the Scientific Council of the Russian Academy of Sciences on the Problems of Motion Control and Navigation with the participation of the Research and Technology Organization of NATO. The Conference was sponsored by the State Research Center of the Russian Federation – Central Scientific and Research Institute "Elektropribor". This international conference, which is held annually in Saint Petersburg, brought together some 150 engineers and scientists from 17 different countries. The Conference was co-chaired by Prof. Vladimir G. Peshekhonov, (Russia), and Dr John Niemela (USA).

The Conference was organized into 4 sessions covering the main topics involved in integrated navigation systems and their applications :

I. Inertial Sensors and Systems
II. Satellite Navigation
III. SatNav/INS Technology
IV. Integrated Navigation

This scope is very broad and, of course, the Conference could only partially cover the status of research and development in this area. Speakers from 11 countries presented 25 papers of which 12 from Russia, 1 from the Ukraine and 14 from other countries : Canada, France, Germany, Italy, Netherlands, South Africa, South Korea, Turkey, United States.

In the following review, papers are referred to by their index in the conference's program and proceedings.

SESSION I - INERTIAL SENSORS AND SYSTEMS

The first session consisted of 9 papers and was the largest one. It was divided in 2 sub-sessions : Inertial Sensors with 7 papers and Inertial Systems with 2. This separation may appear somewhat arbitrary because some papers might have been in either one of the two sub-sessions.

In the Inertial Sensor part, a Russian overview paper [1] on optical gyroscopes first retraced the history of the laser gyroscopes which started about 35 years ago. It emphasized their current domination in most strapdown military applications, and an almost total domination in commercial aviation. Even fundamental research is interested in laser gyroscopes : "superlarge" laser gyroscopes, a few meters in perimeter, have been made for investigation of the relativity theory, the earth's tides and the earth's angular and seismic motions. This paper also stressed the hopes placed in fibre optics gyroscopes (FOG), for which the R&D started about 25 years ago. It pointed out the large number of countries where significant efforts are being carried on, demonstrating the strong interest placed in this technology. The FOG technology was presented as more mature and competitive in the low grade segment. Promising work was reported on slightly larger gyroscopes for very accurate submarine navigation taking advantage of the very benign environment and of the very low random drift provided by the FOG technology.

Three interesting papers presented different important aspects of the theory and technology of vibrating gyroscopes, the mechanical challengers of optical gyroscopes. One of them from South Africa/Russia [5] discussed a new method for the mathematical analysis of the influence, on the gyroscope dynamics, of stresses in the resonator. The understanding of this influence is essential for the development of a good vibrating gyroscope. Inner stresses result from the machining process of the resonator or from the inertial motion. They create micro cracks which degrade the performance. The analysis was done for a Hemispherical Resonator Gyroscope (HRG) but the method could be applied to a large variety of vibratory gyroscopes.

The second paper [6], by Russian authors, discussed the difficult technological problems posed to HRG by the very strong isotropy requirements. A few of these problems were analysed for the pick-off and forcing electrodes and their associated electronics. Solutions to optimize them were proposed, based on the experience gained with HRG prototypes of two different sizes.

The third paper [7], from the Ukraine, was about the manufacturing of a low cost HRG. The sensing element is the hemispherical resonator; it is usually made of one single part in fused quartz which is complex and, therefore, expensive to produce. A resonator in two parts, much simpler and thereby easier to produce, was proposed to lower the cost. The manufacturing process described uses spherical and cylindrical optics production techniques.

A triad cluster of orthogonal accelerometers was proposed in a Russian paper [4] for the inclinometry of oil and gas wells. The rapid temperature changes during well surveys create large instabilities of the scale factor difficult to model and, therefore, to compensate. A common magnet and a common DC current source are used to solve the problem. This triad cluster is also proposed for strapdown inertial systems.

Two complementary and original presentations, one from the U.S. the other from Russia, discussed the integration algorithms of angular dynamics in strapdown systems. The two speakers referred to each other's work. The accuracy of the integration algorithm of a strapdown inertial navigation system is a key element of its performance, especially in response to angular coning motion. Usually, these algorithms assume a flat transfer function for the incremental outputs of the gyroscopes, but this is not true for many gyroscopes. The U.S. paper [2] proposed a simple modification of the coefficients of the usual algorithms to tune them to non flat transfer functions. The Russian paper [8] was more general and described a new approach for three classes of integration problems which apply to strapdown systems but also to other real time systems : multiple integration of a signal, multiple integration of a product of two signals and signal smoothing by polynomial fitting. This approach uses increments of the signal multiple integrals over the sampling interval of the high rate preprocessing. These integrals accurately compress the raw sensor data for the low rate processing. This approach was said to solve the problem of non flat sensor transfer function and to be very efficient in filtering the gyroscopes' noise, therefore to attenuate the pseudo-coning errors.

Two Russian papers dealt with redundant strapdown systems. The first one [3] described two generations of redundant strapdown attitude references used in the Alpha International Space Station. The first generation, developed for MIR, is mounted on the Functional Cargo Module of Alpha and is already in orbit. It has classical triads of orthogonal gyroscopes. The second generation is on the Service Module which should be launched this year. It has tetrads of non orthogonal gyroscopes. Both use floated gyroscopes with rotor gas bearings and float magnetic centering. Both have a 10 year service life. The second generation was necessary to meet the Service Module's more demanding requirements of 0.003 deg/h drift and 30 ppm scale factor stability.

The second paper [9] presented a comparative analysis of known and new algorithms for failure detection and isolation in fault tolerant redundant strapdown systems. Linear metric space formalism was used to describe, simply and uniformly, these algorithms. The comparison made shows that the sensitivity of the failure detection of all the considered algorithms is equivalent.

SESSION II - SATELLITE NAVIGATION

The second Session had 5 papers. 2 of them were about GPS receivers signal processing and the 3 others about experimental results.

A French paper [11] described a new technique for increasing the accuracy and robustness of GPS receivers by reducing the acquisition and tracking thresholds. This is done by using the strong tracking channels of the receiver for aiding the other channels to acquire or track the satellites presenting low signal to noise ratio. This is also done by reducing the predetection bandwidth to the lowest possible value. This paper also presented a review of other known threshold's reduction techniques.

A U.S. paper [12] analysed the comparative performance of a GPS code delay lock loop for several new suggested GPS waveforms. Better tracking results for the current P(Y) and C/A

signals were shown as compared to the new more complicated subcarrier modulated waveforms.

A Russian paper [10] presented the initial results concerning a GPS/GLONASS receiver, tested on SOYUZ in orbital flight, and intended for the station ALPHA. These results show the benefits of using the two constellations to improve accuracy and availability. They also show the importance of ionospheric errors at negative elevations.

A German paper [13] described the experiment of a GPS receiver carried out on the Small Satellite Equator-S. The purpose of this experiment was to demonstrate the feasibility of the positioning of geostationary satellites by using GPS. The highly elliptical orbital flight demonstrated it was possible to make measurements up to 61,000 kilometers, i.e. at an altitude much greater than the altitude of the GPS constellation. This result was achieved with a receiver not especially optimized for that mission.

A paper from the Netherlands [15] described an experimental GPS/GLONASS reference station, used to monitor the integrity of measurements and to detect and correct errors in real time. Experimental results were shown to illustrate the monitoring as well as multipath modeling and ionosphere estimation capabilities. Future work is intended to generate differential GPS and GLONASS corrections for differential positioning. Relative positioning should also be explored. Ultimately, a network of GPS/GLONASS reference stations is envisaged, depending on GLONASS development.

SESSION III - SATNAV/INS TECHNOLOGY

The third Session had 6 papers

A German paper [20] made a brief review of existing navigation systems with emphasis on satellite systems like GPS and GLONASS. It mentioned their excellent accuracy capabilities but also their poor dynamic response and fairly poor reliability. It underlined the exceptional complementarity between inertia and satellite navigation and the benefit of their complete integration to overcome the weaknesses of both systems in such demanding applications as Cat III landing and car guidance.

Some interesting results were presented by a Russian author [16] on the use of a gimbaled inertial system for survey with zero velocity updates for positioning and vertical deflexion estimation. The system was also used for in-flight gravimetry, associated with a differential GPS receiver. For these measurements, an accurate gravimeter was installed on the Z axis of the inertial platform. The author showed ground positioning accuracy of 0.4 to 0.6 meters and gravimetry accuracy of 1 mGal per 3 kilometers.

A Canadian paper [17] described a low cost integrated strapdown IMU/DGPS system developed for flight testing. The loosely coupled integration is achieved by a 42 state Kalman filter with state feedback to correct the inherently large errors of the low cost IMU. Simulations using real flight data were very valuable to test the filter and detect the need to expand the error models of the gyros and accelerometers to meet the performance goal.

An Italian paper [18] presented preliminary work for either using, for flight testing, a DGPS receiver by itself or a high precision INS/DGPS system. The most interesting part of the paper is the flight test on an agile aircraft of two GPS receivers in differential mode. In particular, this test provided the conditions to be respected for the use of DGPS alone. One of the conclusions of the paper is that the best solution for flight testing is an integrated INS/DGPS system.

A Russian paper [19] presented the preliminary definition of a low cost integrated strapdown INS/GPS/GLONASS system for civil aircraft.

A Russian paper [21] proposed a neural technique for rugged adaptive monitoring of the integrity of an inertia/satellite navigation/barometric altitude system with Kalman filtering.

SESSION IV - INTEGRATED NAVIGATION

The fourth and final session had 5 papers.

Optimal algorithms were proposed by an interesting Russian/South Korean paper [22] for solving the integration of a car navigation system using map matching/odometer/ directional gyro/SatNav. The difficulty lies in the non-linear problem posed by turns. A non-linear Markovian filter was suggested as a solution to this non-linear

problem during turns. During straight lines, this non-linear filter is replaced by a linear Kalman filter. This method allows the use of all the available information. Real information was used to evaluate the map-matching accuracy.

A paper from Germany [23] presented an experimental INS/GPS P(Y) code /terrain correlation system for demonstrating low altitude flight and Cat I landing capabilities in precision and safety. Integration uses a bank of 4 Kalman filters corresponding to the normal mode and 3 back-up modes. The Kalman filters are identical and use an index vector to configure them to the proper sensors. The system allows fault detection and post-fault reconfiguration. The latter is done simply by switching to the back-up filter which is not affected by the failed sensor. In this way, even a delayed failure detection does not degrade the back-up modes. The flight test results presented comply with the low altitude flight and Cat I landing requirements, even under loss of GPS and comply with Cat II landing with GPS.

An original Russian paper [24] proposed the synthesis of federated filters by analogy with the transformation of electric circuits. In particular, this approach explains, in a simple way, the reasons of the loss of optimality of federated filters.

A Russian paper [25] presented theoretical results concerning the use of stereoscopic images for 3D navigation and observation. It described the observation equations to be used in a Kalman filter for computing position fixes from one stereo image or from a sequence of them. Then it gave the observation equations and an algorithm for automatically detecting and identifying small objects.

A Turkish paper [26] described a preliminary study of an integrated system for supervision of the position of an offshore platform. The measurements are provided by hydroacoustic, inclinometric and dynamometric sensors. Two Kalman filters are used to separate the low frequency motion, made by the wind and current, from the high frequency motion caused by the sea-way. The parameters of the first filter are automatically adapted to the sea state estimated by the second filter.

CONCLUSION

The progress made through the years by the Conference in quality and interest has been noticed by all the participants in the previous Conferences. The growing number of participants, the improvement of the content and oral presentation of the papers, the increased comfort of the Conference facilities, the good local organization and the efficiency and friendliness of the Elektropribor staff were particularly appreciated. One very welcome innovation this year was the use of excellent Russian/English and English/Russian interpreters. This must continue in the future to further increase the reputation and the attendance of the Conference.

The general topic of the Conference – Integrated Navigation Systems – will certainly stay an important one in the future for manned and unmanned vehicles. Inertial sensors, inertial systems, satellite navigation and correlation techniques of sensor measurements with data bases will probably remain the major ingredients of integrated navigation for a long time.

Laser and fiber-optic gyros: the status and tendencies of development

D.P.Loukianov
The State Electrotechnical University, St.Petersburg
5, Prof. Popova Str., Saint Petersburg, 197376, Russia
Tel./Fax (812) 234 45 66

The analysis of the status and tendencies of development of laser (RLG) and fiber-optic gyros (FOG) is based on the study of the research, development and production of the leading companies in Russia, the USA, France and Germany.
RLGs of different architectures are considered: those with planar and nonplanar resonators, mirrors and totally reflecting prisms, employing various methods for linearizing the output characteristic. The development of ring lasers with superlarge perimeter from 4 to 14 m for fundamental research is discussed.
Notable achievements in the field of optical waveguides and integrated optics provided revolutionary progress in the development of FOGs which are conquering the domestic and international markets. The present-day architecture of FOGs, their main components and error models are considered. Particular attention is given to a promising line in the development of FOGs that uses the effect of controlled pseudorandom polarization and depolarization of counter-propagating waves which allowed the FOG's drift to be reduced to the level of 0.005 deg/h and better.

Introduction

This survey of the results and industrial development of the leading experts and companies of Russia, the USA, and France will come out to the reader on the boundary of the XXI century as a direct result of the radical social reformation which has been under way in Russia since perestroika began in the former Soviet Union and the socialist countries. The new social processes have destroyed nearly an age-long ideological, political and military confrontation of the leading states of the world and have created realistic pre-conditions for effective international cooperation which has proved to be fruitful in the field of new navigation technologies as well.

At least three aspects are of paramount importance here:

- expansion of users' network and updating of the global satellite navigation systems GPS and GLONASS;
- development and production of a radically new type of sensitive elements – the RLG and FOG which, in their turn, gave impetus to vigorous development of SINS;
- the synthesis of adaptable architectures of integrated SINS/GLONASS/GPS-systems for a wide scope of military and commercial applications.

It is no secret that the most significant scientific and practical results in laser and fiber-optic gyroscopy have been achieved in the depths of military and industrial complexes of the leading states. Although there is a great number of publications devoted, first of all, to the theory of optical gyros, many basic details concerning the production and new technologies are still classified, and the period between the laboratory/prototype models and commercial production turned out to be rather long. The history of Honeywell, one of the leading companies in the world in laser gyroscopy is the evidence that considerable commercial progress can only be made through effective mastering of military and commercial markets or, in other words, dual technologies. Simultaneous development and introduction of a great number of devices and systems provides decrease in their cost which, in its turn, results in increase of sales. That can be illustrated by the results of Honeywell activities in the period from 1965 to 1994 [1].

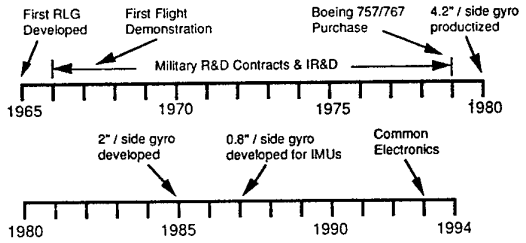


Fig.1. Honeywell RLG technology development history (1965-1994)

Indeed, as follows from Fig.1, the first stage of the investigations and developments which lasted rather a long time from 1965 to 1979, could only have been realized under the government financing which was evidently provided until the first deliveries of the laser navigation equipment for the Boeing 757/767. Subsequently, the results obtained enabled a new series of RLGs with a more reduced size and weight and, as will be seen from Fig.2, the quantity of the production

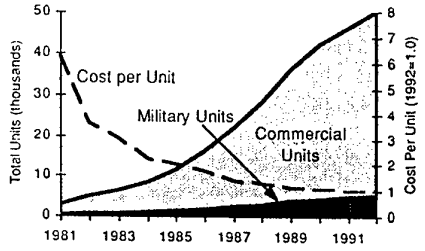


Fig.2. RLG production history and cost trend

Paper presented at the RTO SCI International Conference on "Integrated Navigation Systems", held at "Elektropribor", St. Petersburg, Russia, 24-26 May 1999, and published in RTO MP-43.

sharply increased, at the same time the costs became lower. By 1992 the ratio between the commercial and military production was more than 10:1, whereas the cost per unit was reduced by a factor of 6.5 from 1981 to 1992. As T.Cunningham remarks in [1], this is representative of nothing else but of the fact that the cost and the quantity are inversely proportional to each other.

The history of laser gyroscopy development at one of the leading enterprises in the former Soviet Union, the Central Design Bureau (CDB) and the plant 'Arsenal' is different to some extent. Here the development of both prototype and serial products were systematically financed as governmental orders which had little or no provision for development of civil commercial technology. The planned production and tight control were targeted to make the research and development periods shorter, to improve the tactical characteristics of the devices developed, as follows from Fig. 3.

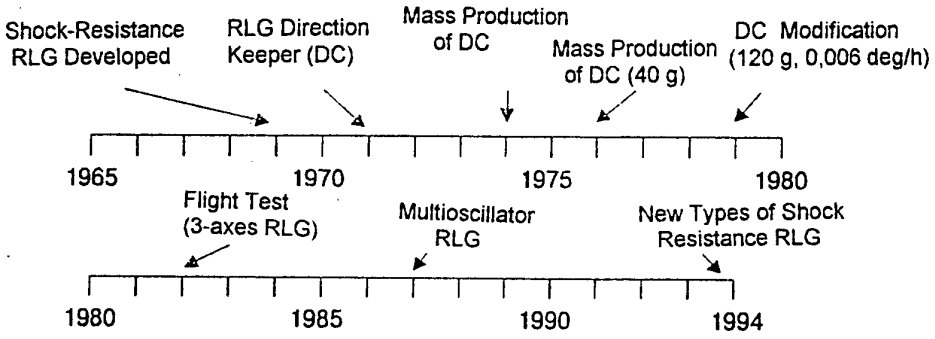


Fig.3. CDB and 'Arsenal' RLG technology development history (1969 - 1994)

The world process of designing the FOG that made wide use of achievements in laser gyroscopy, seems to be more dynamic. If the beginning of the intensive research on FOGs dates back to the late 70s and early 80s, then it took the FOG less than 10 years to be developed from laboratory prototypes to quantity-produced devices. The average-accuracy FOGs of 0.1-5 deg/h stability for tactical missiles [2,3], aircrafts, ground-based and marine vehicles [4,5] had been developed by the mid-90s. The FOG modifications of minimum configuration *are* employed on the Dornier-328 - small short-range aircrafts and the Boeing 777 for orientation, in case the main navigation aids fail.

In 1996 Honeywell was reported to have developed FOGs with the random and systematic drift of $2 \cdot 10^{-4}$ deg/h and $3 \cdot 10^{-4}$deg/$\sqrt{h}$, correspondingly [6,7]!

The Russian government has greatly reduced financing of large research and industrial enterprises recently. This resulted in establishment of a number of small enterprises which have to earn the money on their own and look for their niches in the national and international markets. Of course, it couldn't but affect the development of the optical gyroscopy in Russia.

'Electrooptica' is one of such small research and production complexes. Over the 10 years of its existence the enterprise has managed to design a number of prototype RLG models and start producing them in small batches. Laser gyros provided the basis for development of airborne SINS and the aids for monitoring of railway track profiles.

Another small enterprise 'Fizoptica' is also known for its high-efficient activities. Since 1989, when they had to start with choosing the concepts for designing and mastering the new technologies of FOG manufacturing, they have been able to put the VG941 and VG 951, the well-known gyros now, into full production. The list of the companies that sell the products of 'Fizoptica' in the international market, among them TANAKA TRADING (Japan), SPAR ENGINEERING Ltd. (Ireland), SENSORTEC (Germany), SAMYANG RADIO Co. Ltd. (Korea), FIBERSENSE CORPORATION (USA), CHINA NORTH OPTICAL CORPORATION (China), is the convincing evidence of great and steady progress of the enterprise.

The present status and tendencies of development of laser gyroscopy

Methods and techniques for linearizing RLG input-output characteristic. Two-mode ring laser gyros. The history of the laser gyroscopy development is, first of all, the history of overcoming the lock-in effect (mutual synchronization of counter-propagating waves) which causes nonlinearity of the data output, scale factor instability and zero bias. A typical input-output characteristic for a two-mode RLG is shown in Fig. 4

At low rates two counter-propagating waves in the resonator are very close in frequency and the scattering

effect provides coupling between the two beams. As a result, the frequencies of independent generators become equal, thus resulting in a dead zone $\pm\omega^L$ [8].

Two principally different technological approaches stand out against the numerous decisions aimed at reducing or eliminating the lock-in effect:
mechanical 'dither' – sinusoidal motion of large amplitude applied to the RLG;
the use of magnetooptical and electrooptical effects to induce artificial non-reciprocity.

The first approach involves either the forced rotation of the RLG at a constant angular rate $\omega^D >> \omega^L$ when the operating point of the input-output characteristic is forced out of the synchronization zone (Fig.4), or the reversal (oscillating) angular motion (Fig.5), known as dither, when the RLG passes rapidly through the synchronization zone and the input-output characteristic is linearized (Fig.6).

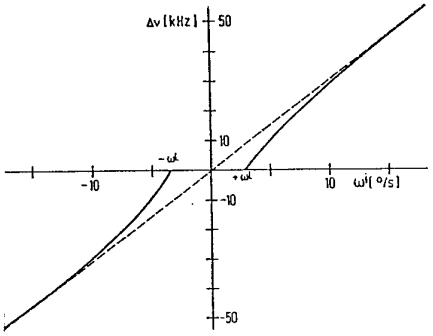


Fig.4. Input-output characteristic for a two-mode RLG for $S = 0.017$ m^2, $L = 0.6$ m, $\lambda = 0.63\mu$m

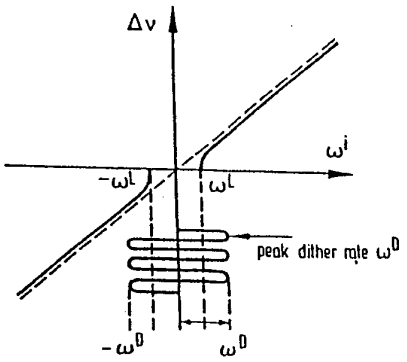


Fig.5. Lock-in compensation with dither: $\omega^L = 300$ deg/h

The last approach proved to be the most productive, - dither gyros were and have been developed and used in the former USSR and Russia (the Research Institute 'Polyus', the Research and Production Association (NPO) 'Arsenal', MIEA, NPO 'Rotor', 'Electrooptica', et al.), in the USA (Honeywell, Litton, Rockwell, et al.), Germany (Crouzet-Sfena), and Great Britain (Sperry). It is safe to say that dither gyros, highly accurate, reliable, relatively low-cost, stimulated the development of strapdown inertial navigation systems.

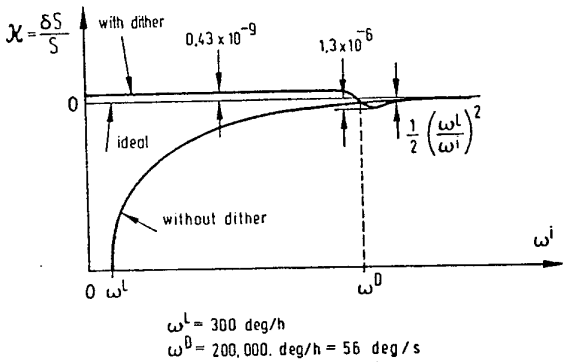


Fig.6. Scale factor linearity with dither

In spite of the obvious advantages due to dithering as a means of linearizing the RLG input-output characteristic, the experience gained over a long period of operation made it possible to reveal a number of serious disadvantages:

- a considerable tangential acceleration of RLG component parts, close to 30 g;
- excitation of attendant mechanical oscillations;
- a considerable level of the noise zero bias preventing the RLG from approaching its potential accuracy.

The quest for elimination of the drawbacks mentioned above has led to the development of a modified reversible rotation of the RLG such that the RLG makes a few revolutions at a constant angular rate in one direction, for example, clockwise, and then 'instantaneously' (over 0.1 s) it starts rotating in the opposite direction and makes the same number of complete revolutions counterclockwise. In fact, the output data is modulated by a bipolar sequence of rectangular pulses – meanders. In literature abroad the latter is known as a carouselled or rate bias method. The main advantage of this method is less time over which the RLG has to stay in the lock-in zone. If N is the number of passes through the lock-in zone, then, with **conventional** dithering at a frequency of 400 Hz, the value of N per hour will be as high as

$N_{osc} = 400 \cdot 2 \cdot 3600 \approx 3 \cdot 10^8$!

At the same time for the reversible (carouselled) mode of motion at an angular rate of 50 deg/s and 2 revolutions in each direction [9] N_{rev} over an hour will make up:

$$N_{rev} = \frac{50}{720} \times 3600 = 250.$$

As a result, the random drift can be reduced more than

an order of magnitude. At present the reversible method is one of the most effective ways of approaching the potential accuracy of the RLG. It is used in airborne and marine SINS of the accuracy class ranging from 1 mile/h [9] to 1 mile/24 hours [10].

Now let us consider the possibilities and the results achieved due to the use of magnetooptical and electrooptical effects. The general conception of using the above effects consists in creating nonreciprocal phase properties for counter-propagating waves in the RLG resonator without any additional mechanical motion. At the first stages of the development of laser gyroscopy great promise was shown by nonreciprocal phase-shifting units (NPU) based on the Faraday effect. The diagram of an RLG with an NPU is shown in Fig 7.

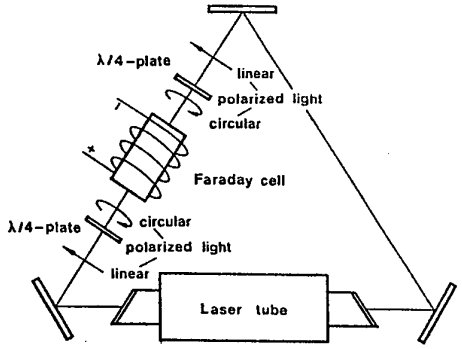


Fig.7. An RLG with the Faraday cell bias technique

The NPU represents an optically transparent cylinder between two quarter-wave plates. They fulfil the function of polarization converters which convert the linearly polarized radiation in the RLG resonator into the circularly polarized one inside the cell [11]. The longitudinal magnetic field applied to the NPU leads to the nonreciprocal phase-shift $\Delta\varphi$ which causes the frequency shift of the counter-propagating waves by a value of

$$\Delta f = \frac{c\Delta\varphi}{\pi\Sigma l_i n_i},$$

where $\Delta\varphi = VH\,l$; V is the Verdet constant; H is the intensity of the longitudinal magnetic field, l is the length of the NPU, $\Sigma l_i\, n_i$ is the optical perimeter of the RLG.

However, this elegant method of creating the initial frequency shift is not devoid of disadvantages, namely, significant dependence of the nonreciprocal phase shift on external magnetic fields and temperature variations.

Multilayered screening is the natural solution to this problem but the need to reduce the residual fields to the value of about $5\cdot10^{-5}$E makes reliable screening an intricate technological problem [11]. The differential circuit including two NPUs, as shown in Fig.8, seems to be a more effective decision [11].

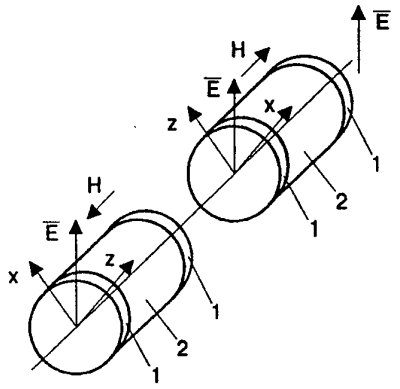


Fig.8. An NPU with two separate Faraday cells: 1 – λ/4 plate, 2 - glass

Here the crystallographic axes of the input quarter-wave polarization converters x and z of each NPU are turned relative to each other by an angle of 90°. As a result, the circularly polarized waves in the neighbouring cells rotate in the counter directions, and the magnetization fields are oriented towards each other. In this case, the influence of external homogeneous magnetic fields is eliminated.

A similar RLG design found practical realization at the Kiev plant 'Arsenal' in the quantity-produced COG-1.2 (КОГ-1,2).

The structure shown in Fig.8 can be modified. In this case the two parts of the differential circuit of the NPU are connected by a half-wave plate thus eliminating two boundaries [12], as shown in Fig.9.

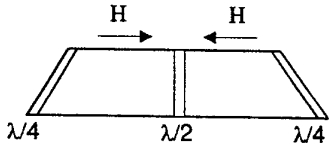


Fig. 9. A differential circuit with two Faraday cells

A serious impediment to the further wide use of the Faraday NPU in double-frequency RLGs is the temperature dependence of the Verdet constant which makes it difficult to gain the desired stability of the initial frequency difference of the counter-propagating waves.

Multi-oscillator laser gyros. A situation is different for multi-frequency RLGs whose general optico-physical block diagram is shown in Fig.10.

The plane resonator of the RLG contains two elements: a Faraday cell (nonreciprocal element) and an optically active quartz cell (reciprocal element).

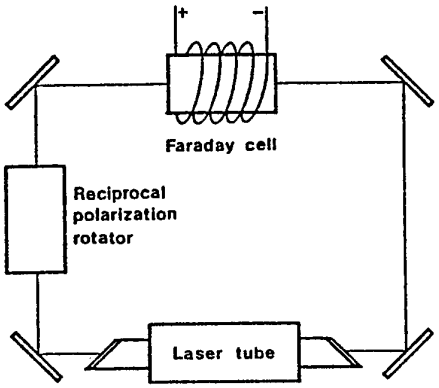


Fig.10. A block diagram of a 4-frequency RLG:

The natural waves of this system is a pair of counter-propagating left-hand and right-hand polarized waves. When passing through the reciprocal element, each of them experiences opposite phase shifts which causes frequency split by a value of

$$\Delta\nu = \frac{\theta c}{\pi \Sigma l_i n_i},$$

where $\pm\,\theta$ is a phase shift between the left-hand and right-hand polarized waves.

In particular, with $\theta=\pi/2$, $\Delta\,\nu = c/2L$, as shown in Fig 11.

Now a couple of counter-propagating circularly polarized frequency-separated waves are travelling in the RLG resonator in each direction (clockwise and counterclockwise). That corresponds to two independent laser subgyros whose natural waves are separated in frequency and the state of polarization. The operation of each of these subgyros with circular polarizations opens up favourable conditions to reduce the coupling between the counter-propagating waves in comparison with the RLGs with linear-polarized waves.

Really, when the circularly polarized wave is back scattered, the direction of the circular polarization is reversed. In an ideal case the direct and scattered waves do not interact.

However, in real conditions counter-propagating waves are not strictly circularly polarized but elliptically polarized. Thus the back-scattered wave includes a component of the same polarization as that of the direct wave. An NPU, the Faraday cell, is introduced in the RLG resonator to eliminate the residual influence of the coupling of counter-propagating waves. It shifts the generating frequencies of the counter-propagating waves in the opposite directions, as shown in Fig.12. A similar nonreciprocal phase shift caused by the RLG rotation results in additional separation of frequencies.

For each of the subgyros with orthogonal circular polarizations the frequency separation of the counter-propagating waves will make up:

for the subgyro with left-hand circular polarization (LCP)

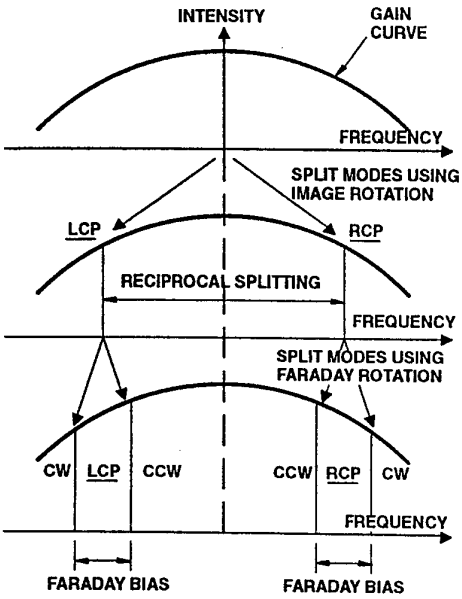


Fig.11. Mode splitting for 4-frequency RLG (nonrotating gyro)

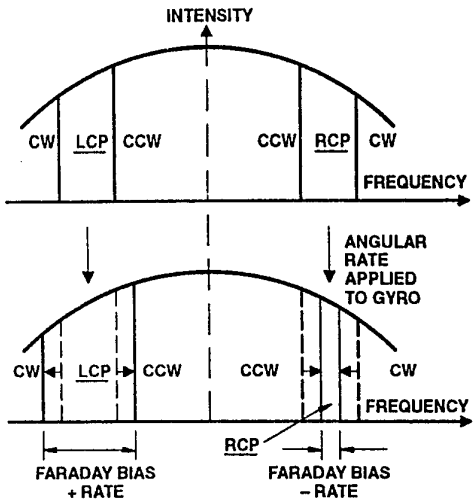


Fig.12. Mode splitting for 4-frequency RLG (rotating gyro)

$\Delta \nu_{lc} = \Delta \nu^F - K\Omega$;
for the subgyro with right-hand circular polarization (RCP)
$\Delta \nu_{rc} = \Delta \nu^F + K\Omega$,
where $K = \frac{4S}{\lambda L}$ – the RLG scale factor.
So, the effect proportional to the angular rate of the RLG rotation can be presented as $\Delta \nu_{r.h} - \Delta \nu_{l.h} = 2K\Omega$, from which it follows that the scale factor has doubled and that the RLG output signal does not depend on frequency variations (e.g., thermal) introduced by the Faraday NPU.

H. de Lag [13] was probably one of the first to suggest using the four-wave mode in the RLG to reduce the coupling between the counter-propagating waves. The US patent [14] issued for the differential laser gyro system adopted the well known abbreviation DILAG for four-oscillator gyros. One of the first DILAGs was designed in 1977 [8].

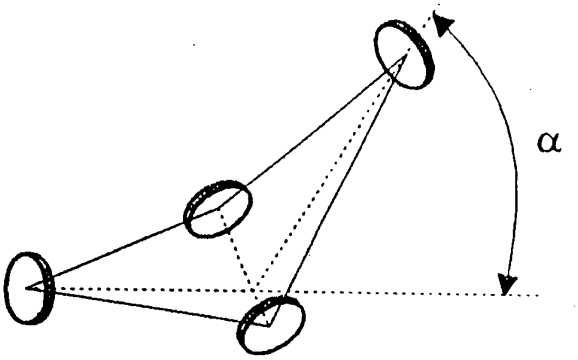


Fig.13. A block diagram of a nonplanar resonator

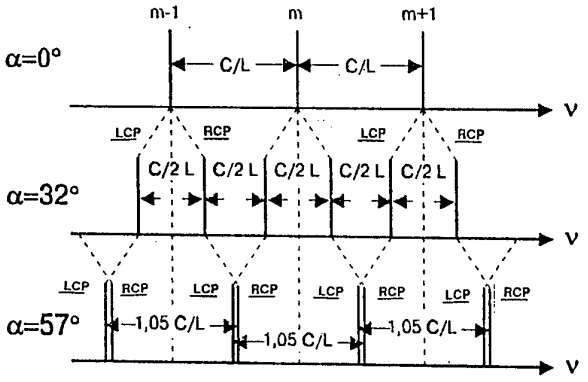


Fig.14. The natural frequency spectra for nonplanar resonators for different bending angles α

Multi-oscillator RLGs with a non-planar resonator stand apart [16-19]. They provide conditions for generation of circularly polarized waves. A block diagram of a circuit like that is shown in Fig.13. The frequency separation of the left-hand and right-hand circular waves depends on the bending angle α of the circuit. The changes in the spectrum of the natural frequencies of RLG resonators for different angles α are represented in Fig.14. Thus, the RLG circuit bending turns out to be a very effective means in attaining two aims simultaneously, as it provides the conditions for:

- generation of circularly-polarized waves;
- geometrically regulated frequency separation of left-hand and right-hand circularly-polarized waves without using an optically active quartz cell in the resonator.

The commercial production of four-frequency RLGs, called zero lock laser gyros (ZRLG), was started by Litton in 1991 [20].

The conception of the RLG architecture with a non-planar resonator proved to be very fruitful and gave rise to another series of devices based on counter-propagating circularly polarized waves [19]. Among them Zeeman laser gyros in which the magnetic field is applied directly to the active medium, making the spectral line of Ne split in the magnetic field into two components with orthogonal circular polarizations according to the relation
$\Delta \nu_z = 2g\beta H/\hbar$,
where g is the Lande factor, β- the Bohr magneton, $\hbar$ – the Planck constant. The Zeeman effect, as well as the Faraday effect, is nonreciprocal and causes the frequency of the counter-propagating waves to split by the value of
$\nu_{rc} - \nu_{lc} \approx (2g\beta H/\hbar)(\Delta \nu_r/\Delta \nu_D)$,
where $\Delta \nu_r$ and $\Delta \nu_D$ are the frequency bands of the resonator and of the Doppler loop of the spectral line of Ne, correspondingly.

Either the two-frequency or four-frequency modes may be realized depending on the bending angle of the resonator and the width of the spectral line [19]. No optical elements are needed in this case which is an obvious advantage of Zeeman RLGs.

The survey of methods and procedures for overcoming lock-in effects of counter-propagating waves in the RLG would not be complete unless the possibility to use magnetic mirrors [21-25], based on the nonreciprocal transverse Kerr effect [23-25], for frequency separation of the counter-propagating waves is mentioned here. The effect is maximum when two beams are linearly polarized in the plane of incidence and the magnetization is in the plane of the magnetic mirror and normal to the plane of incidence. The magnetic mirror consists of a very thin transparent layer of specially composed iron garnet, together with a high reflectivity multilayer dielectric stack, all this allowing the reflectivity of over 99.8 %. The nonrecirpocial phase shift for the counter-directed waves is 2 arc

minutes which produces a bias equivalent to 60 degrees per second in a 0.4 m perimeter RLG operating at 1.15 $\times 10^{-6}$m wavelength. The magnetic mirror was reported to have been used as a nonreciprocal element in four-frequency RLGs [26].

Configurations of resonators and their elements. Two modifications of laser gyros, - a planar-resonator and non-planar-resonator RLGs, - have already been partially considered. The discussion about the resonator can be continued by considering the geometry of its configuration which depends on the number of the mirrors used. Three and four-mirrored resonators are commonly used in practice, although five-mirrored ones are also mentioned in literature [27]. In deciding on the number of the mirrors, it is necessary to analyze the following parameters.

The RLG scale factor. It is determined by the ratio between the resonator's area and its perimeter, all other factors being the same. If it is assumed, for comparison, that both resonators (triangular and square) are inscribed in the same circle, D in diameter, these ratios will be as follows:

$S/L\Delta = 0.125\ D$; $S/L\square = 0.175\ D$,

where $L\Delta = 2.6\ D$; $L\square = 3.45\ D$.

Here the values of the perimeters $L\Delta$ and $L\square$ are set out deliberately, as they will be of use in the further analysis of the geometrical factors.

Now, let us consider the coupling between counter-propagating waves. Taking into consideration the Lambert model of scattering on the mirror inhomogeneity, a number of the authors consider [28, 29] that the share of the energy scattered into the counter-propagating wave at an angle of 45° with the mirror (four-mirrored resonator) is much less than at 30° (three-mirrored resonator). As a result, the total back scattering on four mirrors turns out to be less than on three mirrors. It is also noted that for the S-component of linearly-polarized radiation the reflection coefficient is higher for the angle of 45° than for 30° [29].

The potential accuracy of the RLG. There are two main noise sources that affect the potential accuracy of the RLG: spontaneous emission of the active medium and random loss of information in dithering The last source of noises sets the limit of the real accuracy rather than the potential one, and only with dithering used. So attention must be focussed on the estimation of the potential accuracy.

In order to analyze the potential accuracy's of the RLG [30–34], recourse was made to the final results in [35] which considers the spontaneous radiation caused by random phase non-coherent photons. As a consequence, the spectral lines of the counter-propagating waves get wider.

If one of the counter-propagating waves is expressed as a vector E, the spontaneous emission S, adding up to the vector E in the random phase, as shown in Fig.15, results in random variations of the phase $\Delta\psi$. The results of [35] are used below to express the random drift dispersion as

$$D^{w} = R\sqrt{t}\ ,$$

where

$$R = \frac{60}{2\pi K_{imp}}\left(\frac{C}{L}\right)\left(\frac{fh\nu T}{2P_0}\right)^{1/2}\ [\text{deg}/\sqrt{h}\ ];$$

K_{imp} is the scale factor [imp/deg]; γ - the total losses; T – the radiation losses; P_0 - the output power [erg/s] , $f = \left(\frac{n_1 g_2}{n_2 g_1}\right)^{-1}$; $n_1 g_1$ and $g_2 n_2$ are the populations of the lower and upper energy levels, correspondingly.

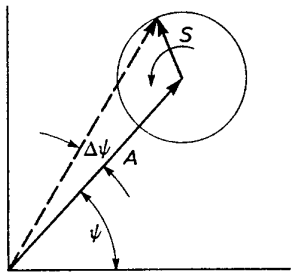


Fig. 15. A vector diagram to the estimation of an RLG potential limit

For the Honeywell GG1300 RLG the measured value of the coefficient R was $1.6 \cdot 10^{-4}$ deg/$\sqrt{h}$ under the following parameters: λ = 0.56 μm; L = 43.4 cm; $K_{imp} = 2.29 \cdot 10^3$; P_0 = 12 μW; $T = 1.6 \cdot 10^{-4}$; $\gamma = 5 \cdot 10^{-4}$.

Dependence of K_{imp} and P_0 on the resonator's perimeter shows that $R \equiv L^{-2/5}$ from where the relation for the random drift coefficients of four- and three-mirrored resonators takes the form :

$$\frac{R_\square}{R_\Delta} = \left(\frac{L_\square}{L_\Delta}\right)^{-5/2} \approx 0.5$$

Thus, the random drift coefficient for the four-mirrored RLG is half as much as that of the three-mirrored RLG.

Reflectors of RLG resonators. The terms a 'three-mirrored' and 'four-mirrored' RLG resonator do not imply that only multilayered dielectric mirrors are used as reflectors.

The designs developed in the former Soviet Union and now in Russia employ totally reflecting prisms, making

it possible to create different configurations of resonators, as shown in Fig 16.

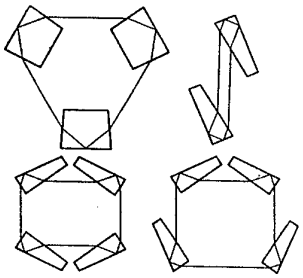


Fig.16. Configurations of RLG resonators with totally reflecting prisms

The chief advantage of prism reflectors is the possibility to use traditional optical technologies in their manufacture which makes them relatively cheap. It should be noted that the present-day high technologies used in manufacturing multilayered dielectric mirrors contribute significantly to the total cost of the RLG. Hence, prism RLGs cost lower than mirrored ones.

However, using prisms involves a number of constructional features. Special methods and aids have to be employed to control the perimeter as it is impossible to use piezoceramic converters. Besides, as the Verdet constant for the totally reflecting prisms is high, it is necessary to take special steps to eliminate the influence of external magnetic fields.
Nevertheless, prism RLGs enable low losses, high resistance of the prisms' inner surfaces to discharge in the active medium, and, as a consequence, sufficient accuracy characteristics and reliability. More detailed information about RLGs on reflecting prisms can be found in [34].

The material of the resonator. Modern RLGs are of monoblock design. With the material chosen properly, they provide a stable perimeter and geometry of the resonator. In this country monoblock units are commonly made of SITALL whereas foreign technologies employ CERVIT or ZERODUR. The ultimate value of the linear expansion coefficient for ZERODUR is $1.3 \cdot 10^{-6}$ 1/°C allowing the RLG, with the built-in system for controlling the perimeter, to be operated in the thermal range from - 55 to + 85°C.

Superlarge perimeter laser gyros

Progress in theory and modern laser technologies made feasible not only high-accuracy RLGs but also miniature devices of 5-12 cm perimeter. Though not highly accurate in comparison with the others, they are low cost and quite competitive with other gyroscopic sensors for missiles and shells of various classes, robots, and small navigation systems.

At the same time the further development of fundamental investigations in the field of the relativity theory, refinement of the Earth's angular motion, influence of the Earth tides and seismic disturbances called for new precision measuring technologies.

In this connection in the early 90s an international group of scientists from the USA, Germany and New Zealand launched the research on the development of a stationary He-Ne superlarge perimeter ring laser (RL) [29, 36-39]. The prototype of the first large monoblock RL was a laboratory model of dimensions 0.75×1 m. It was placed at a depth of 30 m in one of the New Zealand caves with constant climatic conditions. It is on this model, called CI, that the main conceptual issues of developing a precision measuring instrument were worked out.

Before proceeding to the next model of the monoblock ring laser CII, it is appropriate to bring the table of the dependence of the main RL parameters on the perimeter L.

Table 1

Scale factor, resonator's Q-factor	Signal-to-noise ratio (single-frequency mode)	Longitudinal mode density	Random drift, scale factor nonlinearity	$1/f$ noise
L	L^2	L^{-1}	$L^{-5/2}$	L^{-4}

The data given in Table 1 clearly demonstrates the significant improvement of the main RL parameters as the perimeter grows larger.

The monoblock RL (CII) design was developed in Germany by Carl Zeiss. In 1996 it was placed in the same cave in New Zealand. The model CII has a 1×1 m resonator in a massive ZERODUR monoblock of the total mass of about 600 kg (Fig. 17). Unique multilayered $SiO_2 - Ta_2O_5$ mirrors with the total losses of the order of $(3\pm1)10^{-6}$ are used as reflectors of the

resonator. The high-Q factor of the resonator (Q = $3.3 \cdot 10^{12}$) allows a very short plasma length (of the order 10mm) to be used at high-frequency pumping power of only 1 W.

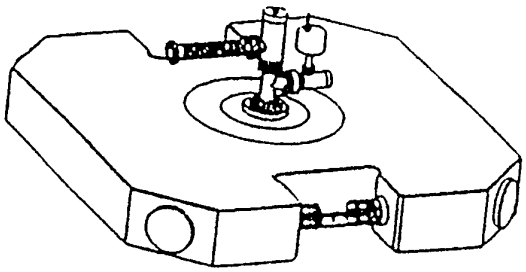


Fig.17. An CII-type RLG

The estimated value for the lock-in frequency is 0.1 Hz (it was 10 Hz for the prototype CI model). The output signal from the diurnal revolution of the Earth at the New Zealand latitude for the RL was 76 Hz. The spectral line band being 0.81 □Hz, it takes more than 21 days to determine 'the gravity center' of the spectral line! The researchers report that the tidal variations of the Earth rate are characterized by time intervals more than 21 days and can be observed with CII.

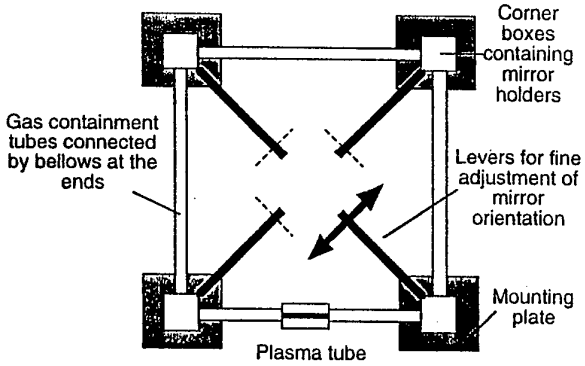


Fig.18. A block diagram of a GO-type RLG

However, the fine structure of the Earth rate variations requires the estimates at the level of 10^{-8}-10^{-9} which is unattainable for the RL CII. So the next step in the development of the superhigh-resolution RL was the completion of the 3.5×3.5 m GO that was placed in the same cave on a vertical concrete wall. A block-diagram of the GO is shown in Fig.18.

The resonator mirrors are mounted in special containers fixed on the wall, whereas the tubes filled with the mixture of He-Ne are connected to the containers with bellows. The diagonal connecting levers are intended for fine adjustment of the resonator mirrors. Fig. 19 represents the spectrum of the GO output signal.

In justice to history it should be noted that it was as early as in 1974 in the former Soviet Union that Kiev Central Design Bureau and the plant 'Arsenal' managed to overcome the problem of estimating angular turns of constructions and direction keeping under seismic disturbances. For this purpose a quantity-produced two-frequency COG-2 RLG with a planar resonator and a differential Faraday NPU was used.

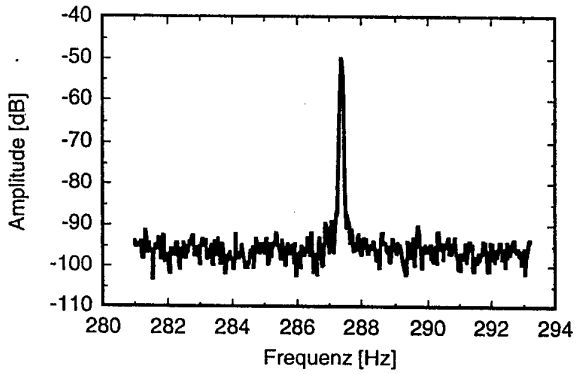


Fig.19. The spectrum of output signal of GO-type RLG

To conclude this section, Tables 2, 3, 4 is given to illustrate the performance data for the RLGs produced by the Research and Production Complex 'Electrooptica' (Moscow), CDB and plant 'Arsenal', Honeywell, Rockwell, Litton and oth. correspondingly.

Table 2

№	Parameter	Units	Research and Production Complex 'Electrooptica'		
			GL-1	GL-2	GL-3
1	Perimeter	cm	45	28	62
2	Mass	kg	4	2	6
3	Dimensions	cm	∅20.6× 10.5	15.4×11.6× 8.8	24.5×19.5× 11.1
4	Material	-	Sitall		
5	Lock-in technique	-	Dither		
6	Number of prisms	-	4	4	4
7	Discharge	-	h.f. transverse		
8	Power	-	24 V/0.6 A	30 V/0.6 A	30 V/0.6 A
9	Max input rate	deg/s	360	500	250
10	Zero bias stability 1-10 h	deg/h	0.007	0.01	0.001
11	Zero bias stability from start to start	deg/ h	0.01	0.01	0.001
12	Random drift	deg / $\sqrt{h}$	0.003	0.005	0.001
13	Scale factor	arc s/imp	1.30	2.24	1.00
14	Scale factor linearity	-	5×10^{-6}	5×10^{-6}	5×10^{-6}
15	Scale factor stability	-	1×10^{-5}	1×10^{-5}	5×10^{-6}
16	Zero bias temperature sensitivity	deg/h/°C, no more than	0.003	0.002	0.001

Table 3

№	Parameter	Measuring units	COG-1 (КОГ-1) (three-channel)	COG-2 (КОГ-2) (three-channel)	FIL (ФИЛ)	TR090 (ТР090)	SG (СГ)	Fanza BLG-56(Фанза БЛГ-56)	Four-frequency
1	RLG perimeter	cm	84	90	40	56	28	56	48
2	RLG mass	kg	62	98	12	28	2	16	2
3	Dimensions	cm	30×19×12	30×19×12	17×10×4	16×16×6	13×16.5×4.7	25×25×21	12×12×12
4	Material	-	Sitall CO-115M						
5	Linearization technique (dithering)	-	Faraday cell (FC)	FC	FC	FC	Mechanical vibration	FC	FC, four-frequency orthogonal polarization
6	Number of mirrors	-	4	5	4	4	4	4	6
7	Number of anodes/cathodes	-	2/1 per 1 channel	2/1 per 1 channel	2/1	2/1	2/1	2/1	2 cathodes, 4 anodes
8	Power	W	120	120	80	80	35	40	40
9	Max angular rate	deg/s	30	60	200	200	200	90	200
10	Zero bias (over 1-10 h)	deg/h	-	-	5 arc min per min	0.5 arc min per min	0.1 - 0.2	0.2 - 0.3 arc min per min	0.05
11	Zero bias (from day to day)	deg/h	-	-	1	0.5	0.02	0.3	0.05
12	Scale factor (SF)	arc s/imp	0.68	0.72	0.8	0.92	0.93	0.23	0.91
13	SF linearity	arc s/imp	10^{-4}	10^{-4}	10^{-4}	10^{-4}	10^{-5}	10^{-4}	10^{-5}
14	SF stability	arc s/imp	10^{-4}	10^{-4}	10^{-4}	10^{-4}	10^{-5}	10^{-4}	10^{-5}
15	Random drift	deg/ $\sqrt{h}$	-	-	0.2	0.2	0.005	0.2	0.01
16	Temperature gradient	deg/h/°C	1.0	1.0	0.5	0.5	0.002	0.5	0.2

Table 4

№	Parameters	Measuring units	Honeywell 11 GG-1342	Rockwell -	Litton LG-2717	Litton LG-2728	Sperry SLIC-7 (3-axis)	Sperry SLG-15	Raytheon RB-25
1	Perimeter	cm	32	43	17	28	20	40	25
2	Mass	kg	1.9	3.63	1.5	1.5	2.3	7.2	-
3	Dimensions	cm	15.7×14.7×5.3	22.9×22.9×7.6	14×11×6.5	14×17.7×5	10.2×11.4	10.2×20.3×27.9	-
4	Material	-	CERVIT	CERVIT	ZERODUR	ZERODUR	CERVIT	CERVIT	CERVIT
5	Linearization technique	-	Dithering				Magnetic mirror		4-frequency Faraday cell
6	Number of mirrors	-	3	3	4	4	3×3	3	4
7	Number of anodes/cathodes	-	2/1	S	2/1	2/1	2/1	2/1	-
8	Power	-	28 V/1.1 A	28 V/0.6 A	3 W	3 W	0.6×3 W	0.6 W	-
9	Maximum angular rate	deg/s	800	-	800	400	1500	750	-
10	Zero bias (1-10 h)	deg/h	0.007	0.01	0.02	0.005	0.2	0.02	0.005
11	Zero bias (from day to day)	deg/h	0.01	0.02	0.04	0.01	0.5	0.05	-
12	Random drift	deg/√h	0.003	0.005	0.015	0.003	0.05	0.006	0.001
13	Scale factor	arc.s/imp	2.0	1.57	3.0	1.8	3.2	1.6	1.5
14	Scale factor linearity		$5 \cdot 10^{-6}$	$10 \cdot 10^{-6}$	$5 \cdot 10^{-6}$	$3 \cdot 10^{-6}$	$200 \cdot 10^{-6}$	$100 \cdot 10^{-6}$	$5 \cdot 10^{-6}$
15	Scale factor stability		$5 \cdot 10^{-6}$	-	$5 \cdot 10^{-6}$	$5 \cdot 10^{-6}$	$200 \cdot 10^{-6}$	$100 \cdot 10^{-6}$	-
16	Temperature gradient	deg/h/°C	0.002	0.02	-	-	0.008	0.002	0.0004

Fiber-optic gyros

The FOGs took a little more than 10 years to advance from the first laboratory models in 1975 to flight tests of the FOG-based navigational system at the end of the 80s. The last decade proved to be the most fruitful as it was marked by the elaboration of the FOG theory, the development of high-quality light-emitting diodes and the components of the integrated optics, all contributing to higher accuracy of the device. At the same time the dimensions, energy consumption and cost were reduced.

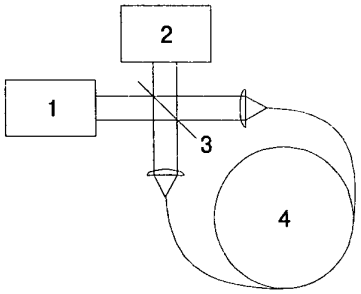


Fig.20. To the principle of FOG operation

The general principle of operation is illustrated in Fig.20 which sketches the basic components of the construction: a light source 1; photodetector 2; beam splitter 3; focusing lenses 4 and 5; interferometer 6 formed of multiturn optical fiber. The light waves introduced into the interferometer after the beam splitter travel through the multiturn interferometer in counter directions (clockwise and counterclockwise). If the FOG is not rotating ($\Omega = 0$), the waves, after travelling the same distance $L = \pi DN$, where D - the coil's diameter, N - the number of turns, after the beam splitter 3 (which now operates as a combiner of the counter-propagating waves) reach the photodetector 2 in phase, with the phase difference of the counter-propagating waves being equal to zero. When the FOG is rotating, the counter-propagating waves travel different distances resulting in the nonreciprocal phase shift

$$\Delta\Phi_s = \frac{2\pi LD}{\lambda c}\Omega,$$

where λ is the wave length of the light source.

The values of the parameters L, D and λ used at present and the FOG's performance data corresponding to them are represented in Table 5.

Table 5

Class of FOG	L, km	D, cm	λ, nm	Sensitivity, $\Delta\Phi_s/\Omega$, s	Drift, deg/h
Middle accuracy	1	10	1550	1.35	0.15
Low accuracy	0.2	3	850	0.15	1.4

The potential limit of the FOG's accuracy is determined by the fluctuation noise which results in the root-mean-square value of the phase noise

$$\overline{\varphi} = \sqrt{\frac{2\hbar f\Delta f}{P_0}},$$

where $\hbar$ – the Planck constant; f – the light wave frequency; Δf – the system bandwidth; P_0 – the light wave power. For the FOG with L = 1 km, D = 10 cm, $f = 3\cdot10^{14}$ Hz it is possible to write

$$\frac{\Omega_{min}}{\sqrt{\Delta f}} \approx 10^{-2} \text{ [deg/h/}\sqrt{\text{Hz}}\text{]} \text{ [40].}$$

Consider the resulting field of the light waves at the input of the photodetector in the form of

$$E_{res} = E_1\exp[j(\omega t + \Delta\Phi_s/2)] + E_2\exp[j(\omega t + \Delta\Phi_s/2)],$$

where E_1 and E_2 – the amplitudes of the counter-propagating waves. For the intensity $I = 0{,}5I_0(1-\cos\Delta\Phi_s)$. Here it is assumed that $E_1 = E_2 = E_0$; $I_0/4 = E_0^2$.

From the expression for the intensity I it is clear that in the FOG's configuration shown in Fig.20 its sensitivity to low values of the angular velocity is determined by the derivative $\frac{dI}{d\Omega_{|\Omega\to 0}} \to 0$.

To increase the sensitivity, the nonreciprocal shift $\pi/2$ between the counter-propagating waves is introduced to the FOG interferometer using the 'reciprocal' phase modulator, as shown in Fig.21.

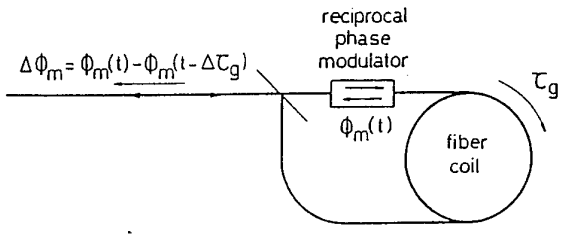


Fig. 21. Introduction of nonreciprocal phase shift with reciprocial phase modulator

Here the wave travelling in the clockwise direction is modulated when entering the interferometer, and the counter-propagating wave – when going out of it . Thus, the counter-propagating waves turn out to be phase modulated with the same signal, shifted in time $\Delta\tau$, which is equal to the difference between the time when the counter-propagating waves reach the modulator. The reverse value $1/\Delta\tau = f_c$ is called the natural frequency of the interferometer [42].

The principle of the phase modulation with the rectangular signals (meanders) is shown in Fig. 22.

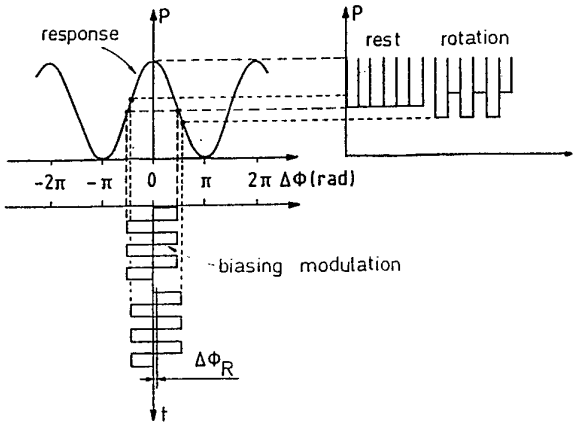


Fig.22. Phase modulation of counter-propagating waves in a FOG

When the FOG is not rotating ($\Omega = 0$), the modulating signal is symmetric about the sinusoidal curve $P(\Delta\Phi)$ and the output signal does not include modulation. The FOG rotation breaks the symmetry, the modulation signal is shifted (to the right in Fig.22) and the output signal is being modulated. The alternating component of the output signal is proportional to the FOG's rate of rotation at the maximum sensitivity.

To reduce the influence of the asymmetry of the modulating signal, the modulation frequency should be selected equal to the natural frequency f_c [43–45] regardless of the form of the modulating signal : either rectangular or sinusoidal.

The nonlinear character of the dependence $P(\Delta\Phi)$, the variations of the intensity of the counter-propagating waves and the depths of modulation cause the harmonics at the modulation frequency which is equivalent to the zero bias of the FOG's input-output characteristic and the change of the scale factor. Such is the case for the change of the input angular rate in a wide range.

A closed-loop signal processing approach is used to solve this problem such that a balancing phase shift is introduced through the negative feedback in order to make the output signal return to the zero operating point, as shown in Fig. 23.

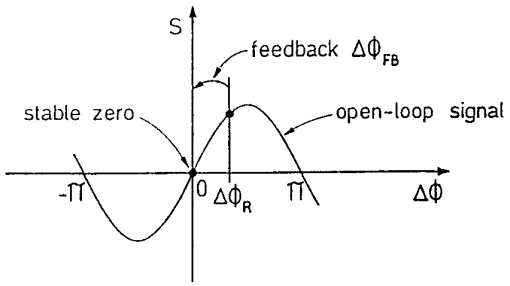


Fig.23. Principal of closed-loop operation of FOG

Different block diagrams can be found in [40, 47].

The term a 'minimum reciprocal configuration' is commonly used for FOGs [40, 41] to describe a FOG architecture with the minimum number of components capable of providing sufficient accuracy. It is not difficult to see that the main peculiarities of the FOG operation are related to reciprocity, the condition when, in the absence of rotation, the counter-propagating waves travel along the same paths in opposite directions retaining specific polarization states and the form of the wave front. For this the light is introduced into the interferometer through the single-mode waveguide filter and leaves it the same way [41]. As a result, the counter-propagating waves which passed through the ring interferometer become strictly cophasal at $\Omega = 0$.

A general block diagram of a FOG of a minimum reciprocal configuration is shown in Fig.24.

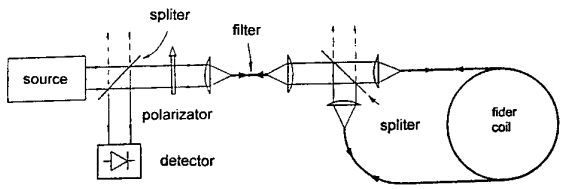


Fig.24. A FOG of the minimum reciprocal configuration

The integral part of most modern FOG architectures is various electronic fast-response components, primarily digital, intended to handle the problems of modulation, synchronous detection, automatic control, etc. On the other hand, the general tendency toward microminiatuarization of multiple optical elements (couplers, combiners, modulators, etc.) and their maximum integration on the basis of integrated optics shows itself quite apparently.

Integrated components of the FOG. The mass production of FOGs using microelectronics technologies calls for production of universal hybrid components that can be employed in different architectures of FOGs, in particular, in those with many types of optical fibers. Indeed, from the general block-diagram of the present-day FOG, shown in Fig.25, one can see that in all the FOGs without exception the light emitted by the radiation source travels to the detector (photodetector) through dividers, linear polarizers and a modulator which can be manufactured on the basis of the hybrid technology.

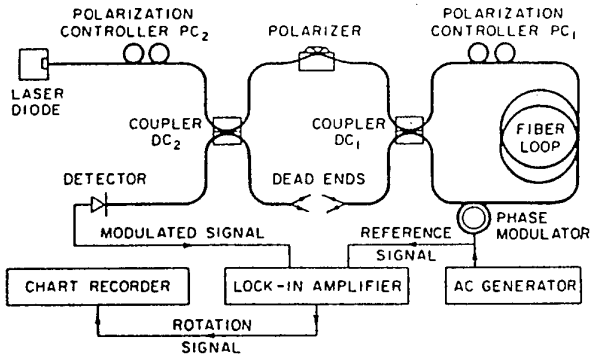


Fig.25. All fiber gyroscope

This design is known abroad as an 'all-fiber' gyro, its 'heart' being two splitters, a polarizer and modulator. If necessary, it can be provided with a mode filter.

As an example let us consider an optimal hybrid configuration of the Y- coupler with a modulator whose block-diagram is shown in Fig.26 [43].

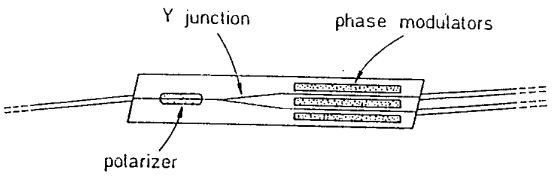


Fig. 26. Multifunctional integrated optical circuit

It is manufactured on an electrooptical lithium niobate substrate ($LiNbO_3$) which provides efficient wideband phase modulation at low control voltages. The lithium substrate makes it possible to fabricate the polarizer with an absorbing metallic overlay or using proton replacement of lithium in the $LiNbO_3$ crystal structure.

The hybrid FOG construction developed by the Jet Propulsion Laboratory in California Institute of Technology has all the elements of a standard FOG, except the multi-turn interferometer, fabricated as integrated optical circuits on lithium niobate substrate: a light source, multiplexer, polarizer, couplers-combiners, frequency-shifting elements, modulator, photodetector [48, 49].

The next step towards microminituarization was the FOG architecture based fully on hybrid integrated microelements developed by Northrop Corporation [50]. In this construction the multi-turn interferometer was replaced by a thin-film single-turn waveguide of titanium applied on the lithium niobate substrate by the diffusion technology.

It is safe to say that this technology opens up the way to the development of a new generation of integrated optical gyros.

Multiple-axis architectures of FOGs. The minimum reciprocal architecture making the most use of integrated optics generally serves as a basic configuration of present-day multiple-axis FOGs.

Let us consider the most important technical performance of a single-axis FOG (FOG-1-B) which was used as a prototype of the multiple-axis gyro [51]:

- as a light source a low-cost multimode laser diode of 820 nm wavelength is used and up to 1.5 mW optical power is launched into the single-mode fiber pigtail. It is mounted in a half-butterfly package with a thermo-electric cooler, temperature sensor and monitor diode;
- the multifunctional integrated-optic module contains a pigtailed IOC with a polarizer, Y-branch and a couple of phase modulators on a z-cut lithium-niobate substrate;
- polarization-preserving fiber used for the pigtails and a 100 m fiber coil allow Sagnac phase shift of approximately 1μrad for 1 deg/h input rate;
- the detector module consists of a Si-PIN photodiode with a preamplifier hybrid circuit;
- square-wave phase modulation at the eigen frequency of 1 MHz with digital demodulation using a 8-bit/4 MHz flash converter provides the required sensitivity improvement. The demodulation technique with the 8-bit converter and without any analog RF mixing and filtering was implemented in 1986 [1];
- a new ASIC with analog and digital functions is now used to generate the modulation signal, to demodulate the detector output, and attain an improved scale factor. Open-loop and closed-loop operation are available. Scale factor calibration as well as modelling of temperature errors of bias and scale factor are performed by a single-chip microcontroller;
- the gyro has an extended built-in test capability, including power-on BIT, initiated BIT, and periodic BIT for monitoring critical parameters during the operation. A minimum number of components is used in the mechanization of the optics and electronics, thus improving the reliability of the device.

The change-over to multiple-axis FOG architectures is based on the conception of maximum integration of separate optical and electronic elements permitting of reduction in energy consumption and size of the module. Thus, the three-axis configuration can be realized with a single common light source (single-source configuration) or with a common light source and detector (single source and single-detector configuration) [51, 52]. Fig 27 illustrates the optical part of the last version.

To separate each of the three channels, phase modulation is realized at different frequencies f, f_2, f_3 with the corresponding filters. However, as the frequencies are different, they need matching to the length of fiber-optical interferometers, and each channel has a different length of light guides.

The field of application of three-axis FOGs, particularly for the integrated SINS/GPS/GLONASS, extends from aircraft and missile flight control to navigation systems of mobile vehicles, practically with no exceptions. As an example, Table 6 gives the performance data for the FOGs produced by ALCATEL.

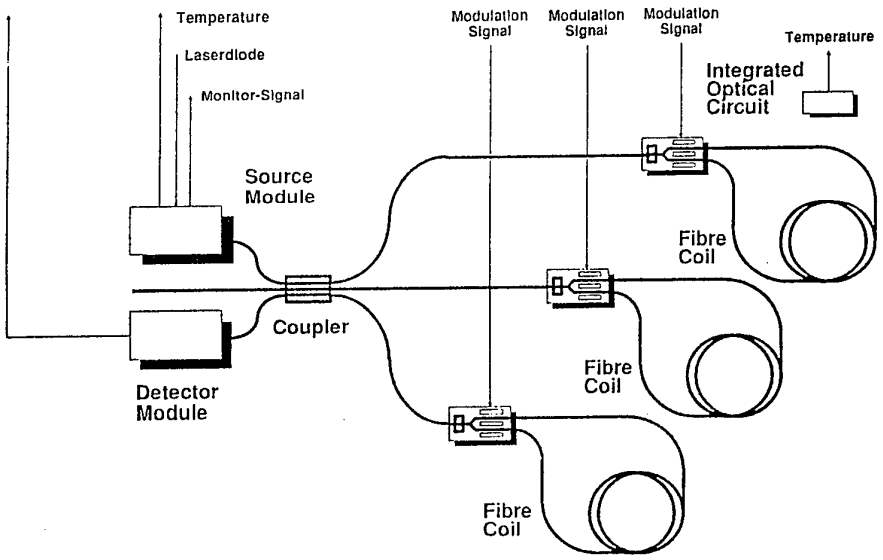


Fig.27. A block diagram of a three-axis FOG

Table 6

Configuration	Bias stability	Scale factor error	Measurement range, deg/h	Application
Rate gyro package (1 axis). 1 FOG	0.1-0.3 deg/h	0.5-1%	100	Single-axis stabilization, angular velocity sensor heading reference for ground vehicle navigation
Rate gyro assembly (3 axis). 3 FOGs or a FOG triad	0.1-0,3 deg/s	0.5-1%	100-400	Multi-axis stabilization, flight control
Inertial measurement unit. 3 FOGs or a FOG triad, 3 accelerometers, IMU processor	1-3 deg/h	$(300\text{-}500)10^{-6}$	400-800	Subsystem for integrated systems: SINS/GPS
Attitude/heading reference system. 3 FOGs or a FOG triad, 3 accelerometers, 1 magnetometer, system processor	1-3 deg/h	$(300\text{-}500)10^{-6}$	400-800	Attitude and heading of a vehicle. Magnetic north reference
Integrated GPS-aided inertial navigation system. 3FOGs or a FOG triad, 3 accelerometers, 1 magnetometer, 1 GPS receiver, system processor	1 deg/h	$(100\text{-}300)10^{-6}$	400-800	Complete position and navigation information ; attitude and heading velocity

An interesting solution in the production of miniature FOGs was found and realized by specialists of 'Fizoptica' (Moscow). The technology they offered uses one light guide, with all the optical components (fiber-optical interferometer, modulator, Y-coupler, polarizer) arranged along it. Thus, there is no need for optical

connectors between the elements which reduces losses and improves reliability. Small-sized cheap FOGs produced by this technology found their place on the domestic and foreign markets. The performance data of the gyros serially produced by 'Fizoptica' are presented in Table 7.

Table 7

Type of FOG	Range, deg/s	Repeatability		Zero bias stability	Output noise, deg/√h/√Hz	Mass g	Application
		Scale factor, %	Zero bias				
VG910	100-300	0.2-0.4	20-60 deg/h	5-15 deg/h	3-6	110	Ground-based navigation, stabilization of VCRs, SINS/GPS/ GLONASS
VG941 Ø2.5×5 mm	up to 500	0.2-0.4	0.1 deg/s	0.01-0.03 deg/s	–	40	Motion control, dynamic tests, inspection
VG951	40	0.2	1-3 deg/h	0.3-1 deg/h	1	650	Motion control, navigation

Progress in FOGs. All the FOGs produced at present for various purposes can be conventionally divided into three categories:
medium-class FOGs with a drift from 0.1 to 5 deg/h and scale factor stability of ~ 10^{-3} for tactical missiles, aircrafts, autonomous vehicles and marine gyrocompasses;
FOGs for navigation systems accurate enough to determine position and orientation when integrated with a triad of accelerometers. They call for the devices better than 0.006 deg/h and scale factor stability at the level of $10 \cdot 10^{-6}$ in the range of angular velocities of up to 1000 deg/s;
precision FOGs for special-purpose aircrafts and submarines with a drift much better than 0.001 deg/h and very low random drift.

Whereas the attainment of the technical requirements to the medium-accuracy FOGs (0.1 - 5) deg/h is not a serious problem, in order to meet the requirements of the two last classes it is necessary to modify the already existing and develop new architectures. Two trends can be distinguished here. The first makes use of the traditional conception of the reciprocal FOG with a light guide capable of retaining polarization of counter-propagating waves. In 1996 A. Cordova and others [53] reported the development of a FOG with a drift of 0.0081 deg/h in the temperature range from – 71 to +54 °C. However the high cost of this FOG, mainly due to the cost of the fiber-optical interferometer, made the specialists from Honeywell employ standard single-mode fiber and the conception of the depolarized FOG. It should be noted the FOG's scale factor fluctuations depend to a great extent on the fluctuations of polarization in the light guide. When the counter-propagating beams propagate from the input to the output, the polarization state can change in a random way. The value of the interference signal is dependent on the relative orientation of polarizations. In the worst case two beams come to the photodetector with orthogonal polarizations and the output signal will be equal to zero. To eliminate this phenomenon, depolarizers are installed at the ends of the interferometer, as shown in Fig.28. They are intended to distribute the light waves randomly to all possible states of polarization.

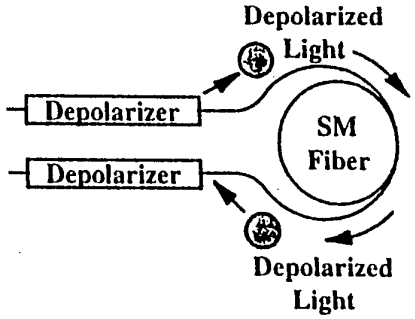


Fig.28. To the concept of the depolarized FOG

The depolarization efficiency can be improved by using a wideband light source. The FOG architecture developed by Honeywell employs either superluminescent diodes or a superfluorescent erbium-doped fiber source to produce broadband light in the 1.53 μm to 1.56 μm wavelength range.

All this led to the development of a precise FOG for spacecraft and submarine navigation. With 4 km fiber and a coil ∅ ~14 cm the random angular drift was

about 0.00019 deg/√h and the zero drift better than 0.0003 deg/h.

Conclusion

Many urgent problems like methods and procedures used to increase the resolution of RLGs and FOGs [54 – 56], prediction of their accuracy characteristics and adjustment following the results of the study of an 'empty' resonator[19, 57], pioneer investigations in the development of microminiature integrated optical gyroscopes and others have not been considered in the paper because of the natural volume limitations of the publication. The papers given in the references and particularly the last publications will be able to partially fill the vacuum.

The author is making grateful acknowledgement to Director – the Chief Designer of the Central Design Bureau 'Arsenal' (Kiev) V.I. Buzanov and Directors of 'Electrooptica' G.N.Zyuzev and 'Fizoptica' V.N.Logozinsky for supplying the information for this paper.

REFERENCES

1. **T.Cunningham.** Dual use: Opportunities, Payoffs and challenges, AGARD conference Proc. №556 Dual usage in Military and Commercial Technology in Guidance and Control, March 1995.
2. **H.Lefevre, P.Martin, J. Morrise, P.Simonopietri, P.Viveont, H.J.Ardity.** High dynamic range fiber gyro with all-digital processing, SPIE. Proc., Vol. 1367, 1990.
3. **R.-Y.Liu, T.F.EI-Waily, M.Olson, G.W.Adams, J.E.Caroll.** Miniaturized, low-cost FOG IMU for tactical weapon applications, SPIE Proc., Vol.2837, 1996.
4. **K.Sakuma.** Application of fiber gyros at JAE. SPIE Proc., Vol.2837, 1996.
5. **M.Ribes, G.Spahlinger, M.W.Kemmler.** 0,1 deg/h DSP - controlled fiber optic gyroscope. SPIE Proc., Vol.2837,1996.
6. **G.A.Sanders, B.Szafraniec, Y.Liu, Laskoskie, L.Stanford and G.Weed.** Fiber Optic Gyros for space, marine and aviation applications. SPIE Proc., Vol.2837, Fiber Optic Gyros: 20th anniversary conference, 1996.
7. **K.Killian, M.Burmenko, W.Holliger.** High Performance Fiber Optic Gyroscope With Noise Reduction, SPIE Proc., Vol.2292, July, 1994.
8. **F.Aronowitz.** The Lazer Gyro. In Lazer Applications. Vol.1, Editor M.Ross, Academic Press, New York, 1971.
9. **A.Matthews and H.Welter.** Cost - Effective, High - Accuracy Inertial Navigation, Navigation Journal of the Institute of Navigation Vol.36, №2, Summer, 1989.
10. **The New High** Accuracy Ship's Inertial Navigation System LITEF PL41 MK4. Naval Forces. Vol.X № V/1989.
11. **Бычков С.И., Лукьянов Д.П., Бакаляр А.И.** Лазерный гироскоп. М.: Сов. Радио, 1975
12. **Лукьянов Д.П., Рогачев А.Ф.** Устройство для создания невзаимного фазового сдвига линейно-поляризованных колебаний. Авт. Свид. №1208029/26-25, Б.И.№14, 1973.
13. **H. de Lang.** Eigenstates of Polarization in Lasers. Phillips Res. Repts 19, 1964.
14. **G.B.Yutema, D.C.Grant, Jr. and R.T.Warner.** Differential Laser Gyro System. U.S. Patent, 3, 862, 803, January 1975.
15. **J.Breman, H.Cook and D.Lysobely.** Differential Laser Gyro Development. Navigation: Journal of the Institute of Navigation, Vol.24, №2.
16. **Лазерный гироскоп/Мельников А.В., Просветов В.К., Рыбаков Б.В., Скулаченко С.С., Хромых А.М., Юдин И.Н.//Патент СССР № 745542, приоритет от 16.05.67.**
17. **I.Smith and T.Dorschner.** Electromagnetic Wave Ring Resonator. U.S. Patent 4.110.045. October 1980.
18. **T.Dorschner and I.Smith.** Clear - Path Four - Frequency Resonator for Ring Laser Gyros, J. Opt. Soc. of Amer., Vol. 68, 1978.
19. **Азарова В.В., Голяев Ю.Д., Дмитриев В.Г., Дроздов М.С., Казаков А.А., Коллодный Г.Я., Мельников А.В., Назаренко М.М., Свирин В.Н., Тихменев Н.В.** Зеемановские лазерные гироскопы. Состояние и перспективы развития //Гироскопия и навигация. - 1997.- №4.
20. **C.H.Volk, I.Longstaff, J.M.Canfield and S.C.Gillespie.** Litton's Second Generation Ring Laser Gyroscope. Proceedings of the 15th Annual Biennial Guidance Test Simposium, 1991
21. **Mc.Clure, R.E.Vaher.** An Improved Ring Laser Bias Element. Proceedings of the IEEE 1978 National Aerospace and Electronics Conference.
22. **W.M.Macek.** Improvements in or Relating to Ring Lasers Patent Specification GB 1406730.
23. **M.Voigt.** Magneto und electrooptic. Druck und Verlag von B.G.Teubner, Leipzig, 1908.
24. **B.Stieler and H.Winter.** Gyroscopic Instrumentation and Their Application to Flight Testing. AGARDograph № 160, vol. 15, 1982.
25. **А.П.Войтович, В.Н.Севериков.** Лазеры с анизотропными резонаторами. - Минск: Наука и техника, 1988.
26. **T.A.Dorshner, I.W.Smith, H.A.Statz.** A Maganese Bismuth Magnetic Mirror for the Raytheon Laser Gyroscope/Proc. IEEE Nat. Earosp. and Electronic: Conf. «NAECON»78, Dayton, 1978, N.Y. 1978, vol.1-3
27. **W.L.Lim, J.P.Hauck and J.W.Raquet,** U.S. Patent №4.705.398 (11.10.87).
28. **R.R.Simpson and R.Hill.** Ring Laser Gyro Geometry and Size. Roy. Aeron. Soc. London, U.K., 15

Feb., 1987.
29. **H.R.Bilger, V.Schreiber, G.E.Stedman.** Design and Application of Large Perimeter Ring Laser. Proc. of Gyro Technology Symposium, Stuttgart, 1996.
30. **W.W.Chow, J.Gea-Banacloche, L.M.Perotti, V.E.Sanders, W.Schleich and M.O.Scully.** Rev. Mod. Phys. 57,61, 1985.
31. **J.N.Simpson,** Proc. IEEE NAECON, 1980
32. **T.Dorschner, H.Haus, M.Holz, I.Smith and H.Statz.** IEEE J. Quantum Electr., QE 16, 1980.
33. **M.Tehrani.** Unpublished analysis at Honeywell, 1979.
34. **Optical** Gyros and Their Applications. Edited by D.Loukianov, H.Sorg and B.Stieler, AGARDograph. (в печати).
35. **F.Aranowitz, et al.** Proc. Symposium Gyro Technology, DGDN, Stuttgart, 1982.
36. **Z.Li.** Optical Supercavity And Precision Ring Laser Measurement. Ph.D. dissertation, University of Canterbury, Christchurch, New Zealand, 1993.
37. **R.J.Neutze, G.E.Stedman and H.R.Bilger.** Reanalysis of laub drag effects in a ring cavity. J. Opt. Soc. Am. B.13, 1996.
38. **R.Rodloff.** Gibtes den optischen Superkreisel? Zeitschrift fur Flugweissenschaft und Weltraumforschung 18, 1994.
39. **H.R.Bilger, G.E.Stedman, M.P.Poulton, C.H.Rowe, Z.Li, P.V.Wells.** Ring laser for precision Measurement of nonreciprocil phenomena. IEEE Trans. Instr. Meas. IM-42, 1993
40. **Шереметьев А.Г**. Волоконный оптический гироскоп. - М.: Радио и связь, 1987.
41. **H.Lefevre** The fiber-optical gyroscope, Artech House, Boston-London, 1993.
42. **J.M.Martin and J.T.Winkler**. Fiber optic lazer gyro signal detection and processing technique. SPEI Proc. vol.139, 1978.
43. **H.C.Lefevre.** Application of the Sagnac Effect in the Interfometric Fiber Optic Gyroscope. Optical Gyros and their Applications. AGARDograph (в печати).
44. **H.C.Lefevre.** Comments about fiber-optic gyroscopes. SPEI Proc. vol.838, 1987.
45. **R.Ulrich**. Fiber optic rotation sensing with low drift. Optic Letters, vol.5, 1980.
46. **B.Gulshaw, J.P.Giles**. Fiber-optic gyroscopes. - J. Phys.E: Sci.Instrum.Vol.16, 1983.
47. **J.G.Mark and D.A.Tazartes**. A survey of optical gyroscopes. Proc. of the 2nd St.Petersburg International Conference on Gyroscopic Technology and Navigation, 1995.
48. **W.C.Goss, R.Goldstein**. Fiber Optic Rotation Sensor Technology - IEEE/OSA Conference of Laser Engineering and Applications, 1979.
49. **W.C.Goss, R.Goldstein**. Integrated optics for fiber gyros. - Proc. SPEI, Vol.412, 1983
50. **J.R.Haavisto**. Thin-film waveguides for inertial sensors. - Proc. SPEI, Vol.412, 1983.
51. **W.Auch**. Progress in fiber-optic gyro development and application. SPEI, Vol.2070. Fiber Optic and Lazer Sensors XI, 1993.
52. **G.A.Sanders and B.Szafraniec**. Progress in Fiber Optic Gyroscope Applications II with Emphasis on the Theory of Depolarized Gyros. AGARDograph Optical Gyros and their Applications (в печати).
53. **A.Cordova, R.Patterson, J.Rahu, L.Lamb, D.Rozelle**. Progress in navigation grade FOG performance. Proc. SPEI, Vol2837, 1996.
54. **Yu.V.Filatov, D.P.Loukianov, and R.Probst**. Dynamic angle measurement by means of a ring lazer. Metrologia, 1997, 34.
55. **M.N.Burnachev, Yu.V.Filatov, K.E.Kirianov, D.P.Loukianov, A.F.Mezentsev, and P.A.Pavlov**. Precision angle measurement in a difractional spectrometer by means of a ring laser. Measurement Science and Technology, V.9, No.7, July 1998.
56. **J.G.Mark and D.A.Tazartes**. A Resolution Enhancement Technique for Laser Gyros. 4th St.Petersburg Intern. Conference on Integrated Navigation Systems, May 26-28, 1997.
57. **M.M.Techrani**. Equivalence of Passive and Active Lock-in Rates in Ring Laser Gyro Symposium Gyro Technology, Stuttgart Germany, 1998.

Application of Coning Algorithms to Frequency Shaped Gyro Data

J.G. Mark and D.A. Tazartes

Litton Guidance and Control Systems Division
5500 Canoga Avenue
Woodland Hills, California, 91367
USA

ABSTRACT

A key parameter of a strapdown inertial navigation system is its response to coning motion. Substantial efforts have gone into the development of sophisticated algorithms which reduce system drift errors in the presence of coning motion. Present-day algorithms use incremental angle outputs from the gyros to form high-order correction terms which reduce net coning errors. These algorithms assume a flat transfer function for the processing of the incremental angle outputs and are structured to yield very high order responses. Techniques such as resolution enhancement (the subject of a previous conference paper) shape the frequency response of the gyro data and consequently degrade the performance of the conventional coning algorithms. Likewise, many gyros exhibit complex frequency responses and violate the assumptions used in deriving the previously developed coning algorithms. The mismatch between the assumed and actual frequency response of the gyro data leads to degradation of performance in a coning environment as well as amplification of pseudo-coning errors. This paper discusses a method of deriving algorithms which are tailored to the frequency response of the particular type of gyros used. These algorithms can be designed to arbitrarily high order and can also supply an extremely sharp high-frequency cutoff to minimize pseudo-coning errors. This work was motivated by the desire to use resolution enhanced ZLG data to form the strapdown attitude solution and was heavily influenced by the most recent work in Russia on coning algorithms by Yury Litmanovich. However, the techniques developed equally apply to mechanical, fiber-optic, and other types of gyros. Extensive simulation of the new algorithms has been performed and we are now in a position to incorporate them in the Litton ZLG product line.

INTRODUCTION

In a strapdown inertial navigation system, angular rotation measurements are processed and integrated to form an attitude quaternion or matrix which describes the attitude of the system with respect to a stabilized reference coordinate system. Attitude integration is complicated by the non-commutativity of rotations and by finite instrument sampling rates. This is not a problem in the case of rotation about a single, unchanging axis. However, if the axis of rotation changes, it can be easily shown that the attitude of a body depends not only on the magnitude but also on the order of the rotations performed. If the order is not properly tracked then attitude errors will result and navigation performance will be seriously degraded. Coning motion is one particular input used to evaluate strapdown inertial navigation systems and attitude integration algorithms. In a coning motion, the axis of rotation is constantly changing directions. Coning is a very demanding type of motion and algorithms which operate satisfactorily in this environment will satisfy most other requirements. In coning motion, one (or more) axes of the system sweeps out a cone in space. The Goodman-Robinson theorem [1] states that the output of a gyro whose sensitive axis is along a coning axis will be equal to the solid angle swept by this axis, even though there is no net rotation produced. An appropriate attitude integration algorithm can process three axes of information from orthogonal gyros and re-constitute the actual motion. However, an additional complication lies in the fact that in practice, instrument (gyro) outputs are only sampled at finite rates. The sampling bandwidth limitation must therefore be taken into account. In the past, coning algorithms have been used to improve attitude integration algorithms using sampled gyro data. These algorithms are discussed in many different references such as [2], [3], [4], [5], [6], [7]. However, these publications deal with ideal gyroscopes. In many instances, gyro outputs have a shaped frequency response or are filtered to reduce other errors such as quantization [8] or vibration-induced errors. In these cases, the traditional coning algorithms do not work to high order accuracy. In 1997, when Reference [8] (Resolution Enhancement for Laser Gyros) was presented, Y. Litmanovitch asked how do you deal with coning algorithm response with Resolution Enhancement. At that time we didn't; this paper presents an approach to solving the problem and generalizes to other types of frequency shaping.

CONING ALGORITHMS

The usual method of generating high-order coning algorithms (i.e. algorithms which approach ideal coning response) begins with a postulated coning motion about

Paper presented at the RTO SCI International Conference on "Integrated Navigation Systems", held at "Elektropribor", St. Petersburg, Russia, 24-26 May 1999, and published in RTO MP-43.

the z axis as described in the first part of the Appendix. Each attitude update requires a vector angle $\Delta\phi$ which is a single-axis rotation describing the net attitude change from the beginning to the end of the interval in question. The problem lies in the fact that the gyro incremental angle outputs $\Delta\theta$ are only an approximation to the components of the required vector angle. In order to improve the approximation, each attitude integration interval is divided into m sub-intervals of equal duration. Gyro data samples from the sub-intervals are processed to yield a high order approximation of the vector angle. The correction $\Delta\phi$ - $\Delta\theta$ is formed by combining cross products of the $\Delta\theta_i$ of the m sub-intervals. Figure 1 shows an example for the case m=4.

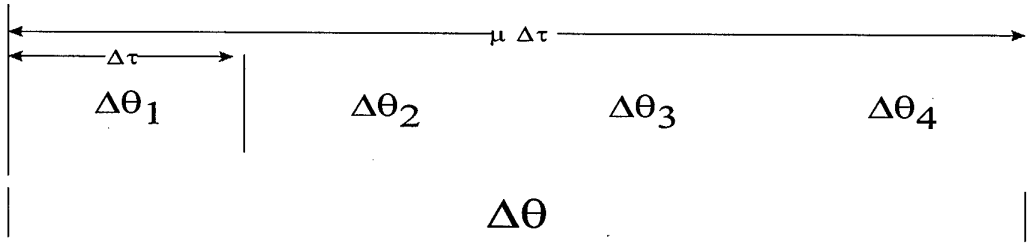


Figure 1: Gyro Data Intervals

In this case, there are 6 possible vector cross-products that can be formed and these are grouped in three different categories depending on the spacing between the data samples:

Spacing 1: $\underline{\Delta\theta_1} \times \underline{\Delta\theta_2}$; $\underline{\Delta\theta_2} \times \underline{\Delta\theta_3}$; $\underline{\Delta\theta_3} \times \underline{\Delta\theta_4}$
Spacing 2: $\underline{\Delta\theta_1} \times \underline{\Delta\theta_3}$; $\underline{\Delta\theta_2} \times \underline{\Delta\theta_4}$
Spacing 3: $\underline{\Delta\theta_1} \times \underline{\Delta\theta_4}$

In general, there are m-1 possible spacings and C_m^2 possible vector cross-products. Each category of cross-products is described by the quantity $C_p(n)$ where:

$$C_p(n) = \left(\Delta\vec{\theta}_{nm} \times \Delta\vec{\theta}_{nm+p}\right)_{\vec{k}\,\text{component}} \qquad (54)$$

It can be shown that all cross-products with the same spacing have the same $\vec{k}$ component and that the result for $C_P(n)$ is independent of n. The AC components (i.e. $\vec{i}$ and $\vec{j}$) have only higher order terms and can be ignored. The coning correction is formed by taking a linear combination of C_P's which approximates the difference between $\Delta\phi$ and $\Delta\theta$. Compensation can be achieved to an order commensurate with m. That is, the relative residual error can be up to the $(2m)^{th}$ power of frequency. Methods of solving for the coefficients of the C_P's are given in several references including [4] and [5]. These methods expand the coning error ($(\Delta\phi - \Delta\theta)_{k^{th}\,\text{component}}$) and cross-product families (C_P) into Taylor series expressed in terms of powers of frequency. The first m terms of the coning error series are equated to the corresponding terms of a linear combination of cross-poducts and the coefficients applying to those cross-products are solved for. Thus, the error corresponding to the first m terms in the coning error expansion are completely compensated. For example, for a four sample algorithm (i.e. m=4), the ideal coning correction is given by:

$$\Delta\phi \approx \Delta\theta + \frac{1}{105}\,(214\ \text{VCP}_1 + 92\ \text{VCP}_2 + 54\ \text{VCP}_3)$$

where:

$$\text{VCP}_1 = a\Delta\theta_1 x\Delta\theta_2 + b\Delta\theta_2 x\Delta\theta_3 + c\Delta\theta_3 x\Delta\theta_4 \qquad a+b+c=1$$
$$\text{VCP}_2 = d\Delta\theta_1 x\Delta\theta_3 + e\Delta\theta_2 x\Delta\theta_4 \qquad d+e=1$$
$$\text{VCP}_3 = f\Delta\theta_1 x\Delta\theta_4 \qquad f=1$$

Table 1 summarizes the VCP coefficients for the first five algorithms. The corresponding coning error relative residual plots are given in Figure 2.

TABLE 1: STANDARD CONING ALGORITHM COEFFICIENTS

NUMBER OF SUBINTERVALS (M)	ORDER	CROSS-PRODUCT DISTANCE	CROSS-PRODUCT COEFFICIENT	RESIDUAL ERROR COEFFICIENT
1	2	N/A	N/A	- 1 / 3 !
2	4	1	2 / 3	- 4 / 5 !
3	6	1	27 / 20	- 36 / 7 !
		2	9 / 20	
4	8	1	214 / 105	- 576 / 9 !
		2	92 / 105	
		3	54 / 105	
5	10	1	1375 / 504	- 14400 / 11 !
		2	650 / 504	
		3	525 / 504	
		4	250 / 504	

Note that the algorithm frequency response follows the (2m) order discussed above. Furthermore, note the algorithms essentially cut-off at the Nyquist frequency ($f\Delta t = \frac{1}{2}$) and some overshoot (error amplification) occurs between the Nyquist frequency and the sampling frequency. This results from the usual aliasing problems occurring with sampled data. In general, we would like as sharp a cutoff as possible to maintain coning rejection at lower frequencies along with the minimum amplification of errors above the Nyquist frequency. Some instruments also exhibit pseudo-coning (i.e. an erroneous high-frequency output which does not reflect true physical motion). It is therefore desirable to sharply cut-off the frequency response of the coning algorithm in order to avoid compensation of an apparent (not real) coning motion.

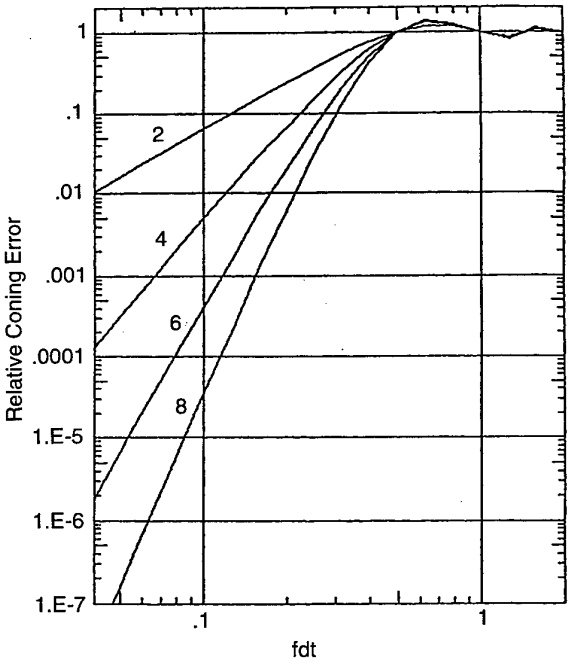


Figure 2: Standard Coning Algorithm Response

Resolution Enhanced Laser Gyroscope

Reference [8] discusses a method of resolution enhancing laser gyroscope data. What is described is a method of reducing quantization noise by using a digital anti-aliasing filter. This filter which operates with high speed sampling therefore also shapes the frequency response of the data. This affects the series expansion of the C_p's, thus requiring different coefficients in the coning algorithm.

The Appendix develops the mathematical basis for the calculation of the coning algorithm coefficients for a wide class of data frequency responses of which the resolution enhancement filter is a special case. The structure of the algorithms remains the same while the coefficients need to be tuned for the particular filter function in use. Table 2 shows the coefficients needed for the first five possible coning algorithms with resolution enhanced data.

TABLE 2: CONING ALGORITHMS FOR RESOLUTION ENHANCED DATA

NUMBER OF SUBINTERVALS (M)	ORDER	CROSS-PRODUCT DISTANCE	CROSS-PRODUCT COEFFICIENT	RESIDUAL ERROR COEFFICIENT
1	2	N/A	N/A	- 30 / 5 !
2	4	1	3 / 4	- 294 / 7 !
3	6	1	124 / 80	- 4920 / 9 !
		2	33 / 80	
4	8	1	17909 / 7560	- 124456 / 11 !
		2	5858 / 7560	
		3	3985 / 7560	
5	10	1	193356 / 60480	- 4638816 / 13 !
		2	66994 / 60480	
		3	65404 / 60480	
		4	29762 / 60480	

Figure 3 shows the relative coning error response for these algorithms as a function of frequency. Note that the residual error coefficients are slightly larger than in the case of unfiltered data but only marginally so. The coning response is again seen to cut-off at the Nyquist frequency. However, the cutoff is much sharper with much less overshoot (i.e. overcompensation) beyond the Nyquist frequency. These algorithms therefore take advantage of the band-limiting provided by the data pre-filtering thereby avoiding aliasing as well as pseudo-coning errors. In order to compare the coning algorithm frequency response (as opposed to relative error response), Figures 4 and 5 were generated.

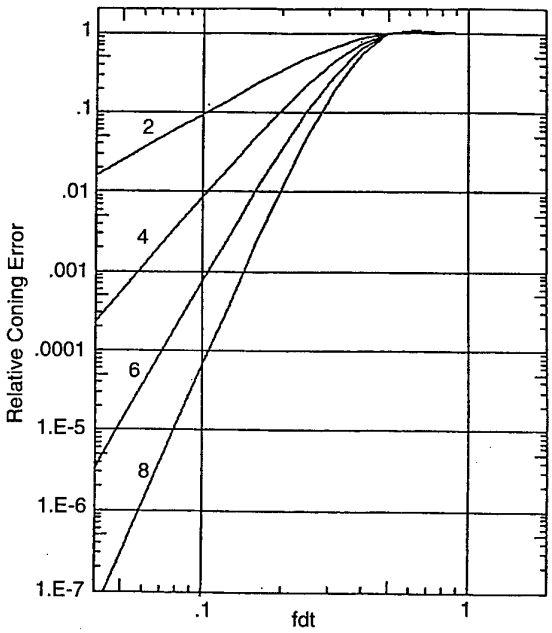


Figure 3: Coning Error Response for Algorithms Tuned for Resolution Enhanced Data

Figure 4 shows the coning algorithm response for the 4 sample 8th order algorithm operating on unfiltered data. Figure 5 shows the 4 sample, 8th order algorithm response using resolution enhanced data. Comparison of these two figures clearly illustrates how the filtered data algorithm provides a much sharper cutoff.

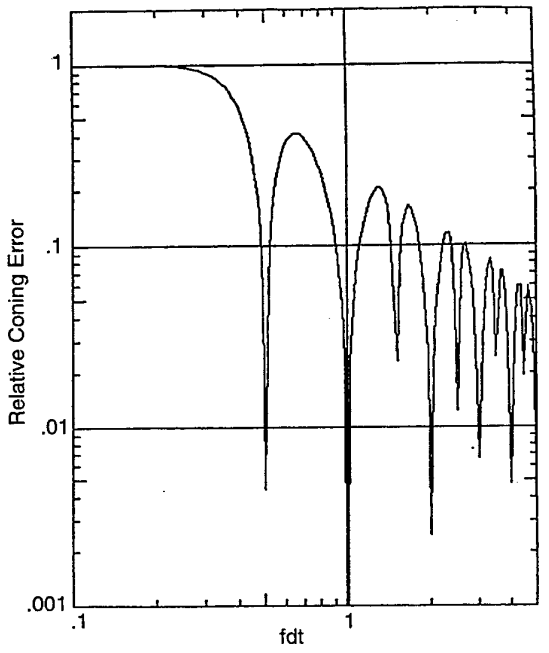


Figure 4: Standard Coning Algorithm Responses

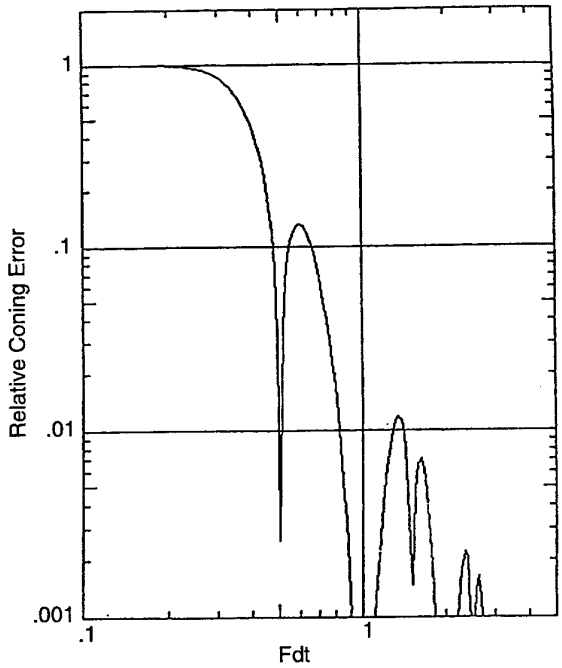


Figure 5: Resolution Enhanced Coning Algorithm Response

Simulation

In order to verify the proper operation of the coning algorithms, a simulation was constructed and a coning motion was applied to the algorithms. The results agree with the predictions. In fact, Figures 1 and 2 are the result of simulation and the slopes agree with the theoretical predictions given in the last column of Table 1 and 2 respectively. In addition, an 8th order algorithm (m=4) simulation with very high coning rates was also attempted. This case is shown in Figure 6 for the resolution enhanced data condition.

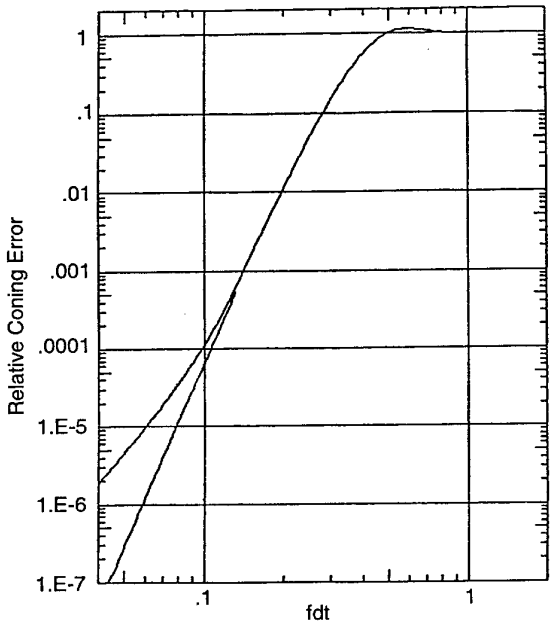


Figure 6: Resolution Enhanced Coning Error Response For Large Motions

This figure shows excellent agreement with the lower coning rate case with deviations only occurring at low frequency where the angle excursions of the cone are extremely large. Even so, the algorithm does not break down but smoothly transitions to a lower order.

Conclusion

The above discussion answers the question posed by Dr. Litmanovich. Coning algorithms for resolution enhanced (or, more generally, filtered) data can be implemented in exactly the same form as for un-filtered data. The only change required is the tuning of the vector cross-product coefficients to match the gyro data frequency response. The use of filtered data with coning algorithms presents the additional advantage of a very sharp cutoff at the Nyquist frequency with substantially reduced overshoot and overcompensation at frequencies above Nyquist.

References

[1] Goodman, L.E., and A. R. Robinson. *Effect of Finite Rotations on Gyroscopic Sensing Devices.* ASME Paper 57-A-30, 1957, 4 pp.

[2] Bortz, J.E. *A New Mathematical Formulation for Strapdown Inertial Navigation.* IEEE Transactions on Aerospace and Electronic Systems, Vol. AES-7-1, January 1971, pp. 61-66.

[3] Miller, R.B. *A New Strapdown Attitude Algorithm.* Journal of Guidance, Control, and Dynamics, Vol. 6, No. 4, 1983, pp. 287-291.

[4] Ignagni, M.B. *Optimal Strapdown Attitude Integration Algorithms.* Journal of Guidance, Control, and Dynamics, Vol. 13, No. 2, 1990, pp. 363-369.

[5] Lee, J.G., Y.J. Yoon, J.G. Mark, and D.A. Tazartes. *Extension of Strapdown Attitude Algorithm for High-Frequency Base Motion.* Journal of Guidance, Control, and Dynamics, Vol. 13, No.4, July-August 1990, pp. 738-743.

[6] Litmanovich, Y. A. *Use of Angular Rate Multiple Integrals as Input Signals for Strapdown Attitude Algorithms.* Proceedings of Symposium Gyro Technology, September 1997.

[7] Savage P.G. *Strapdown Inertial Navigation Integration Algorithm Design Part 1: Attitude Algorithms.* Journal of Guidance, Control, and Dynamics, Vol. 21, No. 1, January-February 1998, pp. 19-27.

[8] Mark, J.G. and D.A. Tazartes. *Resolution Enhancement Technique for Laser Gyroscopes.* Proceedings of 4th International Conference on Integrated Navigation Systems, St. Petersburg, Russia, 1997.

APPENDIX: Derivation of Coning Algorithms

A pure coning motion can be represented in terms of the vector angle $\vec{\phi}$ where

$$\vec{\phi} = \varepsilon \sin(\omega t)\, \vec{i} + \varepsilon \cos(\omega t)\, \vec{j} \tag{A-1}$$

where ε is the cone half-angle, ω is the angular frequency, and t is time. The unit vectors $\vec{i}$ and $\vec{j}$ are orthogonal and normal to the Euler axis. The coning motion can also be represented by the quaternion $\tilde{q}$ where

$$\tilde{q} = [\cos\frac{\varepsilon}{2},\ \sin\frac{\varepsilon}{2}(\sin(\omega t)\, \vec{i} + \cos(\omega t)\, \vec{j})] \tag{A-2}$$

The angular rate vector $\vec{\Omega}$ can be found using the quaternion differential equation

$$\frac{d\tilde{q}}{dt} = \frac{1}{2}\tilde{q}\vec{\Omega} \tag{A-3}$$

or

$$\vec{\Omega} = 2\tilde{q} * \frac{dq}{dt} \tag{A-4}$$

where the asterisk denotes the conjugate of the quaternion wherein the vector components are reversed in sign.

$$\vec{\Omega} = 2[\cos\frac{\varepsilon}{2},\ -\sin\frac{\varepsilon}{2}(\sin(\omega t)\, \vec{i} + \cos(\omega t)\, \vec{j})]\ \omega[0,\ \sin\frac{\varepsilon}{2}(\cos(\omega t)\, \vec{i} - \sin(\omega t)\, \vec{j})] \tag{A-5}$$

$$\vec{\Omega} = \omega[0,\ \sin\varepsilon\cos(\omega t)\vec{i} - \sin\varepsilon\sin(\omega t)\, \vec{j} + 2\sin^2\frac{\varepsilon}{2}\, \vec{k}] \tag{A-6}$$

The gyro incremental angle output is given by: $\Delta\vec{\theta} = \int_{(n-\frac{1}{2})\Delta t}^{(n+\frac{1}{2})\Delta t} \vec{\Omega}\, dt$ (A-7)

$$\Delta\vec{\theta} = \omega\,\Delta t\left[\sin\varepsilon\,\mathrm{sinc}\frac{\omega\Delta t}{2}[\cos(n\omega\Delta t)\, \vec{i} - \sin(n\omega\Delta t)\, \vec{j}] + 2\sin^2\frac{\varepsilon}{2}\, \vec{k}\right] \tag{A-8}$$

In strapdown systems, the quaternion is updated using a transition quaternion derived from the gyro incremental angle outputs. However, the exact transition quaternion can be computed for the coning motion described above using the expressions:

$$\tilde{q}^{n+\frac{1}{2}} = \tilde{q}^{n-\frac{1}{2}}\, \tilde{q}(\Delta\vec{\phi}) \ \text{ or } \ \tilde{q}(\Delta\vec{\phi}) = \tilde{q}^{*n-\frac{1}{2}}\, \tilde{q}^{n+\frac{1}{2}} \tag{A-9}$$

$$\begin{aligned}\tilde{q}(\Delta\vec{\phi}) = &\left[\cos\frac{\varepsilon}{2},\ -\sin\frac{\varepsilon}{2}\left(\sin[(n-\tfrac{1}{2})\omega\Delta t]\, \vec{i} + \cos[(n-\tfrac{1}{2})\omega\Delta t]\vec{j}\right)\right] \\ &\bullet\left[\cos\frac{\varepsilon}{2},\ \sin\frac{\varepsilon}{2}\left(\sin[(n+\tfrac{1}{2})\omega\Delta t]\, \vec{i} + \cos[(n+\tfrac{1}{2})\omega\Delta t]\vec{j}\right)\right]\end{aligned} \tag{A-10}$$

$$\tilde{q}(\Delta\vec{\phi}) = \left[\cos^2\frac{\varepsilon}{2} + \sin^2\frac{\varepsilon}{2}\cos(\omega\Delta t),\ \sin\varepsilon\sin\frac{\omega\Delta t}{2}\left(\cos(n\omega\Delta t)\, \vec{i} - \sin(n\omega\Delta t)\, \vec{j}\right) + \sin^2\frac{\varepsilon}{2}\sin(\omega\Delta t)\, \vec{k}\right] \tag{A-11}$$

Now

$$\tilde{q}(\Delta\vec{\phi}) = [\cos\frac{|\Delta\vec{\phi}|}{2},\ \sin\frac{|\Delta\vec{\phi}|}{2}\ \vec{1}_{\Delta\phi}] \tag{A-12}$$

Hence,

$$\Delta\vec{\phi} = 2\frac{\sin^{-1}\frac{|\Delta\phi|}{2}}{\sin\frac{|\Delta\phi|}{2}}(\sin\frac{|\Delta\phi|}{2}\ \vec{1}_{\Delta\phi}) \tag{A-13}$$

Equating the magnitudes of the vector parts of (A-11) and (A-12) yields

$$\sin\frac{|\Delta\vec{\phi}|}{2} = \sqrt{\sin^2\varepsilon\sin^2\frac{\omega\Delta t}{2} + \sin^4\frac{\varepsilon}{2}\sin^2(\omega\Delta t)} \tag{A-14}$$

Utilizing a series expansion $\frac{\sin^{-1}x}{x} \cong 1 + \frac{1}{3!}x^2 + \ldots$ yields the following approximation for equation (A-13).

$$\Delta\vec{\phi} = SF\ \omega\Delta t(\sin\varepsilon\mathrm{sinc}\ \frac{\omega\Delta t}{2}\cos(n\omega\Delta t)\, \vec{i} - \sin\varepsilon\mathrm{sinc}\frac{\omega\Delta t}{2}\sin(n\omega\Delta t)\, \vec{j} + 2\sin^2\frac{\varepsilon}{2}\mathrm{sinc}\ \omega\Delta t\, \vec{k}) \tag{A-15}$$

where
$$SF = \frac{\sin^{-1}\frac{|\Delta\vec{\phi}|}{2}}{\sin\frac{|\Delta\vec{\phi}|}{2}} \approx 1 + \frac{1}{3!}[\sin^2\varepsilon\sin^2\frac{\omega\Delta t}{2} + \sin^4\frac{\varepsilon}{2}\sin^2(\omega\Delta t)] \tag{A-16}$$

Note that the scale factor SF is basically dependent only on the magnitude of $(\Delta\vec{\phi})^2$ or $(\Delta\vec{\theta})^2$ and is bounded by $(\frac{|\vec{\Omega}|}{2}\Delta t)^2$ and not really by ε or ω, as shown below. From equation (A-6),

$$\frac{|\vec{\Omega}|}{2}\Delta t = \omega\Delta t\sin\frac{\varepsilon}{2} \tag{A-17}$$

and
$$\begin{aligned}\sin^2\varepsilon\sin^2\frac{\omega\Delta t}{2} + \sin^4\frac{\varepsilon}{2}\sin^2(\omega\Delta t) &= (\omega\Delta t)^2\sin^2\frac{\varepsilon}{2}\operatorname{sinc}^2\frac{\omega\Delta t}{2}(1-\sin^2\frac{\varepsilon}{2}\sin^2\frac{\omega\Delta t}{2}) \\ &= (\frac{|\vec{\Omega}|}{2}\Delta t)^2\operatorname{sinc}^2\frac{\omega\Delta t}{2}(1-\sin^2\frac{\varepsilon}{2}\sin^2\frac{\omega\Delta t}{2}) \le (\frac{|\vec{\Omega}|}{2}\Delta t)^2\end{aligned} \tag{A-18}$$

We will ignore the effect of the scale factor in the remainder of this discussion since in practice, the bound is small. Thus, the target $\Delta\vec{\phi}$ desired for use in updating the quaternion is given by

$$\Delta\vec{\phi} = \omega\Delta t\left(\sin\varepsilon\operatorname{sinc}\frac{\omega\Delta t}{2}\cos(n\omega\Delta t)\,\vec{i} - \sin\varepsilon\operatorname{sinc}\frac{\omega\Delta t}{2}\sin(n\omega\Delta t)\,\vec{j} + 2\sin^2\frac{\varepsilon}{2}\operatorname{sinc}\omega\Delta t\,\vec{k}\right) \tag{A-19}$$

The normal method of deriving coning algorithms introduced by R. B. Miller ("A New Strapdown Attitude Algorithm", *Jounal of Guidance, Control and Dynamics*, Vol. 6, No. 4, 1983, pp. 287-291) is to concentrate on the fact that the main difference between $\Delta\vec{\phi}$ and $\Delta\vec{\theta}$ is in the z component.
Thus, we want

$$\begin{aligned}&\Delta\theta_z + \text{ compensation } \cong \Delta\phi_z \text{ , or} \\ &\text{compensation } \cong 2\sin^2\frac{\varepsilon}{2}(\operatorname{sinc}\omega\Delta t - 1)\omega\Delta t\end{aligned} \tag{A-20}$$

The compensation is obtained by employing cross-products of $\Delta\vec{\theta}$'s from sub-intervals.
Now suppose that

$$\Delta\vec{\theta} = \omega\Delta t\left(F(\omega)\sin\varepsilon\ \operatorname{sinc}\frac{\omega\Delta t}{2}[\cos(n\omega\Delta t)\,\vec{i} - \sin(n\omega\Delta t)\,\vec{j}] + 2F(0)\sin^2\frac{\varepsilon}{2}\,\vec{k}\right) \tag{A-21}$$

where $F(\omega)$ is a digital filter function for which:

$$F(0) = 1 \quad ; \quad \lim_{\omega\to 0}F(\omega) = 1 \quad ; \quad |F(\omega)| \le F_{\max} \text{ for all } \omega \tag{A-22}$$

We now consider a coning motion where the cone angle is a function of frequency, i.e. $\varepsilon = \varepsilon_0 g(\omega)$. Then the desired vector angle $\Delta\vec{\phi}$ is approximately given by equation (A-15).

$$\Delta\vec{\phi} = \omega\Delta t\left(\sin(\varepsilon_0 g(\omega))\operatorname{sinc}\frac{\omega\Delta t}{2}\left[\cos(n\omega\Delta t)\,\vec{i} - \sin(n\omega\Delta t)\,\vec{j}\right] + 2\sin^2\frac{\varepsilon_0 g(\omega)}{2}\operatorname{sinc}\omega\Delta t\,\vec{k}\right) \tag{A-23}$$

We now equate the AC parts of equations (A-21) and (A-22) and obtain the following:

$$F(\omega)\sin\varepsilon = \sin(g(\omega)\varepsilon_0) \quad \text{and solving for } g(\omega), \quad g(\omega) = \frac{1}{\varepsilon_0}\sin^{-1}(F(\omega)\sin\varepsilon) \tag{A-24}$$

The function $g(\omega)$ defines the apparent cone resulting from the filtered data. The term appearing in the third (DC) component of equation (A-23) is scaled by the factor:

$$\sin^2\frac{\varepsilon_0 g(\omega)}{2} = \frac{1-\cos(\varepsilon_0 g(\omega))}{2} = \frac{1-\cos(\sin^{-1}(F(\omega)\sin\varepsilon))}{2} \tag{A-25}$$

For either small ω or small ε, the above factor can be approximated in many cases by:

$$\sin^2\left(\frac{\varepsilon_0 g(\omega)}{2}\right) \approx F^2(\omega)\sin^2\frac{\varepsilon}{2} \tag{A-26}$$

Substituting equation (A-24) and approximation (A-26) into equation (A-23) yields the desired $\Delta\vec{\phi}$.

Now, with filtering: $\Delta\vec{\theta} = \omega\Delta t(F(\omega)\sin\varepsilon\,\mathrm{sinc}\frac{\omega\Delta t}{2}[\cos(n\omega\Delta t)\,\vec{i} - \sin(n\omega\Delta t)\,\vec{j}] + 2F^2(\omega)\sin^2\frac{\varepsilon}{2}\mathrm{sinc}\ \frac{\omega\Delta t}{2}\vec{k})$ (A-27)

The compensation is the difference between the $\vec{k}$ components of equations (A-21) and (A-27).

$$\text{compensation} \cong 2\omega\Delta t\sin^2\frac{\varepsilon}{2}\left(F^2(\omega)\ \mathrm{sinc}\ \omega\Delta t - F(0)\right) \cong 2\omega\Delta t\sin^2\frac{\varepsilon}{2}\left(F^2(\omega)\ \mathrm{sinc}\ \omega\Delta t - 1\right) \tag{A-28}$$

It should be noted that equation (A-28) collapses into equation (A-20) when $F(\omega)=1$.

The generalized Miller method breaks up each quaternion update interval into m subintervals of duration Δt. A $\Delta\vec{\theta}$ is obtained in each subinterval and vector cross-products formed between the subinterval $\Delta\vec{\theta}'s$ according to their spacing in time. For example, with m=4, subinterval spacings of 1, 2, and 3 are possible. A one-subinterval spacing results in the category-1 cross-products $\Delta\vec{\theta}_1 \times \Delta\vec{\theta}_2$, $\Delta\vec{\theta}_2 \times \Delta\vec{\theta}_3$, and $\Delta\vec{\theta}_3 \times \Delta\vec{\theta}_4$. A two-subinterval spacing results in the category-2 cross-products $\Delta\vec{\theta}_1 \times \Delta\vec{\theta}_3$ and $\Delta\vec{\theta}_2 \times \Delta\vec{\theta}_4$. A three-subinterval spacing results in the category 3 cross-product $\Delta\vec{\theta}_1 \times \Delta\vec{\theta}_4$.

In general, there will be m-1 possible subinterval spacings and C_m^2 possible cross-products. Each category of cross-products is described by the quantity $C_p(n)$ where:

$$C_p(n) = \left(\Delta\vec{\theta}_{nm} \times \Delta\vec{\theta}_{nm+p}\right)_{\vec{k}\ \text{component}} \tag{A-29}$$

It can be shown that all cross-products with the same spacing have the same $\vec{k}$ component. The AC components (i.e. $\vec{i}$ and $\vec{j}$) have only higher order terms (i.e. ε^3 and $F^3(\omega)$) and can be ignored.

Returning to equation (A-21) and appropriately substituting into the above expression yields:

$$C_p(n) = -(\omega\Delta t)^2 F^2(\omega)\sin^2\varepsilon\,\mathrm{sinc}^2\frac{\omega\Delta t}{2}\sin(p\omega\Delta t) \tag{A-30}$$

Note that the expression above does not depend on n.

In the case of resolution-enhanced data (U.S. Patent 5,485,273),

$$F(\omega) \cong \mathrm{sinc}\ \frac{\omega\Delta t}{2} \tag{A-31}$$

It follows that

$$C_p(n) \cong -(\omega\Delta t)^2\sin^2\varepsilon\,\mathrm{sinc}^4\frac{\omega\Delta t}{2}\sin(p\omega\Delta t) \tag{A-32}$$

Let $\alpha=\omega\Delta t$; then

$$C_p(n) \cong -\alpha^2\sin^2\varepsilon\,\mathrm{sinc}^4\frac{\alpha}{2}\sin(p\alpha) \tag{A-33}$$

Now

$$\mathrm{sinc}^4\frac{\alpha}{2} = \alpha^{-4}\left(6 - 8\cos\alpha + 2\cos(2\alpha)\right) \tag{A-34}$$

$$C_p(n) \cong -\frac{\sin^2\varepsilon}{\alpha^2}\left(6 - 8\cos\alpha + 2\cos(2\alpha)\right)\sin(p\alpha)$$

and

$$\cong -\frac{\sin^2\varepsilon}{\alpha^2}\left\{\sin\left[(p-2)\,\alpha\right] - 4\sin\left[(p-1)\,\alpha\right] + 6\sin(p\alpha) - 4\sin\left[(p+1)\,\alpha\right] + \sin\left[(p+2)\,\alpha\right]\right\} \tag{A-35}$$

$$\cong -\frac{\sin^2\varepsilon}{\alpha^2}\left[\sum_{k=1}^{\infty}(-1)^k\left((p-2)^{2k-1} - 4(p-1)^{2k-1} + 6p^{2k-1} - 4(p+1)^{2k-1} + (p+2)^{2k-1}\right)\frac{\alpha^{2k-1}}{(2k-1)!}\right]$$

Evaluating the k=1 and k=2 terms of the sum reveals that these are zero for any value of p. Thus, the sum reduces to

$$C_p(n) \cong \frac{\sin^2 \varepsilon}{\alpha^2}\left[\sum_{k=1}^{\infty}(-1)^k\left((p-2)^{2k+3} - 4(p-1)^{2k+3} + 6p^{2k+3} - 4(p+1)^{2k+3} + (p+2)^{2k+3}\right)\frac{\alpha^{2k+3}}{(2k+3)!}\right] \quad \text{(A-36)}$$

Returning to equation (A-28) which gives the desired compensation, we obtain:

$$\text{desired compensation} \cong 2\omega m\Delta t \sin^2\frac{\varepsilon}{2}\left(F^2(\omega)\ \text{sinc}\ \omega m\Delta t - 1\right) \quad \text{(A-37)}$$

For resolution-enhanced data, we substitute equation (A-31) to yield:

$$\text{desired compensation} \cong -2\omega m\Delta t \sin^2\frac{\varepsilon}{2}\left(1 - \text{sinc}^2\frac{\omega\Delta t}{2}\ \text{sinc}\ \omega m\Delta t\right) \quad \text{(A-38)}$$

The parenthetical expression in the above equation is expanded by means of trigonometric identities.

$$\begin{aligned}(\ ) &= 1 - \frac{4}{m\alpha^3}\left(\frac{1-\cos\alpha}{2}\right)\sin(m\alpha) \;=\; 1 - \frac{1}{m\alpha^3}\left\{-\sin[(m-1)\,\alpha] + 2\sin(m\alpha) - \sin[(m+1)\,\alpha]\right\} \\ &= 1 - \frac{1}{m\alpha^3}\sum_{k=1}^{\infty}(-1)^k\left[(m-1)^{2k-1} - 2m^{2k-1} + (m+1)^{2k-1}\right]\frac{\alpha^{2k-1}}{(2k-1)!}\end{aligned} \quad \text{(A-39)}$$

For k=1, the expression within the sum vanishes. For k=2, the expression corresponding to the sum is 1. Thus, the parenthetical expression can be rewritten as

$$(\ ) = \frac{1}{m}\sum_{k=1}^{\infty}(-1)^{k+1}\left[(m-1)^{2k+3} - 2m^{2k+3} + (m+1)^{2k+3}\right]\frac{\alpha^{2k}}{(2k+3)!} \quad \text{(A-40)}$$

Finally, substituting (A-40) into (A-38) yields:

$$\text{desired compensation} \;= 2\omega\Delta t \sin^2\frac{\varepsilon}{2}\sum_{k=1}^{\infty}(-1)^k\left[(m-1)^{2k+3} - 2m^{2k+3} + (m+1)^{2k+3}\right]\frac{\alpha^{2k}}{(2k+3)!} \quad \text{(A-41)}$$

Note that for m=1, no compensation is possible, and the net error will be the negative of the desired compensation. The lowest order term, expressed as a rate, will be:

$$\text{error rate} = -\frac{1}{2}\omega\sin^2\frac{\varepsilon}{2}(\omega\Delta t)^2 \quad \text{(A-42)}$$

For m>1, cross-products may be found and applied as compensation. Linear combinations of the (m-1) categories of cross-products are found which equal up to the (m-1)'th term of the desired compensation equation. Thus, coefficients x_p are chosen such that:

$$\sum_{p=1}^{m-1} C_p x_p = 2\omega\Delta t \sin^2\frac{\varepsilon}{2}\sum_{k=1}^{\infty}(-1)^k\left[(m-1)^{2k+3} - 2m^{2k+3} + (m+1)^{2k+3}\right]\frac{\alpha^{2k}}{(2k+3)!} \quad \text{(A-43)}$$

This is equivalent to the following system of equations (using equations (A-36) and (A-43).

$$\begin{aligned}\sin^2\varepsilon\sum_{p=1}^{m-1}\left\{\left[(p-2)^{2k+3} - 4(p-1)^{2k+3} + 6p^{2k+3} - 4(p+1)^{2k+3} + (p+2)^{2k+3}\right]x_p\right\} \\ = 2\sin^2\frac{\varepsilon}{2}\left[(m-1)^{2k+3} - 2m^{2k+3} + (m+1)^{2k+3}\right]\end{aligned} \quad \text{(A-44)}$$

Under the assumption of a small ε, equations (A-44) further reduce to:

$$2\sum_{p=1}^{m-1}\left\{\left[(p-2)^{2k+3} - 4(p-1)^{2k+3} + 6p^{2k+3} - 4(p+1)^{2k+3} + (p+2)^{2k+3}\right]x_p\right\} = \left[(m-1)^{2k+3} - 2m^{2k+3} + (m+1)^{2k+3}\right] \quad \text{(A-45)}$$

For k=1,. . . ,m-1, this leads to m-1 equations with m-1 unknowns (x_1, . . . ,x_{m-1}).

Once the compensation has been applied, a residual error exists which is the remaining portion of the desired compensation. The largest order term is the m'th term of the difference between the compensation series and the desired series. This can be converted to a rate by dividing by $m\Delta t$. The relative error rate (i.e. error rate normalized to coning rate) will be of order $2m$ in $\omega\Delta t$.

The calculated coning coefficients x_p are given in Table 2 for several values of m. Computer simulations of present-day coning algorithms yielded the response graphs shown in Figs. 2 and 4. Computer simulations of the coning algorithms described above yielded the response graphs shown in Figs. 3 and 5. The results for large angular rates are shown in Fig. 6.

Returning to equation (A-38), the desired compensation for small angles can be expressed as:

$$\text{desired compensation} \cong \frac{1}{2} m\alpha\varepsilon^2 \left(F^2(\alpha)\ \text{sinc}\, m\alpha - 1 \right) \tag{A-46}$$

where α has been substituted for $\omega\Delta t$. The applied compensation according to equation (A-30) is:

$$\text{applied compensation} \cong \sum_{p=1}^{m-1} C_p x_p = -\alpha^2\varepsilon^2 F^2(\alpha)\ \text{sinc}^2 \frac{\alpha}{2} \sum_{p=1}^{m-1} \sin(p\alpha)\, x_p \tag{A-47}$$

In comparing equations (A-46) and (A-47), it is apparent that $\alpha\varepsilon^2$ appears in both. Hence, we define the following:

$$z(\alpha) = \frac{1}{2} m \left[F^2(\alpha)\ \text{sinc}\, m\alpha - 1 \right] \tag{A-48}$$

The quantity $z(\alpha)$ is clearly an even function of α and may be expanded into a Taylor series about α=0.

$$z(\alpha) = \frac{1}{2} m \sum_{k=0}^{\infty} \frac{\alpha^{2k}}{(2k)!} \frac{d^{2k}}{d\alpha^{2k}} \left[F^2(\alpha)\ \text{sinc}\, m\alpha - 1 \right] \Big|_{\alpha=0} \tag{A-49}$$

Since $\lim_{\alpha \to 0} F(\alpha) = 1$, the k=0 term of the equation is zero. Consequently,

$$z(\alpha) = \frac{1}{2} m \sum_{k=1}^{\infty} \frac{\alpha^{2k}}{(2k)!} \frac{d^{2k}}{d\alpha^{2k}} \left[F^2(\alpha)\ \text{sinc}\, m\alpha \right] \Big|_{\alpha=0} \tag{A-50}$$

We also define:

$$y(\alpha) = \frac{1}{\alpha\varepsilon^2} \sum_{p=1}^{m-1} C_p x_p = -\alpha F^2(\alpha)\ \text{sinc}^2 \frac{\alpha}{2} \sum_{p=1}^{m-1} \sin(p\alpha)\, x_p \tag{A-51}$$

This is also an even function of α and may be expanded into a Taylor series:

$$y(\alpha) = -\sum_{p=1}^{m-1} \left\{ \sum_{k=0}^{\infty} \frac{\alpha^{2k}}{(2k)!} \frac{d^{2k}}{d\alpha^{2k}} \left[\alpha F^2(\alpha)\ \text{sinc}^2 \frac{\alpha}{2} \sin(p\alpha) \right] \Big|_{\alpha=0} \right\} x_p \tag{A-52}$$

The k=0 term of the internal summation is zero and consequently,

$$y(\alpha) = -\sum_{p=1}^{m-1} \left\{ \sum_{k=1}^{\infty} \frac{\alpha^{2k}}{(2k)!} \frac{d^{2k}}{d\alpha^{2k}} \left[\alpha F^2(\alpha)\ \text{sinc}^2 \frac{\alpha}{2} \sin(p\alpha) \right] \Big|_{\alpha=0} \right\} x_p \tag{A-53}$$

We now equate the first (m-1) terms of the k series of equations (A-50) and (A-53) to yield m-1 equations with m-1 unknowns $x_1, x_2, \ldots, x_{m-1}$. For k=1 to $m-1$,

$$\sum_{p=1}^{m-1} \frac{d^{2k}}{d\alpha^{2k}} \left[\alpha F^2(\alpha)\ \text{sinc}^2 \frac{\alpha}{2} \sin(p\alpha) \right] \Big|_{\alpha=0} x_p = -\frac{1}{2} m \frac{d^{2k}}{d\alpha^{2k}} \left[F^2(\alpha)\ \text{sinc}\, m\alpha \right] \Big|_{\alpha=0} \tag{A-54}$$

The solution to these equations leads to the cancellation of all terms of order $2m$-2 and above, leaving terms of order $2m$ in the relative error rate. This type of analysis may be used for a variety of filter functions $F(\omega)$ including a lowpass filter.

Gyroscopic Instruments for the Russian Segment Attitude Determination System of the International Space Station Alpha

B.A. Kazakov, S.A. Kharlamov, A.P. Mezentsev, V.I. Reshetnikov, I.N. Sapozhnikov

Kuznetsov Research Institute of Applied Mechanics
55, Aviamotornaya street, Moscow, 111123, Russia
Tel. (7095)273-38-76, Fax (7095)273-19-02

Summary

There are several sets of gyroscopic instruments measuring the attitude of the Service Module and Functional Cargo Module which constitute the Russian segment of the International Space Station Alpha. This report presents some important details of the development of the precision gyroscopic instruments, the contribution of which in the ISSA attitude accuracy have the primary significance. The main performance characteristics of gyroscopic instruments are discussed such as the accuracy (the ultimate bias drift uncertainty less than 0.003 deg/hour), service life (more than 10 years), and high reliability.

Introduction

According to the planned assembly on the International Space Station Alpha in orbit [1], the Russian segment should involve two basic elements, namely, the FGB (the Russian acronym for the Functional Cargo Module now coined internationally) and the SM. This time, the FGB had been in orbit since November 20, 1998, the SM launch is planned this year. The RIAM had developed two precision gyroscopic instruments for the attitude determination systems of the FGB and SM. Both instruments are intended to measure the vector of the ISSA angular rate rotation relative to the inertial coordinate system. They have the strapdown configuration, i.e., they have the single degree-of-freedom gyros operating in the torque-to-balance mode which are rigidly attached to the navigational base of each module. An instrument of such type in named as gyroscopic angular rate measuring device. Since this literal translation of its name from Russian is long and inconvenient, hereinafter, the instrument will be called as a Strapdown Gyrometric Unit, or simply, as a GMU. Its output represents the information about the ISSA rotation angle vector in the station body-fixed reference frame for any given time interval. This output is transformed via an interface to a board computer where it is processed and used for the station attitude control.

The GMU installed on the board of FGB had been developed in the begin of eighties for the manned space stations. The instruments of this type operate successfully since February 16, 1986, up today on the board of Space Station MIR. The valuable experience accumulated allows us to get a number of innovative solutions aiming to the further improvement of the gyroscopic instruments. These innovations were implemented in the design of SM GMU, which had been developed in the middle of nineties. Thus, the FGB and SM GMUs belong to the different instrument generations and, therefore, the principal attention will be paid on the SM GMU. The data for FGB GMU are shown for comparison in order to illustrate quantitatively the judgements abut the progress in the gyroscopic instrument development from generation to generation.

Instrument mechanization concept

The GMU mechanization concept was formulated on the base of consideration of the main requirements to the gyroscopic instruments for the long-life manned space stations:

- the long service life more the 10 years;
- the high reliability;
- the high stability of accuracy (e.g., the random variations of gyro bias drift should be less than 0.003 deg/hr, the scale factor uncertainty should be less than 10 ppm):
- the low informational output noise level considerably less than 1 arc second.

The first two requirements were the reason of the selection of the proven MIR station GMU as the FGB GMU especially if we take into account an additional economical constraint to reduce the cost of development. The severe requirements to the accuracy were satisfied in the SM GMU design on the base of the large experience of gyroscopic instrument development for the various spacecraft and launch boosters.

The high requirements to the service life and reliability lead to the increasing emphasis on the redundancy and standby usage in the GMU design, where the redundancy and multiplication may be used on the level of measuring channels, their gyros and electronic devices, as well as electronic components. In contrast to the FGB GMU which has the classical triad of gyros with orthogonal

Paper presented at the RTO SCI International Conference on "Integrated Navigation Systems", held at "Elektropribor", St. Petersburg, Russia, 24-26 May 1999, and published in RTO MP-43.

input axes and the redundant electronic devices, the SM GMU has the tetrad of gyros and each of them, being equipped by the individual electronic feedback loop, forms the independent measuring channel. The orientation of input axes of four SM GMU gyros relative to the Service Module body-fixed coordinate system is shown in fig. 1. The gyro input axes are orthogonal to the four upper faces of an octahedron the vertices of which are located on the axes of SM body-fixed (navigational) frame. Such a configuration of gyros arrangement provides equal contribution of each gyro in the process of SM angular rate measurement when either all the measuring channels are in operation or one of them fails. In addition to the gyro redundancy, the same principle (triplication and voting procedure) is realised in the SM GMU electronic circuitry on the level of devices. Therefore, the GMU has four independent, electrically uncoupled measuring channels.

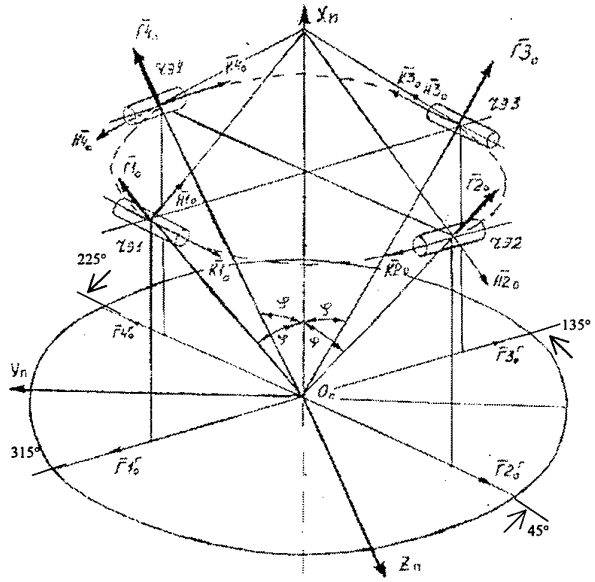


Figure 1. The orientation of gyroscope input axes relative to the Service Module body-fixed coordinate system.

A functional block diagram of CM GMU demonstrating its interrelation with a board computer is shown in fig. 2. The instrument output information provides the computation of the ISSA attitude in the quaternion form. Therefore, the GMU represents the angular deviation receiving part of a strapdown inertial navigation system (SINS). In Russia, the leading role in the development if SINS theory and applications to spacecraft is played by the skilful team in RSC "Energya" which is guided by professor V.N. Branets [2].

The second conceptual aspect of the CM GMU development is the provision of thermally comfortable conditions to the gyros and the precision electronic devices of feedback loop of the measuring channel. The ultimate goal is the fulfilment of the challenging requirements to the accuracy of instrument measuring channels. To this end, the gyros and electronic devices are packaged in a cluster and the thermal control system for the cluster is realised in which the solid state micro-coolers (Peltier elements) are used. The thermal control system provides the precision temperature stabilisation and thermal comfort for the gyros and precision electronics under effect of the environmental temperature variation in the range from zero to +35 deg C. The special attention was paid on the correct organisation of heat flux from the heat generating source such as gyros and electronics to the heat rejecting sink or the navigational base surface. The problem was solved by the incorporation of special heat conducting elements into instrument design and by the application of heat conductive pastes in the gap between the bottom surface of GMU housing and the heat rejecting surface of SM navigational base. These design features are exceptionally important since they provide only the heat rejection in space where the heat convection is absent.

Gyroscopes

The inertial sensor of GMU is a float integrating gyroscope modified to its implementation in the strapdown applications. The gyro is used for the measurement of the angular rate vector projection on its input axis. The key design features which enable to satisfy the strict requirements to accuracy are gas-lubricated gyro wheel spin bearings, a float magnetic centring system, and an individual thermal control system.

The hemispherical gas lubricated rotor spin bearings provide practically unlimited continuous running at 30000 rpm. The special design factors and technological procedures prevent the penetration of extraneous particles more than 1 μm in size into the bearing gap. The special manufacturing process and materials selected for the bearing surface coating enable to get the hard surface protecting film which reduces the wear of surfaces in mechanical contacts during start-stop operations. As a result, the total number of allowable start-stop cycles is more than 2000 which is much more than the number of cycles needed for the total instrument service life. The elimination of mechanical contact between the running and fixed components of gyro promotes also the reduction of mechanical noise level in the output information of gyroscope.

The float magnetic centring system is specially aimed toward a strapdown application in order to unload float journals and pivots under the condition of the ISSA rotation and the gyros rigidly attached to the station. The unloading eliminates the mechanical contact and the boundary sliding friction in the float journal and pivot bearings providing the considerable enhancement of gyro bias drift stability in the angular rate measurement operating mode.

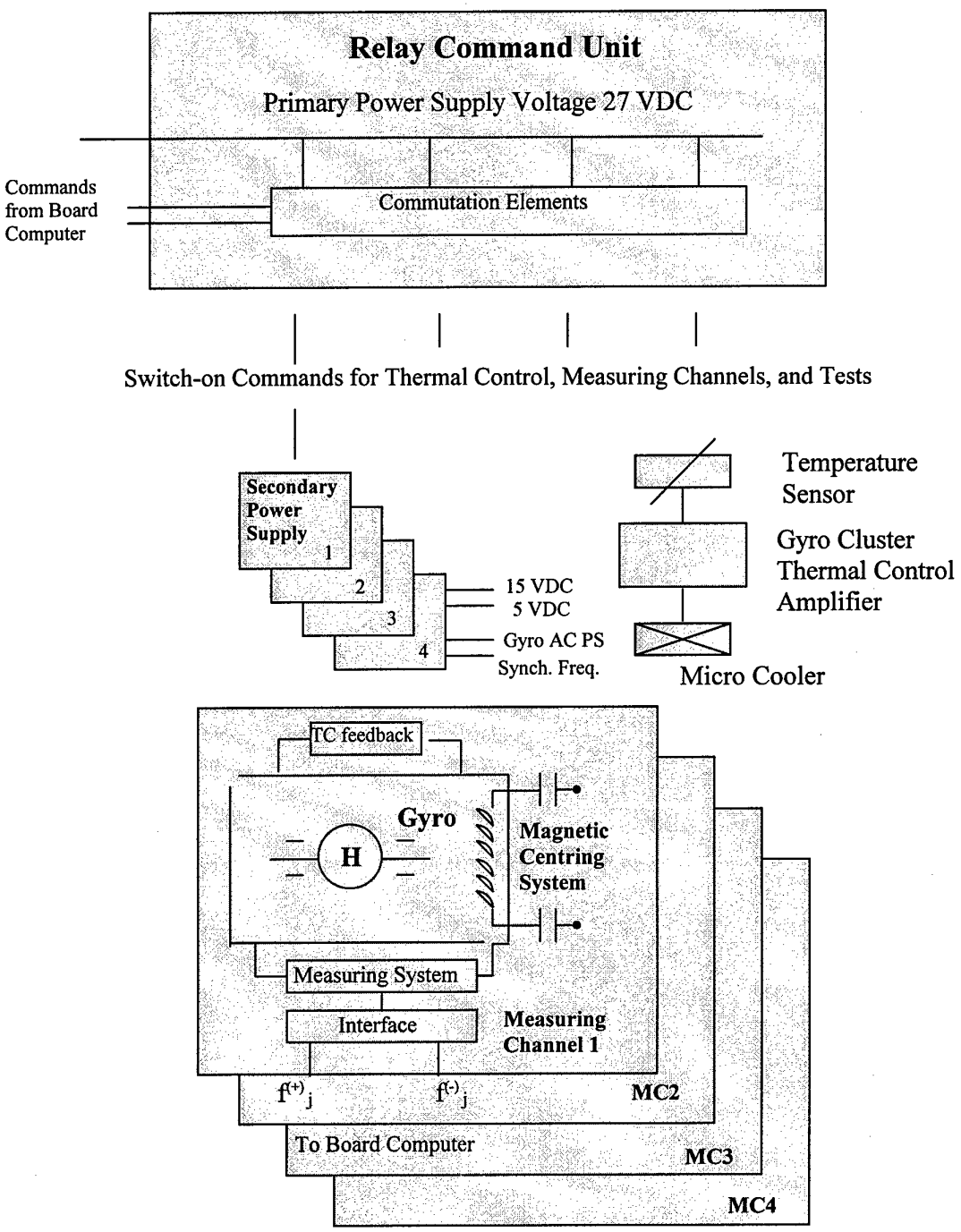


Figure 2. The functional block diagram of the Gyrometric Unit of the Service Module.

Each gyro has an individual precision temperature control system which consist of a temperature sensor, feedback control amplifier, and heater. This system restrains the gyro operational temperature in the range of several hundredths of deg C. The total gyro temperature control system has two stages if one takes into account the thermal control system of cluster.

As a result, the gyros have the high drift stability. The test results for one of SM GMU channels, which are aimed to check the gyro bias drift stability over 24-hour period, are shown in fig. 3 where the time history of the bias drift in deg/hr is plotted. The plot presented demonstrates that the random bias drift variations are in the tolerance limit with an obvious margin.

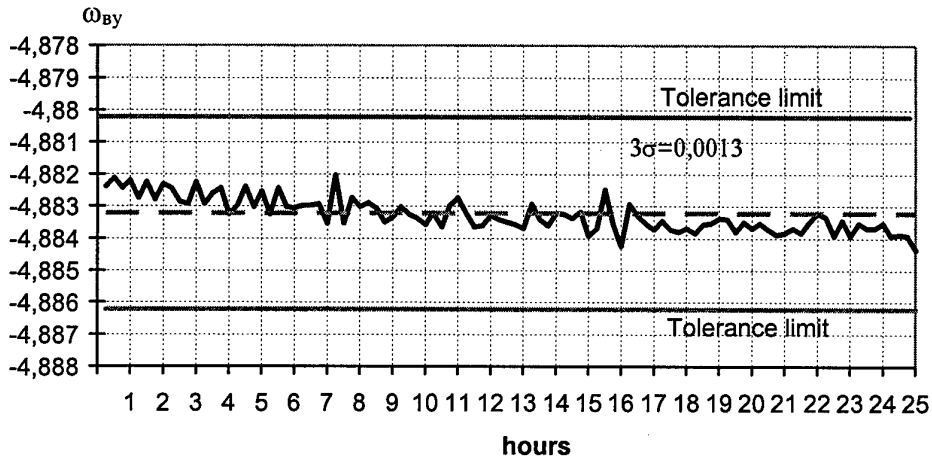


Figure 3. GMU test results for the bias drift rate stability in 24-hour run.

Measuring channel

The float integrating single DOF gyro is equipped by the electronic feedback control system in such manner that it operates in the torque-to-balance mode. The gyro and feedback system make up the measuring channel which puts out the ISSA angle of rotation over a given time interval and transfers this information to the board computer. The crucial requirements to the measuring channel are the highest possible scale factor stability and the lowest possible noise level. To this objective, the pulse-width modulation of the torque generator current is implemented, the rigorous analysis and careful design of all the elements of feedback control system from the angle pickoff to the torque generator and the output of electronic code transformation unit (current-frequency-code) is carried out, and all the principal electronic devices affecting on the accuracy are installed into the temperature stabilized zone of the instrument. The electronic part of instrument has the required redundancy. It is assembled of the highly reliable components selected by the special screening tests.

The quality of scale factor stability problem solution is illustrated by fig. 4 where the stability of information bit value in arc sec over 24-hour continuous run is plotted for ten different measurements, the numbers of which are indicated in abscissa. The plot shows that the ultimate deviation of this parameter from its averaged value is less than 11 ppm.

The noise reduction problem is solved on the base of large volume of experimental studies and tests. In the course of experimental studies of noise generated by many sources of mechanical and electrical origin, the special test rigs (such as stabilized foundations) and high sensitive measurement equipment were used. The nature of noise in measuring channel output is shown in fig. 5 for a sample time interval of 350 s. The sampling period during the tests is equal to 0.1 s (sampling or interrogation rate 10 Hz). For the sample volume of 3500 readouts, the RMS noise level is 0.041 arc sec.

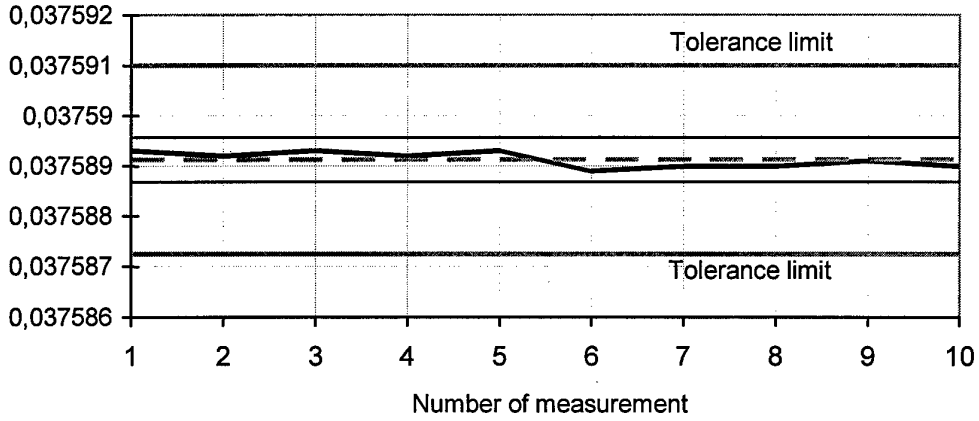


Figure 4. Test results for GMU measuring channel output information bit value.

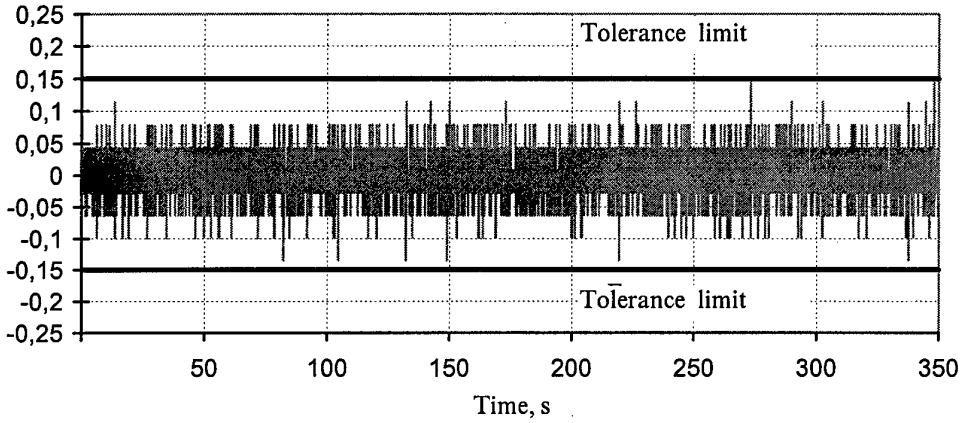


Figure 5. Test results for the measurement of noise in the GMU measuring channel.

In summary, the real scale factor stability over several month period has the level of tens ppm, the value of output information bit is less than 0.04 arc sec, the RMS level of noise is less than 0.1 arc sec. successfully. Thus, the scale factor stability of SM GMU developed is competitive with those of instruments based on other types of gyros (especially the ring laser gyros which are considered usually as having the highest scale factor stability). At the same time, the SM GMU has the advantage over other instruments with respect to the bit value and noise level.

Instrument design

The instrument represents a monoblock having the cylindrical form. Its bottom side provides the strictly defined installation of unit on the SM navigational base and thermal interface. Its internal cavity contains the cluster of gyros and precision electronic devices which serves as a thermostat. The special design measures are realised to provide the high stability of gyro input axes relative to the basic planes of instrument. The cluster supports provide its thermal insulation from the instrument housing, which assists to the precision gyro temperature control operation, as well as the vibration and acoustic protection of cluster during the SM orbit insertion phase. The instrument internal cavity is hermetically sealed by special gasket and filled by dry nitrogen.

The instrument accuracy enhancement is ensured also by the calibration procedures and error corrections. Such procedures are used both in the instrument manufacturing and in real operation.

The problem of significant overall size and total mass reduction has been solved simultaneously with the reliability, accuracy, and performance enhancement.
The table below contains the basic accuracy and performance characteristics of both FGB and SM gyrometric units.

Parameter of Gyrometric Unit	Value for	
	FGB	SM
1.Range of measured angular rate, deg/hr	3	0.5
2. Information bit value (scale factor), arc second	1.8	0.036
3. Bit value uncertainty (scale factor stability), ppm	±200	±30
4. Bias drift stabilitiy over 24-hour period, deg/hr	±0.03	±0.003
5. Gyro input axis misalignment stability, arc second	±60	±30
6. Informational output noise [1σ], arc second	±2	±0.1
7. Temperature on instrument mounting surface, deg C	0÷35	0÷35
8. Service life, years	10	10
9. Mass, kg	32	12.6
10. Power consumption, W	130	75

Conclusion

The presented above basic, but not exhaustive details of the development of precision gyrometric units for the attitude determination system of the Russian segment entering in the International Space Station Alpha demonstrate the efforts of the Kuznetsov RIAM which enable eventually to create the instruments with basic characteristics shown in the foregoing table. All the listed values of parameters were verified by the large volume of ground qualification tests of both the instrument as a whole and the separated subsystems and elements.

The flight GMUs of FGB and SM had been manufactured, passed successfully all the tests, and delivered to the customers for the integration into the attitude control systems of FGB and SM and their checks before the insertion in orbit.

The comparison of parameters of GMUs prepared for the FGB and SM shows the significant progress in the development of precision gyroscopic instruments for space applications gained by the Kuznetsov RIAM engineering team. These instruments should have high reliability, high accuracy, and long service life.

References

1. Tedder, A.J. and Shreve, J.V., "Estimating Space Station Alignment for Shuttle Docking - A US/Russian Joint Venture," The 2nd Saint Petersburg International Conference on Gyroscopic Technology and Navigation, Saint Petersburg, May 24-25, 1995, Part I, pp. 43-52.

2. Branets, V.N. and Shmyglevsky, I.P. Introduction to the Theory of Strapdown Inertial Navigation Systems (in Russian), Nauka, Moscow, 1992.

PECULIARITIES OF CALIBRATING THE TRIAD OF ACCELEROMETERS

S.F. Konovalov, V.V. Yurasov
Bauman Moscow State Technical University
24 Holzunov pereulok 6
11902 Moscow, Russia

Abstract

Key words: triad of accelerometers, inclinometery, navigation

In inclinometery process of oil and gas wells (inclined and horizontal) three accelerometers with orthogonal sensitive axes are used as inclinometer sensors. The process of lowering devices in a borehole is accompanied by fast temperature change which for tens of minutes can be ~100 ...120°C. If accelerometers have different scale factor (SF) temperature coefficients, intolerable large error in determination of zenith angle can arise. The attempts of using algorithmic compensation are inefficient in this case as it is required to carry out this procedure in unsteady temperature mode. Moreover, algorithmic compensation is not capable to eliminate errors connected with irreversible changes of magnetization of accelerometer torquer magnets due to thermal and possible mechanical shocks which accompanies lowering the devices in borehole.The problem is solved by creation of the triad of accelerometers in which strictly coordinated SF change of all devices being incorporated into the triad is provided.

The specified triad of accelerometers is of interest not only for inclinometery of wells, but also for inertial navigation as the use of the triad significantly simplifies and accelerates a process of preflight calibration.

Introduction

One of the problems, involving both inclinometery and inertial navigation, is the keeping scale factors (SF) stability with time of accelerometers which measure projections of acceleration of a vehicle. It is known that owing to a number of the reasons SF stability may vary within limits essentially exceeding the required tolerance which values are of the order from 0,2 ppm to 1 ppm for most perfect modern navigation accelerometers [1]. The main reasons of changing SF consist in the following factors:

1. A change of pendulous moment owing to deformations of movable accelerometer unit (basically due to temperature deformations).
2. A change of a hydrostatic moment owing to temperature change of density of the device filling liquid and presence of a displacement of centre of pressure for movable unit relative to its suspension axis. The given factor is only critical for floated devices.
3. Convertible temperature changes of magnetic flux density in a gap of the device magnetoelectric torquer. They are caused basically by a temperature change of magnetic energy of permanent magnets.
4. A size change of movable coil of a torquer (arm of acting forces) owing to temperature change.
5. A spontaneous demagnetization of permanent magnets while operating.
6. A change of SF of accelerometer analog-to-digital converters (ADC).

From the listed factors the most important are the 2nd, the 3rd, the 5th and 6th ones. Though 2nd and 3rd factors result in convertible changes of SF, but they can be compensated by algorithmic way only in accelerometers of INS, where the temperature mode is close to steady-state. It is a problem to ensure precise algorithmic compensation for inclinometer. It is impossible to eliminate effect of 5th and 6th factors by algorithmic way.

Let us consider influence of the 5th factor. Since here the instant of SF change is unpredictable, a value of the change is accidental and can be rather great. The conventional diagram of the SF change is shown in Fig.1. Jump changing SF are especially probable during lowering and lifting the tool in a borehole. Many researchers give large attention to work on reduction of specified errors. However, the progress here goes very slowly and extremely dearly costs as it is associated with a creation of new precision magnetic alloys and development of special technological techniques of stabilization of magnetic flux density.

The influence of the 6th factor (FS instability of analog-to-digital converter) has a double nature. The first component is a convertible influence of temperature on parameters of silicon Zener-diodes and reference resistors used in a current (voltage) stabilizer of the converter. This influence can be minimized by special tuning the stabilizer or it can be discounted by algorithmic way. The second component of the influence is related to time instability of parameters of such components as Zener-diodes, resistors and zero currents of switches of the converter. This influence is unpredictable, and, though here the changes occur more smoothly and have smaller values as compared with change of magnetic flux density, but, nevertheless, they have remarkable influence on SF stability of accelerometer.

Paper presented at the RTO SCI International Conference on "Integrated Navigation Systems", held at "Elektropribor", St. Petersburg, Russia, 24-26 May 1999, and published in RTO MP-43.

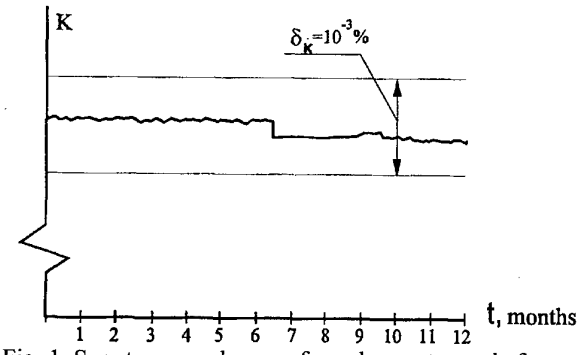


Fig. 1. Spontaneous change of accelerometer scale factor with time.

The best achievement for today is a change of scale factor owing to manifestation of the 5th and 6th factors within limits of 0,005%-0,01 % for a year depending on conditions of temperature changing (in storage mode). And we would like to emphasize once again that an instant of time, when the spontaneous change can occur of magnetic flux density in a torquer gap and conversion coefficient of A/D converter, is accidental. The given problem is generally accepted to solve by calibrating accelerometers of INS before working.

The calibrating is provided by specified turning accelerometric unit in three mutually perpendicular positions, so that accelerometers serially measure precisely known acceleration of gravity in these positions. An accelerometer output signal equivalent g is defined in each position. There are no considerable problems in performing specified turns for inclinometer, but for INS given procedure requires the use of special turn mechanism. Besides, in a number of cases for INS the operation of scaling has inadmissible duration. Therefore, the calibrating is made not directly before working, but as a preventive measure, for example, one time a day. Thus, period of time between calibrations is chosen from test results statistical processing for accelerometer batch and evaluated on allowable probability of scale factor change. In the period between calibrations a scale factor is assumed as a constant value. In this case there is a risk of its change which causes the INS operation on the basis of incorrectly determined scale factors with all following results.

Here the objective to provide the greatest possible value of the allowable period between calibrations. In USA there is a program of development of navigation technology for the period till 2000 [2], which provides creation of accelerometers admitting single-pass scaling during 100 days of operation with providing accuracy of determining a scale factor in order of 1 ppm (10^{-4} %). However, the given problem has no a solution for now.

Search of solving the problem of fast calibration of accelerometers

In navigation technology to use indirect ways in solving technical problems which can not be solved by a direct way has become traditional. Let us consider variant of the solving the problem not requiring creation of superstable magnetic materials and high-precision ADC. The problem of fast calibration of the devices can be solved in a case, when the scale factors of all three accelerometers of INS change in coordination, i.e. for each accelerometer output signal U_i is defined as

$$U_i = K_{i_0} \cdot K(t,t^0) \cdot a_i, \qquad (1)$$

where K_{i_0} - scale factor of i-th accelerometer with ADC precisely determined during manufacturing the device;

$K(t,t^0)$ - coefficient, depending on temperature and time, identical for all accelerometers;

a_i - projection of acceleration to an axis i-th accelerometer.

For realization of dependence (1) it is necessary to ensure coordinated changes of SF of ADC and coordinated changes of transfer coefficient of torquer of all accelerometers.

For inclinometer the first problem is solved by using common ADC with multiplexer. For INS this way is unacceptable, so a common reference element (Zener diode) and either identical or common reference resistors for three ADC.

For electromechanical accelerometer unit coordinated changing scale factors is possible to realize by use of its common magnetic system for every magnetoelectric torquer. If we provide conditions of coordinated changing scale factors for three accelerometers with ideal mutually perpendicular measuring axes, a value of

coefficient $K(t,t^0)$ can be determined on results of simultaneous measurements of output signals of these devices being in arbitrary and unknown position relative to vector g (see Fig.2) as follows

$$K(t,t^0) = \frac{1}{g}\sqrt{\left(\frac{U_1}{K_{1_0}}\right)^2 + \left(\frac{U_2}{K_{2_0}}\right)^2 + \left(\frac{U_3}{K_{3_0}}\right)^2} \qquad (2)$$

As seen from (2), in this case at unspecified accelerometer orientation the process of calibration does not require a change of accelerometer orientation in a field of acceleration of gravity, and it is enough to make measurements of output signals U1, U2, U3. As result we can provide a measuring complex (inclinometer, INS) with data of calibration directly before its operation that gives essential improving in accuracy. Moreover, time needed for calibration is reduced (owing to eliminating the necessity of the turning and also to parallel measuring signals U1, U2, U3 instead of sequential measuring in accordance with a traditional technique).

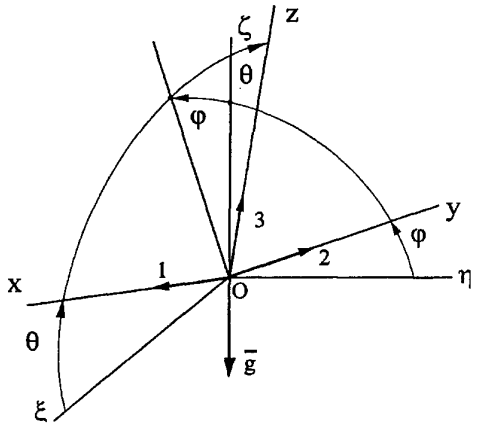


Fig. 2. Orientation of accelerometer 1, 2 and 3 of ideal triad while calibrating.

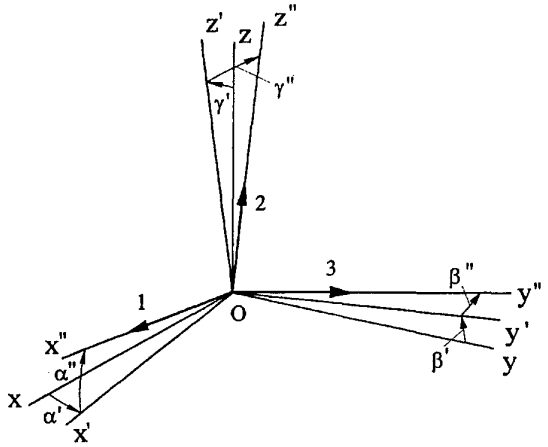


Fig. 3. Position of measuring axes of accelerometers of a real triad at presence of base errors.

The expression (2) is right only when a perpendicularity between measuring axes of accelerometers is provided. Really, accelerometers have base (alignment) errors (Fig.3), which also induce errors in process of calibration. The projections of acceleration g to measuring axes of imperfect accelerometers having regard to minor values of angles γ', γ'', α', α'', β', β'' are equal:

$$a_{x''} = g\sin\theta\cos\varphi + g\sin\varphi\,\alpha' + g\cos\theta\cos\varphi\,\alpha'';$$
$$a_{y''} = g\sin\varphi - g\sin\theta\cos\varphi\beta'' + g\cos\theta\cos\varphi\,\beta';$$
$$a_{z''} = g\cos\theta\cos\varphi - g\sin\varphi\,\gamma' - g\sin\theta\cos\varphi\,\gamma''.$$

In this case the sum of squares of acceleration projections $a_{x''}, a_{y''}, a_{z''}$ will differ from g^2, and to the second order it will be equal:

$$a_{x''}^2 + a_{y''}^2 + a_{z''}^2 = g^2(1 + \sin2\varphi\cos\theta\,\varepsilon_{23} + \sin2\theta\cos^2\varphi\,\varepsilon_{12} + \sin2\varphi\sin\theta\,\varepsilon_{13})$$

where: $\varepsilon_{12} = \alpha'' - \gamma''$; $\varepsilon_{23} = \beta' - \gamma'$; $\varepsilon_{13} = \alpha' - \beta''$
are angles of mutual non-perpendicularity between 2nd and 3rd, 1st and 2nd, 1st and 3rd pair of accelerometers respectively.

Note here that base errors, determined as the coordinated turns of axes 1 & 2, 2 & 3, 1 & 3 relative to mounting plane of the base of the accelerometers' triad, do not exert effect on the accuracy of calibration. When using the expression (2), base errors can result in an essential error of scale factor determining. For example, for a level of an error about 5 arc. minutes the error in scale factor determining by the expression (2) can be as great as 0,1 %. However, it is important to take into account that the angles $\varepsilon_{12}, \varepsilon_{23}, \varepsilon_{13}$ are small, are known (since

for the triad of accelerometers these angles are determined in testing the devices at a stage of manufacturing the triad) and they keep stable values or its change is known (for example, under temperature changing) in service. Hence its presence in calibrating the devices can be discounted by an analytical way.

$$K'(t,t^0) = \frac{K(t,t^0)}{\sqrt{1 + sin2\varphi^* cos\theta^* \varepsilon_{23} + sin2\theta cos^2 \varphi^* \varepsilon_{12} + + sin2\varphi^* sin\theta^* \varepsilon_{13})}} \qquad (3)$$

where values of φ^* and θ^* with accuracy sufficient for the use we shall find using the expressions:

$$\varphi^* = arcsin\frac{U_3/g}{K(t,t^0)}$$

$$\theta^* = arctg\frac{U_1}{U_2} \qquad (4)$$

When calibrating the triad, α', α", β', β" angles are ignored in determining φ^* and θ^*, since they give the correction having no appreciable effect on the value of $K(t,t^0)$. When inclinometering wells, the appropriate correction can be made for accurate determining zenith angles φ and θ.

The stand to determine accelerometers' parameters of the triad

To realize precision calibration of the triad and subsequent determination of acceleration projections it is necessary to have error model of accelerometers of the triad. As while testing it is necessary to confirm the fact of coordinated SF change of devices incorporated in the triad, it requires to provide a possibility of parallel tests of accelerometers' pairs (1,3) and (1,2). For this purpose on the rotary unit of the test stand it is necessary to ensure precision orientation of the triad being tested in the following positions:

1. The measuring axis of accelerometer 2 is parallel to rotation axis of the stand.
2. The measuring axis of accelerometer 3 is parallel to rotation axis of the stand.

These turns should be realized inside the thermal chamber. Measuring axis of accelerometer 1 of the triad is normal to turn axis of the stand in both cases. In detail the technique and test equipment of a the triad a of accelerometers is considered in [3].

Development of design of the triad of accelerometers

Let us consider some variants of a triad design [4], ensuring coordinated changes of scale factor for all three devices. These designs are assumed to use accelerometer electromechanical unit containing a pendulous frame of magnetoelectric torquer. The arm of a pendulous frame can be perpendicular to the frame plane (Fig. 4 a) or lie in the frame plane (Fig. 4 b). The frame 9 is mounted in the device case with the help of precision support means 6 having very small moment of friction. The case is supplied by pick-off which measures turn angle of the frame and consists of movable part 8 and stator 10. The core 11 made of magnetically soft material is placed inside the frame of present devices instead of a permanent magnet commonly used in similar electromechanical units. The hermetically sealed case of an electromechanical element is made of nonmagnetic material. For the variant of design scheme where a plane of a pendulum 7 is perpendicular to a plane of the frame 9, there are three sensitive elements (SE) 4 in gaps of common magnetic system which is presented in Fig. 5.

Planes of the SE frames pass through a common axis of symmetry of magnetic system and ones are oriented relative to each other at angle 120°. Choosing a cone angle α=35°15′52″, we provide mutual perpendicularity of accelerometers' measuring axes. For the given design when a magnetizing force are changed, magnetic flux density of each magnetic core threading the frames of three sensitive elements will have identical change in value, and thus SF of the devices will have coordinated changes. In addition, it should be noted that a magnetic gap in the given design is formed by elements made of magnetically soft material from both sides in contrast to traditional design, where a magnetic gap is formed by a surface of a

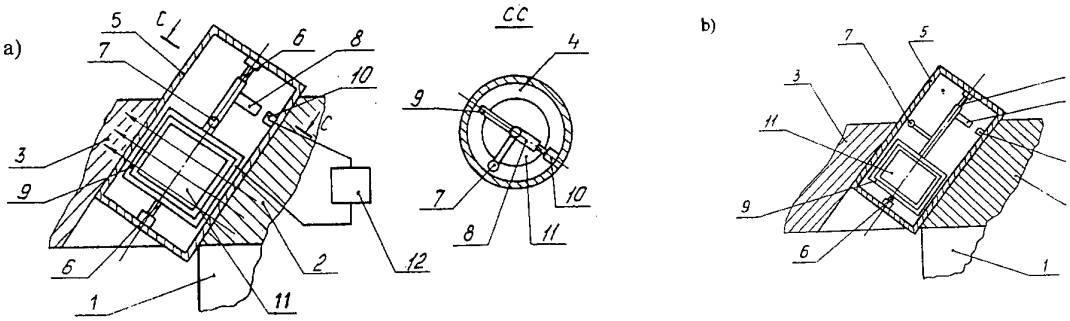


Fig. 4. Design scheme of single-axis accelerometer of the triad:
a - variant with pendulum arm perpendicular to planes of torquer frame;
b - variant where the pendulum arm lies in a plane of torquer coil:
1- common constant magnet; 2 - common pole piece; 3 - common magnetic circuit; 4 - single-axis accelerometer; 5 - nonmagnetic case; 6 - precision support; 7 - pendulum; 8 - movable part of pick-off; 9 - torquer frame; 10 - torquer stator; 11 - torquer core; 12 - amplifier

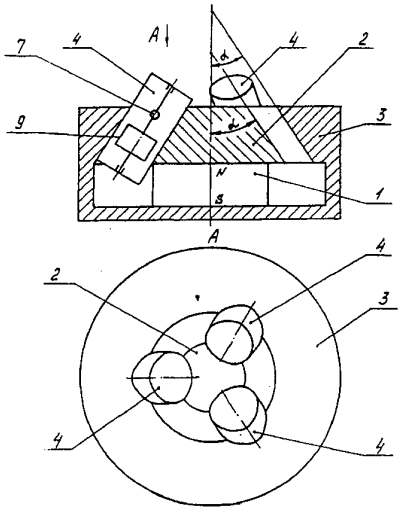


Fig. 5. A design variant of the triad with arrangement of axes of accelerometer support along generatrix of a cone surface with vertex looking onto the opposite side relative to a magnet:

1 - common constant magnet; 2 - common pole piece; 3 - common magnetic circuit; 4 - single-axis accelerometer; 7 - pendulum; 9 - torquer frame

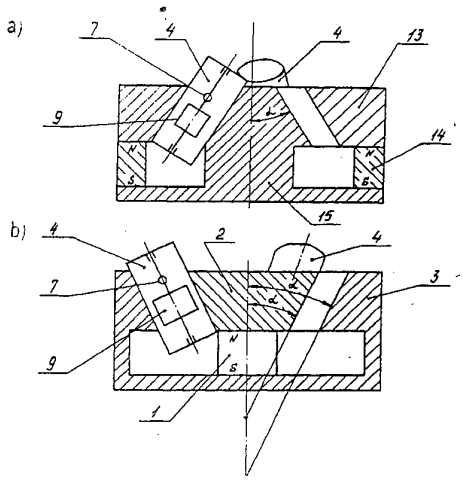


Fig. 6. Design variants of the triad:
a - with magnetic circuit located inside pole piece and magnet rings with axial magnetization; b - with an arrangement of axes of accelerometers support along generatrix of a cone surface with vertex looking onto the side of a magnet:

1 - internal constant magnet; 2 - internal pole piece; 3 - internal magnetic circuit; 4 - single-axis accelerometer; 7 - pendulum; 9 - torquer frame; 13 - external pole piece; 14 - external magnet; 15 - internal magnetic circuit

permanent magnet. It guarantees a uniform distribution and stability of magnetic flux density in gaps and, hence, the absence of dependence of the frame turn angle on torquer transconductance. Moreover, the absence of influence of SE pick-off instability on the device scale factor is provided. Variants of using the sensitive element when a pendulum arm is located in plane perpendicular to the frame are shown in Fig. 6 a,b.

In variant shown in Fig. 7, as opposed to variants considered earlier, the accelerometer measuring axes are not crossing (transversal) lines but skew lines, remaining mutually perpendicular. It does not exerts effect on accuracy of INS work. Here a magnetic system has a cylindrical gap, and angle between suspension axis of sensitive element and generatrix of a cylinder is 35°15′52″.

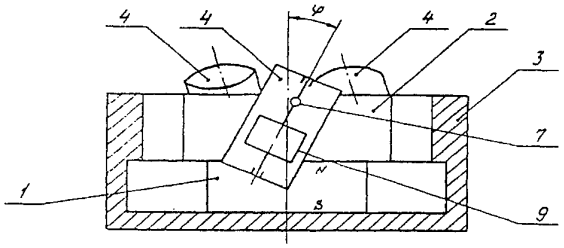


Fig. 7. Design variant of the triad with a cylindrical gap of a magnetic system and arrangement of axes of accelerometer support at angle φ to generatrix of cylindrical surfaces limiting this gap:

1 - internal constant magnet; 2 - internal pole piece; 3 - external magnetic circuit; 4 - single-axis accelerometer; 7 - pendulum; 9 - torquer frame

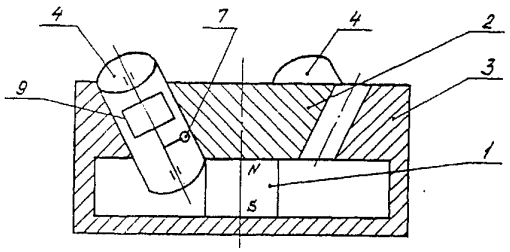


Fig. 8. Design variant of the triad with a conic gap of a magnetic system and arrangement of axes of accelerometer support at an angle a to generatrix, limiting a gap of cone surfaces, with vertex of a cone looking onto the opposite side relative to a magnet

1 - internal constant magnet; 2 - internal pole piece; 3 - external magnetic circuit; 4 - single-axis accelerometer; 7 - pendulum; 9 - torquer frame

Variants of using sensitive elements with pendulums lying in the frame plane are given in Fig. 8,9. Here a magnetic system has conic gaps with angle α, and the axes of symmetry of sensitive elements have additional inclination for angle β relative to generatrix of a conic gap. The measuring axes of sensitive elements are skew lines, remaining mutually perpendicular, if the following condition is provided:

$$\alpha = arccos\, sin\frac{35^{\circ}15'52''}{sin\,\beta}$$

The considered design variant (Fig.9) is preferable, since a more compact arrangement of sensitive elements and consequently smaller dimensions of a triad are provided. The realized design is shown in a photo (Fig. 10). The experiments with the triad completely prove all statements of the paper.

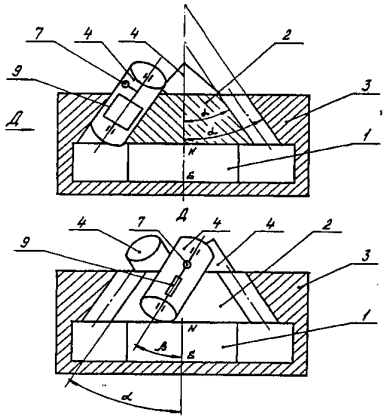


Fig. 9. Design variant of the triad with a conic gap of a magnetic system and arrangement of axes of accelerometer support at an angle to generatrix, limiting a gap of cone surfaces, with vertex of a cone looking onto the side of a magnet:

1 - internal constant magnet; 2 - internal pole piece; 3 - external magnetic circuit; 4 - single-axis accelerometer; 7 - pendulum; 9 - torquer frame

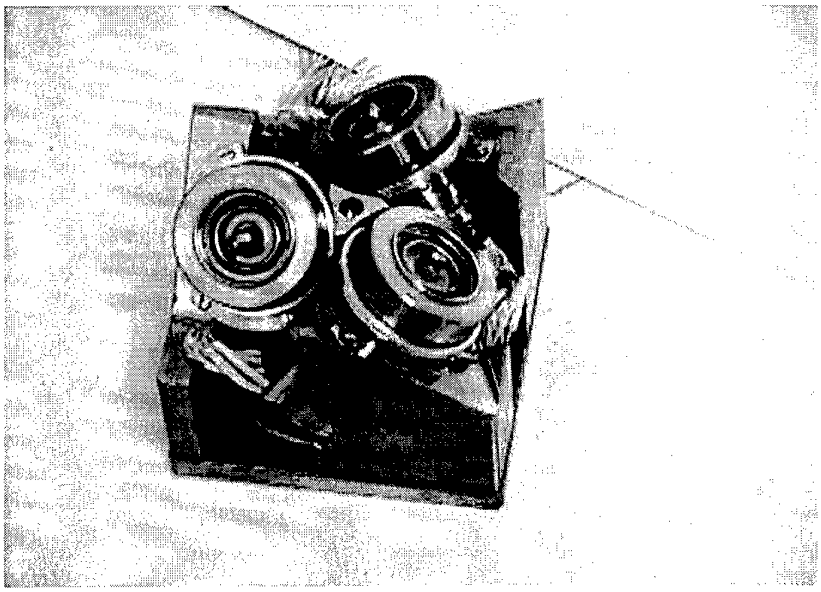


Fig. 10. Realized scheme of the triad of accelerometers

Conclusion

The use of triads with integrated magnetic system and coordinated ADC allows to solve a problem of fast preflight calibration of accelerometers of navigation systems, simplifies and improves INS accuracy. Moreover, and also provides increase in accuracy in making measurements of zenith angles in process of inclinometering oil and gas wells.

References

1. N.M. Barbour, J.M. Elwell, R.H. Setterlund, G.Schmidt. Inertial instruments: where to now? The 1st Saint Petersburg International Conference on Navigation.

2. Inertial Technology for the future. // IEEE Transactions of Aerospace and Electronic Systems. V AES-20 No 4, July 1984.

3. S.F.Konovalov, G.M.Novoselov, A.V.Polynkov, A.A.Trunov, V.V.Yurasov. Methods and facilities of accelerometer triads tests. // The 5th Saint Petersburg International Conference on Integrated Navigation Systems. 1998.

4. Konovalov S.F., Kurtyukov V.A., Yurasov V.V. and others. Three-component (axis) accelerometer. PCT the application, № International publication WO 97/10513 of 20.03.97

INFLUENCE OF PRESTRESS ON DYNAMICS OF HEMISPHERICAL RESONATOR GYROSCOPE

M.Y. Shatalov (PhD),
Council of Scientific and Industrial Research (CSIR),
Materials and Manufacturing Department, Integrated Sensing Systems,
PO Box 395, Pretoria 0001, South Africa

B.S. Lunin(PhD),
M. Lomonosov Moscow State University
Department of Chemistry,
Moscow, Russia

Abstract

Inner prestresses result from the mechanical treatment of a resonator or from inertial motion of the gyroscope, for example, its rotation. Their analysis is complex from the theoretical viewpoint because it is necessary to consider the nonlinear geometry of the shell. It is also difficult from the view point of numerical analysis by means of finite element methods due to infinity of different possibilities for inner prestress distributions. This problem of influence of the prestresses on dynamics of a hemispherical resonator gyroscope is investigated analytically. On the basis of analysis of nonlinear geometry of a hemisphere the terms responsible for the prestresses are introduced in the expression for the strain energy of the resonator and the problem is solved in linear approximation. The example of an important case of a particular distribution of local stresses is considered. The special case of influence of centrifugal forces on the dynamics of a rotating gyroscope is investigated.

Introduction

Inner stresses in resonators of vibratory gyroscopes result from the structure prestress in the process of manufacturing. They can also be generated by intensive motion of the foundation, for example its rotation or linear acceleration. The manufacturing prestress provides conditions for a distributed system of micro cracks especially in the vicinity of the resonator rim, lateral surfaces and stem.These cracs can relax, propagate and sometimes damage the resonator during the gyro operational life. The prestress due to micro cracks results in splitting of natural frequencies of a gyro resonator as well as in inhomogeneous damping of the resonator's oscillations. The inner stresses of the resonator can be tested in the laboratory which leads to the problem of prediction of gyro performance based on these measurements.

Despite the obvious practical importance of the problem it is necessary to stress the lack of relevant theoretical investigations in this field. Some authors ([1], [2], [3], [4]) failed to properly solve the related problem of influence of centrifugal forces on natural frequencies of rotating structures (this problem was correctly solved by Bryan [5] for the case of thin rotating ring in 1890 !). There were some attempts to solve the problems of vibratory gyro dynamics by finite element method (FEM) technique ([6]). The main drawback of FEM - based approach is in lack of theoretical background. ***It is necessary to understand that for proper solution of the linear problem of dynamics of a resonator with prestress the correspondent additional nonlinear terms of the strain components must be taken into consideration.*** This circumstance substantially complicates the problem and was not taken into consideration in the abovementioned papers [1] - [3] and book [4]. As for FEM - technique not all FEM packages support the dynamics of elastic structures with internal prestress. Moreover, in FEM applications it is necessary to change the model and re-mesh it for every particular distribution of inner stresses or input angular rate which is obviously a time consuming operation.

Paper presented at the RTO SCI International Conference on "Integrated Navigation Systems", held at "Elektropribor", St. Petersburg, Russia, 24-26 May 1999, and published in RTO MP-43.

The conclusion is that the theoretical investigation of dynamics of vibratory gyroscopes with prestress is very important. The analysis presented highlights the theoretical background and physical essence of the effects and results in a simple algorithm of measurement of inner stresses with subsequent estimation of gyro performance. The abovementioned problem was solved for the case of a hemispherical resonator gyroscope (HRG) but the approach is applicable for a variety of vibratory gyroscopes, including micromachined ones. An additional advantage of the proposed approach is the possibility of combination of theoretical investigation with "numerical FEM experiment" in the cases where the relevant theoretical solutions are unavailable and results from an approximate procedures of Rayleigh-Ritz type are tedious and cumbersome. In this case one can solve the corresponding FEM problem ***only once*** for an ***ideal*** construction of vibratory gyroscope without any perturbations and rotation. This solution can be used in the subsequent theoretical analysis of the problem.

Mathematical Model

In the model geometrical parameters (radius of mid surface, thickness, rim) and physical ones (mass density, model of elastic isotropic body with constant modulus of elasticity and Poisson's ratio) of the hemispherical resonator are assumed to be ideal. Damping effects are neglected. The dynamics of gyro resonator is considered in terms of linear model with geometrical nonlinearities for proper introduction of the prestress effects. The Novozhilov-Gol'denveizer model of thin shells is used. Imperfections are introduced in the model through an inhomogeneous distribution of inner stresses.

The method of analysis is based on derivation of the expressions of kinetic and strain energies of the resonator as well as the expression for work by inner stresses. The expressions are combined in Lagrangian of the system with subsequent simplification for the case of a single operational mode and derivation of equations of motion. These equations are solved by the method of perturbation and the results are discussed.

Kinetic, Strain Energies and Work of Inner Stresses

Hemispherical resonator geometry and the coordinate system used in the analysis shown in the Fig. 1.

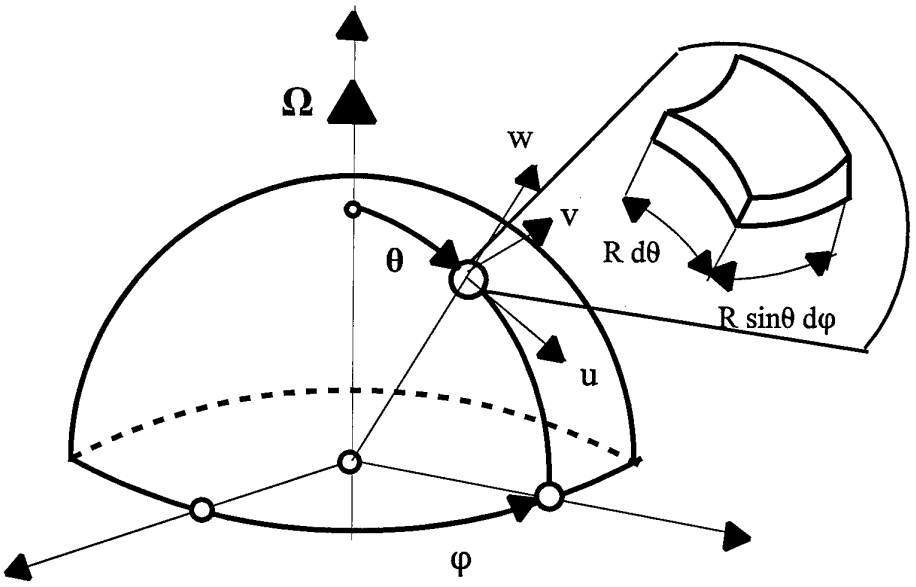


Fig. 1

where u, v - tangential and w - radial displacements of the resonator's element; θ - latitudinal and φ - l ongitudinal (polar) angles of the resonator; Ω - input inertial angular rate of the foundation.

Kinetic energy of the resonator:

$$T = \frac{\rho h R^2}{2} \int_0^{2\pi} \int_0^{\frac{\pi}{2}} V^2(t,\theta,\varphi) \sin\theta \, d\theta \, d\varphi \tag{1}$$

where ρ - mass density; h - thickness; R - radius of the mid surface of the resonator and

$$\begin{aligned} &\vec{V} = \left[\dot{u},\dot{v},\dot{w}\right]^T + \vec{\Omega} \times \vec{r}\,; \quad \vec{\Omega} = \left[-\Omega\sin\theta, 0, \Omega\cos\theta\right]^T; \\ &\vec{r} = \left[u, v, R+w\right]^T \end{aligned} \tag{2}$$

Strain energy of the resonator (Novozhilov-Gol'denveizer model of thin shells):

$$P = \frac{EhR^2}{2(1-\nu)} \int_0^{2\pi} \int_0^{\frac{\pi}{2}} \left\{ \begin{aligned} &\left[(\varepsilon_\theta{}^2 + \varepsilon_\varphi{}^2) - 2(1-\nu)(\varepsilon_\theta \varepsilon_\varphi - \frac{\omega^2}{4}) \right] + \\ &\frac{1}{12}\left(\frac{h}{R}\right)^2 \left[(\kappa_\theta{}^2 + \kappa_\varphi{}^2) - 2(1-\nu)(\kappa_\theta \, \kappa_\varphi \, - \tau^2) \right] \end{aligned} \right\} \sin\theta \, d\theta \, d\varphi \tag{3}$$

where E - modulus of elasticity; ν - Poisson's ratio and the strain-displacements relationships are:

$$\left\{ \begin{aligned} &\varepsilon_\theta = \frac{1}{R}\left(\frac{\partial u}{\partial \theta} + w\right); \quad \varepsilon_\varphi = \frac{1}{R\sin\theta}\left(\frac{\partial v}{\partial \varphi} + \cos\theta \, u + \sin\theta \, w\right); \quad \kappa_\theta = -\frac{1}{R}\left(\frac{\partial^2 w}{\partial \theta^2} - \frac{\partial u}{\partial \theta}\right); \\ &\omega = \frac{1}{R\sin\theta}\left(\frac{\partial u}{\partial \varphi} + \sin\theta \frac{\partial v}{\partial \theta} - \cos\theta \, v\right); \quad \kappa_\varphi = -\frac{1}{R\sin\theta}\left[\frac{1}{\sin\theta}\frac{\partial^2 w}{\partial \varphi^2} + \cos\theta\left(\frac{\partial w}{\partial \theta} - u\right) - \frac{\partial v}{\partial \varphi}\right]; \\ &\tau = -\frac{1}{R\sin\theta}\left[\frac{\partial^2 w}{\partial \theta \partial \varphi} - \frac{\partial u}{\partial \varphi} - \cot\theta \frac{\partial w}{\partial \varphi} - \sin\theta \frac{\partial w}{\partial \theta} - \cos\theta \, v\right]; \end{aligned} \right. \tag{4}$$

Work of inner stresses of the resonator:

$$W = -hR^2 \int_0^{2\pi} \int_0^{\frac{\pi}{2}} \left[\sigma_{\theta\theta}(\theta,\varphi)\Delta e_{\theta\theta} + \sigma_{\varphi\varphi}(\theta,\varphi)\Delta e_{\varphi\varphi} + \sigma_{\theta\varphi}(\theta,\varphi)\Delta e_{\theta\varphi} \right] \sin\theta \, d\theta \, d\varphi \tag{5}$$

where $\sigma_{\theta\theta}(\theta,\varphi)$; $\sigma_{\varphi\varphi}(\theta,\varphi)$; $\sigma_{\theta\varphi}(\theta,\varphi)$ - inner stresses (prestress) of the resonator and $\Delta e_{\theta\theta}$; $\Delta e_{\varphi\varphi}$; $\Delta e_{\theta\varphi}$ -nonlinear components of the strain-displacement relationships are as follows:

$$\Delta e_{\theta\varphi} = \frac{1}{2R^2}\left[\left(\frac{\partial u}{\partial\theta}+w\right)^2+\left(\frac{\partial v}{\partial\theta}\right)^2+\left(\frac{\partial w}{\partial\theta}-u\right)^2\right];$$

$$\Delta e_{\varphi\varphi} = \frac{1}{2R^2\sin^2\theta}\left[\left(\frac{\partial u}{\partial\varphi}-\cos\theta\, v\right)^2+\left(\frac{\partial v}{\partial\varphi}+\cos\theta\, u+\sin\theta\, w\right)^2+\left(\frac{\partial w}{\partial\varphi}-\sin\theta\, v\right)^2\right]; \quad (6)$$

$$\Delta e_{\theta\varphi} = \frac{1}{R^2\sin\theta}\left[\begin{array}{l}\left(\frac{\partial w}{\partial\theta}-u\right)\left(\frac{\partial w}{\partial\varphi}-\sin\theta\, v\right)+\left(\frac{\partial u}{\partial\theta}+w\right)\left(\frac{\partial u}{\partial\varphi}-\cos\theta\, v\right)+\\ \left(\frac{\partial v}{\partial\theta}\right)\left(\frac{\partial v}{\partial\varphi}+\cos\theta\, u+\sin\theta\, w\right)\end{array}\right]$$

The Lagrangian of the system is as follows:

$$L = T - P + W = L\left(\dot{u},\dot{v},\dot{w},u,v,w,u'_\theta,v'_\theta,w'_\theta,u'_\varphi,v'_\varphi,w'_\varphi,w''_{\theta\theta},w''_{\varphi\varphi},w''_{\theta\varphi}\right) \quad (7)$$

On the basis of this Lagrangian the system of partial differential equations as well as the full set of boundary conditions can be formulated. Of course this problem is very difficult and it can be simplified and reduced to the equivalent lumped parameters system by means of consideration of a singule n - operational mode as follows:

$$\left\{\begin{array}{l}\begin{bmatrix}u(n,\theta,\varphi,t)\\ w(n,\theta,\varphi,t)\end{bmatrix} = \begin{bmatrix}X_n(\theta)\\ Z_n(\theta)\end{bmatrix}\left[a_n(t)\cos n\varphi + b_n(t)\sin n\varphi\right];\\ v(n,\theta,\varphi,t) = Y_n(\theta)\left[-a_n(t)\sin n\varphi + b_n(t)\cos n\varphi\right]\end{array}\right. \quad (8)$$

Hence the corresponding kinetic energy (1) can be rewritten as follows:

$$T\left(\dot{a}_n,\dot{b}_n,a_n,b_n\right) = \pi\rho hR^2\left\{I_0\frac{\dot{a}_n^2+\dot{b}_n^2}{2}+2\Omega I_1\left(a_n\dot{b}_n-b_n\dot{a}_n\right)+\Omega^2\left[I_2\left(\frac{a_n^2+b_n^2}{2}\right)\right]\right\} \quad (9)$$

The corresponding strain energy (3) and work of inner stresses (5) are:

$$\left\{\begin{array}{l}P = \frac{\pi Eh}{2\left(1-\nu^2\right)}\left[I_3+\frac{1}{12}\left(\frac{h}{R}\right)^2 I_4\right];\\ W = -2\pi h\left[\begin{array}{l}\sum_{i=1}^{N}\left[\sigma_{\theta\theta}^{(0)i}I_5^{(i)}+\sigma_{\varphi\varphi}^{(0)i}I_6^{(i)}\right]\frac{a_n^2+b_n^2}{2}+\sum_{i=1}^{N}\left[\sigma_{\theta\theta}^{(c)i}I_7^{(i)}+\sigma_{\varphi\varphi}^{(c)i}I_8^{(i)}+\sigma_{\theta\varphi}^{(s)i}I_9^{(i)}\right]\frac{a_n^2-b_n^2}{2}+\\ \sum_{i=1}^{N}\left[\sigma_{\theta\theta}^{(s)i}I_7^{(i)}+\sigma_{\varphi\varphi}^{(s)i}I_8^{(i)}-\sigma_{\theta\varphi}^{(c)i}I_9^{(i)}\right]a_nb_n\end{array}\right]\end{array}\right. \quad (10)$$

where N - number of layers in which the inner stresses are measured and I_i are definite integrals:

$$I_0 = \int_0^{\pi/2}\left[X_n^2(\theta)+Y_n^2(\theta)+Z_n^2(\theta)\right]\sin\theta\,d\theta;\quad I_1 = \int_0^{\pi/2}Y_n(\theta)\left[X_n(\theta)\cos\theta+Z_n(\theta)\sin\theta\right]\sin\theta\,d\theta;$$

$$I_2 = \int_0^{\pi/2}\left[X_n^2(\theta)\cos^2\theta+Y_n^2(\theta)+Z_n^2(\theta)\sin^2\theta+X_n(\theta)Z_n(\theta)\sin 2\theta\right]\sin\theta\,d\theta;$$

$$I_3 = \int_0^{\pi/2}\left\{\left[\tilde{\varepsilon}_\theta(\theta)+\tilde{\varepsilon}_\varphi(\theta)\right]^2-2(1-\nu)\left[\tilde{\varepsilon}_\theta(\theta)\tilde{\varepsilon}_\varphi(\theta)-\frac{\omega^2}{4}\right]\right\}\sin\theta\,d\theta;$$

$$I_4 = \int_0^{\pi/2}\left\{\left[\tilde{\kappa}_\theta(\theta)+\tilde{\kappa}_\varphi(\theta)\right]^2-2(1-\nu)\left[\tilde{\kappa}_\theta(\theta)\tilde{\kappa}_\varphi(\theta)-\tau^2\right]\right\}\sin\theta\,d\theta; \qquad (11)$$

$$I_5^{(i)} = \int_{\theta^{(i-1)}}^{\theta^{(i)}}\Delta e_{\theta\theta}^{(0)\,i}(\theta)\sin\theta\,d\theta;\quad I_6^{(i)} = \int_{\theta^{(i-1)}}^{\theta^{(i)}}\Delta e_{\varphi\varphi}^{(0)\,i}(\theta)\sin\theta\,d\theta;\quad I_7^{(i)} = \int_{\theta^{(i-1)}}^{\theta^{(i)}}\Delta e_{\theta\theta}^{(c)\,i}(\theta)\sin\theta\,d\theta;$$

$$I_8^{(i)} = \int_{\theta^{(i-1)}}^{\theta^{(i)}}\Delta e_{\varphi\varphi}^{(c)\,i}(\theta)\sin\theta\,d\theta;\quad I_9^{(i)} = \int_{\theta^{(i-1)}}^{\theta^{(i)}}\Delta e_{\theta\varphi}^{(c)\,i}(\theta)\sin\theta\,d\theta;\qquad (i=1,\ldots,N)$$

where $\theta^{(i)}$ (i = 0, ... ,N) are latitudinal boundary angles of the layers $\theta^{(0)} = 0$; $\theta^{(N)} = \pi/2$ and

$$\tilde{\varepsilon}_\theta(\theta) = \frac{dX_n(\theta)}{d\theta}+Z_n(\theta);\quad \tilde{\varepsilon}_\varphi(\theta) = -nY_n(\theta)+\cos\theta\,X_n(\theta)+\sin\theta\,Z_n(\theta);$$

$$\tilde{\omega}(\theta) = \frac{1}{\sin\theta}\left[nX_n(\theta)+\sin\theta\frac{dY_n(\theta)}{d\theta}-\cos\theta\,Y_n(\theta)\right];\quad \tilde{\kappa}_\theta(\theta) = -\left[\frac{d^2Z_n(\theta)}{d\theta^2}-\frac{dX_n(\theta)}{d\theta}\right];$$

$$\tilde{\kappa}_\varphi(\theta) = \frac{1}{\sin\theta}\left[\frac{n^2}{\sin\theta}Z_n(\theta)+\cos\theta\left[X_n(\theta)-\frac{dZ_n(\theta)}{d\theta}\right]-nY_n(\theta)\right]; \qquad (12)$$

$$\tilde{\tau}(\theta) = \frac{1}{\sin\theta}\left\{n\left[X_n(\theta)-\frac{dZ_n(\theta)}{d\theta}+\cot\theta\,Z_n(\theta)\right]+\sin\theta\frac{dZ_n(\theta)}{d\theta}-\cos\theta\,Y_n(\theta)\right\}$$

The inner stresses in the i - layer can be expanded in the Fourier series. For our further purposes the constant and 2n harmonic components are necessary:

$$\begin{bmatrix}\sigma_{\theta\theta}^{(i)}(\theta,\varphi)\\ \sigma_{\varphi\varphi}^{(i)}(\theta,\varphi)\\ \sigma_{\theta\varphi}^{(i)}(\theta,\varphi)\end{bmatrix} = \begin{bmatrix}\sigma_{\theta\theta}^{(0)\,i}(\theta)\\ \sigma_{\varphi\varphi}^{(0)\,i}(\theta)\\ \sigma_{\theta\varphi}^{(0)\,i}(\theta)\end{bmatrix}+\begin{bmatrix}\sigma_{\theta\theta}^{(c)\,i}(\theta)\\ \sigma_{\varphi\varphi}^{(c)\,i}(\theta)\\ \sigma_{\theta\varphi}^{(c)\,i}(\theta)\end{bmatrix}\cos 2n\varphi+\begin{bmatrix}\sigma_{\theta\theta}^{(s)\,i}(\theta)\\ \sigma_{\varphi\varphi}^{(s)\,i}(\theta)\\ \sigma_{\theta\varphi}^{(s)\,i}(\theta)\end{bmatrix}\sin 2n\varphi \qquad (13)$$

Hence it follows from (6) and (8) that:

$$\begin{bmatrix}\Delta e_{\theta\theta}(\theta,\varphi,a_n,b_n)\\ \Delta e_{\varphi\varphi}(\theta,\varphi,a_n,b_n)\end{bmatrix} = \begin{bmatrix}\Delta e_{\theta\theta}^{(0)}(\theta)\\ \Delta e_{\varphi\varphi}^{(0)}(\theta)\end{bmatrix}\frac{a_n^2+b_n^2}{2}+\begin{bmatrix}\Delta e_{\theta\theta}^{(c)}(\theta)\\ \Delta e_{\varphi\varphi}^{(c)}(\theta)\end{bmatrix}\frac{a_n^2-b_n^2}{2}\cos 2n\varphi+\begin{bmatrix}\sigma_{\theta\theta}^{(s)}(\theta)\\ \sigma_{\varphi\varphi}^{(s)}(\theta)\end{bmatrix}a_nb_n\sin 2n\varphi \qquad (14)$$

and

$$\Delta e_{\theta\varphi}(\theta,\varphi,a_n,b_n)=\Delta e^{(c)}_{\theta\varphi}(\theta)\,a_n b_n\cos 2n\varphi+\Delta e^{(s)}_{\theta\varphi}(\theta)\frac{a_n^2-b_n^2}{2}\sin 2n\varphi \tag{15}$$

where from (6) and (8):

$$\begin{aligned}
&\Delta e^{(0)}_{\theta\theta}(\theta)=\frac{1}{2}\left\{\left[\frac{dX_n(\theta)}{d\theta}+Z_n(\theta)\right]^2+\left[\frac{dZ_n(\theta)}{d\theta}-X_n(\theta)\right]^2+\left[\frac{dY_n(\theta)}{d\theta}\right]^2\right\};\\
&\Delta e^{(c)}_{\theta\theta}(\theta)=\Delta e^{(s)}_{\theta\theta}(\theta)=\frac{1}{2}\left\{\left[\frac{dX_n(\theta)}{d\theta}+Z_n(\theta)\right]^2+\left[\frac{dZ_n(\theta)}{d\theta}-X_n(\theta)\right]^2-\left[\frac{dY_n(\theta)}{d\theta}\right]^2\right\};\\
&\Delta e^{(0)}_{\varphi\varphi}(\theta)=\frac{1}{2\sin^2\theta}\left\{\left[nX_n(\theta)-\cos\theta\,Y_n(\theta)\right]^2+\left[nZ_n(\theta)-\sin\theta\,Y_n(\theta)\right]^2+\left[-nY_n(\theta)+\cos\theta\,X_n(\theta)+\sin\theta\,Z_n(\theta)\right]^2\right\};\\
&\Delta e^{(c)}_{\varphi\varphi}(\theta)=\Delta e^{(s)}_{\varphi\varphi}(\theta)=\frac{1}{2\sin^2\theta}\left\{\begin{array}{l}-\left[nX_n(\theta)-\cos\theta\,Y_n(\theta)\right]^2-\left[nZ_n(\theta)-\sin\theta\,Y_n(\theta)\right]^2+\\ \left[-nY_n(\theta)+\cos\theta\,X_n(\theta)+\sin\theta\,X_n(\theta)\right]^2\end{array}\right\};\\
&\Delta e^{(c)}_{\theta\varphi}(\theta)=-\Delta e^{(s)}_{\theta\varphi}(\theta)=\frac{1}{\sin\theta}\left\{\begin{array}{l}\left[\frac{dZ_n(\theta)}{d\theta}-X_n(\theta)\right]\left[nZ_n(\theta)-\sin\theta\,Y_n(\theta)\right]+\\ \left[\frac{dX_n(\theta)}{d\theta}+Z_n(\theta)\right]\left[nX_n(\theta)-\cos\theta\,Y_n(\theta)\right]+\frac{dY_n(\theta)}{d\theta}\left[-nY_n(\theta)+\cos\theta\,X_n(\theta)+\sin\theta\,X_n(\theta)\right]\end{array}\right\}
\end{aligned} \tag{16}$$

The form factors $X_n(\theta);\,Y_n(\theta);\,Z_n(\theta)$ can be derived from the solution of the problem on the bases of equations of motion and boundary conditions derived from the Lagrangian (7). There expressions are quite cumbersome and alternative approaches can be used. For example, we can use the Rayleigh-Ritz method, or can simulate the structure by finite element method and interpolate the form factors by the corresponding functions. In our paper we use the third approach which is based on application of Rayleigh solution for inextensional deformations of bell-type shells:

$$X_n(\theta)=-\sin\theta\tan^n\left(\frac{\theta}{2}\right);\quad Y_n(\theta)=\sin\theta\tan^n\left(\frac{\theta}{2}\right);\quad Z_n(\theta)=(n+\cos\theta)\tan^n\left(\frac{\theta}{2}\right); \tag{17}$$

In this case (16) can be rewritten as follows:

$$\begin{aligned}
&\tilde{\varepsilon}_\theta=\tilde{\varepsilon}_\varphi=\tilde{\omega}=0;\quad \tilde{\kappa}_\theta=-\tilde{\kappa}_\varphi=\tilde{\tau}=-\frac{n\left(n^2-1\right)}{\sin^2\theta}\tan^n\left(\frac{\theta}{2}\right);\\
&\Delta e^{(0)}_{\theta\theta}(\theta)=\frac{1}{2}(n+\cos\theta)^2\left(\frac{n^2}{\sin^2\theta}+1\right)\tan^{2n}\left(\frac{\theta}{2}\right);\quad \Delta e^{(c)}_{\theta\theta}(\theta)=\Delta e^{(s)}_{\theta\theta}(\theta)=\frac{1}{2}(n+\cos\theta)^2\left(\frac{n^2}{\sin^2\theta}-1\right)\tan^{2n}\left(\frac{\theta}{2}\right);\\
&\Delta e^{(0)}_{\varphi\varphi}(\theta)=-\Delta e^{(c)}_{\varphi\varphi}(\theta)=-\Delta e^{(s)}_{\varphi\varphi}(\theta)=\frac{1}{2}\left[\frac{n^2\left(n^2+1+2n\cos\theta\right)}{\sin^2\theta}-\left(2n^2-1\right)\right]\tan^{2n}\left(\frac{\theta}{2}\right);\\
&\Delta e^{(c)}_{\theta\varphi}(\theta)=-\Delta e^{(s)}_{\theta\varphi}(\theta)=n\left(n+\cos\theta\right)\left[\frac{n(n+\cos\theta)}{\sin^2\theta}-1\right]\tan^{2n}\left(\frac{\theta}{2}\right)
\end{aligned} \tag{18}$$

For the particular case **n = 2** the integrals (11) have the following values:

$$\begin{aligned}
&I_0=1.5296;\quad I_1=0.4237;\quad I_2=1.0592;\quad I_3=0;\quad I_4=42(1-\nu);\quad I_5=3.053;\\
&I_6=-I_8=2.1589;\quad I_7=1.947;\quad I_9=4
\end{aligned} \tag{19}$$

Particular case of axisymmetric deformation due to centrifugal forces of external rotation

The system of equations with accuracy $O((h/R)^2)$ for determination of displacements of the hemispherical shell:

$$\begin{cases} \dfrac{du}{d\theta} + \cot\theta\, u + 2w &= \dfrac{1-\nu}{2}\dfrac{\rho R^3}{E}\Omega^2 \sin^2\theta; \\ \sin\theta \dfrac{dw}{d\theta} &= \dfrac{1-\nu}{2}\dfrac{\rho R^3}{E}\Omega^2 \sin^2\theta \cos\theta \end{cases} \tag{20}$$

with boundary conditions: u(0) = 0; w(0) = 0. The solution is:

$$u(\theta) \equiv 0; \quad v(\theta) \equiv 0; \quad w(\theta) = \frac{1-\nu}{4}\frac{\rho R^3}{E}\Omega^2 \sin^2\theta \tag{21}$$

Hence the strains (4) and stresses of the hemisphere due to external rotation are:

$$\begin{aligned} &\varepsilon_\theta = \varepsilon_\varphi = \frac{w}{R} = \frac{1-\nu}{4}\frac{\rho R^2}{E}\Omega^2 \sin^2\theta; \quad \omega \equiv 0; \\ &\sigma_{\theta\theta} = \sigma_{\varphi\varphi} = \frac{E}{1-\nu}\varepsilon_\theta = \frac{E}{1-\nu}\varepsilon_\varphi = \frac{\rho R^2}{4}\Omega^2 \sin^2\theta; \quad \sigma_{\theta\varphi} \equiv 0 \end{aligned} \tag{22}$$

The correspondent component of the work of inner stresses due to centrifugal forces of external rotation:

$$W^{(\Omega^2)} = -\pi\rho h E R^2 \Omega^2 I_{10} \frac{a_n^2 + b_n^2}{2} \tag{23}$$

where

$$I_{10} = \int_0^{\pi/2} \Delta e_{\theta\theta}(\theta) \sin^3\theta\, d\theta \tag{24}$$

In the case n = 2 : $I_{10} = 2.7178$. Let us compare this result with $I_2 = 1.0592$ (see (19) and third term of (9)):

$$T^{(\Omega^2)} + W^{(\Omega^2)} = -\pi\rho h R^2 \Omega^2 (I_{10} - I_2) \frac{a_n^2 + b_n^2}{2} \tag{25}$$

and hence the strain energy term dominates the kinetic energy one.

Lagrangian and System of Equations

The Lagrangian of the system and dynamical equations of motion

$$L = J_0 \frac{\dot{a}_n^2 + \dot{b}_n^2}{2} + 2\Omega J_1 \left(a_n \dot{b}_n - b_n \dot{a}_n\right) - \left(J_2 - \Omega^2 J_3\right) \frac{a_n^2 + b_n^2}{2} - J_4 \frac{a_n^2 - b_n^2}{2} - J_5 a_n b_n \tag{26}$$

where

$$J_0 = \pi\rho hR^2 I_0;\ J_1 = \pi\rho hR^2 I_1;\ J_2 = \frac{\pi Eh}{1-\nu^2}\left[I_3 + \frac{1}{12}\left(\frac{h}{R}\right)^2 I_4\right] + 2\pi h\sum_{i=1}^{N}\left[\sigma_{\theta\theta}^{(0)\,i} I_5^{(i)} + \sigma_{\varphi\varphi}^{(0)\,i} I_6^{(i)}\right];$$

$$J_3 = \pi\rho hR^2\left(I_{10} - I_2\right);\quad J_4 = 2\pi h\sum_{i=1}^{N}\left[\sigma_{\theta\theta}^{(c)\,i} I_7^{(i)} + \sigma_{\varphi\varphi}^{(c)\,i} I_8^{(i)} + \sigma_{\theta\varphi}^{(s)\,i} I_9^{(i)}\right]; \qquad (27)$$

$$J_5 = 2\pi h\sum_{i=1}^{N}\left[\sigma_{\theta\theta}^{(s)\,i} I_7^{(i)} + \sigma_{\varphi\varphi}^{(s)\,i} I_8^{(i)} - \sigma_{\theta\varphi}^{(c)\,i} I_9^{(i)}\right];$$

The equations of motion can be represented in matrix form:

$$J_0\begin{bmatrix}1 & 0\\ 0 & 1\end{bmatrix}\begin{bmatrix}\ddot{a}_n\\ \ddot{b}_n\end{bmatrix} + 2\Omega J_1\begin{bmatrix}0 & -1\\ 1 & 0\end{bmatrix}\begin{bmatrix}\dot{a}_n\\ \dot{b}_n\end{bmatrix} + \left\{\left(J_2 + \Omega^2 J_3\right)\begin{bmatrix}1 & 0\\ 0 & 1\end{bmatrix} + \begin{bmatrix}J_4 & J_5\\ J_5 & -J_4\end{bmatrix}\right\}\begin{bmatrix}a_n\\ b_n\end{bmatrix} = 0 \qquad (28)$$

Solution of Equations of Motion

Solution of the system of equations (28) can be achieved by means of transformation of coordinates $\left(a_n, b_n, \dot{a}_n, \dot{b}_n\right) \to \left(P_n, Q_n, \theta_n, \psi_n\right)$:

$$a_n \cos n\varphi + b_n \sin n\varphi = P_n \cos n\left(\varphi - \theta_n\right)\cos\left(\lambda - \psi_n\right) + Q_n \sin n\left(\varphi - \theta_n\right)\sin\left(\lambda - \psi_n\right) \qquad (29)$$

the second equation is achieved from the system (28). In the new coordinates the averaged solution can be represented as follows:

$$\begin{cases}
\dot{P}_n = -\dfrac{1}{2}\sqrt{\dfrac{1}{J_0\left(J_2 + \Omega^2 J_3\right)}}\, Q_n\left[J_4 \sin\left(2n\theta_n\right) - J_5 \cos\left(2n\theta_n\right)\right];\\
\dot{Q}_n = \dfrac{1}{2}\sqrt{\dfrac{1}{J_0\left(J_2 + \Omega^2 J_3\right)}}\, P_n\left[J_4 \sin\left(2n\theta_n\right) - J_5 \cos\left(2n\theta_n\right)\right];\\
n\dot{\theta}_n = -\Omega\dfrac{J_1}{J_0} - \sqrt{\dfrac{1}{J_0\left(J_2 + \Omega^2 J_3\right)}}\,\dfrac{P_n Q_n}{P_n^2 - Q_n^2}\left[J_4 \cos\left(2n\theta_n\right) + J_5 \sin\left(2n\theta_n\right)\right];\\
\dot{\psi}_n = -\sqrt{\dfrac{1}{J_0\left(J_2 + \Omega^2 J_3\right)}}\,\dfrac{P_n^2 + Q_n^2}{P_n^2 - Q_n^2}\left[J_4 \cos\left(2n\theta_n\right) + J_5 \sin\left(2n\theta_n\right)\right];
\end{cases} \qquad (30)$$

This equation can be solved for a particular set of initial conditions by numerical methods. The solution describes an evolution of vibrating pattern in terms of its main and quadrature amplitudes (P_n and Q_n) respectively; angle of orientation (θ_n) and the phase (ψ_n). It can be found from (28) and (30) that the inner stresses stipulated the split of natural frequencies ($\Delta\omega$) and change orientation of the main mass-stiffness axes(angle of location of the "maximum frequency" axis ψ):

$$\Delta\omega = \frac{1}{2}\sqrt{\frac{J_4^2 + J_5^2}{J_0\left(J_2 + \Omega^2 J_3\right)}};\quad \cos 4\psi = \frac{J_4}{J_4^2 + J_5^2};\qquad \sin 4\psi = \frac{J_5}{J_4^2 + J_5^2} \qquad (31)$$

Example: Let us consider a hypothetic resonator of radius R = 30 mm; h = 1.5 mm made from isotropic material with mass density $\rho = 2.65 \cdot 10^3\ N s^2 m^{-4}$; $E = 8 \cdot 10^{10}\ N m^{-2}$; ν = 0.27, rotating with angular velocity $\Omega = 3 \cdot 2\pi\ s^{-1}$ and oscillating at an elliptic (n = 2) mode. Expression (28) shows that the natural frequency increment is only $\Delta^{(\Omega^2)}\omega \approx 2.5 \cdot 10^{-3}\ Hz$ and hence the main influence of input angular rate Ω are the gyroscopic terms (see the second term in (28)). Further we will consider Ω = 0. Suppose that we estimate influence of inner stresses in five layers of the hemispherical shell defined by the angles: $\theta^{(0)} = 0$; $\theta^{(1)} = \pi/10$; $\theta^{(2)} = \pi/5$; $\theta^{(3)} = 3\pi/10$; $\theta^{(4)} = 2\pi/5$ and $\theta^{(5)} = \pi/2$. The correspondent integrals I5 - I9 for Rayleigh solution (17) are (see (11), (14), (16) and (18)):

$$\begin{bmatrix} I5 \\ I6 \\ I7 \\ I8 \\ I9 \end{bmatrix} = \begin{bmatrix} 2.81 \cdot 10^{-3} & 0.0458 & 0.2287 & 0.7468 & 2.0288 \\ 2.75 \cdot 10^{-3} & 0.0421 & 0.1905 & 0.5566 & 1.3669 \\ 2.73 \cdot 10^{-3} & 0.0404 & 0.1750 & 0.4953 & 1.2336 \\ -2.75 \cdot 10^{-3} & -0.0421 & -0.1905 & -0.5566 & -1.3669 \\ 5.48 \cdot 10^{-3} & 0.0825 & 0.3634 & 1.0342 & 2.5144 \end{bmatrix} \quad \textbf{(32)}$$

Distributions of the polar components of inner stresses in these levels are shown in Fig. 2 (other stress components are supposed to be zero):

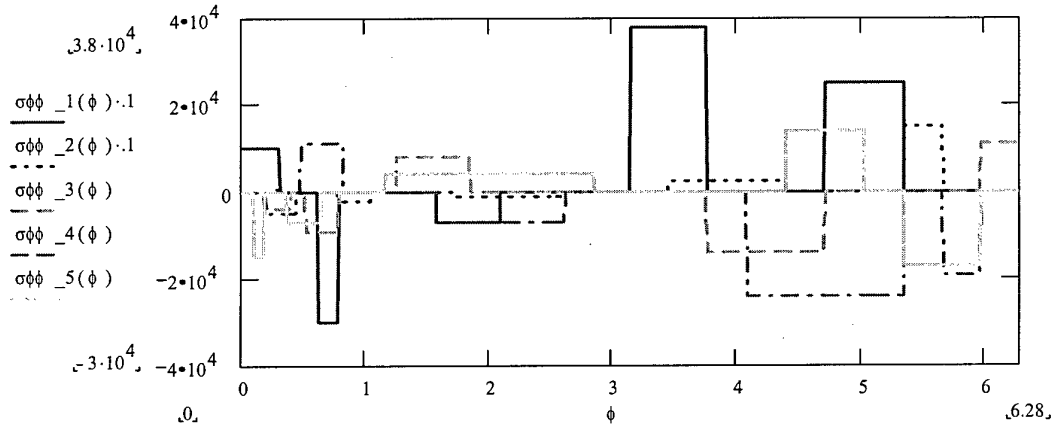


Fig. 2.

Fourier expansions results in the following values of constant and four fold cosine and sine components (see (14)):

$$\begin{bmatrix} \sigma_{\varphi\varphi}^{(0)} \\ \sigma_{\varphi\varphi}^{(c)} \\ \sigma_{\varphi\varphi}^{(s)} \end{bmatrix} = \begin{bmatrix} 5.4747 \cdot 10^4 & 9.452 \cdot 10^3 & -1.284 \cdot 10^3 & -6.514 \cdot 10^3 & 286.7 \\ 4.6225 \cdot 10^4 & -1.537 \cdot 10^3 & 3.184 \cdot 10^3 & -2.269 \cdot 10^3 & 4.346 \cdot 10^3 \\ 8.3283 \cdot 10^4 & 143.7 & 726.3 & 1.054 \cdot 10^3 & 647.1 \end{bmatrix} \quad \textbf{(33)}$$

Using the expressions (31) one can estimate the frequency split as well as orientation of the main axes:
$\Delta\omega = 0.1139\ s^{-1} = 0.018\ Hz$; $\psi = 5.29°$ (angle of location of the "maximum frequency" axis).

Algorithm of Estimation of Gyrodynamics Stipulated by Inner Stresses

1. Choose the number of operational mode (n) and number of layers (N), in which the stresses are measured, and their boundaries $\theta^{(i)}$ (i = 0, 1, ..., N);
2. Estimate the integrals $I_0 - I_{10}$ (see expressions (11) and (24)), using Rayleigh solutions (17) - (19) or interpolated values of FEM realization for the whole resonator and particular layers;
3. Measure stress distributions in N layers, expand it in a Fourier series and take into consideration the constant values and 2n harmonics (see (13));
4. Estimate the core values $J_0 \ldots J_5$ (see (27));
5. Substitute $J_0 \ldots J_5$ into expressions (30), define split of natural frequencies, orientation of the main axes of the vibrating pattern (31) and solve this system numerically for estimation of HRG dynamics for a given set of initial conditions.

Conclusions

1. A method of analysis of distributed vibratory gyroscopes with initial prestress is proposed;
2. An algorithm of estimation of gyrodynamics produced by inner stresses in the resonator is proposed;
3. Tensile inner stresses increase natural frequencies of vibratory gyroscopes; compressive inner stresses decrease these frequencies;
4. For the n^{th} operational mode the 2n-harmonics of inner stresses distribution are essential and produce a natural frequencies split as well as change the orientation of main frequency axes;
5. Centrifugal forces always generate tensile inner stresses and increase natural frequencies of vibratory gyroscopes.

References

1. Burdess J., The Dynamics of a Thin Piezoelectric Cylinder Gyroscope, Proc. Inst Mech. Eng., 1986, Vol. 200, No. C4, pp. 271-280;

2. Burdess J., Wren T., The Theory of a Piezoelectric Disc Gyroscope, IEEE Trans. on Aerosp. and El. Syst., July 1986, Vol. AES-22, pp. 410-418;

3. Putty, M., A Micromachined Vibrating Ring Gyroscope, PhD Dissertation, University of Michigan, 1995, pp. 59-67;

4. Soedel W., Vibrations of Shells and Plates, 2nd revised edition, Marcel Dekker, Inc, New Yok, 1993, pp. 340-385;

5. Bryan G., On the Beats in the Vibrations of a Revolving Cylinder or Bell, Proc. Cambridge Phil. Soc., Math Phys. Sci., Nov. 24, 1890, Vol. 7, Pt. III, pp. 101-111;

6. Kagawa Yu., et.al., Finite Element Simulation of Piezoelectric Vibrator Gyroscopes, IEEE Trans. on Ferroel. and Freq. Cont., July 1996, Vol. 3, No. 4, pp. 509-518.

Hemispherical Resonator Gyro Technology. Problems and Possible Ways of their Solutions.

(May 1999)

E.A. Izmailov, M.M. Kolesnik, A.M. Osipov, A.V. Akimov.
The Moscow Institute of Electromechanics and Automatics
5 Aviatzionny per. Moscow, 125319
Russia
Phone: 155-06-47. Fax: 152-26-31

Abstract

Key words: hemispherical resonator gyro, fundamental wave, quadrature wave.

Hemispherical Resonator Gyro (HRG) technology is now considered as one of the promising gyroscope technologies. However, to implement HRG potential capabilities one should solve a number of problems connected both with manufacturing an isotropic resonator having a high mechanical Q-factor and with conserving its parameters within the gyro.

The article, basing on the experience of developing HRGs of two standard sizes, presents some of the mentioned problems and the results of mock-up tests.

Introduction

All known gyro types are based on using either the law of conservation of momentum or moment of momentum of solid bodies, material points or fundamental particles or on using main statements of the theory of relativity. In any case, gyro quality is determined by the conservatism level of its proper sensing element and a pickup system. The more energy dissipates within the sensing element and the pickup system under other similar conditions, the lower the gyro accuracy and reliability are. From this point of view the HRG has good potential characteristics. Using inertial properties of a standing wave in a high Q-factor resonator, which is an axis-symmetrical shell, in combination with an electrostatic system for excitation, support and control of vibrations on a resonance frequency, and with a similar pick off system, the HRG is characterised by extremely low specific energy losses.

The above mentioned HRG feature is also a basis of its unique property - to preserve information on angular movement of an object during momentary interruptions in power if the HRG is used as an integrating gyro (IG). Depending upon tolerable error, this interruption can be of several dozens of seconds.

The same sensing element, depending upon switching external electronic circuits, can be used both as an IG and an angular rate sensor (ARS) [1].

As the majority of gyros, existing and being developed, the HRG is an integral system consisting of a sensing element and functional electronics which not only generates an output information signal but also provides conditions for sensing element operation. Depending upon the conditions of HRG application, this electronics can be of various complexity (also from the point of view of implemented functions), however, there are common problems and without their solving there is no sense to speak about the complexity of electronics types.

Simple design of the HRG sensing element consisting of 2-3 rigidly connected parts, and potentially low cost are also attractive. The current HRG development experience allows to predict its cost which is at least five times less than that of a laser gyro provided they have similar accuracy parameters.

Practical works on HRG creation proved the validity of main statements of fundamental investigations conducted in this field [2], [3], [4], [5], and revealed some application matters presented in this article.

Sensing Element

The HRG's sensing element is a hemispherical resonator in which a standing wave is excited by electrostatic forces. As the HRG is based on inertial properties of the standing wave, it is clear, that minimisation of energetic connection of this wave with structural elements moving in the inertial space is the main way to increase HRG accuracy. The value of this connection is subdivided into two components: dissipation of the resonator's vibratory energy and position force action on the vibrating shell.

In case of an ideally made resonator, the first component includes: energy losses in the resonator and energy losses in the electrostatic pick off and wave control systems. This means, that the resonator material and the whole process of resonator manufacture must finally provide a vibratory loop with minimum possible damping. As the resonator is a spatial shell in which the presence of "precedence" directions is prohibitive, only isotropic material can be used for its manufacturing. In addition, this material must feature minimum possible losses for internal friction. Fused quartz most fully

Paper presented at the RTO SCI International Conference on "Integrated Navigation Systems", held at "Elektropribor", St. Petersburg, Russia, 24-26 May 1999, and published in RTO MP-43.

meets these requirements. However to retain potentially high characteristics of the material in a manufactured resonator one must solve such technological problems as full removal of internal stress and a defective layer. These defects appear in the process of mechanical treatment, mainly in the process of shaping. It means that in these processes the forces appearing in the treatment zone shall be tangentially directed to the surface being treated. However, in this case internal stress and a defective layer also appear. Therefore, chemical and thermal treatment is the necessary condition to obtain a high-quality resonator.

Creation of the technological process for manufacturing a resonator on the basis of the above requirements allowed to produce resonators with the mechanical Q-factor of $(12\text{-}15)\times10^6$ for a 60-mm diameter and $(6\text{-}9)\times10^6$ for a 20-mm diameter (Figure 1).

As for electrostatic systems, there are problems connected with pick off channel optimisation and damping minimisation due to currents induced in the resonator metal coating. The latter problem is not in conflict with the requirement to minimum vibration damping in the resonator because it complies with the requirement to minimum thickness of the film provided it retains its homogeneity. But this implies additional requirements to the quality of treatment and preparation of the resonator surface to applying the metal film as well as to the process of film applying. In particular, surface quality shall comply with the optical class of cleanness for which chemical and mechanical treatment is used at the final stage.

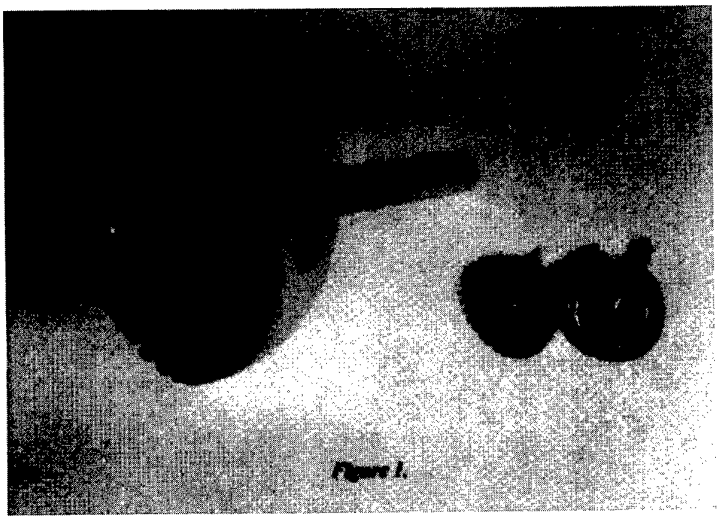
Fused quartz features the presence of nonsaturated radicals on the surface which causes absorption of different substances, including water, from the atmosphere by the surface layer. This phenomenon not only reduces the Q of the resonator but also negatively influences the quality of the applied metal film which finally deteriorates Q as well. Therefore, not only surface layer passivation is necessary but observance of technical hygiene as well.

In case of electrovacuum application of the metal film, the most promising method is the one which potentially provides maximum possible plasma homogeneity provided it has enough energetic parameters. In this case, deterioration of the resonator reference Q after metal film application does not exceed (5-10)%.

It is apparent, that the ideal resonator can not be made. As it was mentioned in well-known publications [6], this leads to resonance frequency difference and vibrations of resonator center of mass. The methodology for eliminating difference in resonance frequencies is fully elaborated at present. It includes both adjustment operations at the phase of sensing element manufacturing and structural compensation methods implemented in functional electronics. However, it should be taken into consideration that the emergence of the compensation loop of the resonance frequency residual difference requires solving the task of phase error minimisation. Elimination of the resonator mass center vibrations is a more complicated task. Solving this task is necessary because when the resonator mass center is vibrating the vibratory energy is substantially dissipated at the resonator fastening points and in body members including metallic ones. As far as this process is principally instable, the standing wave energetic connection with structural elements moving in the inertial space also becomes instable. The experience of development and investigation of 60-mm diameter and 20-mm diameter HRG mock-ups shows that with the reduction of the resonator diameter the influence degree of the resonator mass center vibrations becomes prevailing.

As it is well-known, the process of resonator symmetrization consists in mass elimination along the resonator rim perimeter. Availability or nonavailability of teeth on the resonator rim depends upon its manufacturing cost and mass elimination procedure as well, which may be much more expensive than the resonator manufacturing. If the resonator does not have teeth, mass elimination shall not cause change in material structure (at the place of elimination) and emergence of internal stress. For example, in the course of comparative experiment minor effect of the CO_2 laser radiation in the balancing zone of the resonator without teeth caused anisotropy of its mechanical Q-factor of $\delta Q \sim 40\%$ under simultaneous decrease of maximum value of $\sim 20\%$. Similar effect on the resonator with teeth practically did not influence its reference Q. Thus, if toothless resonator is used, mass must be eliminated only by chemical or ion-chemical processes. The resonator with teeth can be effectively symmetrized using CO_2 laser radiation.

The second component in the mentioned case emerges due to: force electrode system asymmetry, phase errors of control signal generation, and induced "wandering" potentials. Symmetrization of the force electrode system is implemented both at the design stage and by selecting appropriate procedures for sensing element assembly and adjustment. As the residual asymmetry (if the above procedures are rationally selected) is a small and stable magnitude, it can be eliminated algorithmically. "Wandering" potentials can be eliminated both by implementing equipotentiality for the resonator metal film and stability of permanent potentials on all electrodes. The most complicated problem is minimisation of phase errors of control signal generation. As it is connected not only with the electronic circuits errors and their instability but with noise in the data channel as well, the best decision can



be made only when control action generation algorithms, functional electronics building structure and component capabilities are considered as a whole.

Pick off Electrode System

The pick off electrode system generates both the necessary control actions on the resonator and the HRG output signal as well. It contains 8 electrodes symmetrically located relative to the resonator; the electrodes together with the resonator surface form the same number of capacitors (Figure 2). Of course, in reality neither absolute symmetry of electrodes nor their absolute identity are feasible. In addition, a real gap between electrodes and the resonator surface is not similar either. Therefore, each capacitor will have its own C_i magnitude and its deviation from an ideal angular position φ_i, where i = 1,2,..., 8. Let us write C_i in the form of an equivalent flat capacitor:

$$C_i = \frac{dS_i}{\delta_i}, \qquad (1)$$

where: d - dielectric penetrability;
S_i - an equivalent area;
δ_i - an equivalent gap.

Here and further the second resonator vibration mode is considered. Assuming that resonator vibration amplitude is $\Delta << \delta_i$, the change of the capacitor magnitude can be presented in the form of:

$$A_i = K_i \times \Delta_i, \qquad (2)$$

where: $K_i = -\frac{C_i}{\delta_i}$;

Δ_i - gap change under resonator vibrations.

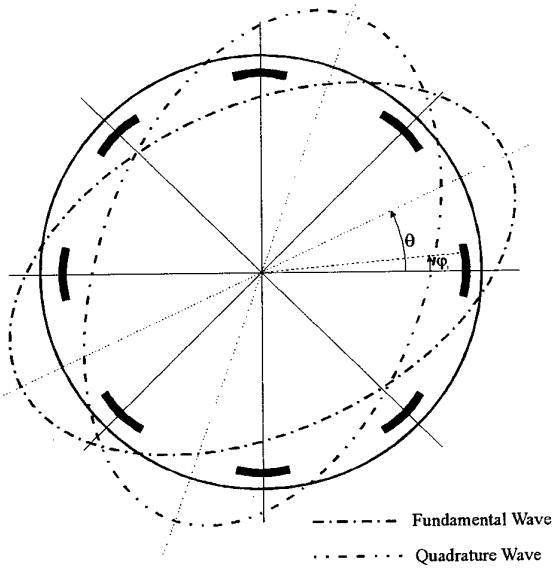


Figure 2. *The pick off electrode system.*

Let us assume that there is only the fundamental wave in the resonator and that the pick off signal is generated due to constant voltage U applied to the resonator surface. Assuming that all capacitors are connected to similar active resistors and there is no relation between them, output signals over sin and cos channels can be written with an accuracy of up to the scale factor:

$$u_s = U\Delta_f(E_1\cos 2\theta + E_2\sin 2\theta)\cos\omega t; \qquad (3)$$
$$u_c = U\Delta_f(E_3\sin 2\theta - E_4\cos 2\theta)\cos\omega t,$$

where: $E_1 = K_1\cos\varphi_1 + K_3\cos\varphi_3 + K_5\cos\varphi_5 + K_7\cos\varphi_7$;
$E_2 = K_1\sin\varphi_1 - K_3\sin\varphi_3 - K_5\sin\varphi_5 + K_7\sin\varphi_7$;
$E_3 = K_2\cos\varphi_2 + K_4\cos\varphi_4 + K_6\cos\varphi_6 + K_8\cos\varphi_8$;
$E_4 = K_2\sin\varphi_2 - K_4\sin\varphi_4 - K_6\sin\varphi_6 + K_8\sin\varphi_8$;
ω - angular frequency of resonator vibrations;
t - time.

Signals (3) are used to control resonator vibration amplitude and to determine angular position of the wave. An expression for estimation of the amplitude square will be obtained in the form of:

$$\Delta_f^2 = 0{,}25U^2\Delta_f^2\left[B_1 + B_2\left(4\theta - \xi_f\right)\right]. \qquad (4)$$

where: $B_1 = E_1^2 + E_2^2 + E_3^2 + E_4^2$;

$$B_2 = \sqrt{\left(E_1^2 + E_4^2\right)^2 + \left(E_2^2 + E_3^2\right)^2 + 2\left(E_1E_2 - E_3E_4\right)^2};$$

$$\xi_f = arctg\frac{2\left(E_1E_2 - E_3E_4\right)}{E_1^2 + E_4^2 - E_2^2 - E_3^2}.$$

Thus, in case of pick off channel non-identity the amplitude stabilisation loop will produce a false signal with a period of 4θ.

It is not difficult to show that an evaluation error of wave angular position is determined by the relation:

$$\theta' = -0{,}5arctg\frac{B_3 + B_4\sin\left(4\theta + \xi_1\right)}{B_5 + B_4\cos\left(4\theta + \xi_1\right)}, \qquad (5)$$

where: $B_3=E_2+E_4$;

$$B_4 = \sqrt{\left(E_1 - E_3\right)^2 + \left(E_4 - E_2\right)^2};$$

$B_5=E_1+E_3$;

$$\xi_1 = arctg\frac{E_4 - E_2}{E_1 - E_3}.$$

Expression (5) describes the HRG scale factor error which is also a periodic function of 4θ. In addition, there is a certain constant shift defined by a relative angular error of electrode location.

Now let us consider the display of the same asymmetry under parallel existence of fundamental and quadrature waves having different angular frequencies of ω_f and ω_q vibrations. It is evident that such a mode can exist only in case of orientation of the mentioned waves over orthogonal axes of the resonator frequency anisotropy. Subject to this factor, expression (3) can be rewritten in the following form:

$u_s=U[\Delta_f(E_1\cos2\theta+E_2\sin2\theta)\cos\omega_f t-\Delta_q(E_1\sin2\theta-E_2\cos2\theta)\sin\omega_q t];$ (6)
$u_c=U[\Delta_f(E_3\sin2\theta-E_4\cos2\theta)\cos\omega_f t+\Delta_q(E_3\cos2\theta+E_4\sin2\theta)\sin\omega_q t].$

Let us introduce an instant angular frequency ω defined by the relation:

$$\omega_f \le \omega \le \omega_q,$$

then: $\omega_f = \omega + \omega_1$ and $\omega_q = \omega - \omega_2$. (7)

Subject to (7), equation (6) is transformed into:

$u_s=U(z_1\cos\omega t - z_2\sin\omega t);$ (8)
$u_c=U(z_3\cos\omega t - z_4\sin\omega t),$

where:
$z_1=(\Delta_f E_1\cos\omega_1 t-\Delta_q E_2\sin\omega_2 t)\cos2\theta+(\Delta_f E_2\cos\omega_1 t+\Delta_q E_1\sin\omega_2 t)\sin2\theta;$
$z_2=(\Delta_f E_1\sin\omega_1 t-\Delta_q E_2\cos\omega_2 t)\cos2\theta+(\Delta_f E_2\sin\omega_1 t+\Delta_q E_1\cos\omega_2 t)\sin2\theta;$
$z_3=(\Delta_f E_3\cos\omega_1 t-\Delta_q E_4\sin\omega_2 t)\sin2\theta-(\Delta_f E_4\cos\omega_1 t+\Delta_q E_3\sin\omega_2 t)\cos2\theta;$
$z_4=(\Delta_f E_3\sin\omega_1 t-\Delta_q E_4\cos\omega_2 t)\sin2\theta-(\Delta_f E_4\sin\omega_1 t+\Delta_q E_3\cos\omega_2 t)\cos2\theta;$

As for the real resonator $\frac{\omega_q - \omega_f}{\omega} << 1$, z_i are slowly changed periodic time functions.

Phase shift between u_s and u_c signals is obtained from (8) as an angle between corresponding vectors in cosωt, sinωt planes (Figure 3).

$$\Psi = arctg\frac{2\Delta_f\Delta_q B_6\cos(\omega_f - \omega_q)t}{(\Delta_f^2 + \Delta_q^2)B_7 + [(\Delta_f^2 - \Delta_q^2)B_8 - 2B_9\Delta_f\Delta_q\sin(\omega_f - \omega_q)t]\sin4\theta - \overline{-[(\Delta_f^2 - \Delta_q^2)B_9 + 2B_8\Delta_f\Delta_q\sin(\omega_f - \omega_q)]\cos4\theta}}, \qquad (9)$$

where: $B_6=E_1E_3+E_2E_4$;
$B_7=E_2E_3-E_1E_4$;
$B_8=E_1E_3-E_2E_4$;
$B_9=E_1E_4+E_2E_3$;

It follows from (9) that ψ is a nonshifted estimation of Δ_q for the considered asymmetry of pick off electrodes and therefore is suitable to be used in an appropriate control loop. Using ψ as a control signal of the electronic balancing loop (quadrature wave suppression), equation (9) can be simplified subject to $(\omega_f-\omega_q)\to0$:

$$\Psi = arctg\frac{2\Delta_f\Delta_q B_6}{(\Delta_f^2 + \Delta_q^2)B_7 + (\Delta_f^2 - \Delta_q^2)B_{10}\sin(4\theta - \xi_2)}, \qquad (10)$$

where: $B_{10} = \sqrt{(E_1^2 + E_2^2)(E_3^2 + E_4^2)}$;

$$\xi_2 = arctg\frac{B_9}{B_8}.$$

It should be mentioned, that in case of pick off electrode system symmetry i.e. $E_2=E_4=0$ and $E_1=E_3$, the ψ function described by the relation:

$$\Psi = arctg\frac{2\Delta_f\Delta_q B_6}{(\Delta_f^2 - \Delta_q^2)\sin4\theta}.$$

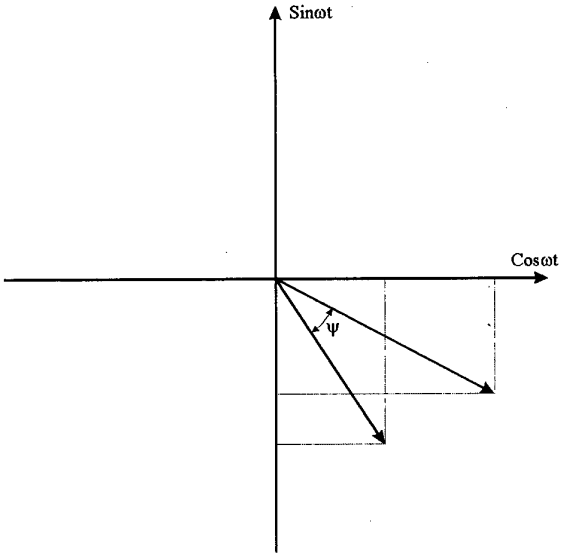


Figure 3.

has the first order break $\psi = \pm\pi/2$ either in case of $\Delta_f = \Delta_q$ or under $\Delta_f \neq \Delta_q$ in case of $\theta=(n-1)\pi/4$. In real conditions full symmetry of pick off electrodes is slightly probable, but it is necessary to take into consideration the mentioned feature of the ψ function when the control loop is being formed.

Let us consider free vibrations of the resonator described by equations (8). It is not difficult to show that in the u_S, u_C plane (Figure 4) they describe ellipse [7]:

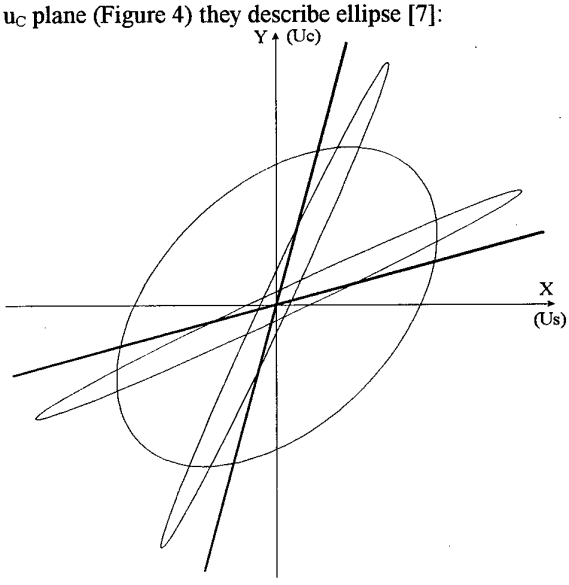


Figure 4.

$$u_s^2(z_3^2+z_4^2)+ u_c^2(z_1^2+z_2^2)-2u_su_c(z_1z_3+z_2z_4)=(z_2z_3-z_1z_4)^2.$$

Substituting corresponding symbols, we obtain:

$$u_s^2\{(\Delta_f^2+\Delta_q^2)B_{11}+[(\Delta_f^2-\Delta_q^2)B_{12}+4\Delta_f\Delta_qB_{15}\sin(\omega_f-\omega_q)t]\cos4\theta-2[(\Delta_f^2-\Delta_q^2)B_{15}-\Delta_f\Delta_qB_{12}\sin(\omega_f-\omega_q)t]\sin4\theta\}+u_c^2\{(\Delta_f^2+\Delta_q^2)B_{13}+[(\Delta_f^2-\Delta_q^2)B_{14}-4\Delta_f\Delta_qB_{16}\sin(\omega_f-\omega_q)t]\cos4\theta+2[(\Delta_f^2-\Delta_q^2)B_{16}+\Delta_f\Delta_qB_{14}\sin(\omega_f-\omega_q)t]\sin4\theta\}-2u_su_c\{(\Delta_f^2+\Delta_q^2)B_{17}+[(\Delta_f^2-\Delta_q^2)B_8-2\Delta_f\Delta_qB_9\sin(\omega_f-\omega_q)t]\sin4\theta-[(\Delta_f^2-\Delta_q^2)B_9+2\Delta_f\Delta_qB_8\sin(\omega_f-\omega_q)t]\cos4\theta\}=2U^2\Delta_f^2\Delta_q^2B_6^2\cos^2(\omega_f-\omega_q)t. \quad (11)$$

where: $B_{11}=E_3^2+E_4^2$;
$B_{12}=E_4^2-E_3^2$;
$B_{13}=E_1^2+E_2^2$;
$B_{14}=E_1^2-E_2^2$;
$B_{15}=E_3E_4$;
$B_{16}=E_1E_2$;

It follows from (11) that coefficients which define ellipse parameters are periodic time functions. Hence, we can watch some ellipse evolution's in the u_S, u_C plane (for example, the screen of a double-beam oscillograph). For the purpose of simplification, the further consideration will be provided for symmetric electrode system. Taking into account, that fundamental axes of the ellipsoid are orthogonal we can determine them by evaluating the expression for one of them. In this case the desired relation is as follows:

$$\varphi = 0{,}5arctg\frac{(\Delta_f^2-\Delta_q^2)\sin4\theta-2\Delta_f\Delta_q\sin(\omega_f-\omega_q)t\cos4\theta}{(\Delta_f^2-\Delta_q^2)\cos4\theta}, \quad (12)$$

Thus, fundamental axes of the ellipse will make angular movements. In this case, at the moments of $t=(n-1)\pi/(\omega_f-\omega_q)$ φ will exactly correspond to the angular position of the fundamental wave (in the u_S, u_C plane $\varphi=2\theta$).

It follows from (11) that at the moments of $t=(2n-1)\pi/2$, the ellipse degenerates into direct lines:

$$u_c = \frac{\sin(4\theta - arctg\frac{\Delta_f\Delta_q}{\Delta_f^2-\Delta_q^2})}{1+\sin(4\theta+arctg\frac{\Delta_f^2-\Delta_q^2}{\Delta_f\Delta_q})}u_s; \quad (13)$$

$$u_c = \frac{\sin(4\theta + arctg\frac{\Delta_f\Delta_q}{\Delta_f^2-\Delta_q^2})}{1+\sin(4\theta-arctg\frac{\Delta_f^2-\Delta_q^2}{\Delta_f\Delta_q})}u_s.$$

Thus, ellipse evolution consists in its angular vibrations, accompanied by its ellipticity change and degeneration at the boundaries into direct lines (13). Ellipse vibration period precisely equals $\omega_f-\omega_q$, and an angular amplitude of its vibrations is determined by the amplitude

relationship Δ_f and Δ_q. If $\Delta_f=\Delta_q$ at the moments of $t=(n-1)\pi/(\omega_f-\omega_q)$ the ellipse is changed into a circumference. In this case direct lines (13) exactly coincide with angular position of fundamental and quadrature waves.

It follows from the above, that availability of two different resonance frequencies under free vibrations of the resonator do not cause wave drift. This fully agrees with conclusions of the publication [2]. In addition, the obtained result is an information basis when the resonator is balanced for elimination of the 4-th harmonic of the resonator elasticity and mass inhomogeneity. We should mention here that it is not difficult to provide practically ideal symmetry of the pick off electrode system in the production equipment.

In case of generating an pick off signal by applying alternating voltage of angular frequency Ω_1 to the sin channel electrodes and frequency Ω_2 to the cos channel electrodes [8], the signals taken from the hemisphere to the purely resistive load for each frequency can be represented in the form of:

$$u_s=U_1\Omega_1[E_5+\Delta_f(E_1\cos2\theta+E_2\sin2\theta)\cos\omega_f t-\Delta_q(E_2\sin2\theta-E_1\cos2\theta)\sin\omega_q t]\sin\Omega_1 t;$$
$$u_c=U_2\Omega_2[E_6+\Delta_f(E_3\sin2\theta-E_4\cos2\theta)\cos\omega_f t+\Delta_q(E_3\cos2\theta+E_4\sin2\theta)\sin\omega_q t]\sin\Omega_2 t. \quad (14)$$

where: $E_5=K_1+K_5-K_3-K_7$;
$E_6= K_2+K_6-K_4-K_8$.

Technically It is not difficult to provide $U_1\Omega_1=U_2\Omega_2$ condition. Having provided an ideal synchronous detection of signals (14) in accordance with corresponding carrier frequencies we will obtain signals different from (6) by availability of permanent members E_5 and E_6 which express modulation deviation from 100%. Upon liquidation of these members, we obtain the above considered expression (6). Thus, ignoring nonideality of the above mentioned operations we can make a conclusion that errors which occur from asymmetry of the pick off electrode system do not depend upon pick off signal generation procedure.

Force Electrode System

The force electrode system includes a ring electrode and 16 discrete electrodes. The ring electrode is used for parameter stabilisation of the resonator vibration amplitude. The discrete electrodes opposite the ring one are combined into pairs and provide resonator excitation on its resonance frequency, electronic balancing of the resonator, and wave position control in case of using the HRG as an ARS. As the force is proportional to the square of potential difference between the resonator surface and an electrode, control with the help of alternating potentials will generate a certain permanent force.

Practically, the force electrode system features asymmetry similar to that of the pick off electrode system. An exception is the ring electrode the asymmetry of which is generated by gap irregularity. In the result, permanent force acting on the resonator shell from the side of vibration amplitude preserve loop will be the function of θ. As this force changes shell elasticity, the resonator elasticity-mass parameters will also become a certain function of θ, which can be expanded into a Fourier series according to θ. It can be easily shown that the most characteristic gap irregularities of eccentricity and ovality type generate just the first four expansion members or first four inhomogeneity harmonics. Thus, even in case of absolutely symmetrical resonator the functioning of the vibration amplitude preservation loop under asymmetric gap will cause two resonance frequencies and occurrence of resonator mass center vibrations, which is more unpleasant [6]. The latter will inevitably cause losses of the resonator vibrating energy in the HRG structural elements and consequently deterioration of the resonator Q and its asymmetry. It is well known [2] that the resonator δQ is the reason of HRG drift. Thus, ring electrode asymmetry is the source of HRG additional drift. If to take into account the error of generating information signal of the vibration amplitude preservation loop (4), the drift caused by this loop will be the function of 4θ.

An additional drift is proportional to the square of potential difference (necessary to preserve vibrations), vibration amplitude, relative asymmetry of the gap and in inverse proportion to Q. Hence, Q increase and vibration amplitude decrease will cause decrease of the additional drift. Decrease of relative asymmetry of the gap under other similar conditions is possible only by increasing its nominal value which will require increase of potential difference. In addition, gap increase and vibration amplitude decrease requires increase of hemisphere potential to keep the required signal-to-noise ratio in the pick off channel. Thus, there is a task of rational selection of the said parameters.

Figure 5 and Figure 6 present drift velocity dependencies of one of the HRG mock-ups with a 20-mm diameter resonator. These dependencies were obtained in different conditions: the results shown on Figure 6 were obtained for a reduced vibration amplitude and an increased potential on the hemisphere. As vividly seen from the comparison of the presented curves, they comply with the said above.

Electronic balancing is made by providing appropriate constant potential difference between the hemisphere and two pairs of orthogonal discrete electrodes. In case

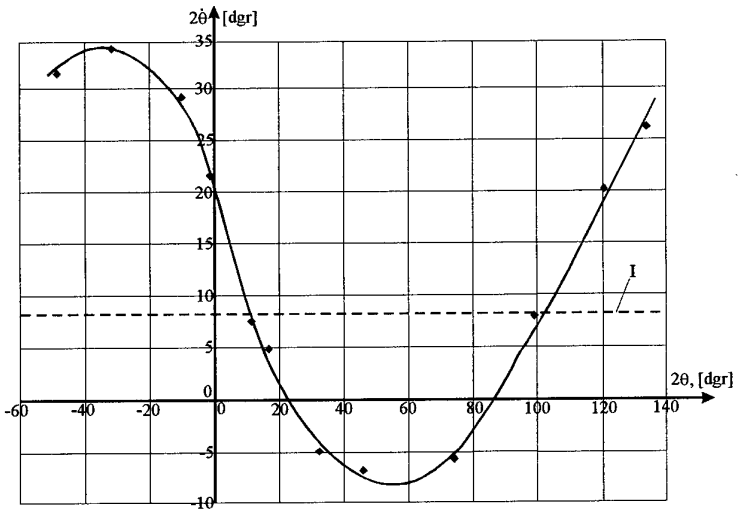


***Figure 5**. Case-Oriented Drift as a Function of Pattern Angle.*
I - vertical component of Earth rotation.

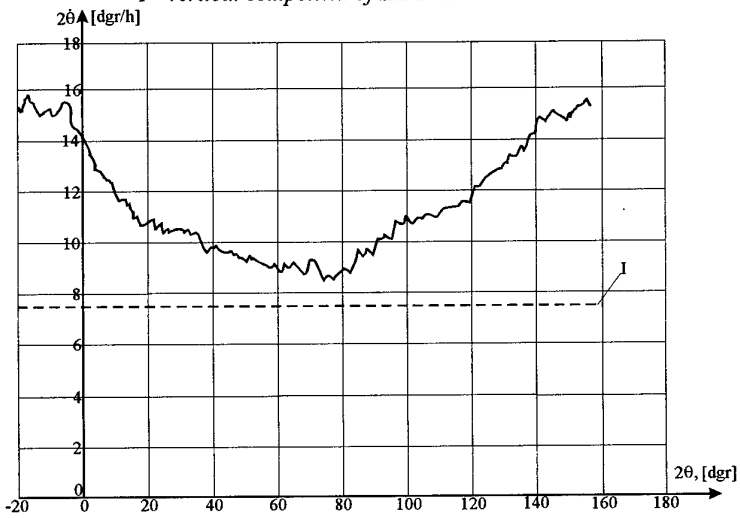


***Figure 6.** Case-Oriented Drift as a Function of Pattern Angle.*
I - vertical component of Earth rotation.

of full symmetry, these potentials create the 4-th harmonic of resonator elasticity-mass parameter inhomogeneity which compensates for its initial inhomogeneity which becomes apparent in availability of two resonance frequencies. Subject to discrete electrode system asymmetry the mentioned potentials will also create lower harmonics of inhomogeneity, i.e. we shall obtain the result described above.

Equation (9) obtained before, which is an information signal of electronic balancing loop (quadrature wave suppression) does not take into account phase errors of electronic circuits which generate ψ assessment. Availability of these errors will cause the shift of the assessment. In the result, in the resonator there will be kept two resonance frequencies, the difference of which is determined by the value and the sign of the ψ assessment shift. Availability of vibration amplitude stabilisation loop under these conditions will lead to the drift independent of θ [2]. By this fact we can explain availability of constant shift of the curves on Figure 5 and 6 relative to the Earth angular velocity. As phase errors of the electronic circuits can not be considered as stable unlike electrode system asymmetry, full or partial calibration of these circuits is expected to become apparent in the drift characteristic. Unfortunately, in the course of experiments it was possible to calibrate only electronic circuits of input amplifiers of pick off signals u_s and u_c. Figure 7 shows drift plots of the said above mock-up. The plots are drawn by wave return into the initial point by the mock-up body turn. No calibrations were made. Figure 8 shows similar plots, but each time calibration of input amplifiers in initial state was made. The results obtained are not in conflict with the said above.

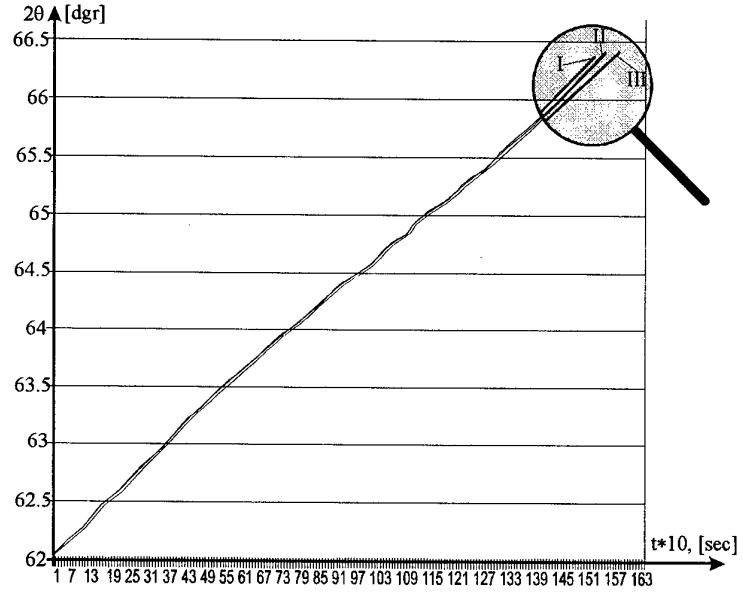


Figure 7. *Three 0.5h drift lines.*
The calibration was not performed.
Trend 2θ̇, [dgr/h] : I - 10.13; II - 10.09; III - 10.04

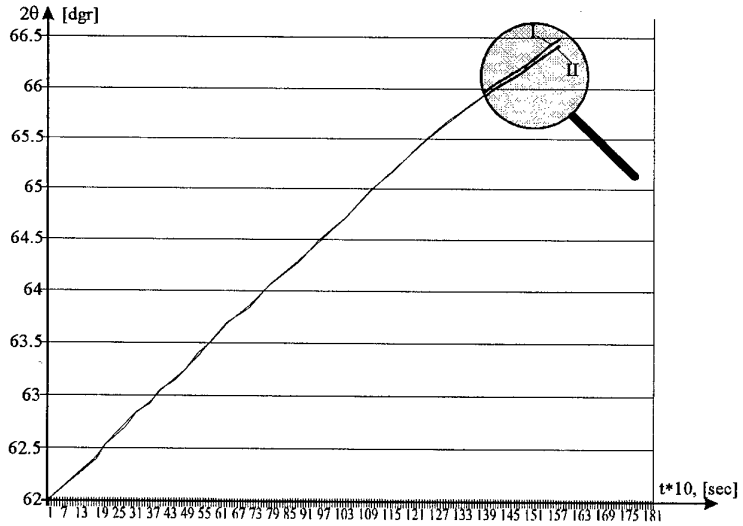


Figure 8. *Two 0.5h drift lines.*
Two calibrations was performed.
Trend 2θ̇, [dgr/h] : I - 10.13; II - 10.11

Summary

Based on the development experience of HRG mock-ups of two standard sizes, some essential problems of HRG technology are raised and methods of their solving are proposed. Errors in pick off signal generation and assessments of vibration parameters used in control loops have been analysed. It is shown that free vibrations of the resonator having two resonance frequencies do not cause drift. A model of the drift additional components is given which is defined both by pick off electrode system symmetry and force electrode system asymmetry. Experimental data comply with the statements given above.

References

1.Lynch D.D., Matthews A., Dual-Mode Hemispherical Resonator Gyro Operating Characteristics, 3-rd Saint Petersburg International Conference on Integrated Navigation Systems, 1996, part 1, pp. 37-44.

2. Журавлев В.Ф., Климов Д.М., Волновой твердотельный гироскоп, М., Наука, 1985, с. 123.

3. Zhuravlev V.F., Theoretical Foundation of Hemispherical Resonator Gyro (HRG), Izv. RAN, Mekhanika Tverdogo Tela, Allerton Press Inc., Vol. 28, N3, 1993, pp. 3-15.

4. Lynch D.D., Vibratory Gyro Analysis by the Method of Averaging, 2-nd Saint Petersburg International Conference on Gyroscopic Technology and Navigation, 1995, part II, pp. 26-34.

5. Zhuravlev V.F., Lynch D.D., Electric Model of a Hemispherical Resonator Gyro, Izv. RAN, Mechanika Tverdogo Tela, Allerton Press Inc., Vol. 30, N5, pp. 12-24.

6. Zhbanov Yu.K., Theoretical Aspects of Balancing the Hemispherical Resonator Gyro, 2-nd Saint Petersburg International Conference on Gyroscopic Technology and Navigation, 1995, part II, p. 88.

7. Shatalov M.Y., Elements on General Control of Vibratory Gyroscopes, 5-th Saint Petersburg International Conference on Integrated Navigation Systems, 1998, pp. 204-213.

8. Lynch D.D., Matthews A., Varty G.T., Transfer of Sensor Technology from Oil-Drilling to Space Application, 5-th Saint Petersburg International Conference on Integrated Navigation Systems, 1998, pp. 27-36.

TECHNOLOGICAL ASPECTS OF MANUFACTURING OF COMPOUND HEMISPHERICAL RESONATORS FOR SMALL-SIZED VIBRATORY GYROSCOPES

Yu.A. Yatsenko, S.F. Petrenko
Ukrainian Center for Optical Instrument Technology – Lileya Ltd.
8 Moscowskaya str., 252010 Kyiv, Ukraine

V.V. Vovk
Ukrainian Center for Optical Instrument Technology – Arsenal Corp.
8 Moscowskaya str., 252010 Kyiv, Ukraine

V.V. Chikovani
Ukrainian Center for Optical Instrument Technology (NTU/KPI)
8 Moscowskaya str., 252010 Kyiv, Ukraine

Summary

Technological processes for manufacturing of low-cost compound hemispherical resonators are presented in this paper. The compound resonator consists of hemispherical shell (meniscus) and stem, which are made separately with use of the well fulfilled technology of spherical and cylindrical optics. Manufacturing technology of compound resonator elements is described in the paper. Geometrical and dynamic characteristics of meniscus are given. This work was conducted within the frame of the STCU joint project with Lileya Ltd., "Arsenal" Corp. and Litton Guidance and Control Systems as collaborators.

Introduction

The development of navigation devices in many respects is defined by the development of their sensing elements, and its applicability is defined, in the main, by reliability, cost and dimensions. This is especially actual when developing low-cost and compact attitude systems for civil application. To meet these requirements highly effective technology and new design principle for sensing element are used.

Use of highly effective technology of optical production when manufacturing resonators for hemispherical resonator gyroscope (HRG) allows one to considerably lower the price for their manufacture and to keep their dynamic characteristics in the sufficiently high level.

A design features of the compound resonator

With the purpose of increasing of adaptability to manufacture and cost reduction we offer a design of the compound resonator [1-3], which consists of hemispherical shell (meniscus) and stem. The components of the resonator are made separately. Besides the design data of the resonator elements are supervised after each stage of a technological chain during manufacture, that should reduce a technological error component of resonator manufacturing. At this, the well fulfilled methods of the control are applied. Connection of the design elements is made with the help of the specially developed technological adaptations. Meniscus and stem connection is produced with the use of deep optical contact (in case of a quartz glass КУ-1) or with a glue. It is expected, that the good quality of the developed resonators will be high enough, and Q-factor and frequency mismatches will answer to middle class devices.

The basic reasons of energy dissipation in resonators are the internal losses: losses in a surface layer and losses connected to fastening of the resonator (caused unbalanced radial forces, resulting in transfer of energy in the resonator stem and further in the gyro case [3,5]). Besides in the offered design of the compound resonator there are additional losses in a transitive layer meniscus-stem.

Paper presented at the RTO SCI International Conference on "Integrated Navigation Systems", held at "Elektropribor", St. Petersburg, Russia, 24-26 May 1999, and published in RTO MP-43.

However, as our investigation has shown for 60 mm compound resonators [3], these losses are in acceptable tolerances in comparison with one-piece resonators.

Therefore, basic task when developing compound resonator design is the optimization of design data with the purpose of providing of the maximal Q-factor, achievement of the minimal technological defects of manufacturing (and connection) and providing of the minimal influence of external disturbances, transmitted through the gyro basis and stem.

The development of a compound resonator design was carried out on the basis of the chosen mathematical model, which allowed us to define the most important parameters of a design with the purpose of separation of resonant frequencies of hemisphere and stem to avoid their mutual influence, and to achieve maximum Q-factor of the working mode. The frequencies of a design elements, as a first approximation, were defined with the use of ideal hemisphere shell model [4] and cantilever model with mass on its free end (hemispherical shell with the moment of inertia M_i) by the formulas:

$$f_{(i)} = \frac{J(i)}{2\pi}\frac{hc_T}{R^2}(1+6\zeta_c\frac{R^2}{h^2}+...), \qquad (1)$$

where $J(i) = i(i^2-1)\sqrt{\frac{\xi_{a(i)}}{3\xi_m}}$, ξ_m- dimensionless coefficient of inertia, $\xi_{a(i)} = \int_{\alpha}^{\pi/2}\frac{\tan^{2i}(\frac{\theta}{2})}{\sin^3\theta}d\theta$ - dimensionless stiffness coefficient, taking into account shell (meniscus) and stem connection, $(\alpha = \frac{d}{2R})$, $\zeta_{c(i)} = \frac{\eta_{ac}}{4\pi i^2(i^2-1)\xi_{a(i)}}$ coefficient, that takes into account membrane stiffness of a shell and depending on connection with stem as well.

$$f_{(1)} = \frac{d^2c_T}{16\pi RL}\sqrt{\frac{3(1+\nu)}{hL}(1\mp 3\frac{i_0^2}{L^2})}, \qquad (2)$$

where $f_{(1)}$ is a flexural frequency of a stem, R, h - radius and thickness of a hemispherical shell, L, d- length and diameter of a stem, i_o - radius of inertia of a shell relative to a centre of gravity (± - concerns to internal or outside stem, accordingly), c_T, ν- propagation speed of flexure and Poison coefficient.

Other criterion when choosing design and technological parameters of resonators is their Q-factor, which depends both on the chosen material (internal losses ξ_e), and on quality of its surface (superficial losses ξ_n), and also on total rigidity of a design, on its balancing and a way of fastening of the resonator. Besides, since the considered design of the resonator is represented in view of interaction of two oscillatory systems: meniscus and stem, which frequencies can be considered as independent, the resulting Q-factor depends on their mutual coupling. It should be noted that coupling of frequencies of oscillatory system depends on a difference of partial frequencies and on a elastic coupling coefficient, which, in our case, carries complex character because of presence of losses in a transitive layer.

For a qualitative estimation of influence of internal and superficial losses on resonator Q-factor for the working mode, the following expression is applicable:

$$Q_i = \frac{1}{\omega_i(\xi_e+\xi_n)}, \qquad (3)$$

where $\xi_e = \frac{4}{3}\frac{\delta}{2\pi^2 f}$, δ - logarithmic damping factor, $\xi_n = \frac{T\alpha^2\rho c_T a^2}{\kappa}$, κ- coefficient of thermal conductivity, T- temperature, α - coefficient of linear expansion, a - size of superficial heterogeneity.

Investigations of various materials carried out by the different authors [2,5,9] show that the most suitable one is the fused quartz combining good acoustic parameters, high durability and thermal

stability of parameters. The logarithmic damping factor for working frequencies is $(1.3 \ldots 3.4) \times 10^{-7}$, that corresponds to potential good quality $Q = (1 \ldots 2.4) \times 10^{7}$ for glass КУ-1 (Russian classification) and $Q = (8 \ldots 16) \times 10^{6}$ for glasses КВ and КРЛ-1 (Russian classification) [9].

The superficial condition of resonators is defined by processing technology of a glass and, in particular by its finishing stage - etching in chemical solutions with the purpose of removal of the broken superficial layer. The absorption of elastic waves in a superficial layer is connected with the thermal conductivity losses. This mechanism of losses is considered in [7], where it is noted, that this is applicable, when the relaxation time is much less, than the period of oscillation, i.e. $\omega \leq \chi/a^2$, where $\chi = \kappa/c$ is a factor of temperature conductivity (for a quartz c=0.0180 kg/(m·s)), that is obviously valid for kilohertz of frequencies.

The calculations, which have been carried out with the use of the formula (3), show (see fig. 1), that if the sizes of superficial heterogeneity $a \leq 2$ microns, superficial losses prevail. This fact limits Q-factor at a level of $(1 \ldots 10) \times 10^{6}$.

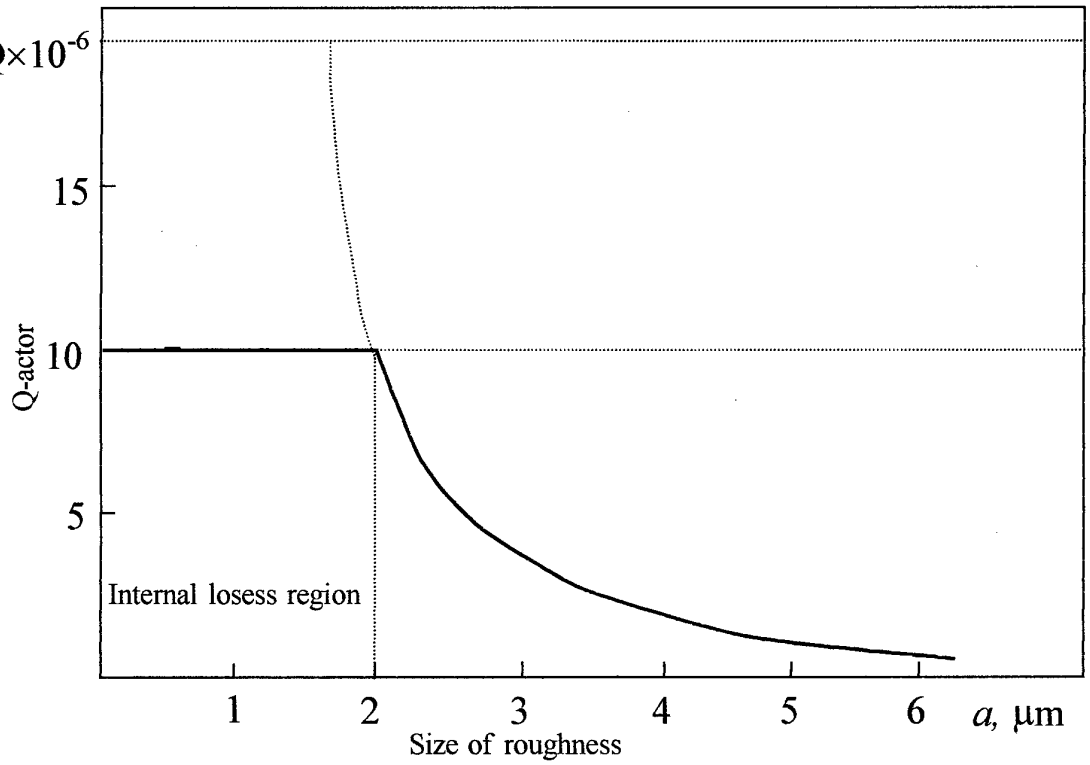


Fig. 1. Dependency of superficial losses on heterogeneity size.

Manufacturing technology of compound resonator elements.

The meniscus manufacturing from a fused quartz was carried out with the use of method of deep grinding and polishing, that is typical technological process. After preliminary forming of meniscus surfaces from КУ-1 they were etching in a mix of hydrofluoric and orthophosphoric acids with the purpose of stabilization of a cracked layer (which can make 150 ... 250 microns). The depth of etching layer was lead up to 50 microns. At grinding by electrolytically produced corundum, removal of a material to 250 microns were lead up. This way the cracked layer was removed. However, at grinding a surface by abrasive materials with grains of the various sizes there are imperfections of the certain depth (at operational development of a surface with the help of polishing paste and water, the size of

imperfections in a superficial layer can reach some microns). The finishing etching of meniscus was made in assembly of the resonator.

It should be noted, that when etching a meniscus, radiuses of a curvature actually vary and the surfaces "go off on hillock". Therefore sphericity control was made by taking into account of this "go off on hillock", and respectively were calculated allowances for "transitions". In order to glass removal on "transitions" was identical over all surface of meniscus, the control by a set of test glasses was carried out. So, when polishing by electrolytically produced corundum M10, the radiuses of test glasses differed from final by 30 microns. Then, when polishing by electrolytically produced corundum M10 on moistened polishing tools, the control test glasses differed from final by 20 microns. And only then at final polishing by polishing paste, the control was made on nominal test glasses. Thus the deviation from sphericity on meniscus did not exceed two interference bands $N \leq 2$, and the local deviations did not exceed $\Delta N < 0.2$ interference bands. In a place of connection of meniscus with a stem the total deviation from sphericity of meniscus and stem surfaces did not exceed 0.2 interference bands for providing a deep optical contact [6].

The heterogeneity on the polished surface were measured with the help of heterogeneity indicator TALYSURF (maximum resolution 0.02мкм). The profilogramm decoding (see fig. 2) shows, that height of heterogeneity of a structure estimated by ten points (represents average distance between 5 maximum points of ledges by 5 lowest points of hollows within the limits of a base line) is $R_z = 0.02$ microns. It is necessary to note, that for a quantitative estimation of superficial losses as the average size of heterogeneity it is necessary to take into account average meaning of a heterogeneity step S_m, which is determined within the limits of the base length of profilogramm (base length for 14 classes of surface finish is l=0.08 mm), which is calculated with the use of relationship:

$$S_m = \frac{l_i}{10 K_u N}$$

Here N – the number of maximums over base length, l_i and K_u – the length of processed part of the profilogramm and factor of increase of heterogeneity indicator. As one can see from decoding of profilogramm the average step of imperfection (fig. 2) is ~ 2 ... 3,4 microns, that corresponds to Q-factor of meniscus at a level of $Q = (2 \ldots 10) \times 10^6$.

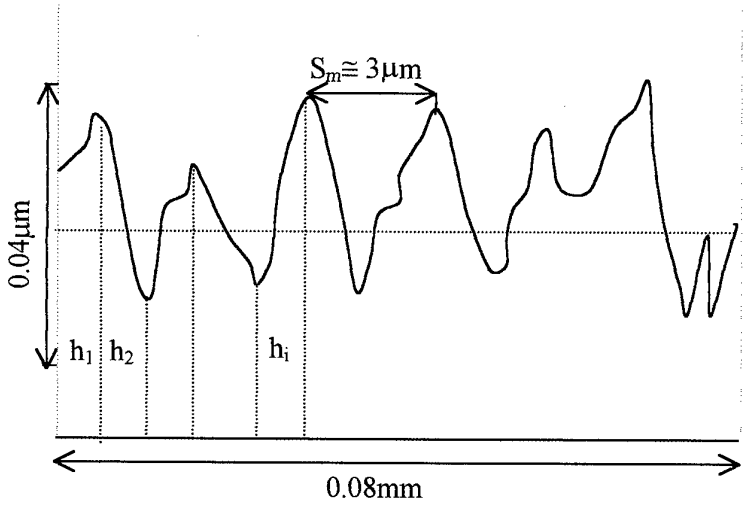


Fig.2 Meniscus surface imperfection profilogramm

The manufacturing of stems consists of a preliminary forming, grinding and polishing. For grinding and polishing, with the purpose of increase of productivity, separator method of processing was applied. Grinding was made by metallic separators, and polishing by organic one. Thus, non-

roundness and conicity did not exceed 0.1 microns. For providing beating ≤ 0.5 microns of a spherical surface of a stem relative to a base lateral surface, processing was carried out with the use of the adaptation, which had necessary radius and diameter 4 times more. The control of beating was carried out with the help of telescope caliper under support of a lateral surface of the certificated 90° prism.

The other major reason of errors in vibratory gyro are the technological defects of the resonator, such as difference in thickness of meniscus, eccentricity, tilt of a meniscus axis relative to an axis of a stem. These defects, except for difference in thickness, are connected to accuracy of assembly of resonators and are defined by an error of technological adaptations used for assembly and an error of methods of the control of assembly. In this connection before final assembly of the resonator the control of geometrical and dynamic parameters of elements of the resonator was carried out. The control of meniscus dynamic parameters was made by the same technique of resonator control. Doing this, meniscus was fixed in special fluoroplastic holders fixing its position relative to the case. The dynamic parameters of the stem were estimated by results of its piezoelectric excitation.

In the table 1 the basic parameters of the manufactured elements of the resonator are given.

Table 1

Parameters	Value
Meniscus diameter, mm	30
Difference in thickness, μm	0.5
Meniscus non-sphericity, interference band	2
Frequency of mode n=2, Hz	3600…5600
Q – factor	$(2…3)\times10^5$
Frequency mismatch, Hz	0.01…0.1
Frequency of mode n=3 Hz	14000…14200
Q – factor	$(3…5)\times10^6$
Frequency mismatch, Hz	0.01…0.1
Frequency of mode n=4 Hz	24900…25200
Q – factor	$(2…3)\times10^6$
Frequency mismatch, Hz	0.1…0.4
Flexural frequency of the stem, Hz	3750

As follows from the analysis of an imperfect symmetric meniscus dynamic equation [8,9] for the mode n=2, first three harmonics of defects of the resonator result in oscillation of the centre of gravity. Therefore the part of wave energy is transferred to the stem and dissipates. And the oscillation of the centre of gravity of imperfect hemispherical meniscus has the most essentially influence on the n=2 mode. This was also confirmed by experimental results: the Q-factor of n=3 mode has appeared much more than n=2 mode because of smaller influence of the centre of gravity of the unbalanced resonator (see table 1).

CONCLUSION

The manufactured elements of the compound resonator have potentially high Q-factor, that exceeds the one obtained in experiment. In the experiment much losses were gained in the device of meniscus fastening.

The obtained parameters of meniscus frequency mismatches, despite of their wide scatter from meniscus to meniscus, caused, in main, by method of fastening, confirm adaptability to manufacture of resonator elements based on technology of spherical optics and demonstrate their high potential accuracy.

The preliminary investigation of dynamic parameters of meniscus show, that Q-factor for n=3 mode is less sensitive to the methods of its fastening. This fact requires additional detail investigation.

References:

1. V.V. Vovk, V.P. Maslov, E.A. Portnova "Resonator elements connection technique"./ Patent USSR No.1577192, G01 C 25/00, 12, April 1991. (in Russian).
2. Yu.A. Yatsenko, S.F. Petrenko, V.V. Chikovani, V.V. Vovk "Comparative Investigation of Design Particularities of Resonators of Vibratory Gyros"./ IV S.Petersburg Conference on Integrated Navigation Systems, editor V.G. Peshekhonov, 26-28, May 1997, pp. 276-281.
3. Yu.A. Yatsenko, S.F. Petrenko, V.V. Chikovani, V.V. Vovk "Influence of a Transient Layer of a Combined Hemispherical Resonator of a Solid-State Wave Gyro Upon Its Dinamical Characterestics"./ V S.Petersburg Conference on Integrated Navigation Systems, editor V.G. Peshekhonov, 25-27, May 1998, pp. 214-217.
4. G. V. Strett (Lord Rayleigh) "Sound Theory".- M.: Izdat.Teoreticheskoy Literatury, v.1, 1955 (in Russian).
5. G.S. Pisarenko, A.P. Yakovlev, V.V. Matveev " Resistance of Materials Handbook".- Kyiv: Naukova Dumka, 1988 (in Russian).
6. R.K. Fatykhova, "Control of Optical- Mechanical Devices".- M.: Mashinostroenie, 1988 (in Russian).
7. A.G. Smagin, M.I. Yaroslavsky "Piezo-Electricity of Quartz and Quartz Resonators".- M.: Energia, 1970 (in Russian).
8. N. E. Egarmin, V.E. Yurin. "Introduction to Theory of Vibratory Gyroscopes". Moscow, Binom Co.,1993 (in Russian).
9. B.P. Bodunov, V.M. Lopatin, B.S. Lunin, D.D. Lynch, A.R. Voros "Low-Cost Hemispherical Resonator for Miniature HRG-based Navigation Systems"//IV S.Petersburg Conference on Integrated Navigation Systems, editor V.G. Peshekhonov, 26-28, May 1997, pp. 19-25.

Signal Processing Using the Increments of Signal Multiple Integrals: From Strapdown INS to Other Real–Time Systems

Yury A.Litmanovich, Vladimir M.Lesyuchevsky, Valery Z.Gusinsky
Central Scientific and Research Institute "Elektropribor",
Malaya Posadskay 30, St. Petersburg, 197046, Russia

Summary

A new approach to signal processing in strapdown INS is presented and examined with a view to apply it to the other real–time systems. The solutions for three problems of the strapdown INS software, which are typical for other systems are expressed via the increments of the signal multiple integrals over the iteration interval. The possibility and utility of the signal multiple integrals generation while the signal pre-processing is discussed.

1. Introduction

One of the issues of the inertial navigation system (INS) design that is typical for any real–time system is in rational separation of the net computation process into primary and final processing of the sensor measurement signals. The former is performed in the sensor digital electronics and the latter - in the system central computer. Primary processing is conventionally aimed to compress the raw sensor data and to put them in a form suitable for the final processing and is mainly responsible for the measurement noise smoothing to decrease the sensor sampling rate. The incremental sensors, which generate the increments of signal one-time integral over the sampling interval are one such example. The incremental gyros and accelerometers are in wide use in the modern-day INS.

The main goal of this paper is to review and present in a general form a new approach to signal processing in strapdown INS, which was developed and implemented in Elektropribor, with a view to transfer it to other real-time systems (RTS). There are at least three problems in the strapdown INS software, which could be considered as a special case of the general problems that can be met in RTS. They are: (1) multiple integration of a signal, (2) multiple integration of a product of two signals, and (3) signal smoothing by polynomial fitting, .

In the paper the solutions for all three problems are presented in such a mathematical form that permits the using of the increments of signal multiple integrals as input data for the final processing. In this case, the signal primary processing is reduced to the generation of the increments of signal multiple integrals over the sampling interval with the sampling rate specified by the measurement signal's dynamics. For each of the problems under consideration we start with the formulation of the problem as a general one, then derive the solution via the signal multiple integrals and, finally, specify it as applied to the strapdown INS and evaluate the advantages of the proposed solution over the conventional ones. In conclusion the aspects of the signal multiple integral generation in strapdown INS, in particular, and in RTS at all are discussed.

2. Multiple Integration of a Signal

Consider a problem of discrete two-time integration of a signal over time taken as an example of a more general problem of signal multiple integration. As applied to the INS the problem under consideration forms the basis for the velocity and position computation from the specific force components in the navigation frame that is common for the INS of all types. Let a is a measurement signal to be doubly integrated. A two-level scheme for computation of signal one-time (V) and two-time (S) integrals is presented by the following discrete algorithm which is readily apparent from a continuous one:

$$V(k+1) = V(k) + \delta V(k, k+1) \qquad (1)$$

$$S(k+1) = S(k) + V(k) \cdot \Delta T + \delta S(k, k+1) \qquad (2)$$

where $t_k = k \cdot \Delta T (k = 0, 1, 2...)$; ΔT is the computation iteration interval; k is the iteration index and δV, δS are the increments of a one-time and two-time integrals over the iteration interval, respectively, calculated as:

$$\begin{aligned} \delta V(k, k+1) &= \int_{t_k}^{t_{k+1}} a(t)dt \\ \delta S(k, k+1) &= \int_{t_k}^{t_{k+1}} \int_{t_k}^{t} a(\tau)d\tau dt \end{aligned} \qquad (3)$$

According to this algorithm the process of both one-time and two-time signal integration is separated into two stages: a primary continuous (quasicontinuous)

Paper presented at the RTO SCI International Conference on "Integrated Navigation Systems", held at "Elektropribor", St. Petersburg, Russia, 24-26 May 1999, and published in RTO MP-43.

integration over the iteration interval, and a secondary integration with the iteration interval over the system operation time. A possibility of mechanizing the algorithm (1),(2) and its accuracy completely depend on the possibility and the accuracy of generating the quantities $\delta V, \delta S$. A key feature of the derived discrete algorithm lies in the fact that it is free from the numerical integration method errors whatever iteration rate is taken.

As to the INS, Eq.(1) if written in a vector form defines the conventional discrete velocity generation algorithm. It goes from the local-level INS in which the navigation frame specific force components are generated directly by the accelerometers. For this reason Eq.(1) can be readily realized with the incremental accelerometers in hand. As regards the position computation, the algorithm (2) wasn't used inspite of its clarity. It was conventional to calculate the navigation frame position components by numerical integration of the obtained velocity discrete values, say, as follows:

$$S(k+1) = S(k) + V(k) \cdot \Delta T \qquad (4)$$

Eq.(4) represents in a scalar form a rectangular integration scheme, though a trapezoid or higher-order integration scheme can be used. The important point is that in this case the position components will be determined with iteration interval dependent errors whatever numerical integration method is used. So, a restriction is imposed to the iteration rate - it must be higher than the frequency of the most high-frequency significant component in the velocity spectrum. The navigation iteration rate is typically chosen based on the vehicle's dynamics. It turns out that in this case a high INS rapid alignment accuracy can't be provided under the linear vibration with the traditional position calculation algorithm of Eq.(4) type [3]. Consequently, a true discrete velocity integration algorithm (2) is to be used. To implement it in the local-level INS the "second-order" incremental accelerometers should be used that generate the increments of both one-time and two-time specific force integrals over the iteration interval. Going to the strapdown INS, it results in extending the formulation of the specific force transformation problem as opposed to the conventional one - in addition to the calculation of the integral

$$\delta \mathbf{V}(k, k+1) = \int_{t_k}^{t_{k+1}} C(t) \cdot \mathbf{a}_b(t) dt \qquad (5)$$

the following integral is to be calculated:

$$\delta \mathbf{S}(k, k+1) = \int_{t_k}^{t_{k+1}} \int_{t_k}^{t} C(\tau) \cdot \mathbf{a}_b(\tau) d\tau dt \qquad (6)$$

where $\mathbf{a}_b$ is the specific force vector, expressed in the body-fixed frame coordinates; C is the transformation matrix from the body axes to the navigation axes; $\delta \mathbf{V}, \delta \mathbf{S}$ are the vectors of the velocity and position increments over the iteration interval expressed in the navigation frame coordinates.

The results presented above were obtained in "Elektropribor" in mid-80s and widely published in 1994. Note that it was in the most recent paper by Paul Savage of 1998 [2] when he came to the same conclusion.

3. Multiple Integration of a Product of Two Signal

Now let us consider a problem of discrete one–time and two–time integration over time of a product of two signals. Clearly, this problem can be reduced to the previous one but with the quantities $\delta V, \delta S$ expressed as defined by Eqs.(5),(6). Eventually the problem is to calculate these definite integrals over the iteration interval. As it has already been mentioned, the problem of specific force transformation in strapdown INS is just of such type. To use the latter as an illustrative example we'll derive the solution in a matrix-vector form (as it is formulated in this case).

The numerical algorithms for calculation the integrals under cinsideration are conventionally derived using the polynomial expansion for both cofactors (the acceleration and the attitude matrix in strapdown INS) with respect to some reference point (the start or center) at the iteration interval. In doing so the signal derivatives at the reference point in the Tailor series representation are calculated via the current and past signal discrete values (or current and past data samples for the incremental sensors) using one numerical method or other.

It turns out that the general solution for the problem under consideration can be easily derived using the polynomial expansion for anyone of the cofactors alone. In fact, let us represent the matrix C as a Taylor series expansion over the iteration interval ΔT with the start point of the iteration interval taken as the reference one, namely:

$$C(t_k + \tau) = C(k) + \dot{C}(k) \cdot \tau + \ddot{C}(k) \cdot \frac{\tau^2}{2} + \cdots \qquad (7)$$

where $0 < \tau \leq \Delta T$; $C(k), \dot{C}(k), \ddot{C}(k) \cdots$ are the values of matrix C, its first, second etc. derivatives calculated at the time point t_k.

Substituting this relation into the formulas (5), (6) and integrating by parts, it results (see Appendix):

$$\delta \mathbf{V}(k,k+1) = C(k+1)\int_{t_k}^{t_{k+1}} \mathbf{a}_b(t)dt$$
$$-\dot{C}(k+1)\int_{t_k}^{t_{k+1}}\int_{t_k}^{t} \mathbf{a}_b(\tau)d\tau dt + \cdots$$
$$= \sum_{r=1}^{\infty}(-1)^{r-1}\cdot C^{(r-1)}(k+1)\cdot \boldsymbol{\Delta}\bar{\mathbf{V}}^r(k,k+1)$$
$$\delta \mathbf{S}(k,k+1) = C(k+1)\int_{t_k}^{t_{k+1}}\int_{t_k}^{t} \mathbf{a}_b(\tau)d\tau dt$$
$$-2\dot{C}(k+1)\int_{t_k}^{t_{k+1}}\int_{t_k}^{\tau_1}\int_{t_k}^{\tau_2} \mathbf{a}_b(\tau_2)d\tau_2 d\tau_1 dt + \cdots$$
$$= \sum_{r=1}^{\infty} r\cdot(-1)^{r-1}\cdot C^{(r-1)}(k+1)\cdot \boldsymbol{\Delta}\mathbf{V}^{r+1}(k,k+1) \tag{8}$$

where

$$\boldsymbol{\Delta}\mathbf{V}^r = \underbrace{\int_{t_k}^{t_{k+1}}\cdots\int_{t_k}^{t_{r-1}}}_{r} \mathbf{a}_b(\tau_r)d\tau_r\cdots d\tau_1 \tag{9}$$

Setting truncation in the general quadrature formulas (8) and choosing the numerical differentiation methods for the matrix C derivatives calculation the algorithms of different accuracy can be derived. So, the derived general formulas specify a class of the algorithms for the integrals of type (5),(6) calculation with the increments of one signal multiple integrals as the input parameters. These algorithms possess a new fundamental property as opposed to the conventional ones - they are invariant to the frequency content of the signal that is multiply integrated. As a result the computation iteration rate is specified by the second cofactor dynamics only.

As applied to the strapdown INS this approach resulted in the development of a new family of transformation algorithms with specific force multiple integrals as input signals which were used in "Elektropribor" since mid-80s. It was shown [3] that under the sculling (remind that the sculling motion refers to a combined oscillatory angular rotation and linear acceleration when a non-zero average acceleration error is rectified) the proposed algorithms are of the same order of accuracy as the corresponding conventional ones. Fig.1 shows the sculling response (the relative sculling error as a function of the relative frequency) of three simplest transformation algorithms of a new class obtained both analytically and by the simulation. The algorithms under consideration were derived when one, two and three terms have been held in the quadrature formulas (8), i.e. for three values of parameter $R: 1, 2, 3$. These results well correlate with those presented in the paper [5] for the

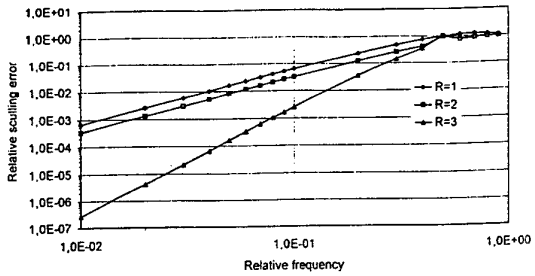


Figure 1: Algorithm sculling response (log-log plot)

algorithms, which use the conventional accelerometer outputs.

The advantage of the new algorithms was manifested in the case when the so–called "pseudosculling" errors can arise with the high-frequency components in the accelerometer outputs due to the instrument noise and/or the linear vibration. Fig.2 shows the rel-

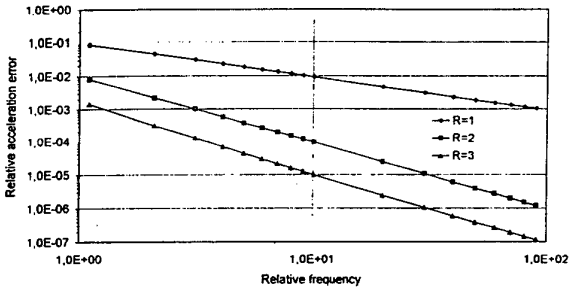


Figure 2: Algorithm pseudosculling response (log-log plot)

ative acceleration errors of three new transformation algorithms as a function of the relative frequency with the high-frequency acceleration component combined with the angular motion. In the simulation the navigation iteration rate was selected so that the aliased vibration frequency was exactly equal to the angular motion frequency. It is clearly seen from the figures that with the new sculling algorithms (of the second order and higher) the strapdown INS accuracy can be provided with low navigation iteration rate specified by the vehicle's angular motion dynamics: the higher is the vibration frequency, the more effectively the pseudosculling errors are attenuated. It was shown [3] that under these conditions the conventional transformation algorithms demonstrate poor performance of the first–order algorithm.

4. Signal Smoothing by Polynomial Fitting

In a lot of real-time systems the designers are faced with the problem of signal smoothing by polynomial fitting. As an example let us consider the problem of evaluating the signal polynomial model's coefficients against a background of an additive noise. It looks quite essential to apply the least-squares method (LSM) in this case. As an example, let the signal ω is approximated by the square-law polynomial model over the time interval $(T, T+\delta T)$ as follows:

$$\omega(T+\tau) = a + 2b\tau + 6c\frac{\tau^2}{2} \tag{10}$$

where

$$\begin{aligned} a &= \omega(T) \\ 2b &= \dot{\omega}(T) \\ 6c &= \ddot{\omega}(T) \end{aligned} \tag{11}$$

Asume that the measurements of the signal are available with an interval δt over the interval $(T, T+\delta T)$. According to the LSM algorithm, the polynomial equation (10) is represented in terms of measurements Y_i and state vector X as follows:

$$Y_i = H_i \cdot X \tag{12}$$

where

$$\begin{aligned} Y_i &= \omega_i \\ H_i &= (1, i, i^2) \\ X &= (a, 2b\delta t, 3c\delta t^2)^T \end{aligned} \tag{13}$$

$()^T$ is the matrix transpose index.

The LSM solution for the vector X estimates is given by:

$$\widehat{X} = \left[\sum_{i=1}^{N} H_i^T H_i\right]^{-1} \left[\sum_{i=1}^{N} H_i^T Y_i\right] \tag{14}$$

Substituting H_i from (13), the general solution (14) can be rewritten in a scalar form:

$$\begin{pmatrix} \widehat{a} \\ 2\widehat{b}\delta t \\ 3\widehat{c}\delta t^2 \end{pmatrix} = K_\omega(3\times 3) \begin{pmatrix} S_0 \\ S_1 \\ S_2 \end{pmatrix} \tag{15}$$

where

$$\begin{aligned} S_0 &= \sum_{i=1}^{N} \omega_i \\ S_1 &= \sum_{i=1}^{N} \omega_i i \\ S_2 &= \sum_{i=1}^{N} \omega_i i^2 \end{aligned} \tag{16}$$

K_ω is the quadratic matrix with constant elements which are the known functions of the total number of measurements $N(N = \delta T/\delta t)$, i is the current index of the signal ω measurement. It is evident that the smoothing effect can be obtained if a number of measurements Y_i is greater than the number of ω model's coefficients to be determined. (In the case when they are equal the statistical formulation of the problem is reduced to the deterministic one). To smooth over the high-frequency noise components a great number of additional (relative to the deterministic formulation) measurements should be involved that leads to a proportional growth of the primary processing algorithm's complexity.

Meanwhile it turns out that the LSM algorithm can be put in a form, which permits the using of the signal multiple integrals

$$\Delta\Theta^r = \underbrace{\int_{t_k}^{t_{k+1}} \cdots \int_{t_k}^{t_{r-1}}}_{r} \omega(\tau_r) d\tau_r \cdots d\tau_1 \tag{17}$$

In order to get the desired expressions one can expand the multiple integrals $\Delta\Theta^1$, $\Delta\Theta^2$, $\Delta\Theta^3$ via the original measurements ω_i. It gives:

$$\begin{aligned} \Delta\Theta^1 &= \delta t \cdot \tilde{S}_0 \\ \Delta\Theta^2 &= \delta t^2 \cdot \tilde{S}_1 \\ 2\Delta\Theta^3 &= \delta t^3 \cdot \tilde{S}_3 + \delta t \cdot \Delta\Theta^2 \end{aligned} \tag{18}$$

where

$$\begin{aligned} \tilde{S}_0 &= \sum_{i=1}^{N} \omega_i \\ \tilde{S}_1 &= \sum_{i=1}^{N} \omega_i (N+1-i) \\ \tilde{S}_3 &= \sum_{i=1}^{N} \omega_i (N+1-i)^2 \end{aligned} \tag{19}$$

It is easy to notice that the sums $\tilde{S}_j$ can be expressed via the sums S_j (16), and hence the relationship between the sums S_0, S_1, S_2 and the multiple integrals $\Delta\Theta^1$, $\Delta\Theta^2$, $\Delta\Theta^3$ can be established. So the LSM-estimates of the coefficients a, b, c can be obtained using the signal multiple integrals.

Now let us establish the relationship between the so derived estimates of the coefficients a, b, c and those that can be obtained by solving an appropriate set of three linear inhomogeneous algebraic equations. In fact, based on the model (10) the appropriate polynomial models for $\Delta\Theta^1$, $\Delta\Theta^2$, $\Delta\Theta^3$ can be derived and a set of three algebraic equations can be obtained and solved. One can derive the solution for this case as follows:

$$\begin{aligned} \widehat{a} &= \tfrac{3}{\delta T}(\Theta^1 - \tfrac{8}{\delta T}\Theta^2 + \tfrac{20}{\delta T^2}\Theta^3) \\ \widehat{b} &= \tfrac{12}{\delta T^2}(-\Theta^1 + \tfrac{7}{\delta T}\Theta^2 - \tfrac{15}{\delta T^2}\Theta^3) \\ \widehat{c} &= \tfrac{10}{\delta T^3}(\Theta^1 - \tfrac{6}{\delta T}\Theta^2 + \tfrac{12}{\delta T^2}\Theta^3) \end{aligned} \tag{20}$$

It is easy to verify from the analysis of the matrix K_ω elements that the estimates (20) approach those obtained by the LSM as N grows (the difference between the estimates didn't exceed 1% for $N = 100$). So, the expressions (20) can be used as the LSM-estimates for the large N. As usual this is the case.

This approach was applied by one of the coauthors to the problem of the angular rate polynomial model coefficients evaluation that forms the basis for deriving the strapdown attitude algorithms [6]. It was conventional to calculate the angular rate model coefficients via the gyro incremental outputs in a deterministic way with the number of involved gyro samples being equal to the number of model coefficients to be determined. The proposed approach resulted in deriving a new class of attitude algorithms that are identical in form to the corresponding conventional ones of Miller-type [7], but use several sequential angular rate multiple integrals instead of the same number of conventional gyro samples over the iteration interval. The derived algorithms provide effective smoothing of the gyro instrument noise and by these means attenuate the pseudoconing errors. Fig.3 shows the

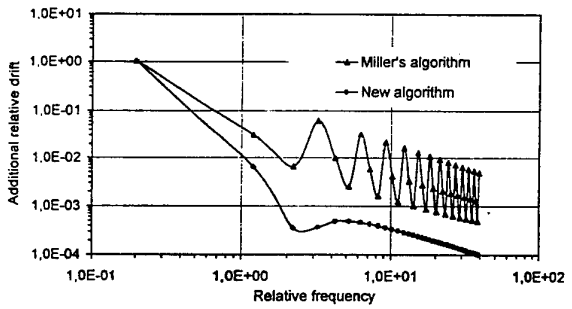


Figure 3: Algorithm pseudoconing response (log-log plot)

relative pseudoconing drift as a function of the relative frequency of both conventional Miller's algorithm and the appropriate new one with the angular motion combined with the high-frequency gyro noise components obtained by simulation. In the simulation the iteration rate was selected so that the aliased noise frequency was exactly equal to the angular motion frequency. It can be seen from the figure that the improvement of the new algorithm over the Miller's one is in approximately 2 orders in all reasonable frequency range.

In that study another unique property of the new attitude algorithms was analytically justified. It was shown that the derived algorithms provide a peculiar kind of the error "autocompensation" when the most significant terms of two rotation vector error components (one, caused by discarding the higher terms in the expression for the rotation vector, and other - by the errors in calculating the angular rate model's coefficients) compensate each other. It results in an accuracy improvement of the derived algorithms over the corresponding conventional ones in 2 orders under the general angular input. So, an accuracy of the corresponding conventional algorithm optimized for the coning input is provided with the new algorithms under the general angular input as well.

5. Signal Multiple Integrals Generation

Let us examine the possibility of the signal multiple integrals generating for both signal and signal-integrating sensors. As an example of the signal-integrating sensor we'll take a single-axis ring lazer gyro (RLG).

In RLG the increments of the angular rate (ω) one-time integral are generated by counting the number of angle quantums (n_i^1) and multiplying the accumulated number of quantums (n^1) over the sampling interval (δT) by the quantum value ($\Delta\Theta_0$), i.e.

$$\Delta\Theta^1 = \int_T^{T+\delta T} \omega(t)dt = n^1\Delta\Theta_0 \qquad (21)$$

The upper index at $\Delta\Theta$ denotes the multiplicity of the angular rate integral.

To generate the angular rate multiple integrals over the sampling interval we only need to take the current counter readings (n_i^1) with the fixed frequency ($1/\delta t$), being divisible by the sampling frequency ($1/\delta T$), and to perform the summing of these quantities as follows:

$$\begin{aligned} n^2 &= \sum_{i=1}^{N} n_i^1 \\ n^3 &= \sum_{i=1}^{N} n_i^2 \\ &\vdots \quad \cdots \\ n^r &= \sum_{i=1}^{N} n_i^{r-1} \end{aligned} \qquad (22)$$

where $N = \delta T/\delta t$

Then for the angular rate multiple integrals we have:

$$\Delta\Theta^r = n^r \cdot \delta t^{r-1} \cdot \Delta\Theta_0; \quad j = 1, 2, \cdots \qquad (23)$$

In the case of the signal sensor the additional outputs can be generated directly by multiple integration of signal discrete values. It should be noted that integrating can be replaced by summing (as in the case

of RLG). In doing so the relations (22) can be used, provided that in this case

$$n^1 = \sum_{i=1}^{N} \omega_i \qquad (24)$$

Then for the multiple integrals it gives:

$$\Delta\Theta^r = n^r \cdot \delta t^r; \quad j = 1, 2, \cdots \qquad (25)$$

Today, the trend is to integrate the digital signal processors (DSP) into the sensor electronics units of the real-time systems. DSP allows to mechanize any pre-processing algorithm and to modify it if necessary. The price one must pay for this benefit is in adding complexity to the sensor electronics that results in a price raise. Besides, the RTS designer has to buy and maintain the technology for developing, debugging and loading the DSP software that could not be economically attractive in particular cases.

Meanwhile, there is another way to generate the increments of signal multiple integrals: the series adders are to be built into sensor electronics. This can be implemented using the simplest digital electronic components without adding noticeable complexity to the sensor electronics (i.e. without using DSP) and in our opinion is not difficult.. This scheme of signal pre-processing could be economically promising for a wide class of RTS, in particular, when the sensor digital electronics has the only function of analog-to-digital converter (ADC). To implement the suggested approach in this case a special type of ADC should be designed, which have the additional outputs over each measurement channel that are the increments of signal multiple integrals. Such a modification of the existing ADC of general applicability would cost almost nothing for the ADC manufacturers, moreover the noticeable effect could be achieved with a few (2-3) additional outputs.

Broadly speaking, the generation of the signal multiple integrals is a kind of signal pre-processing that allows to effectively solve a few typical problems of RTS, but beeing very simple in implementation (only summing) it can be performed in a sensor digital hardware with low expences.

6. Conclusion

The approach for signal processing using the increments of signal multiple integrals over the sampling interval was represented as a general one for the real-time systems that are to solve one of the following problems: multiple integration of a signal, multiple integration of a product of two signals, or signal smoothing by polynomial fitting. Originally it was developed for the strapdown INS and resulted in the new classes of navigation and attitude algorithms. The proposed approach meets one of the basic principles of the signal processing practice that is to accurately compress the raw sensor data before the final processing. The solutions for all the problems under consideration were derived in such a mathematical form that the signal pre-processing is reduced to the signal multiple integration over the computation iteration interval. In this case all the problems under consideration can be correctly solved under the high-frequency sensor instrumental noise components with low sensor sampling rate specified by the benign environment dynamics. Besides, it may result in the appriciable cost savings thanks to the fact that the multiple integrals can be easily generated directly in the sensor hardware with low expenses.

Appendix: Derivation of the General Quadrature Formulas

Substituting (7) into Eq.(5) for the current values of the velocity increments over the interval $(t_k, t_k + \Delta T)$ yields:

$$\begin{aligned}\delta\mathbf{V}(t_k, t_k+\tau) = C(k)\int_0^\tau \mathbf{a}_b(t_k+\tau)d\tau \\ +\dot{C}(k)\int_0^\tau \tau\,\mathbf{a}_b(t_k+\tau)d\tau \\ +\ddot{C}_k\int_0^\tau \tfrac{\tau^2}{2}\,\mathbf{a}_b(t_k+\tau)d\tau + \ldots\end{aligned} \qquad (A1)$$

where $0 \le \tau \le \Delta T$

With the introduced designations (9) expressions for the integrals in $(A1)$ can be transformed by integrating by parts as follows:

$$\begin{aligned}\int_0^\tau \mathbf{a}_b(t_k+\tau)d\tau &= \boldsymbol{\Delta}\mathbf{V}^1(t_k, t_k+\tau) \\ \int_0^\tau \tau\,\mathbf{a}_b(t_k+\tau)d\tau &= \tau\,\boldsymbol{\Delta}\mathbf{V}^1(t_k, t_k+\tau) \\ &\quad -\boldsymbol{\Delta}\mathbf{V}^2(t_k, t_k+\tau) \\ \int_0^\tau \tfrac{\tau^2}{2}\,\mathbf{a}_b(t_k+\tau)d\tau &= \tfrac{\tau^2}{2}\boldsymbol{\Delta}\mathbf{V}^1(t_k, t_k+\tau) \\ &\quad -\tau\,\boldsymbol{\Delta}\mathbf{V}^2(t_k, t_k+\tau) \\ &\quad +\boldsymbol{\Delta}\mathbf{V}^3(t_k, t_k+\tau)\end{aligned} \qquad (A2)$$

Substituting $(A2)$ into $(A1)$ we obtain:

$$\begin{aligned}&\delta\mathbf{V}_n(t_k, t_k+\tau) \\ &= \left[C(k) + \dot{C}(k)\tau + \ldots\right]\boldsymbol{\Delta}\mathbf{V}^1 \\ &- \left[\dot{C}(k) + \ddot{C}(k)\tau + \ldots\right]\boldsymbol{\Delta}\mathbf{V}^2 \\ &+ \left[\ddot{C}(k) + C^{(3)}(k)\tau + \ldots\right]\boldsymbol{\Delta}\mathbf{V}^3 - \ldots\end{aligned} \qquad (A3)$$

In view of (5) and the polynomial expansion for the matrix C and its derivatives, $(A3)$ can be rewritten as follows:

$$\begin{aligned}&\int_0^\tau C(t_k+\tau)\,\mathbf{a}_b(t_k+\tau)d\tau\\&=C(t_k+\tau)\boldsymbol{\Delta}\mathbf{V}^1(t_k,t_k+\tau)\\&-\dot{C}(t_k+\tau)\boldsymbol{\Delta}\mathbf{V}^2(t_k,t_k+\tau)\\&+\ddot{C}(t_k+\tau)\boldsymbol{\Delta}\mathbf{V}^3(t_k,t_k+\tau)-\dots\end{aligned}\qquad(A4)$$

For $\tau=\Delta T$, $(A4)$ transforms into the first equation of (8).

To derive the second equation of (8) let us represent the quantity under examination in the following form:

$$\delta\mathbf{S}_n(k,k+1)=\int_0^{\Delta T}\delta\mathbf{V}_n(t_k,t_k+\tau)\,d\tau\qquad(A5)$$

Substituting $(A4)$ into $(A5)$ yields:

$$\begin{aligned}&\delta\mathbf{S}_n(k,k+1)\\&=\int_0^{\Delta T}C(t_k+\tau)\boldsymbol{\Delta}\mathbf{V}^1(t_k,t_k+\tau)\,d\tau\\&-\int_0^{\Delta T}\dot{C}(t_k+\tau)\boldsymbol{\Delta}\mathbf{V}^2(t_k,t_k+\tau)\,d\tau\\&+\int_0^{\Delta T}\ddot{C}(t_k+\tau)\boldsymbol{\Delta}\mathbf{V}^3(t_k,t_k+\tau)\,d\tau-\dots\end{aligned}\qquad(A6)$$

As the structures of the polynomial expansion for the matrix C and its derivatives are identical, each integral in $(A6)$ can be represented in a form similar to $(A4)$, namely:

$$\begin{aligned}&\int_0^\tau C\,\boldsymbol{\Delta}\mathbf{V}^1d\tau\\&=C\,\boldsymbol{\Delta}\mathbf{V}^2-\dot{C}\,\boldsymbol{\Delta}\mathbf{V}^3+\ddot{C}\,\boldsymbol{\Delta}\mathbf{V}^4-\dots\\&\int_0^\tau\dot{C}\,\boldsymbol{\Delta}\mathbf{V}^2d\tau\\&=\dot{C}\,\boldsymbol{\Delta}\mathbf{V}^3-\ddot{C}\,\boldsymbol{\Delta}\mathbf{V}^4+C^{(3)}\,\boldsymbol{\Delta}\mathbf{V}^5-\dots\\&\int_0^\tau\ddot{C}\,\boldsymbol{\Delta}\mathbf{V}^3d\tau\\&=\ddot{C}\,\boldsymbol{\Delta}\mathbf{V}^4-C^{(3)}\,\boldsymbol{\Delta}\mathbf{V}^5+C^{(4)}\,\boldsymbol{\Delta}\mathbf{V}^6-\dots\\&\cdot\\&\cdot\\&\cdot\end{aligned}\qquad(A7)$$

In view of expressions $(A7)$, $(A6)$ can be transformed to the form of the second equation of (8).

Acknowledgments

This study was partially sponsored by the Russian Foundation for Basic Research under its grant *No* 97-01-01134.

REFERENCES

1. Lesyuchevsky,V.M., and Litmanovich,Yu.A., "Specific Force Transformation and Navigation Discrete Algorithms of Inertial Navigation Systems (additional paper)," *Proceedings of the 1st Saint-Petersburg International Conference on Gyroscopic Technology, post-conference re-edition*, St.Petersburg, Russia, May 25-26, 1994.

2. Savage,P.G., "Strapdown Inertial Navigation Integration Algorithm Design Part 2: Velocity and Position Algorithms," *Journal of Guidance, Control, and Dynamics*, Vol.21, No.2, 1998, pp.208-221.

3. Litmanovich,Yu.A., Lesyuchevsky,V.M., Gusinsky V.Z., "Study of Strapdown Transformation Algorithms with Specific Force Multiple Integrals as Input Signals," *Proceedings of the 5th Saint-Petersburg International Conference on Integrated Navigation Systems*, St.Petersburg, Russia, May 25-27,1998, pp.45-55.

4. Litmanovich,Yu.A., Lesyuchevsky,V.M., Gusinsky V.Z., "Use of Specific Force Multiple Integrals as Input Signals for Strapdown Navigation Algorithms ," *Proceedings of Symposium Gyro Technology 1998*, Stuttgart, Germany, September 17-18,1998,pp.15.0-15.13.

5. Mark,J.G., and Tazartes,D.A., "On Sculling Algorithms," *Proceedings of the 3d Saint Petersburg International Conference on Integrated Navigation Systems*, Part II, St.Petersburg, Russia, May 28-29, 1996, pp.22-26.

6. Litmanovich,Yu.A., "Use of Angular Rate Multiple Integrals as Input Signals for Strapdown Attitude Algorithms ," *Proceedings of Symposium Gyro Technology 1997*, Stuttgart, Germany, September 16-17,1997,pp.20.0-20.9.

7. Miller, R.B., "A New Strapdown Attitude Algorithm," *Journal of Guidance, Control and Dynamics*, Vol.6, No.4, 1983, pp.287-291.

Fault-tolerant Strapdown Inertial Measurement Unit: Failure Detection and Isolation Technique

L.V. Vodicheva
Scientific and Research Association of Automatics
145, Mamina-Sibiryaka street, Ekaterinburg, Russia, 620 075,
e-mail: larisa@gduma.mplik.ru

Abstract

Familiar and new methods of self-contained failure detection and isolation technique in respect to sensors of redundant strapdown Inertial Measurement Unit are analyzed and systematized in the paper. Conditions of non-sensitivity of FDI algorithms to failures are obtained. Measurement unit with any number of sensors with input axes arbitrary arranged in three-dimensional space is under consideration.

Introduction

The use of redundant number of angular velocity and linear acceleration sensors in inertial navigation systems has been considered for many years by researchers and engineers. Inertial sensor redundancy is regarded as a way to improve reliability and accuracy of a lower level of a system that is Inertial Measurement Unit (IMU). Special attention was paid to redundant IMU of strapdown Inertial Navigation Systems (SINS) and Attitude and Heading Reference Systems (AHRS). The same reliability and accuracy improvements are obtained under essentially lower expenditures for strapdown systems than for platform systems. This is not the only reason for attention to redundant strapdown IMU. The problem of optimal use of redundant measurements is exciting itself. It is connected with some interesting and difficult mathematical problems in the fields of linear algebra, many-dimensional geometry, optimization theory, probability theory and mathematical statistics.

In the latest 60-th and particularly in the early 70-th strapdown IMUs were widely discussed in the papers. American researchers and engineers were among the first to present their papers in non-confidential proceedings and journals. The papers of J.P. Gilmore, R.A. Mckern, J.C. Wilcox, J.E. Potter, J.G. Mark, R.E. Ebner, E. Gai presented in the 70-th depend on the early studies of the 60-th. They are devoted to actual systems with given number of defined sensors and their input axis configuration. Six measurement instruments considered mainly in these papers allow to localize simultaneously up to two fault sensors and to detect three failures. The special cases of optimization procedure for four, five and six inertial instrument configurations are considered in the paper [6]. Optimization technique and Failure Detection and Isolation (FDI) algorithms for arbitrary number of sensors were not discussed. On the other hand in American papers even in early 70-th not only theoretical results, but the results of design and tests were presented.

Simultaneously Russian researchers considered redundancy as a more general problem of system quality improvement without application to a concrete system. Various fields of redundancy use were studied. Functional diagnosis and fault search in various technical systems were among the main of them. In 1980 L.A. Mironovskiy presented a survey of dynamical systems functional diagnosis technique [9] based on 124 references. There were papers devoted to various methods of diagnosis and fault search, automatic error correction, optimal estimation of state vector under redundant number of observations. In some papers connection between redundancy and symmetry was considered. The most of the papers considered in the survey were published in the early 70th and do not directly concern navigation and orientation systems. They were devoted to technical systems in general.

A.D. Epiphanov is one of the founders of technical system reliability theory in Russia. In his book [10] the problems of optimal use of redundancy are considered in connection with aircraft control systems. This is fundamental theoretical and methodological work. But there is no sufficient consideration of optimization problem for sensor axis arrangement. Optimization procedures are complicated and special. Optimization is implemented on restricted set of configurations. There are no general conditions of optimality. FDI algorithms described in the book are special ones as well.

An interesting concept of optimal orientation of sensor input axes is proposed by E.A. Belov [11]. He was using technique of design of experiment theory stated in the books [12]-[14]. This allowed him to obtain optimality conditions and to construct some sets of optimal configurations for arbitrary number of measurement instruments. Unfortunately his results were not widely presented at the conferences and in proceedings and journals on navigation and gyroscopic technology. Some years later I obtained the same optimality conditions and optimal configuration for arbitrary number of axes with the help of another technique [15].

Thus, significant and various experience has been accumulated in theory and practice of redundant fault-tolerant strapdown inertial measurement units. This experience needs to be analyzed for using in today's system design. New generation of low cost miniature

Paper presented at the RTO SCI International Conference on "Integrated Navigation Systems", held at "Elektropribor", St. Petersburg, Russia, 24-26 May 1999, and published in RTO MP-43.

mechanical inertial sensors appeared in early 90-th finally changed our image of gyroscope, and the problem of redundancy does not lose its actuality.

Various methods for failure detection and isolation technique ensuring fault-tolerance operation of strapdown inertial measurement unit with loaded reserve (all the operable sensors are continuously used to estimate vector components in a body-fixed orthogonal reference frame) are analyzed in the paper. The problem is considered with respect to arbitrary number and any kind of inertial measurement instruments. This work prolongs the study with results presented by the author at the Saint Petersburg International Conference on Integrated Navigation Systems in May, 1997 [15] and uses the results of Belov's work [11].

1. Basic problems

There are following main concepts of redundancy use depending on input axis orientation [10]:

- multiplicity usage of each single-axis sensor;
- multiplicity usage of each non-redundant vector measuring instrument;
- structural redundancy, which means that redundant number of sensors forms single vector measuring instrument with non-orthogonal arrangement of their input axes, any three of input axes are non-coplanar.

The most efficient is structural redundancy. In particular to ensure fault-tolerant operation under single point failure the first and the second redundancy concepts require nine single axis sensors whereas the third concept requires only five sensors. Further we shall consider only structural redundancy.

There are four main theoretical problems to be solved for Strapdown IMU with structure redundancy. They are

1) to determine the minimal number of sensors needed to ensure IMU efficiency under not more than specified number of failures;
2) to construct failure detection and isolation algorithms;
3) to obtain algorithms for estimation of vector components in an instrument orthogonal frame; these algorithms must minimize an estimation error;
4) to determine how sensor input axes are to be optimally arranged to ensure measurement errors to affect system accuracy minimally and to ensure FDI effectiveness.

Basic problems are connected to each other. However detailed consideration of all these problems in one paper would be overloading. So only the first two problems are considered in this paper. The third one is considered in connection with failure detection and isolation only. The problem of optimization of input axis configuration dependent on chosen estimating algorithms is worth to be considered in a special paper.

Let us introduce some definitions and restrictions to consider the chosen problems in detail without loss of generality. We shall consider measurement unit based on single-input-axis inertial sensors measuring projections of angular velocity and linear acceleration. Methods considered in the paper could be used for any kind of such sensors. Assume sensor errors to be independent. Independence of measurement error is a feature of a well-designed system [2]. Let instrument orthogonal frame be any body-fixed orthogonal right frame. Under failures we mean catastrophic faults including temporary ones, i. e. cases of sudden and rough degradation of accuracy. So if a sensor error is over permissible threshold information of such sensor is either excluded of further processing at all or is not used on a current updating interval. In this case an additive error component is enough to be taken into consideration, a more detailed instrument error model is not required. We shall not consider step-by-step degradation of sensor accuracy caused for example by depending on time changes and wear. We shall not consider the problem of calibration with the help of redundant measurements either. These problems are worth to be considered separately. Data processing algorithms for redundant inertial measurement unit under step-by-step degradation are considered for example in the paper [2]. There is one more serious problem being out of consideration in this paper. This is measurement noise influence upon FDI effectiveness. It could be solved by means of pre-filtering [1],[2],[7]. Filtering algorithms depend on a sensor error model and we are going to devote consideration to the algorithms not depending on an instrument type and an instrument error model.

Let us go to formalized description of considered problems. Under assumed restrictions measurement equations for 3-dimensional vector $\bar{x}$ are known [20] to be written as:

$$Z = H \cdot X + E \text{, where} \qquad (1.1)$$

X is 3×1 matrix; its components represent orthogonal projections of a measured vector;

H is $N \times 3$ matrix; its rows consist of direction cosines of input axes in an instrument orthogonal frame; any three rows are linear independent and

$$h_{n1}^2 + h_{n2}^2 + h_{n3}^2 = 1, n = 1,2,...N; \qquad (1.2)$$

Z is $N \times 1$ matrix representing the result of vector $\bar{x}$ measuring with N instruments;

E is $N \times 1$ matrix with additive measurement errors as components.

For linear measurement equations estimation of a state vector is linear as well:

$$\hat{X} = G \cdot Z \text{, where} \qquad (1.3)$$

G is $3 \times N$ matrix with weight coefficients depending on estimating algorithm as components.

Now the problems considered in the paper could be presented as follows:

1. Under given N^*_{max} find such N that under any $N^* \le N^*_{max}$ unambiguous identification of N^* failures can be ensured.
2. Under given matrix H satisfying (1.2) and having any three rows linear independent construct failure detection and isolation algorithms ensuring failure localization under $N^* \le N^*_{max}$ failures and failure detection in any other case, and construct algorithms of matrix G reconfiguration.
3. Under given matrix H find matrix G(H) minimizing matrix X estimation errors.

We shall consider mainly the second problem in this paper. Minimal number of sensors required for ensuring specified fault-tolerance level depends on a chosen FDI algorithm. The third problem will be concerned, as it is needed for consideration of FDI algorithms. It was analyzed for example in [4]. We shall base here on using the least square method (LSM). In assumption of measurement errors to be additive independent random variables with zero mean values the least square method is advisable to be used in processing of redundant data. Under mentioned assumptions LSM ensures the minimal estimation error variance for each component on the set of all the linear unbiased estimates.

2. FDI methods

Measurement redundancy allows to detect a fault sensor and to eliminate its outputs in vector estimation algorithm without external measurements. This could be done in various ways. FDI effectiveness is defined by a minimal number of sensors ensuring localization of a specified number of possible failures, FDI algorithm complexity with respect to implementation in board computer, probability of error arise. In addition clearness of physical interpretation is also worth to be taken into consideration. It is important while debugging and testing. According to [10] FDI errors are of I type and II type. The I type error means that unfailed instrument is indicated as failed. The II type error means that failed instrument is indicated as unfailed.

Probability of the I type error depends on measurement noise characteristics and specified error threshold. Theoretically filtering could reduce this type of error to zero. In addition, if updating frequency of FDI algorithm is rather high these errors do not result in failure indication. They result only in excluding information of appropriate sensors from data processing on some FDI updating intervals. The II type errors are inherent in principle. It will be shown while analyzing FDI methods.

All self-contained FDI methods are based on comparison of some functions of sensor outputs to specified threshold. Failure localization is implemented either by comparison to threshold or in non-threshold way. For both variants of failure localization we shall consider the problem connected with the case of an error beyond the threshold on FDI updating interval and with identification of fault sensor set. We shall not consider how to choose a threshold and how to evaluate the errors of the I and II type. These problems are interconnected and are worth to be solved for a concrete system with a given error model.

Failures are divided to unit and group failures. Group failure means that all the N^*_{max} sensors may appear to be false simultaneously. Under unit failure sensors are assumed to get out of order one by one, i.e. on single updating interval only one possible failure is assumed. On single updating interval unit failure could be considered as a special case of a group failure. Let us consider familiar and new methods of failure detection and isolation in connection with specified number of group possible failures N^*_{max} for redundant strapdown inertial measurement unit with arbitrary number N of inertial sensors for measuring of each vector.

3. Method of voting

The method of voting [6],[10] is convenient for failure detection and localization to be physically interpreted. Its origin is connected with voted redundancy of units and instruments for reliability improvement of a system. For a redundant vector measurement unit this method could be described as follows. From the system of N equations with three unknowns (1.1) all possible subsystems of three equations are selected. Solution of each of these subsystems represents an estimate of vector X: $\hat{X}_m$, m = 1,2...M; $M = C_N^3$. Detection of failure and localization of failed sensor is implemented by comparing these estimates between each other.

When $N^* = N - 3$ sensors are failed, i. e. there is only three unfailed sensors, only one of estimates is true. In this case all the M estimates are different and failed sensors are impossible to be selected of unfailed ones by "voting". Only the presence of failure can be detected. With four unfailed sensors there are $C_4^3 = 4$ close estimates and four "votes" are enough in principle to localize a set of failed sensors. Thus, **to ensure fault-tolerant operation of a measurement unit under N^*_{max} possible failures number of sensors has to be no less than $N^*_{max} + 4$ for each measured vector,** i. e. $N^*_{max} + 1$ sensors have to be in loaded reserve (all the operable sensors are continuously used to estimate vector).

Let us show that **if N-dimensional vector of errors is a linear combination of matrix H columns:**

$$\mathbf{E} = \alpha \mathbf{H}^{(1)} + \beta \mathbf{H}^{(2)} + \gamma \mathbf{H}^{(3)}, \text{ where} \qquad (3.1)$$

$H^{(j)}, j = 1,2,3$ is N-dimensional column vector of matrix H, then failure detection and localization are impossible. Errors which FDI algorithm is non-sensitive to form 3-dimensional subspace since matrix H has rank 3.

Estimate $\hat{X}_m$ obtained by processing of information of any three sensors with numbers m_1, m_2, m_3 can be written as follows:

$$\hat{X}_m = H_m^{-1} \cdot (H_m \cdot X + E_m), \text{ where} \quad (3.2)$$

$$H_m = \begin{bmatrix} h_{m_1,1} & h_{m_1,2} & h_{m_1,3} \\ h_{m_2,1} & h_{m_2,2} & h_{m_2,3} \\ h_{m_3,1} & h_{m_3,2} & h_{m_3,3} \end{bmatrix}; \; E_m = \begin{bmatrix} \varepsilon_{m_1} \\ \varepsilon_{m_2} \\ \varepsilon_{m_3} \end{bmatrix} \quad (3.3)$$

If (3.1) is true then

$$\hat{X}_m = X + H_m^{-1} \cdot E =$$

$$= X + H_m^{-1} \cdot \begin{bmatrix} \alpha h_{m_1,1} + \beta h_{m_1,2} + \gamma h_{m_1,3} \\ \alpha h_{m_2,1} + \beta h_{m_2,2} + \gamma h_{m_2,3} \\ \alpha h_{m_3,1} + \beta h_{m_3,2} + \gamma h_{m_3,3} \end{bmatrix} =$$

$$= X + \alpha H_m^{-1} \cdot \begin{bmatrix} h_{m_1,1} \\ h_{m_2,1} \\ h_{m_3,1} \end{bmatrix} + \beta H_m^{-1} \cdot \begin{bmatrix} h_{m_1,2} \\ h_{m_2,2} \\ h_{m_3,2} \end{bmatrix} + \gamma H_m^{-1} \cdot \begin{bmatrix} h_{m_1,3} \\ h_{m_2,3} \\ h_{m_3,3} \end{bmatrix} =$$

$$= X + \alpha \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} + \beta \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} + \gamma \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} = X + \begin{bmatrix} \alpha \\ \beta \\ \gamma \end{bmatrix} \quad (3.4)$$

It means that in this case all the M estimates are close not depending on values of errors. When errors are linear combinations of matrix H columns (3.1) all the sensors may get out of order and "voting" will demonstrate that there are no errors.

If a number of failed sensors is N^*_{max}, there are other sets of errors non-detectable in principle. These sets can be obtained by solving comparison equations with errors as unknowns for all possible sensor quartets. These cases are simpler to be explained for method of algebraic invariants.

Algorithm of voting is rather complicated. It requires to calculate C_N^3 estimates of a measured vector and to analyze a number of comparison equations:

$$\hat{X}_{m_1} - \hat{X}_{m_2} = 0, \; m_1, m_2 = 1,2,...M, m_1 \neq m_2 \quad (3.5)$$

If functional dependency is not taken into consideration a number of these equations is C_M^2. For example for six sensors this number is 190.

Comparison equations (3.5) can be presented as linear combinations of sensor errors:

$$\hat{X}_{m_1} - \hat{X}_{m_2} = H_{m_1}^{-1} \cdot (H_{m_1} \cdot X + E_{m_1}) -$$

$$- H_{m_2}^{-1} \cdot (H_{m_2} \cdot X + E_{m_2}) = H_{m_1}^{-1} \cdot E_{m_1} - H_{m_2}^{-1} \cdot E_{m_2} \quad (3.6)$$

It could be shown that for failure localization it is sufficient to select only those comparison equations that contain all possible sensor quartets with one equation for each quartet. Then number of comparison equations will reduce to C_N^4 and the method of voting transforms to the method of algebraic invariants.

4. Method of algebraic invariants

Measurement redundancy allows constructing linear functions of sensor outputs to be invariant to measured vector. These functions are linear combinations of measurement errors. In the method of algebraic invariants a set of linear functions is constructed such a way that an analysis of their values allows localizing unambiguously a set of failed sensors.

This method could be formulated as follows. Find $M \times N$ matrix V complying with the conditions:

$$V \cdot H = 0 \quad (4.1)$$

Multiplying from the left measurement equation (1.1) to matrix V and taking into account (4.1) we shall have

$$\Delta = V \cdot Z = V \cdot E, \text{ where} \quad (4.2)$$

Δ is $N \times 1$ matrix with linear combinations of sensor errors as components. Matrix V is to ensure localization of a maximal number of possible failures with a minimal number of rows.

This method was named as method of algebraic invariants in [10]. It is one of the most widely used for failure detection and localization in strapdown measurement units [1],[4],[7],[16],[17]. The problem was considered many times in respect to six gyroscopes or accelerometers. The form of matrix V was obtained for two optimal sensor configurations, which are 1) conic arrangement of sensor axes with half-angle of the cone equal to $\alpha = \arccos\sqrt{\frac{1}{3}}$ [16],[17]; 2) arrangement of sensor axes as normals to nonparallel faces of dodecahedron [1]. However there is no derivation procedure of matrix V in any of papers mentioned.

In my paper [15] the method of algebraic invariants is used in respect to measurement unit with arbitrary number N of axes and any possible orientation of their input axes with the only restriction that any three of them are non-coplanar. It is shown that in this case as well as in the method of voting a minimal number of sensors required for localization of N^*_{max} possible failures is $N^*_{max} + 4$. Procedure of deriving matrix V from components of matrix H is described. It is shown that a minimal number of rows of matrix V is $C_N^{N^*_{max}}$ and any

row has N^*_{max} zeros; in addition for $N = N^*_{max} + 4$ these rows up to coefficients are determined unambiguously. Failure detection and isolation algorithm is described.

Let us pay our attention to conditions of non-sensitivity of this algorithm to the II type errors. This problem is practically considered neither in [15], nor in other papers with the exception of [7].

It is obvious that if 6-dimensional error vector is a linear combination of columns of matrix H, i. e. meets the expression (3.1) then failure detection and localization are not possible either for method of algebraic invariants. In this case $\Delta = 0$ under any values of errors and any number of failed sensors since rows of matrix V are orthogonal to columns of matrix H.

Another case arises while $N = N^*_{max} + 4$ and a number of failed sensors is equal to maximum number required for fault-tolerant operation N^*_{max}, i. e. only one sensor quartet is unfailed. In this case only one of the relations

$$V_m \cdot Z = 0, m = 1,2,...C^4_N \qquad (4.3)$$

has to be true up to chosen threshold. If in addition errors of some other quartet having any values meet the condition:

$v_{p,i_1}\varepsilon_{i_1} + v_{p,i_2}\varepsilon_{i_2} + v_{p,i_3}\varepsilon_{i_3} + v_{p,i_4}\varepsilon_{i_4} = 0$, where (4.4)

$v_{p,i_k}, k = 1,2,3,4$ are non-zero components of corresponding row of matrix V then separation of unfailed four sensors from failed four sensors is impossible.

If $N \geq N^*_{max} + 4$ rows of matrix V could be chosen in an other way and duplicate analysis of relation $\Delta = 0$ allows reducing error probability. It could be achieved by introducing at least one more sensor in inertial measurement unit with minimal required for a fault-tolerance number of sensors and by complicating algorithms.

For unit failure the algorithm could be modified to a non-threshold version [11]. In this version there is only one variable to be compared with given threshold at the stage of error detecting. This is a function of errors. FDI algorithm is constructed as follows. In advance matrix V is obtained. For unit failure model $N^*_{max} = 1$, $C^1_N = N$, matrix V is a square one with diagonal components equal to zero: $v_{n,n} = 0, n = 1,2,...N$. It means that n-th linear function is independent on information of n-th sensor.

While system operation following steps are implemented on FDI updating interval in a board computer:

1. Algebraic invariant square sum is calculated:

$$\delta^2 = \Delta^T \cdot \Delta = Z^T V^T V Z \qquad (4.5)$$

2. For $\delta^2 \leq \delta^2_{spec}$ all sensors are considered to be unfailed.

3. For $\delta^2 > \delta^2_{spec}$ sensor number n* is recognized to be failed if

$$\frac{(V_{n^*}, Z)^2}{\|V_{n^*}\|^2} = \min_{1 \leq n \leq N} \frac{(V_n, Z)^2}{\|V_n\|^2}, \text{ where} \qquad (4.6)$$

(V_n, Z) means scalar product of vectors V_n and Z;

$\|V_n\|^2 = \sum_{i=1}^{N} v^2_{n,i}$ means square of Euclidean norm of vector V_n.

It could be interpreted as a search for sensor having minimal relation of square appropriate linear function of errors to variance of this function considered as a random variable under assumption of normalized and non-correlated measurement errors.

Similar algorithms could be constructed if in (4.5) algebraic invariants module sum $\sum_{n=1}^{N} |V_n Z|$ or maximum of algebraic invariants modules $\max_{1 \leq n \leq N} |V_n Z|$) were used instead of square function. Yet to my mind linear space with Euclidean distance is more convenient to solve the problem.

5. Methods of comparison with average

For this group of methods comparison equations are formed as functions of residuals that are differences between measurements and their estimates [1],[10]. Measurement estimates and function of residuals used in comparing with threshold depend on a type of optimal estimate of state vector. "Average" means measurement estimate corresponding to the optimal estimate of state vector. The form of optimal estimates is determined by optimization criterion. Theoretically in this method any optimization criterion could be used. The choice of criterion is usually connected with assumed probability distribution of errors. In real navigation systems measurement errors are usually assumed to have normal probability distribution. In this assumption the most effective estimates are those obtained with the help of the least square method (LSM) [18]. In addition LSM is well interpreted for redundant measurements geometrically. Measurement equations can be presented as follows [19]:

$$Z = x_1 H^{(1)} + x_2 H^{(2)} + x_3 H^{(3)} + E, \qquad (5.1)$$

i. e. up to errors measurement vector could be presented as linear combination of matrix H columns. LSM-estimate presents a vector belonging to the subspace spanned on column vectors of matrix H. The distance between this vector and measurement vector is minimal in N-dimensional Euclidean space. We use here the least square method based on minimizing sum of residual

squares. In this case comparison equations are square functions of residuals.

Measurement errors in equation (1.1) can be supposed to be normalized and uncorrelated; this does not hurt our concept. In addition measurement equations (1.1) can be transformed to make measurement errors meet these conditions [20]. Probability model could not be used at all. LSM-estimate based on all the equation (1.1) can be written in following way:

$$\hat{X} = (H^T \cdot H)^{-1} \cdot H^T \cdot Z \qquad (5.2)$$

Respectively estimate of n-th measurement can be written as follows:

$$\hat{z}_n = H_n \cdot (H^T \cdot H)^{-1} \cdot H^T \cdot Z, \text{ where} \qquad (5.3)$$

H_n is n-th row of matrix H.

Matrix of residuals is a linear function of measurement errors:

$$\Delta = Z - \hat{Z} = (E_N - H \cdot (H^T H)^{-1} H^T)(H \cdot X + E) = \\ = (E_N - H \cdot (H^T H)^{-1} H^T) \cdot E, \text{ where} \qquad (5.4)$$

E_N is identity matrix $N \times N$.

It is easy to show that matrix $W = E_N - H \cdot (H^T \cdot H)^{-1} H^T$ is symmetric and idempotent, i. e. $W^T = W$ and $W^2 = W$. This is followed by

$$W^T W = W = E_N - H \cdot (H^T \cdot H)^{-1} H^T \qquad (5.5)$$

The form of FDI algorithm for this method is like the form of non-threshold version of FDI algorithm for the method of algebraic invariants.

1. Residual square sum is calculated:

$$\delta^2 = \Delta^T \cdot \Delta = Z^T W^T W Z = \sum_{n=1} (z_n - \hat{z}_n)^2 \qquad (5.6)$$

2. For $\delta^2 \le \delta^2_{spec}$ all sensors are considered to be unfailed.

3. For $\delta^2 > \delta^2_{spec}$ sensor number n* is recognized to be failed if

$$D_{n^*} = \frac{(z_{n^*} - \hat{z}_{n^*})^2}{1 - H_n \cdot (H^T H)^{-1} H_n^T} = \\ = \frac{(W_{n^*}, Z)^2}{\|W_{n^*}\|^2} = \max_{1 \le n \le N} \frac{(W_n, Z)^2}{\|W_n\|^2} \qquad (5.7)$$

and its information is ignored on current FDI updating interval. Adaptation algorithm is rather simple. Reconfiguration of matrix G amounts to reducing to zero appropriate row of matrix H while obtaining estimate of state vector (5.2) and estimate of measurement (5.3). The expression (5.7) as well as for non-threshold version of method of algebraic invariants could be interpreted as relation of square residual to variance of this residual under assumption of normalized and uncorrelated measurement errors.

4. Steps 1,2,3 are repeated for the rest sensors.

5. If after N^*_{max} iterations $\delta^2 > \delta^2_{spec}$ IMU failure is stated; localization of a set of unfailed sensors is impossible.

To increase the failed sensor error weight the algorithm of comparison with average could be constructed in some other way. In the n-th comparison equation vector X estimate and appropriate measurement estimate are obtained with the help of the rest sensors (and not with all the sensors). In this case vector X estimate looks as follows:

$$\hat{X}_n = (\tilde{H}_n^T \cdot \tilde{H}_n)^{-1} \cdot \tilde{H}_n^T \cdot \tilde{Z}_n, \; n = 1,2,...N, \text{ where} \qquad (5.8)$$

$\tilde{H}_n$ is $(N-1) \times 3$ matrix obtained from matrix H by excluding of n-th row;

$\tilde{Z}_n$ is $(N-1) \times 1$ matrix obtained from matrix Z by excluding of n-th measurement. The appropriate estimate of the n-th measurement can be written as

$$\hat{z}_n = H_n \cdot (\tilde{H}_n^T \cdot \tilde{H}_n)^{-1} \cdot \tilde{H}_n^T \cdot \tilde{Z}_n \qquad (5.9)$$

It can be shown that relation (5.7) transforms for this algorithm to

$$D_n = \frac{(z_n - \hat{z}_n)^2}{1 + H_n \cdot (H^T \cdot H)^{-1} H_n^T} \qquad (5.10)$$

The algorithm itself is similar to above described one.

For both types of algorithms residuals are linear functions of measurements, i. e. $\Delta = W \cdot Z$. They can be shown to be invariant to the measured vector.

$$\Delta = W \cdot Z = W \cdot (H \cdot X + E) = W \cdot H \cdot X + W \cdot E \qquad (5.11)$$

With zero errors $\Delta = 0$, so $W \cdot H \cdot X = 0$ under any X. Hence $W \cdot H = 0$ and rows of matrix W are orthogonal to columns of matrix H. It means that these algorithms (like to above described algorithms) are non-sensitive to measurement errors in the case that these errors are linear combinations of columns of matrix H, i. e. comply with the expression (3.1).

E. Belov in the paper [11] proposed and studied in detail for single-point failures the algorithm named by him "the method of generalized likelihood relation". In his paper the failure detection and localization of fault sensor are considered as problems of the test of statistical hypothesis. It allows to chose a threshold and to estimate probability of FDI II type errors depending on assumed sensor error probability characteristics. This algorithm could be considered also as method of comparison with average. It is as follows. The procedure of failure detection is the same as for the first type algorithm of comparison with the average. It is implemented with the help of relations (5.4), (5.6), all the measurements are used to form measurement estimate. To localize failure N decision functions are constructed as follows:

$$D_n^* = (\tilde{Z}_n - \hat{Z}_n)^T \cdot (\tilde{Z}_n - \hat{Z}_n) = \tilde{Z}_n^T \cdot L_n \cdot \tilde{Z}_n, \\ n = 1,2,...N, \text{ where} \qquad (5.12)$$

$\hat{Z}_n = \tilde{H}_n \cdot (\tilde{H}_n^T \tilde{H}_n)^{-1} \tilde{H}_n^T$ and accordingly

$$L_n = E_{N-1} - \tilde{H}_n \cdot (\tilde{H}_n^T \tilde{H}_n)^{-1} \tilde{H}_n^T \qquad (5.13)$$

The decision function represents the sum of squares of N – 1 residuals. It can be obtained by using the least square method to process information of measurement unit without sensor number n. It is analogous to expression (5.6) for all the measurement instruments. Decision function D_n^* is minimal for sensor recognized as failed. Further it is shown in the paper [11] that

$$D_n^* = Z^T \cdot L \cdot Z - \frac{(z_n - H_n \cdot (H^T H)^{-1} \cdot H \cdot Z)^2}{1 - H_n \cdot (H^T H)^{-1} \cdot H_n^T}, \text{ where} \qquad (5.14)$$

$$L = E_N - H \cdot (H^T H)^{-1} H^T$$

Taking into consideration that the first term in the right part of expression (5.14) is the same for all the indexes n the author suggests replacing search for sensor with minimal D_n^* by search for sensor with maximal function:

$$D_n = \frac{(z_n - H_n \cdot (H^T H)^{-1} \cdot H \cdot Z)^2}{1 - H_n \cdot (H^T H)^{-1} \cdot H_n^T}. \qquad (5.15)$$

Expression (5.15) is completely the same as (5.7). It means that the method of generalized likelihood relation suggested by E. Belov is equivalent to the method of comparison with the average based on all the measurements.

6. Method of error space using

The method considered below was suggested in the paper [9] for unit failures and was named "the generalized likelihood test". The authors also consider failure detection and localization in this method as tests of statistical hypothesis. I would like to interpret this method in terms of linear algebra and many-dimensional geometry.

Measurement errors in equation (1.1) can be presented as vectors of N-dimensional space R^N, $E \in R^N$. This space is a direct sum of two subspaces [21]:

$$R^N = H \oplus M, E = E_H \oplus E_M, \text{ where} \qquad (6.1)$$

H is three-dimensional subspace formed by column vectors of matrix H; all the algorithms being considered are non-sensitive to error component E_H. If $E \in H$ failure detection and localization are impossible for familiar algorithms processing only sensor outputs;

M is (N – 3)-dimensional subspace orthogonal to H; considered FDI algorithms detect and localize failed sensors in respect to E_M only.

Let us construct orthonormal basis of space M, i. e. find N – 3 N-dimensional vectors M_k such that

$$M_k \cdot H = 0; \|M_k\| = \sqrt{\sum_{i=1}^{N} \mu_{k,i}^2} = 1; \qquad (6.2)$$

$$(M_k, M_j) = \sum_{i=1}^{N} \mu_{k,i} \mu_{j,i} = 0; k, j = 1,2,...K, K = N - 3$$

This can be done with the help of the known orthogonalization procedure [21]. These vectors are used as rows to form matrix M. Rows of matrix M represent N – 3 independent on each other linear functions invariant to a measured vector. Matrix product $M \cdot Z$ is representation of error component E_M in orthonormal basis $\{M_k\}$:

$$\Delta = M \cdot Z = M \cdot (HX + E) = M \cdot E = \begin{bmatrix} (M_1, E) \\ (M_2, E) \\ \dots \\ (M_K, E) \end{bmatrix} \qquad (6.3)$$

Matrix M columns $M^{(j)}$ could be interpreted as K-dimensional failure vectors of j-th sensor, j=1,2...N. These vectors are collinear to vector Δ if j-th sensor error differs from zero and the rest sensor errors are equal to zero.

The algorithm described in the paper [8] consists of following procedures. If

$$\delta^2 = \Delta^T \Delta \le \delta_{spec}^2, \qquad (6.4)$$

hypothesis of lack of failures is accepted, otherwise a failure is recognized to occur. If hypothesis of a failure is accepted sensor number n* is recognized to be failed if

$$D_{n^*} = \frac{(\Delta, M^{(n^*)})^2}{\left\|M^{(n^*)}\right\|^2} \qquad (6.5)$$

has maximal value.

Geometrically it could be interpreted as follows. In the procedure of failure detection (6.4) error vector norm is compared to threshold. In the procedure of failure localization (6.5) a value proportional to square cosine of angle between error vector and j-th failure vector is determined for each sensor. A sensor is marked as failed if this value is maximal and the corresponding angle between vectors is minimal.

7. Relations between FDI methods

Different FDI methods are interrelated. Let us consider these relations for non-threshold versions of methods. Non-threshold methods are considered mainly for unit failures. In these methods comparison with specified is implemented only on the stage of failure detection. Failure localization is realized by either minimization or maximization of some functions determined on associated finite sets.

Non-threshold methods include:

1) non-threshold version of method of algebraic invariants;

2) methods of comparison with average;
3) method of error space using.

E. Belov has proved equivalence of these methods for single-point failure [11]. Some of the results obtained by him are presented below.

Equivalence of comparison with the average to the method of error space using is proved as follows. In the stage of failure detecting in the method of comparison with the average the following value is compared with a threshold:

$$\Delta^T\Delta = Z^T(E_N - H\cdot(H^TH)^{-1}H^T)Z, \quad (7.1)$$

in error space method this is the following value:

$$\Delta^T\Delta = Z^TM^TMZ \quad (7.2)$$

In proving the equality

$$M^TM = E_N - H\cdot(H^TH)^{-1}H^T \quad (7.3)$$

the relations

$$M\cdot H = 0 \quad (7.4)$$

$$M\cdot M^T = E_K, K = N-3 \quad (7.5)$$

are used.

Let us denote right part of equality (7.3) by A, left part by B. Matrix B is idempotent. Matrix A is resulted from (7.5) to be idempotent too. It can be shown from (7.4) that $A\cdot B = B\cdot A = A$. A trace of matrix A is equal to a trace of matrix B since trace of symmetrical idempotent matrix is equal to its rank [22]:

$$SpurA = SpurB = K = N-3 \quad (7.6)$$

Let us introduce C = B – A, which is idempotent as well. It is resulted from (7.6) that SpurC = SpurC^2 = 0. But

$$SpurC^2 = \sum_{i=1}^{N}\sum_{j=1}^{N} c_{i,j}^2 = 0 \quad (7.7)$$

It means that C = 0 and A = B.

Let us consider now the numerator of the value (6.5) used in the error space method for failure localization:

$$(\Delta, M^{(n)})^2 = (e_n^T\cdot M^TM\cdot Z)^2 = \\ = (z_n - H_n\cdot(H^TH)\cdot H^T\cdot Z)^2 = (z_n - \tilde{z}_n)^2, \text{ where} \quad (7.8)$$

e_n is n-dimensional vector with n-th component not equal to zero and the rest components equal to zero. Denominator of (6.5) can be transformed to

$$\left\|M^{(n)}\right\|^2 = e_n^T\cdot M^TM\cdot e_n = 1 - H_n\cdot(H^TH)^{-1}\cdot H_n^T \quad (7.9)$$

Thus expression (6.5) is equal to expression (5.7) and the equivalency of considered methods is completely proved.

Equivalence of the method of comparison with the average (named by Belov as method of generalized likelihood relation) to method of algebraic invariant is proved in [11] for the case of minimal redundancy, i. e. for five sensors in measuring three-dimensional vector. In this case row vectors of matrix V up to norm are determined identically. Matrix V with normalized rows meets the following expressions:

$$V\cdot H = 0 \quad (7.10)$$

$$v_{n,n} = 0, n = 1,2,...N \quad (7.11)$$

$$\|V_n\|^2 = 1 \quad (7.12)$$

$$v_{n,m} \neq 0, n, m = 1,2,...N; n \neq m \quad (7.13)$$

Since rows of matrix M form a basis orthogonal to columns of matrix H space, matrixes V and M are connected as follows:

$$V = S\cdot M, \text{ where} \quad (7.14)$$

S is $N\times 2$ matrix. Rows of matrix S meet the following conditions:

$$(S_{(n)}, M^{(n)}) = 0, n = 1,2,..N \quad (7.15)$$

$$S_{(n)}M^TMS_{(n)}^T = S_{(n)}S_{(n)}^T = 1 \quad (7.16)$$

In non-threshold version of algebraic invariants for failed sensor value of function

$$\frac{(V_{(n)}, Z)^2}{\left\|V_{(n)}\right\|^2} = (V_{(n)}, Z)^2 = (MZ)^T\cdot(S_{(n)}^TS_{(n)})\cdot(MZ) \quad (7.17)$$

is minimal. Belov has shown that

$$S_{(n)}^TS_{(n)} = E_2 - \frac{1}{\left\|M_{(n)}\right\|^2}M^{(n)}M^{(n)T} \quad (7.18)$$

and hence

$$(V_{(n)}, Z)^2 = \|MZ\|^2 - \frac{(MZ, M^{(n)})^2}{\left\|M^{(n)}\right\|^2} \quad (7.19)$$

It means that minimization of (7.17) is equivalent of maximization of (6.5). So for minimal redundancy the method of algebraic invariants is equivalent to the method of error space using, hence all three methods being considered are equivalent.

Conclusion

Linear metric space formalism is rather convenient in consideration of failure detection and localization technique for vector redundant inertial measurement instruments. It allows simply and uniformly describing various, at first sight, FDI algorithms. This could be used for analysis of existing FDI algorithms and development of new ones for a measurement unit with any number of arbitrary arranged measurement axes. The proposed concept being geometrically more obvious is well complied with the known solutions presented in the papers. This is with the help of this concept the conditions of algorithm non-sensitivity to failures have been obtained for the first time. These conditions appear to be the same for all the algorithms considered in the paper.

References

1. Gilmore J.P., Mckern R.A. A Redundant Strapdown Inertial Reference Unit. Journal of Spacecraft and Rockets, V. 9, No. 1, 1972.

2. Wilcox J.C. Maximum Likelihood Failure Detection for Redundant Inertial Instruments. AIAA Paper, No. 863, 1972.
3. Chien T.-T. An Adaptive Technique for a Redundant-Sensor Navigation System. AIAA Paper No. 863, 1972.
4. Potter J.E., Deckert J.C. Minimax Failure Detection and Identification in Redundant Gyro and Accelerometer Systems. Journal of Spacecraft and Rockets, V. 10, No. 4, 1973.
5. Hung T.C., Doran B.J. High-Reliability Strapdown Platforms Using Two-Degree-of-Freedom Gyros. IEEE Transaction on Aerospace and Electronic Systems, AES-9, No. 2, 1973.
6. Pejsa A.J. Optimum Skewed Redundant Inertial Navigators. AIAA Journal, V. 12, No. 7, 1974.
7. Ebner R.E., Mark J.G. Redundant Integrated Flight-Control/Navigation Inertial Sensor Complex. Journal of Guidance and Control, V. 1, No. 2, 1978.
8. Daly K.C., Gai E., Harrison J.V. Generalized Likelihood Test for FDI in Redundant Sensor Configurations. Journal of Guidance and Control, V. 2, No. 4, 1979.
9. Mironovskiy L.A. Functional Diagnosis of Dynamical Systems. Survey. Automatics and Remote Control, No 8, 1980 (in Russian).
10. Epiphanov A.D. Redundant Control Systems for Aircraft. Moscow, "Mashinostroenie", 1978 (in Russian).
11. Belov E.A. Development of Algorithms and Methods for Reliable and Precise Estimation of Motion Parameters in Systems with Structural Redundancy. Doctor thesis in technical science. Leningrad Polytechnic Institute, Leningrad, 1981 (in Russian).
12. Nalimov V.V. Theory of Experiments. Moscow, "Nauka", 1971 (in Russian).
13. Fyodorov V.V. Theory of Optimal Experiments. Moscow, "Nauka", 1971 (in Russian).
14. Adler Yu.P., Markova E.V., Granovskiy Yu.V. Design of Experiments in Search of Optimal Conditions. Moscow, "Nauka", 1976 (in Russian).
15. Vodicheva L. Optimization and Failure Detection for redundant Strapdown Inertial Measurement Unit. 4th Saint Petersburg International Conference on Integrated Navigation Systems, May 26-28, 1997.
16. Diagnosis Algorithms for SINS Redundant Measurement Unit. Scientific and technical report on theme "Robins", Scientific and Production Association "Energia", Moscow, 1992 (in Russian).
17. Izmailov E.A., Tchesnokov G.I., Troizkij V.A., Gordasevich A.A. Cheap Small Size Inertial Navigation System with Increased Reliability. The 2nd Saint-Petersburg International Conference on Gyroscopic Technology and Navigation, Part I, Saint-Petersburg, May 24-25, 1995.
18. Mudrov V.I., Kushko V.L. Methods of Measurement Processing. Moscow, "Radio i svyaz", 1983 (in Russian).
19. Shilov G.E. Calculus. Finite-dimensional linear space. Moscow, "Nauka", 1969 (in Russian).
20. Brammer K., Ziffling G. Kalman-Bucy Filter. Moscow, "Nauka", 1982 (in Russian, translated from German).
21. Maltzev A.I. Fundamentals of Linear Algebra. Moscow, "Nauka", 1970 (in Russian).
22. Bellman R. Introduction to Matrix Analysis. McGraw-Hill Book Company 1960.

"Soyuz" - "Mir" Orbital Flight GPS/GLONASS Experiment: First Results

Sergey Klyushnikov*, Sergey Filatchenkov*, Nicolai Mikhailov**, Sergey Pospelov**, Mikhail Vasilyev**,

* Russian Institute of Radionavigation and Time
2 Rastrelli Square,
193124 St.Petersburg,
Russian Federation

** Soft Nav Ltd.
19 Shkiperski Protok,
199106 St.Petersburg,
Russian Federation

Abstract

The combined GPS/GLONASS receiver ASN-2401P has been installed on the manned space ship "Soyuz-TM28" and was used to obtain experimental data during its flight to the space station "Mir" in August – November 1998. The ASN-2401P receiver is based on the ASN-22 eighteen-channel C/A-code avionics receiver module: a joint development of Dasa NFS (Germany, Ulm) and RIRT (Russia, St.Petersburg). The ASN-22 receiver module is described in brief. The receiver used in the experiment together with the antenna will become core elements of navigation system of Russian module of International Space Station "Alpha" and Russian space ships.

Raw pseudorange and carrier phase measurements, along with the position, velocity and time (PVT) results have been recorded during the autonomous flight of "Soyuz-TM28", rendezvous operations, as well as during the docked to "Mir" flight. The receiver installation, space ship attitude orientation modes, receiver control and data recording are described. Analysis of flight data is presented in the paper.

The accuracy assessment of snapshot PVT solutions with GPS and combined GPS/GLONASS measurements compared to the data of the orbit determined from the same measurements is presented. An orbit determination (OD) SW primarily developed by Soft Nav Ltd. for autonomous OD on-board a satellite with GPS measurements was used during the assessment. The features of this OD SW are a precise satellite dynamic model, and a stochastic "random walk" model based on the Kalman filter for clock modeling. Processing of the results has demonstrated that weighted least-square GPS/GLONASS PVT solutions become significantly more accurate with the increase of the number of GLONASS signals involved in the solutions.

The error sources limiting the ultimate accuracy of GPS/GLONASS orbital navigation are examined and discussed. Multipath errors have been experimentally evaluated. For this, differences between C/A-code measured pseudoranges (PR) and accumulated delta-ranges (ADR) measured on signal carrier have been analyzed. The analysis demonstrated an episodic growth of multipath errors up to 5-7 m (1σ). Such error values may become the dominating factor limiting the accuracy of relative spacecraft navigation with code measurements. Spectral characteristics of the multipath error were found to be favorable for PRs smoothing, together with ADR measurements. Unfortunately, PR-ADR differences have been found to diverge very fast (up to 0.7 m/s at negative elevations). A necessity for orbital applications of more sophisticated PR smoothing algorithms, than normally are used on the ground, e.g. utilizing inertial aiding, is discussed in brief. The experimental results will help to develop these smoothing algorithms reducing the errors of relative navigation required for spacecraft rendezvous operations.

Introduction

Navigation information provided by the Global Navigation Satellite Systems (GNSS) allows to tackle a wide range of tasks arising from space exploration and to cut down on the use of sufficiently complex and expensive classical ground-based orbit determination systems [1].

The Rocket Space Corporation (RSC) "Energia" plans to use GNSS based equipment on board space stations and space transport vehicles for autonomous determination of position, velocity and time (PVT) of space vehicles (SV). It is also planned to perform an autonomous attitude determination and relative navigation during rendezvous operations in space.
RSC envisages installing of combined GPS/GLONASS receivers on board Russian SV as the simultaneous use of both GPS and GLONASS diminishes dramatically positioning errors due to partial obstruction of navigation satellites by construction elements of the SV. This advantage of the combined GPS/GLONASS receiver

Paper presented at the RTO SCI International Conference on "Integrated Navigation Systems", held at "Elektropribor", St. Petersburg, Russia, 24-26 May 1999, and published in RTO MP-43.

becomes evident when possible mutual obstruction of navigation satellites by two SVs during a rendezvous is considered. Moreover, the redundant number of signals arising from the use of GLONASS enhances the autonomous integrity monitoring. Finally, due to absence of the Selective Availability (SA) in GLONASS the accuracy of the PVT solutions is higher than that of the GPS only receivers especially with the increase of the number of GLONASS signals used in the PVT solution.

In a framework of the development of new generation space-borne GPS/GLONASS receivers orbital flight experiment has been performed with the ASN-2401P GPS/GLONASS receiver.

The Receiver Description

The ASN-2401P is a prototype for a navigation receiver to be used by RSC as a part of the navigation equipment of the Russian segment of the International Space Station (ISS) "Alpha" and in some other space applications. The ASN-2401P receiver has been developed by the Russian Institute of Radionavigation and Time (RIRT) together with the Navigation Company Bering Ltd. to the specifications of RSC. The receiver consists of an active patch microstrip antenna, a navigation box and a notebook computer, which is responsible for controlling the receiver, storing the output data and visualization of the results. "Libretto" notebook made by Toshiba was used.

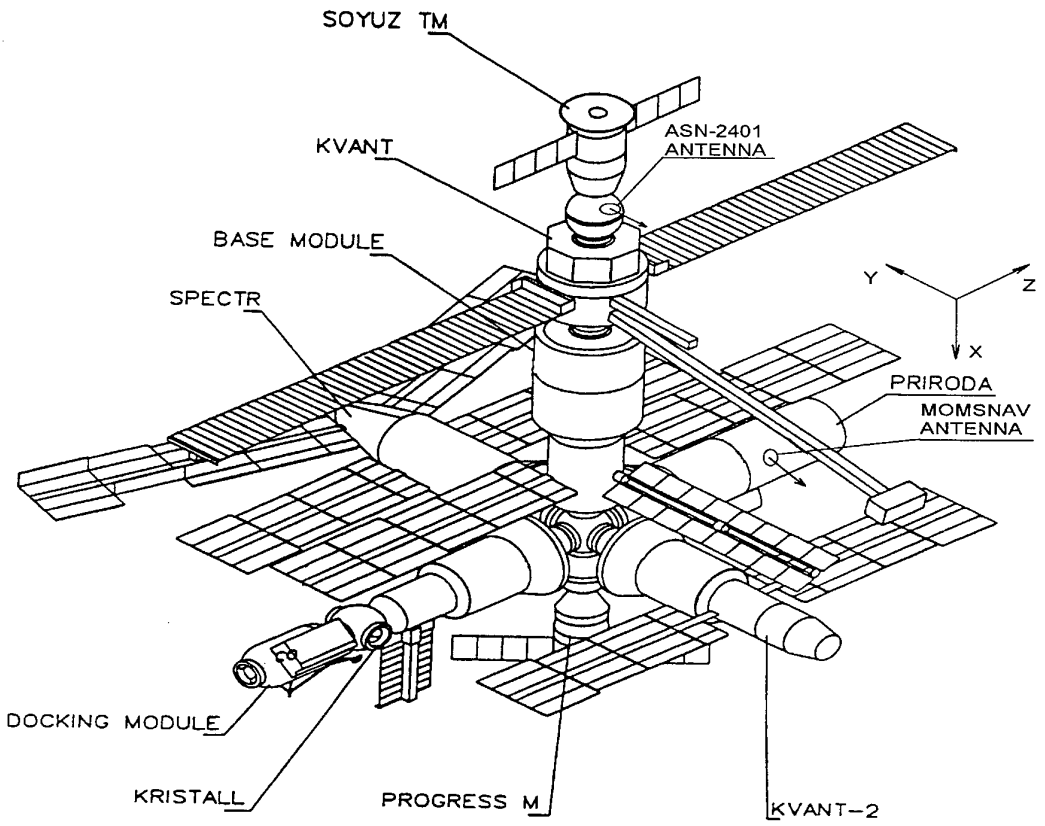


Fig.1. "Mir" Space Station

The navigation box contains an ASN-22A receiver card, a power supply module and an interface card.

The core functionality of the ASN-2401P is ensured by the ASN-22A receiver card. The ASN-22A receiver card is software upgraded ASN-22 avionics receiver card [2].

The ASN-22 receiver module is the joint development of DASA NFS (Ulm, Germany) and RIRT.

The ASN-22 is a high performance, differential-ready, multi-channel receiver capable of receiving and independently tracking the C/A code and carrier phase of up to 12 GPS satellites and 6 GLONASS satellites. It consists of a replaceable module containing all functions

needed to receive, decode and manage GPS and GLONASS satellite signals except antenna/pre-amplifier functions. The ASN-22 has a high sensitivity, high interference immunity and high code and carrier phase tracking accuracy. The 18-channel architecture provides rapid Time-To-First-Fix (TTFF) logic under all start-up conditions and robust performance in applications that require high dynamics. Rapid satellite acquisition is guaranteed under all initialization conditions as long as visible GPS or GLONASS satellites are not obscured. The ASN-22 fulfills all aviation requirements on accuracy and integrity including local differential correction processing, code/carrier phase smoothing and receiver autonomous integrity monitoring (RAIM). The unit accepts differential corrections for GPS and GLONASS satellites in an expanded RTCA SCAT-1 message type 1 format. The ASN-22 is able to meet all performance requirements as specified in TSO-C129a, Class B1, and RTCA DO-208 documents if implemented in a suitably configured unit.

The ASN-2401P was installed on board the SV "Soyuz-TM 28" and used as a navigation equipment during the flight in August December 1998. The navigation data were provided in various phases of the SV's orientation in autonomous flight, while approaching the "Mir" station and during the joint flight of the orbit complex "Mir-Soyuz". Fig.1 shows the joint flight of the orbit complex "Mir-Soyuz" and the siting of the ASN-2401P antenna. As it is seen from Fig.1 the "MOMSNAV" equipment developed by Kayser-Threde GmbH (Munich, Germany) was installed in "Priroda" module of the "Mir" station. A comparison of the data of ASN-2401P with the data of Motorolla's "Viceroy" GPS receiver included in the "MOMSNAV" will be performed later.

Experimental Data Description

Seven GPS-only and GPS/GLONASS combined navigation receiver sessions recorded from 13.08.1998 to 24.08.1998 have been involved into the first stage of experimental data processing. The total duration of these experimental sessions was more than 14 hours. The sessions differed in the space ship attitude orientation. The orientation types were:

- Orbital orientation (zenith pointing) – the space ship is attitude stabilized, the GPS/GLONASS antenna is pointed nearly to the zenith, satellite signals can be received nearly from the whole upper hemisphere;
- Sun-twist orientation - the space ship is slowly rotating, the GPS/GLONASS antenna changes its pointing with time that complicates signals reception;
- "On-duty" (inertial) orientation of the space station "Mir" – the station and the space ship docked to it are attitude stabilized; the GPS/GLONASS antenna of the space ship is pointed nearly horizontally; satellite signals can be received only from about one half of the upper hemisphere. Fig.1 shows the attitude of the space station "Mir" and the space ship Soyuz-TM28; the X-axis is directed along the flight vector, the Z-axis is pointing approximately to the zenith.

The condition for a satellite selection into constellations used for navigation was its elevation higher than minus 15° measured from the local horizon. This elevation threshold value resulted in great number of visible satellites and satellites used in navigation solutions. For example, during the docked to "Mir" flight (unfavorable "on-duty" orientation) on 24.08.1998 average figures were:

Visible GPS	- 13.7
Visible GPS+GLONASS	- 20.8
Used GPS	- 7.5
Used GPS+GLONASS	- 10.2

The report on relative GPS navigation experiment on Shuttle/Mir rendezvous [3] stated that during the critical phases of flight (approach, departure) there was a lack of valuable GPS data due to non favorable antennas pointing resulting from SVs' attitude. During the Soyuz/Mir experiment this was not a serious problem, as GLONASS satellites have been involved, and low elevation satellites were received.

The receiver and its active antenna system provided favorable conditions of satellite signals reception. The measured carrier-to-noise ratio histogram is shown in Fig.2. The average value of measured carrier-to-noise ratio was 47 dBHz.

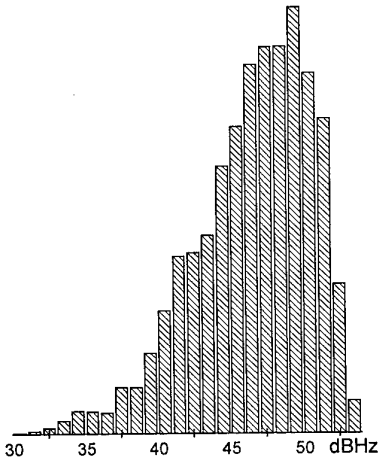


Fig.2. Measured Carrier-to-Noise Histogram Averaged over Seven Sessions.

The recorded data contained:

- Raw measurements: time stamped pseudorange and integrated carrier phase for each of the eighteen GPS

and GLONASS receiver channels at the rate of 1 Hz;
- Position, velocity and time (PVT) accompanied by the status and RAIM information at the rate of 1 Hz;
- Ephemeris and almanacs data for both GPS and GLONASS systems.

PVT solutions in ASN-2401P were combined GPS/GLONASS snapshot weighted minimum squares solutions. The troposphere and ionosphere models were not applied during the PVT solutions.

Raw Measurements Analysis

One of the primary goals of the experiment was the investigation of the peculiarities of GPS and GLONASS satellite signal reception on-board a spacecraft during its autonomous flight, rendezvous operations, as well as during its flight having been docked to an orbital station. To assess the quality of measurements and, in particular, to evaluate multipath reception and ionosphere refraction influence on pseudorange measurements, the differences between, firstly, the pseudoranges (PR) measured with signal code and, secondly, the accumulated delta-ranges (ADR) measured on signal carrier, has been analyzed.

The analysis of data registered during the rendezvous operations and the docked to "Mir" flight are of greater interest due to two reasons. The 1st is that the reflections from "Mir" constructions could cause especially significant multipath errors. The 2nd is that in this case the normal "Mir" attitude orientation allows the GPS/GLONASS antenna of the ASN-2401P to receive satellite signals only from a part (about one half) of the upper hemisphere but, at the same time, the reception of satellites below the horizon is possible.

Figures 3-6 present the most specific graphs of PR-ADR differences together with satellite elevation angle and carrier-to-noise ratio C/N during the docked to "Mir" flight on 24.08.1998. Satellite elevation was calculated during the processing from the user PVT and the satellite ephemeris or almanac data. The gaps in elevation graphs correspond to absence of PVT solutions.
The orbit height varied from 370 km to 381 km.

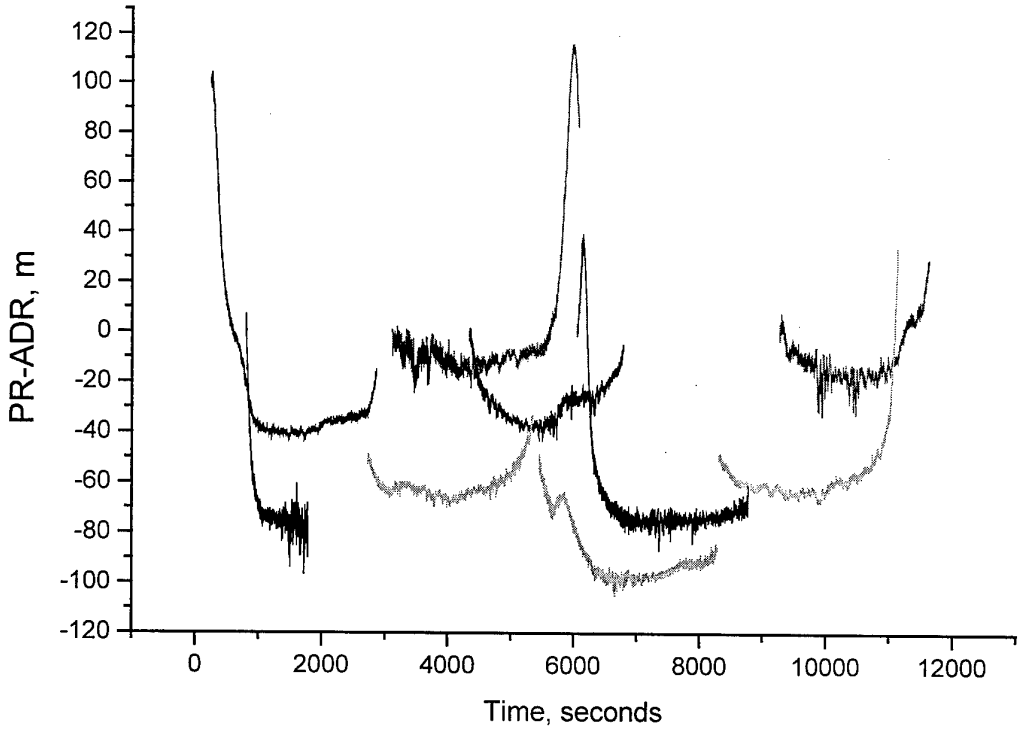


Figure 3. Pseudorange Minus Accumulated Delta-Range (PR-ADR) for Selected Satellites

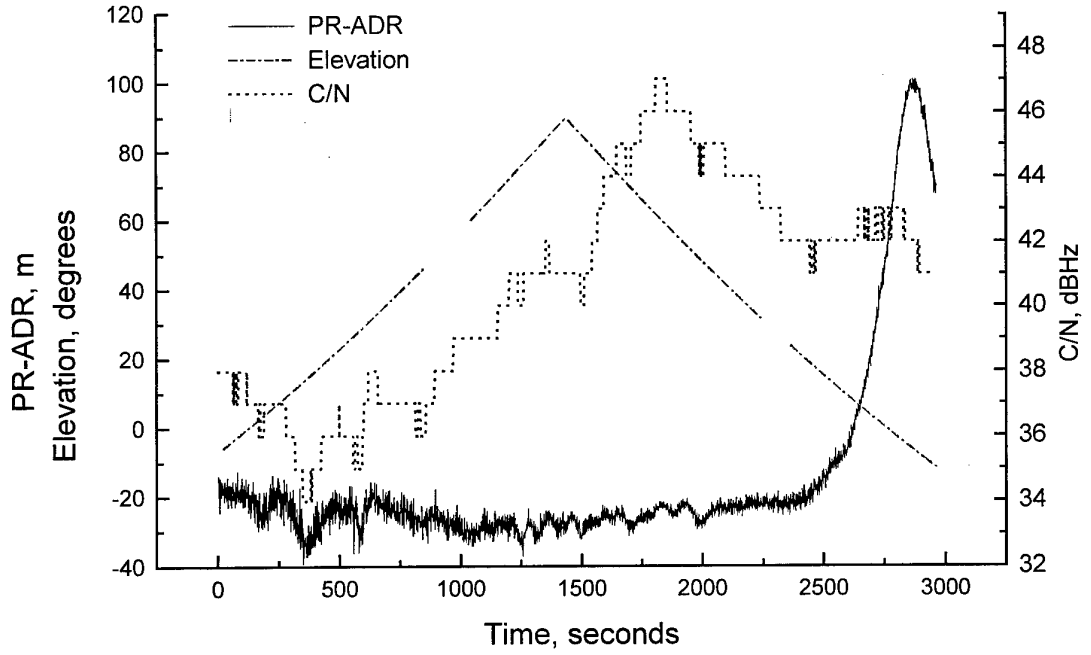


Figure 4. Example 1 of PR-ADR, Elevation Angle and C/N for a Selected Satellite

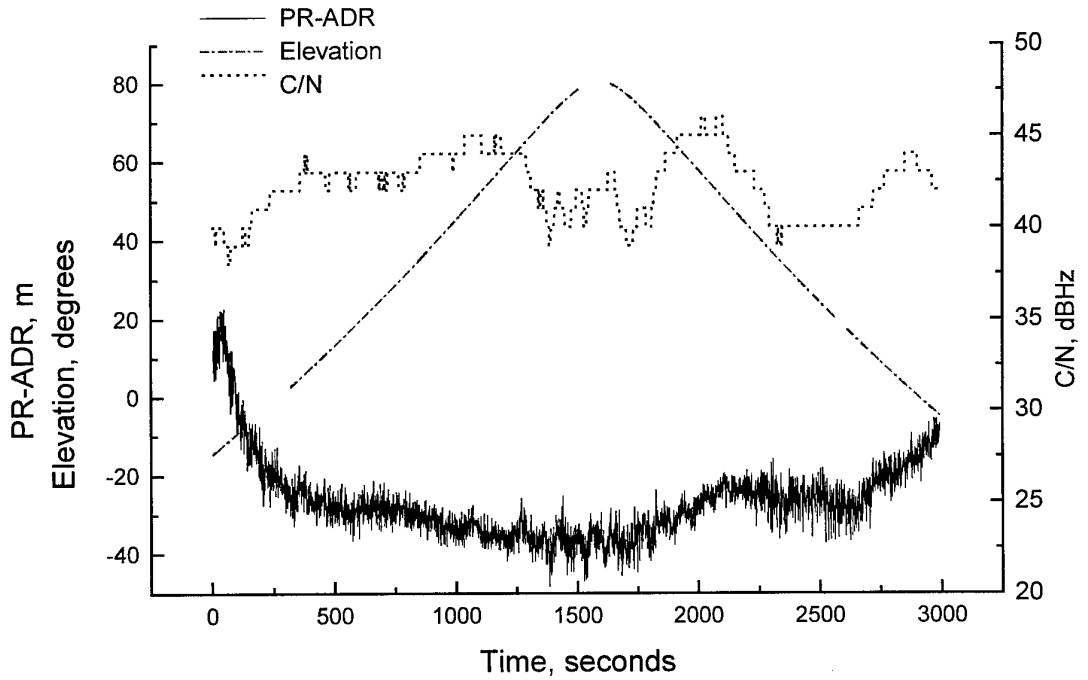


Figure 5. Example 2 of PR-ADR, Elevation Angle and C/N for a Selected Satellite

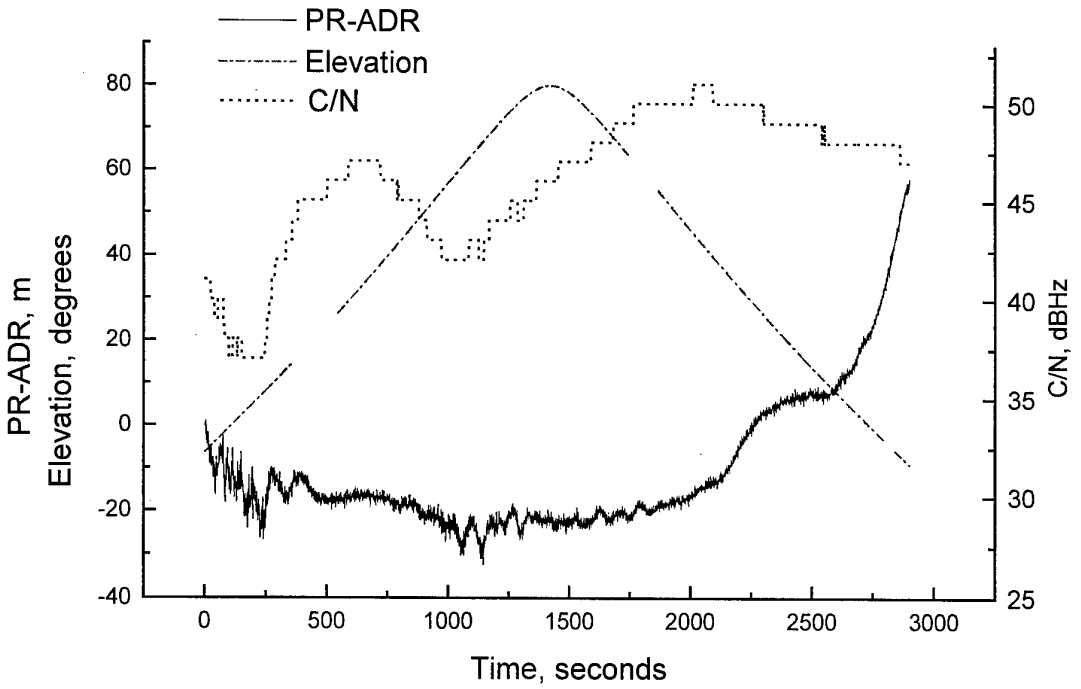


Figure 6. Example 3 of PR-ADR, Elevation Angle and C/N for a Selected Satellite

The analysis has shown that multipath errors episodically reached values up to 5 - 7 m (1σ) for duration of dozens of seconds. These errors are greater than those predictable from the DLL pure noise behavior at the registered carrier-to-noise ratio values no more than 1.5 – 2 times. Such level of multipath errors shall not cause severe problems for autonomous navigation of spacecraft, but can become the dominating factor limiting the accuracy of relative spacecraft navigation with code measurements. The largest part of PR-ADR span is obviously caused by the divergence of code and carrier measurements due to opposite signs of ionosphere refraction influence on these two types of measurements. The example graphs on Fig. 4-6 show that fast PR-ADR variations take place for satellite signals received at low elevations (<5°) and, especially, below the horizon. A detailed analysis of the dependence of the PR-ADR variation velocity upon the satellite elevation, and a comparison with the values predicted by ionosphere modeling are still to be carried out. At the time being, it is possible to state, nevertheless, that even during the periods of moderate solar activity at the altitudes that manned orbital stations normally operate the divergence of code and carrier measurements due to ionosphere refraction can be as fast as 0.7 m/s for satellite signals at elevations below 5°.

This result is important from the point of view of pseudorange measurements potential enhancement implementing for space-borne receivers of the widely used in on-ground GNSS receivers technique that usually is called the carrier smoothing of pseudoranges. Spectral characteristics of the multipath error were found to be favorable for PRs smoothing together with ADR measurements. Unfortunately, fast PR-ADR divergence makes practically impossible any sufficiently long smoothing with simple algorithms. The conclusion is that for orbital applications more sophisticated PR smoothing algorithms than normally are used on the ground, become necessary. These could include the ones utilizing inertial aiding. The experimental results will help to develop these smoothing algorithms reducing the errors of relative navigation required for spacecraft rendezvous operations.

Navigation Accuracy Assessment

One of the remarkable features of the experiment was that the accuracy of PVT data provided by ASN-2401P receiver was higher than the analogous data from all other available sources. But this positive result, as it is often the case, leads to some problems in estimation of real accuracy of the PVT data. The only one technique appears to provide more accurate orbital data compared with the experimental one - this is the method of orbit determination (OD) using a full scale dynamic model of

the satellite orbital motion. It is commonly accepted now that such model shall take into account the following effects: the Earth gravity potential, the lunar and solar gravitational attraction, the radiation pressure, the atmosphere drag, the solid Earth tides, the ocean tides and the relativistic effects. Note also that there are two available types of experimental data, which can be used for orbit determination in our case:

- The raw pseudorange and Doppler measurements.
- The snapshot PVT solutions calculated by the ASN-2401P receiver.

Since the processing of the second type of measurements was found to be easy for implementation the next steps were performed to construct Mir/Soyuz reference orbits:

1. The first available PVT solution is saved as initial coordinates and velocities for numerical integration of orbital motion equations.
2. The numerical integration of the equations is performed. The results of the integration are saved as an approximation to the reference orbit.
3. The differences between the reference orbit and the ASN-22 PVT solutions are calculated.
4. Corrections to the initial satellite position and velocity are calculated using results of step 2. In case the corrections become small enough the iteration process is interrupted. Otherwise steps 2-4 are repeated.

The residuals between the reference orbit and the snapshot PVT solutions are supposed to be exactly equal to the navigation errors. Some statistics of these errors are given in the Table 1.

Since the accuracy of ASN-2401P PVT data is much better than typical for GPS receivers (the result was expected) and worse than specified for GLONASS system we will discuss here primarily the last phenomena (the lack of accuracy for velocity measurements). First possible explanation is the poor geometry of the observed GLONASS constellations. In case when the number of GLONASS satellites used in the PVT solution is more than four the accuracy of 3-D position navigation is agreed with the PVT solution obtained during on-ground ASN-2401P testing. Significant dilution of precision of 3-D velocity determination can not be explained in the same way, though. It seems that the standard ionosphere and relativistic models [4] are not sufficient for the measurements reduction in case of space applications (see also the previous section for PR-ADR processing results).

Several cases of outliers in the PVT data were registered also. The embedded RAIM algorithm has detected problems and the corresponding PVT data were marked as incorrect. Nonetheless, an interesting conclusion can be done from the analysis. It is known that GLONASS system utilizes frequency division of satellite signals. There are GLONASS antipode satellites, which transmit signals on the same frequency and locate on the opposite sides of the Earth. This causes no problems for users on the Earth surface, as they can not observe two antipode satellites simultaneously, but two antipode GLONASS satellites have been observed simultaneously in all cases of wrong PVT solution (the outliers in the PVT data). Simple modeling has shown that there exist areas in space where two antipode GLONASS satellites are observed simultaneously. Therefore, some additional requirements shall be imposed on the navigation planning procedure to prevent attempt of simultaneous tracking of the antipode GLONASS satellites. This example once more shows that orbital specifics should be taken into account carefully during the development of navigation receivers for space applications.

Table 1.

Number of GLONASS SV used in the PVT	% from total time of navigation	GDOP Range	RMS (1σ) of 3-D position determination (m)	RMS (1σ) of 3-D velocity determination (m/s)
0 - 6	100.0	1.16 - 3.73	19.9	0.204
≥3	78.9	1.16 - 3.73	18.1	0.136
≥4	51.4	1.16 - 3.14	17.1	0.158
≥5	23.6	1.16 - 2.21	12.6	0.081

The Use of the Experimental Data for Orbit Determination

As stated above, the GPS technique is considered now as a preferred navigation tool to minimize space mission support from ground stations and to implement real time autonomous systems for orbit determination (OD). It is well known that the processing of real measurements is the best test for the navigation techniques. The experimental data have been used to test the real time OD software developed by Soft Nav Ltd. The key features of this software are precise satellite dynamic model, and a stochastic "random walk" model based on the Kalman filter for the user clock modeling. The test objectives were:

- Test navigation software using the real pseudorange and Doppler measurements.
- Demonstrate the potential accuracy of LEO OD in case of limited number of channels and poor geometry.

To fulfill the second task the conditions of GPS satellite visibility were impaired by using the next three strategies: valid pseudorange and Doppler measurements from two, four and eight GPS channels were used for OD from each ASN-2401P SV Measurements Message. Note that selection of valid measurements for the first two strategies was realized as a simple scan from the first GPS channel to the last one and GLONASS satellites were not used in processing at all.

It is clear that first two strategies are not optimal at all and can be considered as the worst case scenario for the specified number of simultaneously observed GPS satellites. The last strategy is intended to demonstrate potential OD accuracy in the case when optimal or quasi-optimal GPS satellite appointment procedure is implemented. Only one observation arc has been processed. This arc corresponds to the on-duty space station orientation and was measured on 24.08.98.

For all specified strategies orbital parameters were determined using about half-hour pseudorange and Doppler measurements span. The determined orbits have been compared with the reference one (see previous section for the method of calculation the reference orbit) and with the ASN-2401P PVT data obtained during flight experiment. Results of such comparison are represented at Fig. 7-8. Note that four type of differences are plotted on both graphs:

1. The difference between ASN-2401P PVT data (GPS/GLONASS combined mode) and reference orbit (marked "ASN snapshot")
2. The difference between "strategy 1" orbit and reference orbit (marked "2 GPS SV")
3. The difference between "strategy 2" orbit and reference orbit (marked "4 GPS SV")
4. The difference between "strategy 3" orbit and reference orbit (marked "All GPS SV")

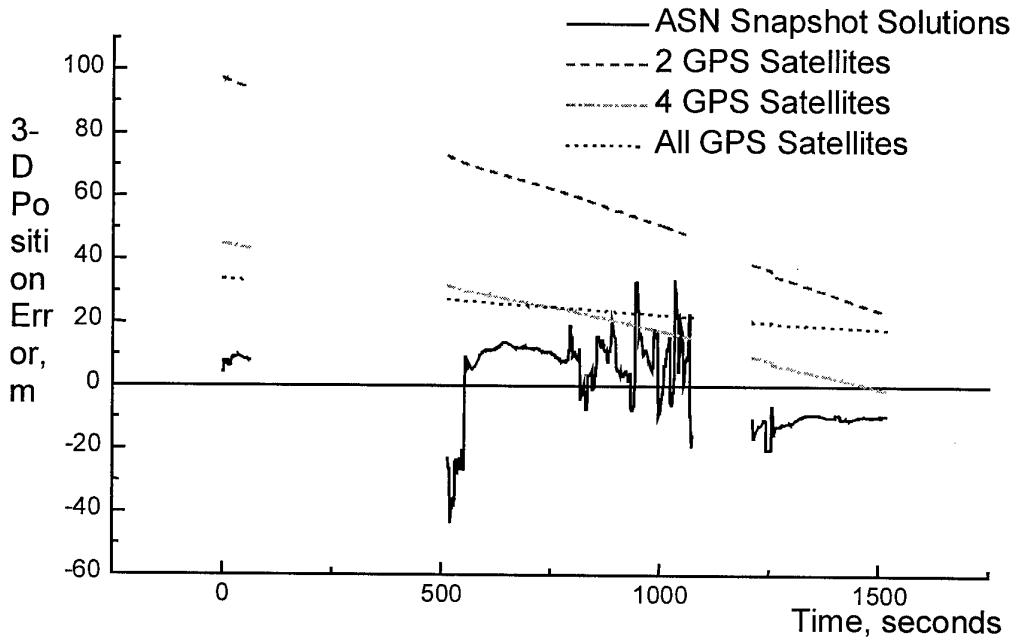


Fig.7. 3-D Position Error

In spite of preliminary character of the results it is already clear that even in the worst case scenario (strategy 1) OD software gives results which are better than it is typical for traditional snapshot solution in GPS ONLY mode. Note that according to standards for civil aviation 3-D position error shall be equal to 100 m (2 σ).

The next conclusion is that an optimal or quasi-optimal satellites appointment strategy together with OD technique allows to obtain precise satellite position

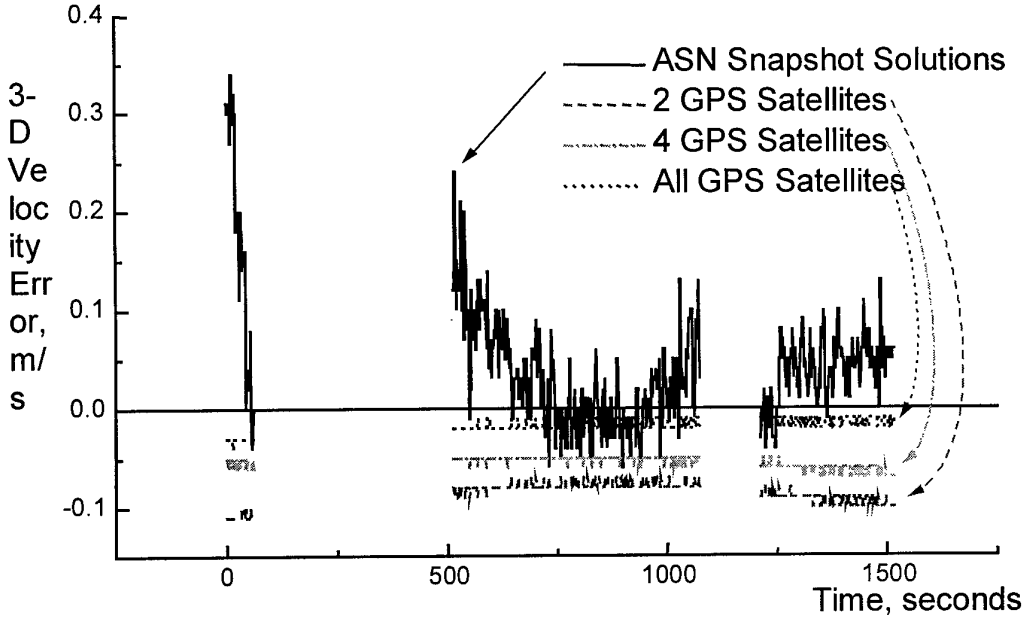


Fig.8. 3-D Velocity Error

estimation even for limited GPS channels number (or number of simultaneously tracked satellites in case of complicated satellite rotation). Moreover, the OD technique for strategies 2 and 3 demonstrates accuracy of satellite position determination, which is comparable with the combined GPS/GLONASS PVT solution.

As it is has been noticed already only one half orbit period was used for the OD. The short arc method usually used for the OD assumes measurements processing for at least one satellite orbital period (the orbital period of the experimental data was about 1.5 hour). It is known that a significant growth of accuracy of the satellite position determinations begins from arc length equals approximately to a half of the satellite orbital period. The authors have no opportunity to check this due to absence of appropriate uninterrupted measurements arc.

Fig. 8 shows also the high accuracy of the satellite velocity determination using the OD technique. This result validates also conclusions of the previous chapter about the considerably high error of the velocity determination observed in the ASN-2401P PVT data. Note that a correct modeling of ionosphere and relativistic effects shall significantly improve accuracy of the velocity and position determination in both cases: OD and traditional snapshot navigation solution techniques.

Conclusions

The experiment on board the space ship "Soyuz-TM28" facilitated validation of core GNSS receiver elements for future Russian segment of ISS and piloted transport SVs.

It has been demonstrated that the combined use of GPS and GLONASS enhances accuracy of in-orbit navigation and availability of navigation satellites during such important phases of flight as rendezvous operations and SV's unfavorable orientation including rotating flight. The errors of snapshot navigation solutions averaged throughout all the sessions with different space ship orientation were less than 20 m in coordinates and 0.2 m/s in velocity (3D, 1σ). With the increase of GLONASS measurements involved in navigation solution, the errors fell down to 12 m in coordinates and 0.08 m/s in velocity (3D, 1σ).

Multipath errors have been experimentally evaluated. The analysis demonstrated an episodic growth of multipath errors up to 5-7 m (1σ). Such error values may become the dominating factor limiting the accuracy of relative spacecraft navigation with code measurements.

The differences between code-measured pseudoranges and carrier-measured accumulated delta-ranges have been found to diverge very fast (up to 0.7 m/s at negative elevations). This effect prevents from use on board the SVs of widely known simple algorithms of pseudoranges smoothing that rely on the above mentioned difference

stability. A necessity for orbital applications of more sophisticated PR smoothing algorithms than normally are used on the ground, e.g. utilizing inertial aiding, becomes evident.

In spite of the fact that the avionics receiver adapted for in-orbit operation has demonstrated acceptable (or even good) performance, the experiment results still show that receivers for space applications may be subject to very specific effects unknown for on-ground applications. This seems to be a strong reason for conducting separate GNSS receiver development targeted to space applications.

The necessity of the development of the dynamic filtering methods for the positioning is shown. The orbit determination technique proposed provides accuracy better than the traditional snapshot solution in GPS/GLONASS mode.

It is also shown that a development of models of ionosphere and relativistic effects is necessary to significantly improve accuracy of the velocity and position determination.

Acknowledgments

The authors would like to acknowledge the invaluable assistance provided by Vladimir Branetz,, Mikhail Mikhailov and Juri Stishov of the RSC Energia.

References

[1] Bauer, F., Hartman, K., Lightsey, E., Spaceborne GPS Current Status and Future Vision. ION-GPS-98, Nashville, TN, September 1998.

[2] Felhauer, T., Tews, T., Botchkovski, A., Golubev, M., Vasilyev, M., "ASN-22 Combined GPS/GLONASS Receiver Module – Architecture, Technical Characteristics and Performance Analysis". ION-GPS-97, Kansas City, MI, September 1997, pp. 81-87

[3] Ortega, G., Mora, E., Carrascosa, C., "GPS Multipath Effects During the Shuttle to Mir Rendezvous for the STS-84 Flight Atlantis". ION-GPS-98, Nashville, TN, September 1998, pp. 1017-1024.

[4] Deines, S. D., Missing Relativity Terms in GPS Navigation: Journal of the Institute of Navigation, Vol.39, No. 1, Spring 1992.

NEW TECHNIQUE TO IMPROVE GPS RECEIVER PERFORMANCES BY ACQUISITION AND TRACKING THRESHOLDS REDUCTION

René Jr. LANDRY

ÉCOLE DE TECHNOLOGIE SUPÉRIEURE

1100 Notre-Dame West, Montreal, Quebec, Canada, H3C 1K3
Tel : +1 (514) 396-8506 - Fax : +1 (514) 396-8684
E-mail : Rene.Landry@ele.etsmtl.ca

KEY WORDS : Autonomous Velocity Aiding, TASAP, Threshold's Reduction, Robustness, GPS Receiver Processing.

SUMMARY

This paper proposes a robust method for threshold's reduction taking into account features both concerning GPS receiver modification and real gain on the performances improvement. This method involves two steps. The aim is to use the strong channels of the GPS receiver which are actually tracking satellites for velocity aiding the other channel trying to acquire or track satellites presenting a low signal over noise ratio due to lower elevation or masking conditions. Second, according to the theory and the characteristics of the digital internal loops of the GPS receiver, the predetection bandwidth is reduced to the lowest value permitted by the velocity aiding accuracy. This technique allows to improve the GPS accuracy and robustness. The paper shows first a large panorama of all potential threshold's reduction techniques both for acquisition and tracking processes. It proposes and identifies the automatic model of a velocity aided loop. Furthermore, to allow the validation of the described tracking threshold reduction, the technique is proposed to be inserted and validated into the new GPS MATNAV simulator [1], which is a generic digital MATLAB GPS receiver model. This work is intended to be used for space and aeronautical applications.

1. INTRODUCTION

The spaceborne GPS receivers are classically used in low earth orbit, with good GPS visibility conditions. However, some space missions require a GPS receiver operating with poor link budget. Such missions are for instance reentry capsule or shuttle (after radio black-out [9]), high altitude spacecraft, GPS attitude determination, degraded pointing modes, radio-occultations and interference environment scenario. These applications need quick reacquisition of the GPS satellites, GPS receivers build-in robustness technologies and better GPS visibility conditions than normal GPS navigation.

The GPS receivers used in aeronautical navigation need also acquisition and tracking threshold's reduction, for the following reasons :

- ✓ improvement of resistance to jammers, navigation availability and satellite's visibility.
- ✓ augmentation of accessible pseudorange measurements (improvement of RAIM).
- ✓ augmentation of accessible carrier phase measurements (improvement of phase tracking navigation).

The integrity is one of the major requirement for aeronautical mission. The objective here is to maximise the number of tracked SV (Satellite Vehicle) that can procure pseudo-measurements, even in the presence of interference. For a space mission where the C/N_0 may be low, we are more concerned with the number of satellites that could be tracked only. But, in orbit, the dynamic (acceleration) is always very small (for free orbital trajectories) and predictable. However, in these 2 domains of application, it is possible to reduce the threshold of GPS signal acquisition and tracking.

Also, in many situations, the acquisition process may be too long and/or the tracking loops of a GPS receiver may loose the lock of the signal during special conditions such as low satellite elevation angle or high dynamic manoeuvring. To improve the GPS acquisition time and the tracking performance an investigation is conducted on the use of additional internal and autonomous GPS velocity aiding information.

The main acquisition and tracking threshold reduction technique presented consist in supplying a pseudovelocity aiding to the carrier and/or the code loop, this pseudovelocity aiding is provided by the navigation filter itself (aeronautical PVT filter, or orbital Kalman filter, such as DIOGENE developed by CNES, where DIOGENE is « Détermination Immédiate d'Orbite par GPS et Navigateur Embarqué », Immediate Orbit Determination with GPS and OnBoard Navigator). The promising technique named TASAP, « Acquisition and Tracking Threshold Reduction Techniques » enables to

Paper presented at the RTO SCI International Conference on "Integrated Navigation Systems", held at "Elektropribor", St. Petersburg, Russia, 24-26 May 1999, and published in RTO MP-43.

reduce the C/N_0 thresholds below 20 dBHz, without any external aiding other than a tight coupling between an aeronautical or orbital navigator and the signal processing loops inside the GPS receiver.

OBJECTIVES AND METHODOLOGY

Many experiments are presently being carried out on the future DGPS-based approach and landing systems to improve the quality of aircraft navigation. The use of C/A-code receivers for aeronautical and spaceborne applications requires high reliability and integrity. Low visibility and satellite elevation angle during phases of flight can present problems for GPS reception of the channels presenting low C/N_0. The study of satellite missions and aeronautical environment show that GPS receiver must be provided with a strategy to reduce the acquisition and tracking thresholds to improve the navigation integrity, availability and performances, required for these missions.

This paper intends to give first an overview of all potential techniques well suited for Threshold's Reduction. It presents a status of the models studied to evaluate the Autonomous Velocity Aiding performances. Moreover, a Generic GPS Receiver has been implemented on the MATLAB software to validate the Threshold's Reduction Technique described in this paper. The aim of the Simulations is to compare the Velocity Aiding Performances with traditional processing.

The threshold reduction performance is evaluated versus the signal processing parameters of the receiver and the pseudovelocity aiding characteristics. Different models of this pseudovelocity aiding are presented, for the aeronautical and spaceborne applications. The main characteristics and theoretical performances of the aided processing loops are derived from these models. Numerical simulations of a GPS receiver provided with internal velocity aiding are also presented and compared with normal operation.

2. THRESHOLD'S REDUCTION TECHNIQUES PANORAMA

The following analysis brings some details of potential reduction techniques based on signal processing which can reduce the limitation of present day acquisition and tracking thresholds. The next section identifies and describes some of these techniques for GPS receivers.

2.1 Suppression of Data Demodulation

With a special design of servo loop, using for example a discriminator which can work without the data demodulation process, can bring 3dB improvement on the C/N_0 threshold.

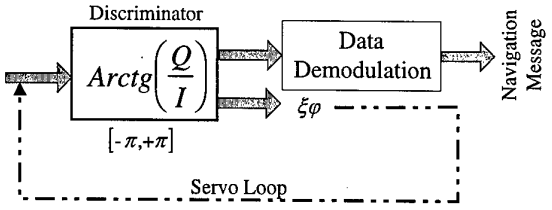


Figure 2-1 : Suppression of Data Demodulation.

The suppression of data demodulation in the loop can remove the C/N_0 limitation of 27dBHz which is necessary to obtain the BER better than 10^{-5}. This data demodulation can be done in parallel at the output of the arctg discriminator, for example, but outside the loop as shown in Figure 2-1.

The extension of the discriminator domain outside $]-k\pi,+k\pi[$ is not any more possible but since the steady state error is small for aeronautical and space missions, this restriction is not embarrassing.

2.2 The Data Wipping Variant 1, 2 and 3

Known as Data Wip, the Data Wipping consist of slightly reducing the predetection filter bandwidth while increasing the linearity of the loop at low SNR. This technique uses the a priori knowledge of the navigation message to reduce the predetection bandwidth B_{FI} of the carrier and code loops reducing at the same time the noise measurements. The carrier noise variance σ_{PLL} in this case is given by the following expression :

$$\sigma^2_{PLL} = \frac{\gamma_o \cdot B_L}{P_s} \cdot \left(1 + \frac{\gamma_o \cdot B_{FI}}{2 \cdot P_s}\right) [rad^2] \qquad \textbf{(2-1)}$$

γ_o : spectrum noise density (input thermal noise),
B_L : loop bandwidth,
B_{FI} : data pre-detection bandwidth,
P_s : power of the useful signal.

In the case of code loop, the code noise variance is :

$$\sigma^2_{DLL} = \frac{1}{2} \cdot \frac{\gamma_o \cdot B_L}{P_s} \cdot \left(1 + \frac{2 \cdot \gamma_o \cdot B_{FI}}{P_s}\right) [chip^2] \qquad \textbf{(2-2)}$$

The term $\frac{\gamma_o \cdot B_{FI}}{P_s}$ corresponds to the quadratic losses even though $\frac{P_s}{\gamma_o}$ represents the signal to noise ratio in 1Hz. It can be shown here that we can reduce the noise variance on measurement by reducing the predetection bandwidth B_{FI} to its minimum value. By this way, the quadratic losses are minimum and the loop will track at lower SNR.

2.2.1 First Variant of the Data Wipping

This technique uses a process which consist in memorising the I and Q samples on more than 20msec and removing the data modulation, depending of the received data demodulated in parallel, before the narrow loops of the GPS receiver.

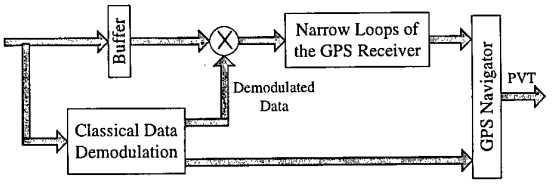


Figure 2-2 : Data Wipping ; Variant #1.

This technique introduces a certain delay in the GPS loops which is important to take into account during the phase correction filter process.

The main advantage of Data Wipping variant #1 is to accept stronger dynamic at same SNR. This variant may brings a good improvement to loop robustness having steady state problem.

2.2.2 Second Variant of the Data Wipping

This version allows to avoid the difficulties linked with the correction phase filter. It is based on the fact that the content of the navigation message is varying slowly and/or the main information can be predictable in a short term.

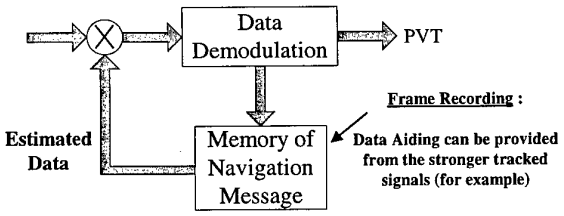


Figure 2-3: Data Wipping ; Variant #2.

The data demodulation is thus removed in real time with an error rate lower enough for cumulating the I and Q samples over 1 word, 600msec.

This variant supposes that one or several channels of the receiver are tracking GPS signal and that the reception quality is decreasing for only some of them.

2.2.3 Third Variant of the Data Wipping

Data Wipping Variant 3 is similar to Variant 2 but this version can be used in a very bad environment (low C/N_o) for some special applications. For a better understanding of its principle, we examine Figure 2-4.

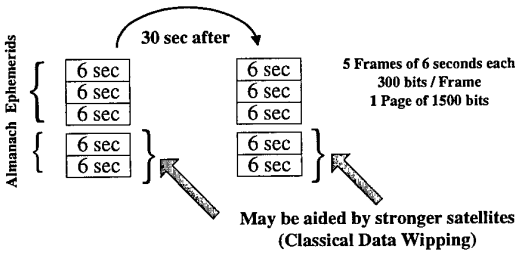


Figure 2-4: Navigation Message Frame for DW Variant 3.

Suppose that the link with a satellite is too bad to receive the data. The idea is to use the other strong channels to collect the almanach and ephemerid data of the weak channel. This can be useful for attitude control application knowing that position can be obtain with an accuracy of about ±500m using only almanach data.

2.3 External Velocity Aiding using INS

Relations (2-1) and (2-2) show that it is possible to reduce the threshold of the loop by reducing the prediction bandwidth. This is of course possible if an external velocity aiding is provided (Figure 2-5). Such a process is detailed in many papers (summary in [2]).

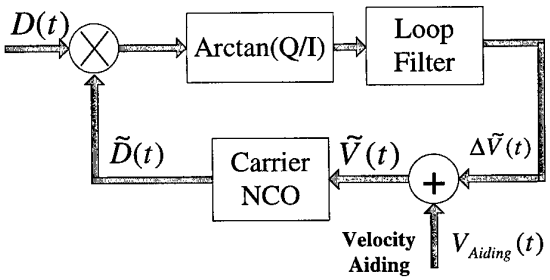


Figure 2-5: Loop with External Velocity Aiding.

This mode of operation is frequently used in military operation using integrated INS. This paper shows that this velocity aiding can be provided by the navigation filter itself (orbital or aeronautical).

3. ANALYSIS OF TRACKING LOOP PERFORMANCES

This section establishes the noise measurement equations of pseudorange and pseudovelocity that will be used to identify the code and carrier loop thresholds.

3.1 Noise Measurements

For the code loop with a programmable chip spacing, the standard deviation of the pseudorange noise measurement standard deviation (1σ) is approximated by :

$$\sigma_{PR}(m) = \frac{c}{R_c}\sqrt{\frac{B_{nm}\cdot C_s}{2\left(\frac{C}{N_o}\right)}\cdot\left[1+\frac{2B_{FI}}{\left(\frac{C}{N_o}\right)}\right]} \qquad \textbf{(3-1)}$$

$c = 3x10^8$ m/s ; speed of light,
R_c = 1.023Mchip/sec ; C/A code speed,
C_s = 1 chip ; chip spacing (1/2, 1/4, 1/8, etc),
B_{nm} = 1 Hz ; code loop bandwidth,
B_{FI} = 50 Hz ; data predetection filter bandwidth,
C/N_o: signal to noise spectral density power ratio.

Numerical Application :

B_{FI} = 50Hz, B_{nm} = 1Hz et C/N_o = 40dBHz,
we obtain, **σ_{PD} = 2meters**.

The standard deviation of the pseudovelocity error measured on the carrier is given by :

$$\sigma_{PV}(m/s) = \frac{c}{\sqrt{2}\pi f_t T_D}\sqrt{\frac{B_{np}}{\left(\frac{C}{N_o}\right)}\cdot\left[1+\frac{B_{FI}}{2\left(\frac{C}{N_o}\right)}\right]} \qquad \textbf{(3-2)}$$

f_t : transmitted frequency (F_{L1}, F_{L2} or others),
T_D = 1 sec ; Doppler integration time,
B_{np} = 5 Hz ; carrier loop filter bandwidth,

Numerical Application :

B_{FI} = 50Hz, B_{np} = 10Hz, C/N_o = 40dBHz et T_D = 0.6s,
we obtain, **σ_{PV} = 0.35cm/s**.

One other possibility to measure the standard deviation of the code pseudovelocity error is given by :

$$\sigma_{PV_{code}} = \frac{\sqrt{\sigma_{PR_k}^2 + \sigma_{PR_{k+1}}^2}}{t_{k+1} - t_k} = \frac{\sqrt{2}\cdot\sigma_{PR}}{\Delta t} \qquad \textbf{(3-3)}$$

with $\sigma_{PR_k} = \sigma_{PR_{k+1}} = \sigma_{PR}$, $t_{k+1} - t_k = \Delta t$
and t_i = date of pseudorange measurement, with instantaneous standard deviation σ_{PR_i}.

Numerical Application : $\sigma_{PVcode} = \frac{\sqrt{2}\cdot 2}{0.6} \cong 4m/s$

3.2 Code Loop Threshold on Tracking Mode

The code loop condition to stay in lock is given by :

$$a\cdot\sigma_{PR} \le \frac{\Delta}{2}(m) \qquad \textbf{(3-4)}$$

From equation (3-1), we obtain the following condition, with $C_s = 1$:

$$\frac{C}{N_o} \ge a^2 B_{nm}\cdot\left(1+\sqrt{1+\frac{4B_{FI}}{a^2\cdot B_{nm}}}\right) \qquad \textbf{(3-5)}$$

For B_{FI} = 50Hz, B_{nm} = 1Hz and a = 3, we have :

$$\left[\frac{C}{N_o}\right]_{Loop} \ge 17 dBHz \qquad \textbf{(3-6)}$$

For B_{FI} = 250Hz instead of 50Hz, we have :

$$\left[\frac{C}{N_o}\right]_{Loop} \ge 20 dBHz \qquad \textbf{(3-7)}$$

These values may be adjusted in function of the velocity aiding. Figure 3-1 shows the tracking threshold versus B_{FI} in function of B_{nm}.

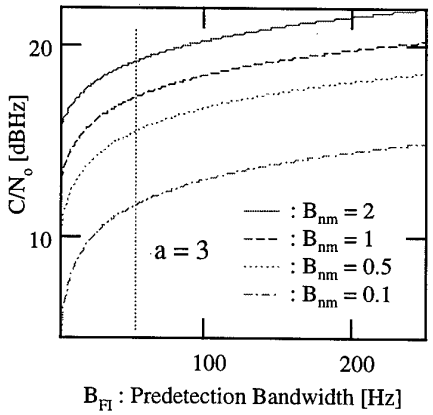


Figure 3-1: C/N_o versus B_{FI} (Variation of B_{nm}).

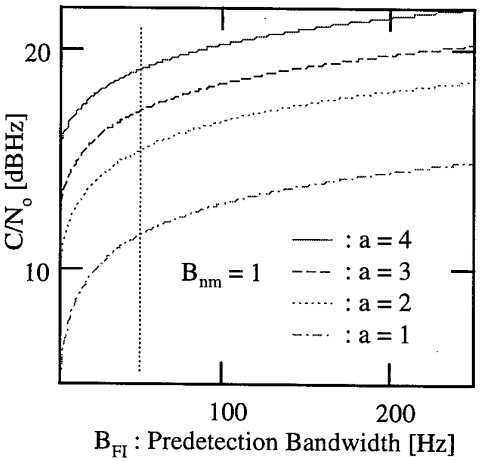


Figure 3-2: C/N_o versus B_{FI} (Variation of a).

From these figures, we see that the C/N_o threshold can be less than 20dBHz depending of the parameters B_{nm} and a.

3.3 Carrier Loop Threshold on Tracking Mode

The behaviour of the carrier loop and C/N_0 threshold are linked to the quality of data demodulation which depends on the required BPSK modulation BER of the following form :

$$BER = \frac{1}{2}\left[1 - erf\left(\sqrt{\frac{C/N_o}{D_r}}\right)\right] \qquad \textbf{(3-8)}$$

where D_r is the data rate (50Hz),
C/N_0 is the signal to noise spectral density power ratio, and the error function is given by :

$$erf(x) = \frac{2}{\sqrt{\pi}}\int_0^x e^{-u^2}\,du \qquad \textbf{(3-9)}$$

For example using (3-8), if it is not tolerate to have more than one bit of error each 30 minutes, with a data rate D_r =50bits/s, the maximum specified BER is 10^{-5} corresponding to a C/N_0 of 26.5dBHz.

Introducing τ, the time response of the loop ($1/\tau = B_L$), and assuming a loop bandwidth of 5Hz with the assumption that one bit error will occur every $3\tau \cdot B_L$ (limit of operation), we obtain a limit to the bit error probability of 6.67×10^{-3} (20ms/3τ·5). This corresponds to a $C/N_0 \geq 24$dBHz (3-8). Bellow this acquisition threshold, the instance of bit error cause the loop desynchronisation. Using an operational margin of 3dB explain the reason why the GPS receiver switch to signal reacquisition or Code-Only mode when C/N_0 is bellow 27dBHz. This C/N_0 level is the actual acquisition threshold for typical GPS receivers.

4. MATHEMATICAL MODEL OF AN AVIA LOOP

To understand more in details the behaviour of the GPS loops in presence of autonomous velocity aiding, a model of a tracking DLL aided by a velocity predetection coming from a local navigator is proposed hereafter.

4.1 Definition and Elaboration of the AVIA Model

Our simplified model of AVIA Loop (Internal and Autonomous Aiding Velocity), using direct modelisation of the velocity aiding errors is described hereafter.

The expression of the pseudorange measurement (PR) is generally written as :

$$PR_i^{Useful} = R_{r_i} + c \cdot \Delta T_i \qquad \textbf{(4-1)}$$

where c is the speed of light, R the radial distance between the GPS satellite and the mobile and ΔT the time difference between the receiver and satellite clock.

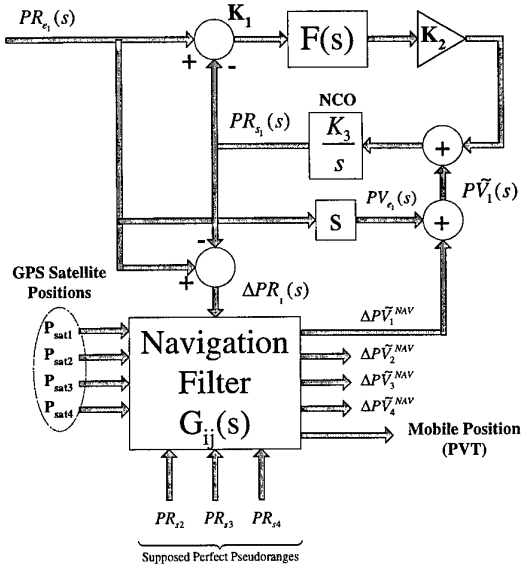


Figure 4-1: Modelling of the Loop with velocity aiding providing from GPS navigator (Autonomous Code ONLY).

After derivation, we obtain the expression of the real useful pseudovelocity (i.e. without error) :

$$PV_i^{Useful} = \frac{dPR_i^{Useful}}{dt} = \frac{dR_{r_i}}{dt} + c \cdot \frac{d\Delta T_i}{dt} = V_{r_i} + c \cdot \frac{d\Delta T_i}{dt} \qquad \textbf{(4-2)}$$

where V_r is the radial speed of the receiver versus the transmitter,
and $c \cdot d\left(\frac{\Delta T_i}{dt}\right)$ is the pseudovelocity due to the drifting shift of the clock between the receiver and the satellite i, noted also PV_i^{Clock}.

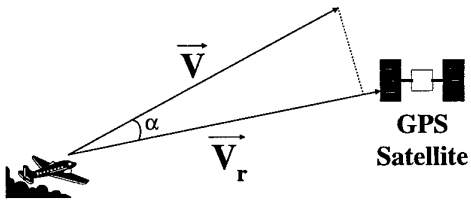


Figure 4-2: Mobile Radial Velocity vs GPS Satellite.

$$PV_i^{Useful} = V_{r_i} + PV_i^{Clock} \qquad \textbf{(4-3)}$$

The orbital navigator supplies an estimation $P\tilde{V}_i$ of the pseudovelocity corresponding to the aiding velocity noted :

$$PV_i^{Navigator} = P\tilde{V}_i^{Useful} = \tilde{V}_{r_i} + c \cdot d\left(\frac{\Delta\tilde{T}_i}{dt}\right) \qquad \textbf{(4-4)}$$

where $c \cdot d\left(\frac{\Delta \tilde{T}_i}{dt}\right)$ is the estimated clock pseudovelocity of satellite i noted also $P\tilde{V}_i^{Clock}$.

Moreover, we introduce the raw measurement from the GPS receiver with the notation $PV_i^{Observable}$ corresponding to the pseudovelocity which can effectively be observed by the loops of the receiver, perturbed by the global error sources represented by the notation $\Delta PV_i^{Observable}$.

$$PV_i^{Observable} = V_{r_i} + PV_i^{Clock} + \Delta PV_i^{Observable} \quad \textbf{(4-5)}$$

where $V_{r_i} + PV_i^{Clock} = PV_i^{Useful}$ is the real useful pseudovelocity, from the navigator estimation.

The term $\Delta PV_i^{Observable}$ represents notably the perturbations due to SA (Selective Availability), to ionosphere, to multipath and jammers.

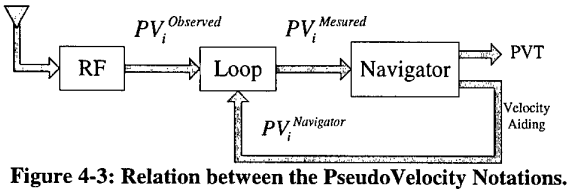


Figure 4-3: Relation between the PseudoVelocity Notations.

Reminding that the $PV_i^{Observable}$ is a characteristic of the received signal, it consists of physical pseudovelocity while $PV_i^{Measured}$ contains the steady state error of the loop such as the potential error reductions due to multipath coming from the loop and the loop thermal noise (N.B. : the measurement $PV_i^{Measured}$ is of course available only when the loop is in tracking mode).

Let's have a look on the GPS receiver PV_i visibility :

$$PV_i^{Navigator} = \tilde{V}_{r_i} + P\tilde{V}_i^{Clock} + \Delta PV_i^{Navigator} \quad \textbf{(4-6)}$$

where $\Delta PV_i^{Navigator}$ is the estimation error of the useful pseudovelocity. The error on the aiding velocity supplied by the navigator and seen by the loops of the GPS receiver is notified by ΔPV_i^{Loop}. We obtain :

$$\Delta PV_i^{Loop} = PV_i^{Observable} - PV_i^{Navigator} \quad \textbf{(4-7)}$$

after some developments, we obtain :

$$\Delta PV_i^{Loop} = V_{r_i} + PV_i^{Clock} + \Delta PV_i^{Observable} - (V_{r_i} + PV_i^{Clock} + \Delta PV_i^{Navigator}) \quad \textbf{(4-8)}$$

after simplification :

$$\Delta PV_i^{Loop} = \Delta PV_i^{Observable} - \Delta PV_i^{Navigator} \quad \textbf{(4-9)}$$

The $\Delta PV_i^{Observable}$ represents the physical perturbation of the GPS signal notably due to the SA.

Examples of $\Delta PV_i^{Navigator}$ are presented in Table 4-1. It is known that the orbit calculation performance of the Orbital Navigator such as DIOGENE are dependent on the type of orbit. Generally, we consider three types of orbit which are :

- Low Earth Orbit (LEO considered here at an altitude of 1000Km),
- Geostationary Orbit (GEO),
- Geostationary Transfer Orbit, after injection (GTO).

The performance of orbit calculation is estimated hereafter :

Type of Orbit	Clock Class	Precision
LEO	Short Term : $\Delta F/F=10^{-7}$	Position 100m (3σ), Speed 0.1m/s (1σ).
GEO	$\Delta F/F=10^{-9}$	Position 250m (3σ), Speed 0.015m/s (1σ).
GTO	$\Delta F/F=10^{-9}$	Position 500m (3σ), Speed 0.08m/s (1σ).

Table 4-1 : Performance of Navigator vs Type of Orbit.

4.2 Analysis of the Developed AVIA Loop Model

Our last analysis is summarised in Figure 4-4.

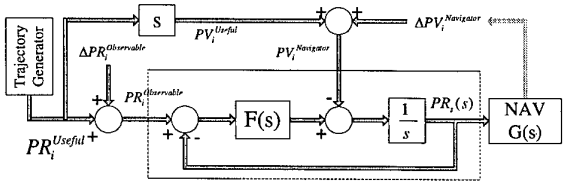


Figure 4-4: Mathematical Model of Autonomous Velocity Aiding Loop (AVIA Loop).

We would like to obtain the following transfer function :

$$H_1(s) = \frac{PR_s(s)}{PR_i^{Useful}(s)} \quad \textbf{(4-10)}$$

where:

$$PR_s(s) = \frac{\left[PR_e(s) - PR_s(s)\right] \cdot F(s)}{s} - \frac{PV_i^{Nav}(s)}{s} \quad \textbf{(4-11)}$$

$$PR_s(s) = \frac{PR_e(s)\cdot F(s)}{s} - \frac{F(s)\cdot PR_s(s)}{s} - \frac{PV_i^{Nav}(s)}{s} \quad \textbf{(4-12)}$$

This gives:

$$\left[1+\frac{F(s)}{s}\right]\cdot PR_s(s) = PR_e(s)\cdot\frac{F(s)}{s} - \frac{PV_i^{Nav}(s)}{s} \quad \textbf{(4-13)}$$

We have also :

$$PR_e(s) = \frac{\left[s\cdot PR_i^{Useful}(s) + \Delta PV_i^{Obs}(s)\right]}{s} \quad \textbf{(4-14)}$$

and :

$$PV_i^{Nav}(s) = s\cdot PR_i^{Useful}(s) + \Delta PV_i^{Nav}(s) \quad \textbf{(4-15)}$$

After some development, using (4-14) and (4-15), we obtain :

$$\left[1+\frac{F(s)}{s}+\frac{G_i(s)}{s}\right]\cdot PR_s(s) = \left[\frac{F(s)}{s}-1\right]\cdot PR_i^{Useful}(s) + \frac{F(s)\cdot\Delta PR_i^{Obs}(s)}{s} \quad \textbf{(4-16)}$$

Where:

$$G_i(s) = \frac{\Delta PV_i^{Navigator}(s)}{PR_s(s)} \quad \textbf{(4-17)}$$

is the transfer function of the Navigation Filter for one channel.

The function can be analysed for several type of sinusoidal perturbations, noted :

$$p_i(t) = b_i\cdot\sin(\omega_i t) \quad \textbf{(4-18)}$$

Assuming the following perturbations :

$$\Delta PV_i^{Nav}(s) = \frac{b_1\cdot\omega_1}{s^2+\omega_1^2} = P_1(s) \quad \textbf{(4-19)}$$

This perturbation can represent the theoretical navigator harmonic error, for instance.

$$\Delta PV_i^{Obs}(s) = \frac{b_2\cdot\omega_2}{s^2+\omega_2^2} = P_2(s) \quad \textbf{(4-20)}$$

The same sinusoid perturbation can represent the selective availability included in the observable pseudovelocity error.

Using $F(s) = \frac{K\cdot(s+a)}{s}$, and combining last equations with (4-16), we have :

$$\left[\frac{s^2+Ks+a}{s^2}\right]\cdot PR_s(s) = PR_i^{Useful}(s)\cdot\left[\frac{-s^2+Ks+a}{s^2}\right] + P(s) \quad \textbf{(4-21)}$$

where

$$P(s) = \frac{1}{s^2}\cdot\left[\frac{K\cdot(s+a)}{s}\cdot P_2(s) - s\cdot P_1(s)\right] \quad \textbf{(4-22)}$$

ie :

$$P(s) = \frac{1}{s^2}\cdot\left[\frac{K\cdot(s+a)}{s}\cdot\left(\frac{b_2\cdot\omega_2}{s^2+\omega_2^2}\right) - s\cdot\left(\frac{b_1\cdot\omega_1}{s^2+\omega_1^2}\right)\right] \quad \textbf{(4-23)}$$

The simulation of this model have been perform on the MATLAB GPS Receiver Simulator named MATNAV V2.1.

4.3 MatLab GPS Receiver Simulator

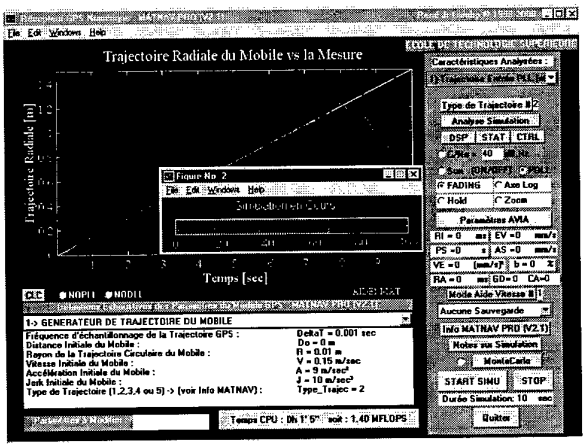


Figure 4-5: MATNAV V2.1 GPS MatLab Simulator.

The MatLab software simulator is fully described in [1]. The simulator includes 5 modules that are as close of reality as possible. The function and interaction of each module are shown in Figure 4-6.

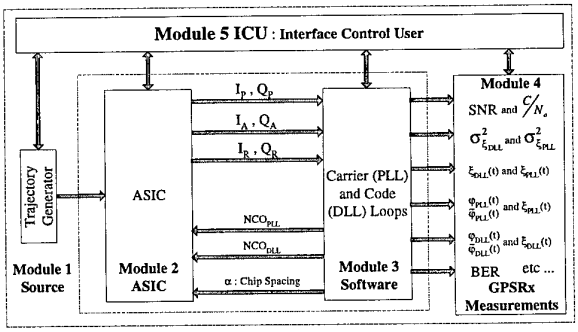


Figure 4-6: Software Modules in MATNAV.

This model was used to simulate a wide range of aeronautical and space scenarios. Figure 4-7 represents the autonomous aiding velocity that is used to push the PLL of the GPS receiver simulator.

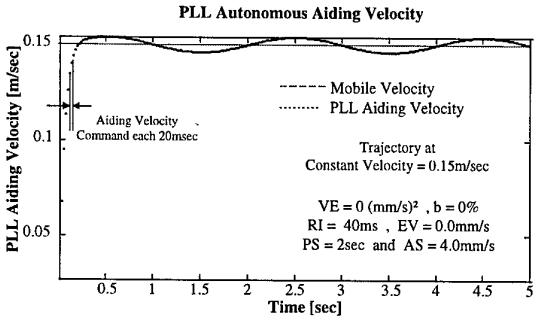


Figure 4-7: Aiding Velocity at each 20msec.

The main results are represented in Figure 4-8. The lower the velocity error (EV) is, the larger the C/N_0 measurement from the GPS receiver will be.

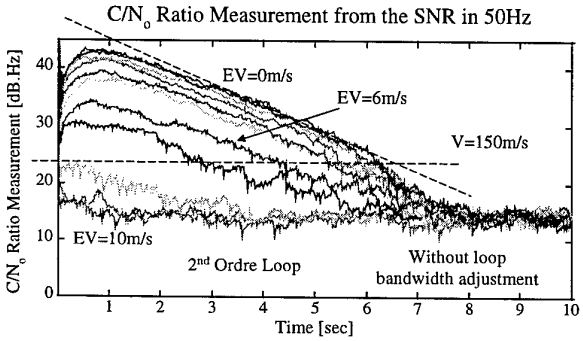


Figure 4-8: C/N_0 Measurements Scenario.

From a theoretical point of view, our simulation shows that the Code Only loss threshold are compatible. Figure 4-9 shows that when the velocity error is increasing up to 9m/sec, the Code Only loss is following the theoretical expectation. This result brings the conclusion that tracking threshold can be reduce as low as the predicted theory and this will depend upon the velocity error that can be achieved.

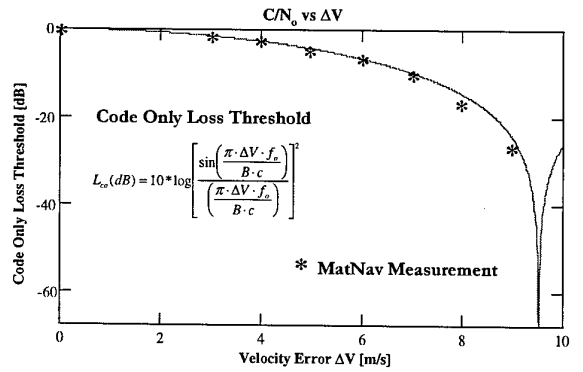


Figure 4-9: MATNAV Code ONLY Loss Measurement.

5. PERFORMANCES OF TASAP IN ACQUISITION MODE

It is not only necessary to reduce the tracking threshold but also the threshold during the phase of acquisition. This section intends to detail the process included in the TASAP Technique and to bring other innovative methods.

5.1 Synthesis & Improvement of Signal Acquisition

5.1.1 Cold Start or Classical Acquisition Mode

We remind that the acquisition thresholds are mainly linked to the value of the integration constant, τ, and to the chip sweeping speed α.

The classical acquisition process is a complete and continuous search for energy among the 1023 chip positions of a GPS C/A code (Figure 5-1).

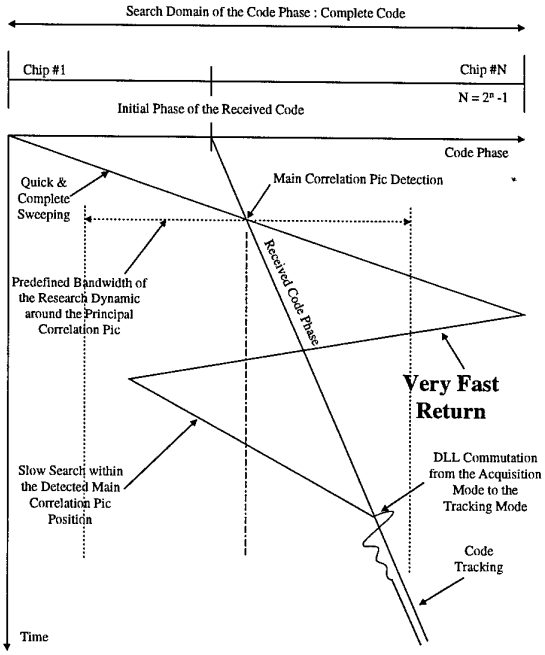


Figure 5-1: Cold Start or Classic Acquisition Process.

This concerns both cold start or classical aided acquisition. The difference between these 2 cases is mainly the number N_{db} of explored Doppler positions.

5.1.2 Aided Start with Code Prepositioning (Direct Acquisition)

The acquisition threshold performance can be improved in reducing the local code sweeping time. This is possible if the number of PN code chip positions to explore is reduced. The DLL local code has to be prepositionned within a predicted value of the received code phase (Figure 5-2).

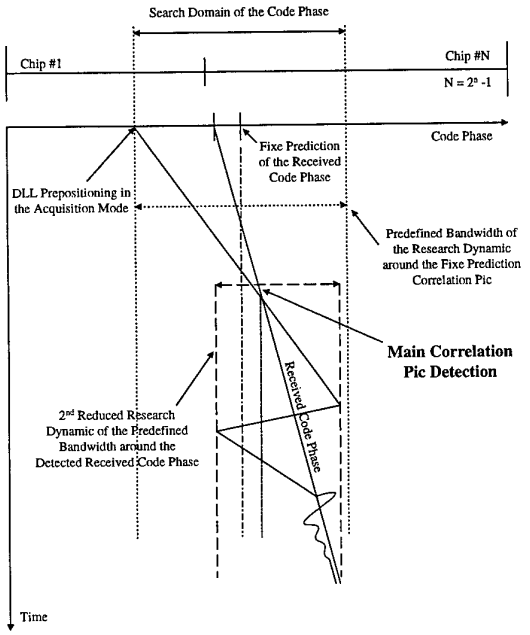


Figure 5-2 : Acquisition with Local Code Prepositioning.

Moreover, the acquisition performance can be still improved, if the local code is prepositionned in a shorter code phase range, pushed by a velocity aiding driving the code NCO (Figure 5-3). This velocity aiding is the same than for the « code only tracking » mode.

5.1.3 Aided Start with Code Prepositioning and Precise Velocity Aiding

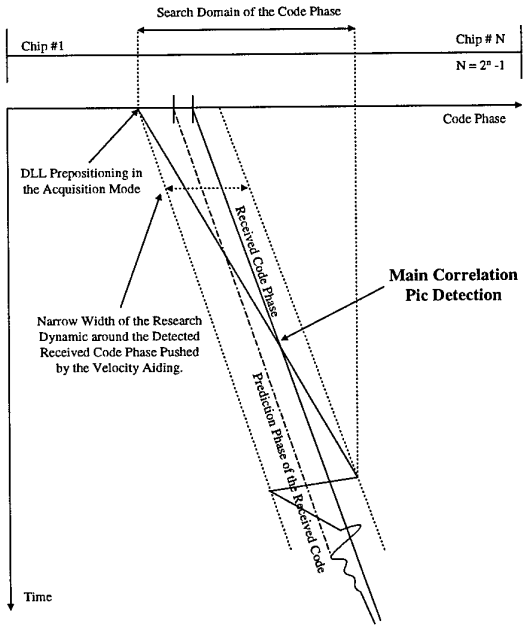


Figure 5-3 : Code ONLY Acquisition Mode.

5.1.4 CODE ONLY Performances during Acquisition

The « Code Only Acquisition » threshold $[C/N_o]_{co}$ is approximated using the following formula [6] :

(5-1)

$$\left[\frac{C}{N_o}\right]_{co} > \frac{2a}{f(N_{cb})\cdot\tau}\left(\sqrt{B_{FI}\cdot\tau\cdot f(N_{cb})}+a\right)\cdot L_{co}\cdot L_{ss}$$

B_{FI} : Pre-Detection Bandwidth,
B_{nm} : Noise Bandwidth (PR measures),
L_{co} : Code Only threshold losses due to pseudovelocity error ($L_{co}>1$),
L_{ss} : Losses due to the sweeping speed α of the DLL local code,
N_{cb} : Number of channels of the correlator,
$F(N_{cb})$: Function of the number of correlator branch ; $F(N_{cb})=N_{cb}$ or $f(N_{cb})=\sqrt{N_{cb}}$.
ΔPV : PseudoVelocity error, ΔPV_{Loop}.

with :

$$L_{co}(dB) = 10*\log\left[\frac{\sin\left(\frac{\pi\cdot\Delta PV\cdot f_o}{B_{FI}\cdot c}\right)}{\left(\frac{\pi\cdot\Delta PV\cdot f_o}{B_{FI}\cdot c}\right)}\right]^2 \quad \textbf{(5-2)}$$

For $L_{co} < 3dB$:

$$abs[\Delta PV] < \frac{0.443\cdot B_{FI}\cdot c}{f_o} \quad \textbf{(5-3)}$$

and

$$L_{ss} = 1-\frac{\alpha}{2}+\frac{\alpha^2}{12} \quad \textbf{(5-4)}$$

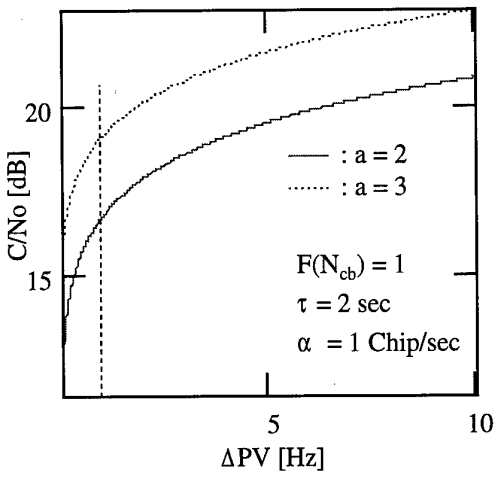


Figure 5-4: C/N_o vs Aiding Velocity Error.

The smallest the pseudovelocity error can be reach (from the Navigation Filter itself), the lower the C/N_o threshold will be during the acquisition process.

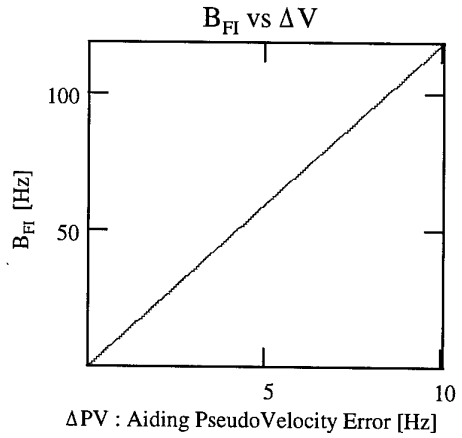


Figure 5-5: B_{FI} vs ΔV.

There is actually other investigations on the acquisition threshold's reduction for the carrier and code loops that are simulated within the MATNAV GPS Simulator. The next section concerns one of our current study.

5.2 Improving GPS Acquisition in Software

Figure 5-6 shows a typical carrier loop configuration where the I and Q samples from the ASIC at 1msec (1KHz predetection bandwidth) are integrated on 20msec.

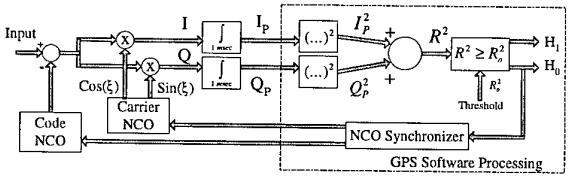


Figure 5-6 : Typical GPS Carrier Loop.

During the acquisition process, if no aiding velocity is available, one can process the I and Q channels to divide the complete predetection bandwidth in a smaller slot (Figure 5-7).

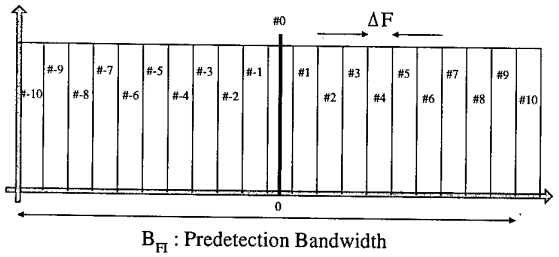


Figure 5-7 : Predetection Bandwidth Reduction.

This can be done at low rate using the following algorithm inside the GPS signal processor :

$$I_{P20}^{N} = I_P \cdot \cos(N \cdot \Delta f \cdot t) + I_P \cdot \sin(N \cdot \Delta f \cdot t)$$
$$Q_{P20}^{N} = Q_P \cdot \cos(N \cdot \Delta f \cdot t) + Q_P \cdot \sin(N \cdot \Delta f \cdot t) \quad \textbf{(5-5)}$$

N is all integer between -10 and +10.

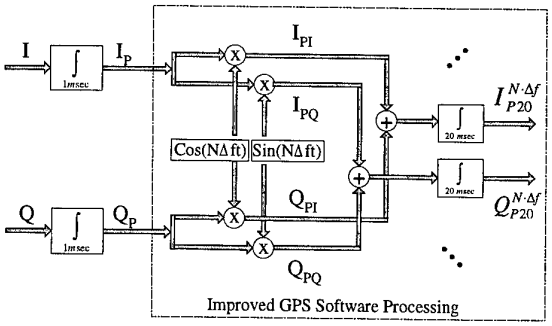


Figure 5-8 : Improved Acquisition Technique.

The final improved acquisition structure is shown in Figure 5-8.

6. GENERAL CONCLUSION

One of the main concerns with the use of GPS is the integrity and availability of the receiver being used. The paper intends to bring a solution to improve the performance of the receiver for the channels presenting low signal to noise ratio. Using a velocity aiding compensation which matches the code-loop dynamic characteristic, the steady state error performance on the measured pseudorange is improved. The existing external carrier-aided code loop is typically designed with loop-bandwidth B_L = 0.5Hz, and is thus less sensitive to velocity aiding errors. The results obtained show that sensibility to code-loop error is considerably reduced using internal velocity aiding. The limitations of the technique concern mainly the mission of the mobile.

Potential Threshold's Reduction Techniques	ADVANTAGES	DRAWBACKS
Suppression of Data Demodulation	2 or 3 dB of Threshold Reduction.	GPS Receiver Architecture Modification.
Data Wipping #1	Tracks High Dynamic Scenario	Needs SNR > 12dB and Phase Correction Filters.
Data Wipping #2	No Phase Correction.	Limited Applications
Data Wipping #3	Attitude Control	Never used.
External Velocity Aiding using INS	Tracking High Dynamic.	Expensive No Space Application.
TASAP	Autonomous	Needs Precise Velocity Aiding

Table 6-1 : Summary of Potential Threshold's Reduction.

Table 6-1 shows a summary of all potential threshold's reduction techniques for aeronautical and space applications.

Some applications are actually requiring precise measurements and large integrity. By reducing the acquisition and tracking thresholds the GPS signals will allow to increase the range of potential applications.

The paper intends to bring some elements concerning a solution to improve the performance of the receiver for the channels presenting signals with low SNR. The status of an autonomous code only processing modelling has been presented. More simulation results will be presented in a future paper.

ACKNOWLEDGEMENTS

The author would like to give special thanks to Patrice Guillard for his support during the phase of the MatLab GPS simulator design, to Abdelahad Benhallam from Ecole Nationale d'Aviation Civile (ENAC), to M.Jean-Luc Issler from the CNES and M.Jean-Louis Jonquière from Technical Service for Aircraft Navigation (STNA), for all their interests in this work. This study was made possible in part by a research grant from the STNA, and by a PostDoctoral grant from CNES, in association with ENAC and the technical support of SEXTANT Avionique.

BIOGRAPHY

René Jr. Landry was born in Montreal, Quebec, Canada in 1968. After graduating (B.Sc.) from the Ecole Polytechnique of Montreal (Canada) in 92, he got the M.Sc. in Satellite Communication Engineering at the University of Surrey (UK) in 93. He received a Diplôme d'Etudes Approfondies (DEA) in Microwave Engineering with a Mastere in Space Electronics in 94, and his Ph.D. degree in Signal Processing in 97 at the Ecole Nationale Supérieure de l'Aéronautique et de l'Espace, ENSAE (Toulouse, France). His dissertation topic concerned anti-jamming technologies applied to the problem of civil aircraft interference using GPS receivers as a primary equipment of navigation. He has been involved in the design of the Amplitude Domain Processing (ADP) Filter and the patented PIRANHA Filter integrated within 2 GPS receivers as a new anti-jamming technology. In 1998, he was a Post-Doc Research Staff in the RadioNavigation Department of the French Space Agengy, CNES (Toulouse, France) where his major interest was concerning signal processing for digital receiver technology. M.Landry holds a Professor position in the Department of Electrical Engineering at the University of Ecole de Technologie Supérieure, in Montreal.

REFERENCES

[1] R.Jr.Landry, « ***Modélisation d'un Récepteur GPS sur MATLAB : MATNAV V2.1*** », Post-Doc Report at CNES, Version 2.1, Toulouse, 5th September 1998.

[2] R.Jr.Landry, ***«Techniques d'Abaissement des Seuils d'Acquisition et de Poursuite pour les Récepteurs GPS - Présentation de la Technique TASAP - »***, Post-Doc Report at CNES, Version 2.1, Toulouse, 20th December 1998.

[3] R.Jr. Landry, ***«Interference Mitigation Techniques for GPS Receivers»***, PhD Dissertation, European Label, SUPAERO - ONERA/CERT, Toulouse, France, 16th December 97.

[4] R.Jr.Landry, ***«Modélisation Générique d'un Récepteur GPS sur COSSAP»***, ONERA-CERT / SUPAERO Rapport, #Sup5b, March 1996.

[5] R.W.Caroll and W.A.Mickelson, ***«Velocity Aided of Non-Coherent GPS Receiver»***, Proceedings of the IEEE 1977 National Aerospace and Electronics Conference, NAECON, 1977.

[6] J.-L. Issler, J.Fourcade, L.Lestarquit, C.Mehlen, G.Garnier, ***«High Reduction of Acquisition and Tracking Thresholds of GPS Spaceborne Receivers»***, ION GPS 98, 19th January 1998.

[7] Hugh Wightwick and Robert Wade, ***«The Design, Implementation and Performance of a GPS Resident Close Coupled GPS-INS Integration Filter»***, Proceedings of National Technical Meeting of Institute of Navigation, Anaheim, California, pp.673-681, USA, 1995.

[8] He Xiufeng and Chen Yongqi, ***«Analysis of the Performance of a GPS Receiver with Inertial Velocity Aided»***, Navigation 2000, National Technical Meeting, Long Beach, California, 21-23th January, 1998.

[9] Gérard Bréard, ***«Atmospheric Reentry Demonstrator : Spécification des Essais de Synthèse du Système Electrique et Logiciel sur l'Exemplaire de Vol»***, AeroSpatiale Espace & Défense, Centre Opérationnel Systèmes, Draft Report, Réf. HT-ST-B-11-AS/M, 13th January 1998.

[10] W.A.Gardner, ***«The Role of Spectral Correlation in Design and Performance Analysis of Synchronizers»***, IEEE Transactions on Communications, COM-34, No.11, pp.1089-1095, November, 1986.

[11] William M.Lear & Moises N.Montez, ***«The Effect of Selective Availability on Orbit Space Vehicules Equipped with SPS GPS Receivers»***, Charles Stark Draper Laboratory & NASA Johnson Space Center, pp.825-840.

[12] William M.Lear, ***«Range Bias Models for GPS Navigation Filters»***, NASA Report N°. JSC-25857, Lyndon B. Johnson Space Center, Houston, Texax, June 93.

[13] J.Potti, P.Bernedo & A.Pasetti, ***«Applicability of GPS-Based Orbit Determination Systems to a Wide Range of HEO Missions»***, ION GPS, 12-15th September 95.

[14] M.Serizawa & H.Suzuki, ***«An Adaptive Carrier Tracking Loop for Digital Satellite Mobile Communications»***, IEEE, CH2424-0/87/0000-1754, 1987.

Analysis of Tracking Performance of a Delay Locked Loop for Newly Proposed GPS Signal Waveforms

Alexandr Draganov and James Stafford[1]
ORBCOMM Global
21700 Atlantic Blvd
Dulles, VA 20166, USA

Abstract.

Recently, several new GPS waveforms have been suggested to enhance the GPS signal [1]. This paper presents a comparative theoretical analysis of Delay Locked Loop (DLL) tracking performance for different waveforms in the presence of the ambient white noise, and band-limited, shaped spectrum noise.

Incoherent (power) DLL is selected as a baseline design. For the purposes of comparative analysis, gain is treated as an arbitrary parameter and is selected to provide desired (and uniform) dynamic tracking capabilities of the DLL for different waveforms. The dynamic differential equation for the code phase being tracked by the DLL is examined by a means of the Fokker-Planck formulation. The Fokker-Planck equation is a partial differential equation describing the evolution of statistical characteristics of the tracking error. Coefficients for the Fokker-Planck equation are derived analytically for all waveforms under consideration. The tracking performance is linked to eigenvalues and eigenvectors of the Sturm-Liouvile problem for the Fokker-Planck equation. Eigenvalues and eigenvectors are found numerically yielding two major results: the average time to lose lock and the root mean squared (RMS) tracking error. Results for different waveforms and different signal to noise (SNR) ratios are presented. They show superior tracking capabilities for P(Y) and C/A signals as compared to more complicated subcarrier modulated waveforms, if the latter are tracked using a typical "early squared minus late squared" DLL. Tracking of subcarrier modulated waveforms can be improved substantially if the subcarrier is tracked separately, using a PLL-type loop.

In addition to tracking signals in the presence of the Gaussian white noise, the paper analyses tracking in the presence of band-limited, shaped spectrum noise (including possible interference from other GPS waveforms) and tracking in the presence of a tone jammer. The quantitative effect of a band-limited noise and of a tone jammer is mapped to a spectral density of a white Gaussian noise which would cause the same tracking degradation as the band-limited noise and/or jammer under consideration. Thus, for different types of interference, an equivalent white Gaussian noise density is determined, and tracking characteristics can be determined using results previously obtained for the white noise case.

Results may serve as a basis for estimating SNR thresholds for tracking different GPS waveforms, including cases of white noise, signal-to-signal interference, and band-limited and tone jamming.

1. Introduction

In this paper, tracking performance of a DLL is analyzed in application to newly suggested GPS signal waveforms. The objective of this analysis is to determine statistical characteristics of tracking, such as the root mean squared (RMS) tracking error and the mean time to lose lock. The tracking performance of a DLL is affected by two separate opposite phenomena, namely the discriminator signal which pulls the estimated code phase of the signal in accordance with the code phase of the received signal, and noise, which is the cause of the tracking error and of the loss of lock. Mathematically, tracking is described by dynamic stochastic differential equations with a noise term.

There are several ways of solving dynamic equations for DLLs, ranging from simple analytical formulas to full-scale numerical simulations. Relative merits and disadvantages of these approaches are as follows.

Canned approximate analytical solutions are available for some traditional waveforms in the presence of white noise. However, such analytical solutions must be rederived if new waveforms and new DLL discriminator functions are being used. Derivation could become complicated for complex discriminator characteristics, as in the case of signal waveforms with subcarrier modulation. Approximations which are required to get analytical formulas, may decrease the fidelity and accuracy of the final result.

On the opposite end of the spectrum of available approaches, there are numerical simulations. Numerical simulations may potentially provide accurate assessment of tracking performance, and are less subject to

[1] James Stafford is with Stanford Telecommunications, Inc. This paper was written when both authors were with Stanford Telecommunications, Inc

Paper presented at the RTO SCI International Conference on "Integrated Navigation Systems", held at "Elektropribor", St. Petersburg, Russia, 24-26 May 1999, and published in RTO MP-43.

limitations of analytical solutions. To obtain statistical (mean) values, simulations must be run multiple times, for different realizations of the random noise process (Monte-Carlo simulations). Moreover, to estimate time to lose lock, dynamic equations must be solved over long time intervals. Hence, numerical simulations may become very computationally intensive.

A third possible approach is to use Fokker-Planck equation to obtain statistical characteristics of tracking performance. The Fokker-Planck technique has been widely used to analyze DLL and PLL tracking in various applications. In the Fokker-Planck equation, statistical characteristics are determined directly, so that there is no need for Monte-Carlo simulations. On the other hand, Fokker-Planck equation is directly derived from original dynamic tracking equation, thus eliminating the need for some simplifying approximations typical for analytical formulas. For the purposes of tracking analysis for different waveforms, Fokker-Planck equation appears to be the optimal compromise. It is less restrictive than some analytical techniques, and less computationally intensive than Monte-Carlo simulations. In this paper, the Fokker-Planck equation is used as a primary mathematical tool for analysis of tracking. The advantages of Fokker-Planck analysis come at a relatively low price of one key assumption: the noise term in the stochastic equation must have short autocorrelation time as compared to the system time scale. This assumption proves to be valid for a majority of analyzed cases. Two exceptions (when the Fokker-Plamck equation may be invalid) are noted in the text below.

2. GPS Frequency Reuse

Recently, several new GPS waveforms have been been proposed to be added to the next generation of GPS satellites. The chief motivations for to suggest new GPS waveforms this effort have been

- to increase the security of the military code,
- the capability to deny adversaries in a war zone the use of the civilian accessible code,
- to provide a second frequency for civilian use to allow dual-frequency ionospheric correction of measured ranges.

The primary criteria that must be met by the new signals is that of backward compatibility. That is, the new signal must share the same frequency band as the existing signal without interfering with it to the extent of degrading current performance. The usual means of accomplishing this has been locate the bulk of the power of the new signals near the nulls of the existing signals.

Also, the new signals must also be resistant to jamming. There are two types of jamming to be considered. The first would be by an adversary trying to degrade the navigation performance of the new signal, and the other by a friend employing whatever means is proposed to jam the civilian signals in-theater. Some of these jamming proposals are well-placed tones and others are wideband jamming signals.

Another consideration that does not specifically relate to signal structure, but bears on tracking performance is the means by which the new signals are combined with the existing ones. The simplest method may be to simply provide an additional antenna onboard the spacecraft to broadcast the new signals.

There are drawbacks to adding another antenna to the spacecraft and other single-aperture methods have been proposed. It was desirable to combine the signals such that the resultant composite signal was of constant power envelope. This enabled the use of non-linear amplifiers and better power efficiency. At least three methods were examined, among them were hard-limiting the linear combination of the signals, weighted majority-voting among the component signals, and interlacing, or time-division multiplexing the signals.

All of these approaches result in a certain amount of power inefficiency, and constrain the power allocation between signals. Effectively, each of the component signals is broadcast at a certain power along with an undesired intermodulation signal. The interference of this intermodulation signal to the desired signals must be examined.

3. Proposed Signal Structures

3.1 Binary Offset Carrier Signals

Most of the proposed signals that were examined belong to a class of signals called Binary Offset Carrier (BOC) signals. A BOC signal is specified by the chipping rate of its PRN code as well as the frequency of a sub-modulating square-wave, in units of MHz. Basically, this technique creates a null in the center of the spectrum, the higher the frequency of the sub-modulating square wave, the wider the null. The existing GPS signals also belong to this class of signals. For instance, the C/A signal is a BOC(0,1) signal and the the P code is a BOC(0,10) signal. Of particular interest for GPS frequency reuse applications were a Manchester chip signal: BOC(5,5), a split-C/A signal: BOC(10,1), and split military signals: BOC(8,4), BOC(9,2), and BOC (10,2). In addition, there was another signal that does not belong to the BOC class proposed by James Spilker of Stanford Telecommuncations, Inc, Sunnyvale, CA. This signal consisted of a 1 MHz chipping rate long-cycle PRN submodulated with a 10 MHz chipping rate 10-chip PRN.

3.2 Autocorrelation and Spectral Properties of Proposed Signal Structures

In order to assess the tracking performance potential of these proposed signals, their autocorrelation and spectral properties must be understood. Since all of these proposed signals are at baseband essentially a sequence of ±1s, their autocorrelations are piecewise linear between transition boundaries. Thus, the autocorrelations can be expressed as the linear superposition of symmetric "triangle" functions whose Fourier transforms are easy to compute. By this means several of the autocorrelations and corresponding power spectra of the proposed signals were computed for use in the tracking performance analysis.

4. Dynamic Equations

The delay locked loop is designed to track code delay of the input signal using autocorrelation properties of the signal. The output of the delay locked loop is the estimated delay of the input signal. In a typical application (so-called coherent DLL), a numerically controlled oscillator generates an "early-minus-late" code $S_0\left(t-\hat{\theta}\right)$ (or a similar code with suitable correlation properties) delayed by the estimated delay $\hat{\theta}$. The input signal is correlated with this code. The result of this correlation may have different polarity depending on whether the estimated delay is less or greater than the input signal delay. It is used to control the rate of the numerically controlled oscillator, thus adjusting the estimated delay. Some of the formulas below are applicable to the coherent DLL case only; however their generalization to other DLLs (e.g., power DLL and dot-product DLL) is straightforward. Coefficients of the Fokker-Planck equation are derived for all three cases.

Let us assume that the input signal contains code $S(t-\theta)$ and noise $N(t)$, and the numerically controlled oscillator generates the waveform $S_0\left(t-\hat{\theta}\right)$, where $\theta,\hat{\theta}$ is the signal delay and estimated delay, respectively. Then the change in estimated delay $\hat{\theta}$ is described by the equation:

$$\frac{d\hat{\theta}}{dt}=\frac{1}{T_0}\int_{t-T_0}^{t}\left(S(\tau-\theta)+N(\tau)\right)GS_0\left(\tau-\hat{\theta}\right)d\tau \quad (1)$$

where T_0 is the DLL integration time, and G is the gain.

To analyze statistics of solutions of equation (1), it is transformed it to a stochastic differential equation in the form which is a starting point for deriving the Fokker-Planck equation. Namely, it is shown that temporal evolution of the tracking error is described by the following:

$$\frac{d\varepsilon}{dt}=g(\varepsilon(t))+F(\varepsilon(t))n(t) \quad (2)$$

where $n(t)$ is random variable normalized to have its autocorrelation function to be similar to Dirac delta-function. The next section derives formulas for coefficients $g(\varepsilon), F(\varepsilon)$.

5. Coefficients of the Dynamic Equations

The tracking loop performance is analyzed in the static case, when the signal delay θ is constant. Then, (1) can be rewritten as:

$$\frac{d\varepsilon}{dt}=-\frac{1}{T_0}\int_{t-T_0}^{t}\left(S(\tau-\theta)+N(\tau)\right)GS_0\left(\tau-\hat{\theta}\right)d\tau \quad (3)$$

where $\varepsilon=\theta-\hat{\theta}$ is the tracking error.

The integral from the product of the received signal $S(t)$ and code $S_0(t)$ yields their cross-correlation function $C(\varepsilon)$:

$$\frac{d\varepsilon}{dt}=-GC(\varepsilon)-\frac{G}{T_0}\int_{t-T_0}^{t}N(\tau)S_0\left(\tau-\hat{\theta}\right)d\tau \quad (4)$$

Note that one can use the cross-correlation function $C(\varepsilon)$ only if the integration interval T_0 is long enough. Rigorously speaking, integral of the product of the signal $S(t)$ and code $S_0(t)$ can be expressed via their cross-correlation function plus a random variable. The latter will contribute to the random (noise-related) term. However, for typical DLL integration times, a random contribution from signal-code correlation can be neglected. Similarly, random contribution for code correlation with other codes can also be neglected.

Cross-correlation function $C(\varepsilon)$ is often referred to as the discriminator function. Comparison of (4) and (2) shows that $-GC(\varepsilon)$ corresponds to the $g(\varepsilon(t))$ term, and the integral in (4) should correspond to the $F(\varepsilon(t))n(t)$ term.

The integral in (4) is a random variable. This variable must be normalized so that its autocorrelation function resembles Dirac delta function. Equation (4) can be rewritten in the form

$$\frac{d\varepsilon}{dt} = -GC(\varepsilon) - \frac{GN_0}{T_0}v(t) \qquad (5)$$

where

$$v(t) = \frac{1}{N_0}\int_{t-T_0}^{t} N(\tau)S_0\left(\tau - \hat{\theta}\right)d\tau \qquad (6)$$

and N_0 is a normalization constant, selected to have the autocorrelation function of $v(t)$ being similar to Dirac delta function. Expressions for N_0 for three different types of discriminator are derived in the Appendix.

Substitution of (A8) in (5) and comparison with (2) yields the following expressions for coefficients of the stochastic equation for the coherent DLL:

$$g(\varepsilon) = -GC(\varepsilon)$$
$$F(\varepsilon) = GP\sqrt{\frac{2dN_\omega}{T_c}} \qquad (7)$$

It is convenient to express all quantities via the noise spectral density and the signal-to-noise ratio. Formulas

$$R(\varepsilon) = \begin{cases} 1 - \dfrac{|\varepsilon|}{T_c} & \text{for } |\varepsilon| < T_c \\ 0 & \text{for } |\varepsilon| < T_c \end{cases}$$

is the signal autocorrelation function. The most surprising and counter-intuitive feature of the expressions for the coherent DLL is that the magnitude of the noise-related term does not depend on the DLL integration interval T_0. Yet, this result is correct and is in agreement with the discussion of "Physical" approach to non-gaussian processes in [2].

Similarly, for the power DLL:

$$g(\varepsilon) = -G_0'(S/N)C_0'(\varepsilon)$$
$$F'(\varepsilon) = \frac{G_0'}{T_c}\sqrt{\frac{2}{3}\frac{d(T_c - d)}{T_0}} \qquad (9)$$

where $G_0' = GN_\omega P^2$ and
$C_0'(\varepsilon) = R^2(\varepsilon - d) - R^2(\varepsilon + d)$.

If the signal has subcarrier modulation (the case of Manchester, binary offset carrier signals, or split C/A signal) formulas for the power DLL take the following form:

$$g(\varepsilon) = -G_0'(S/N)C_0'(\varepsilon)$$
$$F'(\varepsilon) = G_0'\sqrt{\frac{2}{3T_0}\left(\left(\frac{1}{J} - \frac{2d}{T_c}\right)(J-1) + \left(\frac{1}{J} - \frac{d}{T_c}\right)\right)\left(\frac{2d}{T_c}(J-1) + \frac{d}{T_c}\right)} \qquad (10)$$

for $g(\varepsilon)$, $F(\varepsilon)$ take the following form:

$$g(\varepsilon) = -G_0 C_0(\varepsilon)\sqrt{(S/N)}$$
$$F(\varepsilon) = G_0\sqrt{\frac{2d}{T_c}} \qquad (8)$$

where $G_0 = GP\sqrt{N_\omega}$ is normalized gain, $C_0(\varepsilon) = R(\varepsilon - d) - R(\varepsilon + d)$ is the signal-code cross-correlation function for unit amplitudes of both the signal and the code, and

where J is the number of subcarrier chips in the main code chip (i.e., J=2 for the Manchester code, etc.). Note, that in the limit case J=1 (no subcarrier modulation) formula (10) reduces to (9).

Finally, for the dot-product DLL:

$$g(\varepsilon) = -G_0'(S/N)C_0''(\varepsilon)$$
$$F''(\varepsilon) = G_0'\sqrt{\frac{2}{3}\frac{d}{T_c T_0}} \qquad (11)$$

where $G_0' = GN_\omega P^2$ is the same as for the power discriminator, and

$$C_0''(\varepsilon) = \big(R(\varepsilon - d) - R(\varepsilon + d)\big)R(\varepsilon).$$

The above formulas assume infinite bandwidth of the signal and noise and continuous integration. Hence, pulse shape of the signal and of the code is assumed to be rectangular. In the case of signals with large bandwidth (particularly signals with high-rate subcarrier modulation), tracking will be affected by the filtering.

As it is shown below, traditional way of code tracking of some of signals results in a decreased performance. One possible way to improve the tracking performance of signals with subcarrier modulation is to use code tracking to resolve the ambiguity of the subcarrier, and then use a phase locked loop to track the subcarrier phase. Fokker-Planck formulation can be readily applied to analyze performance of subcarrier tracking. If the subcarrier modulation is nearly sinusoidal due to filtering, equations for the coefficients of the stochastic equation are in the form (power discriminator PLL):

$$\begin{aligned} g(\varepsilon) &= -G_0'(S/N)C_0'(\varepsilon) \\ F'(\varepsilon) &= 2G_0' \sin\left(4\pi\frac{Jd}{T_c}\right)\sqrt{\frac{2}{3T_0}} \end{aligned} \qquad (12)$$

where

$$\begin{aligned} C_0'(\varepsilon) &= R_s^2(\varepsilon - d) - R_s^2(\varepsilon + d), \\ R_s^2 &= \frac{1}{2}\cos\left(2\pi\frac{Jd}{T_c}\right) \end{aligned}$$

The code magnitude P and the noise spectral density N_ω are affecting solutions of the stochastic equation always in a combination with DLL gain G. Since the latter is an arbitrary parameter, it can be chosen to optimize the DLL performance for any values of P and N_ω. If P or N_ω have different values in two different simulations, values of G should also be selected different to ensure the same optimum normalized gain G_0. The numerical analysis below will not use individual values of P, N_ω and G; rather, all calculations will be performed for some values of normalized gain G_0.

6. The Fokker-Planck Equation

Statistical approach for analysis of locked loop performance is described in detail in [2]. The density distribution function $f(\varepsilon,t)$ of the tracking error ε can be analyzed as a function of time t.

According to [2], density distribution function $f(\varepsilon,t)$ obeys the Fokker-Planck equation:

$$\begin{aligned} &\frac{\partial f(\varepsilon,t)}{\partial t} + \frac{\partial}{\partial \varepsilon}[K_1(\varepsilon)f(\varepsilon,t)] - \\ &\frac{1}{2}\frac{\partial^2}{\partial \varepsilon^2}[K_2(\varepsilon)f(\varepsilon,t)] = 0 \end{aligned} \qquad (13)$$

where

$$\begin{aligned} K_1(\varepsilon) &= g(\varepsilon) + \frac{1}{2}\frac{dF(\varepsilon)}{d\varepsilon}F(\varepsilon) \\ K_2(\varepsilon) &= F^2(\varepsilon) \end{aligned} \qquad (14)$$

and $g(\varepsilon)$, $F(\varepsilon)$ are defined above. Substitution of (8)-(11) in (13) yields for coherent DLL:

$$\begin{aligned} K_1(\varepsilon) &= G_0 C_0(\varepsilon)\sqrt{S/N} \\ K_2 &= G_0^2\frac{2d}{T_c} \end{aligned} \qquad (15)$$

The same formula for the dot-product DLL is in the form:

$$\begin{aligned} K_1'(\varepsilon) &= G_0' C_0'(\varepsilon)(S/N) \\ K_2' &= G_0'^2\frac{2}{3T_0}\left(\left(\frac{1}{J} - \frac{2d}{T_c}\right)(J-1) + \left(\frac{1}{J} - \frac{d}{T_c}\right)\right)\left(\frac{2d}{T_c}(J-1) + \frac{d}{T_c}\right) \end{aligned} \qquad (16)$$

Finally, for the dot-product discriminator DLL,

$$K_1''(\varepsilon) = G_0'' C_0''(\varepsilon)(S/N)$$
$$K_2'' = G_0''^2 \frac{2d}{3T_0 T_c} \quad (17)$$

The next section describes the solution of the Fokker-Planck equation.

7. Solution of the Fokker-Planck Equation

7.1 Analytical Treatment

The solution of the Fokker-Planck equation must be found in the interval $[\varepsilon_1, \varepsilon_2]$ where $\varepsilon_1, \varepsilon_2$ are zeroes of the discriminator function. The boundary conditions are as following:

$$f(\varepsilon_1, t) = f(\varepsilon_2, t) = 0 \quad (18)$$

The solution is sought using the separation of variables method:

$$f(\varepsilon, t) = e^{-\lambda t}\phi(\varepsilon) \quad (19)$$

Substitution of (19) in (13) yields

$$\frac{d}{d\varepsilon}[K_1(\varepsilon)\phi(\varepsilon)] - \frac{1}{2}\frac{d^2}{d\varepsilon^2}[K_2(\varepsilon)\phi(\varepsilon)] = \lambda\phi(\varepsilon) \quad (20)$$

Solution of this ordinary differential equation is subject to boundary conditions (18). Equation (20) is an eigenvalue problem for an operator in the left-hand side. According to a theorem by Lindsey and Meyr [see 2], there is infinite number of distinct eigenvalues, and all eigenvalues are positive. For each eigenvalue λ_n $(n = 0,1,\ldots)$, the solution of the ordinary differential equation (20) forms an eigenvector $\phi_n(\varepsilon)$. A general solution of the original Fokker-Planck equation (13) can be expressed via an expansion over all eigenvectors:

$$f(\varepsilon, t) = \sum_n C_n e^{-\lambda_n t}\phi_n(\varepsilon) \quad (21)$$

where C_n are constant coefficients.

Since all eigenvalues are positive, the solution approaches zero in the asymptotics $t \to \infty$. This effect describes "leakage" of the distribution function through the boundaries $\varepsilon_1, \varepsilon_2$ and results in the eventual loss of lock by the DLL.

The solution of the Fokker-Planck equation allows us to quantify such DLL characteristics as the time to get to a steady state, the time to lose lock, and the mean square tracking error.

If the first (smallest) eigenvalue λ_0 is much smaller than all other eigenvalues, than the second and subsequent terms in the sum in (21) become negligible compared to the first term after some time t_S. Time t_S when the second term in (21) becomes small can be estimated as $t_S \sim \frac{1}{\lambda_1}$, where λ_1 is the second eigenvalue. All subsequent terms will vanish even faster than the second term. Thus, after time t_S, the general solution of the Fokker-Planck equation can be approximately described by the first eigenvector $\phi_0(\varepsilon)$

$$f(\varepsilon, t) \approx C_0 e^{-\lambda_0 t}\phi_0(\varepsilon) \quad \text{for} \quad t > t_S \quad (22)$$

Time t_S is referred to as the time of getting to a steady state. Technically speaking, the resulting state is not steady since the solution $f(\varepsilon, t)$ continues to decrease slowly in value due to the exponent in the right-hand side. However, *the shape* of the solution remains constant.

Time to lose lock can be estimated from (22) as

$$t_{ll} \sim \frac{1}{\lambda_0} \quad (23)$$

The probability of keeping lock during time t_{ll} is $1/e$.

Finally, the mean square error in the steady state can be defined as:

$$\left\langle \varepsilon^2 \right\rangle = \frac{\int (\varepsilon - \langle\varepsilon\rangle)^2 f(\varepsilon, t) d\varepsilon}{\int f(\varepsilon, t) d\varepsilon} \quad (24)$$

Note that the mean square error is computed upon the condition that the lock has not been lost (i.e., there is no exponential time-dependent multiplier).

Examples of solutions of the Fokker-Planck equations for the probability distribution functions are shown in Figure 1 for C/A, Manchester, and BOC2 signals.

7.2 Numerical Solution

Eigenvalues and eigenvectors have been found numerically for a coherent DLL with $d = 0.1T_c$.

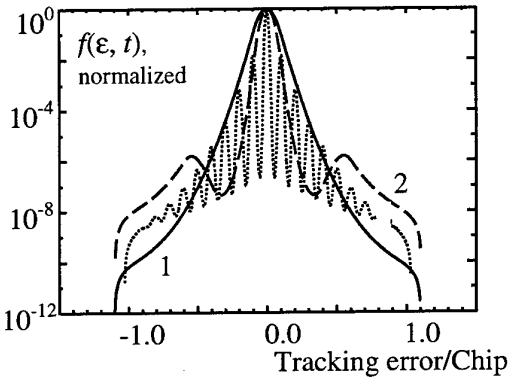


Figure 1. Normalized probability distribution functions for C/A (curve 1), Manchester (curve 2), and Split C/A (curve 3). Oscillations are due to multiple equlibria for the DLL

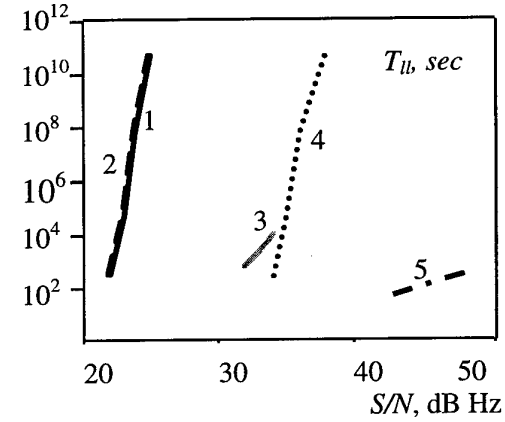


Figure 2. Time to lose lock as a function of signal-to-noise ratio for different modulations. 1 - C/A, 2 - Split C/A, non-coherent tracking, 3 - Manchester, 4 - C/A, majority voting, 5 - Split C/A

Derivatives in equation (20) were approximated by finite differences, thus yielding an eigenvalue problem for a tridiagonal matrix. Algorithms for finding eigenvalues and eigenvectors of a tridiagonal matrix were taken from [3]. The following settings were used:

- Power delay locked loop
- Conventional and majority voting applies to C/A or P(Y) code
- Discriminator separation = 0.1 chip for conventional and Manchester signals; 0.03 chip for the BOC2 and BOC3 signals
- 10 ms loop integration time
- Gain was selected to guarantee 10 ms loop pull-in time (measure of dynamic tracking capability)

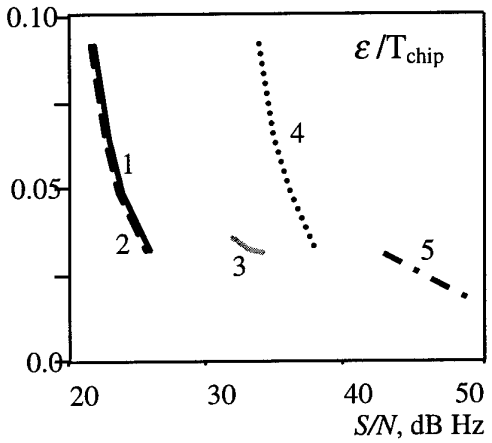


Figure 3. Tracking error as a function of signal-to-noise ratio for same types of modulations, as those in Figure 2

Results are presented in Figures 1-3. If a signal is majority voted, its S/N ratio must be increased by 6 dB.

8. Effects of Band-Limited Noise and Tone Jamming

If a band-limited noise is present, it is possible to define an equivalent spectral density of the infinite-bandwidth white noise, which would have the same effect on the signal tracking. It is assumed that the RMS (root-mean-squared) amplitude of the band-limited noise is the same, as the amplitude of the signal.

Let N_∞ be the desired equivalent spectral density of white infinite-bandwidth noise, $N(\omega)$ be the spectral density of the band-limited noise, S_0 be the amplitude of the signal, and let $S_\pm(\omega)$ to be defined by:

$$|S_\pm(\omega)|^2 = 4|S(\omega)|^2 \begin{cases} \cos^2(\omega d) \\ \sin^2(\omega d) \end{cases} \qquad (26)$$

and $S(\omega)$ be the spectrum of the signal.

After some algebra, the formula for the equivalent spectral density of infinite-bandwidth noise can be derived as

$$N_\infty = \left(\frac{\int |S_+(\omega)|^2 N(\omega) d\omega \int |S_-(\omega)|^2 N(\omega) d\omega}{\int |S_+(\omega)|^2 d\omega \int |S_-(\omega)|^2 d\omega} \right)^{\frac{1}{2}} \frac{2\pi S_0^2}{\int N(\omega) d\omega} \qquad (27)$$

The initial assumption that the RMS amplitude of noise is equal to the amplitude of the signal does not limit the application of formula (27). If the RMS amplitude of noise is higher (lower) than signal amplitude, the equivalent spectral density of the infinite-bandwidth noise should be increased (decreased) proportionally.

For various noise spectra and signals, the equivalent noise density N_∞ can be computed numerically. The result can be used to define the equivalent SNR and all previous Fokker-Planck computations for tracking performance become directly applicable.

The table below shows spectral density of broadband noise which would cause tracking errors equivalent to that caused by a band-limited noise. The mean-squared magnitude of band-limited noise is assumed to be equal to the magnitude of the signal

The case when a tone-jamming signal is present is a bit more complicated. The derivation of the Fokker-Planck equation requires that the stochastic term has short autocorrelation time. This assumption is violated in a case of a tone jammer, since a sine tone signal has infinite autocorrelation time. However, the integral of the tone jammer and the code may have short autocorrelation time, due to the pseudorandom nature of the code. Thus, one can apply the Fokker-Planck formulation for such long-period codes as P(Y). For short-period codes (such as C/A or split C/A), the Fokker-Planck formulation may be invalid. Hence, results for these codes must be used with caution.

Below, formulas are given for the equivalent spectral density of a white, infinite-bandwidth noise, which would cause tracking errors equivalent to that caused by a tone-jammer. The magnitude of the jammer is assumed to be equal to the magnitude of the signal.

$$N_\infty = \frac{2Q\left|\sin(2\Omega d)\right| \sin^2\left(\frac{\Omega T_c}{2}\right) S_0^2}{\Omega^2 T_c \sqrt{\left(\left(\frac{1}{J} - \frac{2d}{T_c}\right)(J-1) + \left(\frac{1}{J} - \frac{d}{T_c}\right)\right)\left(\frac{2d}{T_c}(J-1) + \frac{d}{T_c}\right)}} \qquad (28)$$

where

$$Q = \begin{cases} 1 & J = 1 \\ \dfrac{\sin^2(\Omega T_c)}{\left(1 + \cos(\Omega T_c)\right)^2} & J > 1 \end{cases} \qquad (29)$$

Results for different types of signal are shown in the table. Frequency of the jamming signal is selected to maximize the tracking error.

Signal type	Discr. spacing/chip	Equiv. broad-band noise dB/Hz
C/A*	0.1	-55.4
P(Y)	0.1	-65.4
Manch.	0.1	-62.3
BOC1	0.1	-62.5
BOC2	0.01	-60.3
BOC3	0.01	-63.1
Spl. C/A*	0.01	-56.1

Appendix

For a stationary noise process, code phase $\hat{\theta}$ does not affect statistical characteristics of $v(t)$ as defined by (6). Hence, we will dwell on the following equation:

$$v(t)=\frac{1}{N_0}\int_{t-T_0}^{t} N(\tau)S_0(\tau)d\tau \qquad \text{(A1)}$$

The autocorrelation function of $v(t)$ is defined as:

$$C_V(\Delta t)=\langle v(t)v(t+\Delta t)\rangle \qquad \text{(A2)}$$

where $\langle\cdots\rangle$ denotes the mathematical expectation.

We start from the case $0<\Delta t<T_0$. Other cases can be treated similarly. Substitution of (A1) in (A2) and splitting each integral in two parts yields

$$C_V(\Delta t)= \frac{1}{N_0^2}\left\langle\left(\int_{t-T_0}^{t-T_0+\Delta t} N(\tau)S_0(\tau)d\tau+\int_{t-T_0+\Delta t_0}^{t} N(\tau)S_0(\tau)d\tau\right)\left(\int_{t-T_0+\Delta t_0}^{t} N(\tau)S_0(\tau)d\tau+\int_{t}^{t+\Delta t_0} N(\tau)S_0(\tau)d\tau\right)\right\rangle \qquad \text{(A3)}$$

Next, we find separately mathematical expectations of four products of integrals in (A3). Since noise values are uncorrelated for different times, products of integrals which do not have their integration limits overlapped, are averaged out. The only surviving product yields:

$$C_V(\Delta t)=\frac{1}{N_0^2}\left\langle\left(\int_{t-T_0+\Delta t_0}^{t} N(\tau)S_0(\tau)d\tau\right)^2\right\rangle \qquad \text{(A4)}$$

For a coherent early-mines-late loop, code $S_0(\tau)$ comprises pulses with magnitude $2P(\tau)$, pulse duration $T_p=2d$ and average pulse period $2T_c$, where T_c is the chip period, and d is the early (or late) shift with respect to the sync code. The integration interval on average contains $\frac{T_0-\Delta t}{2T_c}$ of such pulses. Since integrals over each of the pulses are statistically independent, (A4) can be rewritten as

$$C_V(\Delta t)=\frac{1}{N_0^2}\frac{T_0-\Delta t}{2T_c}\left\langle\left(\int_{0}^{2d} N(\tau)2P(\tau)d\tau\right)^2\right\rangle \qquad \text{(A5)}$$

The latter equation has a simpler form when pulses are constant during pulse duration. In this case,

$$C_V(\Delta t)=\frac{2P^2}{N_0^2}\frac{T_0-\Delta t}{T_c}dN_\omega \qquad \text{(A6)}$$

where N_ω is the spectral density of the noise[2].

Other cases (i.e., when $\Delta t>T_0$, $-T_0<\Delta t<0$, $\Delta t<-T_0$) can be treated similarly. Overall, the autocorrelation function $C_V(\Delta t)$ has a triangular form and is plotted in Figure 1.

The autocorrelation function has maximum value $C_v^{\max}=\frac{2P^2}{N_0^2}\frac{T_0}{T_c}dN_\omega$ at $\Delta t=0$. (A3)

The requirement that this autocorrelation function is similar to Dirac delta function, means satisfying two criteria:

1. T_0 is small in comparison with "system reaction time"
2. Integral of $C_V(\Delta t)$ with respect to Δt is equal to 1.

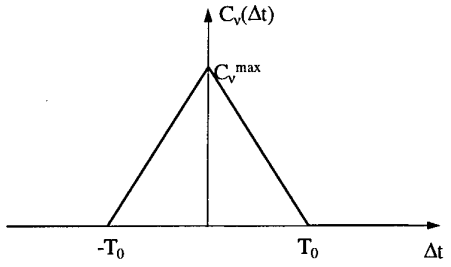


Figure 4. The autocorrelation function $C_V(\Delta t)$

The latter requirement yields the following condition:

[2] Factor 2 appears instead of 4, since we have only in-phase component here and do not account for the quadrature component. Only 1/2 of the total noise power is contained in the in-phase component.

$$\frac{2P^2}{N_0^2}\frac{T_0^2}{T_c}dN_\omega = 1 \qquad \text{(A7)}$$

Thus, the normalization constant must have the following value:

$$N_0 = PT_0\sqrt{\frac{2dN_\omega}{T_c}} \qquad \text{(A8)}$$

Formulas for signals (including signals with subcarrier modulation) tracked by the non-coherent DLL (power DLL and dot-product DLL) can be obtained in a similar way. Since derivation involves considerable algebra, it is omitted in this report. Final formulas are given in the text.

References

1. Anderson, J., *Waveform Development Plan Revision 1*, GPS Modernization: Advanced Signal Development, June 2, 1998.

2. H. Meyr, G. Ascheid, *Synchronization in Digital Communications, Vol. 1, Phase-, Frequency-Locked Loops, and Amplitude Control*, 1990, NY, John Wiley & Sons, 510 p.

3. J. H. Wilkinson, and C. Reinsch, *Linear Algebra, Vol. II of Handbook for Automatic Computation*, 1971, NY, Springer-Verlag, p. 510.

Results from the GPS Experiment on Equator-S

Norbert Lemke
Kayser-Threde GmbH, Wolfratshauser Str. 48, 81379 Munich, Germany
norbert.lemke@kayser-threde.de, http://www.kayser-threde.de

Bernd Eissfeller, Oliver Balbach
Institute of Geodesy and Navigation (IfEN), University FAF Munich, Neubiberg, Germany

Werner Enderle, Michael Schmidhuber
German Space Operations Center (DLR-GSOC), Oberpfaffenhofen, Germany

Summary: An encouraging concept for position determination of geostationary satellites is the use of the Global Positioning System (GPS). Although the geostationary orbit is well beyond the orbits of the GPS Navstar satellites (i.e. 20 000 km), it is possible to use GPS - as demonstrated with the GPS experiment on board of the German Small Satellite Equator-S.

The experiment data have significant impact for the future use of GPS receivers on-board geostationary satellites. In the past, GPS receivers have only been used well below the orbital altitude of the GPS satellites. The reception of GPS signals at greater altitudes has been demonstrated within the Equator-S GPS experiment. As part of this, GPS signals have also been received from the antenna side lobes of the GPS satellites. The maximum altitude where the GPS receiver provided measurements was about 61 000 km.

Introduction: New concepts for efficient use of the geostationary orbit slots require precise, autonomous and reliable orbit and attitude determination and control systems for geostationary satellites. Typical values for current accuracy requirements and future requirements for clustering satellites in geostationary orbits are given in Tab. 1.

With the use of GPS the satellite's position can be determined more precisely with respect to conventional currently used methods including range, rate and angle measurements from one or more ground stations.

Orbital Parameter	Current Accuracy (1σ)	GEO Cluster Requirement (1σ)
semi major axis	$\pm$ 20 m	$\pm$ 3 m
eccentricity	$\pm 5 \cdot 10^{-6}$ m	$\pm 5 \cdot 10^{-6}$ m
inclination	$\pm$ 0.002° (1.4 km)	$\pm$ 0.0005° (0.35 km)
longitude	$\pm$ 0.0060° (4.2 km)	$\pm$ 0.0010° (0.7 km)

Tab. 1: Requirements for GEO Satellite Cluster

The results of the GPS experiment on board of the 230 kg satellite Equator-S are of key importance for the future use of GPS receivers on geostationary and highly elliptical satellites. The use of GPS based orbit and attitude determination and control systems will lead to high accuracy, as well as to onboard autonomy, which will result in a reduction of mission operation costs.

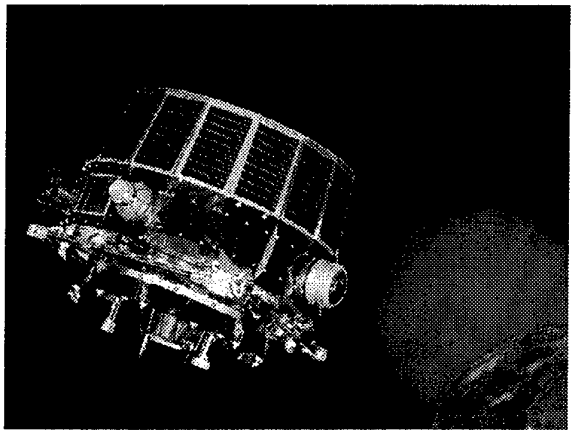


Fig. 1: German Small Satellite Equator-S

The spin-stabilized satellite Equator-S (cf. Fig. 1) was launched on December 2, 1997 into a highly elliptical orbit. The satellite is part of the Inter-Agency Solar-Terrestrial Physics Program. It orbits around the Earth in an highly elliptical orbit. The huge variation in orbital height (between one and 70 thousand kilometers) enabled the investigation of the whole spectrum of GPS space applications with the Equator-S GPS experiment.

The scientific goals of Equator-S are the investigations of the Earth' magnetic field and of the interactions between the magnetic fields of the sun and the Earth. Besides several scientific payloads, Equator-S carries also a GPS technological experiment. The two Equator-S orbits are both highly elliptical (transfer orbit 200 km × 36 000 km, and final orbit 500 km × 67 000 km).

The scientific objective of the GPS Experiment on board of Equator-S was the investigation of GPS data reception conditions in very high altitudes with respect to GPS satellite visibility (geometry, link budget, side lobe usage) and measurement quality (position, pseudo-range, carrier phase) for future applications of autono-

Paper presented at the RTO SCI International Conference on "Integrated Navigation Systems", held at "Elektropribor", St. Petersburg, Russia, 24-26 May 1999, and published in RTO MP-43.

mous orbit / attitude determination and control for especially geostationary satellites.

The GPS experiment was developed with funding of the German Aerospace Center (DLR), by the Institute of Geodesy and Navigation (IfEN) of the University of Federal Armed Forces in Munich, the German Space Operations Center (GSOC) of DLR in Oberpfaffenhofen, and by the Munich based aerospace company Kayser-Threde GmbH.

11-Dec-95	Kick-Off Meeting
18-Mar-97	Delivery of Flight H/W
02-Dec-97	Launch Equator-S
03-Dec-97	Tracking in 34 000 km
13-Dec-97	Tracking using GPS Antenna Side Lobe
31-Mar-98	Tracking in 45 000 km
31-Mar-98	Tracking in 61 000 km
30-Sep-98	Official End of Project

Tab. 2: GPS Experiment Mile Stones

Motivation: The GPS experiment aims at investigating the GPS satellite signals reception conditions at high altitudes, especially those well above the GPS orbital altitude (approx. 20,000 km). The main problem in the signal reception above the GPS altitude is the directivity of the antennas of the GPS satellites. The GPS satellites are Earth oriented and therefore their antennas are directed to the Earth.

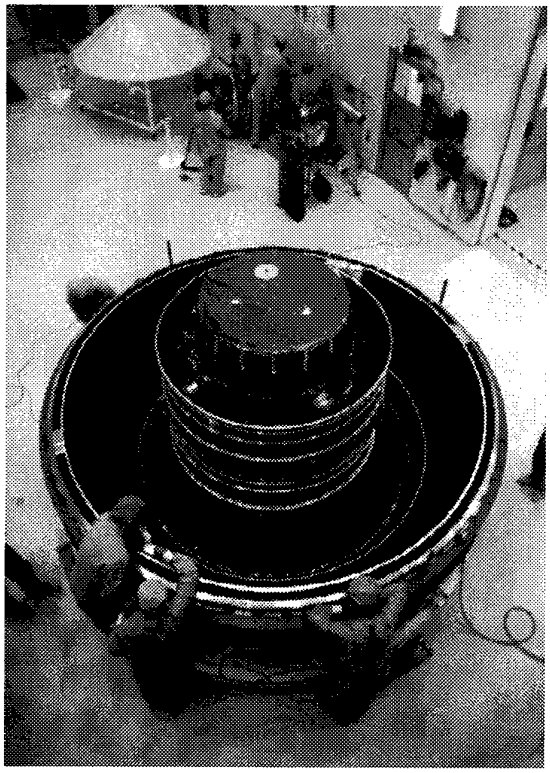
A GPS receiver on a satellite orbiting outside the GPS satellite's orbits consequently can only receive signal from GPS satellites "behind" the Earth, i.e. from GPS satellites, which antennas are directed to the far side (as seen from the GPS receiver) of the Earth. The additional usage of the side lobes of the GPS satellite antennas (the characteristics can be found in [Czop 93]) increases the (RF) visibility of the GPS satellites.

However, the side lobes of the GPS antennas do not graze the Earth. This effect therefore can only by studied in space.

The other challenging aspect to be studied is the very high dynamic of the GPS signal below the GPS orbital altitude due to the eccentricity of the satellite's orbit. At the perigee of the elliptical orbit the velocity of the spacecraft is very high, and therefore also the Doppler shift of the GPS signal.

Hardware: The hardware for the GPS Experiment on-board Equator-S has been provided by Kayser-Threde. The equipment includes the GPS receiver with two antennas and an interface box especially designed for the mission.

GPS Receiver: The Motorola Viceroy™ GPS Receiver is an intelligent GPS sensor intended to be used in spacecraft as part of the on-board navigation system. It is capable of providing autonomous position, velocity and time information. The minimum useable system combines the receiver and an intelligent controller such as an IBM compatible personal computer or a spacecraft



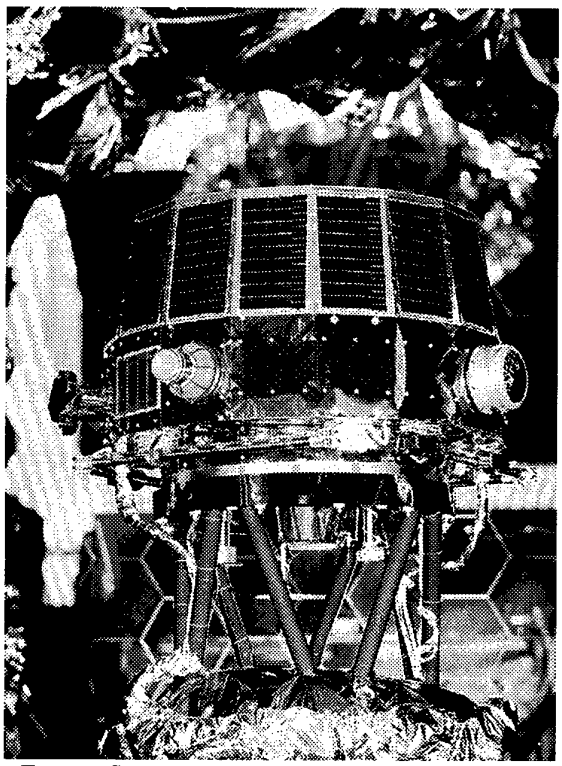


Fig. 2 a + b: Satellite Equator-S

computer.

The Viceroy GPS Receiver is an L1 C/A-code, dual antenna, 12 (2×6) channels receiver. With the dual antenna configuration, the receiver can track up to twelve different GPS satellites. The receiver parts program is based on commercial core design, radiation tolerant parts replace non-tolerant parts, single event upsets are tolerated.

The receiver automatically acquires and tracks GPS satellites, measures the pseudo-ranges and the integrated carrier phases, decodes and collects satellite broadcast data, computes its instantaneous position, velocity and time, and outputs the results.

The use of dual antennas provides significant flexibility in antenna location with respect to shading and space vehicle attitude.

The Motorola Viceroy GPS Receiver will operate when the GPS SA is activated. Accuracy will be consistent with the level of degradation employed and the receiver will introduce negligible additional measurement errors. The GPS AS has no effect on C/A-code operation.

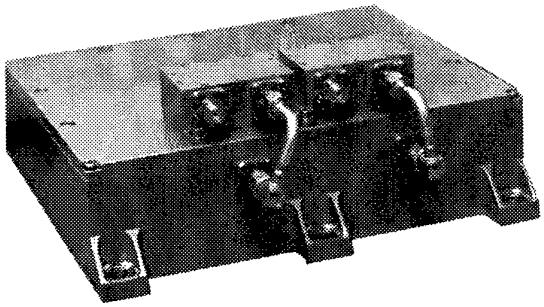


Fig. 4: Motorola Viceroy™ GPS Receiver

Size	approx. 152x132x43 mm
Weight	approx. 1.5 kg
DC Power	4.8 watts (20 to 32 V)
Architecture	continuous dual antenna operation, 6 + 6 channels (6 per antenna), L1, C/A-code, Doppler Dead Band ± 3 kHz
Dynamics	8,000 m/s, 10 m/s²
TTFF	typ. 60 sec.
Interface	RS-422 electrical interface 1 PPS out

Tab. 3: Technical Data GPS Receiver

All commands required for the operation of the Viceroy GPS Receiver can be commanded by telemetry. For this purpose Kayser-Threde has designed an interface box, communicating with both, the GPS receiver and the telemetry and telecommand system of Equator-S.

GPS Interface Box: The GPS I/F Box developed by Kayser-Threde for the GPS Experiment on-board Equator-S is a data converter with an intermediate memory. It converts the PCM data stream from the Equator-S board computer into an RS-422 data transmission which is appropriate for the GPS receiver, and provides power adaptation for the GPS receiver.

Besides the buffered data transmission from 512 Bit/s effective data rate of the PCM system from Equator-S board computer into 19.2 kbit/s for the communication with the GPS receiver, and vise versa, the precise 1 pps clock signal is provided to the spacecraft for time normal correction and later ground processing.

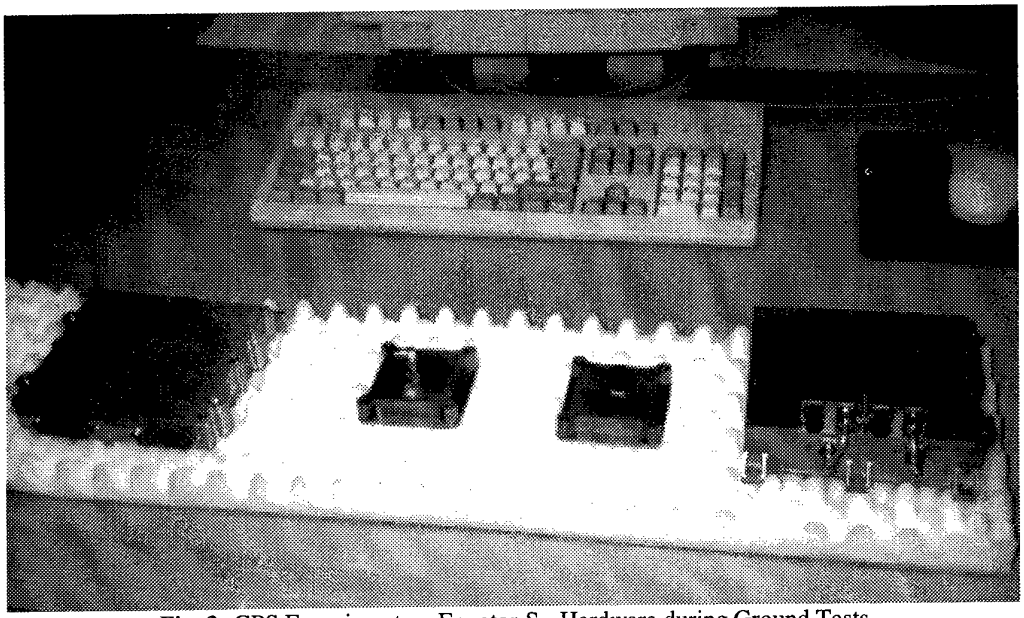


Fig. 3: GPS Experiment on Equator-S - Hardware during Ground Tests

GPS Antennas: One of the two antennas is mounted on the top of the body of the spacecraft at a distance of 14 cm from the spin axis, the other one on the bottom at a distance of 30 cm. Each six channels of the GPS receiver are assigned to the top antenna and the bottom antenna, respectively.

Limitations: Due to costs and time constraints, it was decided to use off-the-shelf equipment to the maximum extent possible for the GPS Experiment on board of Equator-S, in spite of that is was known that the GPS receiver had some limitations which might affect its operational usage in HEOs (the GPS receiver was designed and developed for LEO missions). Its use in this case under experimental conditions was considered as appropriate, especially because it was available very quickly.

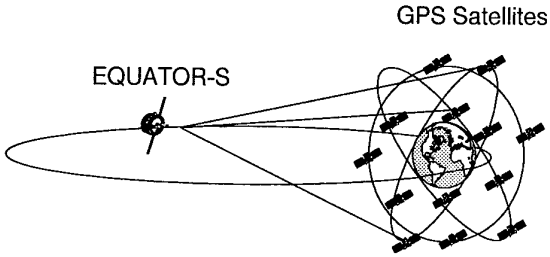


Fig. 5: Principle of the GPS Signal Reception Problem

The usage of the Motorola Viceroy GPS receiver constraints in some cases with the experiment mission profile. Some of the identified constraints are for instance:

- The receiver has standard signal sensitivity characteristics: approx. 39 dBHz acquisition threshold; tracking threshold is typically some few dB below that.
- The receiver satellite selection algorithm does not work above 20 000 km. This indicates that manual satellite selection is necessary.
- The receiver's on-board orbit propagator algorithm is using an orbit model appropriate for LEO but not for HEO. Therefore problems of the Doppler prediction are the results (i.e. problems in signal acquisition will occur).
- The receiver has a dead band from -3 kHz up to +3 kHz in the Doppler shift where no GPS signals can be tracked. This is due to the re-design of a ground receiver.
- The receiver has been developed for low Earth orbit applications and is therefore not well suited with respect to radiation hardness for use in GEO or high altitudes in general.

Fortunately enough, these limitations and constraints can be solved to a very significant degree. However, the price to pay for overcoming the identified limitations lies at higher efforts during the experiment mission operations.

Taking into account the magnitude of the receiver acquisition threshold, it can not be expected that the

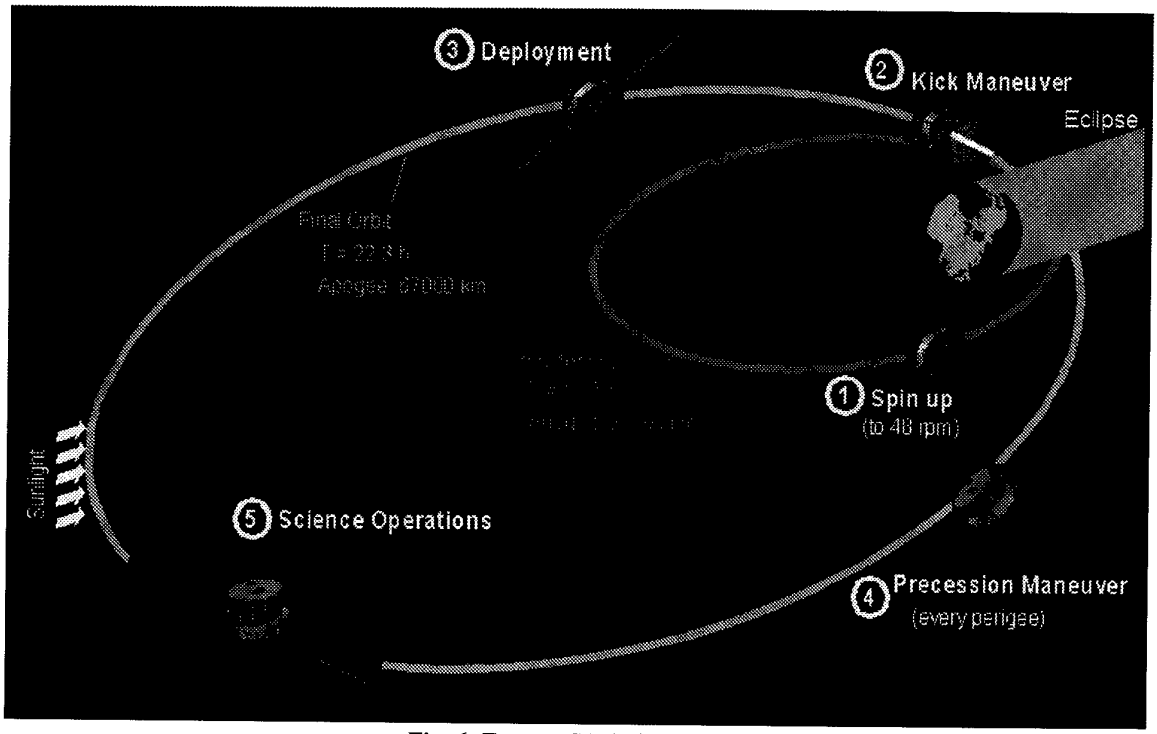


Fig. 6: Equator-S Mission Events

receiver will acquire GPS satellites with very low signal power levels. However, it will be able to keep them in track to a large extent. For this reason, despite Equator-S crosses two times per day the geostationary orbit, only the crossing when moving from perigee to apogee (outbound) has been expected to lead to results of interest for the experiment.

This is due to the receiver's acquisition threshold since it will be much easier for the receiver to keep satellites in track while on the outbound part of the orbit than to acquire them while approaching the Earth (inbound).

However, GPS satellites have been tracked by the receiver at the descending part of the orbit. For instance on 31-Mar-1998: GPS satellite PRN 22 has been tracked for 17 minutes at an altitude from about 47 000 to 45 000 km.

Orbital Element	Initial Orbit (GTO)	Final Orbit
Epoch	997/12/02 23:16:42 UT	997/12/11 18:38:30 UT
Semi Major Axis	24 467.7 km	40 084.9 km
Eccentricity	0.7311	0.8288
Inclination	3.999 deg.	3.999 deg.
R. Ascension of Asc. Node	232.561 deg.	225.667 deg.
Argument of Perigee	178.045 deg.	195.769 deg.
Mean Anomaly	3.840 deg.	6.738 deg.
Perigee Height	200.8 km	484.8 km
Apogee Height	35 978.3 km	67 275.3 km

Tab. 4: Orbital Elements for Equator-S

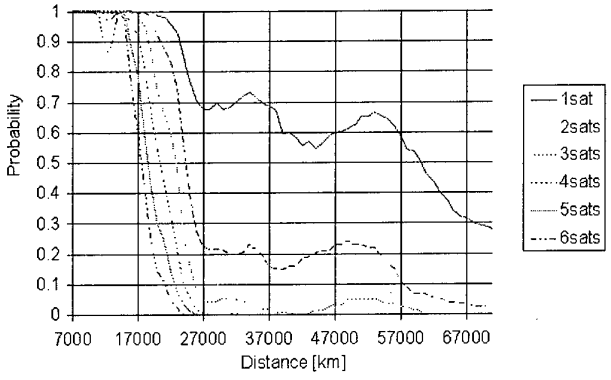


Fig. 7: Probability of GPS satellite visibility (working orbit)

GTO Mission Profile: A typical GTO profile for the GPS experiment is:

- GPS receiver is powered-up and initialized while Equator-S approaches the perigee (at an altitude of approx. 20 000 km). Date, time and user ephemeris are uploaded and the receiver is left in automatic satellite selection mode.
- The receiver will acquire and track satellites normally (shall compute 3D navigation solutions and shall be able to decode and register the current almanac).
- After passing the perigee the updated user ephemeris have to be uploaded.
- At an altitude of approx. 15 000 to 20 000 km after crossing the perigee, a manual selection of the GPS satellites to be tracked will be performed. Of course, the selected GPS satellites must be those which are most appropriate for the experiment, i.e. those visible from Equator-S in altitudes between 15 000-20 000 km and up to 36 000 km. Typically, there will not be more than two or three such satellites so that each of them could be allocated to several channels in order to achieve a more robust tracking performance. Especially important here is the consideration of the Equator-S attitude since each GPS antenna is assigned to a different block of six channels.
- Two more updates of channel-to-GPS satellite assignments by manual select commands after each half an hour.
- It can be expected that from an altitude of 50,000 km (in final orbit) upwards, no GPS satellite will be tracked and thus the receiver may be set to idle mode and powered-down. It is important to notice at this point, that maintaining the receiver powered-up continuously during prolonged periods of time, increases significantly the risk of SEUs and SELs (specially taking into account the high total doses at the Equator-S orbit); for that reason it was not recommended at all to keep the receiver powered-up for longer periods [Moto 96].

Mission Simulation and Mission Training: After integration of the GPS receiver into the Equator-S satellite, the complete experiment has been tested in the final configuration by emitting GPS signal onto the antennas. With two tests the functionality of the GPS receiver, the antenna system (including the harness) and the telecommand and telemetry links between the experiment and the experimenters have been tested. The set-up also allowed the verification of the commanding procedures.

The emitting of GPS signal onto the antennas leads to further advantages:

- The hardware including the cables are already mounted and connected. The final launch configuration (electronic, RF and data links) could be tested and verified.

- More realistic RF conditions (w.r.t. antenna cable connections).
- Disturbances generated by the satellite itself could be excluded

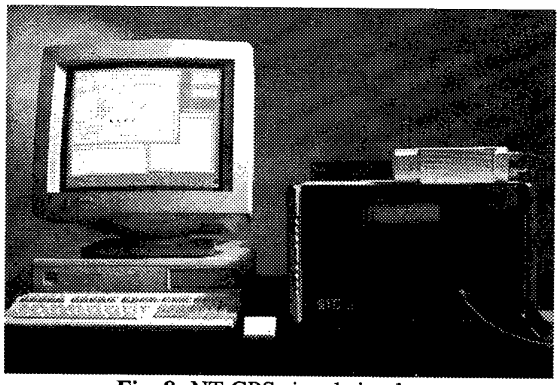


Fig. 8: NT GPS signal simulator

The pre-launch verification of the experiment has been divided into two parts: radiation (re-emitting) of real GPS signal (static scenario) and radiation of GPS signals generated with a GPS signal simulator (Fig. 8) for a GTO (dynamic scenario).

For the calculation of the link budget the thermal noise of the Earth, the space and the components of the transmission line were regarded, as well as losses (including an ionospheric model) and antenna diagrams of both, the GPS transmitting antenna and the receiver antennas. The tool calculates Doppler shift and rate, positions and velocities, antenna directivity losses and the applicable antenna lobe (side lobe or main lobe of the GPS transmitting antenna).

Results: The problems of using GPS beyond the GPS altitude had been discussed in [Eiss 96] and [Ende 98]. Some of the results from the GPS Experiment on-board Equator-S have been presented in [Balb 98] and [Ende 98]. However, the analysis of the so far gathered experiment data has not been finished so far.

03-Dec-1997	PRN24: 34,000 km
03-Dec-1997	PRN5: 25,000 km
04-Dec-1997	PRN5: 34,000 km
04-Dec-1997	PRN19: 34,000 km
09-Dec-1997	PRN3: 23,000 km
11-Dec-1997	PRN3: 24,000 km
13-Dec-1997	Side Lobe Reception
19-Feb-1998	
30-Mar-1998	
31-Mar-1998	PRN22: 45,000 km
31-Mar-1998	PRN30: 61,000 km

Tab. 5: GPS Signal Tracking Altitude Results

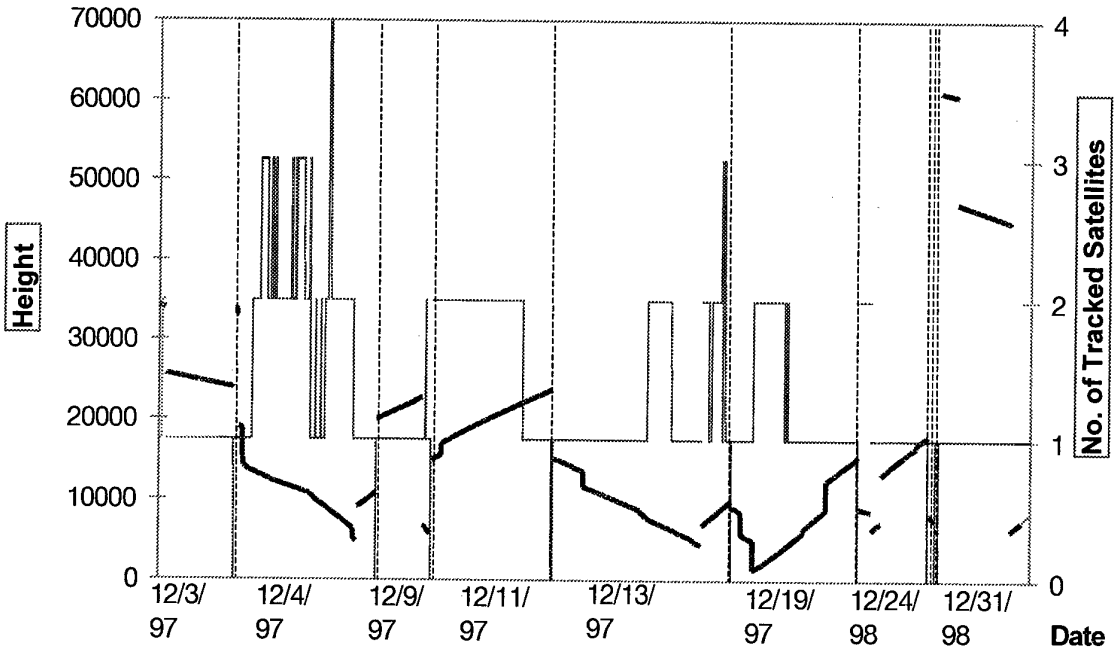


Fig. 9: Overview of GPS Measurements

Here two tracking periods will be presented:

- 03-Dec-1997, first tracking of GPS satellites in GEO altitude
- 31-Mar-1998, tracking at highest altitude

First Tracking of GPS Satellites in GEO Altitude: The first tracking of a GPS satellite (PRN24) took place on December 3, 1997. It started at 05:38 and lasts 03:52 min. This was during the GTO using the bottom antenna.

The receiver locked two channels and reached channel tracking mode 8 "Available for Position".

Fig. 10 a + b show the visibility prediction results from the visibility tool.

Fig. 11 shows the channel assignment and the tracking periods in mode 8 "Avail for Position".

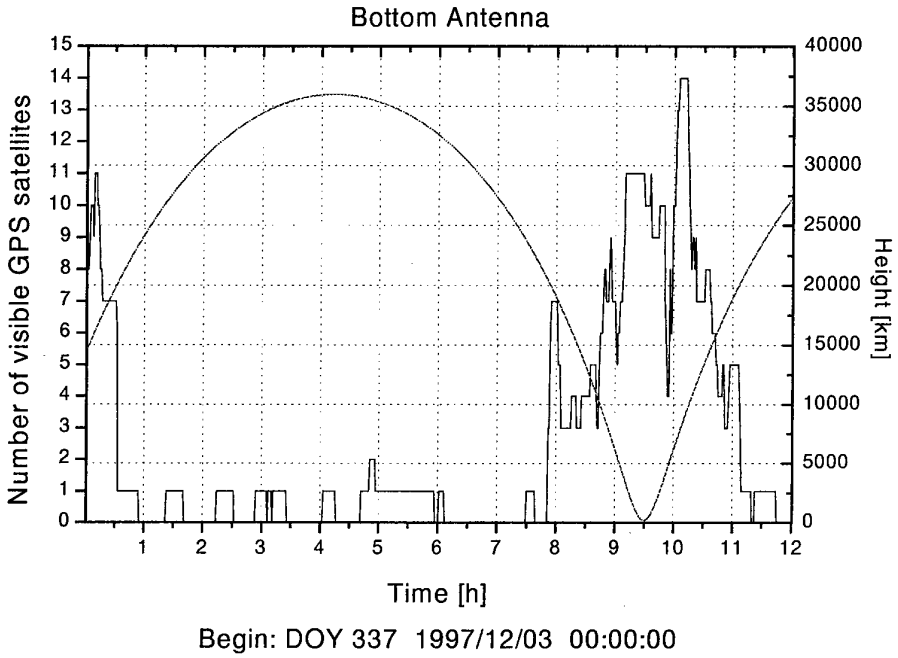


Fig. 10a: GPS Visibility Prediction: Number of Visible Satellites

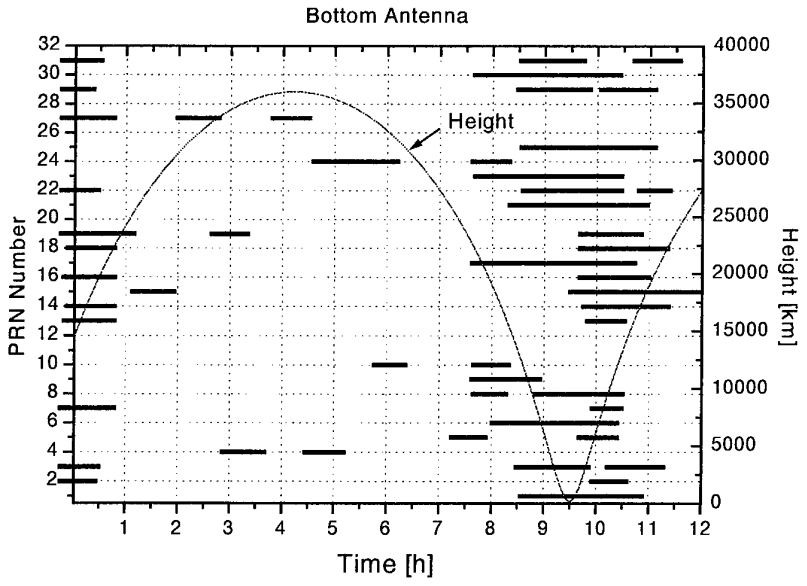


Fig. 10 b: Visible GPS Satellite PRN Prediction

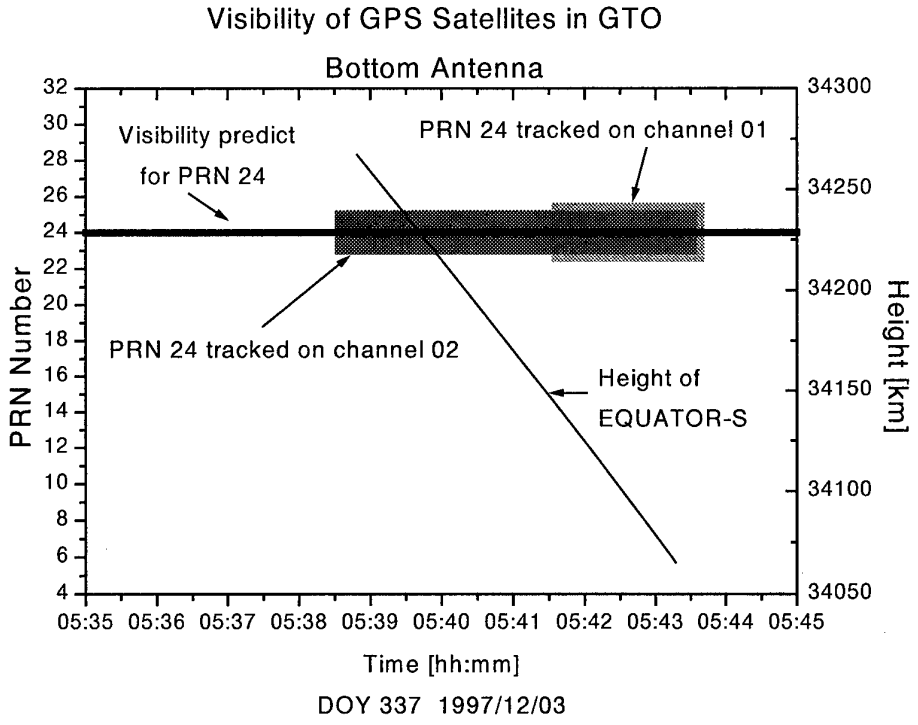


Fig. 11: First Tracking of a GPS Satellite in GEO Altitude

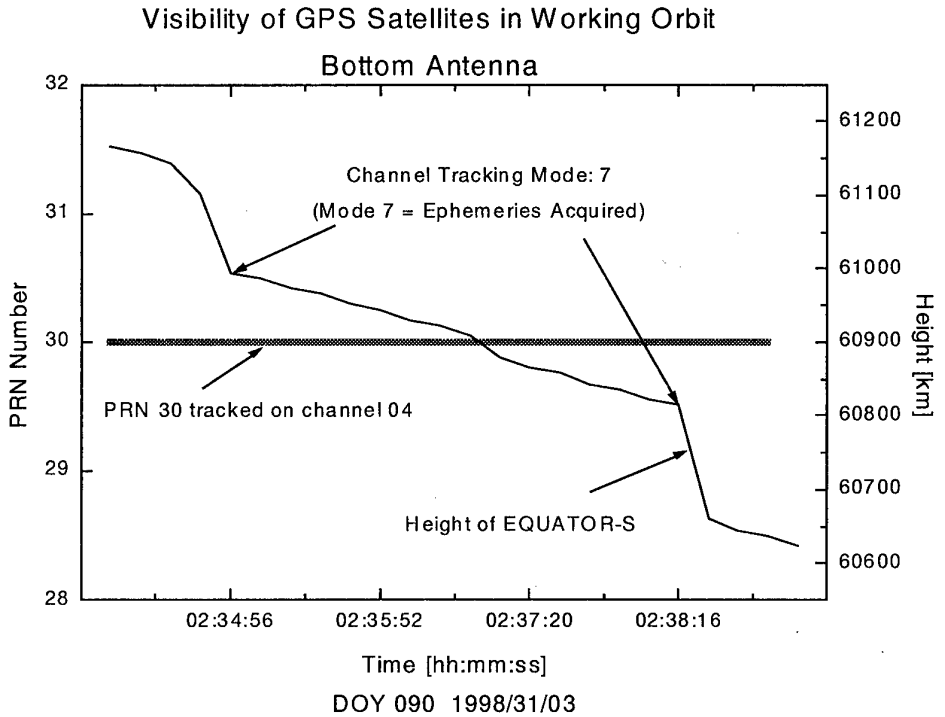


Fig. 12: Tracking of a GPS Satellite in 60,000 km altitude

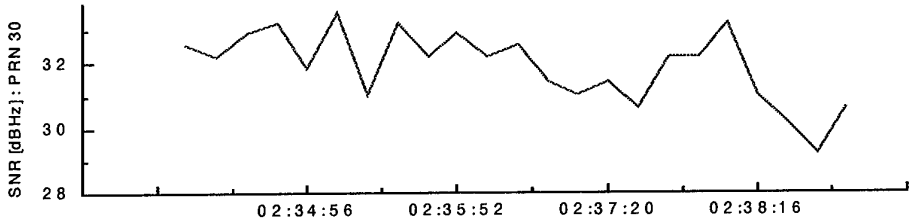


Fig. 13: SNR Values for Tracking in 60,000 km altitude

Tracking at Highest Altitude: The tracking of a GPS satellite (PRN30) at the highest altitude could be measured on March 31, 1998, 02:34. The duration was about 10 minutes, with three minutes in channel tracking mode 7 "Ephemeris Acquire". The altitude at that time was about 61,000 km. However, due to the weak signal the receiver could not reach mode 8 "Avail for Position".

The begin and end of period with tracking mode 7 is marked in Fig. 12 with arrows. The SNR values are given in Fig. 13.

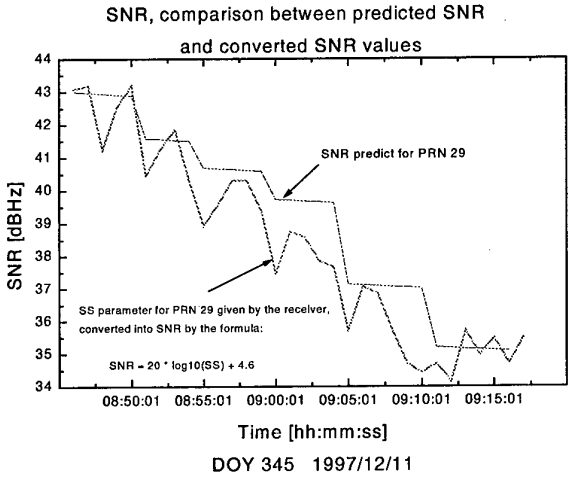


Fig. 14: SNR Comparison for PRN29

Conclusion: The results of the GPS Experiment show that it is possible to acquire and track GPS signals up to the GEO altitude and far beyond. The reception of signals by the experiment's GPS receiver was possible at an altitude of about 61,000 km, although the receiver is not optimal adapted to the mission.

It also could be proved the possibility to track GPS satellites in the first side lobe of the GPS satellite transmitting antenna (cf. [Ende 98] for details). GPS signal reception from the side lobes increase the number of "visible" GPS satellites for high altitude missions.

If the number of instantaneous visible GPS satellites decreases in a particular mission to only three or less GPS satellites it should be possible to continue navigation applying dynamic orbit models and Kalman filtering. For future use of GPS for navigating satellites in higher altitudes back-beams transmitting GPS signals towards anti-nadir will increase the number of visible GPS satellites

Finally it should be noted that the maximum number of simultaneously tracked GPS satellites was three, therefore a position solution determined by the receiver is not available. Unfortunately the satellite Equator-S did not recover from the failure which occurred on May 1st, 1998. Although the hope that Equator-S would heal itself is not completely gone, all signs indicate that the spacecraft is lost. In spite of its short life-time, Equator-S has provided highly valuable data.

Acknowledgement: The project is funded by the former German Space Agency (Deutsche Agentur für Raumfahrtangelegenheiten, DARA GmbH), which is now incorporated into DLR (Deutsches Zentrum für Luft- und Raumfahrt).

References:

[Czop 93] Czopek, F.; Shollenberger, S.: *Description and Performance of the GPS Block I and II L-Band Antenna and Link Budget.* Proceedings of the ION GPS-93, Salt Lake City, Utah, September 1993.

[Moto 96] Motorola GSTG: *Facsimile Communication in the framework of the MOMSNAV program.* KT.MNV.LT.749.IN, November 22, 1996.

[Kays 96] Kayser-Threde GmbH: *GPS Experiment On-board EQUATOR-S, Simulations with a GPS Signal Simulator.,* KT.EQU.TN.001, Issue: 2, Rev.: –, August, 1996.

[Eiss 96] Eissfeller, B.; Balbach, O.; Roßbach, U.; Fraile-Ordónez, J.: *GPS Navigation on the HEO Satellite Mission EQUATOR-S - Results of the Feasibility Study.* ION GPS-96, Kansas City.

[Balb 98] Balbach, O.; Eissfeller, B.; Hein, G.; Enderle, W.; Schmidhuber, M; Lemke, N.: *Tracking GPS above GPS Satellite Altitude: First Results of the GPS Experiment on the HEO Mission Equator-S.* Position Location and Navigation Symposium PLANS, Rancho Las Palmas Resort, April 20-23, 1998

[Ende 98] Enderle, W.; Schmidhuber, M.; Gill, E.; Montenbruck, O.; Braun, A.; Eissfeller, B.; Balbach, O.; Lemke, N.: *GPS Performance for GEO's and HEO's - The Equator-S Spacecraft Mission.* International Symposium on Space Flight Dynamics, GSFC, May 11-15, 1998.

A Permanent GPS/Glonass Reference Station in The Netherlands

C.D. de Jong, N.F. Jonkman
Department of Mathematical Geodesy and Positioning
Delft University of Technology
Thijsseweg 11, 2629 JA Delft, The Netherlands

Abstract
In The Netherlands research on establishing permanent GPS reference stations was initiated in 1993, resulting in a network of five stations together with a central processing facility. This network, the Active GPS Reference System for The Netherlands (AGRS.NL) is used for a wide variety of high-precision GPS applications. They include positioning, sea-level and subsidence monitoring and land-surveying, but also the determination in near real-time of the water vapor content of the atmosphere.

With the development of the Russian Glonass, a second satellite navigation system has become available for similar high-precision applications. The combined use of GPS and Glonass is expected to greatly enhance the availability and reliability of space based navigation systems. In The Netherlands the potential benefits of Glonass were recognized, resulting in a joint research project of the Survey Department of the Ministry of Transportation and Public Works (MD) and Delft University of Technology (DUT), the purpose of which is to gain more insight in the performance of Glonass and in the possibilities of GPS/Glonass integration.

As part of the MD/DUT research project, DUT has established a permanent GPS/Glonass reference station, co-located with one of the GPS reference stations of AGRS.NL. The station is used to monitor the integrity of the Glonass system and the quality of the collected observations. To this end, dedicated integrity monitoring software developed at DUT for the AGRS.NL array has been adapted to incorporate Glonass observations. The software is able to detect, identify and adapt outliers and slips in GPS and Glonass observations in real-time. Moreover, as a byproduct it also allows for the generation of DGPS/DGlonass corrections and ionosphere estimates.

In this contribution a description is given of the GPS/Glonass reference station and the integrity monitoring software that is running at the station. The error detection capabilities of the software are described and demonstrated as well as some of its advanced options like multipath modeling and ionosphere estimation. The contribution will be concluded with a brief outline of future DUT research activities directed towards Glonass and GPS/Glonass integration.

1. Introduction

With the inauguration of a central processing facility in late 1997, the Active GPS Reference System for The Netherlands (AGRS.NL) became fully operational. Originally conceived in 1993 by Delft University of Technology (DUT), the Survey Department of the Ministry of Transport (MD) and the Triangulation Department of the Dutch Cadastre (RD), the AGRS.NL is a permanent GPS array consisting of five stations evenly distributed over The Netherlands. Each of the stations is equipped with a geodetic quality dual-frequency GPS receiver and additional hardware for remote operation of the receiver and data storage and transmission. The data collected at the stations is sent – typically on an hourly basis – to a central processing facility. At this facility the data is "cleaned" using integrity monitoring and network adjustment software after which the data is made available to users via the internet.

The data of the AGRS.NL is used for a wide range of geodetic and non-geodetic applications. One of the original objectives of the AGRS.NL was to serve as a backbone for accurate GPS height determination within the context of the fifth primary levelling of The Netherlands. Other typically geodetic applications are the use of the AGRS.NL for sea-level monitoring along the Dutch coast and land-subsidence monitoring near the Groningen gasfield in the north of The Netherlands. But the AGRS.NL also serves as a reference for precise positioning and surveying in for instance aerial photogrammetry and dike profile determination. As for non-geodetic applications, DUT recently participated in a research project of the Royal Dutch Meteorological Office (KNMI) exploring the use of GPS for the determination of the integrated partial water vapor content (IPWC) of the atmosphere. Originally operated in a campaign-like fashion, IPWC estimates can now be made available in near-real time. For more information on these and other applications, the reader is referred to the internet site www.agrs.nl.

Although the AGRS.NL has thus shown a great potential for a variety of applications, there are still limits on fully realizing this potential. These limits however, are not so much imposed by the design or the facilities of the AGRS.NL, but they are instead inherent to the GPS itself. More in particular, the major limiting factors of the AGRS.NL are the limited number of available GPS satellites and the United States Department of Defense's policy of denial of accuracy and access to unauthorized (civilian) users. As a consequence of the limited number of satellites there are quite a number of prolonged periods during which fewer than six satellites are visible over The Netherlands with an elevation angle above 15 degrees. During these periods the precision of the

Paper presented at the RTO SCI International Conference on "Integrated Navigation Systems", held at "Elektropribor", St. Petersburg, Russia, 24-26 May 1999, and published in RTO MP-43.

positioning results usually suffers from an unfavorable geometry, while the reliability of the results can not completely be ensured either. The denial of accuracy and access by means of Selective Availability (SA) and Anti-Spoofing (A-S) moreover, requires the use of more elaborate procedures and special, more costly equipment to attain the highest possible precision. Examples of this are the need to transmit differential corrections with a high update rate – every few seconds with SA, every few minutes without SA – and the need to use receivers that employ special techniques to reconstruct the observations on the second GPS frequency.

The limitations on the AGRS.NL imposed by GPS, have prompted DUT and MD to start a research project towards integrating the Russian Glonass navigation system in the AGRS.NL architecture. As Glonass is very similar in design and operation to GPS, it holds the promise of doubling the number of navigational satellites. Moreover, Glonass does not suffer from any signal degradation like GPS. Hence, augmenting GPS with Glonass is expected to greatly enhance the precision and reliability of the positioning results.

As a first cautious step towards utilizing Glonass within the context of the AGRS.NL, a single-frequency GPS/Glonass receiver was co-located with one of the AGRS.NL stations in July 1998. Although the primary aim of this receiver was to gain experience with combined GPS/Glonass observations, the receiver was also used to participate in the International Glonass Experiment 1998 (IGEX98). The IGEX98 campaign aims to collect geodetic quality Glonass data from a global network of stations in order to compute amongst others precise Glonass satellite orbits and an accurate transformation between the GPS and Glonass reference frames. In view of the favorable results obtained with the combined GPS/Glonass observations and the prolongation of the IGEX98 campaign until the end of September 1999, the single-frequency receiver was replaced with a dual-frequency GPS/Glonass receiver in February this year, see Figure 1. Moreover, a second dual-frequency receiver was acquired which will be co-located with one of the other AGRS.NL stations. Depending on the results obtained with these receivers and the continuation of the development of Glonass, combined dual-frequency GPS/Glonass receivers may be co-located with all AGRS.NL-stations.

The dual-frequency GPS/Glonass receiver currently installed at the AGRS.NL station in Delft has been equipped with real-time integrity monitoring software. The software was originally developed at DUT for the AGRS.NL array and runs on all AGRS.NL stations as well as on the central processing facility. Primary aim of the software is to detect and correct errors in the observations and to detect errors in the satellite transmitted broadcast ephemeris. As the software operates in real-time, it allows the AGRS.NL operator to immediately discern and act upon problems with the receivers at the AGRS.NL stations as well as with the GPS – and in future also Glonass - satellites. This is of particular importance as the AGRS.NL stations are in principle able to provide real-time differential and kinematic positioning services.

The use of the integrity monitoring software however is not restricted to the AGRS.NL stations alone nor to just the spotting of anomalies in the data. As the integrity monitoring software does not necessarily require a receiver to be installed at a known position, it can also be used at a roving receiver. The software would then act as a sophisticated Receiver Autonomous Integrity Monitoring (RAIM) package. If however a precise position is available for the receiver, the integrity monitoring software can also be used to generate differential GPS and Glonass corrections. Moreover, as the receiver would in that case be stationary, two other options of the software, adaptive multipath modeling and ionosphere estimation, could then also be employed. Multipath modeling will allow the mitigation of multipath effects on the code observations of the receiver and thereby improve quality of the integrity monitoring results as well as of the differential corrections. The ionosphere estimates moreover, can be used to generate interpolated ionosphere corrections for

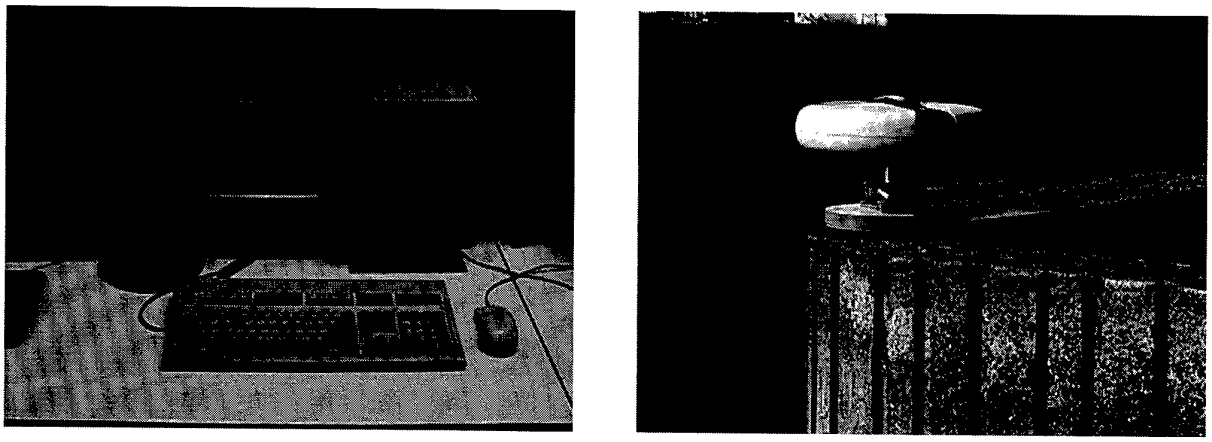


Figure 1: Dual-frequency GPS/Glonass receiver and antenna at Delft University of Technology. The receiver is connected to a PC for real-time integrity monitoring of the data.

long-range differential positioning as is for instance envisaged for the augmentation systems EGNOS and WAAS.

In this contribution, the most recent version of the integrity monitoring software capable of handling both GPS and Glonass data will be described in more detail. In the next section, the models on which the software is based will be explained and a theoretical evaluation of the error detection capabilities of the software will be given. Examples of the software's actual application to GPS and Glonass data will be presented in the following section together with a demonstration of some of the software's advanced options like multipath modeling and ionosphere estimation. Finally, in the last section, a brief overview will be given of the intended future DUT research on Glonass and GPS/Glonass integration.

2. Integrity monitoring of Glonass and GPS observations

In this section, the theoretical concepts on which the integrity monitoring software for GPS/Glonass data is based, will briefly be explained. The approach implemented in the software consists of two steps. In the first step the measured code and carrier phases are validated, while in the second step the satellite ephemeris and clock parameters are checked. Both steps of the procedure will be discussed, but the first step will be emphasized, as it is by far the more intricate. For a more detailed description of the software, the reader is referred to [de Jong, 1996].

2.1. Step 1: Validating the observations

The discussion of the code and carrier phase observation validation starts in the first subsection with a brief review of testing and reliability theory. The observation and dynamic models to which this theory is applied are introduced in the second subsection. Finally, in the third subsection the results of a series of design computations are presented and analyzed, illustrating the strength and efficiency of this first step.

2.1.1. Testing and reliability

The validation of the code and carrier phase observations is based on the recursive Detection, Identification and Adaptation (DIA-) procedure, [Teunissen, 1990], running in parallel to a Kalman filter, [Kalman, 1960]. The DIA-procedure aims to detect possible misspecifications in the observation model of the Kalman filter by means of statistical hypothesis testing. The procedure consists of the following steps

1. *Detection:* An overall model test is carried out to diagnose whether unspecified model errors have occurred.

2. *Identification:* If a model error is detected, its potential source is identified by testing the original or nominal observation model against models extended with bias parameters.

3. *Adaptation:* After the identification of the most likely source for the model error, the observation model is adapted to eliminate the biases in the filters' state vector.

The nominal observation model is indicated as the observation model under the null-hypothesis; the models extended with bias parameters are indicated as models under alternative hypotheses. In the integrity monitoring software only one-dimensional alternative hypotheses, describing outliers or integer cycle slips in the observations are considered. These biases are assumed to be described by a known vector c_k and an unknown scalar ∇, as $c_k\nabla$.

The test statistics associated with the DIA-procedure are based on the filter's predicted residuals or innovations sequence v_k, and its corresponding covariance matrix Q_{v_k}. The predicted residual is defined as the difference between actual and predicted observations. Under the null-hypothesis, the expectation of v_k is zero; under the alternative hypothesis the expectation equals $c_k\nabla$.

With these test statistics, the Detection and Identification step of the DIA-procedure can be described in the following manner. A model error is detected and the null-hypothesis is rejected in the Detection step, if the overall model test statistic exceeds a certain critical value. The critical value is set, based on the distribution of the overall model test statistic under the null-hypothesis. In the subsequent Identification step, the test statistics of the alternative hypotheses, indicated as the local slippage (LS) test statistics, are computed and the alternative hypothesis corresponding to the largest statistic is said to describe the most likely misspecification of the observation model. For this misspecification to be sufficiently likely however, the LS test statistic also has to exceed a critical value, which again is set based on the distribution of this statistic under the null hypothesis. If the largest slippage test statistic remains smaller than the critical value, then a misspecification other than the ones described by the alternative hypotheses is thought to be present.

The size of the model error ∇ that can be detected in the Identification step with a probability γ, the so-called power of the test, is referred to as the Minimal Detectable Bias (MDB), [Salzmann, 1991]. The MDBs can be computed once two reference probabilities have been specified: the probability of rejecting the null-hypothesis when it is actually true (α_0) and the probability of rejecting the null-hypothesis when an alternative hypothesis is true (γ_0). The probabilities α_0 and γ_0 determine together with the power of the test γ,

the value of the non-centrality parameter λ_0. With this non-centrality parameter the MDBs can be expressed as

$$MDB = \sqrt{\frac{\lambda_0}{c_k^T Q_{v_k}^{-1} c_k}} \tag{1}$$

MDBs provide an important diagnostic tool for inferring how well particular model errors, such as outliers and cycle slips, can be detected. The MDBs are said to describe internal reliability of a system. In Section 2.1.3, the MDBs will be presented for different observation scenarios and measurement models.

2.1.2. Measurement and dynamic models

The approach to integrity monitoring presented in this paper is based on the philosophy that it should be applicable irrespective of the application(s) for which the data may originally have been collected. The integrity monitoring software does therefore not require any external information, like satellite and receiver positions, velocities and clock behaviour or information on the atmospheric effects. In this subsection, the Kalman filters' observation and dynamic models are derived that manage to realize the aim of this philosophy.

A total number of five basic observables are available from the GPS and GLONASS satellites: two carrier observations (L1, L2) and three code observations (P1, P2, C/A). In addition, for GPS with Anti-Spoofing switched on, a derived code observable, consisting of the difference between the encrypted P1 and P2 codes, may be available instead of the P2 code observation. The reconstructed P2 code observation is then obtained as the sum of the C/A and P2-P1 observations, resulting in (additional) correlation between the C/A and P2 code observation, see [Teunissen et al, 1998]. The receiver tracking scenarios considered in this paper are summarized in Table 1.

Assuming all five code and carrier observations C and ϕ are available at time k, the observation or measurement model for these five observations can be written as

$$\begin{pmatrix} C_1 \\ \phi_1 \\ \phi_2 \\ C_2 \\ C_{C/A} \end{pmatrix}_k = \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{pmatrix} L_k + \begin{pmatrix} 1 \\ -1 \\ -\kappa \\ \kappa \\ 1 \end{pmatrix} I_k + \begin{pmatrix} 0 \\ \lambda_1 N_1 \\ \lambda_2 N_2 \\ 0 \\ 0 \end{pmatrix} + \begin{pmatrix} \eta_{C_1} \\ \eta_{\phi_1} \\ \eta_{\phi_2} \\ \eta_{C_2} \\ \eta_{C_{C/A}} \end{pmatrix} \tag{2}$$

where $L = \rho + c(\delta^s t + \delta_r t) + \Delta T$ indicates the sum of the geometric distance ρ between satellite and receiver, the satellite and receiver clock errors, $\delta_r t$ and $\delta^s t$, multiplied by the speed of light c, and the combined effect ΔT of the biases due to the troposphere, orbital errors and SA; I is the first order L1 ionospheric effect; N_1, N_2 are the L1 and L2 carrier ambiguities; η are hardware delays (assumed constant for a satellite pass of several hours); λ_1, λ_2 are the wavelengths of the L1 and L2 carriers; the coefficient κ, finally, equals $\kappa = (\lambda_2/\lambda_1)^2$. For GPS, we have $\kappa \approx (9/7)^2$, whereas for Glonass $\kappa = (9/7)^2$ for all satellites.

For the ionospheric unknown I and the biased range term L, linear and quadratic models are introduced respectively. Substitution of these models into measurement model (2) yields

$$\begin{pmatrix} C_1 \\ \phi_1 \\ \phi_2 \\ C_2 \\ C_{C/A} \end{pmatrix}_k = \begin{pmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix} \begin{pmatrix} L \\ \dot{L} \\ \ddot{L} \end{pmatrix}_k + \begin{pmatrix} 1 & 0 \\ -1 & 0 \\ -\kappa & 0 \\ \kappa & 0 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} I \\ \dot{I} \end{pmatrix}_k + \begin{pmatrix} 0 \\ \lambda_1 N_1 \\ \lambda_2 N_2 \\ 0 \\ 0 \end{pmatrix} + \begin{pmatrix} \eta_{C_1} \\ \eta_{\phi_1} \\ \eta_{\phi_2} \\ \eta_{C_2} \\ \eta_{C_{C/A}} \end{pmatrix} \tag{3}$$

Table 1: Receiver tracking scenarios. Cn (n=1,2,3,4,5) refers to code-correlation, X4 to cross-correlation tracking.

	Carrier		Code				
No. of obs.	L1	L2	C/A	P1	P2	P2-P1	Identifier
2	×	-	×	-	-	-	C2
3	×	×	×	-	-	-	C3
4	×	×	-	×	×	-	C4
4	×	×	×	-	-	×	X4
5	×	×	×	×	×	-	C5

This is a singular model, i.e., not all parameters are estimable. The unknown parameters are therefore transformed according to

$$\begin{pmatrix} S \\ \dot{S} \\ \ddot{S} \end{pmatrix}_k = \begin{pmatrix} L \\ \dot{L} \\ \ddot{L} \end{pmatrix}_k + \begin{pmatrix} I \\ \dot{I} \\ 0 \end{pmatrix}_k + \begin{pmatrix} \eta_{C_1} \\ 0 \\ 0 \end{pmatrix} \tag{4}$$

$$\begin{pmatrix} I_{1,k} \\ I_{2,k} \\ I_{3,k} \\ \Delta \end{pmatrix} = \begin{pmatrix} I_k - \dfrac{\lambda_1 N_1}{2} \\ I_k - \dfrac{\lambda_2 N_2}{\kappa+1} \\ I_k \\ 0 \end{pmatrix} + \begin{pmatrix} -\dfrac{\eta_{\phi_1} - \eta_{C_1}}{2} \\ -\dfrac{\eta_{\phi_2} - \eta_{C_1}}{\kappa+1} \\ \dfrac{\eta_{C_2} - \eta_{C_1}}{\kappa-1} \\ \eta_{C_{C/A}} - \eta_{C_1} \end{pmatrix} \tag{5}$$

After applying these transformations the measurement model reads

$$\begin{pmatrix} C_1 \\ \phi_1 \\ \phi_2 \\ C_2 \\ C_{C/A} \end{pmatrix}_k = \begin{pmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix} \begin{pmatrix} S \\ \dot{S} \\ \ddot{S} \end{pmatrix}_k + \begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & -2 & 0 & 0 & 0 \\ 0 & 0 & -\kappa-1 & 0 & 0 \\ 0 & 0 & 0 & \kappa-1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \dot{I}_k \\ I_{1,k} \\ I_{2,k} \\ I_{3,k} \\ \Delta \end{pmatrix} \tag{6}$$

The corresponding covariance matrix of the observations will be denoted by R_k.

The above measurement model is valid for any number of receiver tracking scenarios. For example, if only L1 code and carrier observations are available, the parts which apply to the parameters I_2, I_3 and Δ are simply omitted from measurement model (6). In addition, it should also be noted that the carrier observations appear explicitly in model (6), rather than that carrier aiding of the code observations is performed. The problem of ionospheric divergence due to the opposite effect of the ionosphere on carrier and code observation is thereby avoided.

The dynamic or transition models for the range parameter L and the ionospheric parameter I are given as

$$\begin{pmatrix} L \\ \dot{L} \\ \ddot{L} \end{pmatrix}_k = \begin{pmatrix} 1 & \Delta t_{k,k-1} & \frac{1}{2}\Delta t_{k,k-1}^2 \\ & 1 & \Delta t_{k,k-1} \\ & & 1 \end{pmatrix} \begin{pmatrix} L \\ \dot{L} \\ \ddot{L} \end{pmatrix}_{k-1}$$

$$\begin{pmatrix} I \\ \dot{I} \end{pmatrix}_k = \begin{pmatrix} 1 & \Delta t_{k,k-1} \\ & 1 \end{pmatrix} \begin{pmatrix} I \\ \dot{I} \end{pmatrix}_{k-1} \tag{7}$$

with $\Delta t_{k,k-1} = t_k - t_{k-1}$. These dynamics are assumed to consist of zero-mean white noise processes with process noise matrices Q_k^L and Q_k^I, characterized by the spectral densities $q_S^{(3)}$ and $q_I^{(2)}$, respectively. The process noise matrix Q_k^S follows from applying the covariance law.

Measurement model (6) is very well suited for the generation of differential corrections, [Jin et al, 1995], [Jin, 1996]. The differential corrections can be derived from model (6) by eliminating the geometrical range ρ and the satellite clock error $c\delta^s t$ from model parameter S. The range can be computed from the known reference receiver coordinates and the satellite ephemeris transmitted in the navigation message, while the satellite clock parameter can be derived from correction parameters that are also included in this message. The term that results from S after elimination of range and satellite clock error reads $c\delta_r t + \Delta T + I + \eta_{C_1}$, which conforms the definition of a differential correction according to [RTCM SC-104, 1994].

It should be noted however, that in the case of GPS observations, measurement model (6) will only be valid for short time intervals Δt. The dithering of the GPS satellite clocks as part of the US policy of Selective Availability, will cause the parameter S and thereby also the differential corrections, to change quickly with time. For Glonass, such a restricted access policy is not used and model (6) will remain valid for longer time spans.

Another prerequisite for model (6) to be practically applicable is that parameter S changes smoothly with time. This is often hard to realize, due to the irregular behavior of the receiver clock. Hence, if the receiver is not connected to a stable clock, parameter S and its derivatives have to be eliminated from model (6). The simplest way achieve this, is by subtracting one observation from the other observations. If, for instance observation C_1 is chosen as pivot, we get

$$\begin{pmatrix} \phi_1 - C_1 \\ \phi_2 - C_1 \\ C_2 - C_1 \\ C_{C/A} - C_1 \end{pmatrix}_k = \begin{pmatrix} 0 & -2 & 0 & 0 & 0 \\ 0 & 0 & -\kappa-1 & 0 & 0 \\ 0 & 0 & 0 & \kappa-1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \dot{I}_k \\ I_{1,k} \\ I_{2,k} \\ I_{3,k} \\ \Delta \end{pmatrix} \tag{8}$$

Although the measurement model is slightly weakened by this sacrifice of an observation, it will be shown in the next subsection, that it is still possible with this model to find cycle slips as small as one cycle in the carrier observations, even for relatively large observation intervals Δt.

In spite of careful site selection, data collected by reference stations may still exhibit excessive multipath effects. The implemented integrity monitoring procedures therefore allow modeling of code multipath in order to suppress its effect on both testing results as well as on the derived products like differential corrections and estimated ionospheric delays.

For the purpose of modeling code multipath, the measurement model of the code observations, is extended with a linear multipath model

$$\begin{pmatrix} C_1 \\ C_2 \\ C_{C/A} \end{pmatrix}_k = \begin{pmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix} \begin{pmatrix} L \\ \dot{L} \\ \ddot{L} \end{pmatrix}_k + \begin{pmatrix} 1 & 0 \\ \kappa & 0 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} I \\ \dot{I} \end{pmatrix}_k + \begin{pmatrix} \eta_{C_1} \\ \eta_{C_2} \\ \eta_{C_{C/A}} \end{pmatrix} + \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} M_1 \\ \dot{M}_1 \\ M_2 \\ \dot{M}_2 \\ M_{C/A} \\ \dot{M}_{C/A} \end{pmatrix}_k \quad (9)$$

Introduction of these additional bias parameters requires, another transformation in addition to (4) and (5),

$$S' = S + M_{1,0} \quad (10)$$

$$\begin{pmatrix} I'_{1,k} \\ I'_{2,k} \\ I'_{3,k} \\ \Delta' \end{pmatrix} = \begin{pmatrix} I_{1,k} \\ I_{2,k} \\ I_{3,k} \\ \Delta \end{pmatrix} + \begin{pmatrix} \frac{M_{1,0}}{2} \\ \frac{M_{1,0}}{\kappa+1} \\ \frac{M_{2,0}-M_{1,0}}{\kappa-1} \\ -M_{1,0} \end{pmatrix} \quad (11)$$

$$\begin{pmatrix} \Delta M_1 \\ \Delta M_2 \\ \Delta M_{C/A} \end{pmatrix}_k = \begin{pmatrix} M_1 \\ M_2 \\ M_{C/A} \end{pmatrix}_k - \begin{pmatrix} M_1 \\ M_2 \\ M_{C/A} \end{pmatrix}_0 \quad (12)$$

The extended measurement models (6) and (8) including code multipath, follow from substituting (10)-(12). The dynamic model for the code biases is similar to that of the ionospheric effects, see (7) and characterized by the spectral density $q_M^{(2)}$.

As can be seen from (12), not the multipath effects themselves, but only their changes with respect to some reference value can be estimated, [Jin, 1996], [de Jong, 1996]. Moreover, it should also be noted that for receiver tracking scenario C2, it is not possible to model code biases and estimate them as separate parameters.

2.1.3. Design computations

In this subsection the effectiveness of the first step of the integrity monitoring software will be analyzed by means of the Minimal Detectable Bias (MDB)-measure introduced in the subsection on testing and reliability. MDBs will be given for the models derived in the previous subsection and the five receiver tracking scenarios of Table 1. Default parameters used for the computations are given in Table 2 and are based on the values given in [Jin, 1996] and [de Jong, 1996].

In Figures 2 and 3 the MDBs are presented for the measurement models (6) and (8), i.e. the models with and without a stable receiver clock respectively. Judging by the figures, the MDBs for both models are equivalent, except for scenario C2. Hence, if dual-frequency carrier observations are available, then it is possible to find slips as small as one cycle in the carrier observations for both model (6) as well as model (8).

Several other observations can be made based on Figures 2 and 3. It can for instance be seen that the code MDBs are approximately equal to $\sigma_{Code}\sqrt{\lambda_0}$ or, using Table 2, about four times their standard deviation. Another general conclusion that can be drawn from the figures is that additional code observations do not result in a significant decrease of the MDBs, or in other words single-frequency code observations are sufficient to ensure the internal reliability of the observations. In contrast, the availability of dual-frequency carrier observations does result in a significant improvement of the internal reliability compared to the case only single-frequency carrier observations are available.

Table 2: Default parameters for the computation of MDBs

Standard deviations (m)		Spectral densities		Testing parameters	
ϕ_1	0.003	$q_I^{(2)}$	10^{-8} m^2/s^3	α_0	0.001
ϕ_2	0.003	$q_S^{(3)}$	10^{-5} m^2/s^5	β_0	0.80
C_1	0.3	$q_M^{(2)}$	10^{-4} m^2/s^3	λ_0	17.05
C_2	0.3				
$C_{C/A}$	0.3				
$C_{P2\text{-}P1}$	0.4(*)			Obs. interval	1 s

(*)St. dev. σ computed as: $\sigma_{C_{P2-P1}} = \sqrt{\sigma_{C_1}^2 + \sigma_{C_2}^2}$

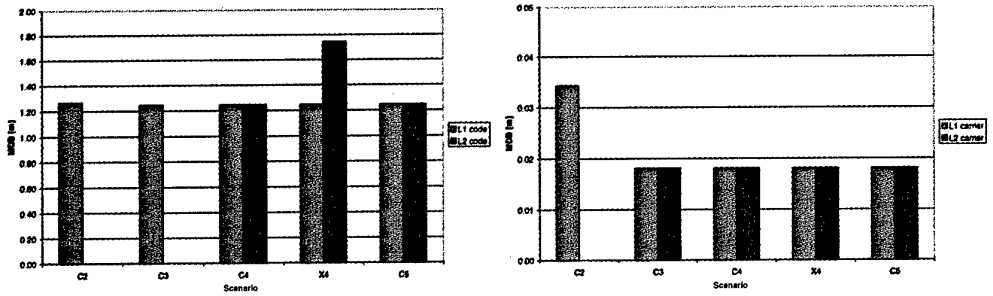


Figure 2: L1 code (left) and carrier MDBs for model (6) for the receiver tracking scenarios of Table 1

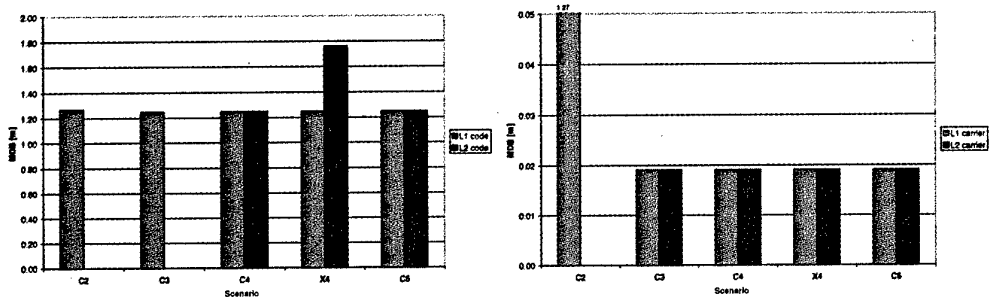


Figure 3: L1 code (left) and carrier MDBs for model (8) for the receiver tracking scenarios of Table 1.

For measurement model (8), the dual-frequency carrier MDBs were also computed for observation intervals Δt up to 60 seconds. The results shown in Figure 4 indicate that even for these long observation intervals, the carrier MDBs are small enough to allow for the identification of even the smallest cycle slip.

MDBs were also computed for measurement model (8), extended with code multipath parameters, according to the models of (9)-(12). Comparing Figures 3 and 5 it can be concluded that the carrier MDBs are not affected by the introduction of the additional parameters, whereas the code MDBs are. The general conclusion on the influence of additional code and carrier observations, drawn above, still holds.

2.2. Step 2: validating the satellite ephemeris and clock parameters

In the first step of the integrity monitoring approach, the code and carrier phase observations are validated. For real time navigation applications however, one has to make use of parameters of the satellite transmitted navigation message, i.e., satellite ephemeris and clock correction parameters, in addition to the observations. Hence, these parameters need to be validated as well, particularly as they are known to contain gross errors on occasion.

The second step of the integrity monitoring is dedicated to the validation of the parameters in the navigation message, using the validated observations from the previous step. Although this step has as yet not been implemented in the integrity monitoring software, procedures to carry out the validation have already been designed. These procedures will briefly be described in this section.

With the satellite ephemeris and clock parameters and the reference station's known position, the distance between satellite and receiver can be computed. The difference between observed (and validated) and computed code range consists of the receiver clock bias, tropospheric and ionospheric effects and biases due to orbit errors and SA. The tropospheric and ionospheric effects can be removed using one of the many available models and by forming a linear combination of the L1 and L2 code observations. The most significant remaining term (for GPS) is the effect due to SA. It is expected that this bias is small compared to biases due to errors in the broadcast parameters. If two satellites are tracked, it is possible to detect errors, if more than two satellites are tracked, it is possible to also identify the satellite(s) with the erroneous navigation data. Detection and identification of biases is again based on statistical hypothesis testing, like those included in the DIA-procedure.

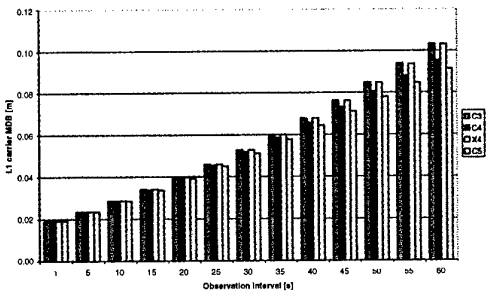


Figure 4: L1 carrier MDBs as a function of the observation interval for model (8) for the receiver tracking scenarios of Table 1

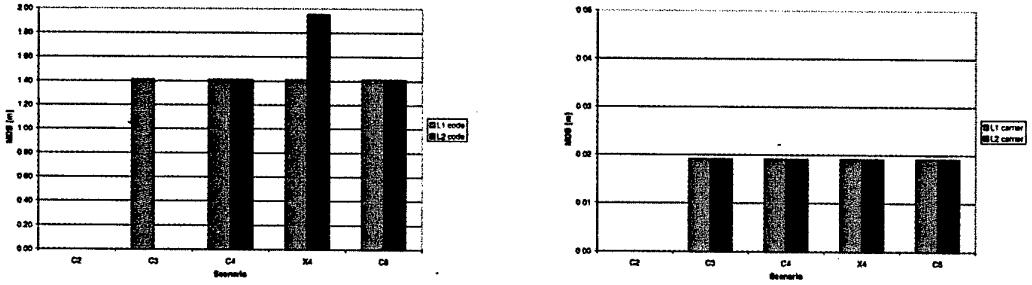


Figure 5: L1 code (left) and carrier MDBs for model (8), *extended with a model for code multipath, eqs (9)-(12), for the receiver tracking scenarios of Table 1.*

3. Integrity monitoring results

In this section, the error detection capabilities of the integrity monitoring software are illustrated by means of some examples. In addition, the beneficial effect of modelling code multipath is demonstrated. Finally, attention will also be paid to a byproduct of the integrity monitoring software, i.e., estimates of the ionospheric delay.

The results in this section were obtained using data from a JPS Legacy GPS/Glonass receiver, located at the IGEX tracking station in Delft. The JPS Legacy is a twenty channel dual-frequency receiver, which, for the purpose of IGEX, is currently operating at an observation interval of 30 seconds. Data analyzed applies to a single day of IGEX tracking, in this case 31 March 1999.

3.1. Error Detection

In Figure 6, the overall model test statistic of the Detection step in the DIA-procedure is shown for a complete pass of a GPS and a Glonass satellite. It can be seen from the figure that the statistic remains well below the critical value of 3.4 for the majority of the observation epochs. However, it can also be seen that the statistic shows a distinct dependence on the satellite elevation, i.e. the statistic tends to decrease with increasing satellite elevation. Consequently, a number of model errors is falsely detected at the beginning and end of the satellite passes.

The dependence of the overall model test statistic on the satellite elevation is due to the well-known phenomenon of deteriorating code and, to a lesser extent, carrier precision with decreasing satellite elevation, [Jin, 1996].

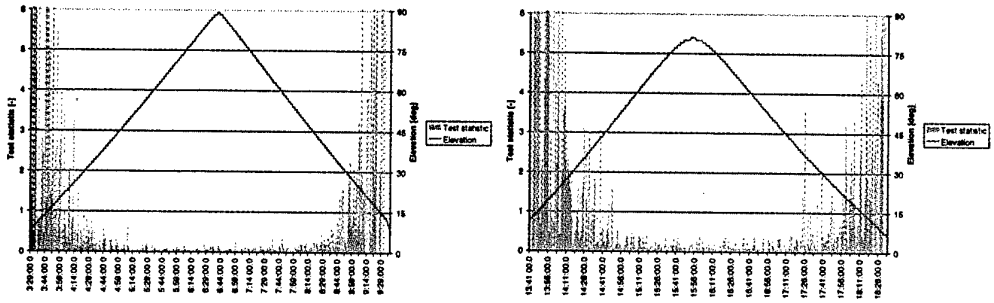


Figure 6: Overall model test statistics and elevation for GPS satellite 9 (left) and Glonass satellite (slot number) 3 (Delft, 31 March, 1999).

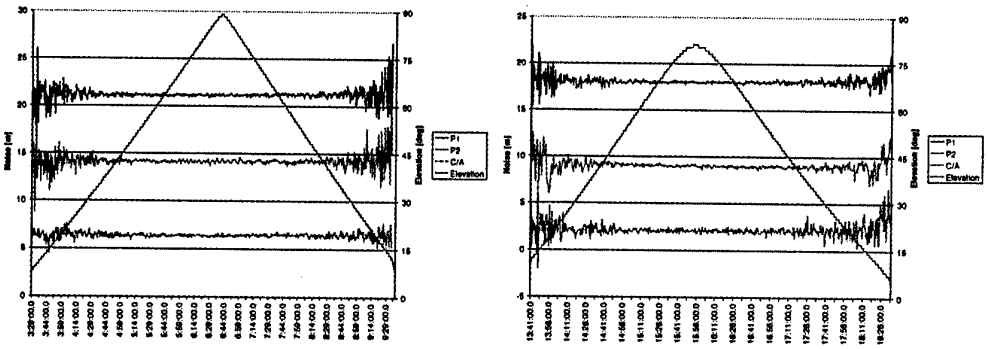


Figure 7: Precision of P1, P2 and C/A code observations for GPS satellite 9 (left) and Glonass satellite (slot number) 3 (Delft, 31 March, 1999).

The phenomenon can illustrated by considering the following linear combinations of one code and two carrier observations

$$M_{L1,k} = C_{L1,k} - \frac{f_1^2 + f_2^2}{f_1^2 - f_2^2}\phi_{L1,k} + \frac{2f_2^2}{f_1^2 - f_2^2}\phi_{L2,k}$$
$$M_{L2,k} = C_{L2,k} - \frac{2f_1^2}{f_1^2 - f_2^2}\phi_{L1,k} + \frac{f_1^2 + f_2^2}{f_1^2 - f_2^2}\phi_{L2,k} \qquad (13)$$

As all time-dependent effects (receiver-satellite distance, clocks, atmosphere, orbits, SA) are eliminated from these combinations, they are mainly a function of the precision of the code observations.

The results of forming the linear combinations for C/A, P1 and P2 code observations are in shown in Figure 7. The elevation dependence of the code precision can clearly be distinguished for both GPS and for Glonass. However, it can also be seen from this figure that the precision of the Glonass P1 and P2 code observations is generally better than for GPS, which is understandable, since the Glonass P codes are not encrypted. For the C/A code however, GPS performs better, which is due to the higher chipping rate of the GPS C/A code as compared to Glonass.

The elevation dependence of the overall model test statistics can be suppressed by implementing an elevation dependent weighing of the code precision in the observation model. However, although such a weighing is possible in the integrity monitoring software, it requires careful tuning, particularly as the weighing function tends to depend strongly on the receiver make and type. In the examples presented in this section, no weighing was applied.

In order to appreciate the performance of the DIA-procedure, a cycle slip of one cycle was added to the L1 carrier observations of GPS satellite 9, starting at 6:07:00 (GPS time). Data was processed using only a dynamic model for the ionospheric effects; code multipath was not accounted for. Shown in Figure 8 are the overall model test statistics of the central part of the satellite pass. A model error is clearly detected. Moreover, in the resulting identification step a cycle slip was correctly identified as the cause of the model error, see also Table 3. From 6:07:00 onwards, the adaptation step of the DIA-procedure automatically corrected the L1 carrier observations by subtracting the estimated integer slip. The results of a similar test in which an outlier of 3 meter was added to the P1 code observation is also presented in Table 3.

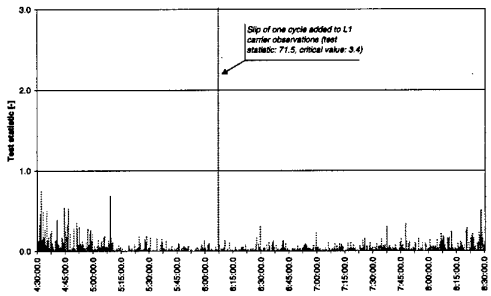


Figure 8: Overall model test statistics of GPS satellite 9 (Delft, 31 March, 1999); slip of one cycle added to L1 carrier data at 06:07:00.

Table 3: Example of performance of the dual-frequency integrity monitoring functions: in the first processing run, a slip of one cycle was added to the L1 carrier data of GPS satellite 9, in the second run, an outlier of three meters was added to the P1 code data.

Time		31/03/99 6:07:00-cont	31/03/99 6:07:00
Type of bias		L1 cycle slip of 1 cycle	P1 outlier of 3 m
Critical value overall test	3.4		
Critical value LS test	3.3		
Detection	Overall model test statistic	71.5	7.73
Identification	Alternative hypothesis	LS test statistic	LS test statistic
	L1 cycle slip	***16.9***	0.14
	L2 cycle slip	16.8	0.04
	P1 outlier	1.5	***5.53***
	P2 outlier	1.6	0.36
	C/A outlier	1.5	0.44
Adaptation	Estimated bias	1.02 cycle	2.80 m
	Cycle slip after rounding	1.00 cycle	-

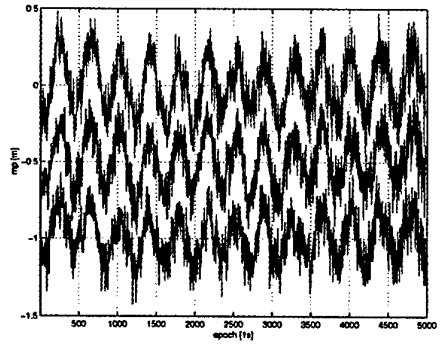


Figure 9: P2 code multipath characteristics, according to (13) for GPS satellite 6 for three consecutive days in Delft, March 1999.

3.2. Multipath modeling

In order to illustrate the beneficial effect of modelling code multipath in the integrity monitoring software, data was collected at a site, notorious for its multipath susceptibility. This location does not coincide with the IGEX reference station, although it is close to it. Moreover, a 1-second rather than a 30-second observation interval was used while collecting the data.

In Figure 9 the results of forming the linear combination (13) for the P2 code observations of GPS satellite 6 are shown for three consecutive days. The repeatability of the oscillations clearly indicates the presence of multipath in the data.

The data of GPS satellite 6 was processed using both the nominal measurement model (8) and an extended version of the model, i.e., with (9)-(12) included to model code biases. In Figure 10 (left) the computed overall model statistics of the first processing run are depicted together with the P2 code multipath estimated in the second processing run. The figure clearly shows the strong correlation between the test statistic and the code multipath. Such correlation is virtually absent in Figure 10 (right), in which the overall model test statistic of the second processing run are shown together with the P2 code multipath estimates. This illustrates that the extended measurement model can cope with even the severest code multipath conditions.

3.3 Ionospheric delay estimation

It follows from Section 2.1 that the integrity monitoring routines provide estimates for the ionospheric delays. Delay I_3, together with the delay, computed from the P1 and P2 code observations, are shown in Figure 11, for a GPS and a Glonass satellite. As can be seen from this figure, the estimated, filtered delays are much more precise than those computed from the code data alone. They are therefore particularly suitable for the generation of wide-area differential ionospheric delay corrections, as is for instance envisaged in the GPS/Glonass augmentation System EGNOS, see [Benedicto et al, 1998] and the GPS augmentation system WAAS, see [Enge et al, 1996].

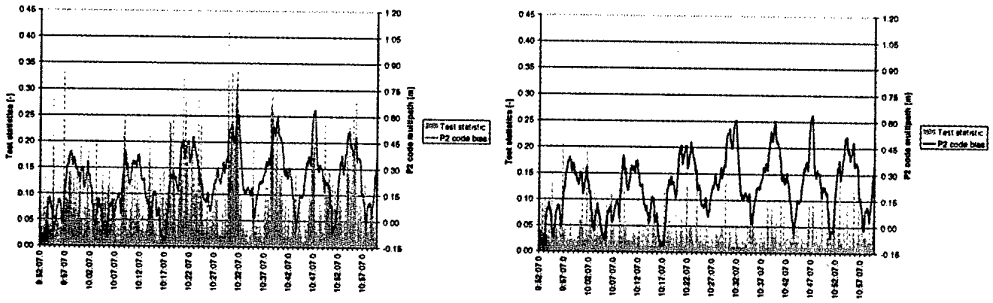


Figure 10: Overall model test statistics resulting from processing with model (8) (left) and with model (8), extended with (9)-(12), to model code multipath. The estimated P2 code multipath is shown in both figures as reference.

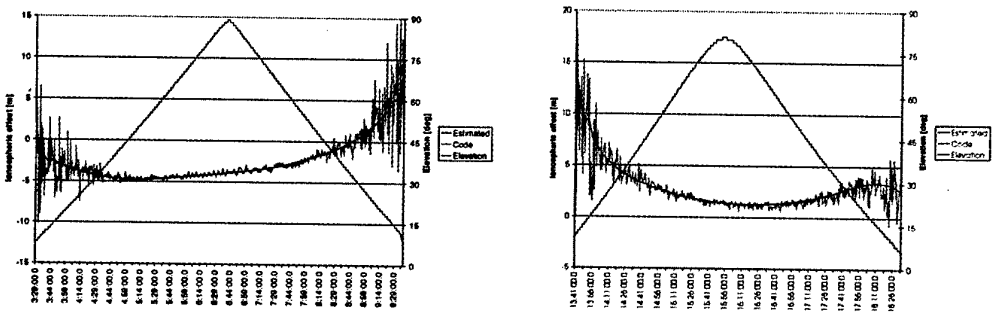


Figure 11: Estimated and computed ionospheric effect and elevation for GPS satellite 9 (left) and Glonass satellite (slot number) 3 (Delft, 31 March, 1999).

4. Conclusions and future work

In this contribution a description was given of the permanent GPS/Glonass reference station in The Netherlands. The contribution centered on the integrity monitoring software running at the station. The software is based on the recursive Detection, Identification and Adaptation (DIA-) procedure developed at DUT for the validation of observations in real-time. Both a description of the DIA-procedure in general and its implementation in the integrity-monitoring software in particular were given, as well as some examples of the software's error detection capabilities.

In the near-future, the integrity monitoring software of the reference station will be extended with procedures to detect and identify anomalies in the satellite transmitted navigation message, in addition to errors in the code and carrier phase observations. The algorithms are also based on a DIA-like approach, but have as yet to be implemented in the software.

The now established reference station will serve as a stepping stone for further research. One area of research is to explore and exploit the capabilities of the integrity monitoring software for generating differential GPS and Glonass corrections. Plans to use the software within the context of the hybrid DGPS/LORAN navigation system EUROFIX, see [van Willigen et al, 1998], are in an advanced stadium. This is a particularly promising development as the department of Electrical Engineering of DUT, at which the EUROFIX-concept was first conceived, is looking to extend the concept to Glonass and Chayka, the Russian equivalent of LORAN.

In addition to the research on differential positioning, research will also be started on the of combined use of Glonass and GPS for precise relative positioning. This is a particularly challenging area of research as the Glonass signal design differs somewhat from the GPS signal design. The resulting differing characteristics of the Glonass observations require careful attention, particularly if a seamless integration with GPS observations is aimed at.

The research on differential and relative positioning requires an expansion of the current single reference station into a network of stations. The first step towards establishing such a network has already been made by acquiring a second dual-frequency GPS/Glonass receiver which will be co-located with one of the stations of the already existing GPS array in The Netherlands, the Active GPS Reference System (AGRS.NL). Ultimately, a completely integrated GPS/Glonass array is envisaged, but the time frame in which this will be achieved, will to a large extent depend on the continued development of Glonass.

Acknowledgments

The support of the Survey Department of the Ministry of Transportation and Public Works is gratefully acknowledged.

References

Benedicto, J., P. Michel, J. Ventura-Traveset (1998): *EGNOS: Project Status Overview.* Proceedings 2nd European symposium on Global Navigation Satellite Systems GNSS 98, Toulouse, vol. 1, paper no. II-0-01, 10 pp.

Enge, P.K., A.J. Van Dierendonck (1996): *Wide Area Augmentation System.* In: Progress in Astronautics and Aeronautics, vol 164, 'Global Positioning System: Theory and Applications', (Parkinson and Spilker Eds.), vol II, pp. 117-142.

Jin, X.X., H. van der Marel, C.D. de Jong (1995): *Computation and quality control of differential GPS corrections.* Proceedings ION GPS-95, Palm Springs, pp. 1071-1079.

Jin, X.X. (1996): *Theory of carrier adjusted DGPS positioning approach and some experimental results.* PhD thesis, Delft University Press, Delft, ix+163 pp.

de Jong, C.D. (1996): *Principles and applications of permanent GPS arrays.* PhD thesis, Department of Geodesy, Technical University of Budapest, vii+96 pp.

Kalman, R.E. (1960): *A new approach to linear filtering and prediction problems.* Journal of Basic Engineering, Transactions of the ASME, Vol. 82, pp. 35-45.

Salzmann, M.A. (1991): *MDB: A design tool for integrated navigation systems.* Bulletin Géodésique, Vol. 65, No. 2, pp. 109-115.

RTCM SC-104 (1994): *RTCM recommended standards for differential NAVSTAR GPS service.* Radio Technical Commission for Maritime Services, Washington.

Teunissen, P.J.G. (1990): *An integrity and quality control procedure for use in multi sensor integration.* Proceedings ION GPS-90, Colorado Springs, pp. 513-522.

Teunissen, P.J.G., C.C.J.M. Tiberius, N.F. Jonkman, C.D. de Jong (1998): *Consequences of the cross-correlation measurement technique.* Proceedings 2nd European symposium on Global Navigation Satellite Systems GNSS 98, Toulouse, vol. 2, paper no. IX-0-11, 6 pp.

van Willigen, D., G.W.A. Offermans, A.W.S. Helwig (1998): *EUROFIX: Definition and current status.* Invited paper, IEEE 1998 Position, Location and Navigation (PLANS) Symposium, Palm Springs.

Inertial Navigation Systems in Geodetic Application: L.I.G.S. experience

Prof. Oleg S. Salychev, Dr. Vladimir V.Voronov, Dr. Vadim V.Lukianov
The Laboratory of Inertial Geodetic Systems
The Bauman Moscow State Technical University
5, 2-nd Baumanskaya st., Bauman MSTU, IU-2, 107005 Moscow, Russia
e-mail : ligs@glasnet.ru

April 15, 1999

Summary

Different applications of inertial technology to the Geodesy and Geophysics is considered: from precise positioning to airborne gravimetry. Actually, the materials of inertial navigation systems (INS) applications and testing is based on the long time experience got by the Laboratory of Inertial Navigation Systems (LIGS) at the Bauman Moscow State Technical University in the field of the real-time navigation and inertial geodesy. The wide spectrum of the testing results obtained in the different countries where our equipment was applied is considered and interpreted. [1, 2]

1 Precise positioning

Let consider at first the one of the oldest stand alone INS application to the surveying. In this case the INS is installed into the truck which periodically stops at the regular interval in order to carry out zero velocity update (ZUPT). At this moment the INS indication in the stop is proportional to the INS's velocity error. The software task for this case is to restore the velocity error behavior between neighboring stops and using above information to compensate the errors in output of INS. High precision inertial platform ITC-2 was used as an equipment for this purpose. The ITC-2 includes high precise sensors which are installed in a compact gyroplatform. It implements three-axes-four gimbal platform equipped with two floated gyroscopes and three accelerometers. Moreover, the ITC-2 includes on-board high productive computer to solve the navigation problem and standard

Paper presented at the RTO SCI International Conference on "Integrated Navigation Systems", held at "Elektropribor", St. Petersburg, Russia, 24-26 May 1999, and published in RTO MP-43.

PC computer with powerful software for the real-time filtering and smoothing of system errors.

1.1 ITC–2 Hardware Configuration

It consists of an inertial surveying system (ISS), contained a Russian precise gimbal INS, navigation indicator, control unit (CU), and optionally GPS receiver and altimeter which can be linked to **ITC–2**.

Moreover for gravity determination the special gravimeter with new electronics has been installed.

1.2 Software

Special adaptive real-time scalar algorithm has been developed to take into account the ZUPT measurements. Above algorithm has following advantages:

— low sensitivity to a priori information on the INS error model;

— backward time restoring of INS error in real-time calculation procedure;

— high precision of coordinates estimation in case of large time interval between ZUPT measurements (10 minutes).

Another type of scalar approach is utilized to smooth the surveyed points by exclusively using the position updates.

Moreover, non-statistical (wave) estimation algorithm can be realized optionally in **ITC–2**.

Besides, **ITC–2** includes software for INS/GPS integration. Wave algorithms with high prediction properties could be used in case when cycle slip or losses of lock occurred.

Special software has been designed for gravity anomalies and deflection of the vertical determination in vehicle semi-kinematic and kinematic modes as well as for airborne gravity applications.

1.3 Performances

The main performance characteristics of **ITC–2** are summarized in the Table 1.

1.4 Field Testing Results

More than 20 runs of survey vehicle were carried out in urban area (Moscow-city) for the **ITC–2** testing. The maximum real time errors of estimation position was $1.9 \div 2.2$ meters. Smoothing procedure provides the accuracy $0.4 \div 0.6$ meters.

In order to check the ITC–2 real-time accuracy in case of large time interval between ZUPT's (10 minutes), the series of survey runs were carried out in Beijing environs

Table 1: *ITC–2 Performance characteristics*

Attitude	0,7 mrad	Yaw, pitch, roll
Horizontal position	$9 \cdot 10^{-5} \times$ HD RMS	Real time
Altitude	$9 \cdot 10^{-5} \times$ HD RMS	Real time
Horizontal position	$3 \cdot 10^{-5} \times$ HD RMS	After position update
Altitude	$3 \cdot 10^{-5} \times$ HD RMS	After position update
Alignment	15 minutes	
ZUPT	$<$ 20 sec	
Output rate	10 Hz	
Power consumption	$<$ 800 W	Cold start;
Size	$710 \times 280 \times 295$	1.5 kW during 3 min
Weight	45 kg	

HD — is horizontal distance between last position update and the present position.

(China). The initial part of the traverse (about 10 km) was used for the field azimuth calibration. After the calibration, the ITC–2 calculated the position coordinates in the real time. In these tests the time between neighboring ZUPT was 10 minutes. The accuracy of 10 m (CEP) was achieved in considered tests.

Four survey vehicle runs of ITC–2 system were carried out in Kananaskis mountain area (Alberta, Canada).

The plots on Fig. 1, Fig. 2 summarize the coordinates errors in real-time data processing for each run after the field azimuth calibration. In this case the time between ZUPT's was 3 minutes. The obtained results show the perfect accuracy of ITC–2 real-time position calculation. The results of Ferranti system application [1] on the same traverse are represented on Fig. 3.

The comparison between two system test results shows that the ITC–2 has more high accuracy than the Ferranti system. There are two reasons to explain this effect. The first one connected with the utilization in ITC–2 scalar estimation algorithm used the real-time restoring (not only filtering) calculation scheme instead Kalman filter (Ferranti system).

The second reason is based on the accurate azimuth misalignment compensation (better than 30 arc sec) due to the special field azimuth calibration procedure.

2 Determination of Deflection of the Vertical

[1] Data was received from The University of Calgary.

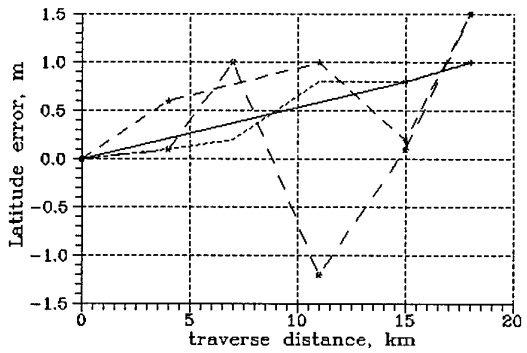


Figure 1: ITC–2 real-time positioning errors

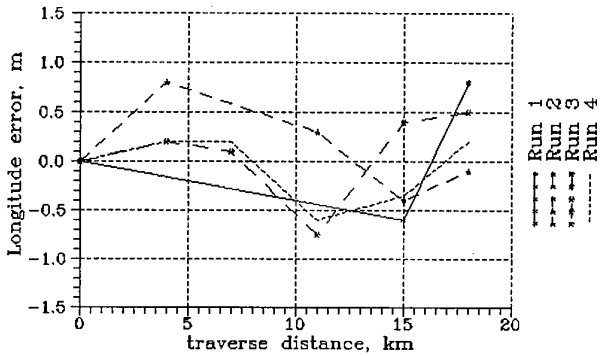


Figure 2: ITC–2 real-time positioning errors

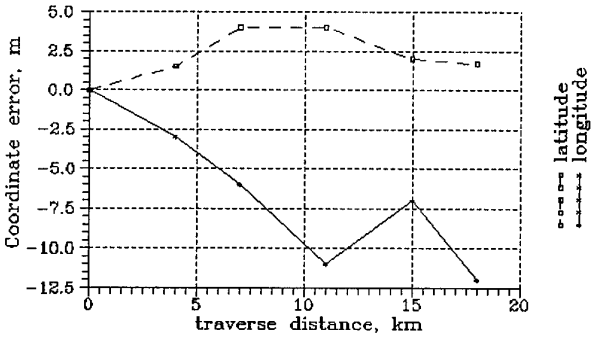


Figure 3: Ferranti real-time positioning errors

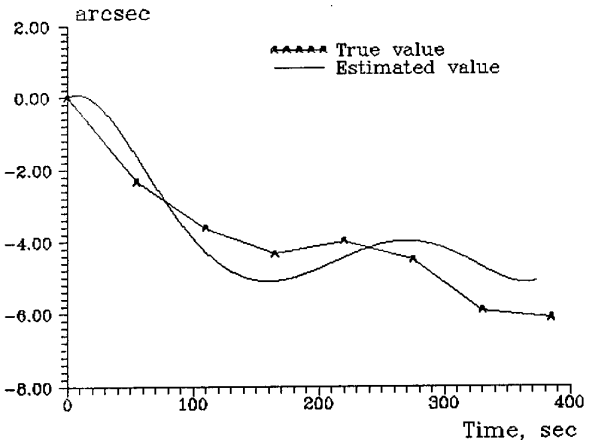


Figure 4: Testing results of deflection of the vertical estimation

Two different ways can be suggested for the determination of the deflection of the vertical depending on the vehicle motion behaviour.

The first procedure is a semi-kinematic application, when GPS data are not used. Only INS velocity measurements at ZUPT points are needed. As is well know, the main problem of the vertical deflection precise estimation is tp separate two signals: deflection increments and gyro drift rates. For this purpose the special preliminary drift calibration has to be applied. Moreover, the INS's gyro sensors must be extremely stable.

The scheme of the semi-kinematic procedure is illustrated in Fig. 5.

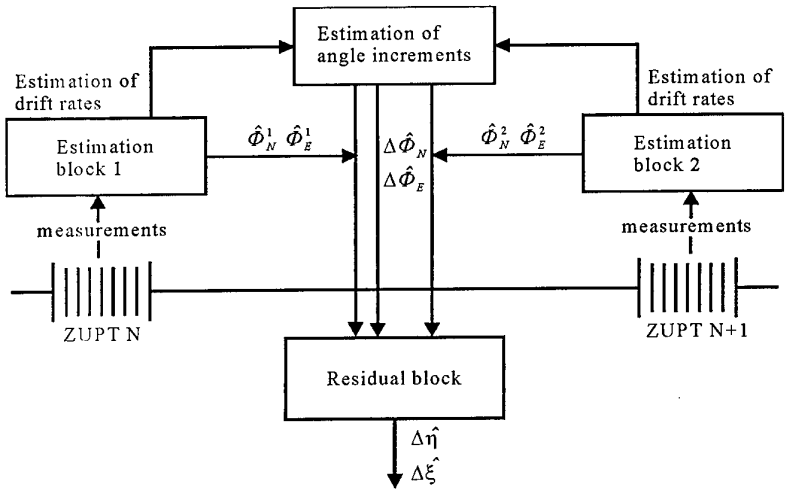


Figure 5: Scheme of vertical deflection determination in semi-kinematic mode

At each stop measurements of INS horizontal velocities are used for the estimation of the angle between the platform and the local-level frames as well as deflection of

the vertical. Comparing the above estimates in neighboring stops and evaluating the angular increments caused by INS velocity errors and drift rates, it is possible to determine estimates of deflection increments for corresponding points. A series of test runs has been executed for vertical deflection determination in the Kananaskis mountains area using a truck as a carrier. Typically the time interval between neighboring stops did not exceed 5 minutes. The results of the above tests are illustrated in Table 2.

Hence, the performed tests indicate than an accuracy of better than 1 arc sec is achievable typically over time intervals of 5 minutes (4 km) using stand-alone INS.

Table 2: *Results of Vertical Deflection Estimation in Semi-kinematic Mode*

	RMS of vertical deflection estimation errors		
Run #	Length (km)	η (arc sec)	ξ (arc sec)
1	4	0.4	0.80
2	4	0.6	0.84
3	4	0.7	0.95

For the kinematic mode application, GPS measurements are needed. In this case the difference between INS and GPS velocities is used during vehicle motion. The main problem in estimating vertical deflection lies in the separation of the useful signal from the increments in gyro drift. Total time of each vehicle run was 7 to 10 minutes. The variation of the calibrated gyro drift during the vehicle mode did not exceed $1 \div 2 \times 10^{-3}$ deg/hour. Drift aberrations lead to estimation errors of $0.8 \div 1$ arc sec in the deflection at the end of the motion interval. Another problem encountered in kinematic gravity determination were the large difference between INS and GPS velocities due to uncompensated nonlinearity of accelerometer scale factor influence. In order to compensate above nonlinearity influence the special calibration procedure was designed and used. The suggested processing method includes multiple smoothing procedure. First, the smoothing algorithm used previously, was reformulated as a forward-backward time algorithm for estimating the useful signal with minimum distortion. For this purpose, the total measurement interval was subdivided into periods of 50-60 seconds. A forward estimation was applied which provided estimates of the useful signal at the end of each period. Using these estimates, it is possible to obtain estimates for the total period by backward smoothing. These estimates still contain errors caused by the high level of measurement noise. To reduce these errors a second and third smoothing step is necessary. They use a priori knowledge of the frequency behavior of signal and noise, to separate the useful signal from the measurement noise. For the actual kinematic tests an Ashtech GPS receiver was used. Data acquisition and collection during the run was done by Portable PC-computers. Average carrier speed was 60 km/hour.

The results of the Russian INS (ITC-2) kinematic tests for vertical determination are

shown on Fig. 4. The results are promising, indicating that accuracies of 1 arc sec and 1 mGal can be achieved over the time interval mentioned above.

3 Airborne gravimetry

3.1 Principal Model of Airborne Gravimetry

The principle of airborne gravimetry is based on Newton's equation of motion in the gravitational field of the Earth. Using the general navigation equation in projections on the local-level frame, the expression for the gravity anomaly can be defined as:

$$\delta g = -\dot{V}_u + f_{up} + (\frac{V_E}{R_1 + h} + 2U\cos\varphi)V_E + \frac{V_N^2}{R_2 + h} - g_n \qquad (1)$$

where:
$\dot{V}_{up}$ — vertical acceleration of the vehicle;
f_{up} — vertical projection of specific force;
V_E, V_N — vehicle velocity in projections on the local-level frame;
g_n — normal gravity vector;
U — angular rate of the Earth;
R_1, R_2 — prime vertical and meridian radii of curvature.

On the basis of equation (1), it is possible to estimate the gravity anomaly using information from an inertial navigation system (INS) and differential GPS -satellite techniques(DGPS). Thus, the information on $\dot{V}_u$, V_E, V_N, φ can be obtained from GPS, while the inertial unit provides the specific force measurement - f_{up}. Obviously, the GPS measurements must be recalculated to the center of mass of the inertial unit. To accomplish this, the attitude measurements of the INS can be used.

The error model of airborne gravimetry can be defined taking into account the individual errors of all parameters included in the equation (1). Thus, the significant errors affecting airborne gravimetry are given by

$$d\delta g_u = a_E(\Phi_N + \mu_N) - a_N(\Phi_E + \mu_E) + \delta f_u + S(f_u) + \delta\dot{V}_u + f(dT) \qquad (2)$$

where:
Φ_N, Φ_E — attitude errors of the INS;
μ_N, μ_E — nonorthogonality errors of vertical accelerometer installation;
δf_u — random noise of the vertical accelerometer;
$S(f_u)$ — nonlinearity of accelerometer scale factor;
$\delta\dot{V}_u$ — error of GPS-derived accelerations;
$f(dT)$ — time synchronization errors.

Let us analyze possible ways to eliminate the influence of different factors on the estimation accuracy of δg. The attitude errors — Φ_N, Φ_E are constant or slowly

changing over a period of time; they can be estimated with sufficient accuracy if the aircraft accelerations can be measured with a standard deviation of about 5-6 mGal. The nonorthogonality of the installation — μ_N, μ_E can be estimated and compensated using laboratory testing with different inclinations of the gyroplatform. δf_u, $\delta \dot{V}_u$ — can be removed by the filtering procedure because these errors have a negligible values in the frequency range of interest. The nonlinearity of scale factor $S(f_u)$ can be eliminated using a special calibration procedure.

3.2 Gravity survey hardware

The hardware equipment consists in the inertial survey system. The basic system applied for the project realization is a Russian inertial navigation system I-21, which originally is designed to be installed in the heavy aircraft. The I-21 includes highly precise sensors which are installed in a compact gyro platform. It implements a three-axis-four-gimbals platform equipped by two floated gyroscopes and three floated accelerometers. The I-21 is equipped by two built-in high- speed processors for the navigation solution. The digital output of the INS data is accepted by a standard PC/AT computer and provides the navigation parameters: two coordinates, two velocities, four angles (pitch, roll, heading, gyro yaw), and INS time as well. However, in order to apply the navigation system for surveying the original hardware was considerably modified in part of the electronics and interface. The ISS equipment set is combined by the following modules:

- inertial unit;
- power supply unit;
- indication and control units;
- block of the gravimeter electronics;
- set of cables and wires.

The total weight of the equipment set does not exceed 30 kg. The power consumption in warm up mode (10 min) is less then 1,5 kW, in regular navigation regime less then 0,75 kW. The PC/INS data communication is scheduled by means of special signals, which can be accepted directly into the ISA bus and interpreted as a hardware interruption (10 PPS). The form and edge of this signal are identical to those of the GPS 1PPS, that is used for precise INS/GPS data synchronization. In order to be applied for the gravimetry purpose a highly sensitive gravimeter-accelerometer has been designed, manufactured and installed on the INS gyro stabilized platform. This unit is completely independent of any other INS hardware in terms of its power and data output. The available gravimeter output rate is 1 Hz. Data are accepted by PC through serial port. The dimensional and power features of the ISS are enough to fit it into the light airplane like Cessna 310.

During the tests in the 1994-1997 the ISS had proved its reliability. We never had any ISS caused failure on several thousand flown kilometers. In combination with perfect data processing algorithm the system is able to produce almost 100flight conditions). However, the system is rather tolerant with respect to the weather.

3.3 Test results

Series of test flights carried out on different types of aircraft, over different areas and under the different weather conditions have demonstrated that the accuracy achieved is about 1mGal per 3km or better. These test flights were performed during a period from 1993 to 1997 and covered more than 50.000 flown lines-kilometers. The first flight series consisted in the flying of the repeated line only and had a goal to prove the equipment reliability and the quality and repeatability of results.

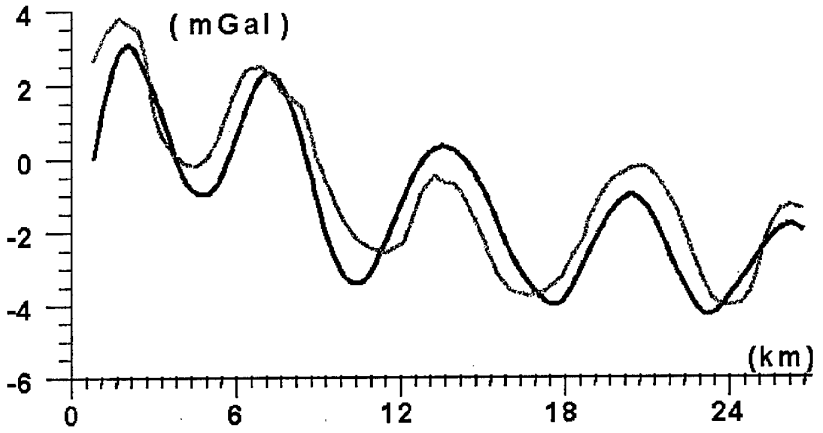


Figure 6: Repeatability flight test results

In order to realize above checking, the comparison with known ground upward continued gravity data were applied. The ground data were provided by the University of Calgary. The gravity estimation results for the repeated line are shown by plots Fig. 6

The gravimetry data processing consists in the original sophisticated multi-step procedure which includes the different kinds of filtering and smoothing. The figure 7 represents the more essentials of the data processing stages.

The next testing step was the grid flying. The assumed result of the airborne gravity survey is the free-air gravity anomalies map of the surveyed area. In the autumn of 1996 Canada celebrated the 50-th anniversary of the oil exploration. The Turner Valley is exactly the place were this happened and the fact that our first grid flight took place over this field in this time looks symbolical. The Turner Valley oil fields is an area with large amount of ground gravity points available. The map built using the upward continued free-air gravity is shown on Fig. 8 (right), while the map based on the system

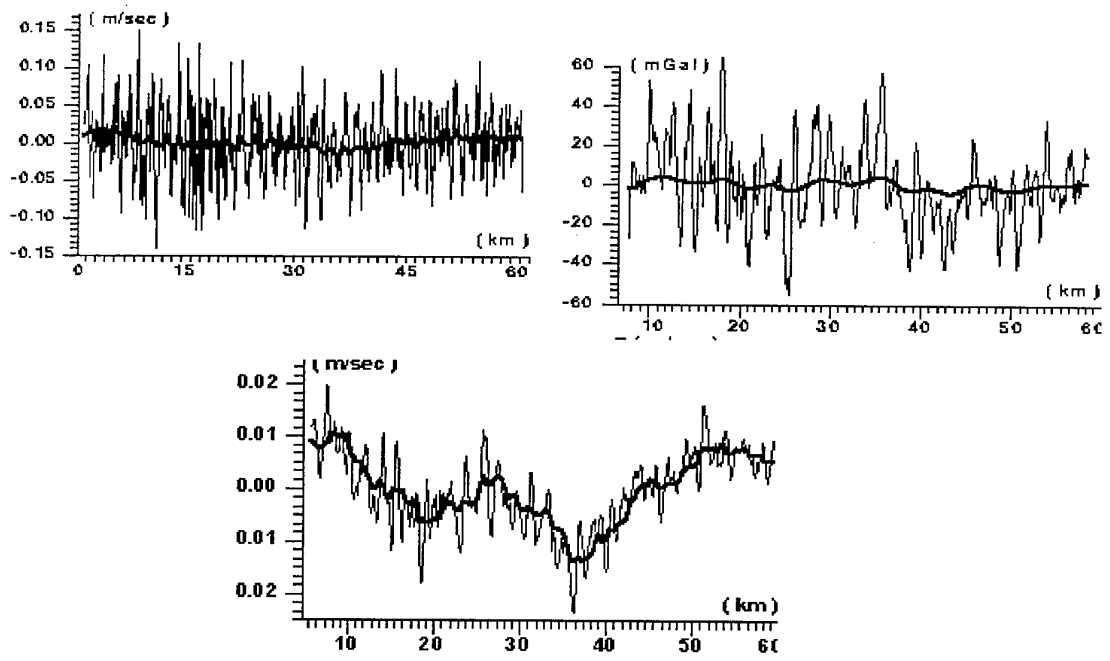


Figure 7: Airborne gravimetry data processing stages

measurements is illustrated by the picture on the left. This was our first airborne gravity survey map. The comparison between these two maps shows the correspondence within 0.6-1mGal per 2-3 km.

4 Real Time Precise Navigation

Another application is the real-time integration of strapdown inertial navigation system and GPS-GLONASS. In this case the Russian laser strapdown navigation system (SINS) I-42 was applied. The Ashtech GG-24 receiver provides the GPS-GLONASS measurements. The tests of such integrated system were carried out in the different countries and the achieved accuracy was on the level of 2..6 m depending on environmental conditions. In the forested and urban area when the GPS shadings are often the position accuracy is about 6..7 meters, in open area — 2 meters.

Another application of INS/DGPS real-time integration has been developed for Newmont Gold Company (CO, USA). The statement of the problem was to provide the navigation guidance of the light aircraft which carries magnetic, electromagnetic or gravity survey equipment. For this purpose a new type of navigation complex was developed. It incudes a low-cost motion sensor (MotionPak^{TM} developed by *Systron Donner Inertial*

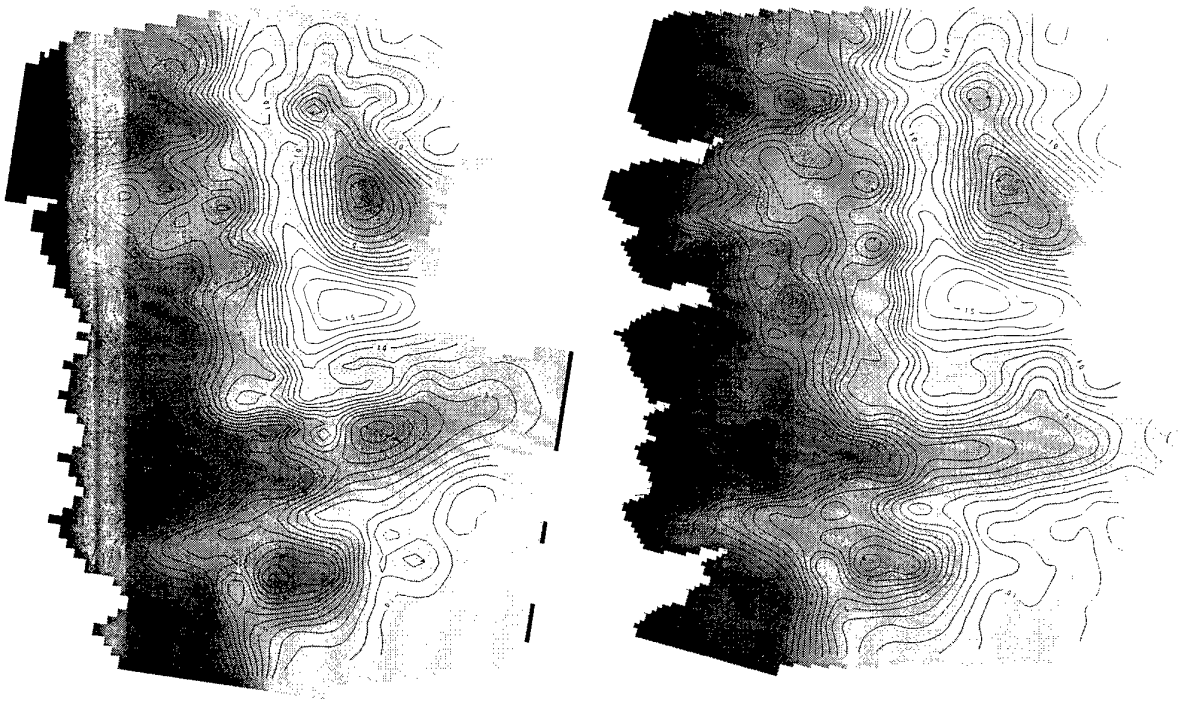


Figure 8: Gravity anomalies flight data (left) and upward continued ground data(right)

Division), real-time DGPS and magnetic compass. The MotionPak^{TM} is a "solid-state" six degree of freedom inertial sensing system used for measuring linear accelerations and angular rates in instrumentation and control applications. It is a highly reliable, compact, and fully self-contained motion measurement package. It uses three orthogonally mounted "solid-state" micromachined quartz angular rate sensors, and three high performance linear servo accelerometers mounted in a compact, rugged package, with internal power regulation and signal conditioning electronics. Following our laboratory test results the resolution of gyros is 14 deg/hrs and day to day drift bias 40 deg/hrs.

The advanced sophisticated software was developed in order to provide the high accuracy of such a complex for both real-time positioning and attitude determination.

Two series of test for this complex have been performed in USA: on the truck and in flight. For the flight test the Cessna 206 (single engine aircraft) carried motion sensor side by side with precise Russian platform INS. The INS attitude indications were used as a reference data for postmission accuracy determination. The following real time accuracy

was achieved during flight test:

velocity determination	< 1 m/s
position determination	< 1m
attitude	0.1 deg
position when GPS data lost for 60 sec	< 15 m

References

[1] Salychev O.S.:— Inertial Surveying: ITC Ltd. Experience, Bauman MSTU Press, 1995

[2] Salychev O.S.:— Inertial Systems in Navigation and Geophysics, Bauman MSTU Press, 1998

Low Cost Strapdown Inertial / GPS Integrated Navigation for Flight Test Requirements

Dr. Barrie W. Leach
Flight Research Laboratory
Institute for Aerospace Research
National Research Council of Canada
Ottawa, Ontario, Canada K1A 0R6
Tel: (613) 998-3149 Fax: (613) 952-1704 E-mail: barrie.leach@nrc.ca

1.0 SUMMARY

This paper describes the development of a capability to integrate low cost strapdown IMU data with differential GPS (DGPS) data, in an optimal fashion, using the principles of Kalman filtering and smoothing. The goal is to create a complete strapdown navigator, based on the low cost IMU, by employing an INS/DGPS Kalman filter in an error state feedback configuration. In this manner, the strapdown IMU's inherently large errors can be corrected in real time to provide a strapdown navigator of sufficient accuracy for all inertial sensing requirements. The raw IMU/DGPS data can also be optimally integrated postflight, using a Kalman filter-smoother, to establish an even more accurate aircraft inertial state time history 'after the fact'.

2.0 INTRODUCTION

The Flight Research Laboratory (FRL) of Canada's National Research Council (NRC) is a national centre for flight test activities of both a civilian and military nature. For on-board navigation and inertial sensing, the FRL utilizes a variety of GPS and differential GPS (DGPS) receivers, medium accuracy Inertial Navigation Systems (INSs) and low cost strapdown Inertial Measurement Units (IMUs). In typical flight research projects (**Figure 1**), the complete 'inertial state' of the aircraft must be measured – i.e., aircraft attitude/heading, angular rates and linear accelerations as well as the usual aircraft position and velocity. For some time now, the FRL has been employing a portable instrumentation / data acquisition system developed at the Laboratory, called Micropak, which has proven to be extremely useful for flight test programs involving aircraft from outside agencies that can only be instrumented on a temporary basis. The heart of Micropak is a strapdown IMU consisting of a triad of Sundstrand Q-FLEX accelerometers together with a pair of Litton (Canada) CSG-2 two-degree-of-freedom tuned rotor gyroscopes. This system, together with an on-board GPS/DGPS receiver, has been used successfully on several collaborative projects involving specialized flight tests. The FRL is currently assessing the capabilities of a tactical grade IMU, Litton's LN-200 (based on fibre optic gyro technology), as a replacement for the Micropak IMU, and one of these LN-200 IMUs is currently being flight tested and evaluated for this role.

This paper describes the development of a capability, as part of a modern, portable flight test instrumentation suite, to optimally integrate strapdown IMU data, from either the Micropak or LN-200 IMU, along with DGPS data, using the principles of Kalman filtering and smoothing. The goal is to develop a complete strapdown inertial navigation system for flight test purposes, but one of relatively low cost and size, based on using either of these IMUs and employing an INS/GPS Kalman filter in an error state feedback configuration. In this manner, the inherently large errors of the navigation system that is based on the low cost strapdown IMU can be corrected in real time, resulting in an inertial navigation system of sufficient accuracy for all aircraft inertial sensing requirements. In many cases, having an optimal *real-time* strapdown inertial navigation capability is not essential; and the raw IMU/DGPS data can be optimally integrated *postflight*, using a Kalman filter-smoother, in order to reconstruct an accurate aircraft inertial state time history 'after the fact'.

Section 3 of the paper gives some of the on-board instrumentation details, including the specifications of the inertial sensors in the Micropak and LN-200 IMUs (i.e., the gyros and accelerometers), the GPS / DGPS equipment being used, as well as the medium accuracy LTN-90-100 IRS used as a reference for evaluating integrated navigation system performance.

Section 4 then describes the error state feedback Kalman filter-smoother algorithm that has been designed to process the IMU and GPS data. By immediately feeding back the error state estimates to correct the various INS system and sensor errors, one can ensure that the errors of the low cost INS remain small and within the linear range required by the Kalman filter. Details of an expanded error state model that accounts for all significant IMU sensor errors are also given.

Section 5 shows some simulation results that predict the performance to be expected when integrating GPS/DGPS

Paper presented at the RTO SCI International Conference on "Integrated Navigation Systems", held at "Elektropribor", St. Petersburg, Russia, 24-26 May 1999, and published in RTO MP-43.

data with inertial data from either the Micropak or LN-200 IMUs. Various simulation runs indicate that inertial velocity errors can easily be maintained within 0.1 m/s RMS, pitch/roll errors within 0.15 degree RMS and heading angle errors within 0.5 degree RMS when using the real-time Kalman filter. The off-line Kalman filter-smoother can reduce these inertial system errors even further in size, by about another factor of two.

Section 6 gives some results based on using this optimal integration design with real flight data from the Micropak IMU and a NovAtel GPS/DGPS receiver. Analyzing actual flight test data confirms that the Kalman filter-smoother performance with real data is generally consistent with what has been predicted using simulated data. Some minor differences do occur and these, and other issues related to using flight test data, are addressed in this section.

In Section 7, the major conclusions are presented and several follow-on tasks are identified to further improve the quality of the integrated data and to develop other application areas at the FRL for this low cost/high accuracy IMU/DGPS technology.

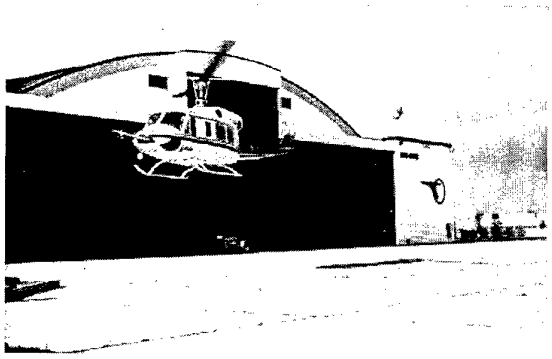


Figure 1. Bell 212 helicopter at FRL for flight test

3.0 ON-BOARD INSTRUMENTATION DETAILS

The Kalman filter error state feedback IMU/DGPS integrated navigator described in this paper has been specifically designed for use with flight test data acquired using FRL's PACNet / Micropak portable instrumentation system. Some of the details of this system, particularly as related to the acquisition of the inertial navigation parameters of interest, are given below.

Data Acquisition

PACNet, the portable instrumentation and data acquisition system used in many of the FRL's flight test programs, is a distributed architecture, digital data acquisition, filtering and recording system. Analog signal sources are sampled at a rate of 1024 Hz, digitally filtered to a bandwidth of 10 Hz, then sparse sampled and recorded at 64 Hz.

Micropak IMU

The Micropak IMU (**Figure 2**) is a strapdown inertial measurement unit consisting of a triad of Sundstrand QA-2000-030 accelerometers mounted on one block, with a pair of Litton (Canada) CSG-2 two-degree-of-freedom tuned rotor gyros (on long-term loan from DREO/DND) mounted on a separate block. According to DREO tests of the CSG-2 gyro[1,2,3], this prototype gyro is capable of excellent inertial grade performance (e.g., bias repeatability of 0.007 deg/hr, scale factor stability of 30 ppm, random drift of 0.003 deg/hr, g-sensitive repeatability of 0.008 deg/hr/g). The manufacturer's specifications for the QA-2000 accelerometer place it in the inertial grade category as well (e.g., residual bias of 60 μg, scale factor stability of 100 ppm, residual misalignment of 40 μrad). The linear accelerations and angular rates (Ax, Ay, Az, P, Q, R) from Micropak are digitized with 16-bit resolution and recorded by PACNet at a 64 Hz rate. Unfortunately, the 16-bit A/D resolution is not nearly accurate enough to match the performance capabilities of the gyros, in particular. The least significant bit value for each of the gyro channels is only 9.0 deg/hr - barely enough to 'see' earth rate - and the least significant bit value for the horizontal accelerometer channels is 15.26 μg (and 61.0 μg for the vertical accelerometer). Because of the poor resolution, and the use of A/D converters that are not up to inertial grade standards, the quality of the recorded IMU data is actually more comparable to that from low cost sensors.

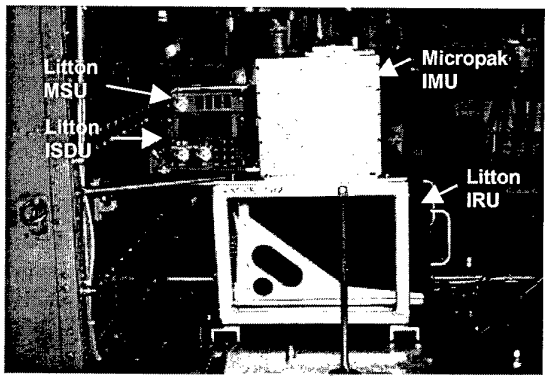


Figure 2. Inertial sensors on-board Bell 212

LN-200 IMU

The Litton LN-200 (**Figure 3**) is a low cost, small, lightweight tactical grade IMU that utilizes fibre optic gyros (FOGs) along with micromachined silicon accelerometers (SiAc's^{TM}) for measurement of vehicle angular rate and linear acceleration. The unit purchased by the FRL for evaluation is specified to have a gyro bias repeatability of better than 3 deg/hr, gyro scale factor

accuracy of better than 300 ppm, gyro random walk of less than 0.15 deg/hr$^{1/2}$ and g-sensitive repeatability of 0.01 deg/hr/g. The accelerometer residual biases are less than 1,500 μg, scale factor stability is 1000 ppm and residual axis misalignment is 300 μrad. Digital input/output is via an RS-485 serial data bus employing the IBM synchronous data link control (SDLC) protocol. Raw digitized delta velocity and delta angle IMU data are available from the bus at a 400 Hz rate. However, the FRL DSP-based interface will normally be configured to average several samples and produce an output data stream at a rate of 80 Hz.

Figure 3. LN-200 IMU on the bench

Litton LTN-90-100 IRS

For a joint CAE / FRL flight test program conducted during the summer of 1997 on a Bell 212 helicopter, a Litton LTN-90-100 RLG IRS (**Figure 2**) was used to obtain supplemental inertial data (i.e., aircraft pitch, roll, heading and groundspeed), recorded by PACNet, in the ARINC 429 digital data format, at 64 Hz. This Litton IRS data also provided redundant inertial reference information for the evaluation of the Kalman filter-based error state feedback strapdown navigator design.

Differential GPS (DGPS)

A NovAtel Model 3151M OEM GPSCard receiver, connected to a laptop computer, served as the airborne GPS station during the Bell 212 flight test program. This receiver had special firmware installed, customized by NovAtel, and optimized and tested by the FRL for use in helicopter flight tests. A NovAtel Model 511 GPS antenna was installed on the Bell 212's centreline, at the junction of the windshield and the cockpit roof (**Figure 4**). A second NovAtel GPS receiver (Model 3151R OEM GPSCard), together with a Model 501 antenna with choke-ring, was connected to another laptop computer and used as a mobile reference station, on FRL's apron, to provide differential GPS corrections. A third NovAtel GPS receiver card, installed in a personal computer and connected to a stationary roof-top antenna, was occasionally used as a second reference station. DGPS data were collected from both the airborne receiver and ground-based reference receiver and post-processed, using the 'P-RTK' software from GeoNav Systems, to obtain earth-referenced aircraft trajectory data (3-D position and velocity), along with GPS time, at a rate of 2 Hz.

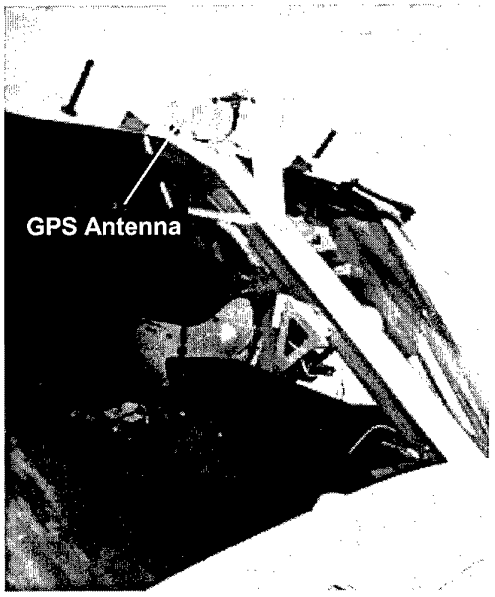


Figure 4. GPS antenna installation on Bell 212 helicopter

4.0 ERROR STATE FEEDBACK KALMAN FILTER-SMOOTHER

Need for Error State Feedback

For gyros having the performance quality of the CSG-2, as configured in the Micropak IMU, simulation studies have determined that the usual error state feedforward Kalman filter-smoother integrated navigation concept is not adequate, due to the very large nonlinear error growth occuring in a stand-alone strapdown INS based on an IMU of that quality. The effective gyro random error (the parameter most affecting overall navigation performance) of the CSG-2 gyro used in the Micropak configuration is at least 9 deg/hr (due to the limited resolution), as compared to 0.01 deg/hr, for example, for the ring laser gyros used in the medium-accuracy LTN-90-100 IRS. The performance accuracy of a stand-alone inertial navigation system that used gyros with the above error characteristics would result in position error buildup in excess of 100 nm/hr and peak velocity errors in excess of 100 m/s. As a consequence, the linearized inertial error state propagation equations used in the Kalman filter would no longer be valid for the size of the errors being encountered.

In such a situation, rather than an error state feedforward Kalman filter design, in which the INS is allowed to function in a 'stand-alone' fashion, with the Kalman filter being used to estimate and track the various inertial error states as they increase over time, integration with DGPS then requires an error state feedback configuration in which all of the error state estimates are fed back to correct the strapdown INS navigation calculations 'on the fly', after each Kalman filter measurement update. By immediately feeding back the error state estimates, to correct gyro and accelerometer errors, platform misalignments, inertial velocities and positions, the strapdown INS errors remain small and within the linear range required by the Kalman filter.

Kalman Filter Equations

The standard Kalman filter recursive update equations are certainly well known[4,5]. Nevertheless, it is worth repeating these equations, to point out some simplifications that arise due to the error state feedback design. As well, the equations being used for a companion R-T-S smoother[6] are specified.

Let $\mathbf{x}(k)$ be the n^{th}-order system error state vector evaluated at discrete time t_k; $\Phi(k,k+1)$ be the n x n state transition matrix over the time interval $t_k \rightarrow t_{k+1}$; $\mathbf{G}(k)$ be the n x r plant noise gain matrix at t_k; and $\mathbf{u}(k)$ be the r^{th}-order vector of zero-mean, white, Gaussian (ZMWG) discrete plant noise processes having covariance matrix $\mathbf{Q}(k)$ at t_k. Let $\mathbf{z}(k+1)$ be the m^{th}-order measurement vector at time t_{k+1}, $\mathbf{H}(k+1)$ be the m x n observation matrix and $\mathbf{v}(k+1)$ be the m^{th}-order measurement noise vector having covariance matrix $\mathbf{R}(k+1)$ at t_{k+1}. Assume that noise vectors $\mathbf{u}$ and $\mathbf{v}$ are statistically independent (i.e., the components of $\mathbf{u}$ are uncorrelated with the components of $\mathbf{v}$); and assume also that $\mathbf{x}(0)$ is independent of both $\mathbf{u}$ and $\mathbf{v}$. Let $\mathbf{x}(k+1^+)$ and $\mathbf{P}(k+1^+)$ be the optimal estimate of the error state vector and its associated error covariance at t_{k+1}, just after the latest measurement update, while $\mathbf{P}(k+1^-)$ denotes the error covariance just before the measurement update at t_{k+1}. Under the foregoing definitions and assumptions, the discrete-time error state feedback Kalman filter equations become:

error dynamics model

$$\mathbf{x}(k+1) = \Phi(k,k+1)\,\mathbf{x}(k) + \mathbf{G}(k)\,\mathbf{u}(k)$$

measurement process

$$\mathbf{z}(k+1) = \mathbf{H}(k+1)\,\mathbf{x}(k+1) + \mathbf{v}(k+1)$$

covariance time update

$$\mathbf{P}(k+1^-) = \Phi(k,k+1)\,\mathbf{P}(k^+)\,\Phi^T(k,k+1) + \mathbf{G}(k)\mathbf{Q}(k)\,\mathbf{G}^T(k)$$

Kalman gain

$$\mathbf{K}(k+1) = \mathbf{P}(k+1^-)\,\mathbf{H}(k+1)^T\,[\mathbf{H}(k+1)\,\mathbf{P}(k+1^-)\,\mathbf{H}^T(k+1) + \mathbf{R}(k+1)]^{-1}$$

state measurement update

$$\mathbf{x}(k+1^+) = \mathbf{K}(k+1)\,\mathbf{z}(k+1)$$

covariance measurement update

$$\mathbf{P}(k+1^+) = \mathbf{P}(k+1^-) - \mathbf{K}(k+1)\,\mathbf{H}(k+1)\,\mathbf{P}(k+1^-)$$

In comparison to the standard Kalman filter update equations, the above equations for full error state feedback have no state time update, and the state measurement update is simplified.

Kalman Smoother Equations

The Kalman filter provides the best estimate of $\mathbf{x}(k)$ based on all measurements available up to discrete time t_k. However, the additional information that is contained in the measurements taken *after* time t_k can be used to improve the error state estimates even further. For optimal off-line error state estimation, this becomes a so-called fixed-interval smoothing problem[6], equivalent to optimally combining (in a minimum variance sense) the estimates of two Kalman filters, one working forward over the data interval and the other one running backward over the same data. A well-known smoother algorithm to perform this efficiently is the so-called Rauch-Tung-Streibel (R-T-S) smoother[7]. The R-T-S smoothed estimate $\mathbf{x}^s(k^+)$ is a linear combination of the associated forward-time Kalman-filtered estimate $\mathbf{x}(k^+)$ and the smoothed estimate at the previous point (in a reverse-time sense) $\mathbf{x}^s(k+1^+)$. Based on the forward-time Kalman filter equations, the complete R-T-S smoother equations, implemented in reverse time, are as follows:

R-T-S smoother gain

$$\mathbf{A}(k) = \mathbf{P}^T(k^+)\,\Phi^T(k+1,k)\,\mathbf{P}^{-1}(k+1^-)$$

smoothed state vector

$$\mathbf{x}^s(k^+) = \mathbf{x}(k^+) + \mathbf{A}(k)\,\mathbf{x}^s(k+1^+)$$

smoothed covariance matrix

$$\mathbf{P}^s(k^+) = \mathbf{P}(k^+) + \mathbf{A}(k)\,[\,\mathbf{P}^s(k+1^+) - \mathbf{P}(k+1^-)\,]\,\mathbf{A}^T(k)$$

with initial conditions for the recursive reverse-time smoothing process given by the final values of the forward-time Kalman filter, namely $\mathbf{x}^s(N) = \mathbf{x}(N^+)$; $\mathbf{P}^s(N) = \mathbf{P}(N^+)$.

The error state feedback version of the R-T-S smoother requires that, after each Kalman filter measurement update, all of the non-zero elements of $\mathbf{x}(k^+)$, $\mathbf{P}(k^-)$, $\mathbf{P}(k^+)$ and $\Phi(k,k+1)$ be stored. However, it does not require the re-processing of the measurements used in the forward-time Kalman filter (unlike some other smoother algorithms). This has practical advantages when handling typical real-world situations in which the Kalman filter design automatically detects and rejects spurious measurement data.

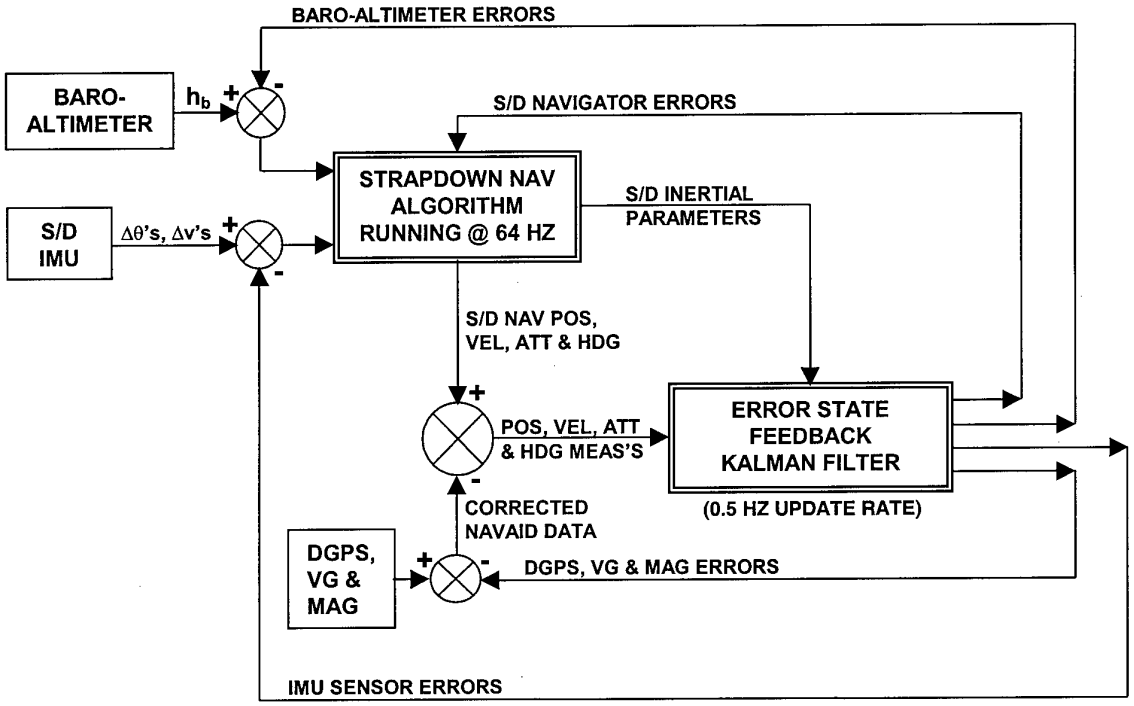


Figure 5. Error state feedback Kalman filter design for optimal integration of strapdown IMU with DGPS and attitude/heading sensors

Kalman Filter Design Details

Figure 5 shows a block diagram representation of the error state feedback Kalman filter integration design being employed to optimally integrate IMU and DGPS data in an error-corrected strapdown inertial navigator. Note that the INS/DGPS integration is of the so-called loosely-coupled form, in which GPS position and velocity components are used as primary measurements rather than the more fundamental GPS pseudoranges and pseudorange rates. With reference to **Figure 5**, raw angular rate, linear acceleration and barometric height data (all available at 64 Hz) are sent to a strapdown navigation algorithm that computes the standard inertial navigation outputs (i.e., position, velocity and Euler angle components). At the same time, an error state feedback Kalman filter processes the differences between strapdown INS and DGPS position and velocity components as the primary Kalman filter measurements (available at 2 Hz). A set of secondary, optional measurements can also be made available by forming differences between strapdown INS pitch/roll attitude and pitch/roll attitude from a vertical gyro, and between strapdown INS heading and heading from a magnetic compass. The Kalman filter estimates the fundamental errors existing in the strapdown navigator, the DGPS, the baro-altimeter, the vertical gyro and the magnetic compass sensors. These estimated errors are then fed back to correct the important navigation quantities 'on the fly' - primarily those of the strapdown navigator, but also those of the auxiliary sensors. There are two types of feedback signals to be provided: i) corrections to the strapdown INS system states (i.e., 3-D position, 3-D velocity, 3 Euler angles and vertical channel damping loop acceleration correction), and ii) corrections to the strapdown IMU sensors, DGPS, baro-altimeter, vertical gyro and magnetic compass for various bias-like errors (e.g., offset errors, scale factor errors and misalignment errors). In the former case, the latest values estimated by the Kalman filter for these quantities are used as corrections to be applied to the corresponding inertial parameters being computed within the strapdown navigation software over the next Kalman filter update interval. However, in the case of the various sensor error states (including those for the DGPS, baro-altimeter, vertical gyro and magnetic compass), the latest correction values are added to the previously accumulated values, and it is the total accumulated value, in each case, that is used as the latest correction to apply to the raw data from each sensor. The accumulated values, as a function of time, then correspond to the time propagation of the bias-like errors in the IMU sensors and the DGPS, baro-altimeter, vertical gyro and magnetic compass measurements.

The companion Kalman smoother attempts to estimate any residual errors in the error state feedback strapdown navigator that the smoothing process is able to identify. These incremental errors, so identified, can be used to

afford a further, off-line, correction to the INS system parameters as well as to the bias-like sensor errors.

The strapdown navigation software uses a wander azimuth navigation frame of reference, whereas the Kalman filter's strapdown INS error state modelling is based on a geographic reference frame[8,9,10]. That being the case, the Kalman filter's estimates of errors in the horizontal inertial velocities are simply transformed into the wander azimuth frame of reference and used to correct v_x and v_y. Also, the Kalman filter-estimated tilt angle errors, in the geographic frame, are not directly useable in the strapdown navigation algorithm. This problem is resolved by using the Kalman filter's estimates of Euler angle errors (rather than tilt angle errors) to correct the latest set of Euler angles computed by the strapdown navigator. Then, a feedback-corrected body-to-wander angle direction cosine matrix (DCM) is used to start the next strapdown navigation iteration. That way, the platform misalignment is adjusted at the beginning of a strapdown navigation interval, and normal attitude updating occurs for the rest of the interval.

A fairly complete Kalman filter error state modelling is employed (especially for the various gyro and accelerometer errors[11]), consisting of a total of 42 error states, with the option to 'switch off' selected error states as desired. The 42-element error state vector consists of: i) 3 INS position errors, ii) 3 INS velocity errors, iii) 3 INS tilt angle errors, iv) baro loop acceleration correction error, v) 3 gyro offsets, vi) 3 gyro scale factor errors, vii) 6 gyro-to-accelerometer misalignment angles, viii) 3 gyro g-sensitive drifts, ix) 3 accelerometer offsets, x) 3 accelerometer scale factor errors, xi) baro-altimeter bias and scale factor, xii) 3 DGPS position offsets, xiii) 3 DGPS velocity offsets and xiv) 3 attitude/heading sensor offsets. All of the error states, other than the so-called strapdown INS system error states (i.e., the first ten), are modelled as first-order Gauss-Markov processes[4].

The Kalman filter measurements consist of up to 9 scalar components, namely: i) 3 strapdown INS - DGPS position differences, ii) 3 strapdown INS - DGPS velocity differences and iii) 3 strapdown INS – attitude / heading sensor differences. The most important of these measurements are the INS - DGPS position differences, since all of the error states can be estimated reasonably accurately using only these 3 measurement components. The Kalman filter is implemented using Bierman's UDU^T mechanization algorithm[12] for efficiency and robustness, and the software has been designed so that the user can choose from Kalman filter-smoother update intervals of 2, 5 or 10 seconds.

5.0 SIMULATION RESULTS

Simulated strapdown IMU and DGPS data were used to test the integration software under various scenarios and contingencies such as IMU/DGPS data time skews, DGPS outages, various complexities of Kalman filter error state modelling and different Kalman filter-smoother update rates. A realistic flight trajectory, with several 180° turns in it, was used to try to depict the typical maneuver-rich environment of a flight test exercise. The IMU sensor errors being simulated were modelled using a fairly complete description of first-order errors[11], taking into account, as well, the limited resolution inherent in the Micropak data acquisition process.

Table 1 shows the average error levels in the various strapdown navigator parameters, predicted from simulation runs when simulating the typical errors to be encountered with Micropak-quality raw inertial data (i.e., 3-D acceleration and angular rate components), both for the real-time Kalman filter and post-flight Kalman filter-smoother updating every 2 seconds (the fastest rate used). With a well-tuned Kalman filter-smoother, the actual error time histories remain consistently within the +/- 1 σ uncertainty bounds shown in **Table 1**.

Table 1.
Error levels predicted from simulation studies

Error State Parameter	1 σ Filter Uncertainty	1 σ Smoother Uncertainty
3-D Position Components	1.5 m	1.5 m
3-D Velocity Components	0.10 m/s horiz, 0.20 m/s vert	0.025 m/s horiz, 0.15 m/s vert
Euler Angles (Pitch, Roll, Yaw)	0.15 deg, 0.75 deg for yaw	0.075 deg, 0.20 deg for yaw
Gyro Offsets	3.0 deg/hr, 7.5 deg/hr for z	2.0 deg/hr, 5.0 deg/hr for z
Gyro Scale Factors	0.005	0.004
Gyro Misalignments	0.09 deg	0.05 deg
Accelerometer Offsets	2000 μg, 4000 μg for z	1500 μg, 3500 μg for z

It can be seen, from **Table 1**, that the errors in the inertial velocity components and Euler angles benefit the most from the off-line Kalman smoothing process. To illustrate this, **Figure 6** and **Figure 7** show typical time histories of the residual errors in the strapdown navigator's calculation of the Euler angles – for Kalman filtering and Kalman smoothing respectively – together with the Kalman filter-smoother's calculation of +/- 1 σ uncertainty bounds (the dashed lines). The improvement in accuracy in going from real-time Kalman filtering to off-line Kalman smoothing is quite evident.

6.0 RESULTS USING REAL FLIGHT DATA

IMU Error Model Investigation

Initial attempts at applying the error state feedback Kalman filter-smoother software to Micropak IMU and NovAtel DGPS flight data led to rather disappointing and inconsistent results. Several performance indicators were available to assess the accuracy of the error-corrected strapdown navigator, namely: i) strapdown navigator position and velocity outputs compared to raw GPS position and velocity parameters, ii) strapdown navigator estimates of groundspeed and Euler angles (aircraft attitude and heading) compared to the equivalent LTN-90-100 IRS parameters (which were being recorded and used in the CAE flight test program), and iii) engineering judgement regarding the various sets of gyro and accelerometer sensor error state estimates. Initially, a fairly basic modelling of the gyro and accelerometer errors was used, namely, first-order Markov error states to represent slowly time-varying biases. This form of sensor error state modelling has proven to be more than adequate when applied to medium-accuracy INSs such as the LTN-90-100 IRS. However, initial comparisons of the Euler angles computed by the strapdown navigator with the LTN-90-100 IRS Euler angles indicated a great deal of variability in the feedback-corrected strapdown navigator residual errors as a function of the turning maneuvers, and a significant offset error in the heading angle, in particular. As well, the estimated accelerometer offset errors were unusually large and very sensitive to the various turning maneuvers – in other words, obviously not physically realizable.

Based on the foregoing observations, and with the knowledge that the Micropak IMU was not nearly as accurate as a ring laser gyro-based IMU (such as that of an LTN-90-100 IRS), the suspicion was that a more complex modelling of gyro and accelerometer errors would be required in order to capture the dominant first-order error effects. To this end, the sensor error modelling was gradually expanded until a fairly complete first-order error model was implemented - for both the gyro and accelerometer error sources. This error model consisted of 15 parameters to model the various gyro errors (i.e., bias errors, scale factor errors, misalignment errors and g-sensitive errors) and 6 parameters to model accelerometer errors (i.e., bias errors and scale factor errors).

This rather complete error model was found to substantially improve performance, and led to consistently good results across several sets of flight data from the Bell 212 flight test program. As it turned out, the dominant source of error came from a rather significant, and completely unexpected, source - a gyro misalignment error of approximately 0.5 deg, in the azimuth direction, relative to the accelerometer axes. Once this misalignment error was modelled properly, it was found that the overall error model could be reduced in size - with accelerometer scale factor errors and gyro g-sensitive errors eliminated - without affecting performance greatly.

General Performance Results

Error state estimation accuracy, based on the Kalman filter's +/- 1 σ uncertainty bounds, depends upon several factors - the amount of aircraft maneuvering, the assumed statistical noise parameters for DGPS, gyro and accelerometer data, and the update rate of the Kalman filter. For the Kalman filter, running on real data and updating its estimates every two seconds (the fastest, and most accurate, rate used), **Table 2** shows the uncertainty bounds predicted, on average, for the errors in various strapdown navigator error state parameters, both for the real-time Kalman filter and post-flight Kalman filter-smoother. With a well-tuned Kalman filter-smoother, the actual error time histories should remain consistently within the +/- 1 σ uncertainty bounds shown in **Table 2**. For the reasons given above, the results in **Table 2** are slightly different from the simulation results previously shown in **Table 1**. As before, the errors in the inertial velocity components and Euler angles benefit the most from the off-line Kalman smoothing process.

The final version of the Kalman filter-smoother software (with the most complete sensor error model), applied to several different sets of Bell 212 flight data, revealed that the error state feedback strapdown navigator performed in a fairly consistent manner, with actual errors that closely approximated the uncertainty bounds being predicted by the Kalman filter-smoother, and consistently small (hence, realistic) estimates of several of the gyro and accelerometer errors. The main method for determining performance accuracy was to compare groundspeed, pitch, roll and heading, as computed by the strapdown navigator, to those same parameters recorded from the on-board LTN-90-100 IRS. It should be noted that the LTN-90-100 IRS parameters do contain errors as well, but these comparisons certainly serve as a good starting point for confirming proper operation of the Kalman filter-based navigator.

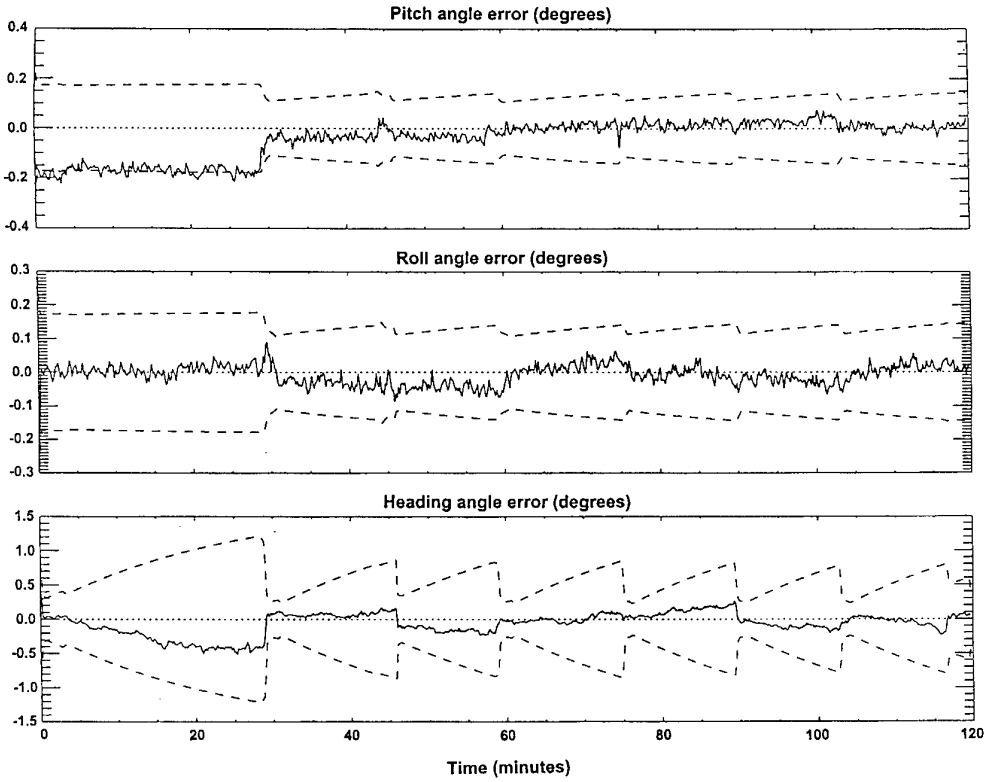


Figure 6. Euler angle error time histories for Kalman filter-corrected strapdown navigator

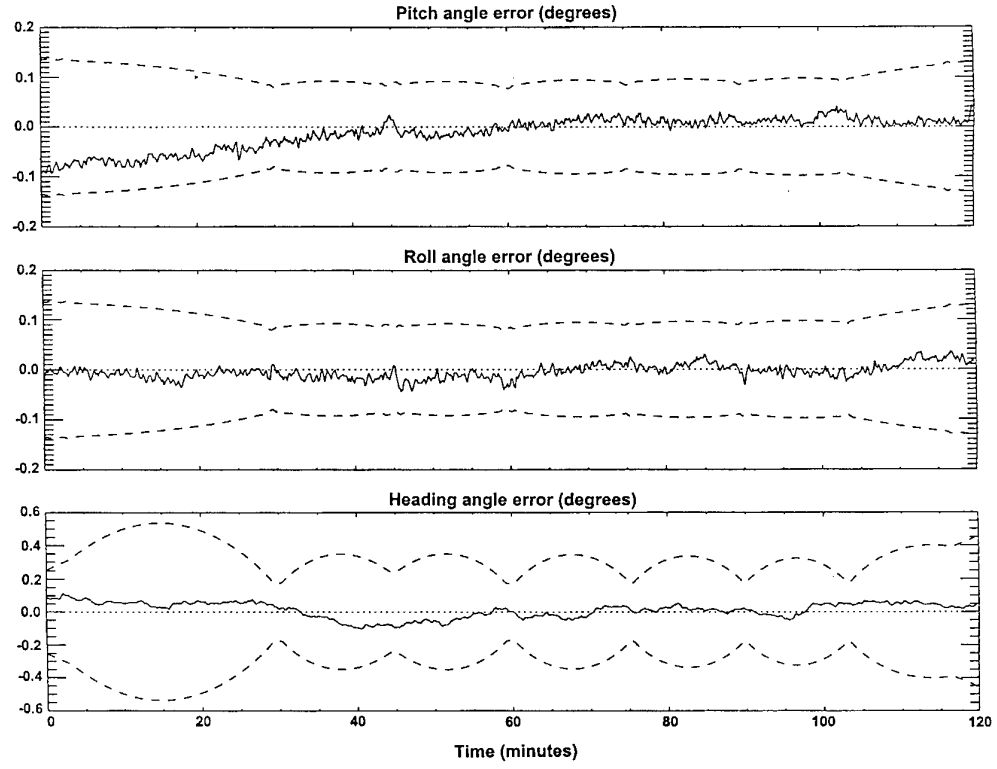


Figure 7. Euler angle error time histories for Kalman smoother-corrected strapdown navigator

Table 2.
Uncertainty bounds predicted from real data runs

Error State Parameter	1 σ Filter Uncertainty	1 σ Smoother Uncertainty
3-D Position Components	4.0 m	2.0 m, 1.0 m for vert
3-D Velocity Components	0.20 m/s	0.075 m/s, 0.20 m/s for vert
Euler Angles (Pitch, Roll, Yaw)	0.10 deg, 0.70 deg for yaw	0.05 deg, 0.30 deg for yaw
Gyro Offsets	8.0 deg/hr, 18.0 deg/hr for z	4.0 deg/hr, 15.0 deg/hr for z
Gyro Scale Factors	0.005	0.004
Gyro Misalignments	0.08 deg	0.05 deg
Accelerometer Offsets	900 μg	900 μg

Probably the best comparisons are those with the three Euler angles (pitch, roll and heading). Normally, with a proper alignment of the LTN-90-100 IRS, one can expect its pitch and roll angles to be accurate to within 0.05 deg, and its heading to be accurate to within 0.1 -> 0.2 deg. Strapdown navigator Euler angle accuracy is not usually that good, so LTN-90-100 IRS Euler angles can serve as a legitimate set of reference values. However, LTN-90-100 IRS groundspeed is not nearly so accurate as a reference because it can contain Schuler-based errors that can peak at 1.0 -> 2.0 m/s.

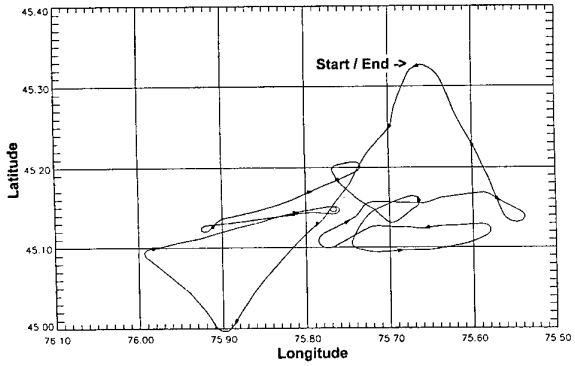


Figure 8. Typical horizontal flight trajectory for Bell 212 flight test program

Specific Performance Results

As a typical example of the performance to be expected from the error state feedback strapdown navigator, **Figure 8** shows the horizontal trajectory of the Bell 212 during Flight # B09 - indicating the rather intense level of maneuvering typical of this flight test program. **Figure 9** then shows the differences between strapdown navigator outputs and LTN-90-100 IRS outputs, together with predicted Kalman filter uncertainty bounds (dashed lines), for groundspeed, pitch, roll and heading angle. For the most part, the differences stay within the predicted error bounds, indicating a consistency between Kalman filter uncertainty estimates and actual error levels. **Figure 10** shows the resulting differences between the Kalman smoother-corrected strapdown navigator outputs and the corresponding LTN-90-100 IRS outputs, again for groundspeed, pitch, roll and heading angle. The Euler angle differences are seen to be significantly smaller than for the Kalman filter results shown previously in **Figure 9**, as would be expected. The improvement in groundspeed error is not so obvious because one now sees the error in LTN-90-100 IRS groundspeed dominating the difference. For many of the Bell 212 flights, the amount of time on the ground prior to take-off was kept to a minimum. In some of these cases, it would appear that there was a thermal transient associated with the gyro outputs that did not damp out before take-off. For those particular flights, **Figure 11** shows a typical set of time histories of the Kalman filter-smoother's estimates of gyro offsets, where the x and y gyro outputs have time varying offsets that start off at large, negative values and eventually settle down to reasonable levels.

Observations Based on Real Flight Data Results

Based on experience with real flight test data from the Micropak IMU and NovAtel DGPS, the following points will also be relevant to the LN-200 IMU integration:

- Although the DGPS data were, in general, of excellent quality, there were, in fact, several instances of individual 2 Hz DGPS records missing from the post-flight solution, as well as obvious cases of incorrect DGPS velocity solutions going to zero when the helicopter was clearly in forward flight. These phenomena almost always occurred during abrupt helicopter maneuvering;
- The Kalman filter had to be tuned very carefully, in particular with respect to specifying DGPS first-order Markov error states and measurement noise values, in order to take into account small time lags in the DGPS data relative to the IMU data and, most importantly, very rapid changes in DGPS accuracy during sudden helicopter maneuvers;

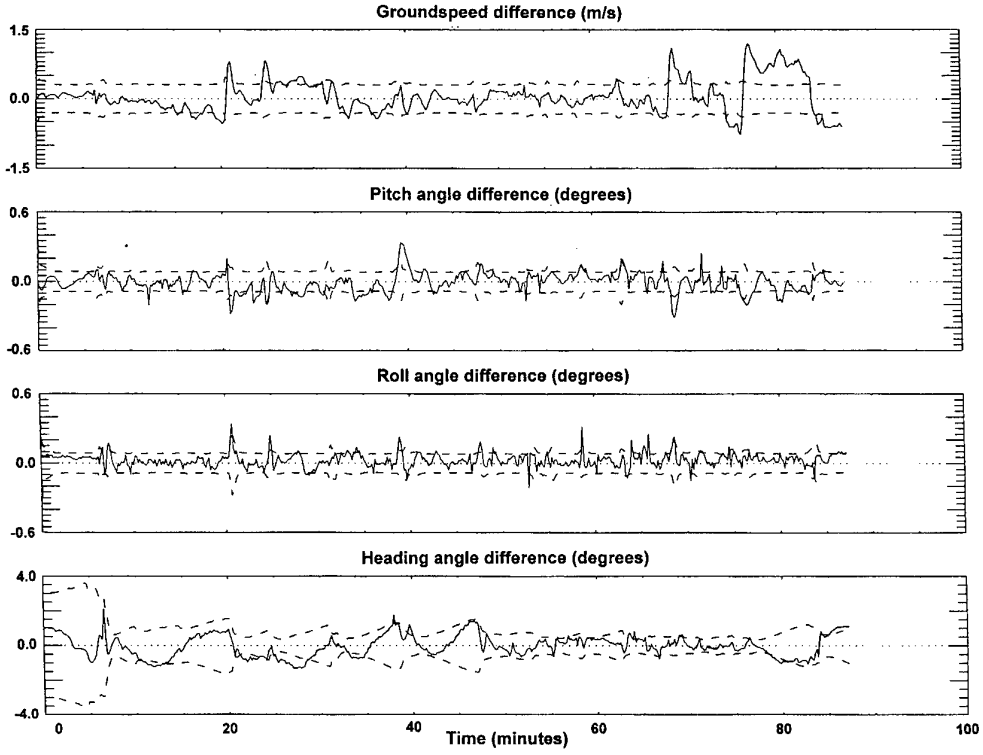


Figure 9. Differences between strapdown navigator outputs and LTN-90-100 IRS outputs for Kalman filter

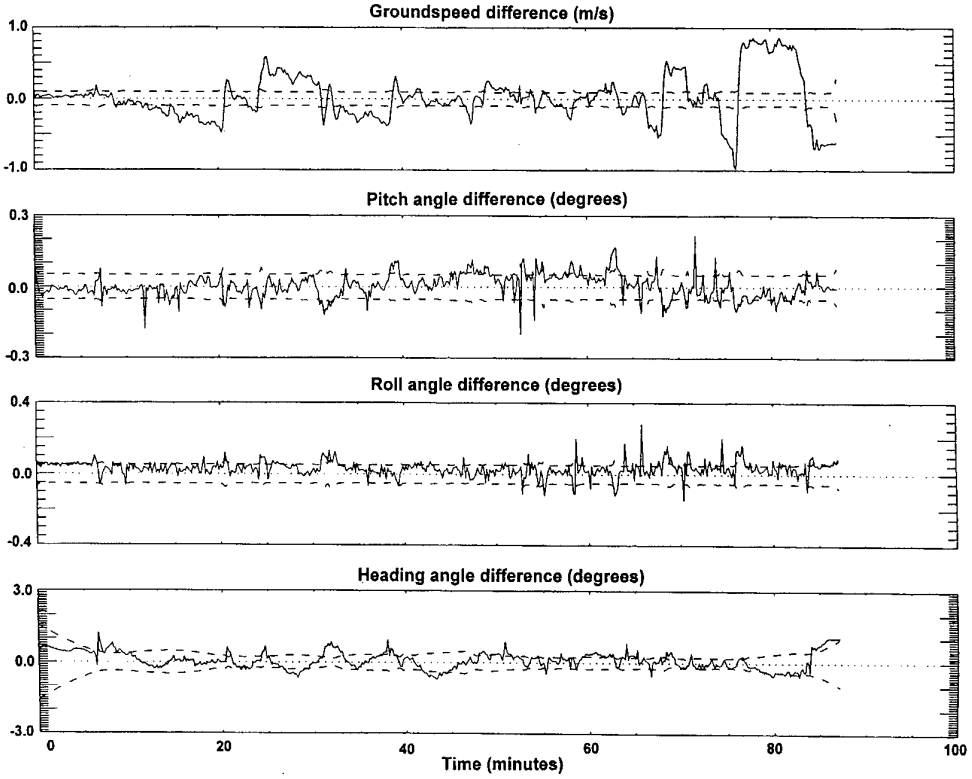


Figure 10. Differences between strapdown navigator outputs and LTN-90-100 IRS outputs for Kalman smoother

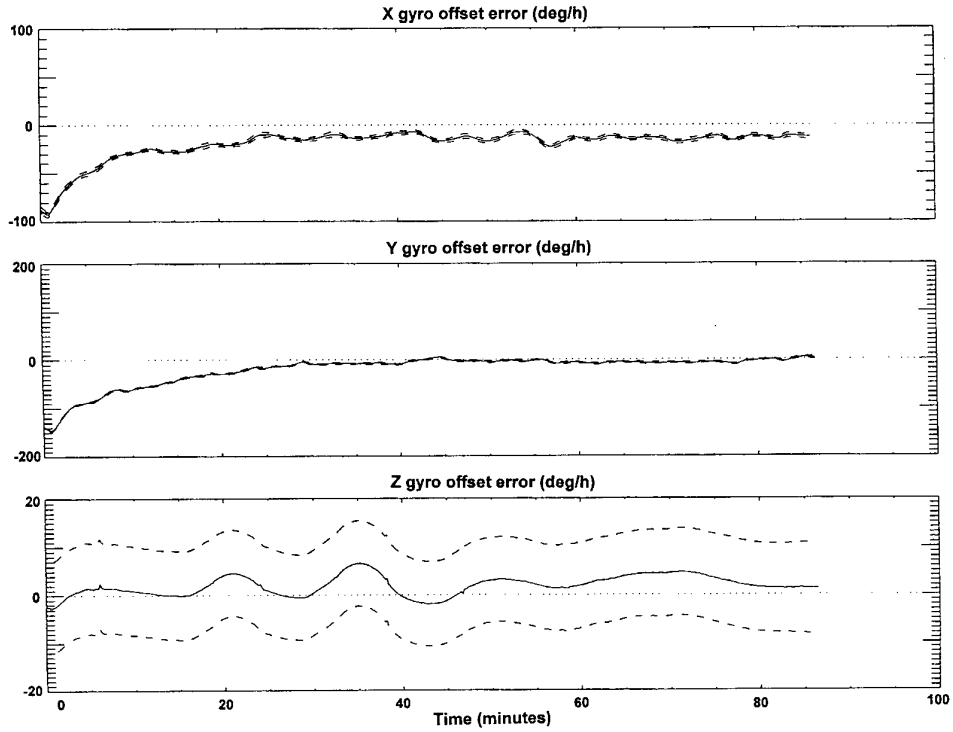


Figure 11. Kalman filter-smoother estimates of time varying gyro offsets

- Degraded DGPS performance during abrupt helicopter maneuvering resulted in sudden, large DGPS velocity Markov error states being estimated;
- The Kalman filter user option of a two second update interval (the shortest one) produced the 'tightest' feedback loop for inertial error corrections and resulted in the fewest problems with Kalman filter divergency caused by DGPS problems;
- As predicted by the various simulation results, the initial Euler angles specified for the strapdown navigator do not have to be very accurate, and starting values from vertical gyro and magnetic heading are certainly adequate;
- The off-line Kalman filter-smoother is useful for fine tuning the IMU sensor error model, including detecting unexpected gyro axis misalignments.

7.0 CONCLUSIONS

The error state feedback IMU/DGPS integration design, when applied to flight data acquired by the PACNet portable instrumentation system, does perform as predicted in simulation studies - in spite of several IMU/DGPS data deficiencies that have been identified. Using only the six Micropak IMU parameters (plus barometric height) together with DGPS, high quality 64 Hz inertial state parameters (i.e., position, velocity, Euler angle, angular rate and linear acceleration parameters) can be reconstructed for the entire flight. A comprehensive inertial sensor error model, coupled with the Kalman filter-smoother, is shown to be very useful for validating gyro/accelerometer calibrations or identifying obvious calibration errors in the IMU.

The following are several possible tasks being considered for the future:

- Improve the accuracy of the strapdown navigation algorithm by running it at a higher rate and, perhaps, including sculling and coning error compensation;
- Flight test the LN-200 IMU and NovAtel DGPS receiver, along with the LTN-90-100 IRS as a reference, in order to completely characterize the performance of an LN-200 / DGPS Kalman filter integration scheme;
- Look at a better way to record and time tag DGPS and IMU data simultaneously so as to avoid the possibility of significant time latencies between the two types of data;
- Study techniques for adaptively optimizing the performance of the Kalman filter as a function of changing DGPS quality - also, to ensure proper 'capture' of DGPS position measurements after there has been a DGPS outage interval;
- Investigate the relative advantages of employing a tightly-coupled IMU/DGPS integration scheme that involves using raw DGPS pseudorange and

pseudorange rate data as Kalman filter measurements, rather than the current loosely-coupled scheme that uses DGPS computed position and velocity data;
- Investigate the potential improvement from employing a measurement averaging scheme for the 2 Hz DGPS data.

There are several immediate applications for this low cost integration technology at the FRL. They include: a) providing real-time positioning and attitude/heading determination for an Enhanced and Synthetic Vision System (ESVS), b) supplying general real-time inertial state sensing for closed-loop control of a fly-by-wire research helicopter, c) providing accurate measurement of aircraft acceleration / braking performance on contaminated runway surfaces and d) reconstructing accurate inertial state time histories off-line and, eventually, on-line for aircraft mathematical model development.

8.0 ACKNOWLEDGEMENTS

The author would like to thank FRL Bell 212 project team members, Ken Hui, Ramesh Srinivasan and Dave Collins, for their assistance in interpreting the flight test data. Thanks are also due to CAE Electronics Ltd. for permission to use flight test data from the Bell 212 program in this paper.

9.0 REFERENCES

[1]Vinnins, M.F., L.D. Gallop and F. Paquet, "Test Evaluation of the Litton (Canada) Canadian Strapdown Gyroscope (CSG-2)', DREO Report No. 914, Ottawa, Canada, September, 1986.

[2]Vinnins, M.F. and L.D. Gallop, "Analysis of Rate and Dynamic Tests on the Litton (Canada) Canadian Strapdown Gyroscope (CSG-2)", DREO Technical Note 88-20, Ottawa, Canada, April, 1988.

[3]Vinnins, M.F. and L.D. Gallop, "Test and Evaluation of the Litton (Canada) Canadian Strapdown Gyroscope (CSG-2) Performance Evaluation Unit (PEU)", DREO Report No. 1010, Ottawa, Canada, September, 1988.

[4]Gelb, Arthur, ed., "Applied Optimal Estimation", The M.I.T. Press, Cambridge, Massachussetts, 1974.

[5]Maybeck, P.S., "Stochastic Models, Estimation, and Control Volume 1", Academic Press, New York, 1979.

[6]Maybeck, P.S., "Stochastic Models, Estimation, and Control Volume 2", Academic Press, New York, 1979.

[7]Rauch, H.E., F. Tung and C.T. Streibel, "Maximum Likelihood Estimates of Linear Dynamic Systems", AIAA Journal, Vol. 3, No. 8, 1965.

[8]Leach, B.W., "A Kalman Filter Integrated Navigation Design for the IAR Twin Otter Atmospheric Research Aircraft", IAR Report IAR-AN-72, National Research Council of Canada, April, 1991.

[9]Schmidt, G.T. (editor), "Strap-Down Inertial Systems", AGARD Lecture Series No. 95, AGARD-LS-95, May, 1978.

[10]Britting, K.R., "Inertial Navigation Systems Analysis", John Wiley & Sons, Toronto, 1971.

[11]Bolduc, C.R.M., M.F. Vinnins and D.A. Staley, "Design and Implementation of the DREO Heading Reference Unit Self-Alignment and Navigation Algorithms", CAS Journal, Vol. 43, No. 2, June, 1997.

[12]Bierman, G.J., "Factorization Methods for Discrete Sequential Estimation", Academic Press, New York, 1977.

HIGH PRECISION DGPS AND DGPS/INS POSITIONING FOR FLIGHT TESTING

Capt Roberto Sabatini

Italian Air Force Research and Flight Test Division (DASRS) - Oficial Test Centre (RSV)
Technical Group (GT) - Avionics and Armament Evaluation Service (SSAA)
Pratica di Mare Airport
Pomezia 00040 (Rome), ITALY

SUMMARY

Historically, test ranges have provided accurate time and space position information (TSPI) by using laser tracking systems, cinetheodolite systems, tracking radars, and ground-based radio positioning systems. These systems have a variety of limitations. In general, they provide a TSPI solution based on measurements relative to large and costly fixed ground stations. Weather has an adverse effect on many of these systems, and all of them are limited to minimum altitudes or to confined geographic regions. The number of participants each system can support is limited, and correlation with other systems is extremely difficult, if not impossible. These limitations greatly increase instrumentation costs and impose severe constraints on test scenarios. Clearly, a more cost-effective TSPI source is needed.

The Global Positioning System (GPS) provides a cost-effective capability that overcomes nearly all the limitations of existing TSPI sources. GPS is a passive system using satellites which provide a universal and accurate source of real-time position and timing data to correlate mission events. The coverage area is unbounded and the number of users is unlimited. The use of land-based differential GPS (DGPS) reference stations improves accuracy to about one metre for relatively stationary platforms, and to a few metres for high performance military aircraft. Further accuracy enhancement can be obtained by using GPS carrier phase measurements, either in post-processing or in real-time. Accuracy does not degrade at low altitudes above the earth's surface, and loss of navigation solution does not occur as long as the antenna has an open view of the sky.

However, DGPS performance in terms of data continuity and accuracy during high dynamic manoeuvres, even if sufficient for many tasks, can not cover the entire flight envelope of modern high performance fighter aircraft. Moreover, the update-rate of GPS receivers is too low for many tasks.

Currently, the integration of GPS with an inertial navigation system (INS) is considered to be the optimal solution to the above mentioned shortcomings. This combination, in fact, can provide the required update rate and have a higher data continuity and integrity. The other advantages of an INS: low short term drift and low noise, are combined with the advantages of GPS: high position accuracy and no long term drift. Moreover, the combination of an INS with GPS is a natural evolution of existing airborne navigation systems, the majority of which is currently based on an INS, updated by other positioning systems to compensate for the shortcomings of the inertial system.

INTRODUCTION

In the early 1996, the Italian Ministry of Defence (MoD) appointed a panel to consider the use of GPS as a TSPI source for flight testing activities. The panel focused on the two major MoD flight-test ranges: Pratica di Mare AFB and Perdasdefogu AFB. The parameters considered included: real-time and post mission requirements, data rate required for mission success, coverage requirements, life cycle costs.

In 1997 a study was undertaken in order to define the requirements of a DGPS-based system for flight testing. The activity involved the Italian Official Test Centre (It-OTC) and the Companies AERMACCHI (AEM) and ALENIA (ALN). The study was addressed on GPS using C/A code, with post-flight differentiation. This has been preferred to GPS with code P due to both simplicity of use and high accuracy attainable notwithstanding its lower cost.

The working team produced a document identifying all technical requirements of the DGPS system.

After contacting many potential suppliers, an initial assessment of different systems was conducted in order to select the DGPS systems best matching the technical requirements identified by the It-OTC, ALN and AEM. Only two of the systems proposed were suitable for flight test applications (ASHTECH and TRIMBLE), and an in-flight evaluation was then conducted on the MB339-C aircraft in order to compare their performance mainly in terms of data quality and continuity during high dynamic manoeuvres.

The system proposed by ASHTECH (i.e., two ASHTECH XII receivers with the RANGER post-processing software), was finally selected, and a second flight test campaign was then conducted on TORNADO in order to:

- identify the critical conditions and related causes limiting the use of DGPS;
- evaluate the DGPS data accuracy;

Paper presented at the RTO SCI International Conference on "Integrated Navigation Systems", held at "Elektropribor", St. Petersburg, Russia, 24-26 May 1999, and published in RTO MP-43.

- identify proper test procedures to both avoid as much as possible limiting conditions during flight and allow best data gathering;
- develop proper post-flight methodologies, including DGPS data merging with information from other available aircraft sensors (inertial system, radar altimeter, baro sensors, etc.), in order to obtain continuous, reliable and accurate positioning data.

The results of the assessment were satisfactory. Therefore, provided an adequate mission planning to avoid masking of the onboard GPS antenna, DGPS techniques could be used successfully by the It-OTC, ALN and AEM in many programs (i.e., TORNADO, AMX, F104 ASA-M, MB339-CD, EF2000, etc.).
However, it was understood that the DGPS system performance in terms of data continuity, update rate and accuracy during high dynamic manoeuvres, could not cover the entire flight envelope of high performance military aircraft (e.g., EF2000).
In consequence of the experience gained in the initial assessment, as well as in the flight test campaigns, a feasibility study has been undertaken in order to investigate the potential of integrated DGPS/INS systems. Together with real-time C/A code DGPS, also DGPS using carrier phase ambiguity resolution on-the-fly has been investigated, while different options for the INS mechanisation have been considered. The primary aim of the study was to define the concept and the design characteristics of a high precision PRS system taking advantage of the latest state-of-the-art hardware and software techniques.
The final results of the study for the integrated DGPS/INS system are not available jet, but an optimal architecture has been identified and further investigation will be conducted in order to verify the validity of the concept. Realistic projections allow to believe that the system will meet most of the position accuracy and data continuity requirements for flight test applications. Due to the advent of real-time ambiguity resolution and cycle slip fixing algorithms, real-time carrier phase computation is no longer a dream at the horizon and an evolution of the current concept based on post-processing techniques to a real-time solution is foreseen.

The paper begins with a review of flight test requirements and a presentation of traditional test instrumentation and methods. Then, GPS range applications are discussed and results are presented of flight test activities carried out on DGPS. Particularly, the critical manoeuvres and flight conditions are identified and optimisation criteria applicable to flight test missions with DGPS are presented.
The second part of the paper is dedicated to integrated DGPS/INS systems. Particularly, the various integration algorithms and architectures are discussed and compared, and the optimal layout of a precision PRS based on DGPS/INS integration is presented.
The paper closes with an underline of future work and some concluding remarks.

FLIGHT TEST REQUIREMENTS

Today, aircraft are equipped with a variety of navigation systems depending on the application (Table 1). For long range navigation aircraft are normally equipped with INS, and/or Omega, and/or LORAN C, where LORAN C is only available in certain areas like the continental USA. More and more aircraft already use GPS for the same purpose as well as for medium range navigation. This navigational task is traditionally performed with VOR and DME or, for military aircraft, with TACAN. Instrument (ILS) and microwave landing systems (MLS) provide the guidance signals for landing. The highest horizontal accuracy is required for ILS and MLS. These systems also yield a very accurate vertical position reference. For less demanding vertical positioning, barometric and radar altimeters can be used.

Application	*Range (km)*	*System*	*Accuracy (m)*
Long Range	10.000	INS, Omega	20.000
Medium Range	500	VOR/DME	200
Prec. Approach	50	ILS/MLS	2

Table I. Navigation Aids Today.

All of the mentioned standard navigation systems are in use for flight testing as well. However, many tasks demand a higher precision with accuracy in the meter or even sub-meter range. This is often required for modern avionics testing (e.g., determination of the performance of satellite or integrated navigation and landing systems), as well as for performance data verification (e.g., take-off and landing distance computation), determination of aerodynamic parameters for the evaluation of handling characteristics, and generation of special flight patterns for noise certification. An extreme precision is required for autoland certification and flight inspection of MLS and CAT III ILS installations.
The basic requirements of a generic Time and Space Position Information (TSPI) system for testing and combat training exercises are given in Table 2.

Optical and optronic systems are widely used today for the accurate measurement of flightpath trajectory. Typical solutions include: theodolites, infrared trackers, laser trackers, and inertial systems with camera or manual position update. Even onboard camera systems are in use for autoland certification. However, since optical and optronic systems are limited in range, specially developed radar systems are also available to cover ranges of up to about 100 km.
In some cases, also radio electric ranging systems can be used for aircraft performance and certification measurements, but the effect of great altitude errors in the case of low elevation angles is a factor to be taken into account.

TEST PARAMETER	A-A			A-S			CRUISE MISS.		EW	Air Exercise & Training
	A/C	DRONE	MISS.	A/C	DRONE	MISS.	En-route	Terminal		
Real-Time Accuracy (1σ):										
Position (x, y), (z), ft	25	25	25	12-25	12-25	5	25	10	15-25	50-200
Velocity (x, y), (z) fps	3	3	3	3	3	3	1	3	3	5-15
Timing (msec)	100	100	100	50-100	50-100	100	100	100	100	50-100
Data Rate (#/sec)	10-20	10-20	10-20	10-20	10-20	10-20	10-20	10-20	10	1-10
Post-Test Accuracy (1σ):										
Position (x, y), (z) ft	<25	3-25	3	5	5	1-5	>25	>10	15	>50-200
Velocity (x, y), (z) fps	3	3	3	3	3	3	1	0.1	3	5-15
Scoring Accuracy (ft-1σ Circ)	—	—	3	—	—	1-6	—	10	—	10
No. Test Articles	1-12	1-12	1-12	5-25	5-25	5-25	5-10	5-10	1-50	1-90
Coverage:										
Attitude - kft	0-100	0-100	0-100	1-75	1-75	1-75	0.1-100	0.1-100	0.1-100	0.1-100
Distance - nm (diameter)	60	60	60	30	30	30	VAR	VAR	30-60	30-60

Table 2. TSPI Requirements.

Mathematical methods currently in use for flightpath trajectory determination include:

- calculation of earth-centred Cartesian co-ordinates and transformation into geographical co-ordinates;
- method of least squares adjustment for redundant measurements computation;
- Kalman filtering for integration of INS with optronic or other traditional sensors.

A description of these methods, together with a discussion on there relative advantages and limitations can be found in [Hurrass, 1994].
All of the above mentioned solutions show some disadvantages. Either they are limited in precision, limited in range, weather dependant, contain a high degree of post-processing work, are fixed in location at specific airports, or sometimes require a major modification of the aircraft. All of these shortcomings can be overcome by using GPS.

GPS RANGE APPLICATIONS

With GPS the following four basic positioning techniques are possible:

- ***SPS GPS positioning***. An accuracy in the order of 100 m is guaranteed (C/A code, L1). This technique suits the requirements of the low accuracy class applications of flight testing, such as the evaluation/certification of medium-range navigation sensors and systems (VOR/DME, FMS, NDB, ADF, etc.).
- ***PPS GPS positioning***, with an accuracy of 10-20 m. This technique can meet some of the medium accuracy application requirements: air-to-air applications (including en-route cruise missile testing), most EW trials and air exercise/training.
- ***Differential code range GPS positioning***, with an accuracy of 1 to 5 m. This technique fulfils the requirements of the medium accuracy class of flight testing: air-to-surface applications (including terminal cruise missile testing), fly-over-noise measurements and remote sensing flights.
- ***Differential carrier phase GPS positioning***, with centimetre to decimetre accuracy. This technique can meet the requirements of the high accuracy flight test applications: the evaluation/certification of an automatic landing system and the determination of the take-off and landing performance of an aircraft.

TSPI data can be obtained by using onboard GPS receivers or frequency translators. A receiver processes the satellite signals and outputs either raw or corrected TSPI data which can be recorded or transmitted to the ground through a telemetry data link. A translator receives the GPS signals and retransmits them on a different frequency for detection and processing on the ground. Translators use up telemetry bandwidth rapidly because at least two megahertz is needed for each simultaneously operating translator. A receiver system uses less than 100 kHz of bandwidth, thus permitting large numbers of users to be active at the same time. Receivers are good candidates for aircraft and test articles with high recovery rates so that the GPS equipment can be reused. Being simpler, smaller, and less expensive than receivers, translators are well suited for small test articles that are expendable or likely to exhibit a high attrition rate such

as missiles and drones. In general, the type of onboard equipment chosen for a particular application will be dictated by performance requirements, form-fit factors, and cost relative to the test article itself.

Figure 2 shows an example of DGPS Reference Station suitable for both flight testing and navigation. Two receivers are shown at the Reference Station to increase the station reliability and to provide station integrity. Nominally each receiver will track all satellites in view in order to assure that differential corrections are determined for all satellites. The Reference Station should be able to broadcast data for all satellites in view.

If SPS equipment is used, the broadcast can be unencrypted. If PPS equipment is used, the transmission of SA corrected errors requires the use of an encrypted data link. Methods have also been developed to broadcast SA-uncorrected DGPS data from PPS Reference Stations, thus allowing broadcast over unencrypted data links. Additional details may be found in STANAG 4392.

If the onboard unit is a translator, the processing (determination of X, Y, Z state vector) must be carried out at the ground station (mostly by Kalman Filter). In the case of an airborne receiver the processing can be either carried out onboard the aircraft or at the ground station. The first option is preferred when redundant positioning data are available from other sensors, so that the resulting X, Y, Z solution can be computed and recorded in the aircraft, provided that differential corrections are transmitted from the Reference Station.

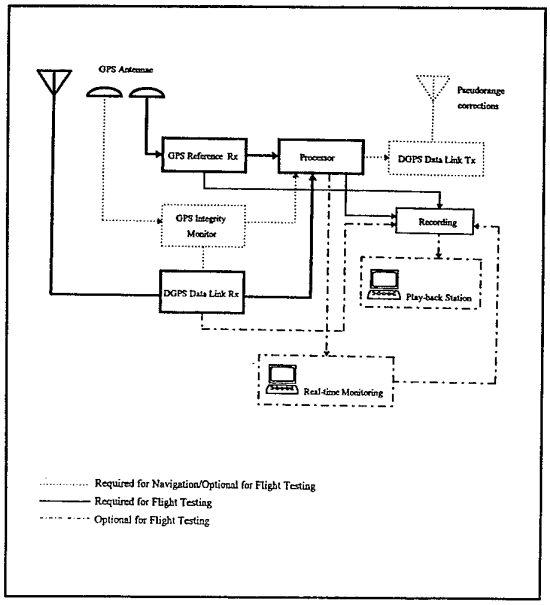


Figure 2. DGPS Reference Station.

A range system concept employing a translator is shown in Figure 3. The carrier for the retransmitted signal is derived from the translator local oscillator and used at the receiving site to aid signal tracking and correct translator local oscillator error [McConnell, 1989].

Vehicle position and velocity are then estimated using special processing algorithms.

Two types of translators, analog and digital, are suitable for test range applications.

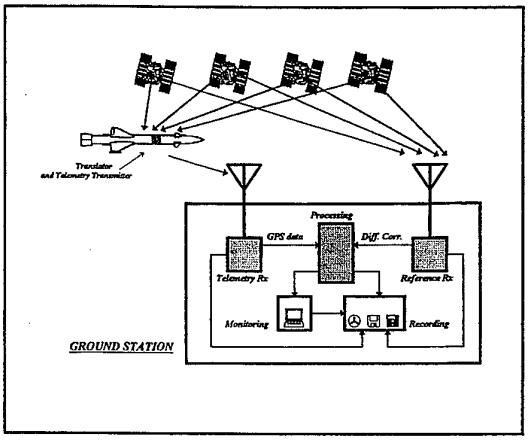


Figure 3. Translator System Concept.

Both types of translators commonly operate with the L1 C/A signals. The analog translator simply receives the satellite signals and retransmits them on another frequency. The digital translator is applicable to situations where the transmitted data must be encrypted or when telemetry and GPS signals are to be simultaneously relayed on the same link. However, the digital translator requires higher transmitted power than the analog translator for equivalent performance.

The advantages of using a translator are:

- it is less complex than a GPS receiver, which can reduce cost, size, and weight of this component up to an order of magnitude;
- it shifts the computational capability to ground base, thus allowing this capability to increase, and enabling Time-to-first-fix (TTFF) of typically less than 5 seconds for range safety applications;
- it allows mission replay, and it gives the system inherent differential GPS accuracy.

There are also some disadvantages when using translators, but they are not a factor for many applications. These disadvantages are:

- only a limited number of targets can be tracked simultaneously because of translator retransmission bandwidth requirements;
- TSPI is not directly available for use by the vehicle;
- recording onboard the vehicle is typically impossible, making a line-of-sight relationship to the master station mandatory (i.e., if loss of satellite signal track or masking of the GPS/Telemetry antennae occurs, position data will be totally lost).

Over the last fifteen years many translators have been built and tested and new processing techniques are now being investigated to increase the performance of existing equipment at the ranges. More information

about translators and techniques for receiving, processing, and recording translated GPS signals using standard telemetry equipment can be found in the literature [McConnell, 1989], [Brown, 1992].

The airborne GPS system can be either podded or installed in the aircraft avionics compartment. In both cases the same concept applies. There are many possible configurations for both the onboard/podded equipment and the Reference Station with supporting test facilities on the ground. Airborne and ground equipment selection criteria are mainly dictated by the task of the test mission and cost-vs-accuracy considerations. Figure 4 shows a typical layout.

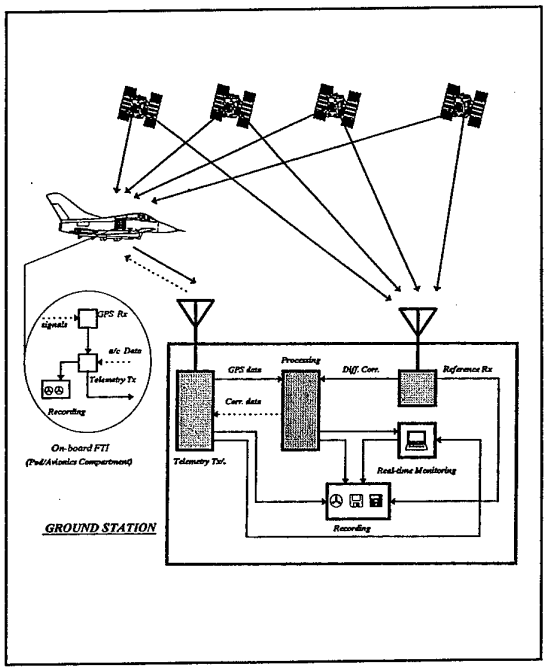


Figure 4. Receiver and Ground Station Concept.

The onboard GPS receiver, together with the other sensors and equipment carried onboard for recording, transmitting/receiving and processing data, are commonly referred to as the airborne "Flight Test Instrumentation" (FTI).
The receiver for use on aircraft have to be designed to minimise reacquisition time (typically multi-channel C/A or P(Y)-code receivers are used) after loss of satellite signals. GPS signal loss occurs for a variety of reasons associated with either extremely high dynamics or platform shading between the antenna and the satellite. In order to minimise the loss of signals different techniques can be used. The strategy usually adopted consists of using a dual antenna system. Each antenna has a hemispherical coverage pattern. The antennas "look" then in opposite directions. In this configuration, the upper antenna looking at the satellite constellation is used most of the time. The lower antenna (looking downward) which could be receiving the interfering multipath signal, should be "de-weighted" in the selection process.

EQUIPMENT SELECTION

The foreseen capabilities of GPS, in terms of data accuracy, quickness of data availability and reduction of cost, moved the Italian Official Test Centre (It-OTC), ALENIA (ALN) and AERMACCHI (AEM), to undertake a study aimed at defining the requirements of a DGPS-based system for flight testing.
The study was mainly addressed to GPS using C/A code, with post-flight differentiation. This has been preferred to GPS using P-code due to both simplicity of use and high accuracy attainable notwithstanding its lower cost.
The working group produced a document identifying all technical requirements of the DGPS system. After contacting many potential suppliers, an initial assessment of different systems was conducted in order to select the DGPS systems best matching the technical requirements set by It-OTC, ALN and AEM.
After a preliminary technical analysis, only two systems were selected. Particularly, the system proposed by ASHTECH (ASHTECH XII for both the airborne receiver and the ground reference station) and TRIMBLE (airborne receiver: TANS, ground station: TRIMBLE 4000SE) fulfilled the requirements stated in the specification document.

DGPS TEST CAMPAIGNS

A preliminary ground trial was performed with the ASHTECH receivers in order to test the accuracy of the DGPS data and to gain a good level of confidence with differential techniques, before performing actual flight trials.
The first flight test campaign was carried out on the MB339-C aircraft. Aim of this assessment was to compare the performance of the TANS III and the ASHTECH receiver in a dynamic environment in order to select the system with the best performance. The final in-flight evaluation was carried out on TORNADO. The primary aim of this assessment was to determine the ASHTECH system accuracy and to identify the critical conditions in which the GPS was likely to lose the signal or provide inaccurate data.

During the preliminary ground session, an ASHTECH XII receiver was installed on an electrically powered trolley and a second ASHTECH receiver located in a definite site to provide data for differential corrections. The trolley covered a well known route of about 3 km in the Turin International Airport (Caselle). The track was completed twice (in opposite directions) along a road about 6.2 metres wide. The trolley track is shown in Figure 5, where also the Reference Station position is evidenced.
After the trial, data stored into the internal memories of the two receivers were downloaded to a personal computer for differential processing and analysis.
Positioning data after differential processing and post-processing noise reduction were very precise, so that all data points were laying within the road.

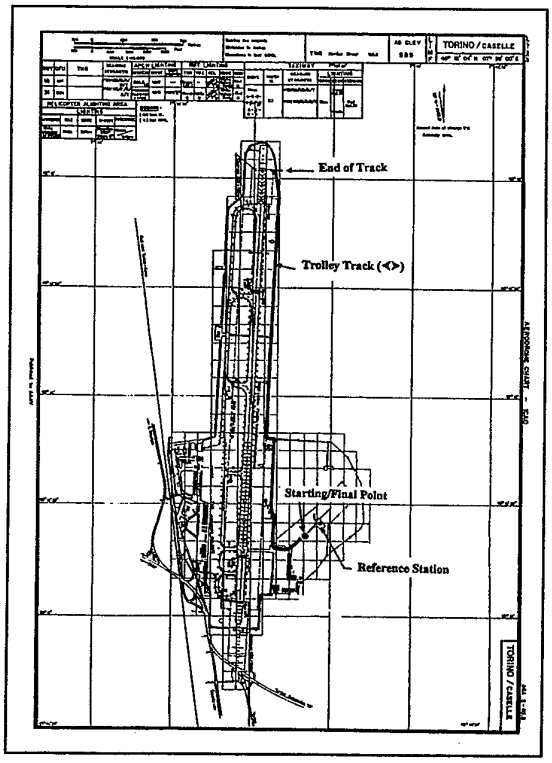


Figure 5. Caselle Aerodrome Chart and Trolley Track.

The stand-alone accuracy of the receiver was within the specification limits (i.e., 100 m 2dRMS with SA). The results obtained with the ASHTECH differential processing were very encouraging with a position data accuracy of 3 m SEP.
Even if the ground test results could not be considered exhaustive for demonstrating the performance of the DGPS system, the activity permitted to gain a certain level of confidence with differential techniques, essential for planning and executing the flight testing phase of the program.

During the MB339-C test campaign, specific dynamic manoeuvres were included in the aircraft flight profiles in order to allow a comparative evaluation of the TANS and ASHTECH receivers. Particularly, the following manoeuvres, representative of the real conditions which can be encountered during operational flight, were performed:

- turns with constant bank angle (to evaluate the antenna masking effect);
- dives followed by a pull up to 4 g's along the 4 cardinal directions (typical weapon aiming manoeuvres);
- high dynamic manoeuvres as tonneau and stick-jerks.

It must be underlined, that a determination of the accuracy provided by the systems was not considered essential in this phase. However, some overshoots of defined ground sites were performed in order to obtain a rough estimation of the systems accuracy.

As expected, during straight-and-level flight both receivers under test performed satisfactorily, with no significant data losses recorded. However, during execution of dynamic manoeuvres both systems frequently lost lock to the satellites. Analysing the data collected in flight, it was understood that this phenomenon was primarily due to shielding of the GPS antenna by the aircraft body (wings, fuselage and tails). Data analysis also confirmed that during dynamic manoeuvres both systems experienced a very significant increase of the PDOP factor.

In Figure 6 are shown some diagrams relative to the variation of the flight parameters (heading, altitude, pitch, bank, etc.) in a period of a few minutes, as given by the aircraft INS during one of the flight trials.

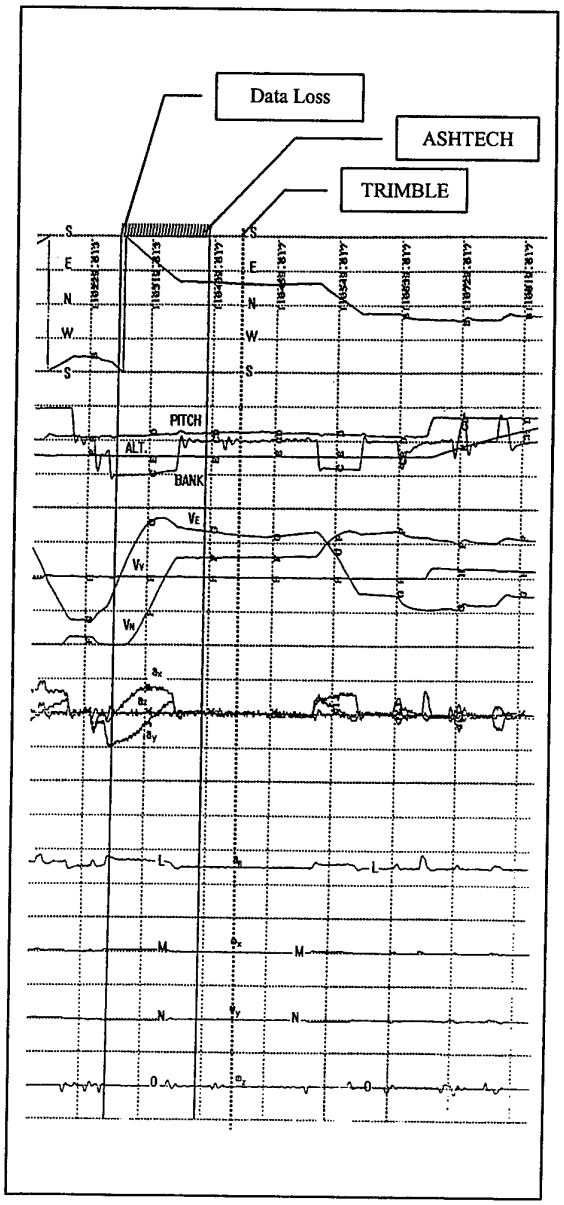


Figure 6. INS data - ASHTECH and TRIMBLE data loss.

Identification of the parameters, together with the various scales, is the following:

- B = Magnetic Heading (-180°÷180°)
- C = Roll Angle (-90÷90°)
- D = Pitch Angle (-40÷40°)
- E = Barometric Altitude (0÷40000 ft)
- F = NORTH Velocity (-800÷800 ft/s)
- G = EST Velocity (-800÷800 ft/s)
- H = Vertical Velocity (-800÷800 ft/s)
- I = Along X Acceleration (-80÷80 ft/s^2)
- J = Along Y Acceleration (-80÷80 ft/s^2)
- K = Along Z Acceleration (-80÷80 ft/s^2)
- L = Normal Acceleration (-2÷6 g)
- M = Angular Velocity around X (-40÷40 °/s)
- N = Angular Velocity around X (-40÷40 °/s)
- O = Angular Velocity around X (-100÷100 °/s)

Particularly, the diagrams are relative to a left turn (-45° bank angle) with GPS signal loss (for both the ASHTECH and the TRIMBLE receivers).
The Signal-to-Noise Ratios (SNR) measured for each satellite are shown in Figure 7.

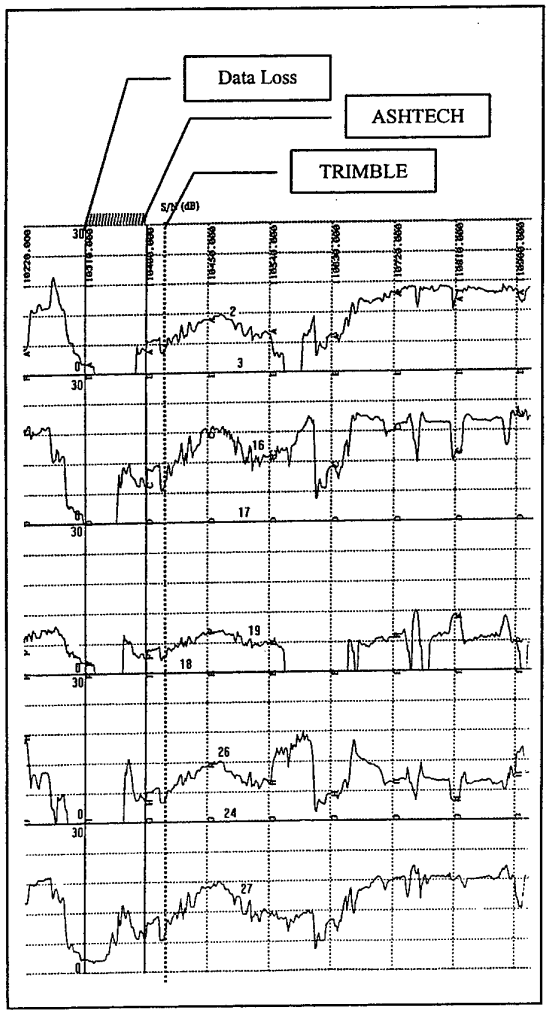


Figure 7. SNR's - ASHTECH and TRIMBLE data loss.

It is evident that the signal intensity did not decrease gradually. This confirms that the signal loss, in this case, was due to interposition of an obstacle between the satellites and the antenna and not to receiver tracking problems.
Figure 8 gives an idea of the relative positions of the aircraft and the satellites during the manoeuvre. It can be noticed that in consequence of the left turn many satellites were masked by the aircraft body. Signal reacquisition took place when the aircraft progressively reduced the bank angle and the heading variation rate.

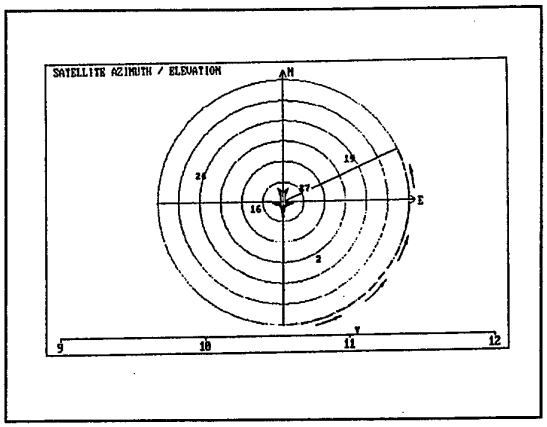


Figure 8. Satellite positions during data loss.

Although both receivers simultaneously lost the GPS data, a difference was noticed between the two receivers in terms of reacquisition time. As the position of the aircraft longitudinal axis during the turn got closer to satellite 2 (in the horizontal plane) the TRIMBLE receiver experienced a loss of track to all satellites, while satellites 19 and 27 were still tracked by the ASHTECH receiver. This was due to the internal processing of the receiver.
Prior flight both receivers were programmed with the same PDOP threshold, but the ASHTECH receiver was able to maintain track to the satellites in view even when their SNR's were very low, while TANS lost track to all satellites (the ASHTECH XII is a 12 channels "All in view receiver" that uses all available satellites in view to generate a solution, while the TANS is a more traditional receiver that choose the four satellites of those available that give the best geometry to perform a position fix). In dynamic conditions, this significantly reduced the time required by the ASHTECH receiver to perform a new position fix as soon as enough satellites were available again.

Other trials also confirmed that the increase of PDOP and the reduction of SNR's during dynamic manoeuvres were responsible for satellite signal losses in the TRIMBLE receiver (even with small variation of the aircraft-receiver relative geometry and in absence of satellite masking), and that the ASHTECH receiver needed shorter periods to compute a new positioning solution after losses of satellite signals caused by masking.

Figure 9 is referred to a stick-jerk manoeuvre.

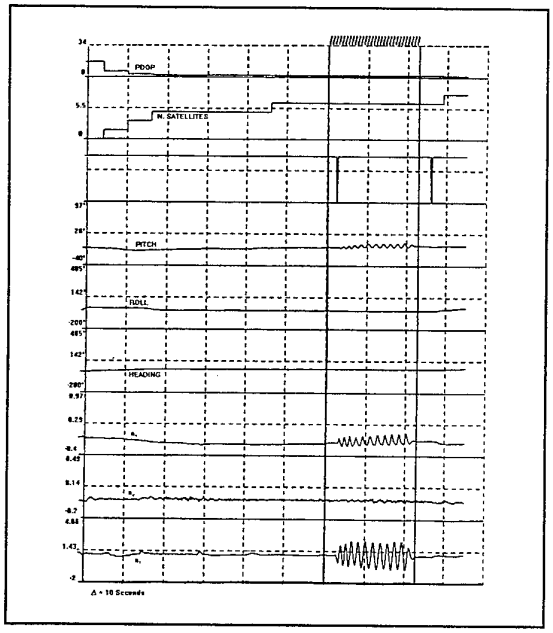


Figure 9. Stick-jerk manoeuvre.

During this manoeuvre the stick is repetitively pulled and pushed in order to obtain high jerks. This can be a very critical manoeuvre for the GPS receivers (especially for the code correlation circuits). The jerk limits were not specified for the ASHTECH receiver, while for TANS a jerk limit of 2 g/s (20 m/s^3) was quoted. During the manoeuvre in Figure 16 a jerk of about 2.8 g/s was obtained but no data loss occurred in any of the two receivers.

The data in Figure 10 were recorded during three pull up manoeuvres at 4 g's (typical in weapon-aiming). While during execution of the manoeuvres lock on to the satellites was kept (between four and three satellites tracked), at the end of each manoeuvre total signal losses occurred. The three manoeuvres were always preceded by straight-and-level flight (about 30 seconds) to allow optimal satellite tracking.

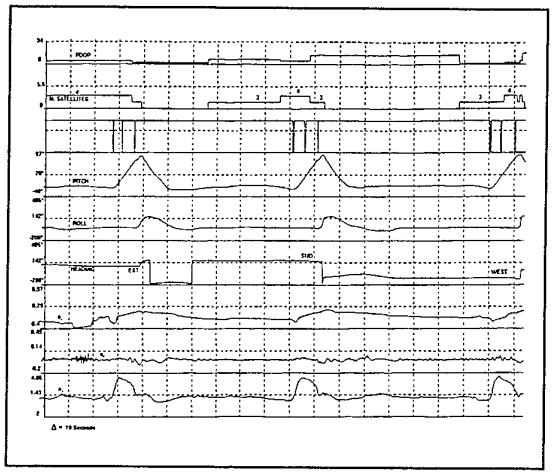


Figure 10. Pull-up manoeuvres (4 g's).

Analysing the data of the MB339-C test campaign, it was concluded that the ASHTECH receiver was better suited than the TRIMBLE receiver for flight testing applications. Particularly, the ASHTECH tracking and reacquisition strategy significantly reduced the time required for a new fix after data losses.
The ASHTECH data continuity during dynamic manoeuvres was significantly better than that of TRIMBLE. On average, the GPS data loss occurred in 25% of the total flight time for the ASHTECH receiver and in 35% of the time for the TRIMBLE receiver. These results were obtained with an identical setting of the threshold parameters for position computation.
Although the TRIMBLE receiver was able to provide a positioning solution even with only three satellites tracked, in this case the accuracy degradation of the horizontal co-ordinates (especially the latitude) was very significant and not acceptable for flight test applications.

FINAL TEST CAMPAIGN

The general layout of the TORNADO installation is shown in Figure 11. The antenna was located on the aircraft skin at about 1.5 m from the cockpit and a few decimetres from the telemetry antenna.
The Electro-Magnetic Compatibility (EMC) of the GPS equipment with the other on-board systems was verified before the flights.

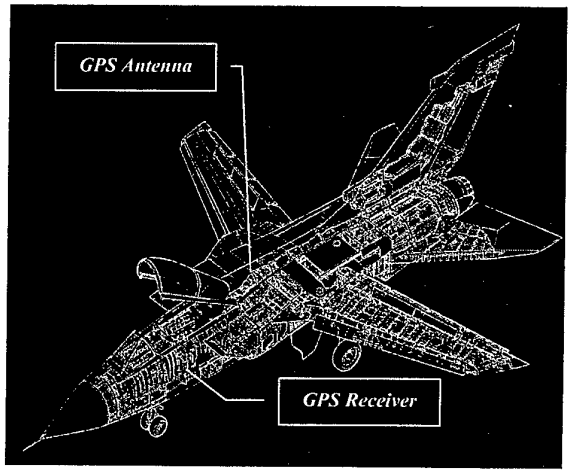


Figure 11. TORNADO installation.

During the trials, the two ASHTECH XII receivers (onboard and ground reference) were used in conjunction with the RANGER and P-NAV post-processing software packages.
Analysing several data from the preliminary test campaign, it was clear that the main disadvantage of the GPS is its vulnerability to signal losses caused by satellites masking. The primary objective of the final test campaign was to thoroughly investigate the masking problem, to test the capability of the ASHTECH receiver to reacquire satellite signals and to provide TSPI even with degraded satellite constellations.

In order to assist in the investigation a special simulation software was used to calculate the global masking effect due to antenna masking and aircraft structural masking.
A preliminary assessment was also carried out of the Doppler and Multipath effects influence on data quality. Moreover, some optimisation criteria were defined for the use of DGPS in flight test missions.
Finally, a data quality assessment was carried out by comparing DGPS with other known references (i.e., radar altimeter, laser range finder and optical tracking system), and a simple integration algorithm was used for merging DGPS and INS data in order to recover DGPS data losses by using INS data.

The main reason of interruption of the GPS signals in flight was the physical masking of the antenna by the aircraft body (especially the wings and the tails) during turns. In order to verify the masking conditions a special simulation software was implemented, called VIEWSAT, which modelled the aircraft shape and accepted input data such as the aircraft heading pitch and bank angles in flight, and relative position of the satellites (azimuth and elevation) tracked during flight. The VIEWSAT software provided a visibility matrix (one dorsal antenna) for the defined flight conditions (Figure 12).

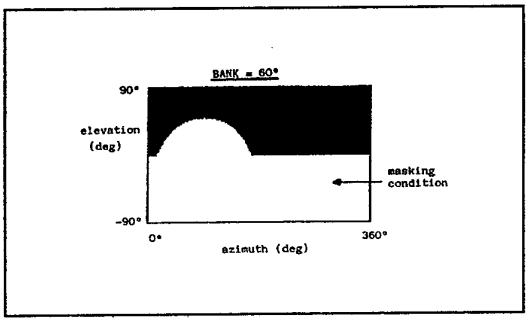


Figure 12. Example of Antenna Masking Matrix.

The software required the definition of a simplified model of the aircraft, as shown in Figure 13.

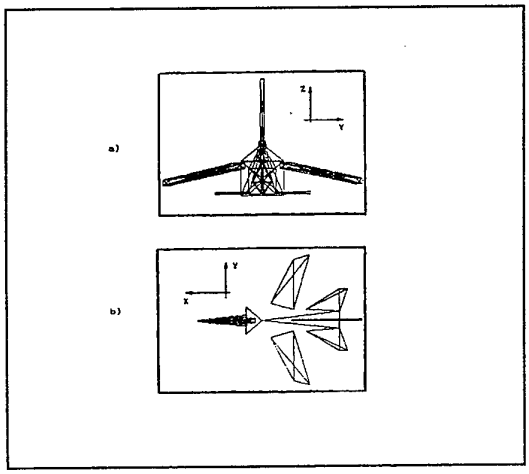


Figure 13. Simplified Model of TORNADO.

The result was a global masking matrix, as shown in Figure 14.

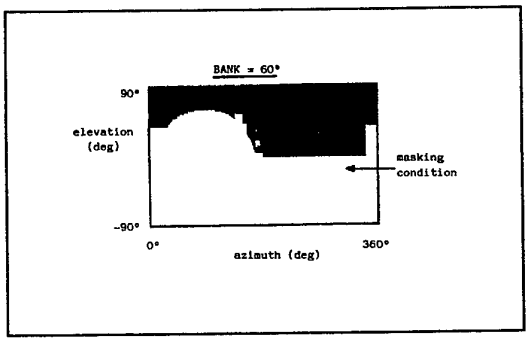


Figure 14. Example of Global Masking Matrix.

The program output was a binary diagram in which for every satellite masked a "0" was shown, while unmasked satellites corresponded to the satus "1". An example of VIEWSAT output, together with the related flight conditions, is shown in Figure 15.

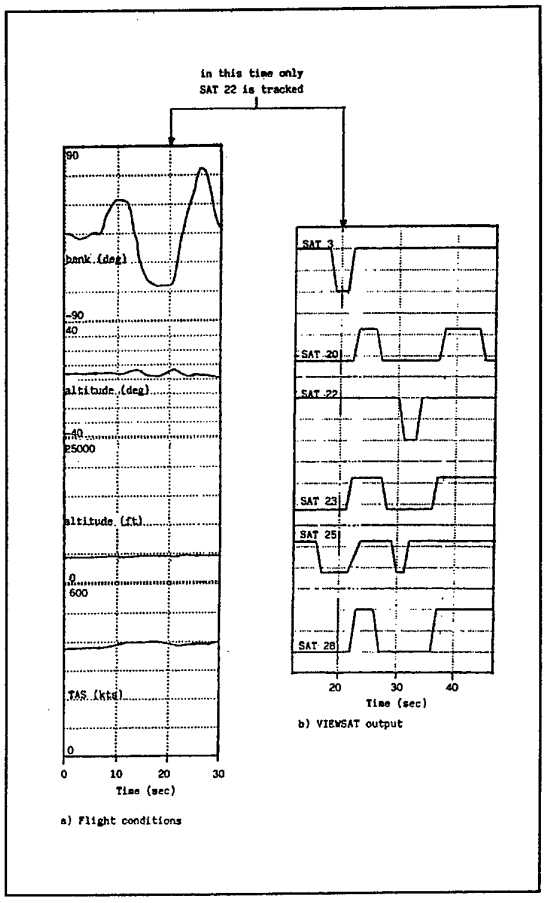


Figure 15. VIEWSAT output and related flight conditions.

For every flight segment with loss of GPS data, satellite masking was investigated using the VIEWSAT. This was done in the attempt to determine the critical aircraft manoeuvres using the time histories of the flight parameters recorded in flight (particularly

Heading, Bank and Pitch Angles, TAS, Baro height and Radalt height).

During a flight carried out with a number of visible satellites between 6 and 7, aircraft assets analysis allowed the identification of some critical manoeuvres and flight conditions. During the trial, the maximum variation of the satellite position was in the order of 30° in azimuth and 20° in elevation. Satellites out of visibility were generally characterised by a low elevation above the horizon (10° to 20°). During manoeuvres, the satellites more likely to be lost were the satellites with an elevation of less than 30°.
The most critical manoeuvres were characterised by a simultaneous non-gradual variation of two asset parameters (pitch, roll or yaw). A fast variation of height (if not associated to a considerable variation in pitch) was not sufficient alone to determine satellite losses (Figure 16).

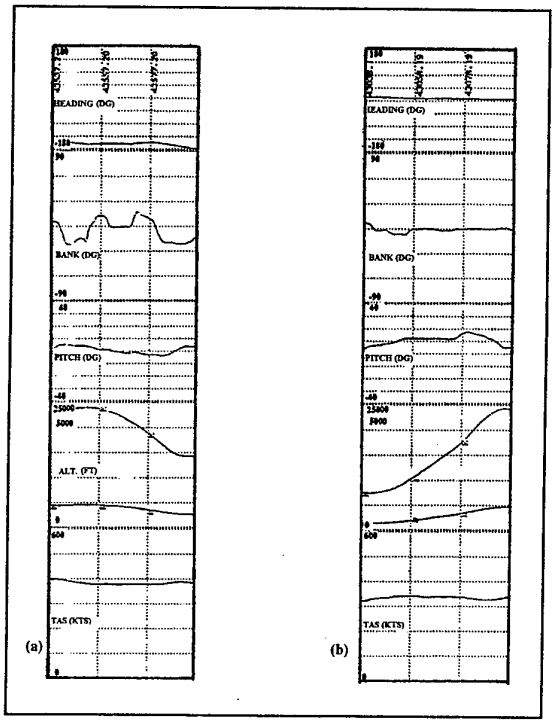


Figure 16. Height variations w/o satellite losses during horizontal manoeuvres (a) and in vertical flight (b).

A manoeuvre that was found to be very critical was the ***turn*** in the following conditions:

- ***bank ≥ 50°;***
- ***heading variation about 180°.***

In Figure 17 two situations are shown in which these elements can be identified.
There are some important consideration to be done about the Critical Bank Angle (CBA). It is in fact necessary to take into account that, during a turn, the satellites lower over the horizon were likely to be masked, but the manoeuvre determined a heading variation which was critical only if performed along a line between two adjacent satellites (Figure 18).

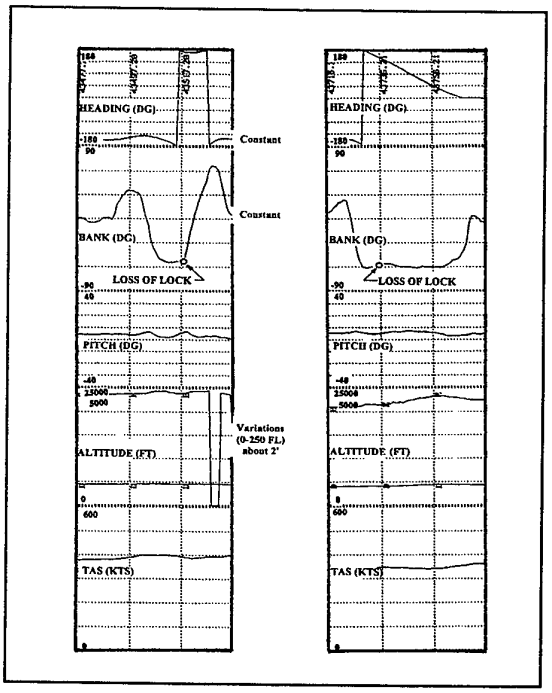


Figure 17. Critical manoeuvres with loss of lock to the satellites.

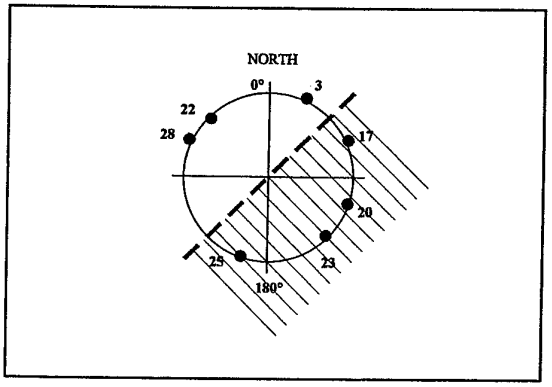


Figure 18. Satellite masking (SVs 17, 20, 23 and 25).

In many situations, however, the same manoeuvre did not determine the loss of lock. The difference between the two cases (loss and no-loss) was represented by the fact that in the no-loss case the turning manoeuvre followed a period of stabilised flight (without significant heading, bank and height variations) of about 40 seconds.

On average, the ASHTECH receiver was able to give a new positioning solution within 20 seconds from reacquisition of the satellite signal. Therefore, it was generally sufficient to maintain a low bank angle for about 20 seconds in order to obtain a new solution. However, if the height was kept constant the time required was reduced to about 10 seconds. It is in fact possible that the receiver was able to store in its internal memory, for about 10 seconds, the last positioning computations and used these data to provide a positioning solution even without inclusion of new measurements. This hypothesis is compatible

with the fact that the receiver uses an adaptive Kalman Filter [Gelb, 1992].
Another significant aspect taken into account was the influence of SNR on satellite reacquisition. Particularly, it was demonstrated by data analysis that the ASHTECH receiver provides a positioning solution only if the SNR of at list 4 satellites is above a pre-defined value (presumably between 15 and 30 dB) which can not be selected by the user at the ground programming stage.

TEST MISSION PLANNING AND OPTIMISATION

As a result of the analysis carried out, some recommendations were formulated in order to optimise the use of DGPS as a datum in flight test missions. Particularly, the following criteria were adopted:

- 4 satellites always in view with an elevation near 50°;
- 50° maximum bank angle;
- at least 20 sec of stabilisation before and after significant flight phases;
- gradual heading variations;
- distance between the aircraft and the ground receiver not greater than 200 NM.

These restrictions, of course, imply operational limitations which reduce the spread of possible applications, unless additional tools are used in support of GPS.

DGPS/INS INTEGRATION FOR FLIGHT TESTING

Although many tasks are fulfilled now by sole means of DGPS, there are still areas where integration with an inertial navigation system (INS) is necessary. Particularly, in flight testing of modern fighter aircraft DGPS cannot provide the necessary information with regard to data rate and data continuity during high dynamic manoeuvres.
A study was therefore undertaken in order to investigate the potential of DGPS/INS integration. As a first step, an assessment was carried using a simple post-processing procedure for recovering DGPS data losses with INS measurements.
Taking into account the satisfactory results obtained and the many benefits of integration documented by the scientific literature available on the subject, the various options for integrating DGPS and INS were investigated and the concept of an integrated Position Reference System (PRF) for flight testing, using state-of-the-art technology, was defined. Together with real-time C/A code DGPS, also DGPS using carrier phase ambiguity resolution on-the-fly was investigated and different options for the INS mechanisation were considered.

A simple method was developed for post-processing recover of DGPS data losses due to antenna masking, hard manoeuvring or bad satellite configuration. The method utilised the direct integration of data provided by the onboard INS starting from an initial reliable position determined by DGPS (i.e., obtained after a stabilisation of at least 20 seconds, with at least 4 satellites in view and PDOP<3). The integration was carried out up to the next reliable DGPS relief.
The difference after integration between the so calculated aircraft position and that provided by DGPS allowed the determination of the average value of the inertial drift derivative with respect to time, which was used to minimise the inertial error during reconstruction of the aircraft trajectory for the whole time slice considered.
So, it was possible to provide an accurate aircraft trajectory even when the manoeuvres led to GPS data losses, without the use of external reference systems (e.g., cinetheodolites, tracking radars, etc.) or pilot fixes (which require good weather conditions and properly instrumented test ranges).
In Figure 19.a, referring to three consecutive orbits lasting for about eleven minutes, DGPS latitude data are compared with inertial latitude corrected by using the method described. The difference between the two sets of data (i.e., DGPS/INS latitude error) is magnified in Figure 19.b and compared with the inertial latitude error without DGPS corrections.

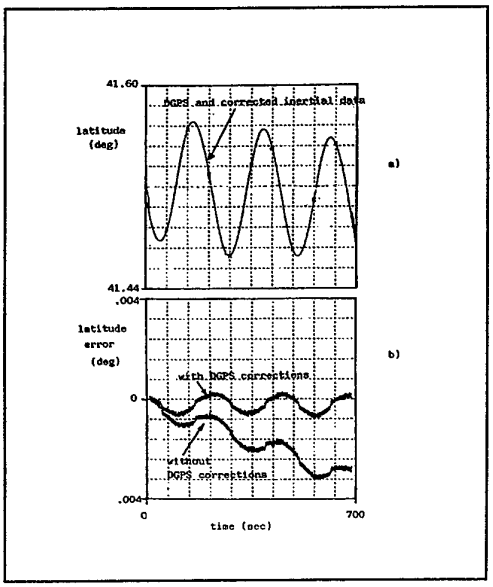


Figure 19. Example of DGPS and INS Data Merging.

The benefits of integrating (D)GPS with an INS are significant and diverse. Basically, each system has important shortcomings. Used in concert through some adequate algorithm, the integrated system solves the majority of these problems. All of the GPS techniques (either stand-alone or differential) suffer from some shortcomings. The most important are:

- the data rate of a GPS receiver is too low and the latency is too large to satisfy the requirements for trajectory analysis, in particular with respect to the synchronisation with external events. In addition, there might be a requirement for high-rate and small latency trajectory data to provide real-time guidance information to the pilot (e.g., for fly-over-

noise measurements). With the need to process radio frequency signals and the complex processing required to formulate a position or velocity solution [Siouris, 1993], GPS data rates are usually at 1 Hz, or at best 10 Hz (an update rate of at list 20 Hz is required for real-time applications);

- high DGPS accuracy is limited by the distance between the Reference Station and the user because of the problem of ionosphere in integer ambiguity resolution on-the-fly [Hein, 1993];
- Selective Availability (SA) degrades differential GPS positioning accuracy over long distances;
- influences of high accelerations on the GPS receiver clock, the code tracking loop (delay lock loop) and carrier phase loop may become significant;
- Signal loss-of-lock and 'cycle slips' may occur very frequently due to aircraft manoeuvres or other causes.

As discussed earlier, there is little to nothing one can do in order to ensure the continuity of the signal propagation from the satellite to the receiver during high dynamic manoeuvres. Due to shadowing of the GPS antenna the receiver will loose track to several and in many cases to all satellites.
Stand-alone INS has its shortcomings as well. The INS is subject to an ever growing drift in position accuracy caused by various instrument error sources that cannot be eliminated in manufacturing, assembly, calibration or initial system alignment [Siouris, 1993]. Furthermore, high quality inertial systems (i.e., platform systems) tend to be complex and expensive devices with significant risk of component failure.
The integration of (D)GPS and INS measurements might solve most of the above mentioned shortcomings: the basic update rate of an INS is 50 samples per second or higher and an INS is a totally self-contained system. The combination of INS and (D)GPS will therefore provide the required update rate, data continuity and integrity. The other advantages of an INS: low short term drift and low noise, are combined with the advantages of (D)GPS: high position accuracy and no long term drift. Therefore, in the case of DGPS/INS integration, the real-time demands on the telemetry data link are not very high because INS errors show a long term drift only and thus do not need to be updated frequently [Kleusberg, 1990].

Technical considerations for integration of DGPS and INS include the choice of system architecture, the integration algorithm (mostly Kalman Filter), and the characterisation and modelling of the measurements produced by the two sensors. Traditionally, GPS has been used to update the position of the INS (in other words, to control the drift behaviour of the INS). When looking to the high accuracy potential of DGPS, it is obvious to go the other way around, namely to get the positioning information primarily from GPS. Thus, the INS can be the secondary sensor enabling higher interpolation in GPS positioning updates, providing the attitude information, damping short periodic influences in GPS, and assisting in cycle slip detection and on-the-fly ambiguity resolution algorithms [Hein, 1993]. Clearly, both types of observations enter the Kalman filter, and only the different weighting of the data decides which sensor mainly contributes to the integrated result.
In principle, integration of DGPS with the INS can be done in levels analogously to the DGPS methods, using:

- DGPS differential pseudorange corrections;
- DGPS carrier phase-smoothed pseudorange corrections;
- DGPS carrier phase corrections [Gloecker, 1992].

These are also the data which can be transmitted by telemetry from the reference station to the user when following the RTCM standardised messages. For a real-time fully integrated DGPS/INS application it is clear that also the appropriate raw data, pseudoranges and carrier phases have to be transmitted, perhaps in compressed form in order to minimise the telemetry load.
The integration of INS with DGPS can be carried out in many different ways depending on the application (i.e., online/off-line evaluation, accuracy required, etc.). In the following paragraph some information are given about integration algorithms for both real-time and post-processing systems.

INTEGRATION ALGORITHMS

Various options exist for the integration algorithm. In general, two main categories can be identified depending on the application: post-processing and real-time algorithms. The complexity of the algorithm is obviously related to both system accuracy requirements and computer load capacity. While for many applications a post-processing solution is acceptable, this is not useful for both navigation and online evaluation during a flight test. The state-of-the-art integration algorithm is the Kalman Filter (KF). In general terms, a KF is a recurrent, optimal estimator used in many engineering applications whenever the estimation of the state of a dynamic system is required. A KF can be used to estimate the errors which affect the solution computed in an INS or in a (D)GPS, as well as in the combination of both navigation sensors.

A number of implementations are possible for the DGPS/INS Kalman Filter. Matrix formulation methods of various types have been developed to improve numerical stability and accuracy (e.g., square root and stabilised formulations), to minimise the computational complexity by taking advantage of the diagonal characteristics of the covariance matrix (U-D factorisation formulation), and to estimate state when the state functions are non-linear (extended Kalman filter). In general, the KF algorithm is optimal under three limiting conditions: the system model is

linear, the noise is white and Gaussian with known autocorrelation function and the initial state is known. None of them are strictly verified in practice. Moreover, the computing burden grows considerably with increasing number of states modelled. Artificial Neural Networks (ANN) could be used to relax one or more of the conditions under which the KF is optimal and/or reduce the computing requirements of the KF. While the total replacement of KF with ANN filters lead to difficult design and unreliable integration functions, a hybrid network in which the ANN is used to learn the corrections to be applied to the state prediction performed by the rule-based module, appears as the best candidate for future sensors integration.

Post-processing integration of (D)GPS and INS measurements can be carried out by means of the Rauch-Tung-Striebel algorithm, which consists of a Kalman filter and a backward smoother. The forward filter can take into account previous measurements only. The smoothed estimate utilises all measurements. It is obvious that the smoothed positions are always at least as accurate as the forward filtered positions. In most cases, however, the smoothed positions are considerably more accurate than the filtered ones. It is evident that backward smoothing can only be done off-line. The filter estimates the state vector together with the error covariance matrix. The state vector contains the error components of the inertial navigation system. The smoothed trajectories and also accurate velocities are obtained by adding the state vector to the measurements delivered by the INS. The equations of the Rauch-Tung-Striebel algorithm can be found in [Gelb, 1992]. With this system, an integration between (D)GPS and INS is possible and the high frequency movements of the aircraft are not smoothed out. Both positions and velocities can be obtained with a high degree of accuracy and are nearly continuously available. This is also true when the time interval between the measurements is relatively long (e.g., 1 minute). There is a need to carry out a great amount of calculations and to store much data for the backward filter, but with modern computers this is not a major problem.

The filtering algorithm most commonly implemented in real-time systems is the Bierman's U-D Factorised Kalman Filter. This algorithm avoids the explicit and computation of the estimation error covariance matrix P_t by propagating in terms of its factors U and D:

$$P_t = UDU^T \tag{5}$$

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 [Bierman, 1977]. The U-D algorithm is efficient and provides significant advantages in numerical stability and precision. Specifically, the factorisation of P_t provides an effective doubling in computer word length in covariance-related calculations, and avoids filter divergence problems which can arise in more conventional filter mechanisations [Gelb, 1992].

Artificial Neural Networks technology has not been fully investigated for application to integrated air navigation systems; however, the suitability of such technology to replace or enhance the performance of the Kalman Filter has been proven in related areas [Dumville, 1994], [Currie, 1992], [Ashkenazi, 1993]. The development of an integration algorithm completely based on neural technology (e.g., Hopfield Networks) appears impracticable at the moment. Furthermore, this solution is even more demanding than the Kalman Filter in terms of computing requirements. Techniques for "On-line Training" would allow for real-time adaptation to the specific operating conditions but further research is required in this field [Lo Conte, 1996].
Hybrid Networks emerge as the best candidate for future applications. In an hybrid architecture an ANN is used in combination with a rule based system (i.e., a complete Kalman Filter or part of it), in order to achieve the required adaptivity and improve the availability of the overall system. A hybrid network provides performances comparable with the Kalman Filter, but with improved adaptivity to non-linearities and unpredicted changes in system/environment parameters. Compared with the Kalman Filter, the hybrid architecture features the high parallelism of the neural structure allowing for faster operation and higher robustness to hardware failures. The use of ANN to correct the state variables prediction operated by the rule-based module avoids computing the Kalman gain, thus considerably reducing the computing burden. Stability may be guaranteed if the output of the network is in the form of correction to a nominal gain matrix that provides a stable solution for all system parameters. The implementation of such a filter, however, would be sensitive to network topology and training strategy. An adequate testing activity would therefore be required.

INTEGRATION ARCHITECTURES

Combining (D)GPS and INS, different depths of integration can be realised. The level of integration and the particular mechanisation of the Kalman filter are dependent on:

- the task of the integration;
- the accuracy limits;
- the robustness and the stand alone capacity of each subsystem;
- the computer time capacity;
- the INS sensor concept (platform/strapdown).

For these tasks, the basic concepts of system integration can be divided into the following topics:

- open loop (D)GPS aided INS (OLDI);
- closed loop (D)GPS aided INS (CLDI);
- fully integrated (D)GPS/INS (FIDI).

Only a brief description of the various architectures is given here. Further information can be found in [Siouris, 1993] and [Jacob, 1989].

The simplest way to combine (D)GPS and INS is a reset-only mechanisation in which (D)GPS is used to periodically reset the INS solution. In this open-loop strategy the INS is not re-calibrated by (D)GPS data, so the underlying error sources in the INS still drive its navigation errors as soon as (D)GPS resets are interrupted. However, for short (D)GPS interruptions or for high quality INS, the error growth may be small enough to meet mission requirements. This is why platform inertial systems have to be used with OLDI systems operating in a high dynamic environment (e.g., military aircraft). The advantage of the open loop implementation is that, in case of inaccurate measurements, just the Kalman filter is influenced and not the inertial system calculation itself. As the sensors of a platform system are separated from the body of the aircraft by gimbals, they are operating normally at their reference point zero [Jacob, 1989]. The attitude angles are measured by the angles of the gimbals. The Kalman filter can run internal or external to the INS, but the errors of the inertial system have to be carefully modelled.

The main advantage of (D)GPS aiding the INS in a closed-loop mechanisation (CLDI) is that the INS is continuously calibrated by the Kalman Filter, using the (D)GPS data. Therefore strapdown sensors can be used in a CLDI implementation.
In contrary to platform systems, sensors of strapdown systems are not uncoupled from the aircraft body. They are operating in a dynamically more disturbed environment as there are vibrations, angular accelerations, angular oscillations which result in an additional negative influence to the system performance. In addition sensors are not operating at a reference point zero. Therefore the errors of the system will increase very rapidly and problems of numerical inaccuracy in an open loop implementation may soon increase. This is why it is advantageous to loop back the estimated sensor errors to the strapdown calculations to compensate for the actual system errors. Consequently, the errors of the INS will be kept low and linear error models can be used. When (D)GPS data is lost due to dynamics or satellite shadowing, the INS can continue the overall solution, but now as a highly precise unit by virtue of its recent calibration. However, the CLDI implementation can be unstable.

The system integration of best accuracy will be of course the full integration of both systems (FIDI) which requires a Kalman filter implementation at a row/uncorrelated measurement level.
As the Kalman filter theory asks for uncorrelated measurements [Gelb, 1992], it is necessary to use either the raw GPS measurements (i.e., the range and phase measurements to at least four satellites, the ephemeris to calculate the satellite positions and the parameters to correct for the ionosphere and troposphere errors). In this case, the receiver clock errors (time offset and frequency) can be estimated as a part of the filter model.
This approach has very complex measurement equations but requires only one Kalman Filter mechanisation. Moreover, its filtering can be most optimal since both (D)GPS errors and INS errors can be included without the instability problem typical of cascade Kalman Filters.

When comparing OLDI, CLDI and FIDI architectures for practical applications, the most important distinction to be made is between cascaded and non-cascaded approaches, which correspond, as mentioned before, to aided (OLDI and CLDI) and fully integrated (FIDI) architectures respectively. In the cascaded cases two filters generally play the role. The first filter is a GPS filter which produces outputs (i.e., position and velocity) which are correlated between measurement times. This output is then used as input for the second filter which is the INS Kalman filter. As time correlation of this measurement input does not comply with the assumptions underlying the standard Kalman filter, it must be accounted for in some way [Napier, 1989]. This will complicate the Kalman filter design (i.e., the potential instability of cascaded filters makes the design of the integration Kalman filter a very cautious task). However, form a hardware point of view aided systems result in the simplest solution: commercially available GPS receivers and INS systems can be used without significant modifications.
In the non-cascaded case there is just a single Kalman filter generally based on an INS error model supplemented by a GPS error model. The GPS measurements, uncorrelated between measurement times, are differenced with the raw INS data to give measurements of the INS errors, also uncorrelated between measurements times.
In an Integrated Position Reference System (PRS) suitable for flight test applications strapdown sensors are preferably used, because the system has to be robust and because compatibility with GPS receivers of different characteristics and accuracy classes is desirable. Therefore, only the CLDI and FIDI integration schemes are considered for such a system. However, the instability problems associated with a closed-loop architecture, even thought to be solvable by adopting adequate techniques [Kerr, 1987], make the CLDI implementation a secondary option.

INTEGRATED PRS CONCEPT

In accordance with flight test requirements, an optimal PRS should be able to cover the following accuracy classes:

- A low accuracy class with stand alone GPS (C/A-code, position accuracy 100 metres) updating the INS. This is required for the evaluation/ certification of medium-range navigation sensors and systems (VOR/DME, TACAN, ADF, etc.);
- A high accuracy class (3-D position accuracy at the decimetre level), in which differential carrier range measurements can be used to update the INS. This

is suitable for a number of applications, such as remote sensing flights, fly-over-noise measurements, evaluation/certification of landing systems and determination of aircraft take-off and landing performance.

A data link would be required to transmit the differential corrections from the ground reference station to the aircraft. This would enable real-time position calculation in the aircraft and provide (optionally) guidance information to the pilot. Without the data link the system will automatically degrade to the low accuracy mode.

The basic hardware setup shown in Figure 20 is an optimal solution for flight test applications.

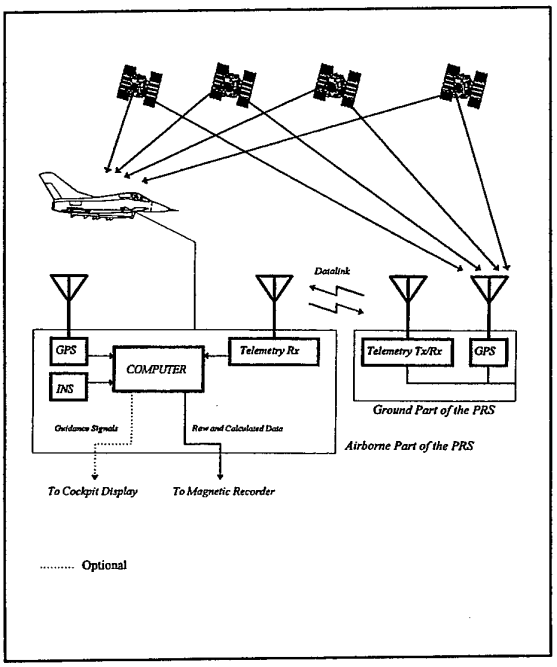


Figure 20. PRS hardware layout.

The computer receives data from the airborne GPS receiver (i.e., pseudoranges, carrier ranges, satellite orbit parameters and/or stand alone position), data from the GPS receiver at the Reference Station (i.e., pseudoranges/carrier ranges and reference position or differential pseudorange/carrier phase corrections), and raw data from the INS.
If only a telemetry uplink is used instead of a bi-directional link, real-time monitoring from the ground would not be performed and differential corrections must be calculated in the airborne computer. Moreover, when using raw data from the ground receiver instead of pseudorange and carrier phase corrections computed at the Reference Station, problems may arise of time sensitivity and telemetry load [Gloecker, 1992]. Using a bi-directional telemetry link and differential corrections from the ground Reference Station, high accurate (D)GPS data can be recorder in flight, used to provide guidance information to the pilot, and transmitted to the ground station for real-time monitoring with high (D)GPS accuracy.
The computer merges the data and sends its solution to a magnetic tape recorder for further processing and analysis after the flight. Optionally, the PRS should be able to send information to a display in the cockpit to provide real-time guidance to the pilot.

The Kalman filter uses an INS-(D)GPS error model, and accepts raw INS and (D)GPS data (i.e., pseudoranges, carrier phases and range rates) differentially corrected using the signals transmitted from the ground Reference Station.
The filter should be implemented in UDU^T form which provides a good balance between numerical performance and computational stability. The structure of the filter should be the same for the low and high accuracy applications. In the first case the on-board GPS delivers the data, in the latter case data should be delivered by a Differential Processing Module (DPM) in which the ground reference corrections are applied. The basic PRS computer software architecture is shown in Figure 21.

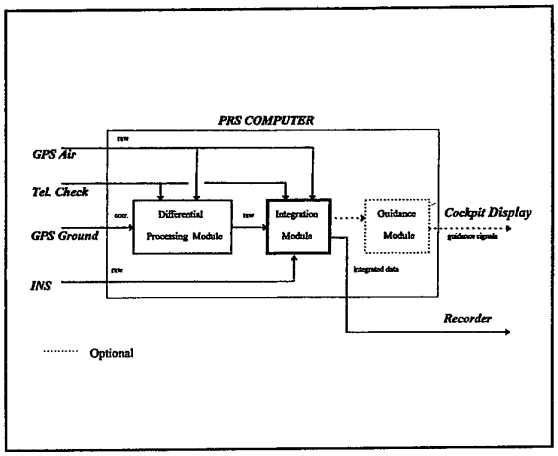


Figure 21. PRS Software Architecture.

The integration module will include the Kalman filter and a sub-module for automatic selection of GPS (onboard) or differentially corrected (DPM) data, depending on availability of the telemetry data link. Moreover, the DPM module should include:

- a sub-module for detecting and fixing cycle slips;
- a sub-module for solving the integer ambiguities;
- a sub-module for calculating the aircraft position.

At the moment, cycle slip fixing and ambiguity resolution algorithms are not mature enough to be used in real-time aircraft applications [Hatch, 1991], [Landau, 1992], [Euler, 1994], [Hansen, 1994]. However, the software has to be developed bearing in mind that an upgrade to the real-time system will be carried out as soon as enough confidence in the real-time performance of the algorithms will be established.

FUTURE WORK

Future work will include the Kalman filter implementation and the proof of the PRF performance with an adequate laboratory testing activity. After that, the system should be tested in flight.
The (D)GPS/INS Kalman filter may include up to 98 error states, but simplifications are possible, and provide acceptable performance with significant state reduction [Liang, 1995]. 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 [Gelb, 1992]:

- an aircraft trajectory generator which provides nominal flight data;
- a reference sensor error truth model which characterises all the sensor errors;
- the reduced-order Kalman filter design to be evaluated.

To conduct sensitivity performance 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 [Liang, 1995].
Once enough confidence has been gained in the performance of the Kalman filter, laboratory testing should be carried out in order to optimise the hardware and software architectures and to give a first estimation of the overall system accuracy.
In order to prove that a PRS based on GPS/INS integration meets its accuracy requirements, an independent reference system is needed with an accuracy of at least a factor 3 (preferably 10). Moreover, the verification should be carried out in the environment where the PRS has to operate (i.e., over the full flight envelope of modern military aircraft). Unfortunately a reference system with the required accuracy in the relevant operational environment is not available. Therefore, the above described philosophy must be traded in for a practical test method, which would still enable the establishment of the system performance[Van de Leijgraaf, 1993]. Since the PRS consists of a number of components (contributing to the total error) the method should be as follows:

- A theoretical evaluation of each PRS component results in the definition of the dominant error sources of the components and their sensitivity to the environment: height, speed, acceleration, attitude and attitude rate, etc. Either by analysis or by measurement the sensitivities are determined.
- A prediction of the total system accuracy over the envelope is calculated from the error contributions of the components.
- The calculated values are verified by carrying out operational system tests in a limited, but sufficiently relevant part of the flight envelope.

CONCLUSIONS AND RECOMMENDATIONS

From the work described, the following conclusions are drawn :

- Comparison of the performance of the ASHTECH XII and the TANS receivers in a military airborne installation, demonstrated that the ASHTECH system is better suited for flight test applications.
- The ASHTECH XII receivers used in conjunction with the RANGER and P-NAV post-processing software packages, can provide a good DGPS solution for flight testing.
- The optimisation criteria to be taken into account during test missions with DGPS are the following:
  - 4 satellites always in view with an elevation near 50°;
  - 50° maximum bank angle;
  - at least 20 sec of stabilisation before and after significant flight phases;
  - gradual heading variations;
  - distance between the aircraft and the ground receiver not greater than 200 NM.
- The DGPS performance are adequate for:
  - performing test missions over wide areas independently from environmental or meteorological conditions;
  - obtaining aircraft present position relieves at least as accurate as those of radar tracking systems;
  - using the system as an effective backup of optical trackers for some applications;
  - reducing pilot workload during test missions;
  - reducing data processing time with respect to other reference systems;
  - speeding up the test activity.
- The DGPS system performance in terms of data continuity and accuracy during dynamic manoeuvres, even if acceptable for many tasks, are not sufficient to cover the entire flight envelope of modern high performance military aircraft.
- Code range DGPS performance in terms of data accuracy are not sufficient for testing aircraft precision landing systems (i.e., ILS, MLS, DGPS).
- The integration with an inertial navigation system (INS) is the optimal solution to the DGPS shortcomings.
- The design characteristics of a high precision DGSP/INS system taking advantage of state-of-the-art hardware/software techniques was defined.

It is recommended that further studies are carried out in order to:

- Evaluate the performance of new post-processing DGPS software packages.

- Investigate on the effect of Doppler shift on DGPS receiver tracking and signal reacquisition strategy.
- Investigate on the effect of Multipath on DGPS data accuracy.
- Investigate on real-time DGPS systems, either for ground monitoring and on-line data analysis.
- Optimise the post-processing integration of DGPS data with measurements provided by the onboard sensors (e.g., INS, radar altimeter, etc.).

Furthermore, it is recommended that the current study for a high precision PRS based on DGPS/INS integration concentrates on the following areas:

- Implementation of simulation tools in order to verify the validity of the concept and to optimise the filtering algorithm;
- Evolution of the current concept based on post-processing to a real-time solution, with the possibility of providing "guidance" information to the pilot;
- Real-time ambiguity resolution and cycle slip fixing algorithms (i.e., real-time carrier phase computation).

REFERENCES

Ashkenazi V., Ffoulkes-Jones G.H., Moore T., and Walsh D., 1993. *"Real-time Navigation to Centimetre Level"*. DSNS93, 2nd International Symposium on Differential Navigation, Amsterdam.

Brown A.K., 1992. *"Test Results of the Advanced Translator Processing System"*. NAVSYS Corporation, Colorado Springs (CO), USA.

Bierman G.J., 1977. "Factorisation Methods for Discrete Sequential Estimation". Academic Press. New York.

Currie M.G., 1992. *"An Optimized Filter Architecture Incorporating a Neural Network"*. International Conference on Neural Networks. Baltimora (USA).

DMA, 1991. Defence Mapping Agency : *"Datums, Ellipsoids, Grids, and Grid Reference Systems"*. DMA TM 8358.1. Fairfax (VA).

Dumville M, Tsakiri M., 1994. *"An Adaptive Filter for Land Navigation Using Neural Computing"*. Proceedings of ION GPS-94.

Euler H.J., 1994. *"Achieving high-accuracy relative positioning in real-time"*. Proceedings of IEEE Plans, Las Vegas (Nevada), April 12-15.

Gelb A., 1992. *"Applied Optimal Estimation"*. The M.I.T. Press, Cambridge, Massachusetts, and London, England.

Gloecker F., Van Dierendonck A.J., Hatch R., 1992. "Proposed Revisions to RCTM SC-104". Proc. of ION-92, 5th Int. Tech. Meeting of the Satellite Div. of the Inst. of Navigation. Albuquerque (USA).

Hansen P., 1994. *"Real-time GPS Carrier Phase Navigation"*. The University of Nottingham, IESSG. Paper presented at the DSNS94 Conference, London.

Hatch R.R., 1991. *"Ambiguity Resolution While Moving, Experimental Results"*. Proc. of ION GPS-91, the 4 th Int. Tech. Meeting of the Satellite Div. of the US Inst. of Navigation, Albuquerque (USA).

Hein G.W., Ertel M.M., 1993. *"High-precision Aircraft Navigation using DGPS/INS integration"*. Institute of Astronomical and Physical Geodesy (IAPG). University FAF Munich, Neubiberg (Germany).

Hurrass K, 1994. *"Measuring of Flightpath Trajectories"*. AGARD-AG-160, Vol. 1 (Issue 2).

Jacob T., Schanzer G., 1989. *"Integrated Flight Guidance System Using Differential GPS for Landing Approach Guidance"*. AGARD-CP-455.

Jacob T., Jachen Meyer, Wacker U., 1993. *"Integrated Navigation and Landing System Using Global Positioning System"*. 2nd International Symposium on Differential Satellite Navigation Systems (DSNS '93). Amsterdam (Netherlands).

Kerr T.H., 1987. *"Decentralized Filtering and Redundancy Management for Multisensor Navigation"*. IEEE Transactions on Aerospace and Electronic Systems. AES-23(1).

Kleusberg A., 1990. *"Kinematic Relative Positioning Using GPS Code and Carrier Beat Phase Observations"*. Marine Geodesy.

Landau H., 1992. *"On-The-Fly Ambiguity Resolution Using Differential P-code Group and Phase Delay Measurements"*.

Lo Conte R., 1996. *"Artificial Neural Networks to Enhance or Replace the Kalman Filter in INS"*. N° 28 GD Aerosystems Course Personal Project. AWC, OD&T, Royal Air Force Cranwell.

McConnell J.B., Greenberg R.H., Pickett R.B., 1989. *"Advances in GPS Translator Technology"*. ION Satellite Division Conference Proceedings.

Napier M., 1989. *"The Integration of Satellite and Inertial Positioning Systems"*. Proceedings NAV-89. The Royal Institute of Navigation. London.

Siouris G.M., 1993. *"Aerospace Avionics Systems"*. Academic Press, San Diego, California (USA).

Van de Leijgraaf R., Breeman J., Moek G. and Van Leeuwen S.S., 1993. *"A Position Reference System for Fokker 70"*. NLR Technical Publication TP 93084L.

STRUCTURES OF INTEGRATED NAVIGATION SYSTEMS BASED ON STRAPDOWN INERTIAL NAVIGATION SYSTEMS (SINS) OF AVERAGE ACCURACY

Alexander M. Tazba / Yuri V. Levi/

Joint-Stock Company "Granit-16"

191014, St.-Petersburg, Gospitalnaya str., 3, Russia

tel. 278 - 98 - 41, fax 271 - 56 - 03,

E-mail : SOKOLOVS@COMSET.NET

Summary.

The possibility of constructing integrated inertial-satellite navigation systems for civil aviation aircrafts based on angular velocity meters and accelerometers of average accuracy is considered.

The ways of construction of such systems and positive results of construction are shown.

Ways of further increasing the navigation system's integratedness level for their upgrading are substantiated.

Introduction.

All known systems, including inertial-satellite, used as a basic means of navigation for civil aviation are based on high-accuracy SINS (error 1 – 2 miles per hour).

In the given work the Inertial-Satellite Integrated Navigation System (ISINS) based on SINS of average accuracy is considered as a basic means of navigation for civil aviation aircrafts.

The analysis of ISINS features and primary information sensors (PIS).

NSI-2000, being at stage of experimental operation on IL-76 aircrafts, is classified as mentioned type.

As PIS in this system serve:

- laser gyroscopes;
-pendulous accelerometers AK-5;
-user equipment of satellite navigation systems (UE SNS) - GG24 unit, working with mixed constellation of "GLONASS" and "NAVSTAR" systems.

Besides, the system receives information from the system of aerial signals (SAS).

Paper presented at the RTO SCI International Conference on "Integrated Navigation Systems", held at "Elektropribor", St. Petersburg, Russia, 24-26 May 1999, and published in RTO MP-43.

As an airborne computer "Kredo M" is used, based on AMD 486 DX4-100 CPU.

Nowadays, the mentioned laser gyroscopes (LG) have accuracy characteristics, shown in table 1. Accelerometers AK-5 cooperative with voltage-frequency converters have accuracy characteristics shown in table 2. Data is listed according to level "1σ".

Table 1

Parameter name	Dimensions	Range
Instability of drift rate:		
- "from start to start" mode	degrees/hour	-0,5÷+0,5
- in "starting" mode	degrees/hour	-0,25÷+0,25
Change of drift rate in the time interval of 600 s	degrees/hour	-0,03÷+0,03
Fluctuation component of drift rate	degrees/$\sqrt{\text{hour}}$	-0,01÷+0,01
Instability of scale factor	%	-0,005÷+0,005

Table 2

Parameter name	Dimensions	Range
Instability of "zero" signal:		
- "from start to start" mode	m/s^2	$-1x10^{-3}÷+1x10^{-3}$
- in "starting" mode	m/s^2	$-1x10^{-3}÷+1x10^{-3}$
Change of drift rate in the time interval of 600 s	m/s^2	$-2,5x10^{-4}÷+2,5x10^{-4}$
Instability of scale factor	%	-0,05÷+0,05

As a basic means of navigation the system has to make available:

1 pre-flight preparation (under conditions of aircraft loading, machines working and wind affecting), including:

- system control,
- independent determination of SINS angular orientation (initial values of heading in the latitude range of -70°÷+70° , roll and pitch),
- stand position determination;

2 in navigation mode output of all parameters, needed for navigation and aircraft handling;

3 solving navigator's problems automation by cooperative processing of current navigation information and loaded data bases.

Acceptable errors of main ISINS output parameters generation are shown in table 3.

Table 3

Parameter name	Dimensions	Range
Geodesic coordinates of place		
Latitude, longitude when UE SNS information break		
up to 3 min	m	-50÷+50
up to 30 min	m	-2500÷+2500
Altitude	m	-75÷+75
Linear velocity projection		
when UE SNS information break up to 3 min		
North, East Vertical	m/s m/s	-0,5÷+0,5 -0,75÷+0,75
Angular coordinates		
Heading (after aircraft take-off)	degrees	-0,5÷+0,5
Pitch, roll	degrees	-0,1÷+0,1
Angular velocity projections	degrees /s	-0,0005÷+0,0005
Projections of load factor	%	-0,075÷+0,075

In the mode of pre-flight preparation heading can be determined with root-mean-square (r.m.s.) error

$$\sigma_K > \frac{\sigma_\Omega}{\Omega_3 \cos\varphi_0} \quad , \qquad (1)$$

where σ_Ω - r.m.s. value of drift rate of laser gyroscope (LG) (horizontal LG_X and LG_Z);

Ω_3 - Earth rotation velocity ;

φ_0 - latitude of aircraft stand position

Calibration of drift rate of horizontal LG (LG_X и LG_Z) can be supplied only with r.m.s. errors

$$\left.\begin{aligned} \sigma_{\Delta\Omega X} &> \sigma_\Omega \sin K_0 \\ \sigma_{\Delta\Omega Z} &> \sigma_\Omega \cos K_0 \end{aligned}\right\} \quad , \qquad (2)$$

where K_0 - heading of aircraft on stand position. Estimations (1) and (2) considering instrumental errors of LG, shown in table 1, show in the mode of pre-flight preparation:

- independent determination of heading, cannot be realized with needed accuracy;
- calibration of drift rate of horizontal LG cannot supply the needed accuracy of navigation parameters determination when UE SNS information break.

Refinement of LG heading and drift velocities can be realized while aircraft's accelerated motion on the runway because of appearance of these parameters observability.

However, their identification while aircraft's motion on the runway determines stringent enough requirements to the UE SNS information accuracy. So, when the aircraft's acceleration $W \approx 1$ m/s^2 available changes of velocity error, received from UE SNS must be limited with r.m.s. error $\sigma_{AS} < 0{,}002$ m/s^2 . As shown in [1], artificially arranged SA (Selective availability) interference , which is input into "NAVSTAR" satellites clock, has the value of $\sigma_{AS} = 0{,}0037$ m/s^2 . With a little number of "GLONASS" satellites in operating constellation the influence of SA interference on errors estimation while their determination has to be reduced. This task becomes complicated because of the necessity to estimate changing in "starting" mode components of drift rate.

SINS parameters identification

SINS parameters identification (including PIS errors) is traditionally realized by Kalman filter by means of cooperative processing of SINS, UE SNS and SAS information. For reducing UE SNS and SAS errors influence into the estimated state vector X besides sub-vector X_S of SINS errors the sub-vector Xn of UE SNS and SAS errors is input.

$$X^T = \| X_S^T \mid X_n^T \| ; \qquad (3)$$

Vector X_S (dimensionality 15) has the following components:

$$X_S^T = \| X\Omega^T \mid X_\alpha^T \mid X_A^T \mid X_V^T \mid X_G^T \| ; \quad (4)$$

$X\Omega$ - vector of non-stationary components of drift rate LG_X, LG_Y, LG_Z.

X_α - vector of inertial coordinates (IC) position determination errors relative to bound coordinates (BC).

X_A - vector of non-stationary components of "zero" signals drifts of A_X, A_Y, A_Z accelerometers.

X_V - vector of velocity projections errors in geographic coordinates.

X_G - vector of geodesic coordinates errors (arc length in latitude direction, longitude direction and altitude).

Error models X_α , X_V и X_G are represented in the form, suggested in [2]. Incoming in this model LG drift rate and accelerometer "zero" signal shifts according to their structure, represented in tables 1 and 2 are determined in the following form:

$$\left.\begin{aligned} X_i(K) &= X_i(K-1) + W_i(K) \\ Y_i(K) &= X_i(K) + W_{ji}(K) \end{aligned}\right\} \qquad (5)$$

where i - Ω, A;
ji - αΩ, VΩ ;
W_i, W_{ij} - discrete white noises.

Vector Xn (dimensionality 7) contains next components:

$$X_n^T = \left\| X_{NS} \middle| \Delta V_{NS} \middle| X_{ES} \middle| \Delta V_{ES} \middle| X_{HS} \middle| \Delta V_{HS} \middle| \Delta H_B \right\| \quad (6)$$

X_{NS}, X_{ES}, X_{HS} - UE SNS coordinate determination errors;

ΔV_{NS}, ΔV_{ES}, ΔV_{HS} - UE SNS velocity projections determination errors;

ΔH_B - resulting deviation of barometric altitude, measured by SAS from geodesic.

From UE SNS information errors, into vector X_n only components, determined by SA interference, input into "NAVSTAR" satellites clock, are included.

As a simplified model of this interference the following model is used in the filter:

$$\begin{Vmatrix} X_{is}(K) \\ \Delta V_{is}(K) \end{Vmatrix} = \left\| \begin{array}{c|c} FS1 + \mu \cdot FS2 & FS2 \\ \hline -b \cdot FS2 & FS1 - \mu \cdot FS2 \end{array} \right\| \cdot \begin{Vmatrix} X_{is}(K-1) \\ \Delta V_{is}(K-1) \end{Vmatrix} + \\ + \begin{Vmatrix} g_1 \\ g_2 \end{Vmatrix} \cdot W_i(K) \cdot \sigma_i \quad ; \quad (7)$$

where i - N, E, H;

$FS1 = e^{-\mu Tf} \cos(\omega_0 Tf)$

$FS2 = \frac{1}{\omega_0} e^{-\mu Tf} \sin(\omega_0 Tf)$

$b = \omega_0^2 + \mu^2$;

$W_i(K)$ - discrete white noise with the variance $Q_i = 1$;

σ_i - r.m.s. value of coordinate error X_{is} ;

ω_0 - dominant frequency of random process;

μ - irregularity factor;

Tf - filter step;

SA interference model in the form (7) represents next factors:

- introduction of interference into satellite clock causes its exhibition in long-range C/A code and, correspondingly, in Doppler canal in the form of carrier frequencies L1 и L2 shift, what is expressed in the model by differential relationship

$$\dot{X}_{is} = \Delta V_{is}(b);$$

- oscillating nature of the process (with the dominant frequency ω_0);

– r.m.s. value of σ_i is determined by the relationship $\sigma_i = \sigma_0 \cdot C_i(K)$:

where

σ_0 - r.m.s. value of pseudo-distance error for "NAVSTAR" satellites;

C_i - factor of influence of pseudo-distance to "NAVSTAR" satellites errors on general coordinates determination error of aircraft system by mixed constellation of "GLONASS" and "NAVSTAR" systems

Factor C_i is determined by secondary processing algorithm in UE SNS, the current values of geometric factors of operating constellations of "GLONASS" and "NAVSTAR" (Γ_{iG} and F_{iN} correspondingly), and the number of satellites in "GLONASS" (N_G) and "NAVSTAR" (N_N) working constellations.

According to [1] the value

$$\sigma_0 = 30 \text{ m}.$$

ω_0 value was determined by correlation analysis of UE SNS measurements

$$\omega_0 = 0{,}0009 \text{ 1/s}$$

and fully coincided with the value, given in [3].

μ value was also determined by correlation analysis of UE SNS measurements

$$\mu = 0{,}005 \text{ 1/s};$$

For ΔH_B error the model in form (5) is used, which determines the difference between barometric and geodesic altitude and random errors of barometric altitude determination.

UE SNS information is basic for SINS errors determination. Numeric values of SA interference model give the idea of limited possibilities of excluding its influence on accuracy of estimation non-stationary SINS errors.

Figure 1 represents the structure of SINS, UE SNS and SAS information transformation in NSI-2000 system. In the system the hardware-algorithmic synchronizing of measurements is realized. Projections of antenna position velocity (V_{NA}, V_{EA}, V_{HA}) and the coordinates φ, λ, H, calculated by SINS, are interpolated or extrapolated for UE SNS measurements moment and fixed by its pulse, transmitted to airborne computer interruption. Differences between pointed SINS parameters and corresponding UE SNS and SAS parameters ($\delta\varphi$, $\delta\lambda$, δH, δV_N, δV_E, δV_H, δH_B) proceed to the filter input. Filter outputs estimations $\hat{X}_S$ и $\hat{X}_n$ of SINS, UE SNS and SAS. These estimations are output with delay τ, which is determined by UE SNS information delay and filter count time. This time in NSI-2000 reaches 0,5 s.

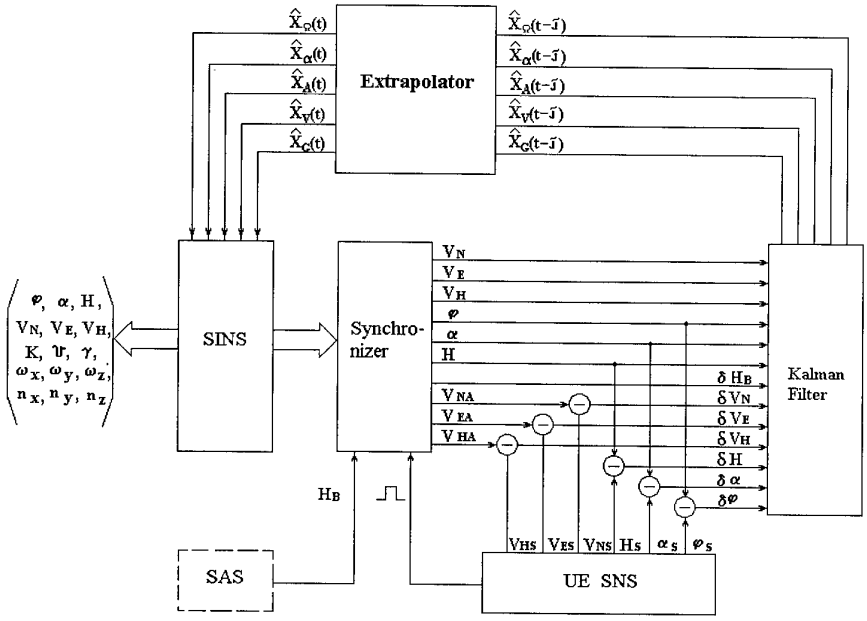


Figure 1

SINS errors correction.

As shown on figure 1, extrapolated estimations of SINS errors $(\bar{x}_s)$ are input into its interior structures. In the general case $\bar{x}_s$ estimations can be used as for error compensation, or for their damping. However for average accuracy SINS the correction of IC position by $\bar{x}_\alpha$ and $\bar{x}_\Omega$ estimations seems to be a matter of principle. Because of high initial values of heading and high instability of drift rate IC determination, errors could reach (after long-duration SINS operation) tens degrees. If this errors are not compensated immediately in SINS

- requirements to SINS error models accuracy are stringented (nonlinear models are needed)
- additional algorithms for compensating all system output parameters, which errors are not involved directly into the estimated vector X_S (heading, roll, pitch etc.) are required.
- vector X_S estimation error increases (filter divergence appears). This effect is displayed evidently when the system is operating after long UE SNS information breaks.

If we call "integratedness" the possibility of an object to replace its characteristics with the help of all other objects, connected with it, then we can estimate the level of the system integratedness on functional level by existence and fullness of connections between SINS, UE SNS and Kalman Filter.

From the NSI-2000 system structure (figure 1) follows:

- full feedback with Kalman filter (full compensation of SINS errors);
- absence of direct connections between UE SNS and SINS (SINS information is not used for satellite signals searching, improving resistance to interferences and compensating UE SNS dynamic errors);
- absence of feedbacks between Kalman filter and UE SNS (UE SNS errors are not being compensated).

Results of NSI-2000 errors examination.

To illustrate specific character of ISINS operation, using SINS of average accuracy it is enough to examine the nature of errors of determining a number of system output parameters and PIS errors.

Features, represented further, were obtained by mathematical simulation of system operating while imitating aircraft movement and PIS characteristics simulation. While NSI-2000 system development and the system ground and flight testing all imitators, used for simulation, were also developed, and that made possible to receive good coordination between real system features and those received while mathematical simulation.

Figure 2 presents heading determination error while aircraft's moving on the runway. Absolute velocity behavior is represented on figure 3.

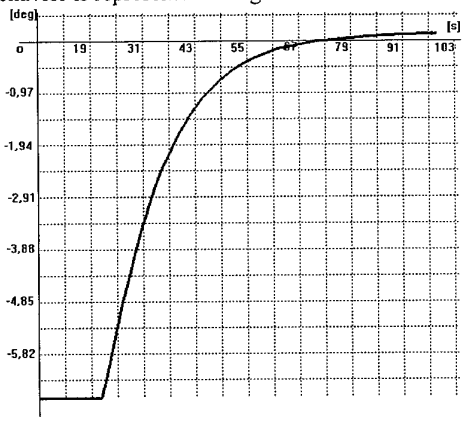


Figure 2

Figures 4 and 5 represent the error of estimation of horisontal LG drift rates (LG_X and LG_Y correspondingly).

Figure 6 represents the error of non-stationary LG drift rate, which behavior is represented on figure 7.

To show the efficiency of SA interference estimation for the constellation, consisting of 7 "NAVSTAR" and 1 "GLONASS" satellites figure 8 represents velocity error, and the value of SA interference is represented on figure 9.

For estimating the influence SA interference characteristics figure 10 represents velocity error for the same constellation, but with the interference dominant frequency ω_0=0,0045 1/s. The value of SA interference itself is represented on figure 11.

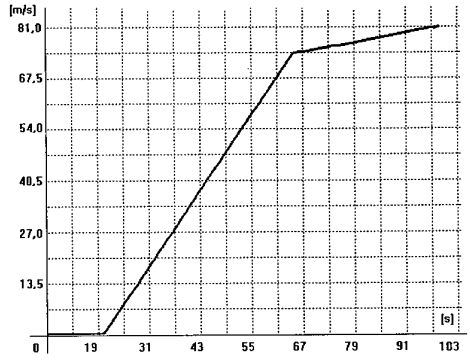


Figure 3

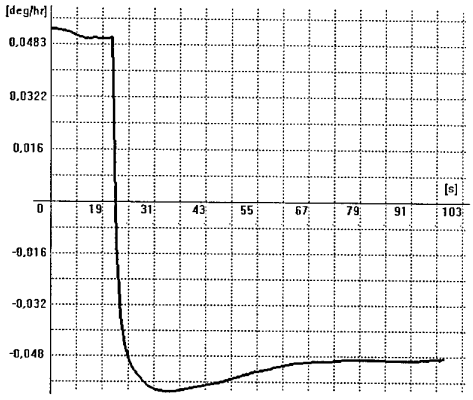


Figure 4

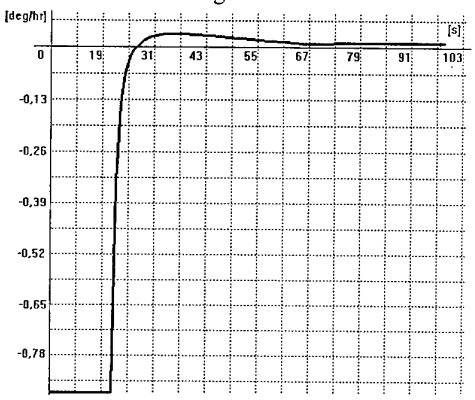


Figure 5

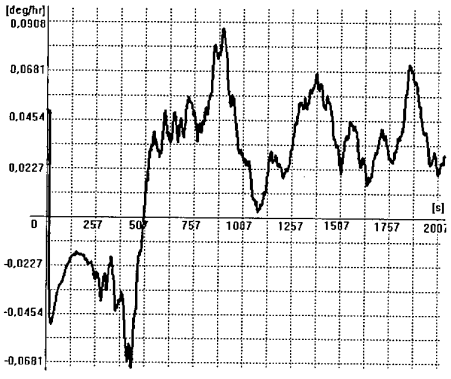


Figure 6

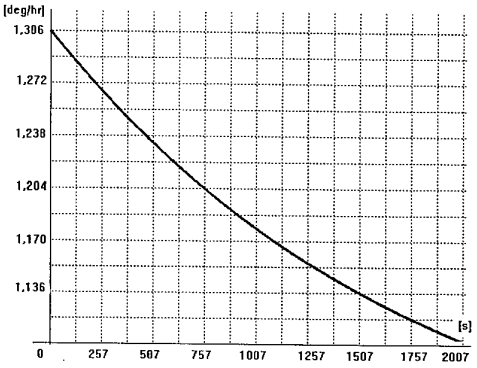


Figure 7

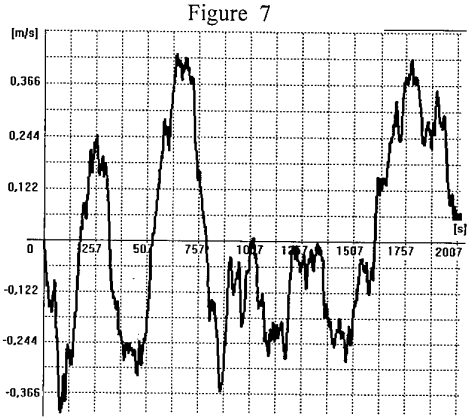


Figure 8

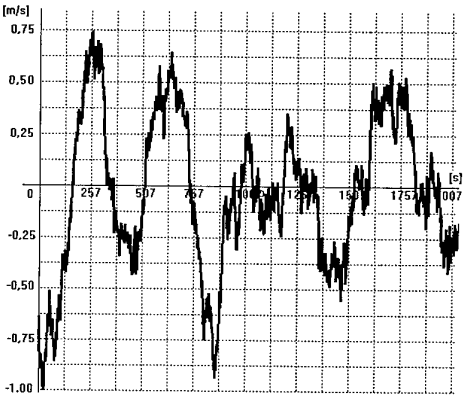


Figure 9

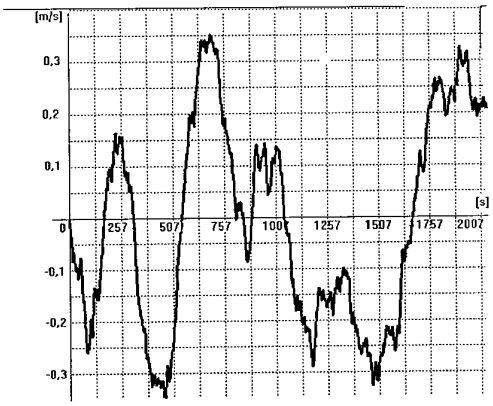


Figure 10

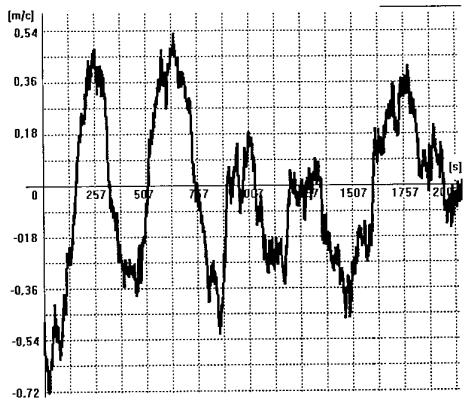


Figure 11

Ways to improve ISINS.

Nowadays the following ways of improving ISINS exist:

- introduction of fiber-optic gyroscopes (FOG);
- increasing the system's integratedness level on hardware and functional levels;
- improvement of UE SNS error models (for their identification and compensation).

Efficiency of these ways of development is based in [4].

The pointed tendencies of ISINS development are realized in the system ISNS-2000 (now on the stage of development). The system uses:

- FOG DUS-500;
- receiving and measuring unit KC-116A (operates with the mixed constellations of "NAVSTAR" and "GLONASS" satellites).

Nowadays the system testing is being performed, with the experimental models slightly distinguished on the functional level from NSI-2000 system.

However, there is a plan of consistent improvement of experimental models for increasing the system's integratedness level.

So, using of the common reference generator is planned, which, combined with the frequency synthesizer , makes possible:

- FOG and accelerometers operation;
- receiving and measuring unit KC-116A operation;
- interruptions, needed for airborne computer.

By this the hardware synchronization of KC-116A measurements is realized.

Substantial obstacle for integratedness level increasing (functional and hardware) is data swapping between KC-116A and airborne computer by the RS-232 channel. If KC-116A unit will have its output straight on the airborne computer wire, the following will be possible:

- using the data adjusted by SINS for compensating dynamic errors in automatic frequency control circuit and tracking the delay of their bandwidth reducing.
- excluding from the construction special communication units.

For increasing the system accuracy ways to exclude the influence of UE SNS significant errors, arising while the operating constellations change, are being developed.

To afford this it is necessary to:

- reduce the time of operating with flicking constellations by optimizing UE SNS settings (allowable rising angle, signal/noise relationship etc.)
- form the message at the moments of constellation change in UE SNS;
- develop the model of UE SNS information errors at the moments of constellation change, which is dependent on the change of geometric factors of sub-constellations of "NAVSTAR" and "GLONASS" systems, and the model of SA interference (fluctuative and ephemerical distortion).

Conclusion.

1. ISINS, used as a basic means of navigation for civil aviation, can be realized on the base of SINS of average accuracy.

2. Using PIS with significant errors causes the necessity to complicate SINS structural corrections.

References.

1 Gregory T.Kremer, RudolPH M. Kalafus, Peter V.W. Loomis, and James C. Reynolds. The Effect of Selective Availability on Differential GPS Correetijns, NAVIGATION, Journal jf The Institude jf Navigation, Vol/37 No.1, Spring 1990.

2. Karakashev V. A. Generalized equations of inertial navigation system errors. //Higher educational establishment news - Devicebuilding - 1973 . -№3.-C. 87-93.

3. DO –229, APPENDIX A, p.2.

4. Shebshaevich V. S. satellite radionavigation systems // Moscow, Radio and connections, 1993

High Precision Integrated Navigation System for Vehicle Guidance

G. Schänzer

Institute of Flight Guidance and Control
Technical University of Braunschweig
Hans-Sommer-Str. 66
D – 38106 Braunschweig, Germany
Phone/Fax: ++49 531 391-3716 / -5954
EMail: G.Schaenzer@tu-bs.de

Introduction

The task of bringing a vehicle safely to its destination is a problem of which a solution has been sought for generations. The compass has been in use for over 1000 years as a navigation instrument for sea voyages. As a result of high safety requirements in aviation, radio navigation systems such as Loran, Decca, VOR, DME were developed and have achieved special significance. The achieved precision ranges from approximately 200m up to 1000m (ref. Fig. 1).

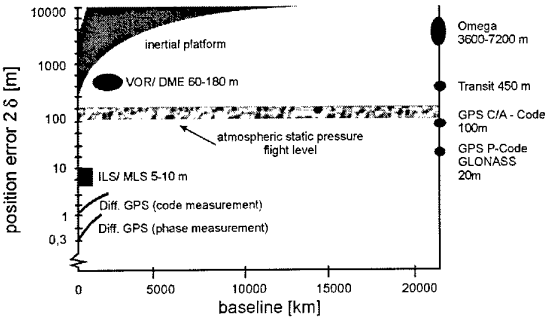


Fig. 1: Comparison of navigation systems [6]

Inertial Navigation Systems (INS), as a rule, do not attain the degree of precision of radio navigation systems, but can function autonomously, for example during long distance transoceanic flight. INS typically have time dependent errors of about 2 km per hour. It is obvious that a high degree of precision and safety is beneficial to aviation and sea navigation, but is of great interest for military applications as well. High demands for precision and reliability are presently posed by all weather approaches in civil aviation. Accuracy of better than 60 cm on the runway threshold must be demonstrated for so called "blind landings" (weather conditions corresponding to CAT III). They are achieved, although at great expense, through the use of Instrument Landing Systems. A modern successor to the ILS, according to a resolution by the International Civil Aviation Organisation (ICAO), is to be the Microwave Landing System (MLS) whose introduction has been delayed for technical and commercial reasons.

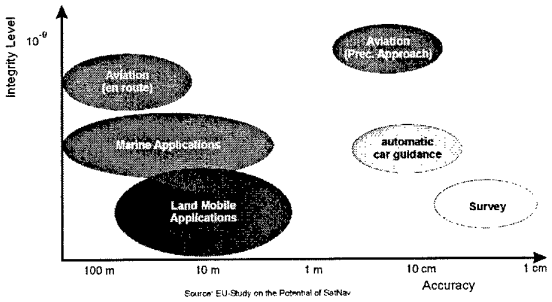


Fig. 2: User requirements [7]

The precision approach systems, ILS and MLS, have a range limited to the airport area. A costly individual system is needed for each runway, as well as for each direction. The ILS and MLS are not suitable for other uses, i.e. for road traffic. The highest demands for precision are related to flight test calibration, flight inspection and geodetic survey. Presently requirements for accuracy better than 10 cm are discussed and validated for automatic car guidance (ref. Fig. 2).

Satellite navigation

Satellite navigation systems were developed in the USA as well as in the Soviet Union for military purposes and make use of an essentially very basic principle. Satellites with an orbit time of about 12 hours (corresponding to a satellite distance from the earth's surface of about 20,000 km) transmit an electromagnetic signal that travels at the speed of light. The transmitted electromagnetic signal is encoded basically with a satellite identifier, the satellite position, the transmission time of the signal and some long term correction parameters. These transmitted microwave signals are received by a high gain receiver onboard the vehicle. The range between satellite and vehicle can be calculated from the signal's travelling time multiplied with the speed of light. The travelling time is the time difference measured by two precise clocks, one in the satellite, another in the vehicle.
The accuracy of the time difference measurement is strongly depending on the quality of the time reference of the receiver. A similar atomic

Paper presented at the RTO SCI International Conference on "Integrated Navigation Systems", held at "Elektropribor", St. Petersburg, Russia, 24-26 May 1999, and published in RTO MP-43.

clock is necessary for the satellite transmitter as well as for the receiver. With a minimum of three visible satellites the range measurement can give the basis for the calculation of the vehicle's position. An additional time difference measurement to a fourth satellite can improve the time reference estimation and the precise atomic clock in the receiver becomes unnecessary.

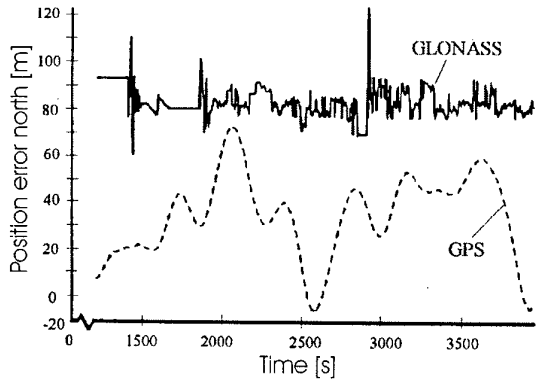


Fig. 3: GPS and GLONASS position deviations

The position calculation becomes more difficult the higher the speed and manoeuvrability of the vehicle is. For this applications, the knowledge of the exact time for each individual range measurement becomes very important. The described procedure requires the knowledge of the position of each of the satellites which includes the corrections of satellite position and atmospheric conditions in real time onboard the vehicle. This improves the accuracy of the range measurement. An additional measurement of the carrier phase gives a range solution within a few millimetres accuracy.

Each satellite navigation system requires at least 24 satellites in the orbit, to assure that a minimum of 4 satellites is visible. This operational status has been reached for GPS in 1993 and was targeted on 1995 for GLONASS. In the beginning of 1999 the system exists of 14 operating satellites. The general usage of both systems will be free of charge 10 years for GPS and 15 years for GLONASS. The specified position accuracy for both systems lies in the range of 16 meters.

Design accuracy of 20 meters for global operation was a challenging target but could be realised without significant problems. Later GPS was offered for civil use, although with a reduced accuracy of several hundred meters. The technique applied for this accuracy degradation is called selective availability (SA) and was introduced in March 1990, with the launch and operation of the Block II satellites. Up to that point GPS could be used by the civil users with almost the same precision as the military P-code users [1]. The aim of the SA technique is to allow only authorised users to gain the full accuracy of GPS. This policy goes back to the year 1982, when interagency study panel, composed of representatives from the Office of the Joint Chiefs of Staff, intelligence agencies and the Defence Mapping Agency recommended to allow a Standard Precision Service (SPS) of only 100 meters of accuracy for the civil users. When SA was put into operation a fierce discussion started within the GPS-user community. The main argument of the SA opponents was, that in the case of a war, it would not matter if a Scud missile is precise to within 10 meters or 100 meters, because that precision is of the same level as the accuracy of most missiles themselves. This is probably true for the majority of applications of satellite navigation in the military environment. It is also very noteworthy that in the Gulf War the SA was turned off. The reason behind this was an insufficient supply of PPS (precision position service) capable GPS receivers in the US forces. The US headquarters therefore had its troops equipped with standard SPS receivers (this step may also have been driven by the fact that PPS receivers need the C/A-Code to lock in quickly). There are no reports about any military action by the enemy in this war that could have been prevented if SA had been switched on. (It is also an interesting side effect that as a civil user one can foresee in the home office at what time a military crisis for the US could arise by simply looking at the level of SA on GPS [2]).

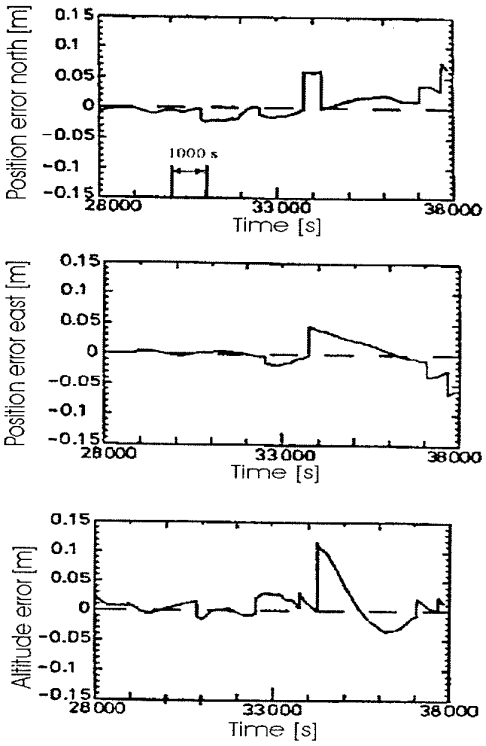


Fig. 4: Differential GPS measurements (v=0 m/s)

Differential Correction Technique

The artificial and natural reduction of accuracy can be overcome easily by differential methods. A fixed based reference receiver can find its own position within minutes with an accuracy of a few millimetres by geodetic methods that are state of the art. All measured deviations in position or range can be defined as errors. If these errors are transmitted to the vehicles, the relevant on-board errors can be eliminated and the navigation accuracy can be significantly improved.

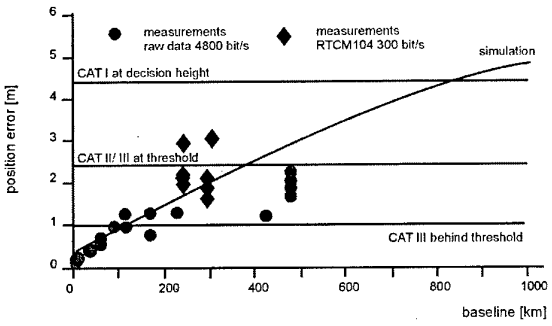


Fig. 5: DGPS horizontal accuracy vs. Baseline

Fig. 4 demonstrates a typical error of less than 10 cm, if the distance between the two antennas is small and both antennas will not be moved. Higher speed of the vehicle and increasing distance between the antennas will increase the position error. Fig. 5 shows the navigation errors as a function of the baseline of differential GPS. CAT III accuracy can be realised even at a distance of approximately 100 km. CAT I is possible with a baseline of 800 km. A technical problem that has to be solved is the broadcasting data link. For quasi visual contact baseline up to 1500 km VLF transmitters (140kHz, 50kW) produce excellent results.

The differential technique is very simple, but extremely powerful. The SA policy of the Pentagon has lead to the undesired emergence of a technology, that can be applied by everyone and offers a level of precision even higher than the standard P-code authorised applications.

The so called Wide Area Differential GPS, that uses communication satellites to broadcast the differential corrections, is still under development. Also an integrity channel is planned. This technique is not only complex and expensive, also is the minimum time to alarm of the integrity channel (6 seconds) too long for most applications (i.e. CAT III requires 2 seconds).

Phase measurement

The original concept of GPS and GLONASS is based on code measurement. The satellite transmits a coded signal, that can be generated similarly in the receiver. The receiver's control loop shifts this coded signal, that both signals, the transmitted and the received will be more or less identical. The time shift is a measure for the range. Each electromagnetic signal is characterised by its wavelength. Counting the passing waves gives a much higher resolution than the code measurement itself. Additionally the phase shift can be measured with still increasing accuracy.

To obtain higher accuracy, the carrier phase relationship between received and internally generated signal is measured. The carrier frequency is higher than the code frequency, leading to much higher resolution in the sub-millimetre range. Unfortunately, since all waveforms look similar, this is only possible to fractions of one full cycle. The unknown number of cycles between user and satellite is called "(full cycle) integer ambiguity" and must be determined before the information can be used for positioning.

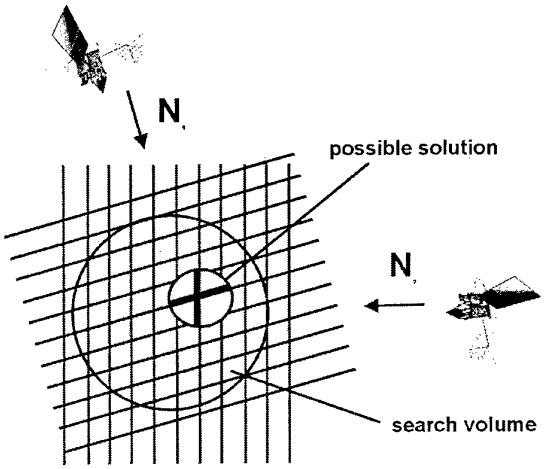


Fig. 6: Principle of ambiguity search

One method to enhance the resolution of the code has been described in [1] and is known as carrier smoothing. Here the code measurement is used to avoid the ambiguity. This method enhances resolution, but not absolute accuracy due to reliance on the code. More accurate results can be obtained by searching through a range of possible values, finally settling on the closest value by some statistical criterion. Many different techniques have been proposed for this search, within the scope of this paper it may suffice to illustrate the general principle.

Usually the code measurement is used to generate an approximate position and together with its uncertainty defines a search volume. In the two-dimensional case this results in a disk on which the search has to be performed. Assuming the true solution lies within this volume one can use the fact that the ambiguities are whole numbers to draw circles with distances of one wave-

length against each other around each satellite. Due to the distance of the satellite relative to the search volume these circles can be approximated as parallel lines as shown in Fig. 7.
Two principles become evident from this figure – the greater the wavelength, the smaller the number of possible solutions and the more satellites are usable (beyond the minimum two needed to generate the two-dimensional solution) the more solution can be eliminated trough redundancy. In the tree-dimensional case a typical initial search volume of about 3 m radius will encompass millions of possible solutions. One can now use the change of geometry caused by the movement of the satellites with time to eliminate further possibilities, until only the correct solution remains.
Obviously the solution quality depends on many factors and the method illustrated here is not optimal with respect to calculation time, use of the measurement information or application to a moving user. Therefore many modifications of the basic algorithm have been published in literature. These modifications change the size or shape of the search volume, establish priorities which solutions should be investigated first or, instead of searching for possible solutions in a fixed volume, try to reduce the search volume over time until only one plausible solution remains. A big reduction in the time needed to come to an unambiguous solution can be obtained by using data from a second frequency (GPS L_2) to generate a combined measurement with significantly increased wavelength. Instead of using the GPS L_1 wavelength of about 19cm, this so called wide lane observable has a wavelength of about 86 cm and leads to a significantly smaller number of possible solutions within the same search volume.
Some results of phase ambiguity measurement are shown in Tab. 1 and Fig. 7.

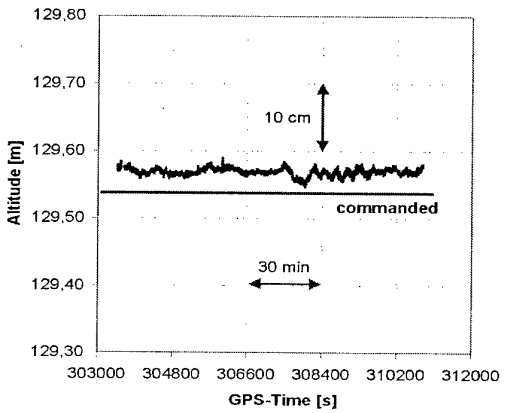


Fig. 7: Stationary Phase Measurement

	Wavelength	*Potential accuracy*	*Oppera-tional Accu-racy / state of the art*
C/A Code	293.1 m/Chip	10 cm	16 m
L_1 Phase	19.05 cm	1 mm	3 cm

Tab. 1: Code and phase measurements

Continuity of Service

Due to the original design of GPS an GLONASS for military applications the coverage is excellent primarily in those areas of the world, where military crises are expected. During the cold war Central Europe, especially Germany, was of main interest and as a consequence has a very good coverage (that is more than 4 satellites visible at a time).

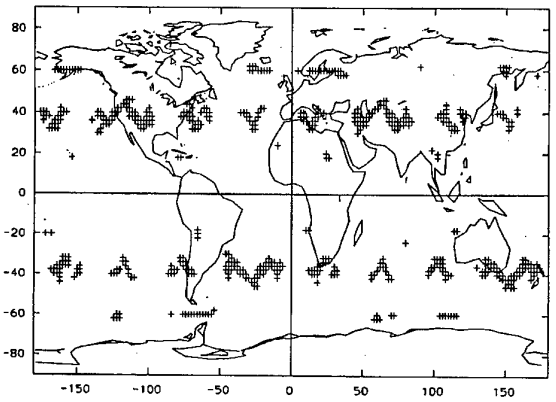


Fig. 8: Coverage gaps of GPS (24h, min < 5 SV) [4]

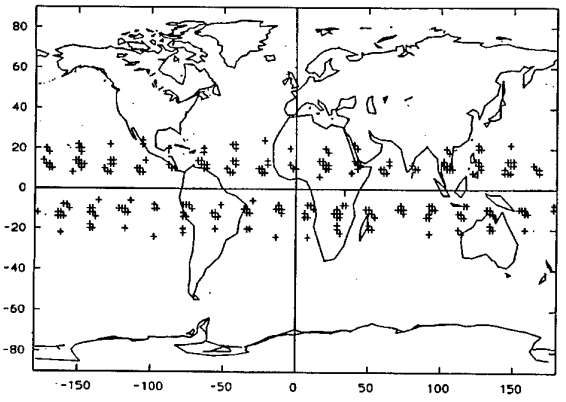


Fig. 9: Coverage gaps of GLONASS (24h, >5 SV)

The orbits of GPS and GLONASS satellites are depending from their launching site. GPS will be launched in Cape Canaveral (28.4N 80.6W) and GLONASS in Baikonur (47N 65E). GPS satellites illuminate primarily the equator (ref. Fig. 10) and have a significant gap in the Mediterranean area and in the US. GLONASS is more polar orientated with gaps close to the equator.

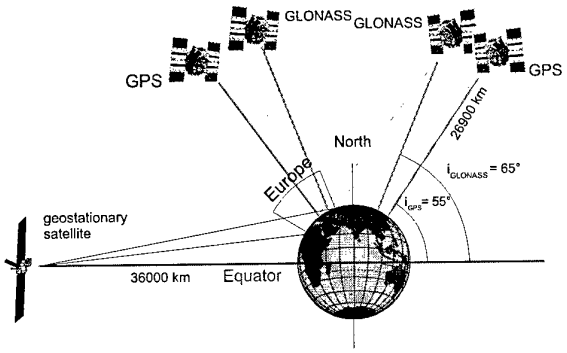


Fig. 10: Coverage lack of GPS, GLONASS and geostationary satellites in northern regions

A free of charge service is guaranteed for 95% of the time by the U.S. ministry of defence (DoD). This means that during 72 minutes a day no service is guaranteed. Operating on a navigation system under these conditions is unacceptable for commercial aviation. To improve the continuity of service additional satellites are required. Investigations show that a minimum of 31 satellites on sufficient orbits is necessary. A combination of GPS and GLONASS produces excellent results [4].
INMARSAT provides additional geostationary satellites with compatible GPS transmitters (ref. Fig. 10). This geostationary satellite illuminates primarily the equator where GPS already gives excellent results. The satellite navigation coverage will be improved in the USA but the improvement in Central Europe is marginal.

System Integrity

As mentioned before, satellite navigation systems complemented with differential techniques are very precise. Much more precise than required for precision CAT III approaches. A significant operational problem is the extremely low level of integrity and reliability of both GPS and GLONASS. As satellite navigation systems are typical radio navigation systems operating in the microwave frequency band, all failures that are known for standard systems also occur with GPS and GLONASS (Fig. 11, Fig. 12):

- Artificial jamming apparatus
- masking, shadowing
- multipath
- interference, jamming due to:
  - military high density communication lines
  - radio amateurs
  - approach radar
  - industrial and medical microwave apparatus

Due to the extremely long distance between the transmitter (satellite) and the receiver (omnidirectional antenna), the received signal is extremely low (-140 dB) and can be easily disturbed. Even permitted side bands of transmitters on other frequency ranges can produce significant problems. Additionally, the GPS and GLONASS frequencies are not protected in most countries.

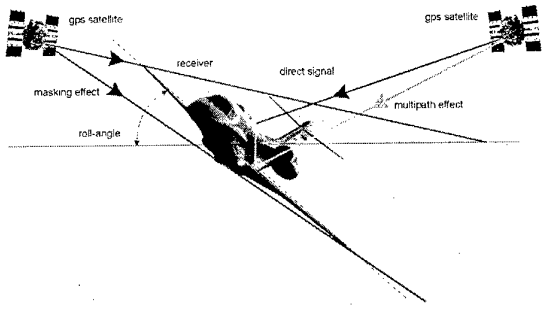


Fig. 11: Multipath and masking effects for an airplane

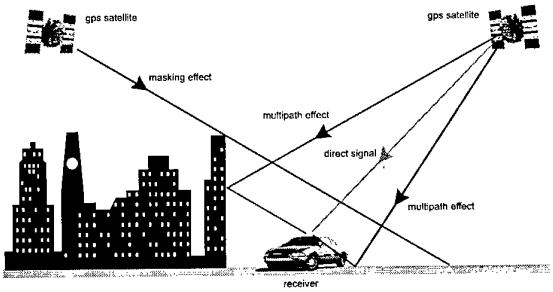


Fig. 12: Multipath and masking effects for surface vehicles

Summarising, GNSS is in principle very unreliable. After eight years of experience in flight and vehicle experiments a rate of approximately 10^{-2} failures per hour was obtained. This is far away from the ILS standard target of 10^{-7} failures per hour. Most failures occur in the local vicinity of aircraft and must be deleted there. The global integrity channels that are under discussion will not improve the situation, as such techniques detect only the very rare satellite errors. Additionally, most of the failures have a systematic nature (i.e. a standard bank of 30° can mask all visible satellites at the same time). Therefore, conventional redundancy techniques will not work. Three parallel GPS receivers with three independent antennas, for example, will produce the same incorrect navigation information. There are many techniques under discussion to detect failures of GPS signals. Most of them use information from additional satellites (more than four are necessary then) for integrity checks. These techniques are known as receiver autonomous integrity monitoring (RAIM). Looking at the situation in Hanover, Tab. 2 demonstrates that the availability of RAIM either with GPS or

GLONASS is not acceptable for precision approaches.

system	area	enroute (%)	Terminal (%)	NPA (%)
GPS	World	2,8	2,8	11,1
	Europe	2,8	2,8	9,7
	North Am.	2,78	2,78	9,72
GLONASS	World	6,9	8,3	33,3
	Europe	0	0	2,8
	North Am.	0	0	2,78
GPS & GLONASS	World	0	0	0
	Europe	0	0	0
	North Am.	0	0	0

Tab. 2: RAIM non-availability per day [2]

A typical terminal flight (ref. Fig. 13) gives an indication of the effect of masking the GPS signal by the wings and the fuselage of the aircraft.

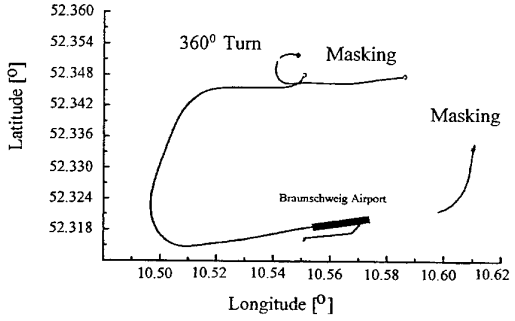


Fig. 13: Terminal Flight with GPS

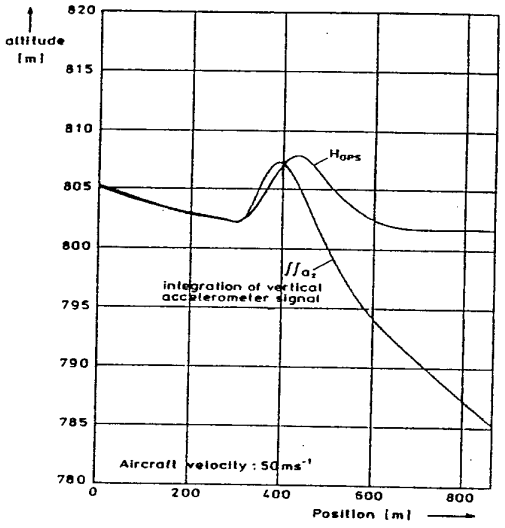


Fig. 14: Aircraft passing the updraft of a cooling tower

Additional problems occur in highly manoeuvrable vehicles. The internal control loops of the satellite navigation receiver are a compromise between accuracy and agility. Especially high precision satellite navigation receivers provide a sluggish response. Fig. 14 demonstrates a significant altitude error and time delay related to an inertial altitude, if aircraft is passing the updraft of a cooling tower. Fig. 14 gives answers to possible solution. The satellite navigation shows lack in the dynamical behaviour and the inertial altitude presents a significant drift.

Integrated Navigation Systems

A well proven technique is to combine the precise but unreliable GNSS signals with complementary inertial sensor, which are reliable but not very precise. Fig. 15 shows an integrated navigation system that produces useful results even in the unfriendly radio navigation environment.

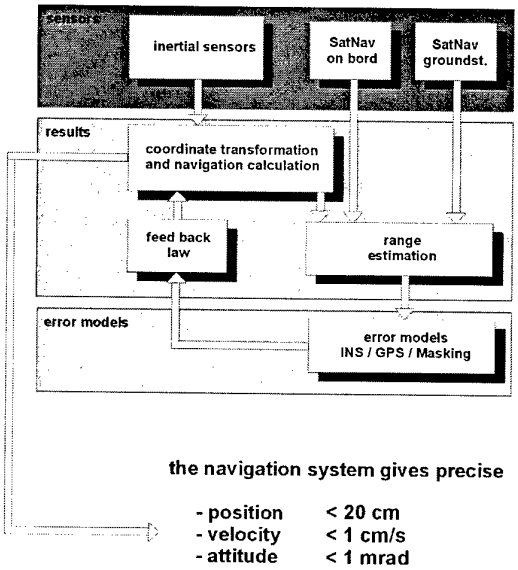


Fig. 15: Integrated Navigation System

Complementary sensors for the excellent accuracy and the poor integrity of satellite navigation are inertial sensors, e.g. gyros, accelerometers. In Tab. 3 some important characteristics are listed.

	Satellite Receiver	*Inertial Sensor*
Self contained	No	Yes
Initialisation required	Yes	Yes
Error behaviour	Rel. high noise, stationary	Low noise, drifts
Dynamics	Time delay, amplitude attenuation	No noticeable time delay or attenuation
Output rate	0.5 – 1 Hz	Up to 100 Hz
Parallel redundancy	Not possible	Succesfull

Tab. 3: Comparison of satellite and inertial navigation

A typical integrated system is demonstrated in Fig. 15. The range measurement to each individual satellite is realised in three independent ways:

- satellite reference station
- satellite receiver
- inertial range calculation

With such a system the range can be estimated more precisely and reliably. Each individual range measurement can be checked out and the failures can be identified and eliminated. All relevant dynamical errors are modelled. The modelling of the internal sensor dynamics has a long successful tradition as well as the simulation of the satellites orbit and the flight path. The propagation errors of the electromagnetic waves as well as the receiver dynamics especially the Costas loop and the code tracking loop can be now described with sufficient error models [2]. But there is still a high potential for improvement. Especially the effects of time and manoeuvre variation of the model parameters are a source for interesting engineering investigations.

If all relevant errors are known, they can be compensated successfully. The basic idea behind the integration of satellite and internal navigation is to estimate the inertial sensor errors online using satellite navigation, whenever available. At the same time the inertial measurement can be used to bridge cycle slips and times of loss of lock, but most important is to bridge the time between two position fixes (ref. Fig. 16).

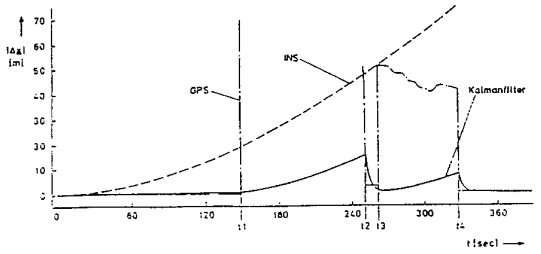


Fig. 16: Position error of an integrated system if GPS is interrupted

The advantages of such an integrated navigation system are:

- high agility and high precision
- real time output
- high frequency data stream (essential for automatic vehicle guidance)
- bridging black outs of satellite navigation
- high accuracy in position, speed, angular rate and attitude angles
- parallel redundancy possible

Applications and results

When demonstrating the extraordinary qualities of a precision navigation system, the lack of suitable reference systems became evident. High precision laser trackers also only achieve accuracy in the sub-meter range and beyond that, can only cover a comparatively small distance. In most cases, one have believe the output of the integrated system and to check the compatibility of the sensor signal in the multisensor system.

Aircraft guidance with integrated navigation systems is comparable simple, when conventional flight path sensors e.g. ILS, MLS, Radar are replaced by this system. Fig. 17 shows the worldwide first automatic landing based on GPS at Braunschweig airport in July 1989.

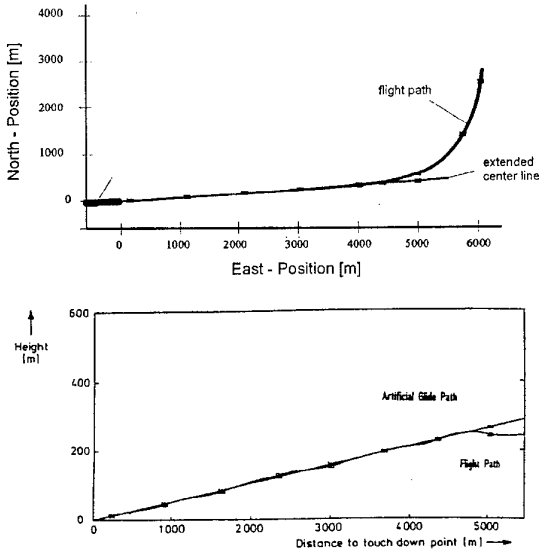


Fig. 17: Automatic GPS-based approach and landing

A flight trial at Hannover airport to compare ILS and Differential GPS demonstrates Fig. 18.

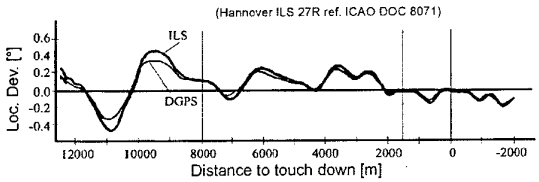


Fig. 18: Comparison of ILS and DGPS

The weather during the approach was fairly bumpy, but the differences between ILS and DGPS are small and within the ICAO CAT III tolerances. This result was the motivation for the Aerodata Company, Braunschweig, to develop a flight inspection system based on Differential GPS as a position reference.

Precise flight path and speed determination improve the quality of the inflight measurement of wind and turbulence. Wind and turbulence velocities with an accuracy of better than 10 cm will

be measured by the Helipod (ref. Fig. 19), developed by Aerodata. This meteorological precision measurement system will be towed by a helicopter.

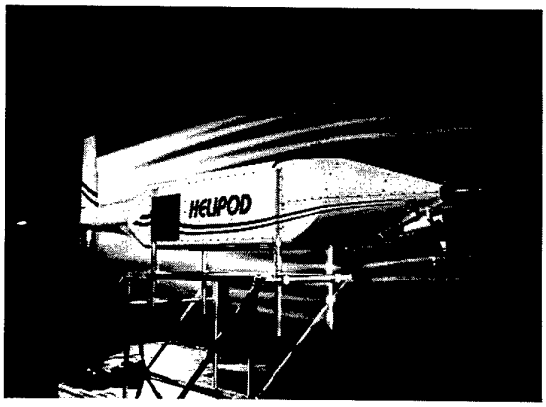


Fig. 19: Helipod for wind and turbulence measurement

An aircraft, equipped with an additional laser distance measurement system (Fig. 20) is able to determine the earth surface contour relative the WGS84 co-ordinate system (Fig. 21), even if the earth surface is covered with forest.

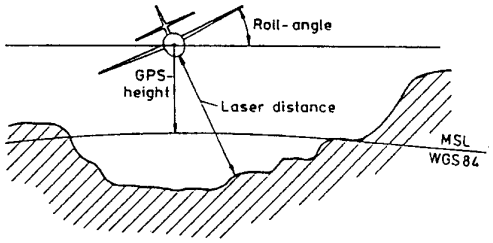


Fig. 20: Earth contour measurement

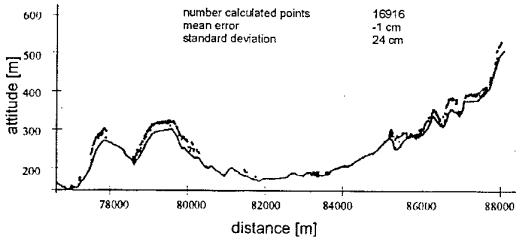


Fig. 21: Laser profile of the Harz mountains

Extreme requirements on accuracy, reliability and continuity of service have been fulfilled for automatic car guidance (Fig. 22). The demands additionally increase, if the roads are narrow and surrounded by high trees. Masking and mulitpath are much more critical compared to aircraft guidance.

Fig. 22: Automatic car guidance

Conclusion

Satellite navigation is extreme powerful and successful for global and local position finding and navigation. With additional procedures and equipment, the accuracy of satellite navigation can be significantly improved. An accuracy of 1 cm in manoeuvrable vehicle is sate of the art and additional improvements are foreseeable. In contrast to this excellent accuracy of satellite navigation is its poor dynamic response and its extreme poor reliability that can probably not overcome by means of radio navigation. Complementary sensors can give the required reliability, if designed in an integrated navigation. Applications e.g. in measurement technique or vehicle guidance are only limited by the constraint vision of developers and inventors.

Acknowledgement

The research activities of the author have been significantly supported by the

- German Research Society (DFG) via the Collaborated Research Initiative 420 "Flight Measurement Technique"
- Ministry of Research and Development (BMFT)
- Ministry of Transport (BMV)
- Lower Saxony Minister for Science and Culture

The author expresses his thanks to these organisations.

References

[1] Hatch, R.; Synergism of GPS Code and Carrier Measurements; Third International Symposium on Satellite Doppler Positioning, Austin TX, 1982

[2] Jacob, T.; Beitrag zur Präzisionserkennung von dynmisch bewegten Fahrzeugen; Thesis; Technical University Braunschweig, 1992

[3] Kayser, D.; Effects of the Specific Military Aspects of Satellite Navigation on the Civil Use of GPS / GLONASS; AGARD GUIDANCE and Control Panel, Rome, 1994

[4] Lechner, W., Vieweg, S.; Realisation of GNSS – Resulats of Combined GPS/GLOANSS data Processing; 6th International Technical Meeting ION GPS 93, Salt Lake City, 1993

[5] Lipp, A., Jagieniak, S.; Experiences Using Carrier-Phase Positioning for Real Time Flight Path Calibration

[6] Schänzer, G.; High Precision Navigation for Vehicles; IECON 1994

[7] Schänzer, G.; Aspects of Safety Critical Applications of Satellite Navigation Systems; GNSS 97

Monitoring and Adaptive Robust Protection of the Integrity of Air Data Inertial Satellite Navigation Systems for Maneuverable Aircraft

(May 1999)

G.I. Djandjgava, A.P. Rogalev,
A.V. Chernodarov
Ramenskoye Design Company
140103, Gurjev Street,2, Ramenskoye, Moscow Region
Tel. (095) 556-2393, 556-5038 ; Fax (09646) 3-1972
e-mail: extpp@cityline.ru
Russia

Abstract

This paper is devoted to the problem of maintaining the integrity of navigation systems (NS's) of maneuverable aircraft under real noise environment. The proposed solution of this problem relies on the potentialities of hardware and algorithmic redundancy when constructing the loops for primary and secondary signal processing. Such a redundancy permits one to improve the reliability of estimation of NS state parameters under uncertainty and also to ensure the mutual support of NS's if critical situations occur. Hardware redundancy is assumed to be attained by the integration, into a unified navigation-time space, of air data, inertial, and satellite measuring channels. It is also assumed that algorithmic redundancy is achieved by the integration, into a single information space, of neural-network procedures for adaptive robust signal processing and combined procedures for detecting and counteracting outliers and failures, too. The effectiveness, as applied to integrated NS's, of the algorithms obtained has been confirmed by the results of half-scale modeling.

Introduction

This paper relating to the domain of noise-immune navigation reveals the possibilities of using the methods of adaptive robust signal processing for improving the objectivity of monitoring the integrity of functionally bound airborne systems and also the reliability of state estimation of such systems.

The term "integrity", as applied to a navigation system (NS), reflects the ability of the NS to maintain the required operational characteristics irrespective of its operation conditions.

The familiar solutions of this problem rely on the detection of failed components, the elimination of such components from the structure, and on the restoration of serviceability of NS's by means of hardware reconfiguration. For example, in inertial satellite navigation systems, the above-mentioned functions are performed by the following systems: the RAIM (Receiver Autonomous Integrity Monitoring) system , the AIME (Autonomous Integrity Monitor Extrapolator) system, the AAIM (Aircraft Autonomous Integrity Monitoring) system, and so on. The implementation of such an approach calls for a substantial hardware redundancy which seems to be impossible in a number of cases. Redundancy reduction may result both in a breach of continuity of navigational support and in a violation of air navigation safety. One possible approach to the solution of this particular problem is based on the integration of NS's, i.e., when their integrity is ensured by the mutual support of measuring means that are physically different in nature. Such an interaction of the NS's with each other allows one to retain or to reduce gradually their performance qualities if critical situations occur and noise conditions change.

At present, the theoretical foundation for the integration of NS's is the mathematical apparatus of the extended Kalman filtering (EKF) and decision theory. However, under the conditions of statistical and parametric uncertainty, the realization of integration characteristics of NS's on the basis of such a mathematical apparatus involves a number of difficulties caused by the possible loss in integrity of the signal processing system itself.

By the integrity of an integrated data processing (IDP) system is meant the state of this system such that the required estimation reliability of navigational parameters is ensured. The reliability, in its turn, is characterized by the no-divergence condition of the EKF, i.e., by the condition where the estimates obtained fit their predicted mean-square values adequately. This gives grounds to include, in the structure of the IDP system, the following loops intended for the protection of the EKF from its divergence [1,2]:

- a guaranteed-protection loop which minimizes the risk of losing the integrity due to the fact that the generalized state parameters of the IDP system fall outside the permissible limits;
- a robust-protection loop which minimizes the risk of losing the integrity in circumstances where the actual and "simulated", i.e., a priori assumed, distribution laws for the generalized state parameters of the IDP system are inconsistent with each other;
- an adaptive-protection loop which provides the parametric tuning, for the actual operation conditions, of the IDP system having a robust risk-oriented architecture.

Paper presented at the RTO SCI International Conference on "Integrated Navigation Systems", held at "Elektropribor", St. Petersburg, Russia, 24-26 May 1999, and published in RTO MP-43.

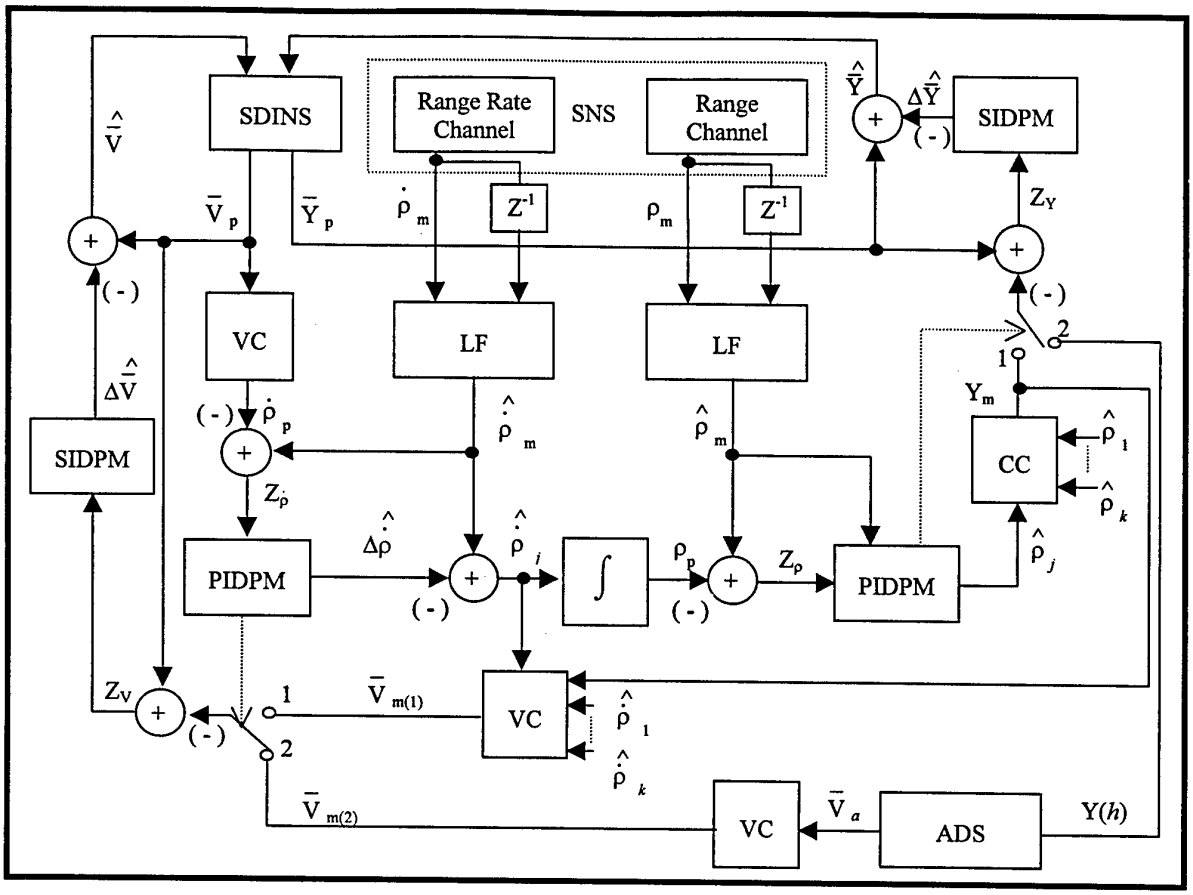


Fig. 1. Block-diagram of an integrated air data inertial satellite navigation system

The technique used to unite, in the interests of counteracting the uncertainty, the above-mentioned loops into an integrated structure relies on the theory of optimization of stochastic systems on the basis of nonclassical cost functionals. This theory is considered in [1,2].

The purpose of this paper is to justify a single methodology of counteracting the uncertainty and protecting the integrity of NS's.

The conceptual basis for the methodology proposed here relies on the solution of the following problems:

1. Provision of functional redundancy for all the hierarchies of integrated NS's.
2. Formation of generalized parameters that characterize both estimation accuracy and the integrity of an IDP system.
3. Formation of a system of state symptoms and decision rules, which make it possible to recognize NS states using generalized parameters.
4. Ranking of states according to their influence on the integrity of NS's.
5. Unification of procedures intended for estimation, integrity monitoring, and control of NS states into an integrated structure with the protection properties that ensure:

- reliable estimation and prediction of NS state parameters under the conditions of uncertainty;
- objective monitoring and prediction of the integrity level of NS's;
- rational control of hardware and/or algorithmic support which provides the required integrity level of NS's;
- guaranteed NS survival as the integrity level and redundancy of the information space reduce.

6. Provision of an intelligent support for the interaction of integrity protection loops.

The solution of the above-mentioned problems is based on the combination of the following possibilities of adaptive robust (AR) and neural network processing of information. In the structure of AR- algorithms, redundant parameters enter. The tuning of this parameters to the actual measurement process makes it possible to improve the reliability of estimating the state of an NS under the conditions of uncertainty. Such a structural parametric redundancy involves a multiple-

choice realization of AR-algorithms, which is the constructive basis for neural networks. At the same time, the neural-network-based architecture of the IDP system implies a successive refinement (by layers) of multiple-choice solutions and "weighting" of these solutions using the corresponding confidence coefficients. The construction of AR-algorithms and their properties allow multilevel formation of such coefficients. The implementation of neural network technique for the monitoring and AR-protection of integrity is considered using an example of an integrated air data inertial satellite NS (ADISNS).

The structure of the ADISNS is shown in Fig. 1, where SDINS is the strapdown inertial NS; SNS is the satellite NS; ADS is the air date system; CC is the coordinate converter; VC is the velocity converter; LF is the lattice filter [2]; PIDPM is the preliminary-integrated-data-processing module; SIDPM is the secondary-integrated-data-processing module; $\bar{Y}$ is the vector of position parameters of an object; $\bar{V}$ is the vector of the relative velocity of an object; ρ is the distance between an object and the satellite; $\dot{\rho}$ is the rate of change of distance between an object and the satellite; Y(h) is the barometric altitude; $\bar{V}_a$ is the vector of airspeed of an object; $Z_{(.)}$ is an observation of the relevant parameters; $\hat{(...)}$, Δ are an estimate and an error, respectively; p, m are subscripts that denote predicted and measured parameters, respectively; 1, 2 are switch position corresponding to the state of health and also to a failed state of SNS channels; Z^{-1} is the operator for one clock-pulse delay.

In the ADISNS, integration of NS's is ensured by their mutual support at the level of both secondary (SDP) and preliminary (PDP) data processing. Here, in the SDP loop, the inertial NS is a master NS, and, in the PDP loop, the satellite NS is a master NS.

1. Monitoring and Adaptive Robust Protection of the Integrity of Primary Navigational Determinations

At present, the necessity for inflight monitoring of the integrity of satellite networks of navigation systems is considered to be justified [3,4]. This type of monitoring provides for the detection of a failed channel and for the exclusion of it from the process of timed navigational determinations (TND).

The topical problem that remains to be solved is, however, the problem of recognizing gradual failures against the background of outliers.

The solution of this problem will make it possible to raise the level of satellite navigation systems availability and TND continuity, because outliers can be counteracted using algorithmic procedures. For this purpose, in [2], the use of the mathematical apparatus of adaptive robust processing of navigational information (ARPNI) is proposed. The kernel of this apparatus involves weigthing functions (influence functions) for the generalized state parameters of an ARPNI system. These functions determine the level of confidence in incoming measurements. The functions mentioned above are formed having regard either to a priori assumptions concerning distribution laws for the valid signal and noise, or to a posteriori statistical characteristics of an innovation sequence. Here the generalized parameters must be chosen so that the regions of states of NS's and the ARPNI system can be put in correspondence with them.

In [2], the following was assumed as a basis for the solution of this problem:

(a) as the basic parameter for the first optimization level that characterizes the accuracy of estimation, the normalized residual β_i between the actual value z_i and the predicted value $\hat{z}_i$ of the measurement was chosen, i.e.,

$$\beta_i = \nu_i / \alpha_i,$$

where $\nu_i = z_i - \hat{z}_i$; α_i is a scaling parameter;

$\hat{z}_i = h\{\varphi[y_{n(i-1)}]\} + \Gamma_i \xi_{i-1}\}$;

y_n is the reference state vector of an NS; h, φ are vector functions having the appropriate dimensionalities; ξ_i is a stochastic sequence with covariance matrix Q_i and transition matrix Γ_i;

(b) as the generalized parameter for the second optimization level that characterizes the integrity of an ARPNI system, the square of the normalized residual β^2_i was chosen;

(c) the necessary condition for the no-divergence state of the EKF, i.e., the χ^2- distribution of the generalized parameter

$$\beta^2_i \in \chi^2(1;2) \qquad (1)$$

was taken into account;

(d) a priori known numerical characteristics of the scalar parameter β^2_i, having the χ^2 – distribution–mathematical expectation $M[\beta^2_i] = 1$ and variance $D[\beta^2_i] = 2$ – were taken into consideration;

(e) multilevel optimization of the estimated state vector (SV) $\hat{y}_i$ was performed on the basis of a nonclassical cost functional which includes the basic parameter and a generalized one, i.e.,

$$\hat{y}_i = \underset{y_i}{argmin} \frac{1}{S} \sum_{i=i_0+1}^{i_f} d^S_i \quad , \qquad (2)$$

$d_i = \rho(\beta_i) + \Delta\rho(\beta_i) - 0.5\gamma^2$;

$\rho(\beta_i) = -\ln f(\beta_i)$; $\quad \Delta\rho(\beta_i) = \ln \Delta f_i$;

$$\Delta f_i = 1 + \frac{K_3}{3!} H_3(\beta) + \frac{K_4}{4!} H_4(\beta) + \ldots \quad ;$$

$f(\beta_i)$ is a probability density function;

K_n, H_n are an n-th-order cumulant and a Hermitian polinomial of degree n, respectively [5];

$\gamma^2_i = M[\beta^2_i] + 3\sqrt{D[\beta^2_i]} \approx 5.2$ is the confidence interval for measurements without outliers.

The solution of Eq.(2), obtained in [2], relies on the Sridhar optimization theory which, in turn, is based on the Pontryagin principle of the maximum and on the method of invariant imbedding [6].

Figure 2 gives this solution in the form of a modified AR-algorithm with layerwise tuning meant for the neural-network-oriented realization. In this figure, the following notation was introduced: M_j, P_j are the values of covariance matrices for estimation errors of the SV at the i-th step after processing the j-th element and all the vector ν_i of residuals, respectively; m_j is an SV estimate at the i-th step after processing the j-th residual ν_j

$H^T_j = \left.\frac{\partial h(y)}{\partial y}\right|_{y=y_n}$ is the vector of coefficients of the connection of the j-th residual with estimation errors of the SV;

Ψ_j, Ψ_j' are an influence function [2, 20] and its partial derivative with respect to β, respectively;

$$\Delta\Psi_j = \Delta f'_\beta(\beta_j)\,\Delta f^{-1}(\beta_j) \quad ;$$

$$\Delta\Psi'_j = \Delta f''_{\beta\beta}(\beta_j)\,\Delta f^{-1}(\beta_j) - \Delta\Psi^2_j.$$

The parameters Ψ, Ψ', $\Delta\Psi$, $\Delta\Psi'$, d, and S are redundant with reference to the EKF. The tuning of these parameters to the actual measurement process makes it possible to improve the reliability of estimating the state of an NS. Also, it is apparently possible to process residuals with the appropriate confidence coefficients which provide resistance to random failures. But, in the structure of the algorithm mentioned above, the mechanism for detecting and counteracting gradual failures is lacking. We propose that such a mechanism should be constructed on the basis of Fisher's statistic criterion [5], i.e.,

$$\hat{F}_{i(j)} = \frac{\hat{\mu}_{2\,i(j)}}{\mu_{2\,i(j)}} \quad , \qquad (3)$$

where $\hat{\mu}_{2\,i(j)}$, $\mu_{2\,i(j)}$ are the current estimate of the residual variance ν_j on a moving time interval and its predicted value, respectively.

For the Gaussian residual $\nu_{i/j}$ and for the adequacy of its predicted variance and an actual one, the random variable (3) possesses a ϑ^2–distribution with N degrees of freedom [7], i.e.,

$$\hat{F}_{i(j)} \in \vartheta^2\{M[F_{i(j)}]; D[F_{i(j)}]\}, \qquad N > 4,$$

where

$$M[F_{i(j)}] = \frac{N}{N-2}; \qquad D[F_{i(j)}] = \frac{4N(N-1)}{(N-2)^2(N-4)};$$

N is the sample size on a moving time interval.

Then, in view of the "three-sigma rule"[8], the necessary condition for the state of operability of the j-th channel of an NS can be determined as

$$\hat{F}_{i(j)} < \vartheta^2_j = M[F_{i(j)}] + 3\sqrt{D[F_{i(j)}]}\,. \qquad (4)$$

Thus, the proposed system for integrated protection of the integrity of channels of the primary navigational determinations (PND) can be represented by the following set of interconnected loops (see Fig.1):

- a loop intended to integrate PND channels that embody dissimilar physical principles of operation and that possess different spectral characteristics of noise;
- a loop intended to detect outliers and to counteract them using adaptive robust procedures;
- a loop intended to detect failures and to counteract them through the reconfiguration of PND channels by hardware;
- a loop intended to select the given number of operable PND channels from a minimum of the square of a normalized residual for secondary navigational determinations (SND).

2. Monitoring and Adaptive Robust Protection of the Integrity of Secondary Navigational Determinations

During the secondary data processing [9] the parameters are formed, which describe the translational and rotational motion of an object in a navigational coordinate frame.

At present, it is considered that, in order for the above-mentioned problem to be solved, the use of dissimilar (in the principle of operation) measuring devices and the unification, into an integrated structure, of these devices on a basis of the procedures of extended Kalman filtering is justified [10]. Therefore, the necessity arose of protecting the integrity of a system for secondary integrated data processing (SIDP). By the integrity of an SIDP system, in this case, is meant the state of the system such that the reqired estimation reliability of navigational parameters is ensured. The reliability, in its turn, is characterized by the no-divergence condition of the EKF [11], i.e., by the condition where the estimates obtained fit their predicted mean-square values adequately. This gives grounds to consider the loops intended for AR-protection of the EKF from divergence as a means for the maintenance of the integrity of an SIDP system.

In order to provide an intelligent support of the interaction of the aforementioned loops, it is apparently advisable to include learning procedures as components of the structure of an SIDP system. The neural network technique [12] used in the construction of such procedures is believed now to have considerable

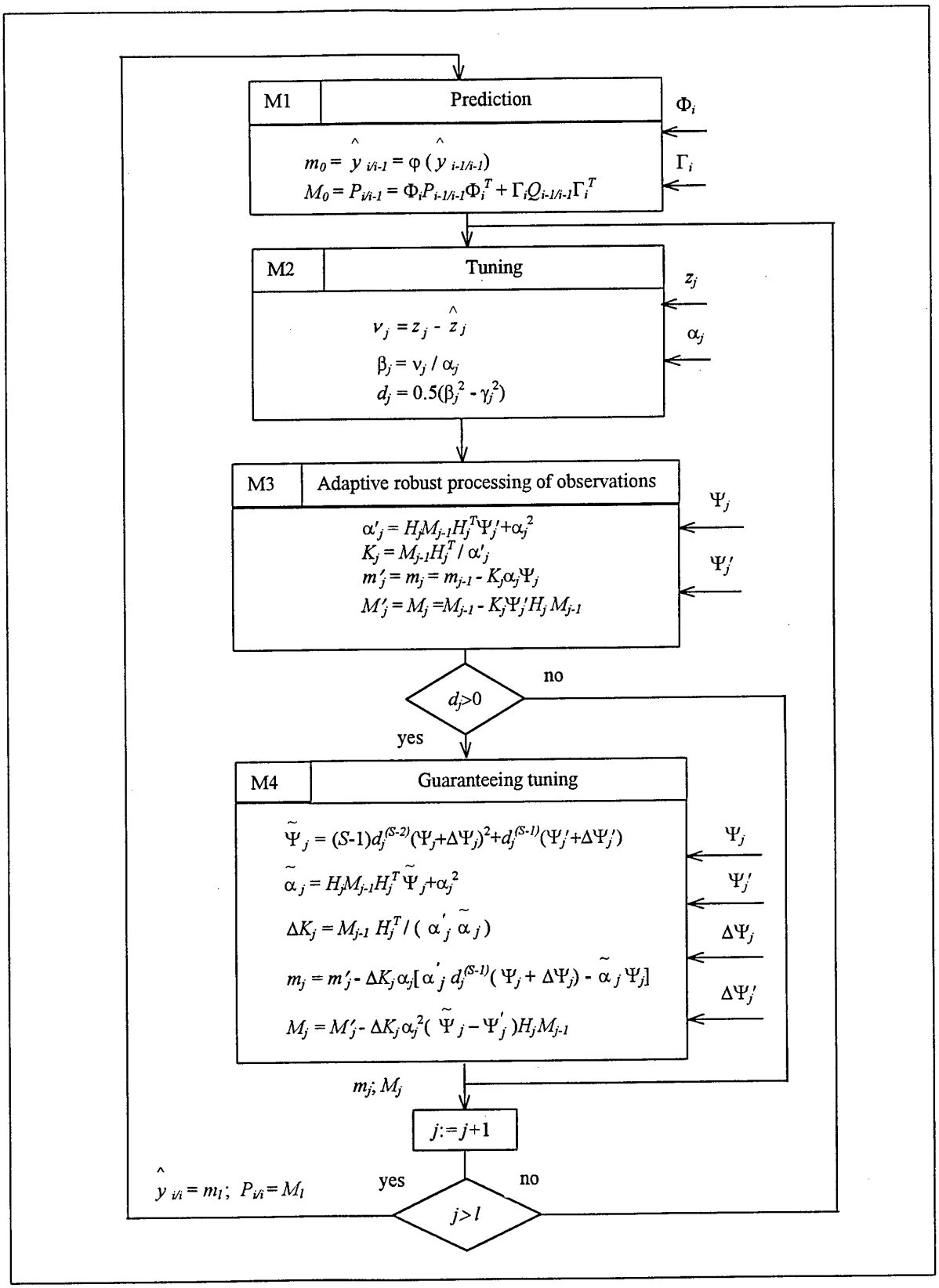


Fig.2. Block diagram of an integrated data processing system with tuning "by layers"

promise. This technique, as applied to problems of state estimation of dynamical systems, includes the following:

- seeking of a solution using multiple-choice procedures;
- "weighting" of the results sought;
- convolution of the weighted solutions;
- tuning of the weighting (learning) function from the results sought.

An analysis of well-known works [13-16], publishished in this field, shows that

(a) the proposed schemes for the mapping of SIDP procedures onto neural networks rely on the construction of the classical Kalman filter. This places a limitation on the possibilities of applying a multiple-choice approach to the seeking of an optimal solution under the conditions of uncertainty, because, in such a case, the above-mentioned approach reduces, in effect, to a multimodel interpretation of the EKF [17];

(b) the following three questions are still open ones : (1) the justification of the construction of a learning function, (2) the justification of the procedure of including this function as a component of the structure of an SIDP system, and (3) the assignment of weighting coefficients when forming the convolution of estimates;

(c) the familiar neural-network modifications of the EKF are not meant for a multilayer realization, and this places a limitation on the possibilities of these modifications as to the use of auxiliary information for lowering the level of uncertainty.

The technique that is proposed here for estimating the performance of channels of the network consisting of the modules M3 and M4, which are built on the sets of the controlling parameters Ψ_j, Ψ'_j, and S_j, relies on the results of [18] . The procedure of pairwise comparisons of the estimation errors δ whith each other in the corresponding network channels as well as the necessary conditions for the no-divergence state of filtering are at the basis of this technique.

Such a procedure can be realized by analyzing the quadratic form

$$J_{j(qk)} = \Delta\delta^{T}_{j(qk)}\, \Delta M^{-1}_{j(qk)}\, \Delta\delta_{j(qk)} \quad , \qquad (5)$$

where $\Delta\delta_{j(qk)} = m_{j(q)} - m_{j(k)}$;

$\Delta M_{j(qk)} = M_{j(q)} + M_{j(k)}$;

$m_{j(k)}$, $M_{j(k)}$ are an SV estimate and its covariance matrix, respectively; the estimate and matrix are obtained by the k-th channel of the network at the j-th instant of time when the j-th observation is processed.

With no divergence of the estimation process ($\delta < 3\sigma$), the random variable (5) must have the χ^2 –distribution with n degrees of freedom, where n is the SV dimension. Considering the characteristics of this distribution [7], i.e., $J_{j(qk)} \in \chi^2(n\,;2n)$ for $\delta < 3\sigma$, the necessary condition for stable operation of the pair of channels, which is to be analyzed, can be formed as

$$J_{j(qk)} \le \eta_j \,, \qquad \text{where } \eta_j = n + 3\sqrt{2n} \,. \qquad (6)$$

Thus, the channels of a neural-network observation processing system can be ranked pairwise, depending on the value of the deviations

$$\Delta\eta_{j(qk)} = J_{j(qk)} - \eta_j \qquad (7)$$

In what follows, information is processed for the pair of channels, for which the value $\Delta\eta_j$ is a minimum, i.e., for $\Delta\eta_j = \Delta\eta_{j(min)}$. If the condition

$$\Delta\eta_{j(min)} \le 0 \qquad (8)$$

holds, then, from the averaged data relating to the selected pair of channels, at the output of the corresponding network layer, the final state vector estimate and its covariance matrix are formed, i.e.,

$$m_j = 0.5[m_{j(k)} + m_{j(q)}] \,; \qquad (9)$$

$$M_j = 0.5\Delta M_{j(qk)} \qquad (10)$$

When condition (8) is violated, the transition is effected from the layer composed of the modules M3 to the layer composed of the modules M4. If, in addition, the generalized parameter β_j^2 is within the permissible limits, i.e., if condition (1) is satisfied, one can say that the level of the predicted uncertainty of SV estimates at the output of the first processing layer is underestimated. The required agreement between the SV estimates and their confidence intervals can be regained by means of the appropriate correction of elements of the "a posteriori" covariance matrix ΔM_j. It is proposed that, in order to ensure the fulfilment of condition (6), such a correction should be made via the U-D components of the matrix ΔM_j. As is known [19], this matrix can be represented as

$$\Delta M_j = U_j D_j U_j^{T} \,, \qquad (11)$$

where U_j is an upper triangular matrix with identity diagonal ; D_j is a diagonal matrix.

Then the quadratic form (5) and condition (6) can be written as follows:

$$J_{jn} = (U^{-1}{}_j\Delta\delta_j)^{T} D^{-1}{}_j (U^{-1}{}_j\Delta\delta_j) = \Delta\tilde{\delta}^{\,T}{}_j D^{-1}{}_j \Delta\tilde{\delta}_j =$$

$$= \sum_{k=1}^{n} \frac{\Delta\tilde{\delta}^{\,2}{}_{jk}}{D_{jk}} \le \eta_j = n + 3\sqrt{2n} \quad ; \qquad (12)$$

$$J_{jr} = J_{j(r-1)} + \frac{\Delta\tilde{\delta}^{\,2}{}_{jr}}{D_{jr}} \le \eta_{jr} = r + 3\sqrt{2r} \,, \qquad (13)$$

where $\Delta\tilde{\delta}_j = U^{-1}_j \Delta\delta_j$;

$$J_{j(r-1)} = \sum_{k=1}^{r-1} \frac{\Delta\tilde{\delta}^2_{ik}}{D_{jk}};$$

$\Delta\tilde{\delta}_{jr}$, D_{jr} are the r-th elements of the vector $\Delta\tilde{\delta}_j$ and the diagonal matrix D_j, respectively.

When condition (13) holds, the estimation process may be considered as a divergence–free process with fiducial probability 0.97. The violated condition for the r-sum of J_{jr} can be restored, however, by correcting the element D_{jr}, i.e.,

$$D_{jr} := \tilde{D}_{jr} = \frac{\Delta\tilde{\delta}^2_{jr}}{[\eta_{jr} - J_{j(r-1)}]} \qquad (14)$$

From this the following procedure of adaptive tuning of the "a posteriori" covariance matrix ΔM_j is inferred:

$$\Delta M_j := \Delta\tilde{M}_j = U_j \tilde{D}_j U^T_j, \qquad (15)$$

where

$$\tilde{D}_{jr} = \begin{cases} \tilde{D}_{jr} & \text{for } J_{jr} > \eta_{jr}; \\ D_{jr} & \text{for } J_{jr} \le \eta_{jr}. \end{cases}$$

If condition (8) is not violated after the observations are processed by the layer composed of the modules M4, then, using relations (9) and (10), the output signals of an SIDP system are formed; otherwise the observations are processed by this layer once more for new values of the parameters S_k. In such a case, the tuning is performed in the regions of those values of the above- mentioned parameters, for which a minimum of the residual $\Delta\eta_k$ corresponds. Such a learning SIDP system can be structurally shown in Fig.3.

3. Half-Scale Modeling and Analysis of the Results

The efficiency of algorithms for the monitoring and adaptive robust protection of the integrity of TND has been tested using the half-scale model of an ADISNS, described in [1]. The reference phase trajectory (RPhT) was formed from the smoothed data of an airborne registrator of a maneuverable aircraft (MA).

This has made it possible to study the effect of realistic dynamics of MA motion on the accuracy qualities of sensors. The output signals of sensors were determined from noisy RPhT parameters. In that case, for a SDINS, root-mean-square errors (RMSE) of calibration of sensors used for simulation were taken as follows: $1\sigma_a = 0.1mg$, $1\sigma_w = 0.1$deg/h, where σ_a is an accelerometer RMS bias; σ_w is a gyro RMS drift. Moreover, it was assumed that actual accelerometer signals possess cross–correlated dynamic errors that are proportional to the corresponding accelerations. The SDINS state vector contained positional parameters, projections of relative velocity and acceleration on the referred triad, attitude, and the angular rate of rotation of the inertial measuring unit (IMU). During the secondary data processing the model of errors, which covered 31 parameters, was put in correspondence with the SDINS state vector.

For the constellation of five satellites, the primary navigational determinations were formed at a frequency of 1 Hz. It was assumed that, for a range channel, to the state of operability there corresponds a Gaussian error with RMSE that is no more than 20m, to a malfunction there corresponds an outlier with RMSE that is no less than 100m, to a failed state there corresponds a random shift with RMSE that is no less than 100m. In that case, the proportion of anomalous signals on a moving time interval was no more than 30 percent. The operation of velocity channels of such an SNS was simulated by the same relationship among possible states and by the Gaussian RMSE of 0.1 m/s. In these channels, loops intended for prediction and for the formation of observations were similar to those of [2]. Accuracy qualities of an air data channel were as follows: $1\sigma_v = 1$ m/s for velocity; $1\sigma_h = 150$m for altitude.

During the processing of observations by layers, the following controlling parameters were given:

- a set of influence functions Ψ_j for the modules M3, which was obtained by forming the convolution of typical distributions [2,20] for "ε - unrectified " noise;
- increment $|\Delta S| = 0.1$.

Some characteristic results of the experiment are presented in Figs. 4-7, where the following circular estimation errors of the position δ_p, σ_p (m) and the velocity δ_p, σ_p (m/s) of an MA are shown:

- the actual error $\delta_p = [\delta^2_{(\varphi)} + \delta^2_{(\lambda)} + \delta^2_{(h)}]^{1/2}$, where $\delta_{(.)} = y_{(.)} - \hat{y}_{(.)}$;
- the potential RMSE σ_i, which was computed as the square root of the sum of the respective diagonal elements of the covariance matrix P_i.

In the case of a conventional EKF, estimation errors are shown in Figs.4 and 5, and for a neural-network AR-algorithm having the property of counteracting outliers and failures, such errors are presented in Figs.6 and 7.

A comparison of the results shows that the estimates obtained by the conventional EKF during a highly maneuvering flight under real noise environment are not stable ($\delta_i > 3\sigma$), while those obtained by the AR filter with neural-network tuning are stable ($\delta_i < 3\sigma$). In addition, the functioning of both these filters is characterized by a comparable level of the predicted potential accuracy (1σ).

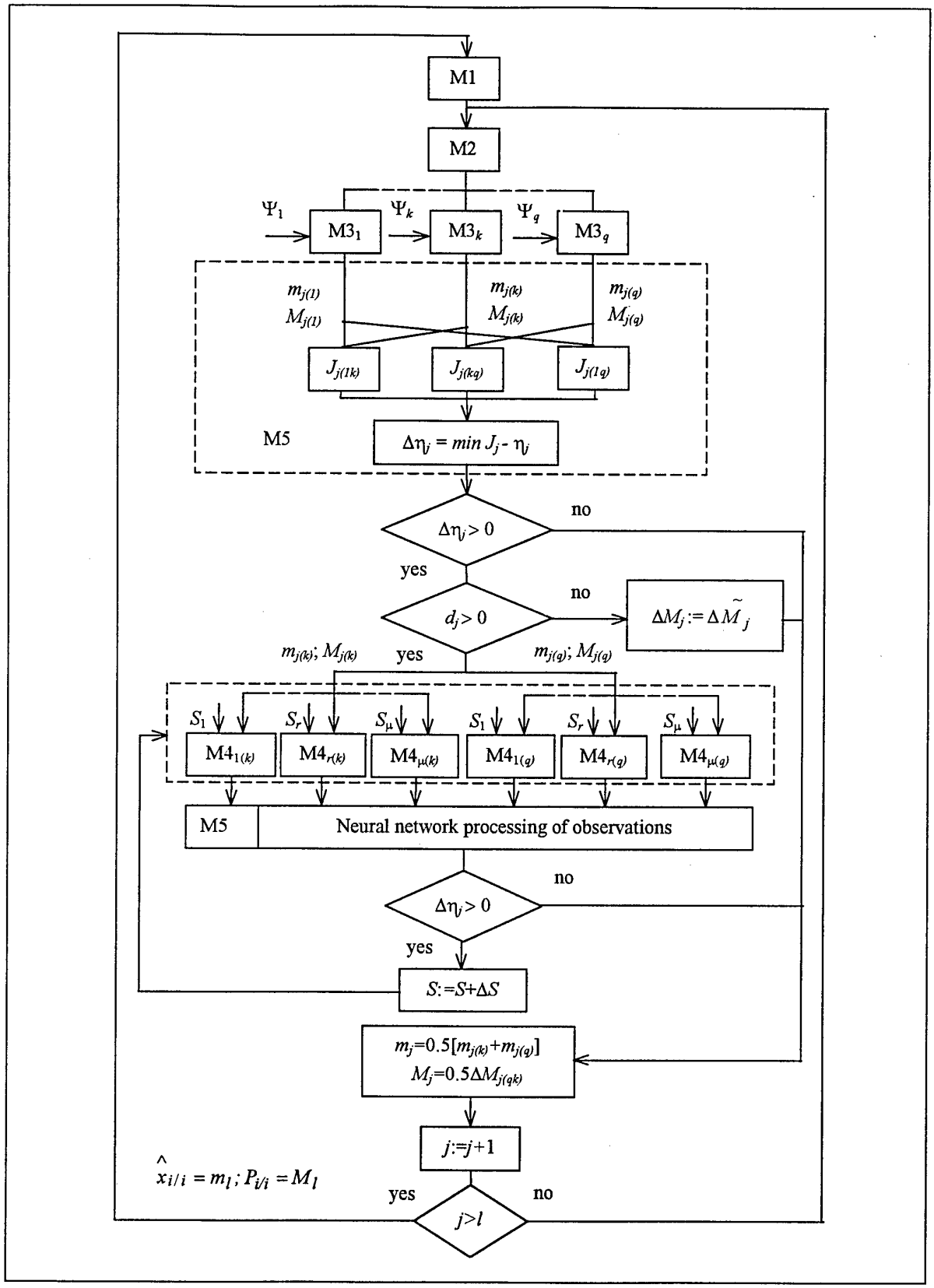


Fig.3. Block diagram of a learning system for integrated data processing

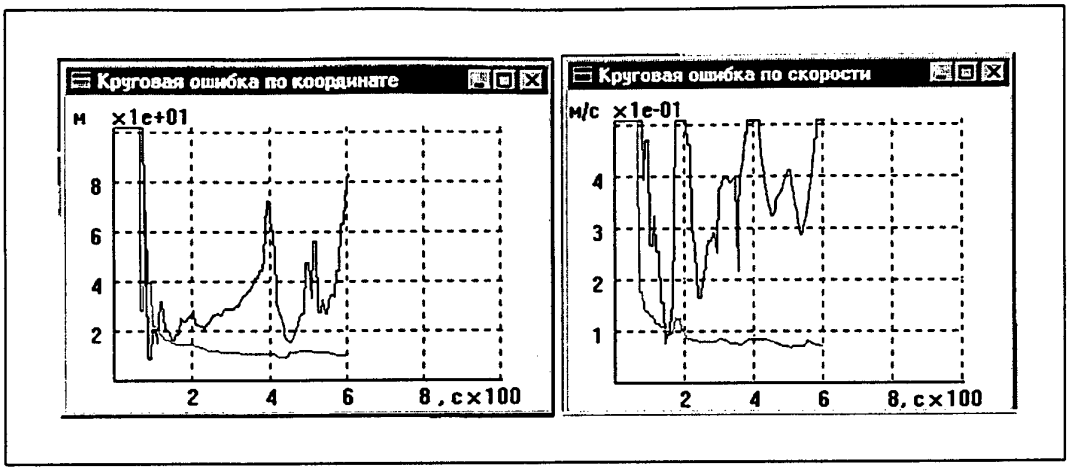


Fig. 4

Fig. 5

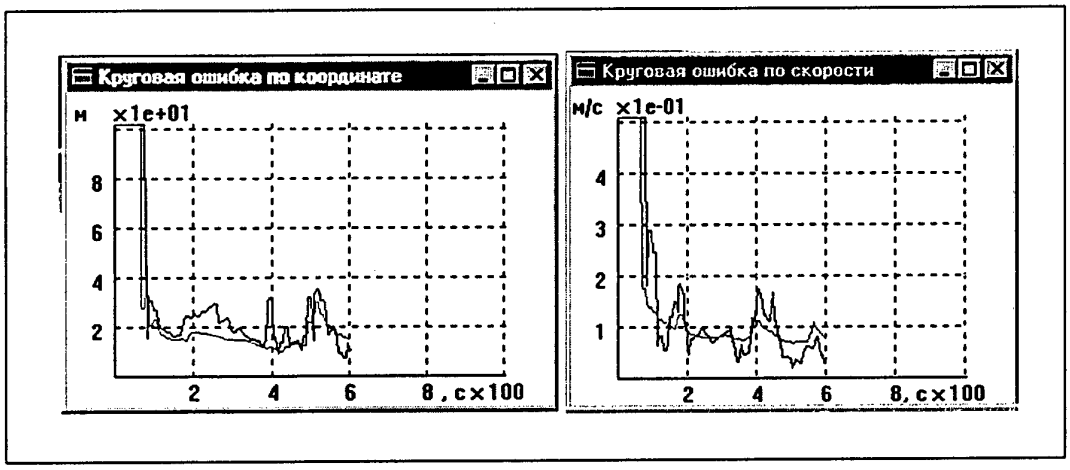


Fig. 6

Fig. 7

Possible loss in the integrity of integrated NS's the kernel of which is an EKF with a loop intended to select failed SNS channels by the χ^2–criterion [3] is caused by the following. A complete exclusion of outliers from data processing and also an "inadequate" response of an IDP system to prefault conditions degrade the availability of an SNS for the correction of SDINS errors of an MA. A nonlinear build-up of these errors, which results from this, brings about a decrease in the signal-to-noise ratio. Under such conditions, the application of the EKF apparatus for the integration of NS's becomes incorrect and inefficient. Moreover, the weak dependence of an EKF gain on the innovation sequence makes it impossible to counteract the loss in the integrity of NS" due to a change in the spectral characteristics of disturbances and noise. This confirms the fact it is expedient to apply combined goodness-of-fit tests for the monitoring of NS status and also to employ adaptive robust approaches to the protection of the integrity of NS's.

Conclusions

The technique proposed here for processing redundant measurements makes it possible

- to detect and to recognize random and gradual failures, and prefault conditions;
- to resist random failures and outliers by means of algorithmic reconfiguration and tuning of the loops that provide adaptive robust protection of the integrity of the DSP systems;
- to resist gradual failures and prefault conditions by means of hardware reconfiguration or on the basis of combined algorithmic–and-hardware redundancy;
- to ensure the integrity of navigational support in critical situations at the expense of both reducing, gradually, the redundancy of the information space and accumulating algorithmic protection means.

In the present paper, a neural-network structure is also proposed for the construction of a loop intended for intelligent support of the integrity of an ARPNI system. To accomplish this, the ARPNI system was brought to a multilayer architecture, performance indices were formed, together with the corresponding technique for controlling the state of each observation processing layer.

Unlike recurrence-search estimation algorithms [21], which are based on an analysis of the input innovation sequence, the proposed technique relies on the properties of differences of "a posteriori" SV estimates in neural-network channels. The utilization of the properties mentioned above has made it possible to replace the process for empirical assignment of unformalizable parameters of the ARPNI system by learning procedures. On the whole, the neural-network-based realization of an IDP system provides for the lowering of the degree of uncertainty of parameter estimates for nonlinear NS's by tuning scalar influence functions for sequentially processed observations, and not by solving the Stratonovich-Kushner equation [9] for "a posteriori" probability density of an n-dimensional state vector.

References

1. Chernodarov A.V., Kozhenkov L.J,. Sorokin G.V., Kovregin V.N. Integrated data processing in SINS/GPS based systems for high maneuvering aircraft positioning and attitude determination. Proc. of the 4th Saint-Petersburg Conference on integrated navigation systems. St. Petersburg: 1997, pp. 388-397.
2. Chernodarov A.V., Djandjgava G.I., Rogalev A.P. Inertial - sensor - signals adaptive robust processing with an H_2 residual bound. Proc. of the 4th Saint-Petersburg Conference on integrated navigation systems. St. Petersburg: 1998, pp. 119-126.
3. Graas F. Signals Integrity. In AGARD Lecture series 207,1996, pp. 7/1-12.
4. Kharisov V.N., Perov A.I., Boldin V.A. (editors). The global satelllite radio-navigational system GLONASS [in Russian]. Moskow: IPRZhR, 1998.
5. Cramer H. Mathematical methods of statistics. Princeton , N.J.: Princeton University Press, 1946.
6. Sage A.P., Melse J.L. Estimation theory with application to communication and control. New York: Mc Graw-Hill, 1972.
7. Korn G.A., Korn T.M. Mathematical handbook. New York: Mc Graw –Hill, 1968.
8. Korolyuk V.S., Portenko N.I., Skorokhod A.V., urbin A.F. A handbook of probability theory and mathematical statistics [in Russian]. Moskow: Nauka, GRFML, 1985.
9. Stepanov O.A. Application of nonlinear-filtering theory in problems of processing of navigational information [in Russian]. St. Petersburg, TsNII "Electropribor", 1998.
10. Philips R.E., Schmidt G.T. GPS/INS Integration. In AGARD Lecture series 207,1996, pp. 9/1-18.
11. Fitzgerald R.J. Divergence of the Kalman filter. IEEE Trans. On Automatic Control, 1971 , vol .AC-16, № 6, pp. 736-747.
12. Simpson P.K. Artificial neural systems. New York: Pergamon Press, 1990.
13. Agarval M. Combining neural and conventional paradigms for modelling, prediction and control. Int. J. Systems Science, 1997, vol. 28, № 1, pp. 65-81.
14. Wu Z.Q., Harris C.J. A neurofuzzy network structure for modelling and state estimation of unknown nonlinear systems. Int. J. Systems Science, 1997, vol. 28, № 4, pp. 335-345.
15. Silva I.N., Arruda L.V.R., Amaral W.C. Robust estimation of parametric membership regions using artificial neural networks. Int. J. Systems Science, 1997, vol. 28, № 5, pp. 447-455.
16. Chaer W.S., Bishop R.H., Ghosh J. A mixture-of-experts framework for adaptive Kalman filtering. IEEE Trans. on Systems, Man, and Cybernetics – p. B: Cybernetics,1997, vol.27, № 3, pp. 452-464.
17. Magill D.T. Optimal adaptive estimation of sampled stochastic processes. IEEE Trans. on Automatic Control, 1965, vol. AC-10, № 10, pp. 434-439.
18. Chernodarov A.V, Sorokin G.V. Application of the methods of adaptive robust estimation in problems of monitoring and diagnosing the systems of aircraft equipment [in Russian]. In: Improvement of Aircraft Equipment, Mezhvuz. Nauchn . Sb. Moscow: MGTU GA, 1996, pp. 63-73.
19. Bierman G.J. Factorization methods for discrete sequential estimation. New York: Academic Press, 1977.
20. Wu W.-R. Target tracking with glint noise . IEEE Trans. on Aerospace and Electronic Systems, 1994, vol. AES-29, №1, pp. 174-185.
21. Beloglasov I.N., Djandjgava G.I., Chigin G.P. The basis of navigation by geophysical fields [in Russian]. Moscow : Nauka, GRFML, 1985.

Optimal Map-Matching for Car Navigation Systems

S. P. Dmitriev, O. A. Stepanov, B.S. Rivkin, D.A. Koshaev.
(30, Malaya Posadskaya str., Saint Petersburg, 197046, Russia, State Research Center of Russia -Central Scientific & Research Institute Elektropribor,
Tel. 7(812) 232 59 15, Fax 7(812) 232 33 76, E-mail: elprib@erbi.spb.SU

D.Chung (Office 705, Building two,1, Bolshoi Gnezdikovskii str., Moscow, 103009, Russia, Mobile Computing Group, Samsung Electronics Co., LTD,
Tel.7(095) 7 97 24 79, Fax 7(095) 7 97 25 01, E-mail:arog@src.samsung.ru)

Keywords: car navigation, map data, nonlinear filtering, map-matching.

Abstract

A statement and a general solution of the problem of determining a car position on a road by using both external measurements (speed, course and coordinates) and maps of roads are suggested and considered within the framework of the Markovian theory of nonlinear filtering. The aim of the problem is to find the most probable road along which a car is moving and to determine its position to the maximum accuracy. Some algorithms are synthesized and the problem of the potential accuracy analysis is solved. The efficiency of the algorithms developed is tested by using real information about the coordinates, speed and course obtained from a satellite system.

Introduction

At present, car navigation systems using digital road maps (DRM), data from dead reckoning and satellite systems are widely applied [1-4]. It is not uncommon that DRM are used not only to display a car position but also to correct it. The correction is performed by comparing a car route calculated from the data obtained from the satellites or dead reckoning with a set of possible routes formed from the map. Using the map data about the road along which a car is moving it is possible to determine the car position more exactly. The peculiarity of the information processing problem in the navigation systems using the map data is its nonlinear character which considerably complicates the synthesis of the algorithms and analysis of their accuracy.

By now no mathematical framework capable of accounting for the nonlinear character of the problem and all the available information entirely has been proposed for the problem of car navigation discussed here. In real systems without any justification the initial problem is substituted, as a rule, for a few separate subproblems: identification (detection) of the road, linear filtering of the data on a straight road, detection of the road turn, and so on. It is essential that only part of the available information is used in solving each of the separate subproblems.

It should be noted that the navigation method based on comparison of measurements and data from a map (mapping or map-matching navigation) has been quite often used in aircraft and marine navigation systems [5-9]. The algorithms of data processing for aircraft and marine map-matching navigation are also nonlinear, though the Markovian filtering theory is effectively used to develop the algorithms for these systems [5,8,9]. By analogy with it this paper suggests using the Markovian filtering theory as a mathematical framework in solving the car navigation problem under consideration. This approach allows taking account of the nonlinear character of the problem and all the available information entirely. It is essential that the problem of selecting (detecting) the road on which the car is most probably located and the estimation of the car's position on this road to the maximum accuracy are solved within a unified statement. Using the described approach as a basis, some algorithms are being synthesized and the problem of the potential accuracy analysis is being solved. The efficiency of the algorithms developed is tested by using real information about the coordinates, speed and course obtained from a satellite system.

1. The statement and the general solution of the problem within the framework of the nonlinear filtering theory

So, assume that the problem of finding the car's position on the road has to be solved using a DRM and horizontal coordinates, speed and course measurements. Let use formulate this problem within the framework of the Markovian filtering theory. Assume that OXY is a rectangular coordinate system on the plane, and X, Y are the coordinates of the car moving along one of the possible road $T^h, h = \overline{1.M}$. For simplicity the car is

Paper presented at the RTO SCI International Conference on "Integrated Navigation Systems", held at "Elektropribor", St. Petersburg, Russia, 24-26 May 1999, and published in RTO MP-43.

assumed to be a point on the road and the width of the road – zero. In this case each of the roads can be described by an implicit, in a general case, nonlinear function $\chi^h(X,Y)$ in the form of

$$T^h = \left\{X,Y : \chi^h(X,Y) = 0\right\},\ h = \overline{1.M}\,. \qquad (1.1)$$

The functions $\chi^h(X,Y)$ can be represented, for example, as a set of the points satisfying (1.1) for each road. This information is stored in the memory of the car computer. Besides, it is assumed that using the car navigation equipment it is possible to measure the course (heading) K_i, speed V_i and coordinates of the car X_i, Y_i:

$$\widetilde{K}_i = K_i + \Delta K_i; \qquad (1.2)$$

$$\widetilde{V}_i = V_i + \Delta V_i; \qquad (1.3)$$

$$\left.\begin{aligned}\widetilde{X}_i &= X_i + \Delta X_i,\\ \widetilde{Y}_i &= Y_i + \Delta Y_i,\end{aligned}\right\} \qquad (1.4)$$

at discrete time i=1,2,...at intervals Δt. Here $\Delta K_i, \Delta V_i, \Delta X_i, \Delta Y_i$ are the measurement errors. For simplicity these errors are assumed to be described by a sum of the Markovian first-order processes x_{Z_i} and white noise $v_{\varepsilon i}$

$$\Delta Z_i = x_{Z_i} + v_{Z_i},\ Z = K, V, X, Y\,. \qquad (1.5)$$

The aim of the problem under consideration is to determine the road number on which the car is most probably located and the car's position on this road to the maximum accuracy. This problem has to be solved using all the map data and measuring information accumulated from the first up to the i-th instant of time. Let us formulate this problem within the framework of the Markovian nonlinear filtering theory [8,9].

The position of the car on the road is conveniently specified by the distance length l_i measured from a certain preset initial point with the coordinates X_0^h, Y_0^h. It is clear that for the current length the equality

$$l_i = l_{i-1} + \Delta t V_{i-1} \qquad (1.6)$$

will hold true.

It is appropriate to define each of the road in this problem by the coordinates X_0^h, Y_0^h and function $K^h(l)$, which describes the dependence of the course angle of the road on its length. In this case the coordinates of the road can be defined as the following functions of the length:

$$X^h(l) = \int_0^l \sin K^h(l)dl, \quad Y^h(l) = \int_0^l \cos K^h(l)dl\,.$$

Assume that the number of the road is a discrete random value (hypothesis) H whose a priori probability distribution density (hereinafter called simply density) is defined as

$$f(H) = \sum_{h=1}^{M} p_0^h \delta(H-h)\,, \qquad (1.7)$$

where $\delta(\cdot)$ is a delta-function; $p_0^h = \mathrm{P}_0(H=h)$ are a priori probabilities of the car being located on the road under number h. In order not to enter a stochastic model for the car's speed, the speed in (1.6) is substituted for the difference $\widetilde{V} - \Delta V$.

With regard to the assumptions made and the designations used the filtering problem can be formulated as follows: to identify (detect) the road number h and estimate the distance l_i which satisfies the equation

$$l_i = l_{i-1} + \Delta t V_{i-1} = l_{i-1} + \Delta t(\widetilde{V}_{i-1} - \Delta V_{i-1})\,, \quad (1.8)$$

using the measurements

$$\widetilde{X}_i = X^h(l_i) + \Delta X_i\,; \qquad (1.9)$$

$$\widetilde{Y}_i = Y^h(l_i) + \Delta Y_i\,; \qquad (1.10)$$

$$\widetilde{K}_i = K^h(l_i) + \Delta K_i\,,\ h = \overline{1.M}\,, \qquad (1.11)$$

accumulated up to the i-th instant of time. Here the functions $X^h(l_i)$, $Y^h(l_i)$, $K^h(l_i)$ are nonlinear in a general case and the errors are described by (1.5). It is clear that this problem is a joint, detection and estimation, problem.

Let us introduce the designations $\overline{\widetilde{K}}_i = (\widetilde{K}_1,...\widetilde{K}_i)^T$, $\overline{\widetilde{X}}_i = (\widetilde{X}_1,..\widetilde{X}_i)^T$, $\overline{\widetilde{Y}}_i = (\widetilde{Y}_1,..\widetilde{Y}_i)^T$. To derive an optimal solution for this problem within the framework of the Markovian filtering theory, it is necessary to know a posteriori density $f(H/\overline{\widetilde{X}}_i, \overline{\widetilde{Y}}_i, \overline{\widetilde{K}}_i)$ of the random value H (number of the road) and a posteriori densities $f(l_i/\overline{\widetilde{X}}_i, \overline{\widetilde{Y}}_i, \overline{\widetilde{K}}_i, H=h)$ of the distance which are determined for every possible number of the road $H=h$, $h=\overline{1.M}$. The density $f(H/\overline{\widetilde{X}}_i, \overline{\widetilde{Y}}_i, \overline{\widetilde{K}}_i)$ is described as

$$f(H/\overline{\widetilde{X}}_i, \overline{\widetilde{Y}}_i, \overline{\widetilde{K}}_i) = \sum_{h=1}^{M} p_a^h(i)\delta(H-h)\,, \quad (1.12)$$

where $p_a^h(i) = \mathrm{P}_a(H=h/\overline{\widetilde{X}}_i, \overline{\widetilde{Y}}_i, \overline{\widetilde{K}}_i)$ are a posteriori probabilities of the car being located on the road

numbered h. The densities $f(l_i / \tilde{\tilde{X}}_i, \tilde{\tilde{Y}}_i, \tilde{\tilde{K}}_i, H = h)$ are determined by the following recurrent relation [8]

$$f(l_i / \tilde{\tilde{X}}_i, \tilde{\tilde{Y}}_i, \tilde{\tilde{K}}_i, H = h) \propto$$
$$f(\tilde{X}_i, \tilde{Y}_i, \tilde{K}_i / l_i, \tilde{\tilde{X}}_{i-1}, \tilde{\tilde{Y}}_{i-1}, \tilde{\tilde{K}}_{i-1}, H = h) \times, \quad (1.13)$$
$$f(l_i / \tilde{\tilde{X}}_{i-1}, \tilde{\tilde{Y}}_{i-1}, \tilde{\tilde{K}}_{i-1}, H = h)$$

where $\propto$ means the proportionality.

Thus, within the suggested approach the solution of the problem reduces to determination of the number of the road *h,* for which the value $p_a^h(i)$ is maximum, i.e.,

$$h^* = \arg\max_h \; p_a^h(i), \quad (1.14)$$

and to estimation of the distance $\hat{l}_i^{h^*}$ for the chosen road

$$\hat{l}_i^{h^*} = \int l_i \; f(l_i / \tilde{\tilde{X}}_i, \tilde{\tilde{Y}}_i, \tilde{\tilde{K}}_i, H = h^*) \; dl_i . \quad (1.15)$$

The integration limits are assumed to be infinite here. The accuracy of this estimate $\hat{l}_i^{h^*}$ is characterized by the conditional covariance calculated as

$$\sigma^2_{\Delta l_i} = \int (l_i - \hat{l}_i^{h^*})^2 \; f(l_i / \tilde{\tilde{X}}_i, \tilde{\tilde{Y}}_i, \tilde{\tilde{K}}_i, H = h) \; dl_i . \quad (1.16)$$

It is clear that $p_a^{h^*}(i) = \mathrm{P}_a^*$ defines the maximum value of the probability that the car is on the road numbered h^*. Knowing the values of $\hat{l}_i^{h^*}$, it is possible to determine the coordinates of the car

$$\hat{X}_i = X^{h^*}(\hat{l}_i^{h^*}), \quad \hat{Y}_i = Y^{h^*}(\hat{l}_i^{h^*}) . \quad (1.17)$$

It is not difficult to understand that they will always lie on the road.

Note that the value $\mathbf{P}_\varepsilon = 1 - \mathbf{P}_a^*$ defines the probability of making a wrong decision thus the maximization of a posteriori probability is equivalent to the minimization of the probability to choose the wrong road. Hence the output of the map-matching algorithm (MMA) proposed for determining the car position provides: h^* - the number of the road; $\mathbf{P}_\varepsilon$ - the probability of making a wrong decision; $\hat{l}_i^{h^*}$ - the optimal estimate of the passed distance and the root mean square (RMS) error $\sigma_{\Delta l_i}$.

Note that the probabilities $p_a^h(i)$ can be also determined by using the recurrent equation [8].

$$p_a^h(i) \propto p_a^h(i-1) \times$$
$$\int [\, f(\tilde{K}_i, \tilde{X}_i, \tilde{Y}_i / H = h, l_i, \tilde{\tilde{K}}_{i-1}, \tilde{\tilde{X}}_{i-1}, \tilde{\tilde{Y}}_{i-1}) \times \quad (1.18)$$
$$\times f(l_i / \tilde{\tilde{K}}_{i-1}, \tilde{\tilde{X}}_{i-1}, \tilde{\tilde{Y}}_{i-1}, H = h) \,]\, dl_i,$$

$p_a^h(0) = p_0^h$.

From this equation it follows that the solution of the problem in this statement reduces, in fact, to solving the partial problems of estimating the length of the distance with fixed the road number, i.e. to determining the densities $f(l_i / \tilde{\tilde{X}}_i, \tilde{\tilde{Y}}_i, \tilde{\tilde{K}}_i, H = h)$, $h = \overline{1.M}$. Notice that in this subproblems the equations for the state vector are linear, and there is nonlinear dependence only for one component of the state vector in measurements – the length of the distance along the fixed road. These peculiarities substantially simplify the solution of the partial problems of nonlinear estimation and, consequently, of the whole problem of map-matching.

To conclude this part, it should be noted that the main advantage of the approach proposed for the solution of the problem considered is the possibility to choose the number of the road h^* on which the car is most probably located (the probability of making a wrong decision $\mathbf{P}_\varepsilon$ is minimum) and to estimate the car's position on this road $\hat{l}_i^{h^*}$ to the maximum accuracy within a unified statement.

2. The algorithm for solving the problem

Concretization of the general structure of the algorithm derived in the previous section is determined by two main conditions. First, it depends on the set of the measuring devices used as they determine the model employed in the description of their errors. Besides the algorithm largely depends on the method of approximation used to describe partial a posteriori densities $f(l_i / \tilde{\tilde{X}}_i, \tilde{\tilde{Y}}_i, \tilde{\tilde{K}}_i, H = h)$ which, in its turn, determines the algorithm for estimation of probabilities $p_a^h(i)$.

For illustration let us concretize the general solution for the simplest models of measurement errors (1.5) which presuppose the presence of only white noises. To what the algorithm is reduced can be shown by an example which presupposes that a car is moving along a fixed rectilinear road towards a crossroad with M different directions. The problem will be solved under the assumption that the a priori density of the car position error at the initial point of motion is gaussin, i.e. $f(l) = N(l; l_0, \sigma^2_{\Delta l})$. Fig.1 represents an example of a road, which corresponds to this variant of motion in the case when the turns at the crossroad are arcs.

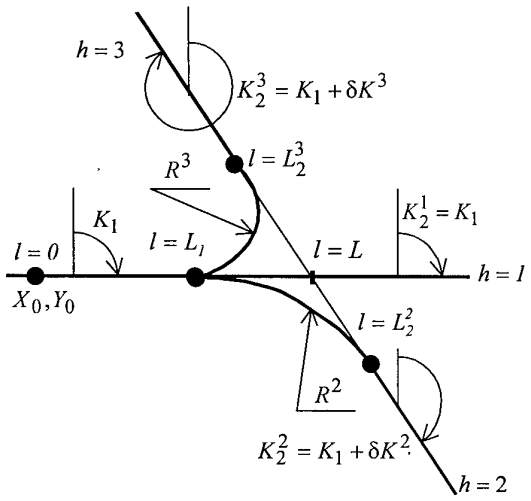


Fig.1. An example of the road with a crossroad

Here the following notation are used: L – a distance from the initial point of the road to the turn; L_1 – a distance from the initial point of the road to the initial point of the turn; L_2^h – a distance from the initial point of the road to the final point of the turn; K_1 – a course angle of the road before the turn; K_2^h – course angles of the road after the turn; R^h - a radius of turns; $h = \overline{1.M}$.

Let us also introduce the domain of the values l, for which the car is located on a leg of the route before the turn Ω_1, at the turn Ω_{12}^h and after the turn Ω_2^h, that is,

$$\left.\begin{aligned} &\Omega_1 = \{l : l < L_1\}; \\ &\Omega_{12}^h = \{l : L_1 \le l < L_2^h\}; \\ &\Omega_{at}^h = \{l : l \ge L_2^h\}. \end{aligned}\right\} \qquad (2.1)$$

It is clear that within each of these subdomains the functions $X^h(l), Y^h(l), K^h(l)$ will be described differently, namely,

$$X^h(l), Y^h(l), K^h(l) = \begin{cases} X_1(l), Y_1(l), K_1(l), \; l \in \Omega_1; \\ X_{12}^h(l), Y_{12}^h(l), K_{12}^h(l), \; l \in \Omega_{12}^h; \\ X_2^h(l), Y_2^h(l), K_2^h(l), \; l \in \Omega_2^h. \end{cases} \qquad (2.2)$$

Using the accepted notation it is not difficult to derive the equations for the functions $X^h(l), Y^h(l), K^h(l)$ which correspond to the legs of the route before the turn, at the turn and after the turn. It is clear that before and after the turn the coordinates will be linearly dependent on the length of the route, for example, up to the turn

$$X_1(l) = X_0 + l\sin K_1, \; Y_1(l) = Y_0 + l\cos K_1, \; K_1(l) = K_1. \qquad (2.3)$$

but at the turn the dependence on l will be linear for the course

$$K_{12}(l) = K_1 + sign(\delta K^h)\left(\frac{l - L_1}{R^h}\right), \qquad (2.4)$$

and nonlinear for the coordinates

$$X_{12}^h(l) = \overline{X}_1 + sign(\delta K^h) R^h \left(\cos K_1 - \cos K_{12}^h(l)\right); \qquad (2.5)$$

$$Y_{12}^h(l) = \overline{Y}_1 - sign(\delta K^h) R^h \left(\sin K_1 - \sin K_{12}^h(l)\right), \qquad (2.6)$$

where $\delta K^h = K_1 - K_2^h$, $R^h = \dfrac{L - L_1}{tg(\delta K^h)}$.

When deriving the algorithm for the solution of the problem it is convenient to consider three stages of its operation corresponding to the sequential legs - before the turn, at the turn and after the turn.

At the first stage, basing on the assumption that the car is located on the rectilinear leg of the road Ω_1 and taking into account the fact that the functions given by (2.3) are linear, the optimal estimate of the distance length can be derived by using Kalman filter (KF). The KF will be generating the estimate of the distance $\bar{l}_i$ and the corresponding $\sigma_{\Delta l_i}^2$ at each instant of time. These parameters define a posteriori density of the distance length, as this density, in this case, is gaussian.

Assume that starting from a certain instant of time i^+ the condition

$$\bar{l}_i + k\sigma_{\Delta l_i} > L_1, \quad i \ge i^+, \qquad (2.7)$$

with $k \ge 3$, holds true. This condition means that the car is coming to the domain of a turn. Then there arises an uncertainty about the number of the road along which the car is moving and nonlinearity of the functions $X^h(l), Y^h(l), K^h(l)$ described by (2.4)-(2.6). Let us consider to what the optimal algorithm reduces at this second stage.

The parameters of a posteriori density $\sigma_{\Delta l^{i^+}}, \bar{l}_{i^+}$ generated by the KF at the time i^+ for the value $l^+ = l_{i^+}$ can be treated as the parameters of the a priori density $f(l^+) = N\left(l^+; \bar{l}^+, \sigma_{\Delta l^+}^2\right)$ for the second stage. Assume that the interval of the time during which the car is located on the turn is short, thus over this interval the effect of the speed errors v_{V_i} can be neglected. In this situation it is possible to write

$$l_i = l^+ + \Delta t \sum_{\mu=1}^{i-1} \widetilde{V}_\mu . \qquad (2.8)$$

As the second summand in (2.8) is known, the problem under consideration can be reduced to the problem of

estimation of the constant value l^+, that is, to finding the densities $f(l^+ / \overline{\overline{\tilde{K}}}_i, \overline{\overline{\tilde{X}}}_i, \overline{\overline{\tilde{Y}}}_i, H = h)$. A recurrent equation of the type (1.13) can be used in finding these densities with due account of the fact that in the given case

$$f(\tilde{K}_i, \tilde{X}_i, \tilde{Y}_i / H = h, l^+, \overline{\overline{\tilde{K}}}_{i-1}, \overline{\overline{\tilde{X}}}_{i-1}, \overline{\overline{\tilde{Y}}}_{i-1}) = f(\tilde{K}_i, \tilde{X}_i, \tilde{Y}_i / H = h, l^+) \propto \quad (2.9)$$

$$\propto \exp\left\{-\frac{1}{2}\left[\frac{\delta K_i^h(l^+)}{\sigma_{\Delta K}^2} + \frac{\delta X_i^h(l^+)}{\sigma_\Delta^2} + \frac{\delta Y_i^h(l^+)}{\sigma_\Delta^2}\right]\right\},$$

where $\delta Z_i^h(l^+) = \tilde{Z}_i - Z^h(l^+ + \Delta t \sum_{\mu=1}^{i-1} \tilde{V})$, $Z = K, X, Y$,

$\sigma_{\Delta K}, \sigma_\Delta$ - RMS errors of the course and coordinates.

Hence the algorithm for the solution of the problem reduces to the following: calculation of $p_a^h(i)$ for all $h = \overline{1.M}$ using (1.13), (1.18), (2.9); choice of the road h^* corresponding to the maximum value $p_a^h(i)$; calculation of the optimal value $\hat{l}_{i^+}^{h^*}$ and the corresponding conditional covariance $\sigma_{\Delta l^+}^2$ using (1.15), (1.16); calculation of the optimal length estimate for the current time

$$\hat{l}_i^{h^*} = \hat{l}_{i^+}^{h^*} + \Delta t \sum_{\mu=1}^{i-1} \tilde{V}_\mu, \quad (2.10)$$

It is evident that under the assumptions made the covariance of this error $\sigma_{\Delta l_i}^2$ is the same as $\sigma_{\Delta l^+}^2$.

In solving the problem it is reasonable to preset a certain level close to unity P_∂ which should be exceeded in order to make a decision that the car is moving along the road numbered h^*, that is, to demand that the condition

$$p_a^{h^*}(i) \geq P_\partial \approx 1 \quad (2.11)$$

should be fulfilled. Attaining this level is necessary for obtaining a "good" solution. The time needed for that is obviously a rather important performance of the solution in the problem under consideration.

In spite of the fact that the functions entering into the integrands (1.13), (1.16) have been determined, the calculation of the corresponding integrals is a very difficult problem as they cannot be determined analytically because the functions $K^h(\cdot), X^h(\cdot), Y^h(\cdot)$ are nonlinear. Special methods of approximation of a posteriori density allowing economical calculation procedures are developed in the theory of nonlinear filtering for calculation of these integrals. Further concretization of the algorithm depends on the method used for the approximation of a posteriori density. For the problem under consideration it is convenient to use approximation of a posteriori density with the use of a set of delta-functions. Such approximation generates the algorithm for calculation of the optimal estimate and the corresponding covariance is known as the method of nets [8]. This method is easy to realize for the problem considered.

It is reasonable to complete the operation of the algorithm at this stage when the following condition is satisfied:

$$\hat{l}_i^{h^*} - k\sigma_{\Delta l_i} > L_2^{h^*}. \quad (2.12)$$

It means that the car has come to another rectilinear leg of the road numbered h^*.

It is obvious that with the use of the values $\hat{l}_i^{h^*}, \sigma_{\Delta l_i}^2$ as initial, the KF whose linear measurements are defined by the equation of the type (2.3) with $h = h^*$ can be used as an algorithm at the third stage.

So, on the whole, the optimal algorithm for the solution of the problem considered reduces to the successive use of the KF which corresponds to Equation (2.3), a nonlinear block realized with the use of the method of nets when the car is at the turn, and the KF which corresponds to the linear measurements with $h = h^*$ after the crossroad. It is essential that the algorithm itself determines the procedure for estimation at the current time.

In conclusion it should be noted that the algorithm will also be of a similar structure for more complicated models used in describing measurement errors, as well as for the case when the car is moving in the area with a few, in particular, parallel streets.

3. Potential accuracy analysis

Using the approach suggested let us analyze the potential accuracy for the problem under consideration. Here it is advisable to consider two different cases. One of them is characterized by the uncertainty about the number of the road along which the car is moving. The time $t_{\min}$ needed for attaining the preset (close to unity) level P_∂ for the value of the probability that the car is on the road numbered h^* is obviously a rather important performance of the algorithm used. This time corresponding to the optimal solution will be used as the quantitative characteristic of the potential performance in this situation. When the road along which the car is moving is known it is obvious that the RMS value $\sigma_{\Delta l_i}$ determined by (1.16) can be used as a characteristic of the potential performance. Let us analyze the main effects that are achieved due to the use of the map data.

This is advisable to assume that the road number along which car is moving is known. This assumption is used here.

It is supposed that the coordinates, speed and course measurements are used from the satellite navigation systems (SNS) or from dead reckoning system using the information from the odometer and the vertical gyro. The model of the measurement errors and their characteristics are represented in Table 3.1

Table3.1

RMS errors of the various measurements

Error components	SNS coordinates	SNS speed	Odometer	Gyro
White noise Δt =1сек	20m	0.3 m/s	-	$3.6\cdot10^2$ °/h
First order Markovian process (τ_{cor})	30m (10min)	-	$7\cdot10^{-2}$m/s (5min)	$3.6\cdot10^2$ °/h (5min)
Random bias	-	-	-	$3.6\cdot10^3$ °/h
Scale factor	-	-	0.03	-

It is essential that the RMS course error for SNS depends on the speed and is determined as

$$\sigma_{\Delta K} = \sigma_{\Delta V} / V, \qquad (3.1)$$

where $\sigma_{\Delta V}$ - the RMS speed error.

First, it is interesting to calculate the errors in cross-track and along-track position when the data from a map is not used. If the data from SNS is used, these errors are equal (RMS=35m.) If the car position is determined by dead reckoning using speed and course measurements (when the data from SNS is not available), these errors are determined by the following equations

$$\Delta\dot{l} = \Delta V_{od}, \ \Delta\dot{s} = -V_{od}\Delta K_g, \ \Delta\dot{K} = \Delta\omega_z, \quad (3.2)$$

where Δl and Δs - the along-track and cross-track position errors, respectively; $\Delta V_{od}, \Delta K_g$ - the speed and course errors; $\Delta\omega_z$ - gyro drift.

RMS values for Δl and Δs calculated by these equations are represented in Table 3.2. The data presented in the table is given for V=50 km/h. It is essential that the cross-track position error increases very quickly, it achieves 500m within 5 minutes. At the same time the along-track position error increases slowly and achieves only 27m.

Now suppose that the map data is used and the car is moving along a straight road. It is evident that in this case there is no error in the cross-track position. This fact results from the assumption that the width of the road is zero. It should be noted that the RMS error for Δl is the same for the cases when the map data is used and is not used.

Table3.2

RMS errors for dead-reckoning [m]

Direction	Time, min No map data/With map data			
	0	1	3	5
Cross-track	23/0	45/0	120/0	484/0
Along-track	23/23	24/24	25/25	27/27

Let us analyze the possibility of increasing the length estimation accuracy when the course of the road is being changed. For simplicity the speed is supposed to be known and coordinates errors and course errors have only white noise components. In this case it is possible to obtain (using the Rao-Cramer inequality) the analytical equations for the lower bounds of the RMS errors of the length estimation [8,10]. For the course and coordinates measurements these equations, respectively, will be as follows

$$(\sigma_{\Delta l_i}^{cour})^2 = (R\sigma_{\Delta K})^2 / i, \qquad (3.3)$$

$$(\sigma_{\Delta l_i}^{coord})^2 = \frac{\sigma_\Delta^2}{\sum_{\mu=1}^{i}\left(\left(\frac{dX(l_\mu)}{dl_\mu}\right)^2 + \left(\frac{dY(l_\mu)}{dl_\mu}\right)^2\right)} = \frac{\sigma_\Delta^2}{\sum_{\mu=1}^{i}\left(\sin^2 K(l_\mu) + \cos^2 K(l_\mu)\right)} = \frac{\sigma_\Delta^2}{i}, \qquad (3.4)$$

where σ_Δ - the RMS error of the coordinates; $l_\mu = l_i - \Delta t(V_{i-1} + \ldots + V_\mu)$ - the length for the μ-th instant of time. The estimation accuracy of the length for course measurements is proportional to the radius of turns. For coordinate measurements this accuracy (for the assumptions made) does not depend on the changes in the road direction, i.e. on the turns it is the same as on a straight road. As the time over which the car is passing the turn is small, it is clear that it is possible not to use the measurements of coordinates on the turn without essential decrease in accuracy. This conclusion is useful, since the processing of coordinate measurements, because of their nonlinear character, involves considerable computational expenses. It is important also

to emphasize that the course errors and R for the real turns are, as a rule, such that the length estimation accuracy on the turn even for one course measurement is much higher than the accuracy provided for several measurements of coordinates.

Now note the peculiarity of course measurements from the SNS. Taking into account that the number of measurements on the turn is determined as $i = R\delta K/(V\Delta t)$ and the RMS error of the course is determined by (3.1), it is not difficult to obtain

$$\sigma_{\Delta l_i}^{cour} = \sigma_{\Delta V}\sqrt{\frac{R\Delta t}{V\delta K}}. \qquad (3.5)$$

So, the accuracy of the length estimation on the turn depends on two factors: the angular rate of the car (V/R) and the angle of the turn.

If slowly varying components of errors are dominant, it is possible to show that the increase of the length estimation accuracy will be equally effective both for the course and coordinate measurements. It is evident that on straight roads the measurements of the coordinates and the course do not effect the length estimation accuracy. Nevertheless, their usage on such legs of the road is advisable, as then there is a possibility to improve slowly varying components of the errors.

The RMS length errors corresponding to the algorithm developed (simulation results) are given in Table 3.3 for the different angle turns.

An example of a posteriori density for the length is shown on Fig. 2 for various values of the course measurements. It is evident that this a posteriori density is non-gaussian.

Table.3.3

RMS length estimation errors after the turn [m]

SNS Only/Gyro+odometer

V, m/s / δK, °	5	10	15
30	1.6/0.8	1.0/0.6	0.7/0.4
60	1.2/0.6	0.7/0.5	0.5/0.3
90	0.9/0.5	0.5/0.4	0.3/0.3

Data of this table confirms the conclusions obtained above about the length estimation accuracy on the turn.

As for the time which is required for identification of the true road on the crossroads, it should be noted that it depends on the angle between roads and on the car speed. When only SNS measurements are used on crossroads, this time, as the simulation results have shown, is equal to (2-6) s for $P_\partial = 0.95$.

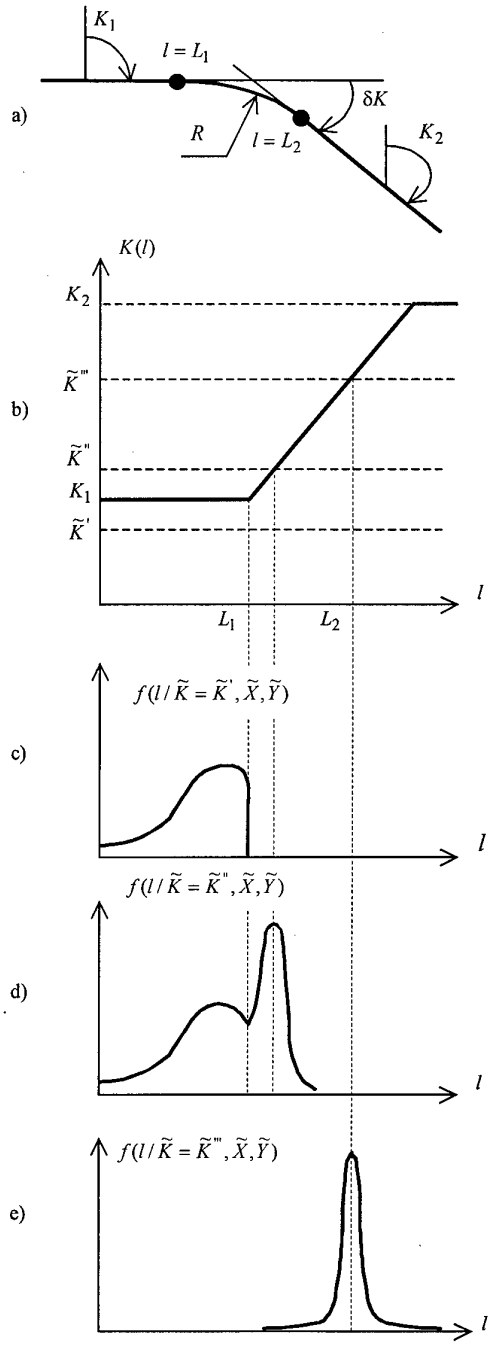


Fig2. A posteriori density for course measurements.

2a –the road; 2b – the function $K^h(l)$;

2c, 2d, 2e –a posteriori density for different course measurements $\tilde{K}', \tilde{K}'', \tilde{K}'''$.

4. The results of the field test using a real map and SNS data

The algorithm developed was checked using only real satellite (without dead-reckoning) measurements and map data. These measurements were accumulated during the car runs in one of St.Petersburg districts. The car

track included road turns and crossroads. The problem was solved in off-line operations using the real satellite measurements and road maps for this district. The map data was presented by piecewise-linear approximation for the roads on the map, the points of turns on the roads and crossroads ("node points") defined by the geographical coordinates $\varphi_i, \lambda_i, \quad i = \overline{0,n}$.

Fig. 3 depicts a part of the map data with the following numbers of the node points:

- node N1 area with a road turn;
- crossroad (node N2) with a small change in movement direction;

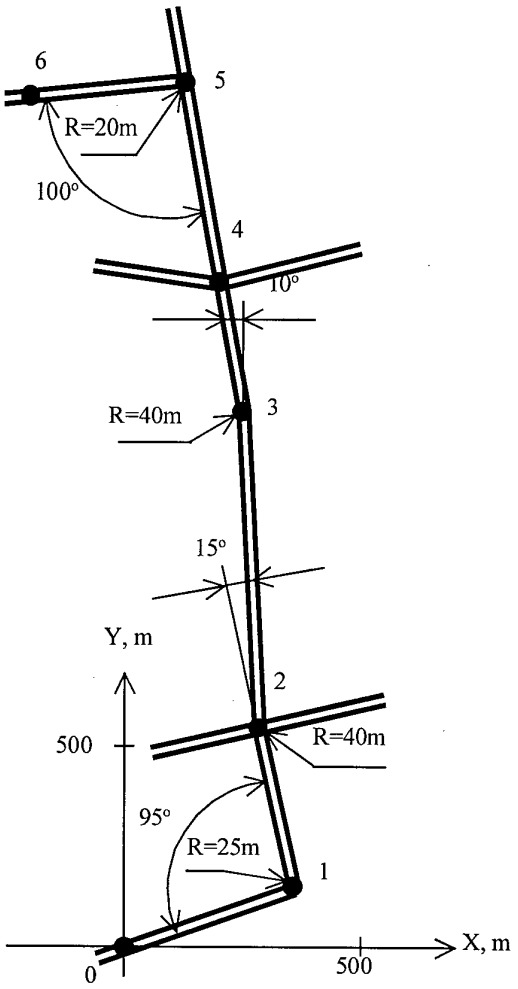


Fig.3. Part of map data.

- node N3 area with a small road turn;
- crossroad (node N4) without a change in the movement direction;
- crossroad (node N5) with a turn to a transverse street.

Fig 4 depicts SNS and MMA coordinates obtained by simulation for one of the roads (node N1 area) used in the field test. All in all 5 similar runs were performed All the data was used to verify the "repeatability"of SNS measurements. The "behaviour" of the course error was analyzed for the stops and slow speed run. On the basis of this investigation two assumptions for the map-matching algorithm were made:

- $\widetilde{V} = 0$ for $\widetilde{V} < 1$m/sec;
- if $\widetilde{V} < 1$m/sec from the t_i, then $\widetilde{K}(t_i) = \widetilde{K}(t_{i-1})$.

Different runs revealed no peculiarities in SNS measurements. One of the runs was chosen for evaluation of the MMA operation and efficiency. The instants of time fixed by the operator when the car is moving through the nodes are used in evaluation of MMA accuracy. For these instants the MMA coordinates and the coordinates of the nodes obtained from the map are compared and, as a result, the MMA error represents a distance between these two points. These errors are given in the Table 4.1. It should be noted that the resulting errors include the operator's mistakes in fixing the instants of the car's going through the nodes. These errors may be as much as 10m.

Table 4.1

Test results of MMA

Node number (car speed)	MMA error, m	The time needed to make a decision, s
1 (10m/s)	13,3	-
2 (4m/s)	18.6	3
3 (14m/s)	33.8	-
4 (3m/s)	39,3	4
5 (5m/s)	8,1	3
6	3,6	-

The large errors in node N2-N4 may be explained by the absence of a turn for N4 and the small turn angles for N2, N3. It should be noted that these errors correspond to the covariance $\sigma_{\Delta l_i}$ determined by (1.16).

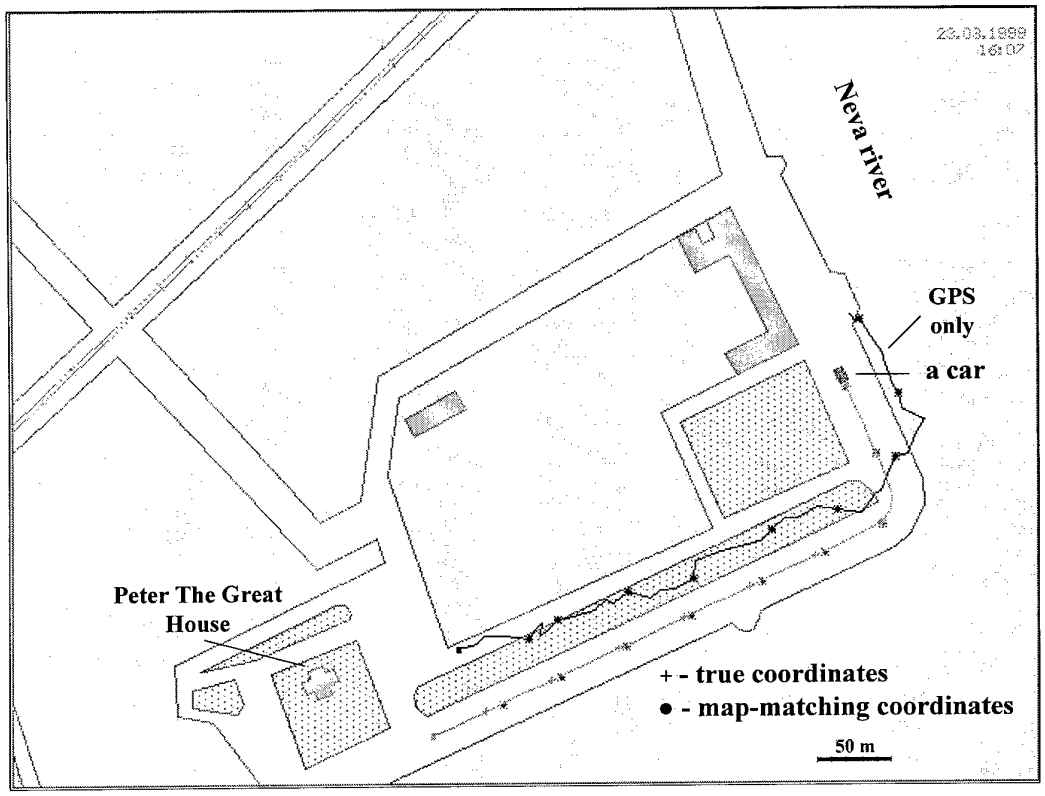


Fig 4. An example of the car navigation using optimal map-matching

References:

1. **French R.L.** Land Vehicle Navigation and Tracking. Global Positioning System: Theory and Applications Volume II, p.275-301. 1996. (Edited by Parkinson B.W.)
2. **Scott C.A.** Improved GPS Positioning for Motor Vehicles through Map Matching. ION GPS 94.
3. **Kim W., Jee G., Lee J.G.** Improved Car Navigation System Using Path-Associated Map-Matching. 9th World Congress of the International Association of Institutes Navigation. 18-21 November 1997. Amsterdam, Netherlands.
4. **Prendergast S.** Mapping and Navigation for the Automotive Cockpit. Navigation News. Sept.Oct. 1997.
5. **Krasovsky A. A., Beloglazov I.N., Chigin G.P.** The Theory of the Extremal Navigation System. M. Nauka, 1979, 448p. (in Russian)
6. **Kayton M., Fried W.R.** Avionics navigation systems. Second edition. New York, 1997, 773p.
7. **Lowrey J. A. III, Shellenbarger J.C.** Passive Navigation using Inertial Navigational Sensors and Maps. Naval Engineering Journal, May, 1997.
8. **Stepanov O. A.** Nonlinear filtering and its application in navigation. CSRI RF Elektropribor. St. Petersburg, 1998. 370p, in Russian.
9. **Dmitriev S. P., Stepanov O. A.** Nonlinear Filtering And Navigation. Proceedings of 5th Saint Petersburg International Conference on Integrated Navigation Systems. Pp.138-149. Saint Petersburg 1998. State Research Center of Russia Elektropribor.
10. **Stepanov O.A.** Estimation Methods of the Potential Accuracy for Extremal Navigation System. Saint Petersburg, State Research Center of Russia Elektropribor 1993, 84 p. (in Russian).

Reliable Autonomous Precise Integrated Navigation RAPIN for Present and Future Air-Vehicles

Dr Thomas Köhler - DaimlerChrysler Aerospace AG Military Aircraft (Ottobrunn)
P.O. Box 801160, D-81663 München, Germany
Tel.: ++49-89-607-24635 Fax: ++49-89-607-32074
E-mail: Thomas.Koehler@m.dasa.de

Franz Turnbrägel - DaimlerChrysler Aerospace AG Military Aircraft (Bremen)

W. Lohmiller, Ph.D. MIT
Dr. Beyer DaimlerChrysler Aerospace AG Military Aircraft (Ottobrunn)

1 Abstract

RAPIN is an integrated navigation system designed for safety and precision in onboard autonomous flight management close to the ground. A digital terrain database is used as an additional information source. Integrating 3 physically independent sensors with complementary behavior LINS, GPS and TRN allows RAPIN to isolate and remove sensor errors. A navigation demonstrator was realized in the RAPIN research and technology project and tested in the Transall C160 test carrier. The flight trials verified functionality and performance under realistic deployment conditions.

Evaluation of the flight trials shows that RAPIN is in conformity with the requirements for tactical low-altitude flight management even under conditions of GPS failure. It was also demonstrated that landings could be performed.

2 Introduction

Advances in processing data and navigation sensors today permit the development of substantially more powerful navigation systems than those currently being deployed. The gradual transfer to the Future Aircraft Navigation System (FANS) will lead to a radical change in civil and military navigation and the fact that some of the navigation sensors being deployed today have been in service for many years supports the hypothesis that there will be widespread demand for a new navigation system that makes full use of the technical possibilities available.

RAPIN is part of a large-scale investigation in the area of low-altitude flight management that was started at Dasa at the beginning of the 1990s with different applications in mind. RAPIN thus combines all development activities in navigation.

The main aim of RAPIN is to develop a reliable, autonomous, precise and integrated navigation system that integrates the P(Y)-GPS, LINS and TRN navigation sensors as a demonstrator. In this context autonomous means that this system can also perform low-altitude and landing operations if GPS reception is not operating.

The aim is to achieve the following objectives:

- Realization of a reliable, precise navigation system as a demonstrator that allows low-altitude flights and landings to be performed without the availability of GPS navigation. The key goal is adequate sensor integration and optimum filtering of input parameters.

- Verification of function and performance in laboratory and flight trials. This includes clarifying the issue of whether the accuracy of GPS P(Y)-Code and/or TRN is sufficient for a military precision landing or whether a D-GPS system is necessary.

- Developing system know-how, particularly in the areas of sensor integration and filtering.

- Developing specification and verification capability.

Considerations of cost restrict the RAPIN project in the first instance to integration of the most important sensors in order to develop basic know-how. This system represents the core of an integrated navigation system that can be expanded with other sensors in future developments. This is particularly relevant to stronger integration with the sensor systems of flight control systems (air data, positional, rotation rate and acceleration information) but also with available reconnaissance and flight-

Paper presented at the RTO SCI International Conference on "Integrated Navigation Systems", held at "Elektropribor", St. Petersburg, Russia, 24-26 May 1999, and published in RTO MP-43.

management sensors (e.g. FLIR or SAR sensors).

The requirement for a navigation system such as RAPIN was identified by the Bundesamt für Wehrtechnik und Beschaffung (BWB, German Ministry for Military Engineering and Procurement). A research and development contract was granted to DaimlerChrysler Aerospace AG with a subcontract to Honeywell Regelsysteme GmbH. The objective and content of the commission is the realization of the navigation system as a demonstrator and performance of flight test trials.

System concept, design, realization, integration and testing of the complete system were carried out by Dasa.

The Wehrtechnische Dienststelle (Military Engineering Office) WTD61 in Manching provided a flight test carrier including the necessary instrumentation, a reference through radar monitoring system and a data recording system for testing the technology.

3 RAPIN system description

This section provides a basic description of the RAPIN system and concept.

3.1 RAPIN Concept

The RAPIN concept is based on the integration and optimum evaluation of navigation instruments that are typically found onboard a military aircraft:

- A Laser Inertial Navigation System (LINS)
- A Global Positioning System (GPS) receiver that operates with the military high-precision Precise Positioning Service (PPS) or P(Y)-Code.
- A Terrain Reference Navigation System (TRN) that compares signals from the Radar Altimeter (R/A) and barometric height signals from the Air Data Computer (ADC) with the stored terrain model to determine and altitude of the aircraft.

These three basic RAPIN measuring instruments have the following mutually complementary advantages that are combined when the sensors are integrated (see Table 1):

- The LINS has a high level of availability and high short-term accuracy. However, over longer periods of deployment the position is subject to drift.
- The GPS P/Y-Code provides extremely high horizontal positioning accuracy. System availability and continuity is restricted here by the possibility of noise interfering with the reception signal (intentional or random)
- TRN provides good positioning accuracy and signal availability.

RAPIN sensor integration permits recognition of sensor errors with a high level of reliability and maintains adequate navigation accuracy even in the event of a failure. This is achieved technically using a Kalman filter bank.

3.2 Integrated system

The RAPIN integrated system is shown in Figure 1. The individual boxes represent the central sensors:

- Laser Inertial Navigation System (LINS),
- Air Data Computer (ADC),
- Radar Altimeter (R/A),
- Terrain Referenced Navigation (TRN) that correlates the radar altitude with the Terrain Database (TBD),
- and Global Positioning System GPS (in the precise military mode PPS or P(Y)-Code).

The instability of LINS in the vertical channel means that the pressure altitude of the Air Data Computer (ADC) must be available. TRN requires a radar altimeter R/A in order to measure the altitude profile with respect to the terrain. This section therefore only deals with the LINS in a LINS/ADC combination and TRN always only in connection with R/A.

3.3 Terrain Referenced Navigation based on LATAN

The LATAN System (Low Altitude Terrain Avoidance and Navigation) constitutes the technological basis for Terrain Referenced Navigation (TRN) in RAPIN. LATAN is a navigation and low-altitude guidance system supported by terrain data. It was developed by DaimlerChrysler AG (formerly MBB) between 1986 and 1993 and successfully tested in the TORNADO. LATAN navigation functions were used in RAPIN, modified to be compatible with the sensors and incorporated in the integrated system.

3.4 Terrain data

The RAPIN terrain database was implemented in the RAPIN computer for supplying terrain data in the RAPIN system.

The terrain database software in the RAPIN computer first uploads the data tiles into the onboard database on the hard drive of the RAPIN computer. This database comprises the actual terrain data and a table with coordinate information and storage addresses of the terrain data.

The real-time component of the RAPIN terrain database comprises three parts (processes):

- Part 1: output of terrain data from the terrain data storage and loading a data window
- Part 2: supply of terrain data for TRN navigation horizontal (NAVHOR)
- Part 3: supply of terrain data for TRN navigation vertical (NAVVER)

Parts 2 and 3 of the terrain database use the data supplied from part 1.

When the terrain data is supplied for the TRN functions, the scanning behavior of the radar altimeter is taken into account. For example, 75% of the entered object height for trees is included in the output altitude values. The scanning behavior of the radar altimeter was determined on the basis of evaluations of the flight trial data and comparisons with the stored terrain data.

An altitude dependent surface evaluation (footprint) is also performed for vertical navigation and the minimum distance between radar altimeter and 'illuminated' ground surface is calculated. The extension of the evaluated ground surface was also determined by evaluation of the flight-trial data and comparison with the stored terrain data. The data determined in this manner are dependent on the type of radar altimeter used and must be adjusted when changing to a different model.

3.5 Terrain-dependent TRN performance

The performance of horizontal TRN navigation is particularly dependent on the significance of the traversed terrain.
Figure 2 shows an evaluation of terrain significance in Germany and the adjacent territories. The significance shown in red makes an excellent TRN navigation possible.

The significance shown as green and yellow in Figure 2 allows medium horizontal navigation at low flight altitude.

The terrain shown in blue is flat. Horizontal TRN navigation is not possible here because the noise amplitude in the measuring signal of the radar altimeter is greater than the amplitude of the terrain.

These restrictions are not applicable to vertical TRN navigation.

3.6 Kalman filter bank

All sensor signals are selected and combined with the Kalman filter bank in Figure 3. A total of 4 independent Kalman filters LINS/GPS/TRN, LINS/GPS, LINS/TRN and TRN/GPS are available. Maximum navigation accuracy can always be attained with the LINS/GPS/TRN filter. The advantage of a Kalman filter bank is that when LINS or GPS drift is recognized at a late stage reference can always be made to a non-contaminated Kalman solution. High signal redundancy is also useful for error recognition and isolation.

Error recognition identifies which sensors have failed. Redundancy management then switches to the Kalman filter without defective measurement. A Kalman filter bank has the advantage that even when errors are recognized too late it is still possible to switch to a correct Kalman filter solution.

Every RAPIN filter uses the same expanded Kalman filter module. The sensors of each Kalman filter are addressed using an index vector. This reduces the computing time and permits simple access to the various Kalman filter modules.

3.7 Kalman filter

The conventional navigation Kalman filter is used in RAPIN for the main filter, the LINS/GPS filter and the LINS/TRN filter. Measured Inertial Measurement Unit (IMU) accelerations and rotation rates $\mathbf{u}$ are integrated in the Strap Down Algorithm (SDA). Additional measuring values $\mathbf{m}$ are also required in order to correct the SDA navigation estimate with Δx navigation state.

- In a closed loop approach the navigation state differences are input directly into LINS.
- In an open loop approach the navigation state differences are calculated separately in the Kalman filter.

An open loop approach is used in RAPIN for reasons of air safety. LINS measurements are regarded as control input u_k, while GPS and TRN measurements at time k are included in the measuring vector TR_k.

The first nine states are calculated in advance in the Strap Down Algorithm as 9-dimensional $\hat{\mathbf{x}}_{kSDA}$ at time k. A state difference vector $\Delta\hat{\mathbf{x}}_k = \hat{\mathbf{x}}_k - \begin{pmatrix} \hat{\mathbf{x}}_{kSDA} \\ \mathbf{0} \end{pmatrix}$ is then propagated with $\Delta\hat{\mathbf{x}}_{k+1} = \mathbf{F}_k \Delta\hat{\mathbf{x}}_k$ where $\mathbf{F}_k(\hat{\mathbf{x}}_k)$ represents the discrete system matrix. This leads to the propagated state vector $\hat{\mathbf{x}}_k = \Delta\hat{\mathbf{x}}_k + \begin{pmatrix} \hat{\mathbf{x}}_k \\ \mathbf{0} \end{pmatrix}$.

The Pk state covariance matrix is then renewed with the system noise covariance matrix **Q**k and $\mathbf{P}_{k+1} = \mathbf{F}_k\mathbf{P}_k\mathbf{F}_k^T + \mathbf{Q}_k$. If there is a measurement, the residue covariance matrix is first determined with the **R**k measurement noise matrix $\mathbf{P}_k = \mathbf{H}_k\mathbf{P}_k\mathbf{H}_k^T + \mathbf{R}_k$. The Kalman feedback matrix then evaluates to $\mathbf{V}_k = \mathbf{P}_k\mathbf{H}_k\mathbf{P}_k^{-1}$. With the estimated measurements $\hat{\mathbf{m}}_k(\hat{\mathbf{x}}_k)$ this leads to the state difference correction $\Delta\mathbf{x}_{k+1} = \Delta\mathbf{x}_k + \mathbf{K}_k\mathbf{e}_k$ to a residue $\mathbf{e}_k = \mathbf{m}_k - \mathbf{m}, \hat{}_k$ $(\mathbf{x}, \hat{}_k)$. The state covariance is then renewed with $\mathbf{P}_k = \mathbf{P}_k - \mathbf{V}_k\mathbf{H}_k\mathbf{P}_k$ and the renewed state calculated

$$\hat{\mathbf{x}}_k = \Delta\hat{\mathbf{x}}_k + \begin{pmatrix} \hat{\mathbf{x}}_k \\ \mathbf{0} \end{pmatrix}.$$

3.8 Failure detection and isolation

Failure detection is performed in every Kalman filter iteration in the following stages:

- **Hardware tests:**
  The built-in test and data transfer test information of the LINS and support sensors are checked for recognized errors.

- **Raw data tests:**
  The individual LINS / support sensor raw data **x**kSDA/**m**k at time k are checked in a range test to ascertain whether they are in a technically possible range **m**mink < **m**k < **m**maxk or **x**SDAkmin < **x**SDAk < **x**SDAkmax. A gradian test then checks whether the time change $\dot{\mathbf{m}}_{,\min k}$ dt < **m**k – **m**k-1 < $\dot{\mathbf{m}}_{,\max k}$ dt or $\dot{\mathbf{x}}_{k\min}$ dt < **x**SDAk – **x**SDAk-1 < $\dot{\mathbf{x}}_{SDAk\max}$ dt is within the technical limits in a time differential dt.

- **Outlier tests:**
  An outlier test is performed after each Kalman filter update. The Kalman update **x**,$\hat{}_k$ (+) is determined first. Proceeding from this statistical mean value an outlier test of update and propagation is performed. In particular, updates are checked with

$$\chi_k = \left| \frac{m_k - m_k(\hat{x}_k(+))}{\sqrt{R_k^*}} \right|,$$

$$\mathbf{R}_k^* = (\mathbf{I} - \mathbf{H}_\kappa\mathbf{K}_\kappa)\mathbf{P}_{ek}(\mathbf{I} - \mathbf{H}_\kappa\mathbf{K}_\kappa)^T$$

where χk has to lie below some given threshold. Propagations are checked with

$$\chi_k = \left| \frac{\hat{x}_k(-) - \hat{x}_k(+)}{\sqrt{P_k^*}} \right|,$$

$$\mathbf{P}_k^* = \mathbf{K}_\kappa\mathbf{P}_{ek}\mathbf{K}_\kappa^T$$

where (-) refers to the propagated state estimate. Repeating outliers in a time interval τ are recognized using the sliding mean

$$\bar{\chi}_{k+1} = \bar{\chi}_k + \frac{1}{\tau}(\bar{\chi}_k - \chi_k)$$

- **State tests:**
  Following successful update all xk Kalman filter states are checked in a range test to ascertain whether they are in a technical possible range **x**kmin < **x**k < **x**kmax. The gradian test then checks whether the state change $\dot{\mathbf{x}}_{k\min}$ dt < **x**k – **x**k-1 < $\dot{\mathbf{x}}_{k\max}$ dt lies within the technical limits in a time differential dt.

It should be noted that errors recognized here can be directly attributed to individual sensors.

3.9 Redundancy management

LINS or support sensor errors are treated as follows:

- **Support sensor management:**
  If a hardware or raw data error occurs when a support sensor is measuring, the measurement is immediately rejected.
  If an outlier exceeds a previously established integrity limit the measurement is rejected by repeating the update without measurement.
  If a repeated outlier or a state error occurs in a support sensor bias the relevant support sensor is rejected for the remainder of the entire flight, the system switches to an appropriate subfilter and a maintenance message is given.

- **LINS management:**
  If a hardware, raw data, outlier, state error occurs for LINS, the LINS is rejected for the remainder of the entire flight, the system switches to the appropriate subfilter and a maintenance message is given.

4 RAPIN test carrier TRANSALL C160

WTD 61 / Manching provided the Transall C160 ANA/FRA 50+90 for the RAPIN trials. The aim was to use as much of the available equipment as possible for the tests. The aircraft already had a P/Y-GPS receiver and a Laser Intertial Navigation System. These and other components were linked up by a MIL-Bus. It was therefore possible to communicate with the RAPIN equipment via this standard bus. The "Dasa measurement cell" also provided the capability for displaying data via the MIL-Bus. Additional special RAPIN equipment was integrated via serial lines such as RS 232, Ethernet or ARINC 429 interfaces.

The aircraft was initially equipped with the following GPS receiver (Collins), ADC, BSIU, CDU, LINS (Honeywell) and Dasa measuring cell.

Other components also had to be installed for RAPIN, such as

- RAPIN computer (Dasa, hardware: mfr. Harris/Concurrent)
- RAPIN operator panel (Dasa, hardware: mfr. Texas Instruments)
- Radar altimeter (Dasa-Ulm)
- Differential GPS (Dasa, hardware: mfr. Ashtec, Fugro, etc.) incl. antennas
- Measuring table, chairs

Figure 4 shows a diagram of the test carrier with the components required for RAPIN.

5 Flight tests

Two flight campaigns with a total of 18 flights and approx. 38 flight hours were carried out on the TRANSALL ANA/FRA 50+90 test carrier in order to test the RAPIN system.

Figure 5 – 7 illustrate exemplarically landing approaches of the LINS/GPS/TRN and the LINS/TRN filter. The horizontal accuracy of the LINS/GPS/TRN lies within the landing threshold. The height channel of the LINS/TRN filter allows CAT II landings, whereas CAT 1 horizontal accuracies are almost achieved. The horizontal accuracy significantly improves with higher terrain significance in front of the airport.

Abbreviations

ADC: Air Data Computer
AGL: Above Ground Level
AmilGeo: Amt für militärisches Geowesen
AS: Anti Spoofing
CAT: Category
DOP: Dilution of Precision
GPS: Global Positioning System
HWILS: Hardware In The Loop Simulation
ICAO: International Civil Aviation Organization
LATAN: Low Altitude Terrain Avoidance and Navigation
LINS: Laser Inertial Navigation System
MTBF: Mean Time Between Failures
NDBS: Navigation Data Bit Stream
PPM: Parts Per Million
RAPIN: Reliable Autonomous Precise Integrated Navigation
SA: Selective Availability (GPS)
SCH: Set Clearence Height
TRN: Terrain Referenced Navigation
UTC: Universal Time Coordinated

Literature

[R1]: Brammer K., Siffling G. Kalman-Bucy-Filter Zweite Auflage, R.Oldenbourg Verlag, München/Wien 1977

[R2]: B. Stieler, Konzept eines Sensorsystems für zukünftige Jagd- und Kampfflugzeuge Flugwiss. Weltraumforschung 9 (1985), Heft 6

[R3]: J. Beyer, Nichtlineare Schätzung inertialer Navigationsgrößen durch die fehlertolerante Verarbeitung zusätzlicher Stützinformation Dissertation TH-Darmstadt

[R4]: GPS Theory and Practice B. Hofmann-Wellenhof, H. Lichtenegger, and J. Collins Springer Verlag Wien New York

[R5]: Beschreibung des Digital Radar Altimeter DRA 100 Deutsche Aerospace AG

[R6]: Decentralized Filtering and Redundancy Management/Failure Detection for Multisensor Integrated Navigation Systems Thomas Kerr in Proceedings of National Technical Meeting of the Institute of Navigation (ION) (San Diego, Calif. Jan 15-17, 1985)

[R7]: Decentralized Filtering and Redundancy Management for Multisensor Navigation Thomas Kerr in IEEE Transactions on Aerospace and Electronic Systems Vol. AES-23, No. 1 January 1987

[R8]: A Fault-Tolerant Multisensor Navigation System Design B.D. Brumback, M.D. Srinath in IEEE Transactions on Aerospace and Electronic Systems Vol. AES-23. No. 6 November 1987

[R9] Agard LS-161, 1988; "The NAVSTAR-GPS System"

[R10] DGON; First international Symposium Real Time Differential Application of GPS, Volume I und II; "DGPS '91"

[R11] Braasch M.S.; NAVIGATION, Winter 90/91, "A Signal Model for GPS"

[R13] Self-Contained GPS Integrity Check using Maximum Solution Separation R. Grover Brown / Paul w. McBurney, NAVIGATION, Spring 1988

[R14] Paul S. Jorgensen.; NAVIGATION, Summer 1989, "An Assessment of Ionospheric Effects on the GPS User"

[R15] AL J. van Dierendonck.; Navigation, Spring 1986, "Recommandations for Differential GPS Service"

[R16] R. Bowen.; NAVIGATION, Summer 1985, "GPS Operational Control System Accuracies"

[R17] Alison K. Brown.; NAVIGATION, Spring 1988, "Civil Aviation Integrity Requirements for the GPS"

[R18] AGARD Flight Test Instrumentation Series Volume 15 on Gyroscopic Instruments and their Applications to Flight Testing B.Stieler, W.Winter September 1982

[R19] ICD-GPS-200 Interface Control Document
[R20] ICD-GPS-059B

[21] Radar seeker based autonomous navigation update System using topography feature matching techniques; H. D. Lerche, F. Tumbrägel AGARD-CP524 Advances in Guidance and Control of Precision Guided Weapons Nov. 1992

[22] Product Specifications for Digital Terrain Elevation Data Defence Mapping Agency 1986

[23] Product Specifications for Digital Feature Analysis Data Defence Mapping Agency 1986

Summary

RAPIN is an integrated navigation system that has been designed with a view to safety and precision for onboard autonomous low-level flight management. The range of applications comprises manned aircraft (fighters, transporters, MPA, helicopters) and unmanned vehicles for reconnaissance and air strike (standoff weapons, drones).

RAPIN integrates 3 physically independent sensors with complementary system characteristics LINS, ADC, PY-GPS and radar altimeter and the TRN system (LATAN). Modeling the error behavior of each individual sensor and additional evaluation of the relevant quality of conformance in a Kalman filter bank optimizes the evaluation of the redundant signals with respect to the required navigation performance (RNP). Residual errors in the sensors (offset and drift) that are also present in calibrated instruments (day to day drift) are determined in the dynamic sensor model (learning the Kalman filter). This permits high quality of conformance in navigation for long periods of time when support information fails (e.g. when GPS reception is subject to phased noise or to terrains with unsufficient terrain significance).
By a focused exploitation of the raw and internal signal redundancies of each Kalman filter (e.g. the residuums) is it possible to isolate and remove component failures. Component failures are simply removed by switching to the sub Kalman filter, that is unaffected by the failed component. The advantage of this so-called Kalman filter bank is that in case of a delayed failure detection we can still access an unaffected Kalman solution.

The RAPIN concept was mainly determined by highly safety critical navigation requirements for tactical missions in a dynamic threat scenario (jamming). RAPIN allows

- autonomous, passive low level flight guidance
- autonomous landings on improvised airfields

The flight tests show that RAPIN fulfills these requirements even without GPS.

		INS + Pressure	GPS	DGPS	TRN	RAPIN
Autonomy		High	Low	-	High	High
Reliability						
	Integrity	Average	Low	Ave.	Average	High
	Availability	High	Average	Low	Average	High
	Continuity	High	Low	Low	High	High
Accuracy						
	Horizontal	Short High Long: Low	High	High	Average	High
	Vertical (MSL)	Average	Average	High	High	High
	Vertical (HGL)	-	-	-	High	High

Table 1 Sensor characteristics

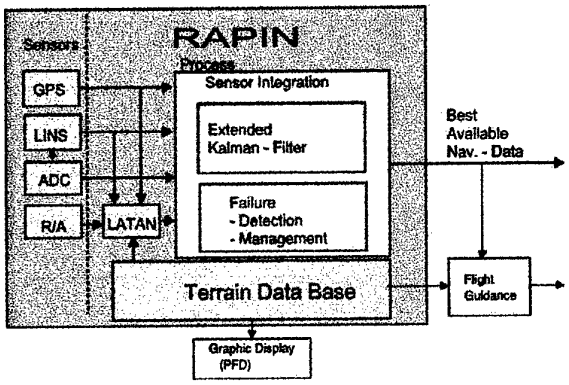


Figure 1 RAPIN Integrated system

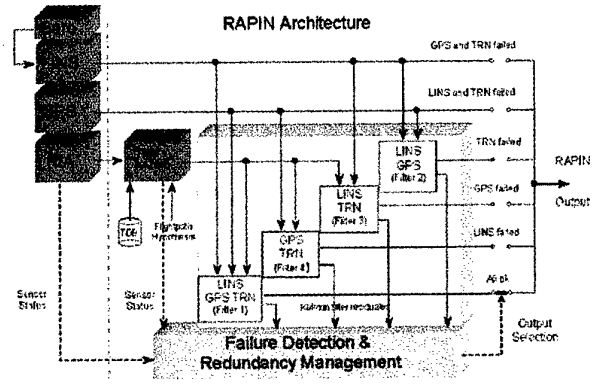


Figure 3 RAPIN Kalman filter bank

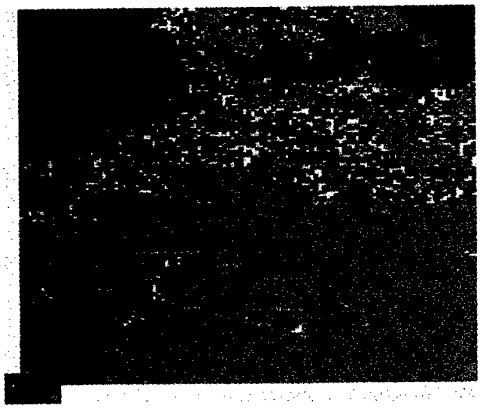


Figure 2 Terrain Significance in Germany

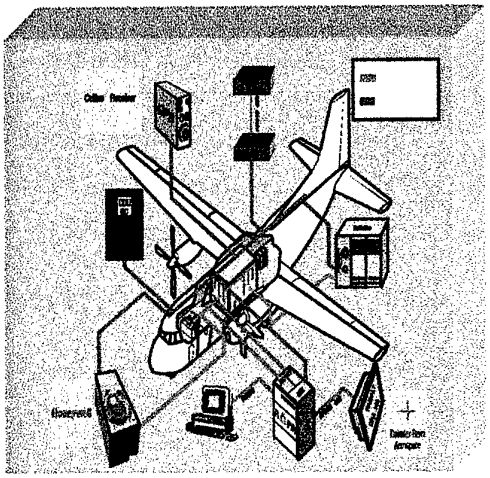


Figure 4 RAPIN components in the Transall

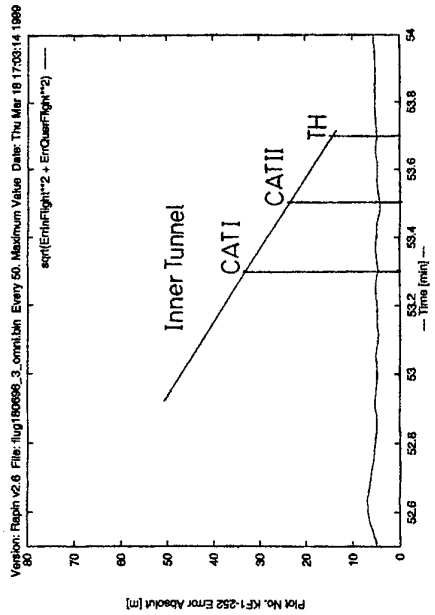


Figure 5: horizontal error LINS/GPS/TRN

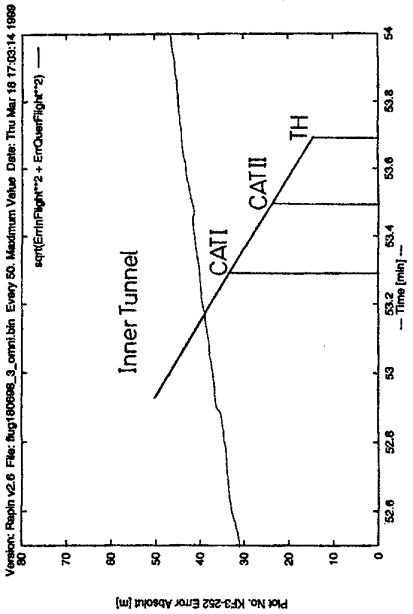


Figure 7: horizontal error LINS/TRN

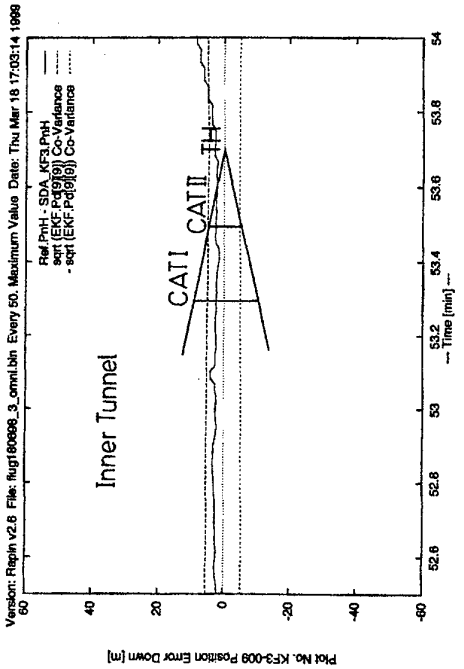


Figure 6: height LINS/TRN

The Synthesis of Federated Filters by Analogy with Transformation of Electric Circuits

V.A. Tupysev
(State Research Center of Russia – Central Scientific & Research Institute Elektropribor
30, Malaya Posadskaya str., Saint Petersburg, 197046, Russia
Tel. (812) 232 59 15. Fax (812) 232 33 75. E-mail: elprib@erbi.spb.su)

1. Introduction

At present when navigational systems are built on a modular principle the methods of federated (distributed) filtering have found a wide application as they allow computational potentialities of measuring modules to be used at a maximum [1-13].

The essence of these methods is that, first, the measurements are processed by using the bank of Kalman-type local filters realized in these modules. Further, local estimates which result from this processing come to the master filter where the best global estimates of the main navigation parameters are generated.

Different approaches to the problem of synthesis and analysis of federated filters are known and a number of important practical and theoretical results have been obtained.

In theory it has been established that as opposed to the centralized processing of measurements when all the measurements can be processed using only one optimal Kalman filter (measurements are assumed to be linear or linearized values), the federated filtering methods are generally suboptimal in relation to the centralized Kalman filter [8,11,14,15].

The conditions for adjusting local filters have also been defined [6,11,15]. It has been shown that the estimated covariance matrix of the estimate error generated in the master filter is an upper estimate for a real covariance matrix of the error estimate [11,15,17].

Various types of federated filters have been analyzed, including those which use the information generated in the master filter both for resetting the local filters [5-8] and for their coordination [16]. The derivation of these and other results involves proofs based on a good mathematical background and sometimes is difficult to understand, that is why in this paper an attempt is made to illustrate the main principles of federated filtering in a simple way, namely, by using the analogy between the process of information processing and the processes typical for electric circuits.

2. Statement of the Problem of Information Processing

Consider the problem of information processing in the following statement:

$$X(k) = F(k)X(k-1) + \xi(k) \tag{1}$$

$$Z_i(k) = H_i(k)X(k) + V_i(k) \tag{2}$$

where $X(k)$ - the vector of the system state, $X(0) \in N\{\hat{X}(0), P(0)\}$; $F(k)$ - the dynamics matrix; $\xi(k)$ - the independent disturbance vector, $\xi(k) \in N\{0, Q(k)\}$; $Z_i(k)$ - the measurement vector from the i-th sensor; $H_i(k)$ - the measurement matrix, $V_i(k)$ - the independent white-noise measurement error, $V_i(k) \in N\{0, R_i(k)\}$.

Equation (1) describes the change of the system state in time, and Equation (2) describes the process of measurement in the measuring modules.

It is known that the optimal solution of the estimation problem in this statement can be derived with the use of the Kalman filter procedures:

$$\tilde{X}(k) = F(k)\hat{X}(k-1) \tag{3}$$

$$L(k) = F(k)P(k-1)F^T(k) + Q(k) \tag{4}$$

$$P(k) = \left(L^{-1}(k) + \sum H_i(k)R_i^{-1}(k)H_i(k)\right)^{-1} \tag{5}$$

$$\hat{X}(k) = \tilde{X}(k) + P(k) \times \\ \times \sum H_i^T(k)R_i^{-1}(k)\left(Z_i(k) - H_i(k)\tilde{X}(k)\right) \tag{6}$$

Suppose that there exist the inverse matrices $P_{mi} = \left[H_i^T R_i^{-1} H_i\right]^{-1}$, then by simple manipulations, omitting the index k, it is possible to represent the filter equation (5,6) as follows:

$$\hat{X} = P\left(L^{-1}\tilde{X} + \sum H_i^T R_i^{-1} Z_i\right) = \\ = P\left(L^{-1}\tilde{X} + \sum P_{mi}^{-1}\hat{X}_{mi}\right) \tag{7}$$

$$P = \left(L^{-1} + \sum H_i^T R_i H_i\right)^{-1} = \left(L^{-1} + \sum P_{mi}^{-1}\right)^{-1} \tag{8}$$

where $\hat{X}_{mi} = \left(H_i^T R_i^{-1} H_i\right)^{-1} H_i^T R_i^{-1} Z_i$.

Now consider the following parallel electric circuit of emf sources shown at the left of Fig.1

Paper presented at the RTO SCI International Conference on "Integrated Navigation Systems", held at "Elektropribor", St. Petersburg, Russia, 24-26 May 1999, and published in RTO MP-43.

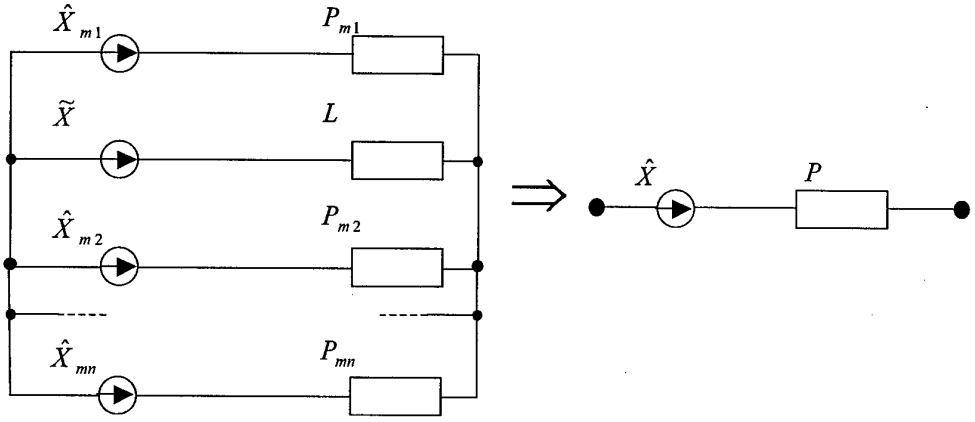


Fig. 1

It is known that the equivalent emf of an ideal source and the equivalent resistance of this circuit, as follows from the theory of electrical circuits, are calculated by the same equations (5,6) as the equations used in calculating the estimates of the state vector and the covariance matrix, if scalar values are used as X and Z_i.

This analogy and the similarity of manipulations on scalar values and matrices give us grounds to consider the circuit given above as a formal illustration of the filtering process for the multidimensional case as well under the assumption that the estimates of state vectors are an analog of the emf, the analog of resistancies are covariance matrices, and the analog of the procedure for the state vector estimation is the calculation of parameters of an equivalent source.

Of course, in view of the assumptions made (substitutions of scalar values for matrices and vectors), the results which will be obtained by analogy with the analysis and transformation of electric circuits can only be used to determine the line of investigations in filtering and they call for strict mathematical proof.

However the use of electric circuits as an analog allows the process of synthesis of various federated filters to be illustrated clearly and in this sense, from our point of view, it proves to be useful.

The analogy between the calculation of the equivalent source and the estimation procedures at one step of filtering allows an assumption that there is a similar analogy between the multistep process of estimation and a certain electric circuit which illustrates this process. Indeed, with account of the prediction of information, the process of estimation for the case of two sensors at the first step can be represented by an electric circuit shown in Fig.2.

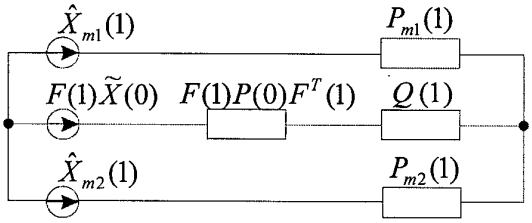


Fig. 2.

Developing the electric circuit further in a similar way and introducing the notation

$$Q^*(j)=\left[\prod_{s=j+1}^{k}F(s)\right]Q(j)\left[\prod_{s=j+1}^{k}F^T(s)\right],$$

$$P^*_{mi}(j)=\left[\prod_{s=j+1}^{k}F(s)\right]P_{mi}(j)\left[\prod_{s=j+1}^{k}F^T(s)\right] \qquad (9)$$

$$\hat{X}^*_{mi}(j)=\left[\prod_{s=j+1}^{k}F(s)\right]\hat{X}_{mi}(j),$$

$$P^*(0)=\left[\prod_{s=j+1}^{k}F(s)\right]P(0)\left[\prod_{s=j+1}^{k}F^T(s)\right]$$

it is possible to obtain the following circuit, presented in Fig. 3, which illustrates the process of optimal estimation starting from the zero step to the k-th step. Generalization of the circuit for the case of n sensors is evident.

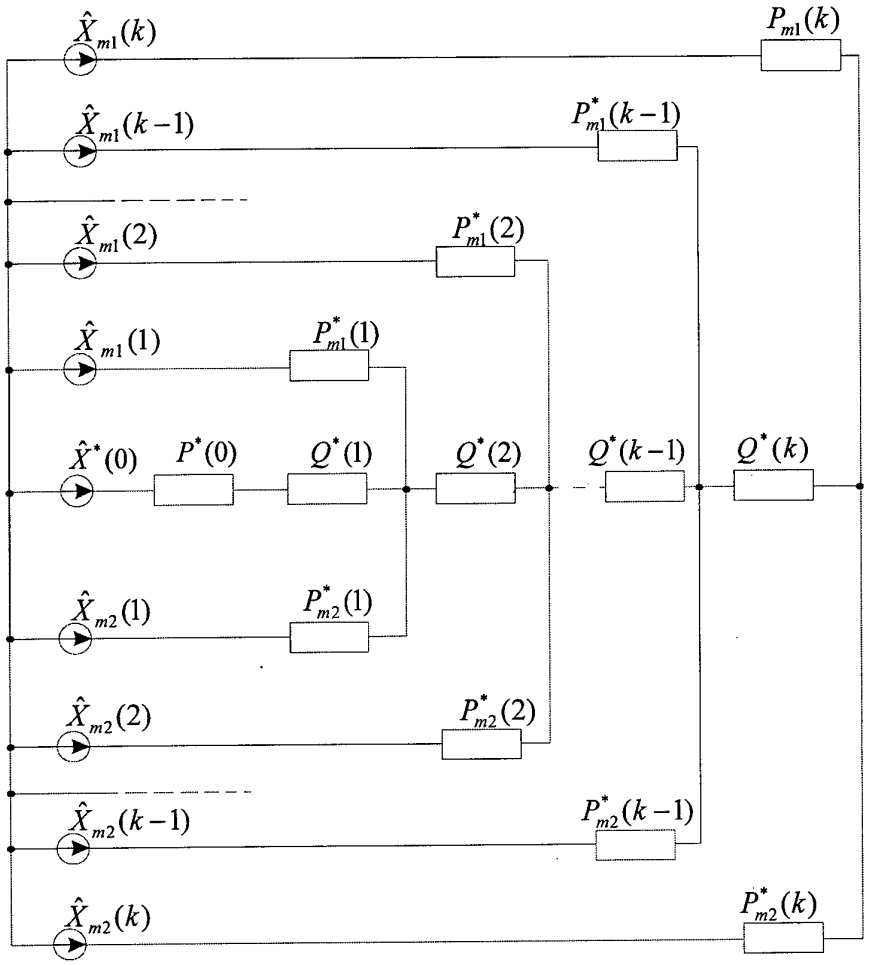


Fig. 3

3. Synthesis of Federated Filters

Now transforming the circuit presented in Fig. 3 let us show the possible approaches to synthesis of federated filters and their connection with the optimal centralized filter. For this, let us transform the central part of the circuit replacing the resistances $Q^*(i)$, $P^*(0)$ by the parallel-connected resistances $Q_1^*(i)$ and $Q_2^*(i)$, $P_1^*(0)$ and $P_2^*(0)$, correspondingly, in such a way as not to break the equivalence of the circuit (Fig.4). From the theory of electric circuits it is known that the equivalence of a circuit will not be broken if the resistances $Q_i^*(k)$, $P_i^*(0)$ will satisfy the conditions:

$$\sum Q_i^{*-1}(k) = Q^{*-1}(k)$$
$$\sum P_i^{*-1}(0) = P^{*-1}(0) \qquad (10)$$

which, in their turn, with regard for the notation (9) are fulfilled if the conditions

$$\sum Q_i^{-1}(k) = Q^{-1}(k)$$
$$\sum P_i^{-1}(0) = P^{-1}(0) \qquad (11)$$

are satisfied.

Note that the equations (11) represent the known conditions for adjustment of local filters [6,11,15], but as before, the circuit illustrates the optimal process of information processing.

To obtain a circuit which will illustrate the process of federated filtering, let us break all the jumpers in the circuit marked *. Indeed, in this case the upper part of the circuit shown in Fig.4 reflects the processing of the measurements $Z_1(k)$ in the local filter 1, the lower part represents the processing of measurements $Z_2(k)$ in the local filter 2. The substitution of the upper and the lower part of the circuit for the equivalent sources with the parameters $P_i(k)$, $\hat{X}_i(k)$, the analogs for which are the parameters generated in the local filters (5) (Fig.5),

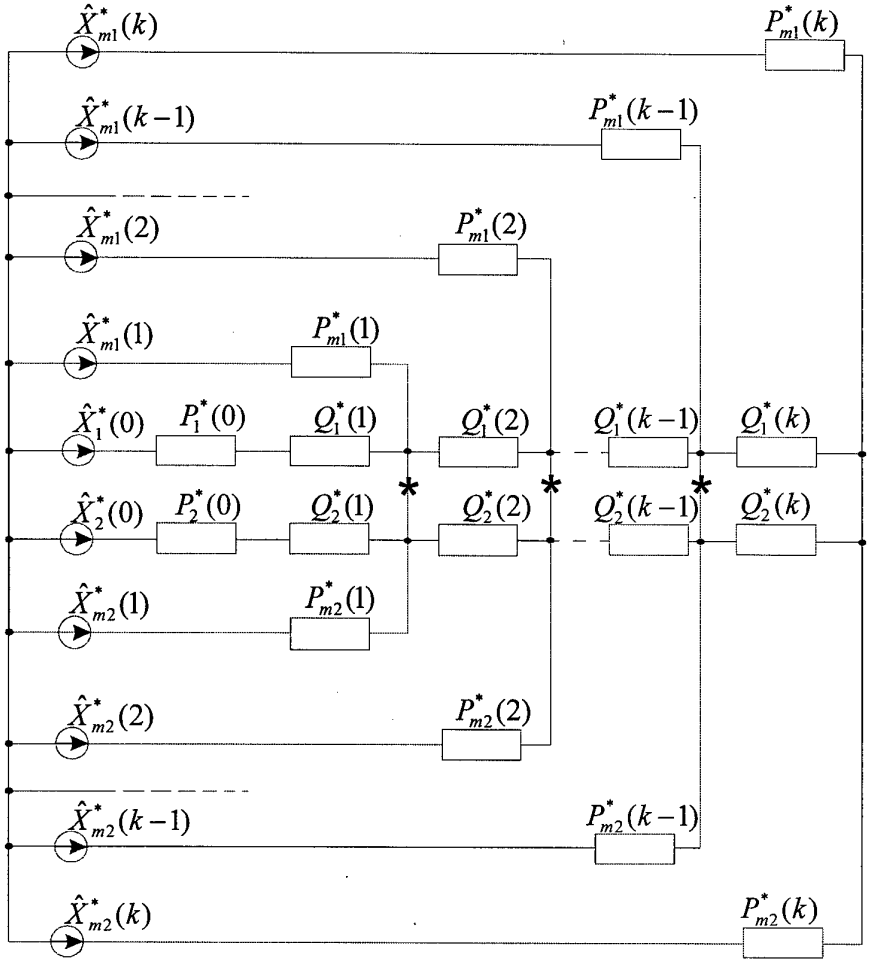


Fig. 4.

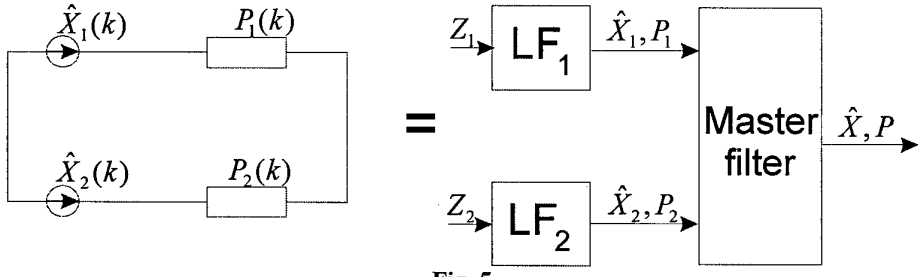


Fig. 5.

makes it clear how a global estimate can be derived in the master filter by the known formulas for weighing local estimates

$$P(k)\sum P_i^{-1}(k)\hat{X}_i(k) = \hat{X}(k)$$
$$\sum P_i^{-1}(k) = P(k) \qquad (12)$$

The difference between the circuits in Fig.3,4 which illustrate the processes of optimal and federated filtering allows the conclusion that federated filters are suboptimal in the general case, and that the calculated covariance matrix $P(k)$ of the federated filter will be bigger than the optimal covariance matrix $P_{opt}(k)$ when the adjustment of the local filters satisfy the conditions (11). Both of these conclusions follow from the fact that a part of the current channels has been eliminated from the electric circuit that can only result in the increase of the resistance of the circuit. A rigorous mathematical proof of the inequality $P(k) > D(k) > P_{opt}(k)$ (where $D(k)$ is a real covariance matrix of the estimate error) which means that $P(k)$ is an upper bound for the real covariance

matrix of the estimate error of the federated filter, is given in [11,17].

4. The Conditions of the Federated Filter Optimality

As it was mentioned above, the federated filter is suboptimal in the general case. However it may be shown that in the absence of limitations on the dimension of local filters the optimality will be valid both for a certain adjustment of local filters and special models.

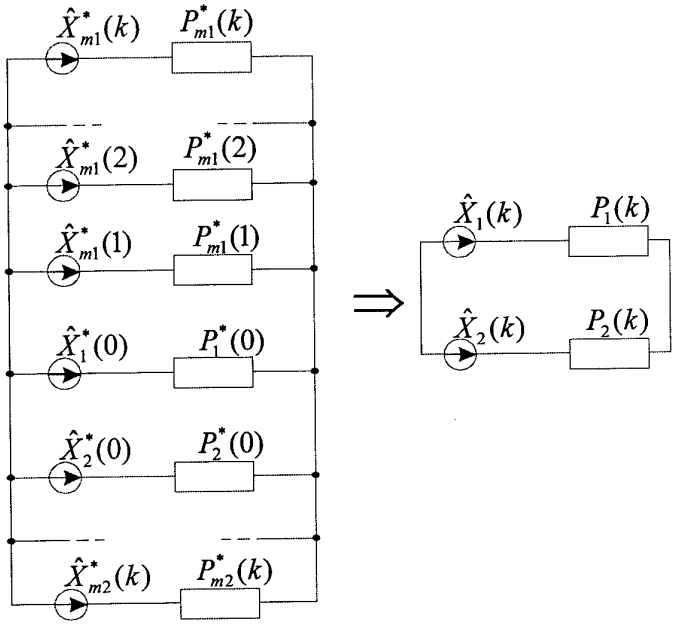


Fig. 6.

The fact that the upper part of the circuit corresponds to the procedures of the local filter 1 and the lower one to the local filter 2 allows the conclusion that the filter is optimal for the quasi-deterministic description as well. The mathematical proof for such a conclusion is given in [9,10].

Now let us illustrate the principle of the coordinated adjustment of local filters [16]. It is evident that the optimality of the federated filter will be valid if the break of the jumpers in the circuit in Fig.4 does not cause changes in the state of the circuit.

For better visibility let us illustrate the essence of the coordinated adjustment of local filters by the example of the electric circuit shown in Fig. 7.

First, let us demand that the fragments of the circuits represented in Fig.8 should be equivalent.
It is obvious that this requirement is met if the following conditions are met

First consider a model with a quasi-deterministic description of the state vector *X(k)*,

$$X(k) = F(k)X(k-1). \qquad (13)$$

This model can be considered as a special case of the model (1) when the matrix of disturbance noise intensities $Q(k)$ is equal to zero.
Under this statement of the estimation problem the circuit illustrating the optimal process of filtering will be as it shown in Fig.6.

$$\left\{(P_1^*)^{-1}+(P_2^*)^{-1}\right\}^{-1}+\left\{(Q_1^*)^{-1}+(Q_2^*)^{-1}\right\}^{-1}= \qquad (14)$$
$$=\left\{P_1^*+Q_1^*)^{-1}+(P_2^*+Q_2^*)^{-1}\right\}^{-1}$$

$$\left\{(P_1^*)^{-1}+(P_2^*)^{-1}\right\}^{-1}\left\{(P_1^*)^{-1}\hat{X}_1^*+(P_2^*)^{-1}\hat{X}_2^*\right\}= \qquad (15)$$
$$=\left\{P_1^*+Q_1^*)^{-1}+(P_2^*+Q_2^*)^{-1}\right\}^{-1}\times$$
$$\times\left\{P_1^*+Q_1^*)^{-1}\hat{X}_1^*+(P_2^*+Q_2^*)^{-1}\hat{X}_2^*\right\}$$

Now let us show that the conditions (14) comply with the matrices Q_1^* и Q_2^* in the form of:

$$Q_i^* = P_i^*(P^*)^{-1}Q^* \qquad (16)$$

where $(P^*)^{-1}=(P_1^*)^{-1}+(P_2^*)^{-1}$,

$Q^* = \left\{(Q_1^*)^{-1}+(Q_2^*)^{-1}\right\}^{-1}$.

Indeed, using (16) a number of simple manipulations of the right-hand part of the equation (14) yields :

$$\left\{\left(P_1^* + P_1^*(P^*)^{-1}Q^*\right)^{-1} + \left(P_2^* + P_2^*(P^*)^{-1}Q^*\right)^{-1}\right\}^{-1} =$$

$$= \left\{\left(E + (P^*)^{-1}Q^*\right)^{-1}\left\{(P_1^*)^{-1} + (P_2^*)^{-1}\right\}\right\}^{-1} =$$

$$= P^*\left(E + (P^*)^{-1}Q^*\right) = P^* + Q^* \qquad (17)$$

Taking into consideration the notation, the left-hand part of the equation (14) is also equal to $P^* + Q^*$. In a similar way it is proved that Q_i^* in the form of (16) also satisfies the equation (15).

It should be noted that the substitution of the fragment of the circuit in Fig. 7 for the fragment of the circuit in Fig.8 on the right results in a circuit of the same structure, as after this substitution the sources $\hat{X}_i^*$ and $\hat{X}_{mi}^*$ can, in their turn, be substituted for equivalent sources with the parameters $\hat{X}_i^*(1)$, $P_i^*(1)$ which are further used in calculating $Q_1^*(2)$ and $Q_2^*(2)$.

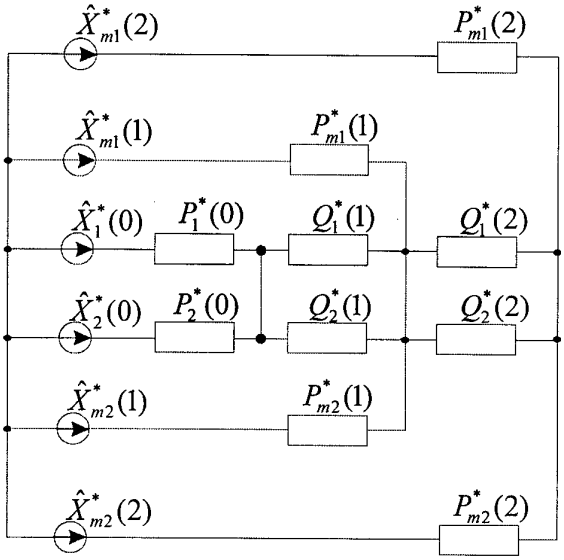


Fig. 7.

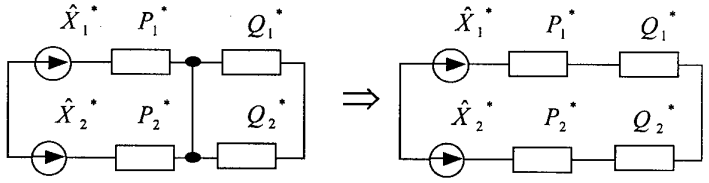


Fig. 8.

Turning back to the circuit which illustrates the process of optimal filtering and carrying out a number of analogous transformations of the circuit, with account of the reverse analogy between the electric circuit and the filtering process it is not difficult to derive the following conditions for adjustment of the local filters at the k-th step which provide the optimal solution of the estimation problem:

$$Q(k) = \left[F(k)P_i(k-1)F^T(k)\right]\times$$
$$\times\left[F(k)P\ (k-1)F^T(k)\right]^{-1}Q\ (k) \qquad (18)$$

where $P\ (k-1)$ is the covariance matrix generated in the master filter at the previous step of filtering; $P_i(k-1)$ are the covariance matrices generated in the local filters.

However it should be noted that the practical realization of the filter, with the adjustment conditions in the form of (16), can present difficulties as the matrix $Q_i\ (k)$ is not symmetric in this case and, consequently, the prediction matrix $L^{-1}(k)$ used in the formulas of Kalman filter (5) will also be asymmetric.

But noticing the fact that Q_i^T is also a solution of the equations (14,15), this shortcoming can be eliminated by substituting the fragment of the circuit at the left of Fig. (8) for the other equivalent fragments (Fig.9):

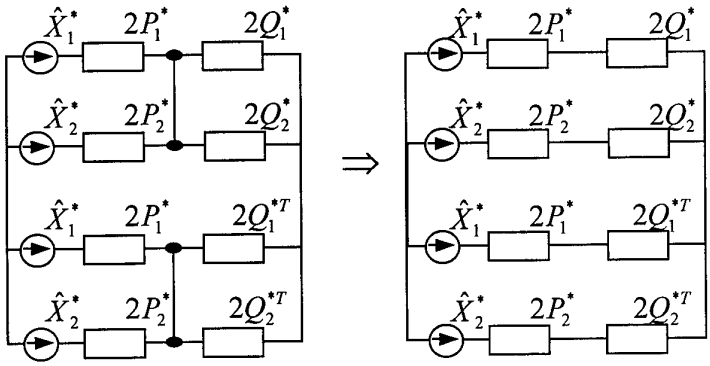


Fig. 9

In practical realization of the federated filter with the coordinated adjustment that means that the matrix L^{-1} used in the formula (5) is calculated by the formula

$$L_i^{-1} = \frac{(L_i)^{-1} + (L_i)^{-T}}{2} \text{ where } L_i = FP_iF^T + Q_i \,.$$

5. The Synthesis of Fussion – Reset Federated Filter

The approach using electric circuits described above can also be employed for the illustration of federated filters with reset of local filters at each step of filtering [5-8]. The core of this method of federated filtering is that the global estimate and the covariance matrix calculated in the master filter is used in the local filters not only to calculate the matrices of disturbance noise intensities $Q_i(k)$ but also to change the state of these filters. Again let us illustrate the procedure of resetting local filters by the example of an electric circuit represented in fig.7.

Let us substitute a fragment of this circuit for a fragment of the same structure but with other parameters of the elements:

From the theory of electric circuits it follows that the equivalence of the circuit will not be broken if in choosing $\hat{X}_i^{**}$, P_i^{**} the following conditions will be fulfilled:

$$\sum (P_i^{**})^{-1} = \sum (P_i^{*})^{-1} = P^{*} \qquad (19)$$

$$\sum (P_i^{**})^{-1} \hat{X}_i^{**} = \sum (P_i^{*})^{-1} \hat{X}_i^{*} = (P^{*})^{-1} \hat{X}^{*} \qquad (20)$$

Note that in a particular case $\hat{X}^{*}$ can be chosen as $\hat{X}_i^{**}$ as with this choice the condition (20) is also fulfilled.

Now substituting the fragment of the circuit in Fig.7 for the fragment of the circuit at the right in Fig.10, breaking the jumper marked "*", and substituting the sources $\hat{X}_i^{**}$ and $\hat{X}_{mi}^{*}$ for the equivalent sources with the parameters $\hat{X}_i^{*}(1)$, $\hat{P}_i^{*}(1)$ it is possible to obtain a circuit of the same structure which makes it possible to use a similar procedure of transformation of electric circuits further.

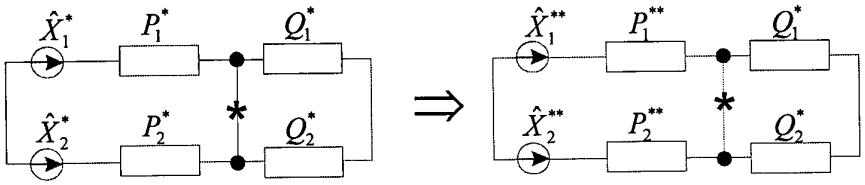


Fig. 10

Now returning to transformation of the circuit representing the process of optimal filtering and making a reverse analogy between the electric circuit and the process of filtering it is possible to show that the conditions for resetting the local filters at each step of filtering will be as follows:

$$\sum (P_i^R(k))^{-1} = \sum P_i^{-1}(k) = P^{-1}(k) \qquad (21)$$

$$\sum (P_i^R(k))^{-1} \hat{X}_i^R(k) =$$

$$= \sum P_i^{-1}(k) \hat{X}_i(k) = P^{-1}(k) \hat{X}(k),$$

where $P(k)$ and $\hat{X}_i(k)$ are the parameters generated in the master filter; $P_i^R(k)$, $\hat{X}_i^R(k)$ are the parameters of the local filters after the reset.

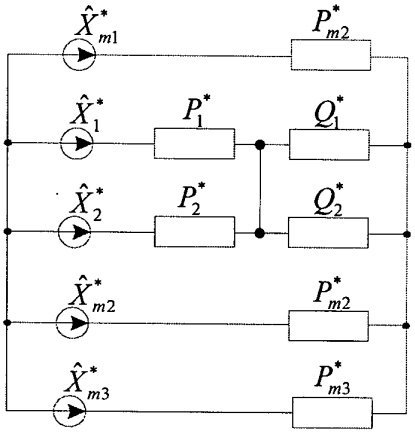
The analysis of the circuit representing the process of federated filtering shows that fusion-reset filters are suboptimal in the general case as the circuit which illustrates their operation has been derived from the circuit for the optimal filter by breaking a part of jumpers between the upper and the lower parts of the circuit.

In order to obtain an optimal procedure of filtering, besides resetting the local filters the latter should be coordinated by the formulas (18) where the calculated covariance matrices of the local filters after resetting are used as $P_i(k-1)$.

6. Illustration of the Principle of Measurement Reproduction

This principle is of great importance in federated filtering as it allows the same measurements to be processed in different local filters. The necessity of this use of information arises, for example, in the process of damping two and more INS from one velocity sensor or in correction of a few INS from one GPS receiver.
Consider the circuit which illustrates the process of optimal information processing at the first step of filtering for the case of 3 sensors (Fig.11, *a*).

a)



b)

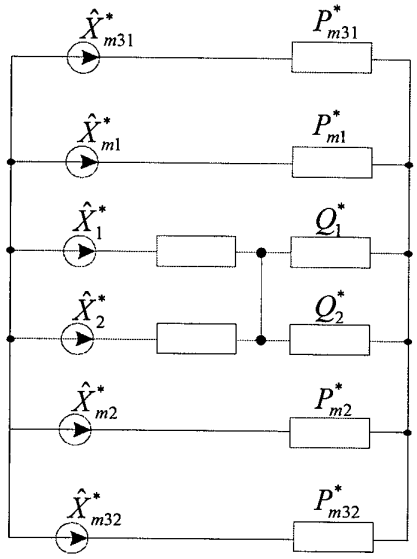


Fig. 11

In this circuit $\hat{X}_3$ represent the measurements which are supposed to be processed in two local filters.

Let us substitute the fragment of the circuit corresponding to the measurement process in the third sensor for a fragment of the circuit equivalent to it (Fig.12).

It is obvious that the conditions of equivalence of this substitution are the conditions:

$$\sum (P^*_{m3i})^{-1} = (P^*_{m3})^{-1} \qquad (22)$$

$$\sum (P^*_{m3i})^{-1} \hat{X}^*_{m3i} = (P^*_{m3})^{-1} \hat{X}^*_{m3} \qquad (23)$$

Note that in a particular case the equal values of $\hat{X}^*_{m3}$ can be chosen as $\hat{X}^*_{m3i}$ as with these values of $\hat{X}^*_{m3i}$ the equation (23) is also valid.

Now substituting the fragment in the circuit shown in Fig.11, *a* for the equivalent fragment of the circuit at the left in Fig.12 it is possible to derive the following circuit which illustrates the process of processing in the centralized filter (Fig.11, *b*).

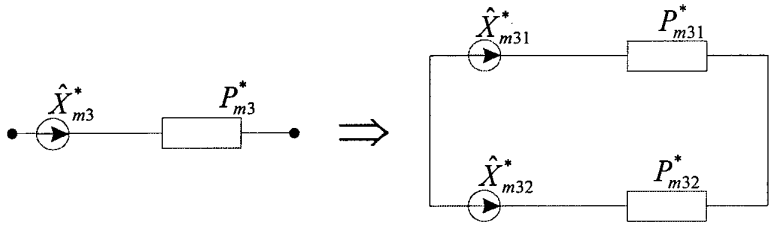


Fig. 12.

Generalization of the circuit for the case of the multistep process of filtering is evident.
Further, eliminating the jumpers between the upper and the lower parts of the circuit it is possible to obtain this or that circuit which illustrates this or that filtering process.

Note that the conditions (22, 23) result directly from the principle of measurement reproduction whose mathematical formulation for the measuring models can be given as follows:

measuring models

$$Z(k) = H(k)X(k) + V(k) \qquad V(k) \in N\{0, R(k)\}$$

and

$$Z_1(k) = H(k)X(k) + V_1(k) \qquad V_1(k) \in N\{0, R_1(k)\}$$
$$Z_2(k) = H(k)X(k) + V_2(k) \qquad V_2(k) \in N\{0, R_2(k)\}$$

are equivalent if the conditions

$$\sum R_i^{-1}(k) = R^{-1}(k)$$
$$\sum R_i^{-1}(k) Z_i(k) = R^{-1}(k) Z(k)$$

are satisfied.

A rigorous mathematical proof of this principle for the case when $Z_i(k) = Z(k)$ is given in [16].

7. Conclusions

The analogy established between transformation of electric circuits and the procedures of optimal and federated filtering has been used in this paper to illustrate in a simple way the approaches to the synthesis of various types of federated filtering. The analysis of electric circuits has revealed the reasons of loss of optimality in the federated filters and allowed us to illustrate the principle of coordinated adjustment of local filters when the optimality remains valid. The illustration of the principle of measurement reproduction makes it clear how the same measurements can be used in two or more local filters. Though the principles of federated filtering were illustrated without any account of limitations on the dimension of local filters, the analogy between the procedures of federated filtering and transformation of electric circuits can also be established for the case with these limitations.

References

1. Speyer, J.L., "Computation and Transmission Requierments for a Decentralized Linear-Quadratic-Gaussian Control Problem," *IEEE Transactions on Automatic Control,* AC-24:2, Apr. 1979, pp.266-269.
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, pp.609-610.
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, pp.799-812.
4. Levy, B.C., D.A. Castanon, G.C. Verghese and A.S.Willsky, "A Scattering Framework for Decentralized Estimation Problems," *Automatica,* 19:4, July 1983, pp.373-384.
5. Carlson, N.A., & Neily, C.M. Jr., "Distributed Kalman Filter Architectures," Phase I Final Report, AFWAL-TR-87-1181, Avionics Laboratory, WPAFB, OH, June 1987.
6. Carlson, N.A., "Federated Square Root Filter for Decentralized Parallel Processes," *Proceedings of National Aerospace Electronics Conference,* NAECON, 1987.
7. Carlson, N.A., "Information-Sharing Approach to Federated Kalman Filtering," NAECON, May 1988, pp.265-280.
8. Carlson, N.A.,"Federated Filter for Fault-Tolerant Integrated Navigation," AGARD-AG-331, Advisory Group for Aerospace Research and Development, Neuilly-sur-Seine, France, June 1995.
9. Tupysev, V.A. "A Symplified Algorithm of Optimal Filtering of Measurements Including Systematic Errors". *"Voprosy Korablestroenia", ser. "Navigation and Gyroscopy"*, Issue 49, 1980, pp. 68-75 (In Russian).

10. Tupysev, V.A. "Optimal Channel Measurement Processing Using Weighing of Local Estimates". *"Voprosy Korablestroenia", ser. "Navigation and Gyroscopy"*, Issue 64, 1982, pp. 77-87 (In Russian).
11. Tupysev, V.A. "Channel Information Processing with a Guaranteed Accuracy of Estimation". *"Voprosy Korablestroenia", ser. "Navigation and Gyroscopy"*, Issue 82, 1985, pp. 54-61 (In Russian).
12. Tupysev, V.A. Tyumeneva G. "Information Processing for a Modular Architecture of a Navigational system". *"Sudostroenie"*, N 8, 1984 (In Russian).
13. Qiu Hongzhuan, Zhang Hongyue "Federated Filer and its Application to Integrated Navigation System" *Proceedings of Second International Symposium on Inertial Technology in Beijing.* Beijing, China, 1998.
14. Levy, L.J., "Suboptimality of Cascaded and Federated Kalman Filters," *Proceedings of the 52nd Annual Meeting, Cambridge,* MA, June 1996, pp.399-407.
15. Da R. "Two-Error Covariance Analysis Algorithms for Suboptimal Decentralized KalmanFilters" *Journal of Guidance, Control and Dynamics, vol.16*, No.5, September-October 1993.
16. Tupysev V. "A General Approach to the Problem of Distributed Kalman Filtering", *Proceedings of AIAA Guidance, Navigation, and Control Conference.* Boston, 1998. Part 2.
17. Tupysev V. "An Approximate Evaluation of Federated Filter Errors" *Gyroscopy and Navigation.* 1996, No.3, pp.68-73.

Stereoscopic Navigation and Observation Systems

I.N. BELOGLAZOV
S.N. KAZARIN
Zhukovsky Air Force Engineering Academy, 125190, Russia, Moscow,
Planetnaja ulitca, 3, tel. 251-2404, kazarins@stk.mmtel.ru

Introduction

The possibility of obtaining the three-dimensional image of an observed scene has been studied in detail in photogrammetry (stereo-photogrammetry). Besides photography facilities, the shooting can be obtained by other iconic systems such as television and infrared systems, radiometers, radar stations, etc., which substantially extends the photogrammetry field. Television, radar, and other kinds of photogrammetry have appeared. Due to the development of powerful computers, the possibility of digital data processing for various iconic systems on moving objects in real time has appeared.

Currently, besides cartography, architecture, and building, the ideas of stereo-photogrammetry are beginning to be applied in systems of stereoscopic computer vision and stereoscopic television systems of future generations. Stereo-photogrammetry has wide potential in navigation. In [1-5], the problem of estimating the location and orientation in three-dimensional space of a mobile object equipped with an iconic system acquiring two-dimensional images is solved. The possibility of using these systems for a vision system of a mobile robot is stressed in [2,3]. The ideas of stereoscopic systems are applied in aircraft navigation [6, 7]. In particular, *Innovative Configuration* (United States) is developing a navigational three-dimensional video system for helicopters to provide safety during low-altitude flights (for instance, over forests).

This paper is aimed at the development and generalization of theoretical results. We derive equations of observation for the use of both a stereo image and a sequence of stereo images, and synthesize optimal rigid and searching algorithms of data processing in stereoscopic navigation systems, study the accuracy of aircraft positioning and determination of the aircraft velocity and orientation, synthesize an algorithm of optimal data processing in combined stereoscopic navigation systems that joins the idea of search and optimal filtration.

A STEREOSCOPIC NAVIGATION SYSTEM

1. Equations of 3-D Observation

The laws of constructing images by different iconic facilities (frame and panoramic aerial cameras, television, infrared cameras, radiometers, radar stations) can be distinguished. But in many cases, mathematical models of iconic facilities appear to be quite precise, which allows us to solve various problem concerning image processing. Although further presentation is aimed at iconic systems that are based on the principle of the perspective projection [8], which is inherent in one or another form in any optic-electronic system with an optic unit, the majority of the final results hold true for all iconic systems.

Consider a stereo image (Fig. 1), i.e., a pair of images that has a common region for the observation W. In the plane of the first image, the uniform grid with the step δ is introduced. The point with coordinates $x^{I}_{i} = i\,\delta$, $y^{I}_{j} = j\,\delta$ is located at the intersection of the i-th column and j-th row.

Equations (1.4), the perspective projection allow us to find the point of the second image x^{II}_{ij}, y^{II}_{ij}, that corresponds to x^{I}_{i}, y^{I}_{j},. First, according to

$$X_{ij} = X^{SI} + [\,h(X_{ij},Y_{ij}) - H^{SI}\,]\,\frac{a_{11}\,x^{I}_{i} + a_{21}\,y^{I}_{j} - a_{31}\,F}{a_{13}\,x^{I}_{i} + a_{23}\,y^{I}_{j} - a_{33}\,F}, \tag{1.1}$$

$$Y_{ij} = Y^{SI} + [\,h(X_{ij},Y_{ij}) - H^{SI}\,]\,\frac{a_{12}\,x^{I}_{i} + a_{22}\,y^{I}_{j} - a_{32}\,F}{a_{13}\,x^{I}_{i} + a_{23}\,y^{I}_{j} - a_{33}\,F},$$

we determine the ground coordinates of the image X_{ij}, Y_{ij}, $h(X_{ij},Y_{ij})$.

The formulas for calculating coordinates of the corresponding point on the second image of the stereoscopic image are as follows:

$$x^{II}_{ij} = F\,\frac{a_{11}\,(X_{ij} - X^{SII}) + a_{12}\,(Y_{ij} - Y^{SII}) - a_{13}\,(H^{SII} - h(X_{ij},Y_{ij}))}{-\,a_{31}\,(X_{ij} - X^{SII}) - a_{32}\,(Y_{ij} - Y^{SII}) + a_{33}\,(H^{SII} - h(X_{ij},Y_{ij}))}, \tag{1.2}$$

$$y^{II}_{ij} = F\,\frac{a_{21}\,(X_{ij} - X^{SII}) + a_{22}\,(Y_{ij} - Y^{SII}) - a_{23}\,(H^{SII} - h(X_{ij},Y_{ij}))}{-\,a_{31}\,(X_{ij} - X^{SII}) - a_{32}\,(Y_{ij} - Y^{SII}) + a_{33}\,(H^{SII} - h(X_{ij},Y_{ij}))},$$

Paper presented at the RTO SCI International Conference on "Integrated Navigation Systems", held at "Elektropribor", St. Petersburg, Russia, 24-26 May 1999, and published in RTO MP-43.

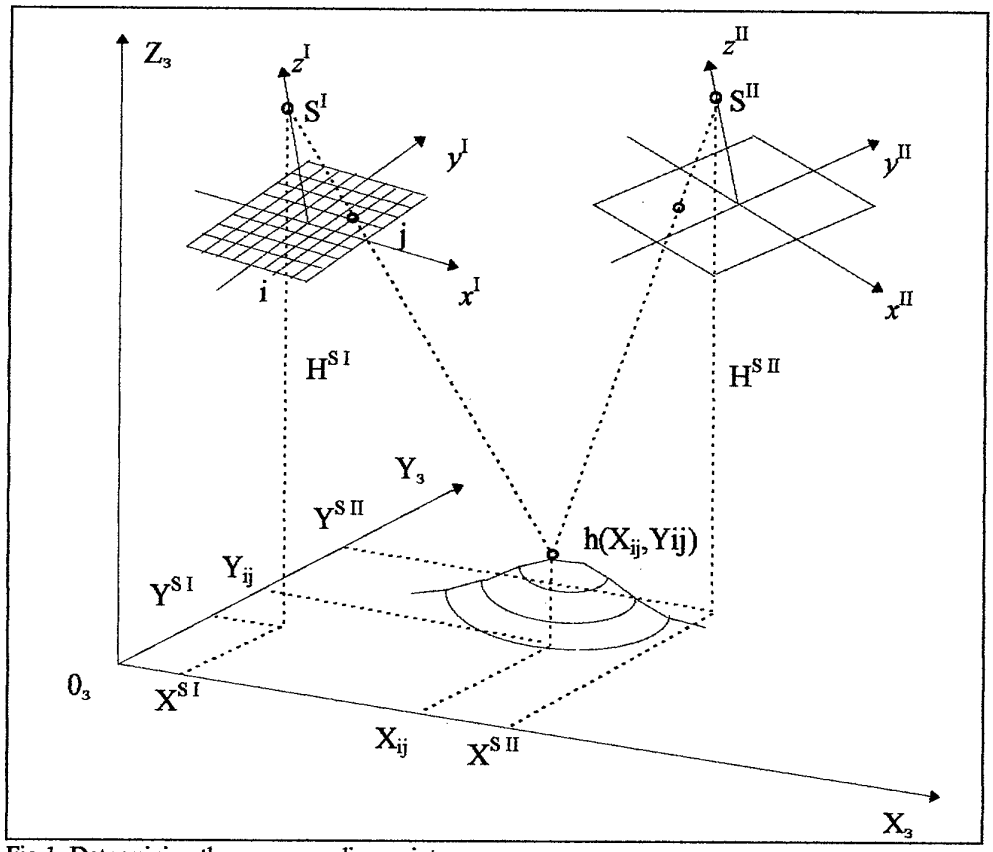


Fig.1. Determining the corresponding points on the first and second images of the stereo image .

$a_{11} = \cos\psi \cos\vartheta$, $a_{12} = -\sin\psi \cos\vartheta$,
$a_{13} = \sin\vartheta$,
$a_{21} = \sin\psi \cos\gamma - \cos\psi \sin\vartheta \sin\gamma$,
$a_{22} = \cos\psi \cos\gamma + \sin\psi \sin\vartheta \sin\gamma$,
$a_{23} = \cos\vartheta \sin\gamma$,
$a_{31} = -\sin\psi \sin\gamma - \cos\psi \sin\vartheta \cos\gamma$,
$a_{32} = -\cos\psi \sin\gamma + \sin\psi \sin\vartheta \cos\gamma$,
$a_{33} = \cos\vartheta \cos\gamma$; ψ , ϑ, γ - are the course angle, the pitch angle and the roll angle.

Relations (1.1), (1.2) connect x^{II}_{ij}, y^{II}_{ij} and x^{I}_{i}, y^{I}_{j} if elements of the orientation for the first and the second images, $X^{SI}_{SII}, Y^{SI}_{II}, H^{SI}_{II}, \psi^{I}_{II}, \vartheta^{I}_{II}, \gamma^{I}_{II}$ and $X^{SII}_{SII}, Y^{SII}_{II}, H^{SII}_{II}, \psi^{II}_{II}, \vartheta^{II}_{II}, \gamma^{II}_{II}$, respectively, are known, as well as the relief h(X, Y).

Let values of coordinates, the velocity, and angular orientation

$$X^{*SII}_{II} = X^{SII}_{II} + \Delta X^{II}, \qquad X^{*SI}_{I} = X^{SI}_{I} + \Delta X^{I},$$
$$Y^{*SII}_{II} = Y^{SII}_{II} + \Delta Y^{II}, \qquad Y^{*SI}_{I} = Y^{SI}_{I} + \Delta Y^{I},$$
$$H^{*SII}_{II} = H^{SII}_{II} + \Delta H, \qquad H^{*SI}_{I} = H^{SI}_{I} + \Delta H,$$
$$\psi^{*II}_{II} = \psi^{II}_{II} + \Delta\psi, \qquad \psi^{*I}_{I} = \psi^{I}_{I} + \Delta\psi,$$
$$\vartheta^{*II}_{II} = \vartheta^{II}_{II} + \Delta\vartheta, \qquad \vartheta^{*I}_{I} = \vartheta^{I}_{I} + \Delta\vartheta,$$
$$\gamma^{*II}_{II} = \gamma^{II}_{II} + \Delta\gamma, \qquad \gamma^{*I}_{I} = \gamma^{I}_{I} + \Delta\gamma,$$

$$V^{*}_{x} = V_x + \Delta V_x, \qquad V^{*}_{y} = V_y + \Delta V_y,$$

be measured onboard the aircraft, where ΔX, ΔY, ΔH, ΔV_x, ΔV_y, $\Delta\psi$, $\Delta\vartheta$, $\Delta\gamma$ are corresponding measurement errors, $\Delta X^{II} = \Delta X^{I} + \Delta V_x \tau$, $\Delta Y^{II} = \Delta Y^{I} + \Delta V_y \tau$, τ is time of movement between the points where the first and second images of the stereo image are obtained. Errors of measurements of the height, velocity, and angles are presumed constant.

If f_{ij} denotes the intensity of the observed terrain point X_{ij} , Y_{ij} , $Z_{ij} = h(X_{ij}, Y_{ij})$, then the signals of the receiver at the corresponding points x^{I}_{i}, y^{I}_{j} and x^{II}_{ij}, y^{II}_{ij} of the first and second images are as follows:

$$f^{I}_{ij} = f(x^{I}_{i}, y^{I}_{j}) = f_{ij} + \delta f^{I}_{ij},$$
$$f^{II}_{ij} = f(x^{II}_{ij}, y^{II}_{ij}) = f_{ij} + \delta f^{II}_{ij},$$

where δf^{I}_{ij} , δf^{II}_{ij} are instrumental errors (noises) of .receivers.

Then, we have

$$f^{I}_{ij}(x^{I}_{i}, y^{I}_{j}) = f^{II}_{ij}(x^{II}_{ij}, y^{II}_{ij}) + \delta f_{ij}\ , \qquad (1.3)$$

$$\delta f_{ij} = \delta f^{I}_{ij} - \delta f^{II}_{ij}\ ,$$

$i = 1, \dots n$, $j = \alpha, \dots \beta$; n, α, β are determined by the region of common observation *W*. We introduce the following notation

$$Z = (f^{I}_{1\alpha}, f^{I}_{1\alpha+1}, \dots, f^{I}_{1\beta}; f^{I}_{2\alpha}, \dots, f^{I}_{2\beta}; \dots; f^{I}_{i\alpha}, f^{I}_{i\alpha+1}, \dots f^{I}_{ij}, \dots, f^{I}_{i\beta}; \dots; f^{I}_{n\alpha}, \dots, f^{I}_{n\beta})^{T}.$$

According to (1.3), this vector can be written as follows:

$$Z = (f^{II}_{1\alpha}, f^{II}_{1\alpha+1}, \dots, f^{II}_{1\beta}; f^{II}_{2\alpha}, \dots, f^{II}_{2\beta}; \dots; f^{II}_{i\alpha}, f^{II}_{i\alpha+1}, \dots f^{II}_{ij}, \dots, f^{II}_{i\beta}; \dots; f^{II}_{n\alpha}, \dots, f^{II}_{n\beta})^{T} + \delta Z,$$

where $\delta Z = (\delta f_{1\alpha}, \dots \delta f_{ij}, \dots \delta f_{n\beta})^{T}$.

Below, we denote the vector of errors of onboard measurements

$X = (\Delta X, \Delta V_x, \Delta H, \Delta\vartheta, \Delta Y, \Delta V_y, \Delta\psi, \Delta\gamma)^{T}$, the vector of measured values of orientation elements of the image

$$X^{*} = (X^{*SII}, Y^{*SII}, H^{*SII}, \psi^{*II}, \vartheta^{*II}, \gamma^{*II}, X^{*SI}, Y^{*SI}, H^{*SI}, \psi^{*I}, \vartheta^{*I}, \gamma^{*I})^{T},$$

the vector of values of relief heights in viewing points of the stereo image

$$h = (h_{1\alpha}, h_{1\alpha+1}, \dots, h_{1\beta}; h_{2\alpha}, \dots, h_{2\beta}; \dots; h_{i\alpha}, \dots, h_{ij}, \dots, h_{i\beta}; \dots; h_{n\alpha}, \dots, h_{n\beta})^{T},$$

the vector of values of heights from the digital map of the available relief

$$h^{k} = (h^{k}_{1\alpha}, h^{k}_{1\alpha+1}, \dots, h^{k}_{1\beta}; h^{k}_{2\alpha}, \dots, h^{k}_{2\beta}; \dots; h^{k}_{i\alpha}, \dots h^{k}_{ij}, \dots, h^{k}_{i\beta}; \dots; h^{k}_{n\alpha}, \dots, h^{k}_{n\beta})^{T},$$

and the vector of errors of cartography by $\delta h = (\delta h_{1\alpha}, \dots, \delta h_{ij}, \dots, \delta h_{n\beta})^{T}$, respectively.

The vector of illumination f^{II} for the corresponding points of the second image depends on X^{*}, X, h

$$f^{II} = F(X^{*}, X, h),\quad F(X^{*}, X, h) = [F_{ij}(X^{*}, X, h)],$$
$$F_{ij}(X^{*}, X, h) = f^{II}(x^{II}_{ij}, y^{II}_{ij}), \qquad (1.4)$$

$i = 1, \dots n$, $j = \alpha, \dots \beta$.

The vector function $F_{ij}(X^{*}, X, h)$ is given by equalities (1.1), (1.2), and (1.4). In new notation, we can write

$$Z = F(X^{*}, X, h) + \delta Z\ .$$

Linearizing the function $F(\dots)$ in a neighborhood of the point $X = 0$, $h = h^{k}$ and taking into account $h = h^{k} - \delta h$, we obtain

$$Z = F(X^{*}, 0, h^{k}) + \left.\frac{dF(X^{*}, X, h)}{dX}\right|_{\substack{h = h^{k} \\ X = 0}} X + \left.\frac{dF(X^{*}, X, h)}{dh}\right|_{\substack{h = h^{k} \\ X = 0}} \delta h + \delta Z\ .$$

Denoting

$$\left.\frac{dF(X^{*}, X, h)}{dX}\right|_{\substack{h = h^{k} \\ X = 0}} = H(X^{*}, h^{k}),$$

$$\delta Z = \delta Z - \delta Z^{*},$$

$$\delta Z^{*} = \left.\frac{dF(X^{*}, X, h)}{dh}\right|_{\substack{h = h^{k} \\ X = 0}} \delta h,$$

we obtain the equation of observation in the form

$$Z = F(X^{*}, 0, h^{k}) + H(X^{*}, h^{k})X + \delta Z\ . \qquad (1.5)$$

Let us determine the covariance matrix $R = M\langle \delta Z\ \delta Z^{T} \rangle$ for the vector of errors for the model of stereo observation. In view of the fact that δZ and δZ^{*} are not correlated, we have

$$R = \sigma^{2}_{\sim} E,\quad \sigma^{2}_{\sim} = \frac{B^{2}}{H^{2}}\sigma^{2}_{f'}\sigma^{2}_{k} + 2\sigma^{2}_{n},$$

E is a unit matrix, $\sigma^{2}_{f'}$ and σ^{2}_{k} are variances of the optic gradient field and the error of the relief map, respectively, σ^{2}_{n} is the variance of the receiver noise.

Consider a stabilized horizontal flight at height $H^{SI} = H^{SII} = H$ with a constant velocity $V_x = V$, zero course $\psi^{I} = \psi^{II} = 0$, and with the vertical optic axis of the frame grabber $\vartheta^{I} = \vartheta^{II} = 0$, $\gamma^{I} = \gamma^{II} = 0$. The best use of stereo takes place for $B = H \tan\varphi$, where φ is half of a viewing angle of the iconic facility (see Fig. 2). In this i case, the common region *W* occupies half of an image, and, in formulas (1.4), (1.5) $\alpha = -n$, $\beta = n$, $Y^{*SII} = Y^{*SI}$, $X^{*SII} = X^{*SII} - B$, where $B = V\tau$ is the stereo base of observation, and $a_{11} = a_{22} = a_{33} = 1$, and the other elements of the matrix of directing cosines equal zero.

Equations (1.1), (1.2) are as follows:

$$X_{ij} = X^{SI} + \frac{h(X_{ij}, Y_{ij}) - H}{F}\, x^{I}_{i},$$

$$Y_{ij} = Y^{SI} + \frac{h(X_{ij}, Y_{ij}) - H}{F}\, y^{I}_{i},$$

$$x_{ij}^{II} = \frac{F}{(H - h(X_{ij}, Y_{ij}))}(X_{ij} - X^{SII}),$$
$$y_{ij}^{II} = \frac{F}{(H - h(X_{ij}, Y_{ij}))}(Y_{ij} - Y^{SII}).$$

Application of a sequence of stereo images supposes serial processing of a number of stereo images.

differences of illuminations for the corresponding points of the first and second images of the stereo image, and the position of the corresponding points on the second image is determined by the data X^* of the corrected navigation system.

The transformations performed give [9]

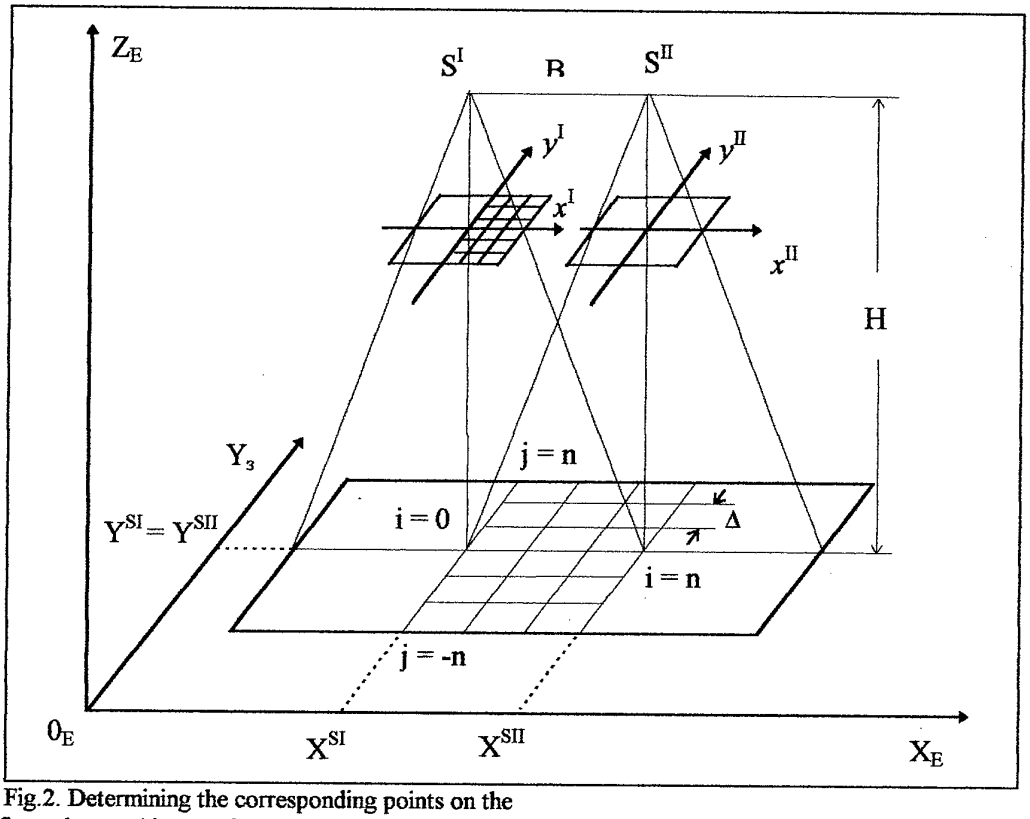


Fig.2. Determining the corresponding points on the fist and second images for stabilized horizontal flight.

2. The accuracy of a rigid stereoscopic navigation system

2.1 Rigid evaluation of coordinates on the basis of one stereo image

Above, we obtained the equation of observation for a stereoscopic navigation system

$$Z = F(X^*, 0, h^k) + H(X^*, h^k)X + \delta Z.$$

If we introduce the notation $\tilde{Z} = Z - F(X^*, 0, h^k)$, then we obtain the equation of observation in the standard form used in the theory of Kalman filtration:

$$\tilde{Z} = H(X^*, h^k)X + \delta Z,$$

where $X = (\Delta X, \Delta V_x, \Delta H, \Delta\vartheta, \Delta Y, \Delta V_y, \Delta\psi, \Delta\gamma)^T$,

and the vector of observation $\tilde{Z}$ is the set of

$$\sigma_X = \frac{2}{\sqrt{5}} \frac{\Delta\sigma_\sim}{\sigma_f H^2 \tan\varphi \sigma_h},$$
$$\sigma_{Vx} = \sqrt{\frac{6}{5}} \frac{\Delta\sigma_\sim V}{\sigma_f H^2 \tan^3\varphi \sigma_h},$$
$$\sigma_H = \sqrt{\frac{6}{5}} \frac{\Delta\sigma_\sim}{\sigma_f H \tan^3\varphi \sigma_h},$$
$$\sigma_\vartheta = \frac{\sqrt{3}}{2} \frac{\Delta\sigma_\sim}{\sigma_f H^2 \tan^3\varphi},$$
$$\sigma_Y = \frac{2}{\sqrt{5}} \frac{\Delta\sigma_\sim}{\sigma_f H \tan^2\varphi \sigma_h},$$
$$\sigma_{Vy} = \sqrt{\frac{6}{5}} \frac{\Delta\sigma_\sim V}{\sigma_f H^2 \tan^3\varphi \sigma_h},$$
$$\sigma_\psi = \sqrt{\frac{6}{5}} \frac{\Delta\sigma_\sim}{\sigma_f H^2 \tan^3\varphi \sigma_h},$$

$$\sigma_\gamma = \sqrt{\frac{3}{2}} \frac{\Delta\sigma_\sim}{\sigma_{f'} H^2 \tan^3 \varphi}.$$

We assumed above that the value of discreteness Δ must not be less than the radius of correlation $\rho_{f'}$ of the optic gradient field. The mean square deviations for the field f (X,Y) and its gradient f'_x(X,Y) are connected by the approximate relation

$$\sigma_f = \rho_f \sigma_{f'}, \qquad (2.1)$$

where ρ_f is the radius of correlation of the field f (X,Y). If the correlation function of the one-dimensional section of the optic field has the form $R_f = \sigma_f^2 e^{-\alpha|\Delta|}$, then equality (2.1) is exactly satisfied, and $\rho_f = \frac{1}{\alpha}$.

Taking this into account, we obtain

$$a = \frac{\Delta\sigma_\sim}{\sigma_{f'}} = \rho_{f'} \sqrt{\sigma_k^2 \tan^2 \varphi + \frac{2\rho_f^2}{(\sigma_f/\sigma_n)^2}}.$$

Fig 3 – 4 show graphs that illustrate the relative accuracy of estimating coordinates in longitudinal and lateral channels of the stereoscopic navigation system for observation of one stereo image ($\tan\varphi = 0.5$) depending on the height of observation H and on the gradient of the relief field $\sigma_{h'}$.

2.2. Rigid estimating coordinates on the basis of a sequence of stereo images

It is necessary to estimate the vector вектор $X_k = (\Delta X_k, \Delta V_{xk}, \Delta H_k, \Delta\vartheta_k, \Delta Y_k, \Delta V_{yk}, \Delta\psi_k, \Delta\gamma_k)^T$ on the basis of observation of a sequence of stereo images. In this procedure, besides the equations of the observation channel, we use equations of the controlled process

$$\Delta X_{k+1} = \Delta X_k + \Delta V_{x\,k}\tau,$$
$$\Delta Y_{k+1} = \Delta Y_k + \Delta V_{y\,k}\tau,$$
$$\Delta V_{x\,k+1} = \Delta V_{x\,k}, \quad \Delta V_{y\,k+1} = \Delta V_{y\,k},$$
$$\Delta H_{k+1} = \Delta H_k, \quad \Delta\psi_{k+1} = \Delta\psi_k,$$
$$\Delta\vartheta_{k+1} = \Delta\vartheta_k, \quad \Delta\gamma_{k+1} = \Delta\gamma_k$$

or in the matrix form $X_{k+1} = A_k X_k$.

Matrix A_k is block diagonal and can be determined by blocks A_k^{long} and A_k^{lat} that correspond to longitudinal and lateral estimation channels:

$$A_k = \begin{bmatrix} A_k^{long} & 0 \\ 0 & A_k^{lat} \end{bmatrix},$$

where $A_k^{long} = A_k^{lat} = \begin{bmatrix} 1 & \tau & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$.

The behavior of the longitudinal and lateral channel is determined by a subvector

$X_k^{long} = (\Delta X_k, \Delta V_{xk}, \Delta H_k, \Delta\vartheta_k)^T$ and a subvector $X_k^{lat} = (\Delta Y_k, \Delta V_{yk}, \Delta\psi_k, \Delta\gamma_k)^T$, respectively.

Fig 5 – 6 show the dependence of relative mean square errors of estimating the navigation parameters of the longitudinal and lateral channels of stereoscopic navigation system on the number of stereo images k for different gradients of the relief $\sigma_{h'}$ and different heights of flight

3. Searching and combined stereoscopic navigation systems

We know from the theory of extremal correlation navigation systems [10, 11] that rigid algorithms can eliminate initial deviations that are not greater than the first extremum - minimum of the correlation. Rigid stereoscopic navigation systems have an analogous property.

Initial errors of the navigation system can de sufficiently large, and the loss of workability of the rigid stereoscopic navigation system based on linearization of observation equations and using the Kalman filtration is possible. In this case, it is necessary to apply other algorithms of processing that do not contain constraints on the value of the initial error of the navigation system. These algorithms are connected with the idea of searching the actual location of the aircraft.

Let us clarify the idea of the searching stereoscopic navigation system for the simplest case of a uniform straight horizontal stabilized flight in the absence of measurement errors of the height, velocity, and angular orientation.

The theory of statistical decisions [10, 11] is the theoretical ground for optimal algorithms of searching stereoscopic navigation systems.

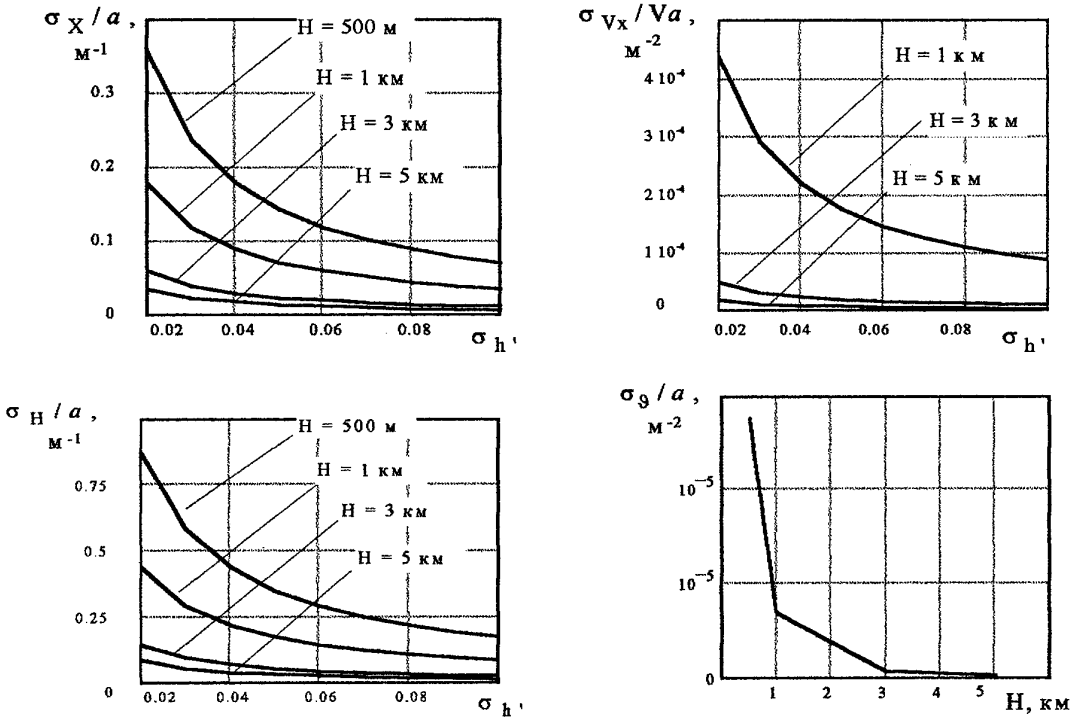


Fig.3 The accuracy of determining navigation parameters in the longitudinal channel on basis of a stereo image.

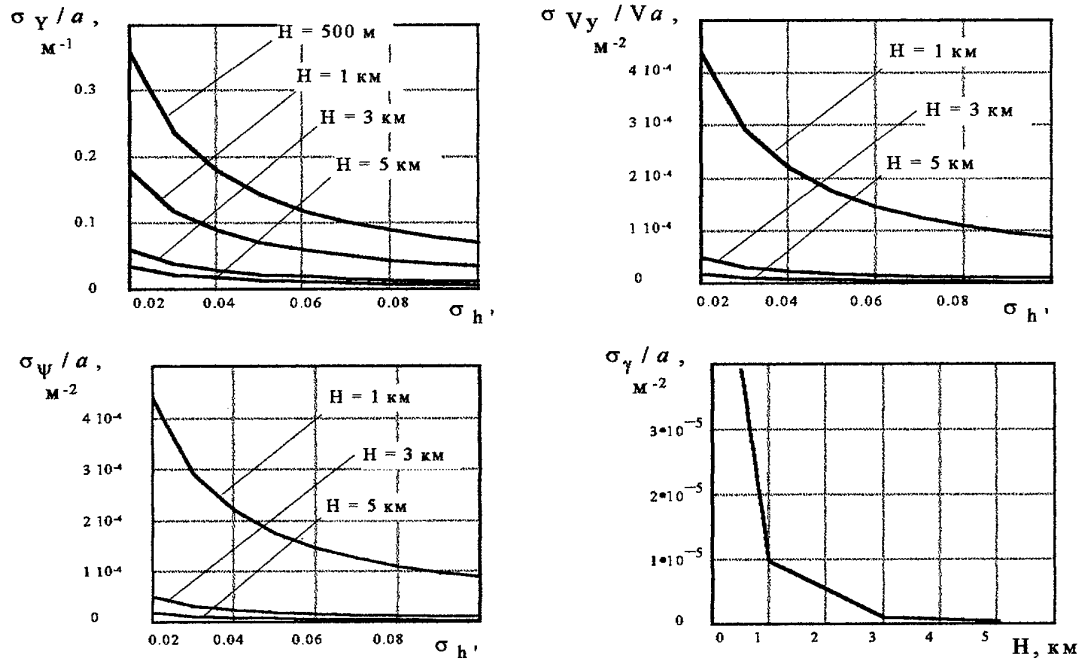


Fig.4 The accuracy of determining navigation parameters in the lateral channel on basis of a stereo image.

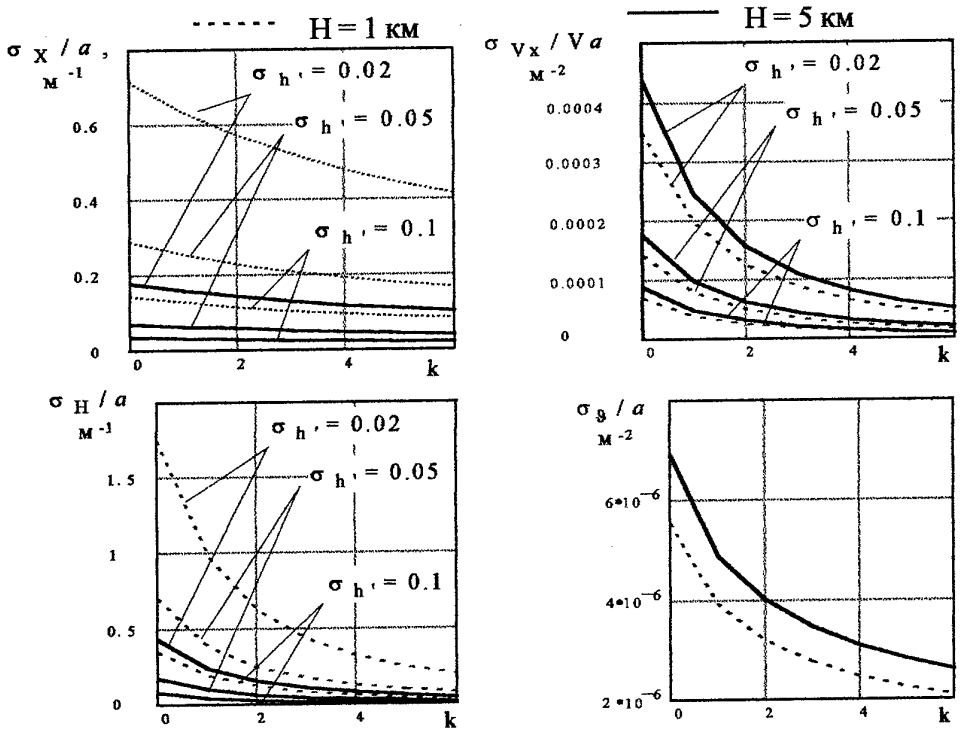


Fig.5 The accuracy of determining navigation parameters in the longitudinal channel on basis of a sequence of stereo images.

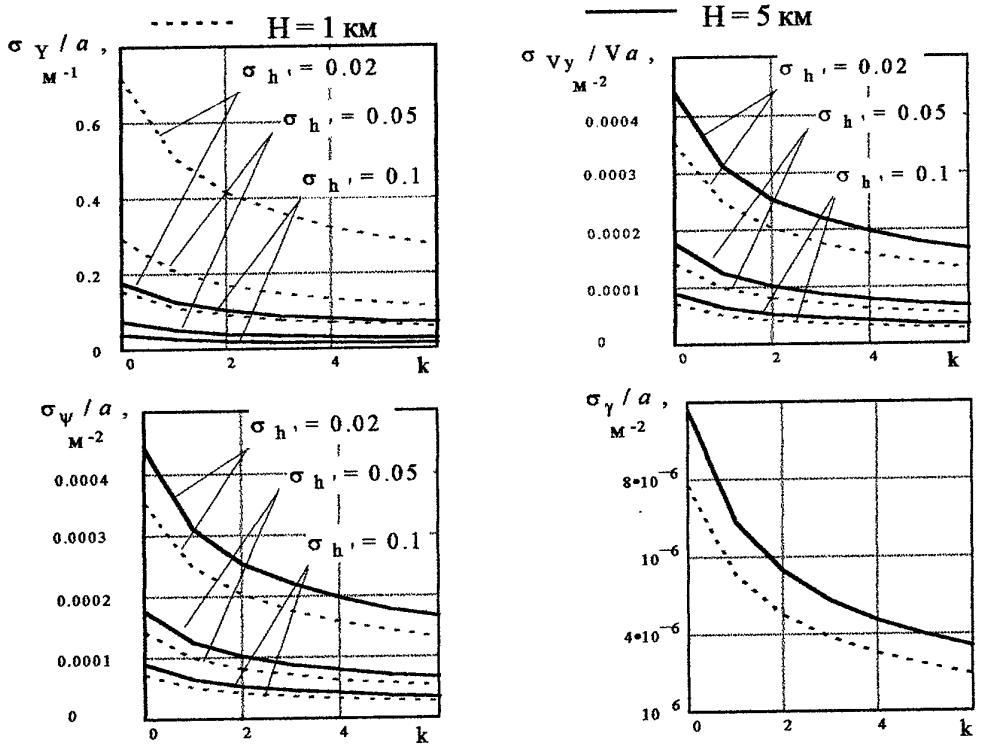


Fig.6 The accuracy of determining navigation parameters in the lateral channel on basis of a sequence of stereo images.

Fig. 7 shows the block diagram of the considered simplified version of the searching stereoscopic navigation system. We briefly describe the mathematical transformations performed by single units.

Memory units I and II store the information about all stereo images. At sequential discrete instants k-1, k, two images of the terrain that have considerable intersection are captured by optic-electronic and/or radar facilities. The set of values $\{f^{I}_{ijk}\}$ are written in memory unit I. The second image is digitized with respect to the coordinates x, y with a step that is considerably less than δ. We do this in order to further control the reading of the second image at an arbitrary point x, y with the required accuracy. The array $\{f^{II}_{\xi\eta k}\}$ storing the information about the second image is written in memory unit II.

The automatic optimizer performs functions generating tested hypotheses D_{pq} about possible location of the aircraft and seeks for the minimum of the functional over possible hypotheses.

The digital map of the relief h(X, Y) in the area of stereo observations is contained in the relief map unit. After formulating the hypothesis D_{pq} the array of the relief values $\{h_{ijpq}\}$ at the node points of the terrain that correspond node points of the first image is taken from the map unit. This array is input in the reading control unit.

In controlling the reading, we find the location x^{II}_{ijpqk}, y^{II}_{ijpqk} of node points on the second image the corresponds to the hypothesis D_{pq}. It is at these points of memory unit II that the hypothetical values of illumination f^{II}_{ijpqk} are read.

The set $\{f^{I}_{ijk}\}$ and also the set of values $\{f^{II}_{ijpqk}\}$ (for the tested hypothesis D_{pq}) come into the unit of calculating functionals for calculating the functional I_{pqk} of their basis. This functional represents the measure of discrepancy for the tested hypothesis about the illumination of points of the first and second images of stereo image. The value I_{pqk} comes into the automatic optimizer and is stored in it. After calculating I_{pql} a new hypothesis is given, and the corresponding value of the functional is calculated. In blind search, the functional are calculated for all hypothesis tested, and then the hypothesis $D_{\mu\nu}$ that corresponds to the functional $I_{\mu\nu N}$ with the least value is determined. The coordinates that correspond to this hypothesis are taken as the actual location of the aircraft.

Modern multifunctional on board complexes require diverse navigation and flight information. Besides the coordinates and the velocity of motion of the aircraft center of gravity, the accurate course, vertical, and other parameters are also often required. Therefore, the extension of the vector of the parameters tested is one of the lines of developing stereoscopic navigation systems. To solve the problems of parametric estimation in stereoscopic navigation systems for large initial errors of the navigation system, as in the case of correlation of extremal navigation systems [20, 11], it is necessary to use combined algorithms of estimation. These algorithms are based on the joint optimal estimation and testing of hypotheses (the so-called recurrent-searching estimation) and combine the ideas of search and filtration [12, 13].

4. The study of the stereoscopic navigation system by numerical simulation

The experiment is aimed at confirming the theoretical conclusions concerning the accuracy of determining the aircraft coordinates on the basis of actual images of terrain and the basis standard of the relief determined by using a topographical map.

We studied four types of relief: plains, low mountains, middle mountains, and high mountains. The mean square deviation of the relief gradient was as follows: 0.02–0.12 for plains, 0.05–0.13 for low mountains, 0.08–0.41 for middle mountains, 0.13–0.75 for high mountains.

The initial data for experiments was based on (1) aerial images of terrain equipped by the stations of the special geodesic network; the overlap, the height of shooting, and the focus length of the aerial camera were 50%, 5 km, and 200 mm, respectively; (2) a topographic map of the terrain shot at scale 1:50000; the accuracy of presenting the relief on the map was not worse then 25 m on the plane and 4 m with respect to the height. The mean square value of the relief gradient was 0.05.

In experiments, we use stereo-comparator SKA-30, framegrabber FEAG-2000, and a personal computer.

The algorithm of the searching stereoscopic navigation system studied in the experiment was described in section 3.

Fig. 8 presents the results obtained for $m=9*10^4$ points (300×300 points, 60 realizations) in the form of a histogram of distribution of the error δR. The mean square accuracy of determining the aircraft coordinates was $\sigma_R=17.5$ m, the step of exhaustion of hypotheses $\Delta_x=20$ m was also sufficient.

Processing all points available of the digitized stereo images ($m=34\times10^4$) for four realizations of the

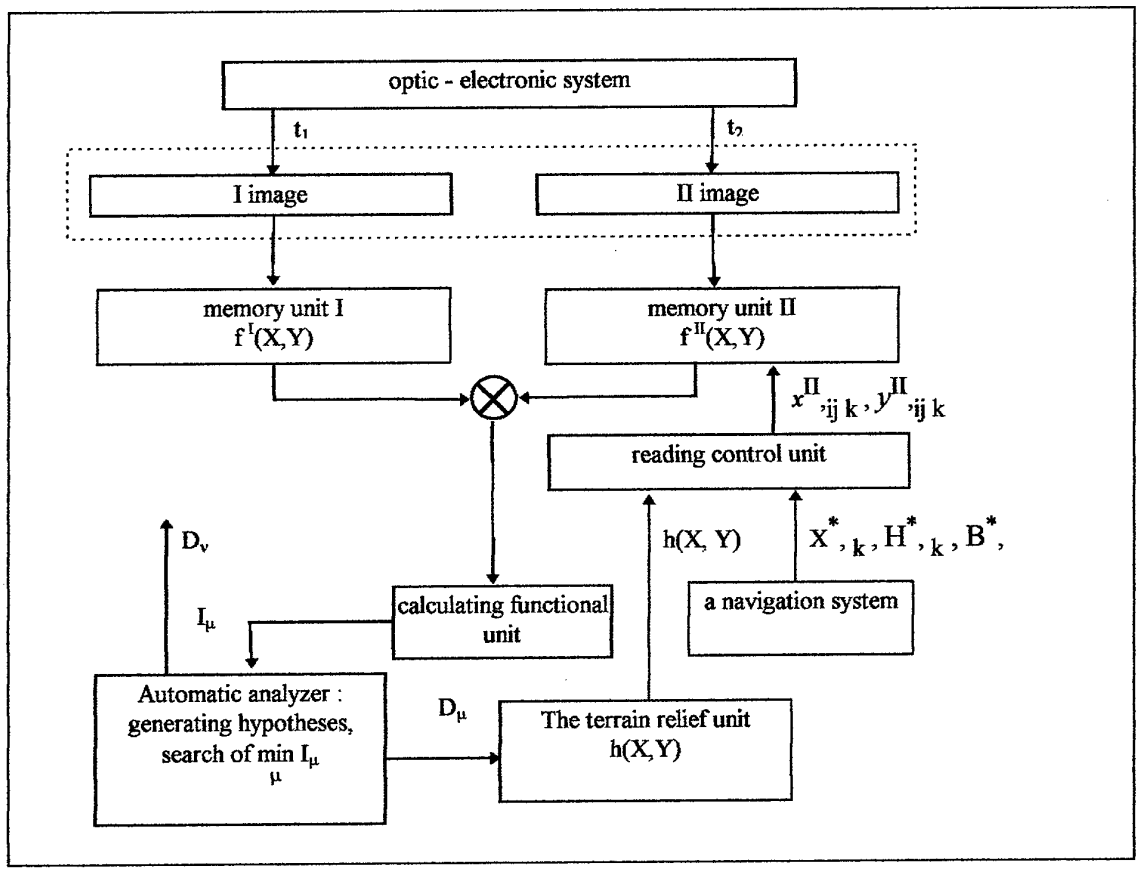


Fig.7. The block diagram of a searching stereoscopic navigation system.

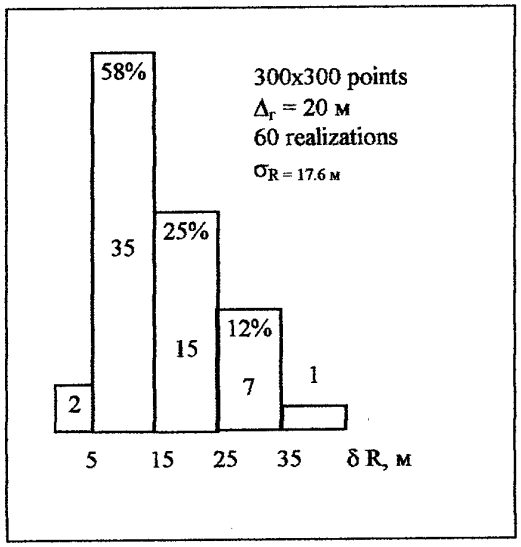


Fig.8. The histogram of the error distribution.

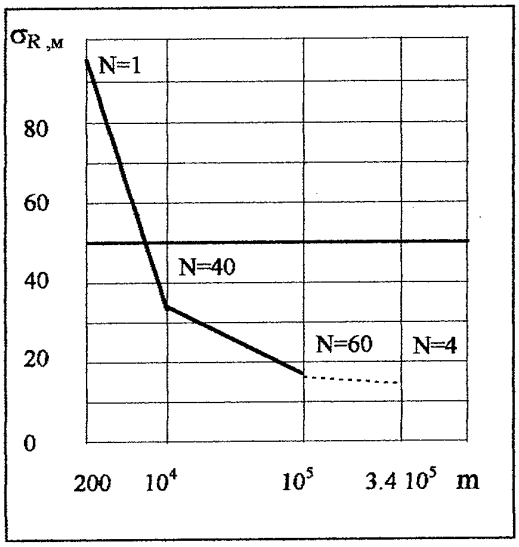


Fig.9. The dependence of accuracy on the number of points processed.

algorithm for Δ_h=10 m led to the following errors of determining the aircraft coordinates: δR=0 m; 15 m; 16 m; 11 m.

Fig . 9 presents the final graph of dependence of mean square accuracy σ_R on the number of the processed points m. The mean square accuracy has been calculated on the basis of N realizations of the algorithm.

Fig 10-11 show the results of the study for the course error and the errors of height measurement, respectively.

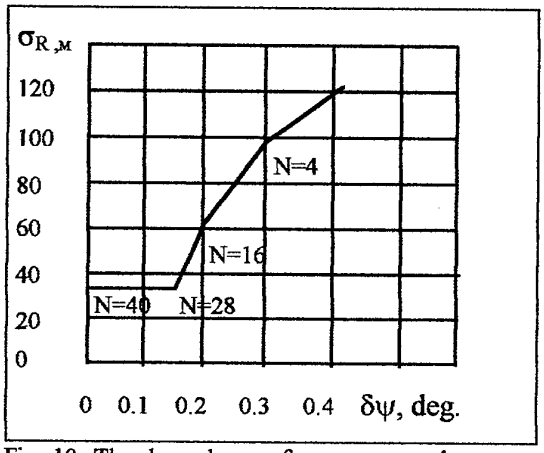


Fig. 10. The dependence of accuracy on the course error.

The obtained results show that the stereoscopic navigation systems allow us to obtain estimation of the aircraft coordinates with accuracy of 10-15 meters.

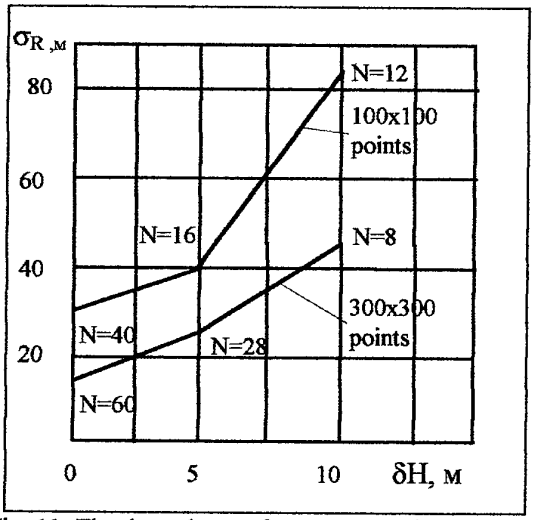


Fig. 11. The dependence of accuracy on the error of height measurement.

It is necessary to note that the mentioned accuracy (about 10-15 m) of determination of the aircraft coordinates was obtained for the step of image digitization on terrain equal to 2.5-3 m. The resolution of 2.5 m is not a limit for the modern level of development of optic - electronic iconic systems. The given limitation is connected only with technological difficulties of experimentation. Therefore, the processing of a greater number of points on stereo images and the achievement of a higher accuracy of stereoscopic navigation systems are possible.

A STEREOSCOPIC OBSERVATION SYSTEM

Stereoscopic observation allows to automize procedure of detecting and identification of small objects .

Correct digital model of observed area describing brightness and size of this area is suggested to be available. Presence of object changes the value of brightness and size in the local range and this is what object indications are. There is some difficulty in value describing of background and object brightness because of its changeability under many factors. Besides, possibilities of application of brightness indication may be reduced greatly in result of masking. Stereo-photographing allows to use brightness as object indication and 3-dimensial form as well.

5. Equation of stereo observation

Let`s have a look at the j-th stereo images cutting (fig. 12 ). Let`s divide the whole field of view into elementary angels which respect to instantaneous field of view of single sensor.

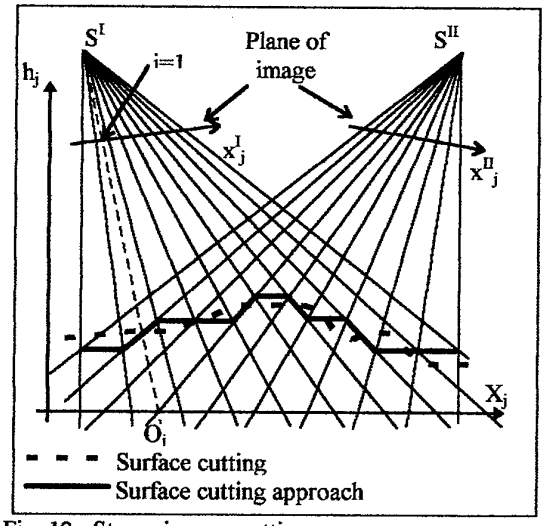


Fig. 12. Stereo images cutting

Beams of the first and the second images defining these angels make the net. From brightness values L_{ji} the vector

$$c_j = [L_{j1}, L_{j2}, \dots, L_{jJ_j}]^T,$$

may be composed , J_j - quantity of elements in j-th cutting .

The signal values f^{I}_{ji} and f^{II}_{ji} in j-th slide combine in vectors.

Let`s combine vectors of the first and the second images of stereo-pairs into common vectors of measures Z and errors δZ:

$$Z^{I}_j = [\, f^{I}_{j1} \quad f^{I}_{j2} \quad \dots \quad f^{I}_{jN^{I}_j}\,]^T,$$

$$Z^{II}_j = [\, f^{I}_{j1} \quad f^{I}_{j2} \quad \dots \quad f^{I}_{jN^{II}_j}\,]^T .$$

$$Z = [Z^{I\,T}, Z^{II\,T}]^T, \quad \delta Z = [\delta Z^{I\,T}, \delta Z^{II\,T}]^T,$$

The equation of stereo-observation may be represented as:

$$Z = F\,c + \delta Z$$

where matrix F depends on object 3-D form.

6. Algorithm of stereoscopic detecting and identification

Hypotheses D_ν , $\nu = 1..I$ of observed object are examined. If we have brightness and 3-D object form model, choosing rule of hypotheses D_μ is as $D = D_\mu$,

$$\mu : \arg\inf_{\nu} \| Z - F_\nu m_\nu \|^2_{(R + F_\nu G_\nu F^T_\nu)^{-1}} - 2 \ln p_\nu ,$$

where m_ν is the expectation and G_ν is the covariance matrix of hypothetical ($D\nu$) brightness distribution, p_ν is an a priori hypotheses probability .

If we have only 3-D form model and $R = \sigma_{\sim} E$, choosing rule is as $D = D_\mu$,

$$\mu : \arg\inf_{\nu} \sigma^{-2}_{\sim} \| Z \|^2_{E - F_\nu (F^T_\nu F_\nu)^{-1} F^T_\nu} - 2 \ln p_\nu .$$

Matrix F_ν correspond to hypothesis D_ν.

7. Stereoscopic detection error probability

We obtain

$$P_{err} = p_0 P_{1,0} + p_1 P_{0,1}$$

for detection error probability P_{err} , where

$$P_{1,0} = \frac{1}{\sqrt{2\pi}\,\sigma_0} \int_{-\infty}^{0} \exp\left(-\frac{(x - m_0)^2}{2\sigma^2_0}\right) dx ;$$

$$P_{0,1} = \frac{1}{\sqrt{2\pi}\,\sigma_1} \int_{-\infty}^{0} \exp\left(-\frac{(x - m_1)^2}{2\sigma^2_1}\right) dx ;$$

$$m_0 = 2 N \sigma^2_f [1 - K_f(\Delta p, \rho_f)] - 2 \sigma^2_{\sim} \ln\frac{p_1}{p_0} ;$$

$$\sigma^2_0 = 16 N \sigma^2_{\sim} \sigma^2_f [1 - K_f(\Delta p, \rho_f)] + 8 N \sigma^4_{\sim} + 4 \frac{L_x}{\Delta} \frac{\Delta p}{\Delta} \sigma^4_{\sim} ;$$

$$m_1 = \frac{L_y}{\Delta} \left[\frac{\Delta p}{\Delta} (\sigma^2_f + m^2_o) + \frac{2L_x - \Delta p}{\Delta} \sigma^2_o - 2 \frac{L_x - \Delta p}{\Delta} \sigma^2_o K_o(\Delta p, \rho_o) \right] - 2 \sigma^2_{\sim} \ln\frac{p_0}{p_1} ;$$

$$\sigma^2_1 = 8 \sigma^2_{\sim} \frac{L_y}{\Delta} \left[\frac{\Delta p}{\Delta} (\sigma^2_f + m^2_o) + \frac{2L_x - \Delta p}{\Delta} \sigma^2_o - - 2 \frac{L_x - \Delta p}{\Delta} \sigma^2_o K_o(\Delta p, \rho_o) \right] + 8 N \sigma^4_{\sim} + 4 \frac{L_x}{\Delta} \frac{\Delta p}{\Delta} \sigma^4_{\sim} ;$$

$$N = \frac{L_x}{\Delta} \frac{L_y}{\Delta} ; \quad \Delta p = h \frac{B}{H} ;$$

Δ is a ground resolution of iconic system; p_0 and p_1 are a priori probability of an object absence and presence, respectively; L_x, L_y - object size , h is an object height, B is a stereo base , H is an observation height, σ_f , σ_o are variances of ground and object surface brightness, $\sigma_{\sim}$ is a noise variance, m_o , m_f is an expectation of object and ground brightness; $K_f(\Delta p, \rho_f)$ and $K_o(\Delta p, \rho_o)$ are auto- correlation functions, correspond to ground and object surface brightness; ρ_f and ρ_o are radiuses of correlation. Ground surface brightness, object surface brightness and noise are uncorrelated together .

Fig. 13-14 shows the probability of small object detection ($N < 2000$).
Fig. 13 correspond to essential difference between ground brightness and object brightness ($\frac{m_o - m_f}{\sigma_{\sim}} > 1$) and illustrate the error probability P_{err} depending on $\frac{m_o - m_f}{\sigma_{\sim}}$ and N. Another conditions are:

$$\sigma_o = 0, \quad \sigma_f = \frac{m_o - m_f}{2}, \quad \rho_o = \rho_f = 2\,\Delta.$$

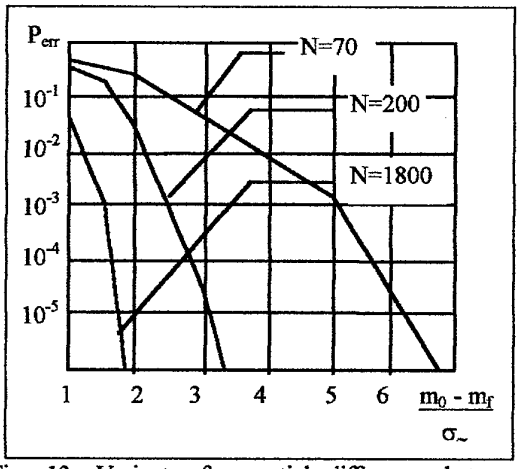


Fig. 13. Variant of essential difference between ground brightness and object brightness

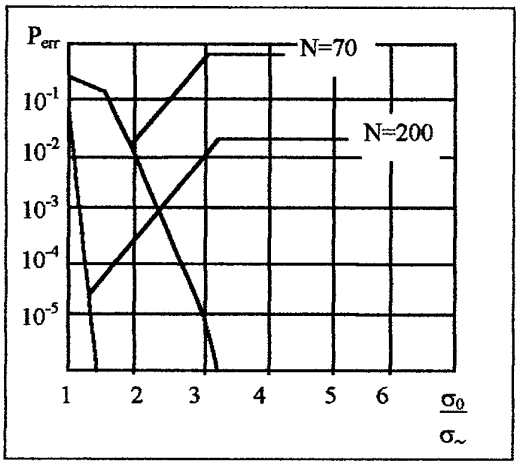


Fig. 14. Variant of small difference between ground brightness and object brightness

Fig. 14 correspond to very small difference between ground brightness and object brightness: $m_o - m_f = 0$, $\sigma_0 = \sigma_f$ and it illustrate P_{err} depending on $\sigma_o / \sigma_\sim$ and N.

References

1. Par Christophe Garcia, Saida Bonakas, Denis Vandorpe. Reseau de hopfield, un outil pour la mise en correspondance. Traitement du Signal 1995, Volume 12, no 2.

2. Hu Dong Chang, Kwang Ick Kim, Timothy Poston. An accurate 3D localization of quick-mark, Pattern recognition Lettes, 16, 1995.

3. Huber E. Object Tracking with stereo vision, American Institute of Aeronautics and Astronautics, 1993.

4. L. Raffo. Adaptive resistive network for stereo depth estimation, Electronics letters 26th October 1995, Vol. 31 No. 22.

5. J.M. Cruz, G. Pojares, J. Aranda, J.L.F. Vindel. Stereo matching technicque based on the perception criterion function, Pattern Recognition Lettes, 16, 1995.

6. Navigation Stereo Video System for Helicopters, Moscow: VINITI (Translated from Defense Electronics, 1995, Vol.27, No.2).

7. Korobkov, N.V., Design of Methods of Processing in Optical Navigation Systems with the Use of the Stereo Vision Principle, Cand. Sci (Tech) Dissertation, Moscow: Moscow Physical Engineering Inst., 1996.

8. Lobanov, A.N., Fotogrammetriya (Photogrammetry), Moscow: Nedra, 1984.

9. Beloglazov I.N., Kazarin S.N., A Stereoscopic Navigation System. Journal of Computer and Systems Sciences International, Vol.36, No.6, 1997, p.p. 846-868.

10. Krasovskii, A.A., Beloglazov I.N., Chigin, G.P., Teoriya korrelyatsionno-ekstremalnykh navigatsionnykh sistem (Theory of Correlation Extremal Navigation Systems), Moscow: Nauka, 1979.

11. Beloglazov I.N., Dzhandzhgava, G.I., Chigin, G.P., Osnovy navigatsii po geofisicheskim polyam (Principles of Navigation by Geophysical Fields), Krasovskii, A.A., Ed., Moscow: Nauka, 1985.

12. Beloglazov I.N., Recursive Search Estimation Algorithms, Dokl. Akad. Nauk SSSR, 1977, Vol.236, No.2.

13. Beloglazov I.N., Kazarin S.N. Joint Optimal Estimation, Identification, and Hypothesis Testing in Discrete Dynamic Systems. Journal of Computer and Systems Sciences International, Vol.37, No.4, 1998, p.p. 534-550.

An Integrated Navigation System for Offshore Platform Coordinates Supervision

Ch. M. Hajiyev* and **F. Caliskan****

* Istanbul Technical University Aeronautics and Astronautics Engineering

** Istanbul Technical University Electrical Engineering
Ayazaga, Istanbul, 80626, Turkey
e-mail: caliskan@elk.itu.edu.tr

Abstract

In this paper, a multichannel Kalman filter design to integrate different measurement systems for offshore platform (OP) coordinates supervision is presented. The complete OP motion is assumed to be composed of the low-frequency motion caused by the wind and undercurrent, and the high frequency motion caused by the sea-way. The mathematical model of the low-frequency OP motion is given by the normal differential equations system, and the high-frequency OP motion is represented by the multivariable autoregression model with the sliding mean. The parameter estimation problem for the model of the low-frequency OP motion, on which the in-service control is performed, is solved through two jointly operating Kalman filters: the first one is used for the estimation of the parameters of the low-frequency motion, and the second one is employed for the high-frequency one. The parameters of the first filter are automatically adapted to variations of the second filter, i.e. they are adapted to variations of the sea. Two algorithms for the OP motion parameter estimation (parallel and with preliminary data compression) which are employed for several measuring channels data estimation, are developed, and simulated on a computer.

1. Introduction

The need for oil and natural gas requires the techniques for their searching and extracting in deep seas. These techniques have yielded particular technical equipments. In recent years, hundreds of floating offshore platforms (OP's) have been designed in various countries. The stabilization of OP's at the point on the exploratory well is very important during drilling in sea. For this purpose, the coordinates of OP's should be measured accurately. The coordinates of OP's are generally measured by the following measurement systems;

- Hydroacoustic navigation system,
- Inclinometric navigation system,
- Dynamometer navigation system,
- Radiotechnique navigation system, and
- Satellite radionavigation system.

The operational principles and accuracy analyses of the above systems have been presented in [1-4]. In [5], a Kalman filter has been used to estimate the state vector of an OP, but this filter uses the measurements from one navigation system only. The coordinates of the OP can only be obtained by more than one navigation systems in real environment.

When the above measurement systems are used all together, the coordinates of the OP may be obtained more accurately. In this paper, a multichannel Kalman filters design to integrate the above measurement systems, is presented.

2. Adaptive filter algorithm for output coordinates of the OP

Researches have shown that, an OP is a dynamic object with multiconnection and nonlinear vibration. This object is subject to periodical and nonstationary random forces. The optimal control

Paper presented at the RTO SCI International Conference on "Integrated Navigation Systems", held at "Elektropribor", St. Petersburg, Russia, 24-26 May 1999, and published in RTO MP-43.

of such an object highly depends on accurate optimal estimation of the state vector. In this section, a solution for the problem of optimal estimation of the OP coordinates is addressed.

The complete OP motion is assumed to be composed of the low-frequency motion caused by the wind and under current, and the high-frequency motion caused by the waves. The mathematical model of the low-frequency OP motion is given by the normal differential equations system as,

$$\begin{aligned}(\frac{D}{g}+\mu_{11})\ddot{\xi}+\nu_{11}\dot{\xi}&=F_{\xi}+P_{\xi}+\varepsilon_{\xi}\\(\frac{D}{g}+\mu_{22})\ddot{\eta}+\nu_{22}\dot{\eta}&=F_{\eta}+P_{\eta}+\varepsilon_{\eta}\end{aligned}\qquad(1)$$

where
D- weight of the platform
g- gravitational acceleration
μ_{11}, μ_{22}- coefficients of united masses
ν_{11}, ν_{22}- coefficients of friction
ξ, η- shift of OP along Ox, and Oy axes
F_{ξ}, F_{η}- forces by the chains (F_{ξ}, F_{η} are nonlinear with respect to coordinates)
P_{ξ}, P_{η}- external forces by wind and undercurrent
ε_{ξ}, ε_{η}- random effects of the wind with zero means and known correlation matrices.

The high-frequency OP motion is represented by a moving, multivariable autoregression model [5,6],

$$\mathbf{A}(q^{-1})\,\mathbf{y}(k)=\mathbf{C}(q^{-1})\,\mathbf{e}(k)\qquad(2)$$

where
$\mathbf{y}(k)$- wave perturbance
$\mathbf{e}(k)$- white noise with zero mean
$\mathbf{A}(q^{-1})=\mathbf{I}+\mathbf{A}_1q^{-1}+.....\mathbf{A}_{na}q^{-na}$
$\mathbf{I}$-unit matrix
$\mathbf{C}(q^{-1})=\mathbf{C}_1q^{-1}+.....\mathbf{C}_{nc}q^{-nc}$, na and nc are known
q^{-1}- delay operator, $q^{-1}y(k)=y(k-1)$.

Measurement equation is,

$$\mathbf{Z}(k)=\mathbf{l}(k)+\mathbf{y}(k)+\mathbf{h}(k)\qquad(3)$$

where

$\mathbf{l}^T(k)=[\xi_l(k),\ \eta_l(k)]$- low-frequency shift of the OP,

$\mathbf{h}(k)$- measurement noise with zero mean and known correlation matrix, i.e.;

$$E\{\mathbf{h}(k)\}=0,\ \text{and}\ E\{\mathbf{h}(k)\mathbf{h}^T(j)\}=D_h\,\delta(kj)$$

where

$\delta(kj)$- Kronecker delta function. Measurement noise has Gaussian distribution.

The information about the high-frequency motion parameters is not useful for control, because the chains have large inertia, and so the high-frequency motion cannot effectively controlled. Therefore, the low-frequency motion to be controlled is estimated in these circumstances [7].

The estimation algorithm is automatically adapted to the changes of the high-frequency motion parameters as well.

The optimization criteria is to minimize the standard error. Let the equations in (1) be stated by finite differences, and the state vector of the OP be shown by $\mathbf{U}$. The equations in (1) are linearized by applying quasi-linearization procedure,

$$\begin{aligned}\mathbf{U}(k)&=f\left[\hat{\mathbf{U}}(k-1),\ \hat{\varepsilon}(k-1)\right]+\\&\mathbf{F}_U\left[\mathbf{U}(k-1)-\hat{\mathbf{U}}(k-1)\right]+\mathbf{v}(k-1)\end{aligned}\qquad(4)$$

where

$$\mathbf{F}_U=\left[\frac{\partial f}{\partial \mathbf{U}}\right]_{\hat{\mathbf{U}}(k-1),\hat{\varepsilon}(k-1)},\ \text{and}$$

$$\mathbf{v}(k-1)=\left[\frac{\partial f}{\partial \varepsilon}\right]_{\hat{\mathbf{U}}(k-1),\hat{\varepsilon}(k-1)}\varepsilon(k-1)$$

$\varepsilon^T(k-1)=[\varepsilon_{\xi}(k-1),\ \varepsilon_{\eta}(k-1)]$.

The measurement equation of the low-frequency OP motion is expressed as,

$$\mathbf{Z}_l(k)=\mathbf{H}(k)\mathbf{U}(k)+\mathbf{h}_l(k)\qquad(5)$$

where
$\mathbf{H}(k)$- system measurement matrix
$\mathbf{h}_l(k)$- measurement error vector.

The low-frequency motion equation may be obtained from,

$$\mathbf{Z}_l(k)= \mathbf{Z}(k)-\hat{\mathbf{y}}(k) \qquad (6)$$

where

$\hat{\mathbf{y}}(k)$- estimated value of the wave perturbance y(k). y(k) is estimated by a Kalman filter designed for evaluating the following autoregression process with moving mean.

Using (6), it can be said that $\mathbf{h}_l(k)$ has the Gaussian distribution with zero mean and with the correlation matrix,

$$\mathbf{D}_{hl}(k)=\mathbf{D}_h(k)+\mathbf{D}_{\hat{y}}(k) \qquad (7)$$

where

$\mathbf{D}_{\hat{y}}(k)$- correlation matrix of the error for the high-frequency OP motion.

The use of Bayes method that has no restriction on linearity and being stationary of the system, is suitable for estimation of the low-frequency OP motion. A posteriori probability density of the estimated parameters is computed by the Bayes formula as,

$$p[U(k)/\mathbf{Z}_l^k]=p[U(k)/\mathbf{Z}_l^{k-1},\mathbf{Z}_l(k)]= \qquad (8)$$
$$\{p[U(k)/\mathbf{Z}_l^{k-1}]p[\mathbf{Z}_l^k/U(k), \mathbf{Z}_l^{k-1}]\}/p[\mathbf{Z}_l(k)/\mathbf{Z}_l^{k-1}]$$

where

$\mathbf{Z}_l^k=\{\mathbf{Z}_l(1),.....\mathbf{Z}_l(k)\}$, $\mathbf{Z}_l^{k-1}=\{\mathbf{Z}_l(1),.....\mathbf{Z}_l(k-1)\}$.

The distribution $p[U(k)/\mathbf{Z}_l^{k-1}]$ is assumed to be approximated by the Gaussian distribution with the mathematical expectation vector $f[\hat{U}(k-1)]$ and the correlation matrix $\mathbf{M}(k)$, i.e.;

$$p[U(k)/\mathbf{Z}_l^{k-1}]=\frac{1}{(2\pi)^{\frac{n}{2}}\sqrt{\det \mathbf{M}(k)}}\exp[-1/2\{$$
$$\{U(k)-f[\hat{U}(k-1)]\}^T\mathbf{M}^{-1}(k)\{U(k)-f[\hat{U}(k-1)]\} \qquad (9)$$

where

$f[\hat{U}(k-1)]$- extrapolation value for one step
n- dimension of $\mathbf{U}(k)$,
$\mathbf{M}(k)$- correlation matrix of extrapolation error, i.e.;

$$\mathbf{M}(k)=\mathbf{F}_U\mathbf{P}(k-1)\mathbf{F}_U^T+\mathbf{D}_v(k-1) \qquad (10)$$

where

$\mathbf{P}(k-1)$- correlation matrix of value error at previous step.
$\mathbf{D}_v(k-1)$- correlation matrix of wind perturbance

As mentioned earlier, since $\mathbf{h}_l(k)$ is white Gaussian noise, $\mathbf{Z}_l(k)$ in $\mathbf{U}(k)$ does not depend on $\mathbf{Z}_l^{k-1}$. Therefore the following likelihood function can be written,

$$p[\mathbf{Z}_l(k)/U(k),\mathbf{Z}_l^{k-1}]=p[\mathbf{Z}_l(k)/U(k)] \qquad (11)$$

Using (5),

$p[\mathbf{Z}_l(k)/U(k)]=\mathcal{N}\{\mathbf{H}(k)U(k), \mathbf{D}_{hl}(k)\}$, or

$$p[\mathbf{Z}_l(k)/U(k)]=\frac{1}{2\pi\sqrt{\det \mathbf{D}_{hl}(k)}}\exp[-1/2\{$$
$$\{\mathbf{Z}_l(k)-\mathbf{H}(k)U(k)]\}^T\mathbf{D}_{hl}^{-1}(k)\{\mathbf{Z}_l(k)-\mathbf{H}(k)U(k)]\} \qquad (12)$$

Since there is no correlation between the state vector $\mathbf{U}(k)$ and the measurement noise $\mathbf{h}_{lk}$, using (5),the following can be written,

$$E\{\mathbf{Z}_l(k)/\mathbf{Z}_l^{k-1}\}=E\{\mathbf{H}(k)\mathbf{U}(k)+\mathbf{h}_l(k)/\mathbf{Z}_l^{k-1}\}$$
$$=\mathbf{H}(k)\, f[\hat{\mathbf{U}}(k-1)]$$

$$E[\{\mathbf{Z}_l(k)-\mathbf{H}(k)f[\hat{U}(k-1)]\}\{\mathbf{Z}_l(k)-\mathbf{H}(k)f[\hat{U}(k-1)]\}^T$$

$$=\mathbf{H}(k)\mathbf{M}(k)\mathbf{H}^T(k)+\mathbf{D}_{hl}(k)$$

Hence the denominator of (8) can be expressed as,

$$p[\mathbf{Z}_l(k)/\mathbf{Z}_l^{k-1}]=\frac{1}{2\pi\sqrt{\det\left[\mathbf{H}(k)\mathbf{M}(k)\mathbf{H}^T(k)+\mathbf{D}_{hl}(k)\right]}}$$
$$\exp[-1/2\{\mathbf{Z}_l(k)-\mathbf{H}(k)f[\hat{\mathbf{U}}(k-1)]\}^T$$
$$[\mathbf{H}(k)\mathbf{M}(k)\mathbf{H}^T(k)+\mathbf{D}_{hl}(k)]^{-1}\{\mathbf{Z}_l(k)-\mathbf{H}(k)f[\hat{\mathbf{U}}(k-1)]\}] \qquad (13)$$

(9), (12), and (13) are substituted in the a posteriori probability density $p[U(k)/\mathbf{Z}_l^k]$ in (8), and necessary arrangements are done to find:

$$p[U(k)/\mathbf{Z}_l^k]=\frac{1}{(2\pi)^{\frac{n}{2}}\sqrt{\det \mathbf{P}(k)}}\exp[-1/2\{$$
$$\{[U(k)-\hat{\mathbf{U}}(k)]^T\mathbf{P}^{-1}(k)[U(k)-\hat{\mathbf{U}}(k)]\}$$

where

$$\hat{\mathbf{U}}(k)=f[\hat{\mathbf{U}}(k-1)]+\mathbf{P}(k)\mathbf{H}^T(k)\mathbf{D}_{hl}^{-1}(k)\{$$
$$\{\mathbf{Z}_l(k)-\mathbf{H}(k)\, f[\hat{\mathbf{U}}(k-1)]\},$$

$$\mathbf{P}(k)=\mathbf{M}(k)-\mathbf{M}(k)\,\mathbf{H}^T(k)$$
$$[\mathbf{D}_{hl}(k)+\mathbf{H}(k)\mathbf{M}(k)\mathbf{H}^T(k)]^{-1}\,\mathbf{H}(k)\mathbf{M}(k)$$

As the optimization criteria is to minimize the standard error, the optimum value is the conditional mathematical expectation of the a posteriori distribution, and value accuracy is characterized by the correlation matrix of the distribution. Taking into account the conditional distribution $p[\mathbf{U}(k)/\mathbf{Z}_l^k]$ is a Gaussian distribution, the following recursive filtering algorithm is obtained to estimate the low-frequency OP motion,

$$\hat{\mathbf{U}}(k)=f[\hat{\mathbf{U}}(k-1)]+\mathbf{P}(k)\mathbf{H}^T(k)\mathbf{D}_{hl}^{-1}(k)\{$$
$$\{\mathbf{Z}_l(k)-\mathbf{H}(k)\,f[\hat{\mathbf{U}}(k-1)]\},$$

$$\mathbf{P}(k)=\mathbf{M}(k)-\mathbf{M}(k)\,\mathbf{H}^T(k)$$
$$[\mathbf{D}_{hl}(k)+\mathbf{H}(k)\mathbf{M}(k)\mathbf{H}^T(k)]^{-1}\,\mathbf{H}(k)\mathbf{M}(k)$$

$$\mathbf{M}(k)=\mathbf{F}_U\,\mathbf{P}(k-1)\mathbf{F}_U^T+\mathbf{D}_v(k-1) \qquad (14)$$

As mentioned earlier, the high-frequency motion of the OP is predicted by a moving-mean autoregression model. The high-frequency OP motion is separated into channels, and the parameters of any i-th channel are evaluated independently.

Assume that,

$$\varphi_i^T(k)=[-y_i(k-1),.....-y_i(k-na),\ e_i(k-1),....e_i(k-nc)],$$

$$\theta_i^T(k)=[a_{i1},\ a_{i2},.....a_{ina},\ c_{i1},c_{i2},.....c_{inc}]$$

For the i-th channel (2) can be rewritten as,

$$y_i(k)=\varphi_i^T(k)\theta_i(k)+e_i(k)$$
$$\theta_i(k)=\theta_i(k-1) \qquad (15)$$

The above linear system takes into account features of the sea-waves, and is controlled by the white noise. For such a system, Kalman filter is employed, and the recursive prediction equations may be written as [8],

$$\hat{\theta}_i(k)=\hat{\theta}_i(k-1)+\mathbf{K}_i(k)[y_i(k)-\varphi_i^T(k)\hat{\theta}_i(k-1)] \qquad (16)$$

$$\mathbf{K}_i(k)=\gamma_i(k-1)\varphi_i(k)/[\sigma_0^2+\varphi_i^T(k)\gamma_i^T(k-1)\varphi_i(k)] \qquad (17)$$

$$\gamma_i(k)=\gamma_i(k-1)-\mathbf{K}_i(k)\varphi_i^T(k)\gamma_i(k) \qquad (18)$$

where

$\mathbf{K}(k)$- Kalman filter gain matrix,
$\gamma_i(k)$- correlation matrix of the estimation error,
σ_0^2- variance of independent unobserved noise.

The parameter estimation problem for the model of the low-frequency OP motion, on which the in-service control is performed, is solved through two jointly operating Kalman filters: the first one is used for the estimation of the parameters of the low-frequency motion, and the second one is employed for the high-frequency one. The parameters of the first filter are automatically adapted to variations of the second filter, i.e. they are adapted to variations of the sea.

3. Parallel, multichannel Kalman filter for the estimation of the OP coordinates

In this section, to estimate the OP coordinates a filtering algorithm that is based on different operational principles and that uses a couple of measurement channel, is derived.

Let the OP coordinates be observed by a N-channel navigation system. The measurement equation of the j-th measurement channel is given as,

$$\mathbf{Z}_j(k)=\mathbf{H}_j(k)\mathbf{U}(k)+\mathbf{y}(k)+\mathbf{h}_j(k)$$

where
$\mathbf{h}_j(k)$- measurement noise of the j-th channel with zero mean and the correlation matrix,

$$E[\mathbf{h}_j(k)\mathbf{h}_j^T(m)]=\mathbf{D}_{hjj}(k)\delta(km)$$

Furthermore, the correlations among the channels are zero. The measurement of the low-frequency OP shift for the j-th measurement channel is,

$$\mathbf{Z}_{lj}(k)=\mathbf{Z}_j(k)-\hat{\mathbf{y}}^j(k)$$

where

$$\hat{\mathbf{y}}^j(k)=[\hat{\mathbf{y}}_1^j(k)\ \hat{\mathbf{y}}_2^j(k)]^T$$

the value of the high-frequency motion according to the j-th measurement channel.

To find the multichannel filtering algorithm for the low-frequency motion, $\mathbf{Z}_l(k)$, $\mathbf{H}(k)$, and $\mathbf{D}_{hl}$ are replaced by,

$$\mathbf{Z}_l(k)=[\mathbf{Z}_{l1}^T(k)\vdots\mathbf{Z}_{l2}^T(k)\vdots.....\vdots\mathbf{Z}_{lN}^T(k)]^T$$

$$\mathbf{H}(k)=[\mathbf{H}_1^T(k)\vdots\mathbf{H}_2^T(k)\vdots.....\vdots\mathbf{H}_N^T(k)]^T \text{ and}$$

$$\mathbf{D}_{hl}(k)=\text{diag}[\mathbf{D}_{hl11}(k),\mathbf{D}_{hl22}(k).....\mathbf{D}_{hlNN}(k)]$$

where

$$\mathbf{D}_{hljj}(k)=\mathbf{D}_{hjj}(k)+\mathbf{D}_{\hat{\mathbf{y}}\,j}(k)$$

$$\mathbf{D}_{\hat{\mathbf{y}}\,j}(k)=\text{diag}[\hat{\varphi}_1^{jT}(k)\gamma_1^j(k)\hat{\varphi}_1^j(k),\hat{\varphi}_2^{jT}(k)\gamma_2^j(k)\hat{\varphi}_2^j(k)$$

$$\hat{\varphi}_i^j(k)=[-y_i^j(k-1),....-y_i^j(k-na),\hat{\mathbf{e}}_i^j(k-1).....\hat{\mathbf{e}}_i^j(k-nc)]$$

After tidying up the above equations, the multichannel filtering algorithm is expressed as,

$$\hat{\mathbf{U}}(k)=f[\hat{\mathbf{U}}(k-1)]+\sum_{j=1}^{N}\mathbf{P}(k)\mathbf{H}_j^T(k)\mathbf{D}_{hljj}^{-1}(k)\{\{\mathbf{Z}_{lj}(k)-\mathbf{H}_j(k)f[\hat{\mathbf{U}}(k-1)]\} \quad (19)$$

$$\mathbf{P}(k)=\mathbf{M}(k)-\{[\mathbf{M}^{-1}(k)+\sum_{j=1}^{N}\mathbf{H}_j^T(k)\mathbf{D}_{hljj}^{-1}(k)\mathbf{H}_j(k)]^{-1}\sum_{j=1}^{N}\mathbf{H}_j^T(k)\mathbf{D}_{hljj}^{-1}(k)\mathbf{H}_j(k)\}\mathbf{M}(k) \quad (20)$$

As seen in (19), the optimal values of the low-frequency OP motion is obtained by weighted sum of the variables $\mathbf{Z}_{lj}(k)-\mathbf{H}_j(k)f[\hat{\mathbf{U}}(k-1)]$ of all channels. Such an algorithm is called parallel, and is used when the measurements from different channels are available synchronously.

4. Multichannel Kalman filter based on preliminary data compression for the estimation of the OP coordinates

The number of the filters may be reduced to minimum if the measurements of different channels are compressed at first. In this case, nonrecursive maximum likelihood values should be found at each measurement step by using the measurements of all channels. Thus, for evaluation of the OP coordinates, only a measurement vector obtained by a linear transformation of different measurement channel vectors, is used, i.e.,

$$\mathbf{y}(k)=\mathbf{Z}(k)-\mathbf{H}(k)\mathbf{U}(k),\ \mathbf{Z}_l(k)=\mathbf{Z}(k)-\hat{\mathbf{y}}(k)$$

where

$$\mathbf{Z}(k)=[\sum_{j=1}^{N}\mathbf{D}_{hjj}^{-1}(k)]^{-1}\sum_{j=1}^{N}\mathbf{D}_{hjj}(k)\mathbf{Z}_j(k) \quad (21)$$

It is assumed that,

$$\mathbf{H}_1(k)=\mathbf{H}_2(k)=.....=\mathbf{H}_N(k)=\mathbf{H}(k)$$

The measurement error variance of the low-frequency motion is,

$$\mathbf{D}_{hl}(k)=[\sum_{j=1}^{N}\mathbf{D}_{hjj}^{-1}(k)]^{-1}+\mathbf{D}_{\hat{\mathbf{y}}}(k) \quad (22)$$

The above transformation is called preliminary data compression. Hence, the number of filters is reduced to minimum, and the OP coordinates are found only by means of an adaptive filter comprising low and high- frequency filters. In this case, $\mathbf{Z}(k)$ and $\mathbf{D}_{hl}$ parameters are replaced by (21) and (22) respectively.

5. Simulations

For the purpose of simulation, the equations in (1) can be rewritten by means of finite difference approximation as,

$$\xi_i=\xi_{i-1}+\Delta t u_{i-1}$$

$$\eta_i=\eta_{i-1}+\Delta t w_{i-1}$$

$$u_i=\left(1-\frac{\Delta t\nu_{11}}{\frac{D}{g}+\mu_{11}}\right)u_{i-1}+\frac{\Delta t}{\frac{D}{g}+\mu_{11}}F_{\xi_{i-1}}+\frac{\Delta t}{\frac{D}{g}+\mu_{11}}P_{\xi_{i-1}}+\frac{\Delta t}{\frac{D}{g}+\mu_{11}}\varepsilon_{\xi_{i-1}}$$

$$w_i = \left(1 - \frac{\Delta t v_{22}}{\frac{D}{g} + \mu_{22}}\right) w_{i-1} + \frac{\Delta t}{\frac{D}{g} + \mu_{22}} F_{\eta_{i-1}} + \frac{\Delta t}{\frac{D}{g} + \mu_{22}} P_{\eta_{i-1}} + \frac{\Delta t}{\frac{D}{g} + \mu_{22}} \varepsilon_{\eta_{i-1}}$$

where Δt is time increment.

The designed Kalman filter is applied to above equations. The simulation results are given in Fig. 1-5.

In the simulations, three sorts of measurement systems have been considered;

1. Hydroacoustic measurement system of which the noise variance is 0.1,
2. Inclinometric measurement system of which the noise variance is 0.26, and
3. Dynamometer measurement system of which the noise variance is 0.72.

These systems are integrated in two forms based on Kalman filtering:

a. Multichannel preliminary data compression Kalman filter, and
b. Parallel multichannel Kalman filter.

In the first form, the variance of the integrated measurement system is calculated as 0.00656 using the maximum likelihood method for the mentioned measurement systems. In the second form, the above measurement systems are integrated through a parallel multichannel Kalman filter which is based on the parallel processing of the three Kalman filters.

The simulation results of the Kalman filter based on hydroacoustic, inclinometric, and dynamometer measurement systems are illustrated in Fig.1, Fig.2, and Fig. 3 respectively.

The first graph in each figure shows the actual and estimated values of the low-frequency OP motion along Ox axis. The error between the actual and estimated values is shown in the second graph. In the third graph, the error variance of the Kalman filter is given. The similar results are presented in Fig.4 and Fig. 5 for multichannel preliminary data compression Kalman filter and parallel multichannel Kalman filter respectively.

The execution time per cycle for each filter, and the variances of the filter values at different iterations (25, 50, 75, 100) are given in the table.

As the simulations imply, the parallel multichannel Kalman filter has the highest accuracy, but it requires longer execution time. The multichannel preliminary data compression Kalman filter has the second highest accuracy, and takes shorter time in comparison with the parallel multichannel Kalman filter.

The simulation results conclude that the OP coordinates could be estimated more accurately when the measurement systems are integrated through the Kalman filter.

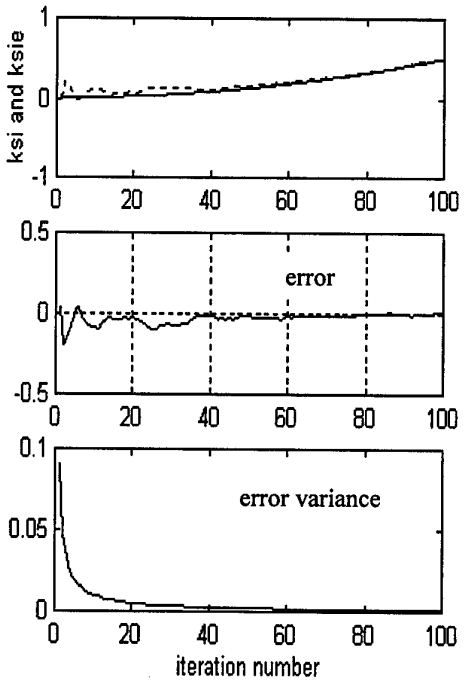


Fig. 1 Results by using Kalman filter based on Hydroacoustic measurement system

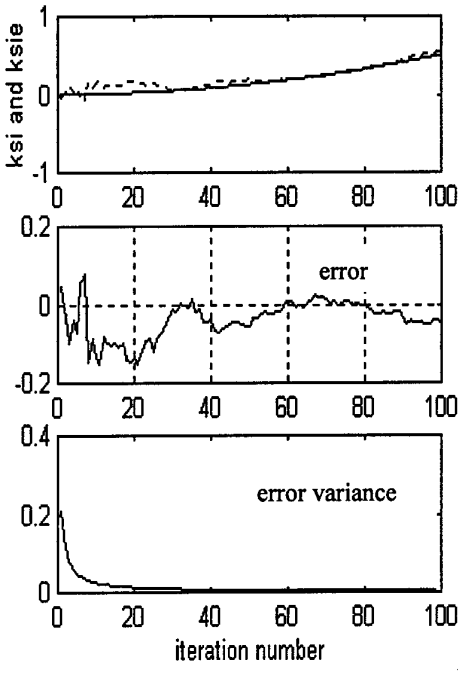


Fig. 2 Results by using Kalman filter based on inclinometric measurement system

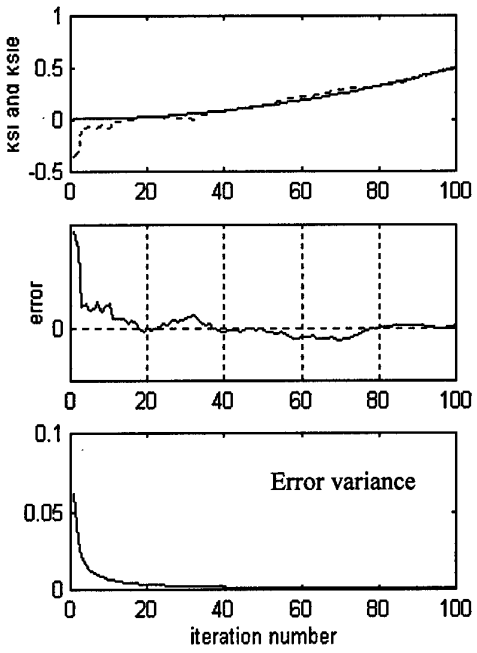


Fig. 4 Results by using preliminary data compression multichannel Kalman filter

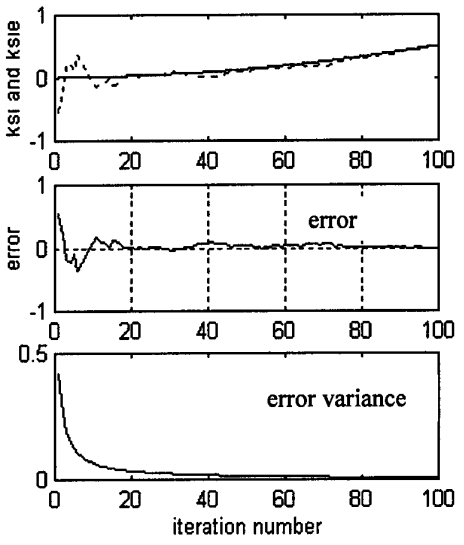


Fig. 3 Results by using Kalman filter based on dynamometer measurement system

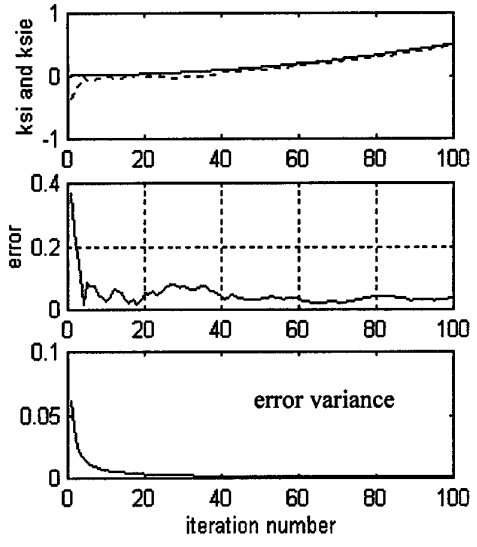


Fig. 5 Results by using parallel multichannel Kalman filter

TABLE

Measurement Syst.	Time for one iter. (sec)	Covariance Matrix of Value Error			
		P(25)	P(50)	P(75)	P(100)
I	0.005	0.004	0.002	0.0013	0.0010
II	0.005	0.010	0.005	0.0035	0.0026
III	0.005	0.028	0.014	0.0095	0.0072
IV	0.006	0.003	0.0015	0.0009	0.0007
V	0.015	0.002	0.0013	0.0008	0.0006

I. Hydroacoustic,
II. Inclinometric,
III. Dynamometer,
IV. Multichannel preliminary data compression,
V. Parallel Multichannel.

6. Conclusions

In this paper, the integration of different measurement systems via Kalman filter to measure the OP coordinates is addressed. For this purpose, two sorts of multichannel Kalman filter are designed: preliminary data compression, and parallel. It is shown that, the integrating the measurement systems enables to increase the estimation accuracy and confidence of the information processing in the evaluation of the OP coordinates. Hence, the high efficiency of the OP in sea is attained.

The comparative analyses of the presented algorithms reveal that the parallel multichannel Kalman filter has the highest accuracy, but takes longer time to execute. On the other hand, the multichannel preliminary data compression Kalman filter has a bit worse accuracy than the parallel Kalman filter, but requires considerably less execution time.

If the OP automatic position control and supervision system has limited computational ability and cannot process the measurements from all the measurement systems in real-time, then the use of the multichannel preliminary data compression Kalman filter is more convenient.

References

1. Krinetski Y.I. and Gadzhiyev (Hajiyev) Ch.M. (1987). Approach to Floating Drilling Rigs Shift Coordinates Evaluation Accuracy Estimation Using Hydroacoustic Steering Systems, "Oil and Gas " USSR High Education Proceedings, No. 4, pp. 37-39 (in Russian).

2. Aliyev R.M. and Gadzhiyev (Hajiyev) Ch.M. (1988). Justification of Requirements to Direct Measurement Result Accuracy for Provision of SFDR Position Supervision Predetermined Certainty by Dynamometrical Method, Dep. in AERT I , 14.11.88, No. 8052-V88, 13 p) (in Russian).

3. Abdullayev A.A. and Gadzhiyev (Hajiyev) Ch.M. (1990). Design of Inclinometrical System of SFDR Position Determination from Preset Supervision Certainty, "Azerbaijanian Oil Economy", No.4, pp. 12-45 (in Russian).

4. Krinetski Ye.I and Gadzhiyev (Hajiyev) Ch.M. (1990). Justification of Requirements to Primary Informatics Sensors Accuracy for Provision of Predetermined Certainty of SFDR Position Supervision by Hydroacoustic Methods, "Oil and Gas" USSR High Education Proceedings, No. 9, pp. 76-80 (in Russian).

5. Fung P.T-K. and Grimble M.J. (1983). Dynamic Ship Positioning Using a Self-Tuning Kalman Filter. IEEE Transactions on Automatic Control, AC-28, No 3, pp.339-359.

6. Hagander P. and Wittenmark B. (1977). A Self-Tuning Filter for Fixed -hag. Smoothing . IEEE Transactions on Information Theory, v.IT- 23 , pp.377-3847

7. Gadzhiyev (Hajiyev) Ch.M. (1990). Adaptive Algorithm of Parameter Identification for Anchored Floating System Model/V1 Conference "Multiconnected Systems Control", Heads of Reports, Moscow, pp.137-138 (in Russian).

8. Fkirin M. (1989). Choice of Least-Squares Algorithms for the Identification of ARMAX Dynamic Systems. Int. J.Systems Sci., 20, No 7, pp.1221-1226.

REPORT DOCUMENTATION PAGE

1. Recipient's Reference	2. Originator's References	3. Further Reference	4. Security Classification of Document
	RTO-MP-43 AC/323(SCI)TP/13	ISBN 92-837-1018-5	UNCLASSIFIED/ UNLIMITED

5. Originator Research and Technology Organization
North Atlantic Treaty Organization
BP 25, 7 rue Ancelle, F-92201 Neuilly-sur-Seine Cedex, France

6. Title
Integrated Navigation Systems

7. Presented at/sponsored by
the Systems Concepts and Integration Panel (SCI) International Conference held at "Elektropribor", St. Petersburg, Russia, 24-26 May 1999.

8. Author(s)/Editor(s)	9. Date
Multiple	October 1999
10. Author's/Editor's Address	**11. Pages**
Multiple	278

12. Distribution Statement There are no restrictions on the distribution of this document. Information about the availability of this and other RTO unclassified publications is given on the back cover.

13. Keywords/Descriptors

Navigation
Integrated systems
Inertial navigation
Detectors
Laser gyroscopes
Optical fibers
HRG (Hemispherical Resonator Gyroscopes)
Accelerometers
Algorithms
Strapped down systems
Space stations
Space navigation
Global positioning system
Orbits
Navigation satellites
Computerized simulation
Flight tests
Kalman filtering
Filters
Markov processes
Signal processing
Fault tolerance

14. Abstract

The meeting proceedings from this 6th International Conference contains the Technical Evaluation Report and 25 papers presented by 61 authors working for 28 organizations from 9 countries. This Symposium on Integrated Navigation Systems was held in St. Petersburg, Russia on 24 to 26 May 1999. It was jointly organized by the Central Scientific and Research Institute "ELEKTROPRIBOR" and the Systems Concepts and Integration Panel of the Research and Technology Organization of NATO.

The papers presented covered the following headings:

- Inertial Gyro Sensors and Navigation Systems
- Satellite Navigation and Navigation Systems
- Integrated Satellite Navigation Systems and Navigation Systems
- Satellite Navigation and Inertial Navigation Systems Technology

NORTH ATLANTIC TREATY ORGANIZATION

RESEARCH AND TECHNOLOGY ORGANIZATION
BP 25 • 7 RUE ANCELLE
F-92201 NEUILLY-SUR-SEINE CEDEX • FRANCE
Télécopie 0(1)55.61.22.99 • E-mail mailbox@rta.nato.int

DIFFUSION DES PUBLICATIONS
RTO NON CLASSIFIEES

L'Organisation pour la recherche et la technologie de l'OTAN (RTO), détient un stock limité de certaines de ses publications récentes, ainsi que de celles de l'ancien AGARD (Groupe consultatif pour la recherche et les réalisations aérospatiales de l'OTAN). Celles-ci pourront éventuellement être obtenues sous forme de copie papier. Pour de plus amples renseignements concernant l'achat de ces ouvrages, adressez-vous par lettre ou par télécopie à l'adresse indiquée ci-dessus. Veuillez ne pas téléphoner.

Des exemplaires supplémentaires peuvent parfois être obtenus auprès des centres nationaux de distribution indiqués ci-dessous. Si vous souhaitez recevoir toutes les publications de la RTO, ou simplement celles qui concernent certains Panels, vous pouvez demander d'être inclus sur la liste d'envoi de l'un de ces centres.

Les publications de la RTO et de l'AGARD sont en vente auprès des agences de vente indiquées ci-dessous, sous forme de photocopie ou de microfiche. Certains originaux peuvent également être obtenus auprès de CASI.

CENTRES DE DIFFUSION NATIONAUX

ALLEMAGNE
Fachinformationszentrum Karlsruhe
D-76344 Eggenstein-Leopoldshafen 2

BELGIQUE
Coordinateur RTO - VSL/RTO
Etat-Major de la Force Aérienne
Quartier Reine Elisabeth
Rue d'Evère, B-1140 Bruxelles

CANADA
Directeur - Recherche et développement - Communications et gestion de l'information - DRDCGI 3
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 (RTO/AGARD Publications)
Carretera de Torrejón a Ajalvir, Pk.4
28850 Torrejón de Ardoz - Madrid

ETATS-UNIS
NASA Center for AeroSpace Information (CASI)
Parkway Center
7121 Standard Drive
Hanover, MD 21076-1320

FRANCE
O.N.E.R.A. (ISP)
29, Avenue de la Division Leclerc
BP 72, 92322 Châtillon Cedex

GRECE
Hellenic Air Force
Air War College
Scientific and Technical Library
Dekelia Air Force Base
Dekelia, Athens TGA 1010

HONGRIE
Department for Scientific Analysis
Institute of Military Technology
Ministry of Defence
H-1525 Budapest P O Box 26

ISLANDE
Director of Aviation
c/o Flugrad
Reykjavik

ITALIE
Centro documentazione tecnico-scientifica della Difesa
Via Marsala 104
00185 Roma

LUXEMBOURG
Voir Belgique

NORVEGE
Norwegian Defence Research Establishment
Attn: Biblioteket
P.O. Box 25
NO-2007 Kjeller

PAYS-BAS
NDRCC
DGM/DWOO
P.O. Box 20701
2500 ES Den Haag

POLOGNE
Chief of International Cooperation Division
Research & Development Department
218 Niepodleglosci Av.
00-911 Warsaw

PORTUGAL
Estado Maior da Força Aérea
SDFA - Centro de Documentação
Alfragide
P-2720 Amadora

REPUBLIQUE TCHEQUE
VTÚL a PVO Praha /
Air Force Research Institute Prague
Národní informační středisko obranného výzkumu (NISČR)
Mladoboleslavská ul., 197 06 Praha 9

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

AGENCES DE VENTE

NASA Center for AeroSpace Information (CASI)
Parkway Center
7121 Standard Drive
Hanover, MD 21076-1320
Etats-Unis

The British Library Document Supply Centre
Boston Spa, Wetherby
West Yorkshire LS23 7BQ
Royaume-Uni

Canada Institute for Scientific and Technical Information (CISTI)
National Research Council
Document Delivery
Montreal Road, Building M-55
Ottawa K1A 0S2
Canada

Les demandes de documents RTO ou AGARD doivent comporter la dénomination "RTO" ou "AGARD" selon le cas, suivie du numéro de série (par exemple AGARD-AG-315). Des informations analogues, telles que le titre et la date de publication sont souhaitables. Des références bibliographiques complètes ainsi que des résumés des publications RTO et AGARD figurent dans les journaux suivants:

Scientific and Technical Aerospace Reports (STAR)
STAR peut être consulté en ligne au localisateur de ressources uniformes (URL) suivant:
http://www.sti.nasa.gov/Pubs/star/Star.html
STAR est édité par CASI dans le cadre du programme NASA d'information scientifique et technique (STI)
STI Program Office, MS 157A
NASA Langley Research Center
Hampton, Virginia 23681-0001
Etats-Unis

Government Reports Announcements & Index (GRA&I)
publié par le National Technical Information Service
Springfield
Virginia 2216
Etats-Unis
(accessible également en mode interactif dans la base de données bibliographiques en ligne du NTIS, et sur CD-ROM)

Imprimé par le Groupe Communication Canada Inc.
(membre de la Corporation St-Joseph)
45, boul. Sacré-Cœur, Hull (Québec), Canada K1A 0S7

NORTH ATLANTIC TREATY ORGANIZATION

RESEARCH AND TECHNOLOGY ORGANIZATION
BP 25 • 7 RUE ANCELLE
F-92201 NEUILLY-SUR-SEINE CEDEX • FRANCE
Telefax 0(1)55.61.22.99 • E-mail mailbox@rta.nato.int

DISTRIBUTION OF UNCLASSIFIED RTO PUBLICATIONS

NATO's Research and Technology Organization (RTO) holds limited quantities of some of its recent publications and those of the former AGARD (Advisory Group for Aerospace Research & Development of NATO), and these may be available for purchase in hard copy form. For more information, write or send a telefax to the address given above. **Please do not telephone.**

Further copies are sometimes available from the National Distribution Centres listed below. If you wish to receive all RTO publications, or just those relating to one or more specific RTO Panels, they may be willing to include you (or your organisation) in their distribution.

RTO and AGARD publications may be purchased from the Sales Agencies listed below, in photocopy or microfiche form. Original copies of some publications may be available from CASI.

NATIONAL DISTRIBUTION CENTRES

BELGIUM
Coordinateur RTO - VSL/RTO
Etat-Major de la Force Aérienne
Quartier Reine Elisabeth
Rue d'Evère, B-1140 Bruxelles

CANADA
Director Research & Development
Communications & Information
Management - DRDCIM 3
Dept of National Defence
Ottawa, Ontario K1A 0K2

CZECH REPUBLIC
VTÚL a PVO Praha /
Air Force Research Institute Prague
Národní informační středisko
obranného výzkumu (NISČR)
Mladoboleslavská ul., 197 06 Praha 9

DENMARK
Danish Defence Research
Establishment
Ryvangs Allé 1, P.O. Box 2715
DK-2100 Copenhagen Ø

FRANCE
O.N.E.R.A. (ISP)
29 Avenue de la Division Leclerc
BP 72, 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

HUNGARY
Department for Scientific Analysis
Institute of Military Technology
Ministry of Defence
H-1525 Budapest P O Box 26

ICELAND
Director of Aviation
c/o Flugrad
Reykjavik

ITALY
Centro documentazione
tecnico-scientifica della Difesa
Via Marsala 104
00185 Roma

LUXEMBOURG
See Belgium

NETHERLANDS
NDRCC
DGM/DWOO
P.O. Box 20701
2500 ES Den Haag

NORWAY
Norwegian Defence Research
Establishment
Attn: Biblioteket
P.O. Box 25
NO-2007 Kjeller

POLAND
Chief of International Cooperation
Division
Research & Development
Department
218 Niepodleglosci Av.
00-911 Warsaw

PORTUGAL
Estado Maior da Força Aérea
SDFA - Centro de Documentação
Alfragide
P-2720 Amadora

SPAIN
INTA (RTO/AGARD Publications)
Carretera de Torrejón a Ajalvir, Pk.4
28850 Torrejón de Ardoz - 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 Center for AeroSpace
Information (CASI)
Parkway Center
7121 Standard Drive
Hanover, MD 21076-1320

SALES AGENCIES

NASA Center for AeroSpace Information (CASI)
Parkway Center
7121 Standard Drive
Hanover, MD 21076-1320
United States

The British Library Document Supply Centre
Boston Spa, Wetherby
West Yorkshire LS23 7BQ
United Kingdom

Canada Institute for Scientific and Technical Information (CISTI)
National Research Council
Document Delivery
Montreal Road, Building M-55
Ottawa K1A 0S2
Canada

Requests for RTO or AGARD documents should include the word 'RTO' or 'AGARD', as appropriate, followed by the serial number (for example AGARD-AG-315). Collateral information such as title and publication date is desirable. Full bibliographical references and abstracts of RTO and AGARD publications are given in the following journals:

Scientific and Technical Aerospace Reports (STAR)
STAR is available on-line at the following uniform resource locator:
http://www.sti.nasa.gov/Pubs/star/Star.html
STAR is published by CASI for the NASA Scientific and Technical Information (STI) Program
STI Program Office, MS 157A
NASA Langley Research Center
Hampton, Virginia 23681-0001
United States

Government Reports Announcements & 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)

Printed by Canada Communication Group Inc.
(A St. Joseph Corporation Company)
45 Sacré-Cœur Blvd., Hull (Québec), Canada K1A 0S7

ISBN 92-837-1018-5