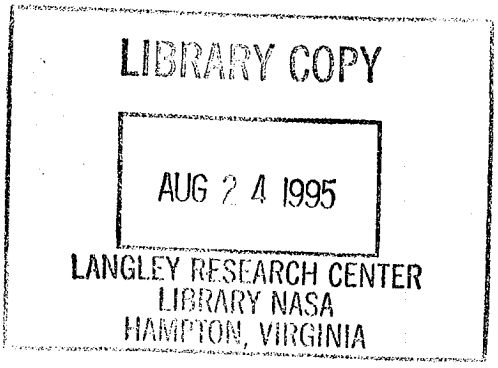
NASA Conference Publication 3305
DOT/FAA/CT-95/39

NASA-CP-3305 19960002161

FAA/NASA Joint University Program for Air Transportation Research 1993–1994

Compiled by
Richard M. Hueschen



Proceedings of a conference sponsored by the Federal Aviation Administration, Washington, D.C., and the National Aeronautics and Space Administration,Washington, D.C., and held in Athens, Ohio
July 14–15, 1994

August 1995

NASA Conference Publication 3305
DOT/FAA/CT-95/39

FAA/NASA Joint University Program for Air Transportation Research 1993–1994

Compiled by
Richard M. Hueschen
Langley Research Center • Hampton, Virginia

Proceedings of a conference sponsored by the Federal Aviation Administration, Washington, D.C., and the National Aeronautics and Space Administration,Washington, D.C., and held in Athens, Ohio
July 14–15, 1994

National Aeronautics and Space Administration
Langley Research Center • Hampton, Virginia 23681-0001

August 1995

This publication is available from the following sources:

NASA Center for AeroSpace Information
800 Elkridge Landing Road
Linthicum Heights, MD 21090-2934
(301) 621-0390

National Technical Information Service (NTIS)
5285 Port Royal Road
Springfield, VA 22161-2171
(703) 487-4650

PREFACE

The Joint University Program for Air Transportation Research is a coordinated set of three grants sponsored by NASA Langley Research Center and the Federal Aviation Administration, one each with the Massachusetts Institute of Technology (NGL-22-009-640), Ohio University (NGR-36-009-017), and Princeton University (NGL-31-001-252). The research conducted under these grants, which were instituted in 1971, focus on the strengths of each institution. The goals of this program are consistent with the aeronautical interests of both NASA and the FAA in furthering the safety and efficiency of the National Airspace System. The continued development of the National Airspace System, however, requires advanced technology from a variety of disciplines, especially in the areas of computer science, navigation, guidance, and control theory, aircraft performance, flight dynamics, communications, and human factors. The Joint University Program (JUP) was created to provide new methods for interdisciplinary education which would develop researchers for solving these large scale problems. Under the JUP, each university submits a separate proposal yearly which is dealt with individually by NASA and FAA. At the completion of each research task, a comprehensive and detailed report is issued for distribution to the program participants. Typically, this is a thesis that fulfills the requirements for an advanced degree or a report describing an undergraduate research project. Also, papers are submitted to technical conferences and archival journals. These papers document the research accomplished under the JUP for national and international audiences.

To promote technical interchange among the students, periodic reviews are held at the schools and at a NASA or FAA facility. The 1993-1994 year-end review was held at Ohio University, Athens, Ohio, July 14-15, 1994. At these reviews the program participants, both graduate and undergraduate, have an opportunity to present their research activities to their peers, to professors, and to invited guests from government and industry.

This conference publication represents the thirteenth in a series of yearly summaries of the program. (The 1992-1993 summary appears in NASA CP-3246, DOT/FAA/CT-94/03). Most of the material is the effort of students supported by the research grants. Four types of contributions are included in this publication: a summary of ongoing research relevant to the Joint University Program is presented by each principal investigator, completed works are represented by full technical papers, research previously in the open literature (e.g., theses or journal articles) is presented in an annotated bibliography, and status reports of ongoing research are represented by copies of presentations with accompanying text.

Richard M. Hueschen
NASA Langley Research Center

CONTENTS

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

AN INVESTIGATION OF

AIR TRANSPORTATION TECHNOLOGY

AT THE

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

1993 - 1994

Robert W. Simpson
Flight Transportation Laboratory

R. John Hansman
Aeronautical Systems Laboratory

MIT, Cambridge, MA 02139

SUMMARY OF RESEARCH ACTIVITIES

1. INTRODUCTION

There are two completed projects and two continuing research activities under the sponsorship of the FAA/NASA Joint University Program as the 1993-94 period ends. There were a number of publications during the year which are referenced in this report. A brief summary of the continuing research projects is provided.

The completed project was:

1. Alhanatis, Robert - Analysis of Aircraft Surface Motion at Boston Logan International Airport, MIT Flight Transportation Laboratory Report, FTL 94-5, June 1994:

2. Achtmann, Eric - A State of the Art Review and Critical Analysis of World Jet Transport Safety and Aviation Fire Safety (to be published).

The active research projects are:

1. ASLOTS - An Interactive Adaptive System for Automated Approach Spacing of Aircraft - Husni Idris.

2. Alerting in Automated and Datalink Capable Cockpits- James Kuchar.

2. REVIEW OF CONTINUING RESEARCH ACTIVITIES

2.1 Human-Centered Automation of ATC Operations in the Terminal Area by Husni Indris, Flight Transportation Laboratory

This research is concerned with development of an interactive decision support system for ATC controllers responsible for arrivals and departures in a busy terminal area. It is focused on a concept called "ASLOTS" for interactive, adaptive spacing of aircraft on final approach to landing as described in last year's report, but has broadened its scope during the last year to be concerned with techniques and methods for Human-Centered Automation of all operations in a busy terminal area which has arrival and departures at multiple airports and runways.

In this year, a review of possible techniques has been conducted, and a paradigm developed for 10 different levels of possible automation based on task allocation. Techniques reviewed are: Knowledge-based Control, Adaptive Control, Learning Control, NeuroControl, Fuzzy Control, and Optimal Control. As a result of this review, a hybrid intelligent control

structure has been developed for ASLOTS where each task can be kept manual or automated to several degrees independently from other tasks. The current specific research goals are now:

1. Implement the automation of the path generation task.
2. Implement automatic Conformance Monitoring
3. Investigate making the reference model for ASLOTS path generation to be adaptive.
4. Create a Rule-Base for the Automated Path Generation & Conformance Monitoring.
5. Introduce fuzzy-logic at points in the Rule-Base.
6. Allow the Rule-Base to interact with the Slot Marker status, and study the associated Human Factors.

The work on developing ATCSIM the simulation capability for air and ground motion has continued. The current ASLOTS logic has been incorporated into the air module, and a ground surface simulation for both Frankfurt and Boston Logan can now be run simultaneously with the airspace operations.

2.2 A Unified Framework for the Evaluation of Hazard Alerting Systems by Jim Kuchar, Aeronautical Systems Laboratory

Given the increasingly complex environment in which aircraft operate, the management of information on the flight deck has become a major issue affecting flight safety. One major component of the modern cockpit is comprised of several alerting systems which have been developed to monitor flight safety and warn the flight crew of potential hazards. Warning systems have seen a rapid growth in complexity, both in terms of the extent of information which now may be used and also in the methods by which alerts may be presented to the crew. To understand these complex system issues, a framework is required that outlines the major issues involved in alerting system performance and that may be applied to various types of alerting systems to evaluate their effectiveness. This paper outlines research that is being conducted toward the development of a generalized framework that can be applied to the design or evaluation of alerting systems.

2.2.1 Problem Statement

The evolution of the aircraft flight deck has seen an enormous increase in the amount of information available to the pilot about the environment around the aircraft. In order to manage this information efficiently, current design philosophy follows the 'dark cockpit' concept where the pilot has easy access to the data required to fly and navigate the aircraft without being overloaded by extraneous information. Alerting systems have been incorporated into cockpit design as a means by which potential hazards may

be monitored silently. Only when an obstacle poses a certain level of risk is the flight crew notified that a potential threat exists. Because they are a sudden intrusion into the 'dark cockpit', alerts are generally treated very seriously and false alarms may reduce crew confidence in the system. For this reason, it is important that alerting systems be designed with care, providing a necessary level of safety while keeping nuisance alerts to a minimum.

Alerting systems on modern aircraft are generally the product of an ad hoc, evolutionary process. Warning systems have historically been added in response to accidents, and incrementally modified when system performance failed to maintain certain safety standards. Modifications have also been made when a system behaved in an overly-conservative manner, resulting in an unacceptable rate of nuisance alerts.

As more information is made available to the cockpit for use in alerting flight crews to threats, the complexity of the alerting systems has grown. Ground Proximity Warning Systems (GPWS), for example, have seen several incremental enhancements. Starting with simple altitude call-outs, sink rate and terrain closure rate warnings were added using radar altimeter data, along with different alerting criteria dependent upon aircraft configuration. Effort is now being applied toward incorporating three-dimensional data from a sensor or database [Bateman, personal communication, 8/93]. As complexity increases, the tradeoffs between providing adequate time to react and being overly conservative become increasingly difficult to manage since changes in system parameters in one area of the alerting system may have large impacts on another area.

In order to efficiently design and evaluate alerting systems, a framework is required that may be used to identify those areas most sensitive to changes in design parameters. Research in the MIT Aeronautical Systems Lab identifies and develops the major issues affecting alerting system performance, and provides a methodology that may be used to evaluate system effectiveness. While alerting systems have historically been created in response to specific problem areas, the research investigates alerting systems in a generalized manner, providing a single framework which may be applied to many specific types of hazard alerting.

2.2.2 Approach

1.2.2.1 Identification of Salient System Components

Although alerting systems are designed and implemented to protect against specific hazards, a generalized model may be developed that provides insight into the interrelationships between system components. Figure 1 is a control system view of the principal elements of a generalized hazard situation with an alerting system. As shown, there are four main blocks in

the control loop: Measurement Sources, Alerting System, Human Operator, and Encounter Dynamics.

Continuing the control system analogy, a state-space representation of a hazard encounter is used. The vector **x** represents the complete set of states that define the dynamics of a hazard encounter. In a terrain alerting system application, for example, **x** might include altitude, descent rate, airspeed, and flap and gear settings. The Measurement Sources block represents the sensors that provide estimates of the states in **x** to the rest of the system. Some of this information is passed directly to the operator through displays and other non-alert sources, and other data are used by the alerting system to determine if an alert is warranted. The alerting system classifies the situation into one of several levels of severity and issues alert information when appropriate. The operator uses the information from the alerting system and other sources to make necessary control inputs to the encounter dynamics. Other information from training, previous experience, established operating procedures, and the environment may also affect the operator's response. Control inputs and uncertainties in the encounter dynamics affect the evolving state vector, **x**, which is available to the measurement sources as the loop repeats.

The generalized structure of Figure 1 enables it to be applied to situations where the operator is controlling one vehicle (e.g., the pilot of an aircraft) and to cases where the operator provides instructions to many vehicles (e.g., an air traffic controller). In addition, the framework may be applied to single-hazard situations or to multiple-hazard encounters and to systems with one or several alerting systems.

Table 1a) Crew Response

Since the end goal of an alerting system is to provide the crew with enough information to avoid a hazard, flight crew reaction is an important consideration. Crew response to an alert is a function of several parameters. The type of display and the quality of data and method of display depiction affect the crew's impression of the severity of the hazard as well as potential escape maneuvers, as has been shown in several simulation studies [Kuchar & Hansman, 1991, 1993]. Crew alertness and training also may have a strong impact on the response. For example, aggressive windshear avoidance training was started in the late 1980's, contributing in part to eliminate windshear-related accidents since 1986 [Bateman, 1991]. In addition, a history of previously-encountered nuisance alerts may decrease crew response to the threat. Numerous Controlled Flight Into Terrain (CFIT) accidents have been attributed to "delayed GPWS response syndrome" [DeCelles, 1991], where flight crews delayed their reaction to an executive "Pull Up!" alert, due in part to past exposure to nuisance alerts.

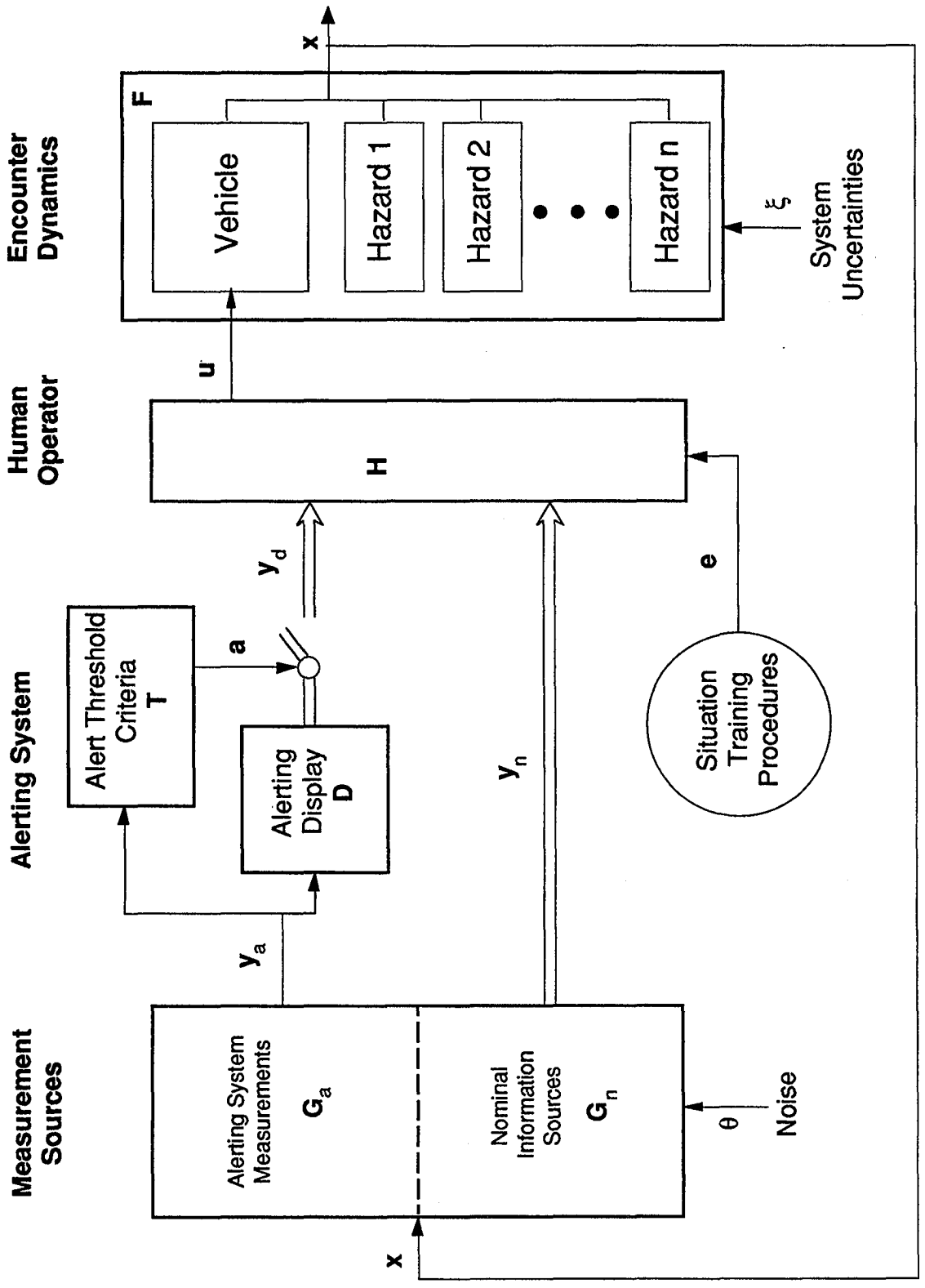


Figure 1: Generalized Hazard Encounter System Block Diagram

b) Alert Thresholds

In a typical alerting system, measurements of a hazard are compared against a set of alerting thresholds to determine whether an alert needs to be issued. However, uncertainties may be present that make a clear distinction between a hazard and a non-hazard difficult. Therefore, the system cannot determine with absolute certainty whether a given situation really warrants an alert. Rather, there is some probability that the situation is truly hazardous, and some probability that it is not.

Given a hazard measurement, the output of an alerting system may be considered to be a deterministic event: either an alert is issued or it is not.[1] Since the alerting system's decision may or may not reflect the true situation, however, there are four possible outcomes of any single-stage alerting decision, as shown in Table 1. If a situation is truly hazardous, an alert is considered to be necessary and is termed a Correct Detection, CD (or true-positive in signal detection theory). Similarly, a failure to alert when a situation is non-hazardous is Correct Rejection, CR (or true-negative). An alert that is issued in a non-hazardous situation is a False Alarm, FA, and likewise, failure to alert to a truly hazardous situation is a Missed Detection, MD.

Alerting System Decision Outcomes

	True Situation	
	Non-Hazard	Hazard
Alert Not Issued	**CR:** Correct Rejection (True-Negative)	**MD:** Missed Detection (False-Negative)
Alert Issued	**FA:** False Alarm (False-Positive)	**CD:** Correct Detection (True-Positive)

More insight may be gained by examining this problem from a mathematical standpoint. Given a hazard level measurement, y, the alerting system makes a mutually exclusive decision: alert or do not alert. The true situation, however, may or may not warrant an alert, and can be treated probabilistically. Accordingly, H is defined as an event in which an alert is needed -- the true situation is hazardous. The probability that the situation is hazardous, given the measurement y, is then denoted $\Pr(H \mid y)$, read "probability of H occurring, given y".

---

[1] Note that in signal detection theory, the true situation is typically treated as deterministic and the measurement is considered to be a random variable. In this discussion, however, the measurement will be considered known deterministically and the true situation is treated as a random variable.

An example graph of Pr(H | y) for an alerting system is shown in Figure 2. The Figure also overlays an alerting threshold: measured hazard levels greater than the threshold produce alerts, and measurements less than the threshold do not. The height of the curve at a given measurement, y, shows the probability that the situation warrants an alert. The four outcome regions (CD, FA, MD, CR) are also shown. For the sample measurement in the Figure, no alert is issued because the measurement is less than the alert threshold. The value of Pr(H | y) at that value of y is therefore the probability of a missed detection, Pr(MD). The value above the Pr(H | y) curve is the probability that the situation is truly not hazardous, Pr(CR).

Similar behavior occurs when the measurement is greater than the threshold. There will exist a probability of correct detection, Pr(CD), as well as a probability of false alarm, Pr(FA).

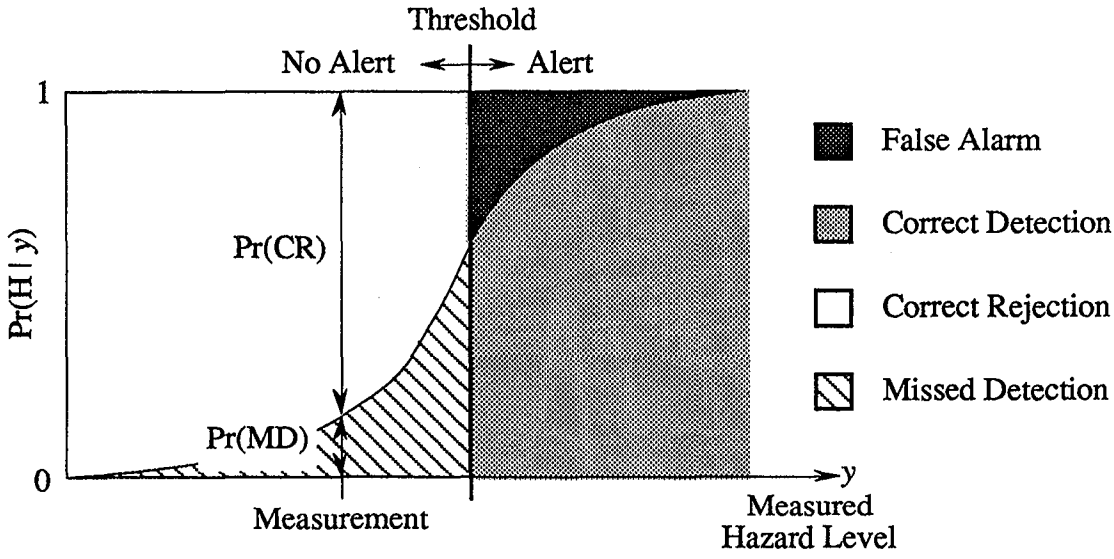


Figure 2: **Alerting System Decision Outcomes**
Example curve of Pr(H | y) with alerting threshold
Measured value > threshold: alert issued
Measured value < threshold: no alert issued

Errors in the alerting decision are generally caused by one or more of three factors. First, due to sensor errors the threat may be measured as inside the given threshold when in reality it is outside the threshold (or vice versa). Second, the threshold itself may be designed in an overly conservative manner -- leading to warnings for targets which do not actually pose a threat.

Finally, the extrapolation of the threat's trajectory may be in error if complete information is not available regarding its intended flight path.

Improvements in one of these three areas may affect the others. For example, increasing sensor accuracy may allow the system designer to use less conservative thresholds. A comparable effect might be achieved, for example, through datalink of an intruder's intended flight path, providing the alerting system with better information with which to extrapolate the hazard's trajectory.

c) **Multi-Stage Alerting**

Many alerting systems currently use a two-stage approach: a potential threat is first highlighted and presented to the flight crew as an advisory caution so that the crew has time to determine whether the threat warrants action. Should the threat persist until corrective action is required, a warning is issued that generally should be followed without hesitation.

It has been observed that crews tend to have low tolerances for false alarms at the warning level, and higher tolerances for false alarms at the cautionary level since a reaction is not mandated [Boucek, personal communication, 2/94]. Therefore, cautionary alerts can be used to identify low-probability potential threats and allow the crew an opportunity to evaluate the situation. Warnings, though, are designed to alert only to high-probability threats when immediate action is needed.

2.2.2.2 Development of Methodology for Estimation of Threshold Effectiveness

The research has also involved the development of a methodology that can be used to estimate the probability of an encounter with a hazard, denoted Pr(E). The methodology is developed in a generalized fashion, allowing it to be extended to a wide variety of hazard encounter situations.

The methodology begins by developing a method for modeling the aircraft and hazard as appropriate to the situation. For example, in a Traffic Alert and Collision Avoidance System (TCAS) application, the aircraft and threat might be modeled as cylinders or spheres. Additional modeling possibilities include 'soft' hazards such as weather or noise-sensitive regions.

Next, the potential trajectories of the hazard relative to the aircraft are calculated, each with its corresponding probability of occurring. The accuracy to which these trajectories may be computed is a function of sensor accuracy and the type of information available to the alerting system. An exclusion zone is then calculated that defines the space in which the hazard must exist

for an encounter to occur. An example situation is shown schematically in Figure 3 for a TCAS application.

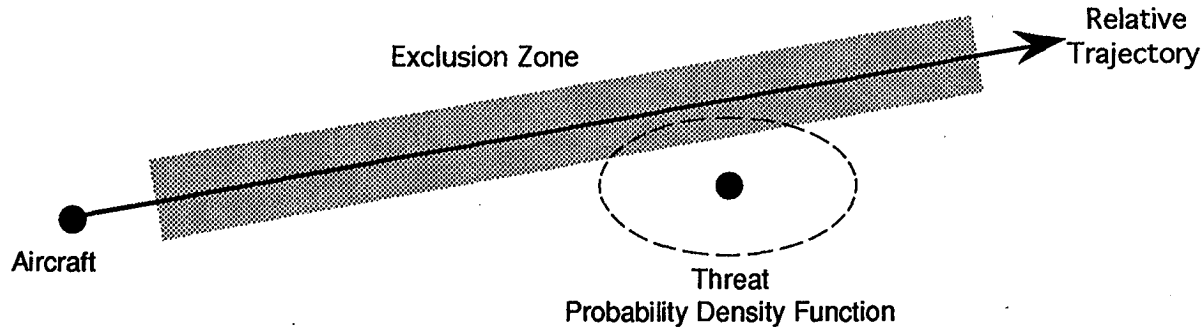


Figure 3: Exclusion Zone for Example TCAS Situation

The probability density function describing the hazard's location (governed by sensor accuracies) is then integrated over the exclusion zone to obtain an estimate of Pr(E) for that trajectory. The overall probability of an encounter, given all possible trajectories, is then given by:

$$\Pr(E) = \int_{\text{all trajectories}} \Pr(E \mid \text{trajectory } i)\, \Pr(\text{trajectory } i)$$

These probabilities then lead to a better understanding of the potential for nuisance alerts and may be used to estimate an overall level of safety for the system.

This method has been applied and demonstrated using an existing TCAS II system in which the aircraft is on a collision course with intruding aircraft at various relative speeds. Figure 4 shows an example of the data obtained from numerical integration using measurement errors typical of a TCAS II system. Curves representing the estimated probability of collision with a threat aircraft are shown as a function of range to the threat. The upper curve represents the case where no avoidance action is taken, while the lower curve represents a case in which a standard TCAS avoidance maneuver is assumed: 5 second delay followed by a 0.25g pull-up until 1500 feet per minute climb rate is attained [RTCA, 1983].

Also shown on the graph is the significance of the choice in alerting threshold location. As shown, given a particular threshold there is a certain probability that even with an avoidance maneuver there will be a collision. In addition, there is a probability that the threat would have been avoided even had no action been taken - the false alarm case. Between the two curves is the effective benefit the alerting system would provide in terms of the improvement in collision probability.

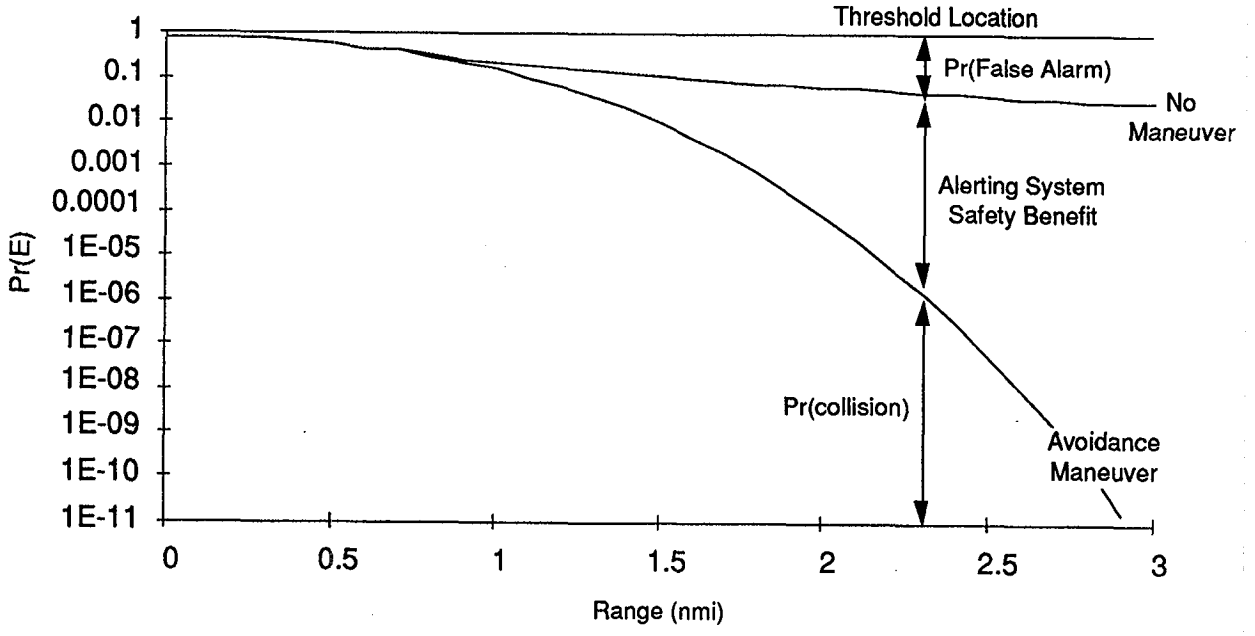


Figure 4: **TCAS Example - Pr(E) Curves**
Coaltitude threat on collision course, 400 kt closure rate.
Top curve: no action taken
Bottom curve: 5 sec delay, 0.25 g pull-up.

2.3 Conclusions

The contributions of the framework for the evaluation of alerting systems are as follows.

1) Major issues relating to alerting system effectiveness have been identified and developed. A framework connecting these issues is then formed that highlights the interactions that are possible between system components.

2) A methodology has been developed that may be applied to complex hazard encounter situations to estimate performance measures such as nuisance alert rate or overall safety level.

3) The framework has been demonstrated through applications to actual alerting system designs.

4) By using a generalized approach to alerting systems, the framework can be applied to varied specific applications including civil and military use, as well as to non-aviation related alerting systems.

A complete report of the research will be available as a doctoral thesis in February, 1995.

2.4 References

Bateman, Don, "How to Terrain Proof the World's Airline Fleet", 44th Annual International Air Safety Seminar, November 14, 1991, Singapore.

Boucek, G., The Boeing Commercial Airplane Company.

DeCelles, J.L., "The Delayed GPWS Response Syndrome", Report for Aviation Research & Education Foundation, Herndon, VA. Also see Air Line Pilot, Jan. 1, 1992.

Kuchar, J.K., and R.J. Hansman, "Advanced Terrain Displays for Transport Category Aircraft", MIT Aeronautical Systems Laboratory Report ASL-91-3, September, 1991.

Kuchar, J.K., and R.J. Hansman, "Part-Task Simulation Study of Candidate Terrain Alerting Displays", MIT Aeronautical Systems Laboratory Report ASL-93-6, June, 1993.

Radio Technical Commission for Aeronautics, "Minimum Operational Performance Standards for Traffic Alert and Collision Avoidance System (TCAS) Airborne Equipment", SC-147, Washington, D.C., 1983.

3. ANNOTATED REFERENCES OF 1993 - 94 PUBLICATIONS

3.1 Alhanatis, Robert; Analysis of Aircraft Surface Motion at Boston Logan International Airport, FTL Report R94-5, September 1994, Flight Transportation Laboratory, MIT, Cambridge, MA 02173

The purpose of this research is to examine the nature of aircraft surface motion on the airport surface during normal operations. Twelve hours of radar data, gathered by MIT Lincoln Laboratories from Logan Airport in Boston, were made available for this study. Specifically, the data included target position reports from the ASDE-3 surface surveillance radar and the ASR-9 radar from the near terminal airspace information. This data covers a variety of runway configurations, weather conditions, traffic levels and high or low visibility conditions.

The study is divided into three sections. The first one focuses on the runway, and examines occupancy times, exit velocities, exit usage and velocity profiles of the final approach and landing phase. The second section, analyzes fourteen runway-taxiway intersections. Results are presented for the crossing times and usage of these intersections. The analysis also focuses on relating crossing direction, runway configuration and aircraft size. Finally,

average taxiway velocities and the overall taxiway usage is measured. Additionally, the role that the location of the taxiway segment, as well as its length, plays in the variation of these velocities are examined. Where possible, this study includes means, standard deviations and sample sizes of the variables in question.

3.2 Achtmann, Eric - A State of the Art Review and Critical Analysis of World Set Transport Safety and Aviation Fire Safety (to be published).

This research examines the safety in air transportation by jet aircraft worldwide in the period 1970-1992 to find the relative occurrence of accidents of different types. It then focuses on the risk of jet aircraft accidents involving fire. Accident types are classified into 11 categories: Bird Strikes, Collisions (Aircraft, Terrain), Weather Decompression, Ditching, Failed Takeoff, Fuel Exhaustion, Ground (Stationary, Taxiway), In-flight explosion, Major Mechanical Failure, Post Crash Fires, and Wake Vortex Upset. Types of accidents are ranked by number of accidents, fatality tolls, and survivability rates.

A case study review is given for the more interesting cases where fire played a major role in a jet transport accident, and a brief review of past and current Aircraft Fire Research activities is provided.

OHIO UNIVERSITY

INVESTIGATION OF AIR TRANSPORTATION TECHNOLOGY AT OHIO UNIVERSITY
1993-1994

Robert W. Lilley
Avionics Engineering Center
Department of Electrical and Computer Engineering
Ohio University
Athens, Ohio

SUMMARY OF RESEARCH

The Joint University Program (JUP) offers students, faculty, and staff of the Avionics Engineering Center the opportunity to conduct basic research relating to the safety and efficiency of the National Airspace System. During the 1993-1994 year research was continued in a variety of GPS-related technologies and also in a hybrid data uplink system.

- Processing techniques developed for the GPS attitude determination have been applied to centimeter-level kinematic GPS aircraft positioning for flight reference systems and aircraft autoland and taxing operations. This resulted in the first real-time kinematic GPS autoland.

- The Fault Detection and Isolation (FDI) algorithm developed under the Joint University Program has been fully adopted by RTCA Special Committee 159 as the baseline algorithm for sole-means navigation of GPS integrated with GLONASS, LORAN-C, or baro-altimeter.

- A Satellite Coverage Research Analysis Model (SCRAM) has been developed which provides outage areas, outage dynamics, and availability information for satellite-based navigation systems. RTCA Special Committee 159 has adopted this model for use in their work.

- A basic receiver structure has been simulated and tested for use with the hybrid data uplink. This uplink utilizes the phase of an existing AM carrier to transmit digital data, thus both amplitude and phase modulation are applied, resulting in a hybrid carrier. The receiver simulation is directly applicable to real-time implementation as a result of the development platform.

This research resulted in two M.S. theses, a Ph.D. dissertation, and various conference papers. An annotated bibliography of these publications can be found on the following pages. Also included is a bar graph which shows the various Ohio University students and their current positions who research and resulting careers were made possible through the funding provided over the twenty-three year history of the JUP program.

ANNOTATED BIBLIOGRAPHY OF 1993-1994 PUBLICATIONS

1. Bernath, G. N.: Fault Detection and Exclusion in Multisensor Navigation Systems. M.S. Thesis, Ohio University, Department of Electrical and Computer Engineering, Athens, OH, March 1994.

In order for a multisensor navigation system to meet integrity requirements, there must be a way of detecting erroneous measurements, using only data from those measurements. This can be accomplished using a parity space estimation algorithm. Erroneous measurements must then be removed from the position solution; the entire process is called Fault Detection and Exclusion (FDE). A baseline FDE algorithm has been determined and is capable of working in real time on present affordable hardware.

2. Diggle, D. W.: An Investigation into the Use of Satellite-Based Positioning Systems for Flight Reference/Autoland Operations. Ph.D. Dissertation, Ohio University, Department of Electrical and Computer Engineering, Athens, OH, March 1994.

An Interferometric Global Positioning System Flight Reference System (IGPS FRS) is implemented and flight tested on a transport category aircraft. This dissertation discusses the IGPS FRS theory and principles of operation, its architecture and integration with the aircraft, and the initial static calibration and flight-test performance results. The validity of the results is established by referencing them to a known ground test point and/or laser tracking system. The principles underlying the operation of the IGPS FRS are similar to those used for kinematic surveying, and are also referred to as differential carrier-phase tracking with integer ambiguities resolved. Flight Reference System objectives include: 0.1 m accuracy rms (each axis); one or more updates per second; UTC (Universal Time, Coordinated) synchronization better than 0.1 ms; real-time, all-weather operation; and, repeatable flight paths. The latter requirement call for full integration with the aircraft flight control system and coupled flight.

3. Mathur, N. G.: Initial Characterization of QBIC IMU and Design of a DATAC Bus Interface. M.S. Thesis, Ohio University, Department of Electrical and Computer Engineering, Athens, OH, November 1993.

With the evaluation of digital computers and strapdown sensors in the 1960's, inertial navigation technology has improved significantly. As compared to the bulky, inaccurate and expensive gimbaled IMU's (Inertial Measurement Unit), development of low cost, medium accuracy strapdown IMU's started with development in GPS (Global Positioning System) technology, towards the later 1980. With good short term accuracy and stability of IMU and good long term stability of GPS, the integration of IMU and GPS is very beneficial. Theory of operation and initial characterization of Q-BIC IMU (Unit

developed by Dr. M. Morrison) concluded that Q-BIC IMU has approximately 2 cm/sec^2 noise on the acceleration data at the rate of 8 samples/sec. For flight evaluation of combined Inertial and GPS position data, a DATAC (Digital Autonomous Terminal Access Communication) bus interface is developed and successfully tested onboard TSRV (Transport System Research Vehicle) Boeing 747-100 by NASA Langley Research center with GPS derived position data.

4. Skidmore, T. A.; van Graas, F.: GPS Integrity Monitoring and Multipath Error Distributions. Proceeding of the ION 49th Annual Meeting, June 21-23, 1993.

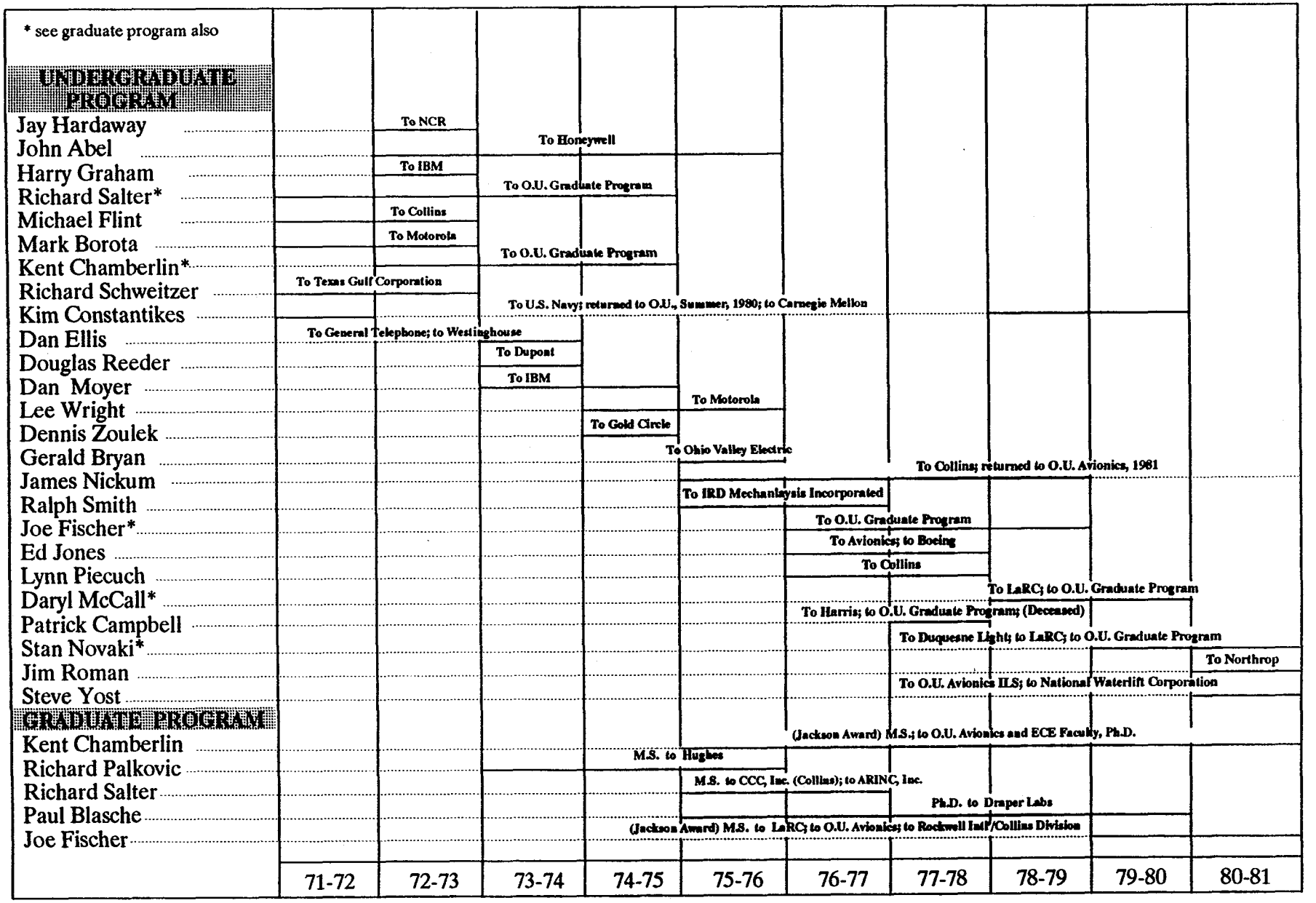
The ability to perform integrity monitoring of a differential Global Positioning System (DGPS) precision approach system is dependent upon the form of the monitor used and the various noise sources acting on the reference station, the monitor, and the approaching aircraft This paper focuses on how much multipath and assumptions about its probability density function (pdf) affect the availability of the integrity monitoring function. Also discussed is a technique specified as Code/Carrier Integrity Monitoring which can be used to greatly improve integrity monitoring availability over methods employing convention C/A-Code operations.

5. Skidmore, T. A.; van Graas, F.: DGPS Ground Station Integrity. Proceeding of the ION National Technical Meeting, January 24-26, 1994.

This paper summarizes the development of a unique Differential Global Positioning System (DGPS) ground station integrity monitor which can offer improved availability over conventional code-differential monitoring systems. This monitoring technique, called Code/Carrier Integrity Monitoring (CCIM), uses the highly stable integrated Doppler measurement to smooth the relatively noisy code-phase measurements. The pseudorange correction is therefore comprised of the integrated Doppler measurement plus the CCIM offset. The design and operational results of a DGPS ground station integrity monitor are reported. A robust integrity monitor is realized which is optimized for applications such as the Special Category I (SCAT-I) defined in the RTCA Minimum Aviation System Performance Standards.

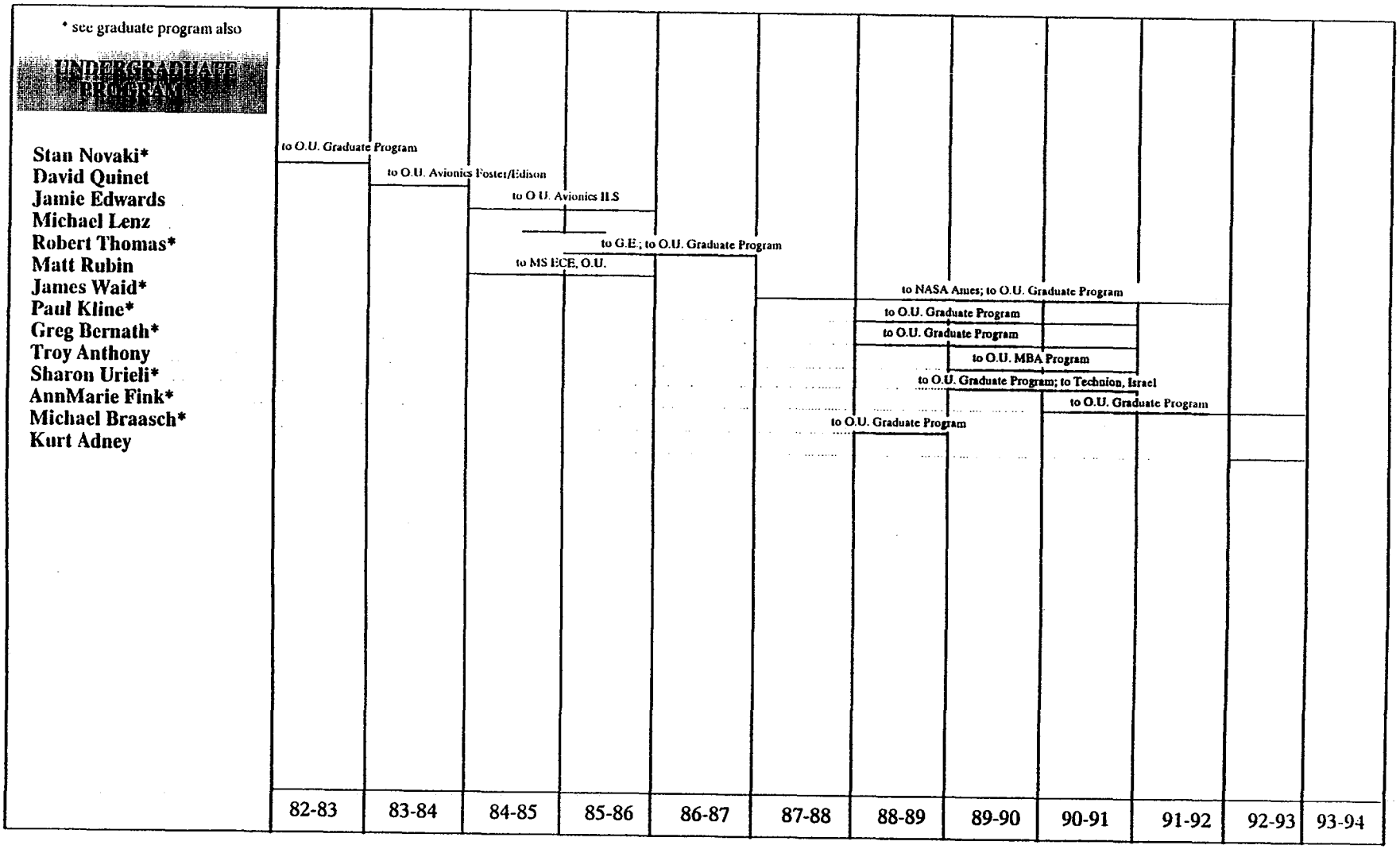
OHIO UNIVERSITY STUDENTS IN FAA/NASA JOINT UNIVERSITY PROGRAM FOR AIR TRANSPORTATION RESEARCH

The First Ten Years



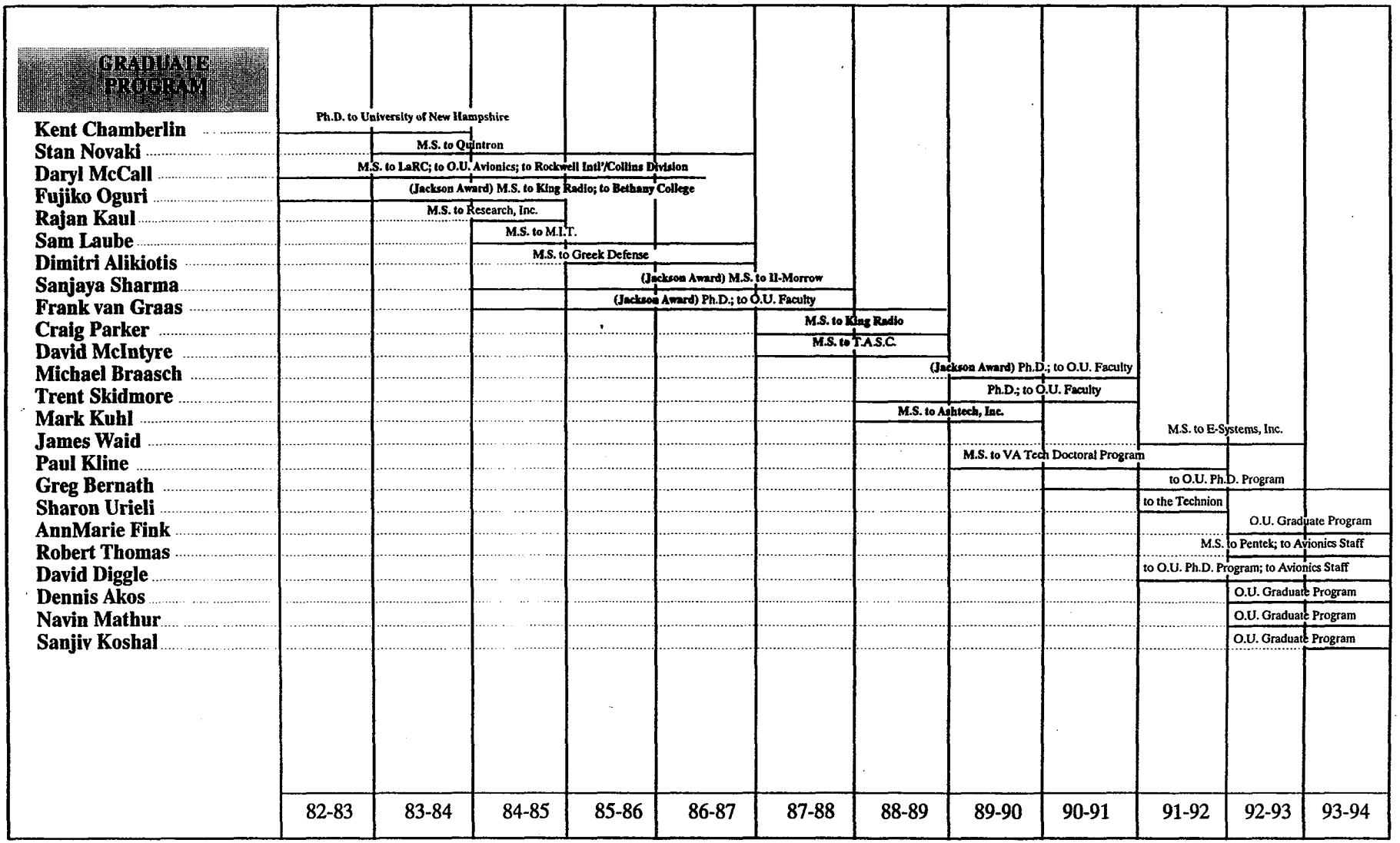
OHIO UNIVERSITY STUDENTS IN FAA/NASA JOINT UNIVERSITY PROGRAM FOR AIR TRANSPORTATION RESEARCH

The Second Ten Years



OHIO UNIVERSITY STUDENTS IN FAA/NASA JOINT UNIVERSITY PROGRAM FOR AIR TRANSPORTATION RESEARCH

The Second Ten Years Continued



FAULT DETECTION AND EXCLUSION IN MULTISENSOR NAVIGATION SYSTEMS

Gregory N. Bernath
Ohio University
Athens, Ohio

SUMMARY

In order for a multisensor navigation system to meet integrity requirements, there must be a way of detecting erroneous measurements, using only data from those measurements. This can be accomplished using a parity space estimation algorithm. Erroneous measurements must then be removed from the position solution; the entire process is called Fault Detection and Exclusion (FDE). A baseline FDE algorithm has been determined [1], and is capable of working in real time on present affordable hardware.

INTRODUCTION

Implementation of the satellite based Global Positioning System (GPS) provides the opportunity to greatly improve the safety and efficiency of air transportation. Although GPS in itself does not provide a sufficient level of availability, GPS augmented with the terrestrial Long Range Navigation System (Loran-C) or the Russian Federation's Global Navigation Satellite System (GLONASS) has the potential to satisfy the Required Navigation Performance (RNP) for all phases of flight except precision approach and landing.

Integrity refers to the ability of a navigation system to always perform within certain performance bounds, or to provide timely warning to the pilot if the system is not within these bounds. GPS presents a difficult problem in integrity checking, because there is currently no external integrity monitor for GPS which can provide timely warnings. Thus, any integrity checking performed by a GPS receiver must use only the existing GPS signals. This process in general is referred to as Fault Detection and Isolation (FDI), while the modified method described in this paper has been named Fault Detection and Exclusion.

Step errors or fast growing ramp errors can be detected with a recursive estimator such as a Kalman filter, which does not require redundant measurements. However, this will not detect slowly growing errors, such as a drifting clock. To detect such errors, a parity space algorithm can be used, which requires redundant measurements, the more the better.

FAULT DETECTION ALGORITHM

Estimation Space and Parity Space

Estimation space contains the real world measurements, and is where the navigation solution is obtained. Due to noise and biases, there is always some position error in the navigation solution. Since exact position is not known, position error is not known. Fault detection is done in parity space, which is a mathematical tool where noise and biases are used to create a parity vector. The parity vector is used to create a detection statistic d_k, which is compared to the detection threshold T_D to determine if an alarm condition exists.

Generally, a growing error in estimation space results in a growing detection statistic in parity space. This relationship will depend on the satellite geometries at that time. With a good geometry, a small position error with give a large (easily seen) detection statistic. The reverse can also be true.

Figure 1 illustrates the case of two different slowly growing ramp errors plotted in parity space vs. estimation space. Two of the parameters of interest are the detection threshold T_D, which is the alarm limit in parity space, and the alarm threshold T_A, which is defined as the allowable horizontal radial position error in the calculated user position. In case I, the detection threshold is crossed before the alarm threshold, resulting in a false alarm. As the error continues to grow, the alarm threshold is crossed, turning it into a true alarm. In case II, the alarm threshold is crossed before the detection threshold, resulting in a missed detection. As the error continues to grow, the detection threshold is crossed, turning it into a correct fault detection.

False Alarms and Missed Detections

A fundamental problem of fault detection is that all measurements contain noise; therefore fault detection always involves a degree of uncertainty. It is never possible to know for certain that an error exists; it is only possible to "guess" within a certain probability by examining a snapshot of parity space. The status of the algorithm is either normal operation, correct fault detection, false alarm or missed detection. The two undesirable events are a false alarm and a missed detection; therefore, probabilities of a false alarm (P_{FA}) and missed detection (P_{MD}) must be minimized. Traditionally, there has been a trade-off between false alarms and missed detections -- reducing one increases the other. However, having a high value for either one is unacceptable. Numerous false alarms reduce user confidence in the system, and missed detections can result in unacceptable and dangerous position errors. Therefore, instead of a trade-off between false alarms and missed detections, both probabilities will be traded off against the protection radius, which is the smallest horizontal position error that is guaranteed to be detected with the given probabilities of false alarms and missed detections. Instead of setting an alarm threshold and then

calculating the corresponding probabilities, first the probabilities are set and then the alarm threshold and protection radius are determined from the probabilities. This guarantees the accuracy of the navigation solution for all time and space points, and allows for changes in the accuracy requirements without the need to make major changes in the algorithm.

The Detection Statistic

A fault is declared if the detection statistic d_k exceeds the detection threshold T_D. The setting of the detection threshold depends on the measurement noise, the desired false alarm rate, and the number of redundant measurements. Consider the case of one redundant measurement, where the parity vector ($\underline{p}$) is reduced to a scalar (p), and $d_k = |p|$. In the case where no measurement bias exists, figure 2 illustrates the probability density function (pdf) of p when no measurement bias exists in any measurement, and measurement noise is normally distributed. The mean is zero because the pdf has not been shifted by any biases. Since no measurements are in error, the system is either in a state of normal operation or false alarm. The probability of a false alarm (P_{FA}) is obtained by integrating the areas outside T_D, which will be a standard Gaussian probability density function.

Figure 3 illustrates the case where a measurement bias does exist, causing real position error. This real position error "translates" into a parity scalar with a nonzero mean (The mathematics of this process are explained in the next section). On figure 3, the mean of this parity scalar is at 100 "parity meters". If the real position error is above the alarm threshold, the system is either in a state of correct fault detection or missed detection, depending on how the measurement noise affects the parity scalar. The probability of a missed detection (P_{MD}) is the integral of the area within T_D, which is again a Gaussian function.

These diagrams may give the impression that P_{FA} and P_{MD} are determined by T_D and σ, when actually the reverse is true -- T_D is determined by P_{FA} and σ, and P_{MD} determines another threshold, the protection radius. It is helpful to look at this on an estimation space/parity space plot. Figure 4 shows an ideal, no-noise case. The satellite geometries are known, so the slope of the bias path is known exactly. The detection threshold and alarm threshold can be adjusted to any position, so long as their intersection is crossed by the bias path. This results in a system with no chance of false alarms or missed detections.

Starting with figure 5, noise is added to the system, represented by the "noise ball". The bias is zero in this case, so the noise ball sits near the origin. The mean of the distribution is not at the origin, since total position error and thus the detection statistic must always be positive. The detection threshold is set far enough to the right so that the right hand edge of the distribution does not overlap into "false alarm" enough to exceed the given probability of false alarm. The corresponding alarm threshold has not been determined yet. Determining T_A is a major part of the algorithm, and will be fully covered later.

One goal of the system is to find how large the detection statistic must be before a

fault can be guaranteed to be identified with only a given small probability of missed detections. This value is called μ_M, the minimum detectable bias. In figure 6, μ_M is set too low, resulting in an unacceptably large P_{MD}. In figure 7, μ_M is set correctly, as almost none of the distribution overlaps into "missed detection". Once T_D and μ_M are known, these are used along with measurement noise and satellite geometry to determine what T_A should be. The following section presents the fault detection algorithm in detail.

Baseline Fault Detection Algorithm

The linearized relation between changes in the measurements and the corresponding change in the user state vector is given by

$$\underline{\delta y} = H\ \underline{\delta x} \tag{1}$$

where: $\underline{\delta y}$ = measurement vector (m × 1)
$\mathbf{H}$ = data matrix (m × n)
$\underline{\delta x}$ = user state vector (n × 1)

$\underline{\delta y}$ is a column vector of m elements containing the changes in measurements from m sources. If only GPS satellites were used, it would consist of the pseudorange to each satellite. $\underline{\delta x}$ is an n-element column vector of the change in user states, consisting of user position coordinates and other navigation state elements such as clock offset with respect to GPS time. $\mathbf{H}$ is an m × n matrix which relates the measurements to the user states.

Given the formulation in (1), three possible cases can occur:
1) m < n : Underdetermined system
2) m = n : Exactly determined system (provided H has full rank)
3) m > n : Overdetermined system

In the underdetermined case, a navigation solution is not possible. In the exactly determined case, a navigation solution is possible, but fault detection is not. Algorithms for managing the redundant measurements in overdetermined systems form the basis of fault detection.

The data matrix $\mathbf{H}$ can be decomposed into the product of a real m × m orthonormal matrix $\mathbf{Q}$ ($\mathbf{Q}^T\mathbf{Q} = \mathbf{I}$) and an m × n upper triangular matrix $\mathbf{R}$ using a "QR" factorization [2],

$$H = QR \tag{2}$$

$\mathbf{R}$ contains (m-n) rows of zeros along the bottom, due to the (m-n) redundant measurements in $\mathbf{H}$. Substituting (2) into (1) and pre-multiplying both sides by $\mathbf{Q}^T$ yields

$$\begin{aligned} \underline{\delta y} &= QR\underline{\delta x} \\ Q^T\underline{\delta y} &= Q^TQR\underline{\delta x} \\ R\underline{\delta x} &= Q^T\underline{\delta y} \end{aligned} \tag{3}$$

Now partition R into an n × n upper triangular matrix **U** and (m-n) rows of zeros, denoted by **0**. Similarly, partition $\mathbf{Q}^T$ into $\mathbf{Q}_x$ (n × m) and $\mathbf{Q}_p$ ((m-n) × m).

$$\begin{pmatrix} U \\ --- \\ 0 \end{pmatrix} \begin{pmatrix} \delta x_1 \\ \vdots \\ \delta x_n \end{pmatrix} = \begin{pmatrix} Q_x \\ --- \\ Q_p \end{pmatrix} \begin{pmatrix} \delta y_1 \\ \vdots \\ \delta y_m \end{pmatrix} \tag{4}$$

Relating the change in measurements to the change in the user state vector, the least squares navigation state solution is

$$U\ \underline{\delta x} = Q_x\ \underline{\delta y} \quad \text{or} \quad \underline{\delta x} = U^{-1}\ Q_x\ \underline{\delta y} \tag{5}$$

and the parity equation is

$$\underline{0} = Q_p\ \underline{\delta y} \tag{6}$$

Each column of $\mathbf{Q}_p$ defines a measurement axis in parity space, and an error in measurement i will lie along the ith measurement axis in parity space.

For measurements corrupted by errors, $\underline{\delta y}$ is replaced by $\underline{\delta y} + \underline{e} + \underline{b}$, where $\underline{e}$ is an m × 1 vector representing zero-mean, normally distributed measurement noise, and $\underline{b}$ is an m × 1 vector containing bias errors. If $\underline{\delta y}$ in equation (6) is replaced by $\underline{\delta y} + \underline{e} + \underline{b}$, the zero can be replaced with the parity vector $\underline{p}$, giving

$$\begin{aligned} \underline{p} &= Q_p\underline{\delta y} + Q_p\underline{e} + Q_p\underline{b} \\ \underline{p} &= Q_p\underline{e} + Q_p\underline{b} \end{aligned} \tag{7}$$

$\underline{p}$ is a function of measurement noise and bias errors. Therefore, it can be used as a detection function for declaring faults.

One Redundant Measurement

If only one redundant measurement is available (that is, m = n + 1), parity space is one-dimensional and $\mathbf{Q}_p$ is reduced to a row vector $\underline{q}$. The parity vector $\underline{p}$ is reduced to a parity scalar p. In the absence of bias errors, p will have a zero-mean normal probability density function given by

$$f_p(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\left(\frac{x}{\sigma\sqrt{2}}\right)^2} \tag{8}$$

If the failure detection is based on exceeding a detection threshold T_D, then the probability of an alarm (P_A) is

$$P_A = P\,(|p|>T_D) = \frac{2}{\sigma\sqrt{2\pi}} \int_{T_D}^{\infty} e^{-\left(\frac{x}{\sigma\sqrt{2}}\right)^2} dx \tag{9}$$

which can also be written as

$$P_A = \text{erfc}\left(\frac{T_D}{\sigma\sqrt{2}}\right) \tag{10}$$

where erfc is the complementary error function,

$$\text{erfc}(z) = \frac{2}{\sqrt{\pi}} \int_{z}^{\infty} e^{-\lambda^2} d\lambda \tag{11}$$

See figure 2 for a pictorial representation of equation (11). Note that if there is no measurement bias, any alarm is necessarily a false alarm, and $P_A = P_{FA}$.

In the presence of a bias error in instrument i (or more than one instrument), the parity scalar has a normal distribution with a mean of

$$\mu_i = |\underline{q}\cdot\underline{b}| \tag{12}$$

If instrument i has a bias, then only the ith elements of $\underline{q}$ and $\underline{b}$ will contribute to the mean. The probability density function of p is then given by

$$f_p(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\left(\frac{x-\mu_i}{\sigma\sqrt{2}}\right)^2} \tag{13}$$

Given a detection threshold T_D, the probability of a missed detection is

$$P_{MD} = P\,(|p|\leq T_D) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-T_D}^{T_D} e^{-\left(\frac{x-\mu_i}{\sigma\sqrt{2}}\right)^2} dx \tag{14}$$

see figure 3 for a pictorial representation of this equation. This can also be written as

$$P_{MD} = \frac{1}{2}\,\mathrm{erfc}\left(\frac{\mu_i - T_D}{\sigma\sqrt{2}}\right) - \frac{1}{2}\,\mathrm{erfc}\left(\frac{\mu_i + T_D}{\sigma\sqrt{2}}\right) \tag{15}$$

Since the integral from $-\infty$ to $-T_D$ is negligible for sufficiently large μ_i, the second term in equation (15) can be ignored, leaving

$$P_{MD} = \frac{1}{2}\,\mathrm{erfc}\left(\frac{\mu_i - T_D}{\sigma\sqrt{2}}\right) \tag{16}$$

Equations (10) and (16) define the performance of the fault detection algorithm in terms of probability of false alarm (P_{FA}) and probability of missed detection (P_{MD}) as a function of:

1. Detection threshold (T_D)
2. Measurement noise standard deviation (σ)
3. The expected value μ_i of the absolute value of the parity scalar p which resulted from measurement bias in instrument i.

A given P_{FA}, P_{MD} and σ can be used to obtain suitable T_D and μ_M (the minimum expected value of the parity scalar required for detection). The detection threshold is obtained using P_{FA} and equation (10)

$$T_D = \sigma\sqrt{2}\ \mathrm{erfc}^{-1}(P_{FA}) \tag{17}$$

Next, using this value of T_D, plus P_{MD}, equation (16) gives

$$\mu_M = T_D + \sigma\sqrt{2}\ \mathrm{erfc}^{-1}(2P_{MD}) \tag{18}$$

Equation (18) implies that P_{MD} is satisfied only if the bias error gives rise to a μ_i such that $\mu_i = |\underline{q}\,\underline{b}| \geq \mu_M$. The vector $\underline{q}$ is known from measurement geometry; therefore, for each measurement i (i = 1 through m), the minimum bias error b_i required to satisfy the probability of a missed detection is calculated from equation (12).

$$b_i = \frac{\mu_M}{|q_i|} \tag{19}$$

So given P_{FA}, P_{MD} and σ, it follows that the minimum detectable bias error is a function of the measurement geometry. The minimum detectable bias error can then be used to calculate the protection radius, the smallest horizontal position error which is guaranteed to be detected within the given probabilities of false alarms and missed detections. This solves the problem of choosing the estimation space alarm threshold T_A.

Two or More Redundant Measurements

If two or more redundant measurements are available, the parity vector will have two or more elements, and will be a vector instead of a scalar. Each element of the parity vector has a normal probability density function as given (for the without bias case) by equation (8). Recall that each column of $\mathbf{Q}_p$ defines a measurement axis in parity space, and an error in measurement i will lie along the ith measurement axis. The number of axes will equal the number of measurements, and the axes will lie in a parity space of (m-n) dimension (that is, dimension = number of redundant measurements).

To maximize the visibility of a bias error, the parity vector is projected on to each of the measurement axes. The norm of the projection is

$$r_i = \frac{|\underline{m}_i \cdot \underline{p}|}{|\underline{m}_i|} \tag{20}$$

where $\underline{m}_i$ is the ith column of $\mathbf{Q}_p$. Each r_i given by equation (20) has a Gaussian distribution, since the above operation is equivalent to a rotation in parity space. Instead of one detection statistic, m detection statistics are obtained. Although m tests are performed, these tests are not independent and have only (m - n) degrees of freedom. Therefore, the detection threshold is given by

$$P_{FA} = (m-n) \times \mathrm{erfc}\left(\frac{T_D}{\sigma\sqrt{2}}\right) \tag{21}$$

$$T_D = \sigma\sqrt{2}\ \mathrm{erfc}^{-1}\left(\frac{P_{FA}}{m-n}\right) \tag{22}$$

In a manner similar to equation (19), the minimum bias error required to satisfy the probability of missed detection for instrument i is

$$b_i = \frac{\mu_M}{|\underline{m}_i|} \tag{23}$$

FAULT EXCLUSION ALGORITHM

The ultimate objective of Fault Detection and Isolation is to ensure that the aircraft is flying with a good position estimate. It is not necessary to isolate a bad measurement; it is only required that the bad measurement is not used in the navigation solution. With this in mind, Fault Detection and Exclusion was devised.

In FDE, once an alarm is raised, the algorithm discards the present combination of satellites and looks for the combination with the next best protection radius. If this set also raises an alarm, the algorithm goes on to the next best set, and so on. Once a set is found that doesn't raise an alarm, that set will be used for navigation. In this manner, the bad measurement is not necessarily isolated, but it is excluded from the navigation solution if it raises an alarm.

HARDWARE RESTRICTIONS

The fault detection algorithm works best when it uses all available satellite signals. Adding more measurements will always improve performance, though sometimes only slightly. Therefore, it should not be a minimum requirement to insist that the receiver tracks all available satellites. In addition, most GPS receivers are not capable of tracking all visible satellites. In order to reduce cost, many GPS receivers are limited to six channels, and can track only six satellites. In order for the algorithm to be useful in practice, it must be functional under this condition.

The solution is to use a "best-of-six" algorithm. This method generates all possible subsets of six satellites from the set of satellites in view, and calculates a protection radius for all of them. The set of six with the smallest protection radius becomes the one the receiver uses to navigate. This may seem computationally prohibitive; however, the calculation of all protection radii only needs to be redone when there is a significant change in the satellite geometry. Recalculating once every five to ten minutes should be sufficient.

Figure 8 shows the protection radius at one different locations over a span of one day using both the best-of-six method and using all satellites in view (based on a 24 satellite constellation). The parameters used were σ = 32m, $P_{MD} = 1x10^{-3}$, and $P_{FA} = 6.67x10^{-5}$. As a reference, with these parameters, non-precision approach requirements would be satisfied by a protection radius of 555m or less [3]. As expected, the all-in-view method is the best, but only slightly more so than the best-of-six. If a receiver can use more than six satellites, it should; otherwise, the best-of-six method would be sufficient.

MEASUREMENT WEIGHTING

Up until now, it has been assumed that all measurements have the same statistics. If this is not the case, the fault detection algorithm can be modified to take differing statistics into account. To accommodate different measurement variances, equation (1) is left multiplied by a weighting matrix W

$$W\underline{\delta y} = WH\underline{\delta x} \qquad (24)$$

The QR factorization is now done on **WH**, and equation (3) is replaced by

$$R\underline{\delta x} = Q^T W\underline{\delta y} \tag{25}$$

This leads to a parity equation of

$$0 = Q_p W\underline{\delta y} \tag{26}$$

and a parity vector defined as

$$\underline{p} = Q_p W\underline{e} + Q_p W\underline{b} \tag{27}$$

In general, **W** can be derived from the measurement noise covariance matrix, but in most applications it is sufficient to simply use a diagonal matrix, where the diagonal elements are the inverse of the measurement noise standard deviation. If **W** is correctly chosen, then the elements of $W\underline{\delta y}$ have unit standard deviation.

One practical way of using measurement weighting is to include the baro-altimeter. Since all aircraft have one, this effectively adds another measurement to the algorithm. Since the noise and biases of the baro-altimeter change during different phases of flight, this has to be taken in to account by changing the **W** matrix. Figure 9 shows a comparison of protection radius for the best-of-six method alone, and the best-of-six method with altimeter aiding. As expected, using the altimeter improves the protection radius significantly.

CONCLUSIONS

The fault detection algorithm has been fully characterized and verified. Any assumptions made have erred towards being overly conservative. To guarantee the integrity of the navigation solution at all times and locations, the probability of a false alarm should not be traded off against the probability of missed detection. Instead, both should be traded off against the protection radius.

The fault detection algorithm is most effective when measurements from all satellites in view are utilized. However, many receivers can only use six satellites. For these cases, a best-of-six algorithm has been developed which performs almost as well, and is thus an acceptable solution.

A measurement from the baro-altimeter can be added by way of a weighting matrix, and will improve performance considerably. Whenever possible, this measurement should be included.

An erroneous measurement which raises the alarm need not be isolated (Fault

Detection and Isolation). Instead, it can be excluded (Fault Detection and Exclusion). As long as the bad measurement is not included in the navigation solution, it isn't necessary to isolate it.

REFERENCES

[1] Bernath, G. N., "A Baseline Fault Detection and Exclusion Algorithm for the Global Positioning System," Masters Thesis, Ohio University, Department of Electrical and Computer Engineering, March 1994.

[2] Golub, G. H. and Van Loan, C. F., Matrix Computations, Second Edition, The Johns Hopkins University Press, Baltimore, MD, 1990.

[3] van Graas, F., and Farrell, J. L., "Receiver Autonomous Integrity Monitoring (RAIM): Techniques, Performance, and Potential," Proceedings of the 47th Annual Meeting of The Institute of Navigation, Williamsburg, VA, June 10-12, 1991.

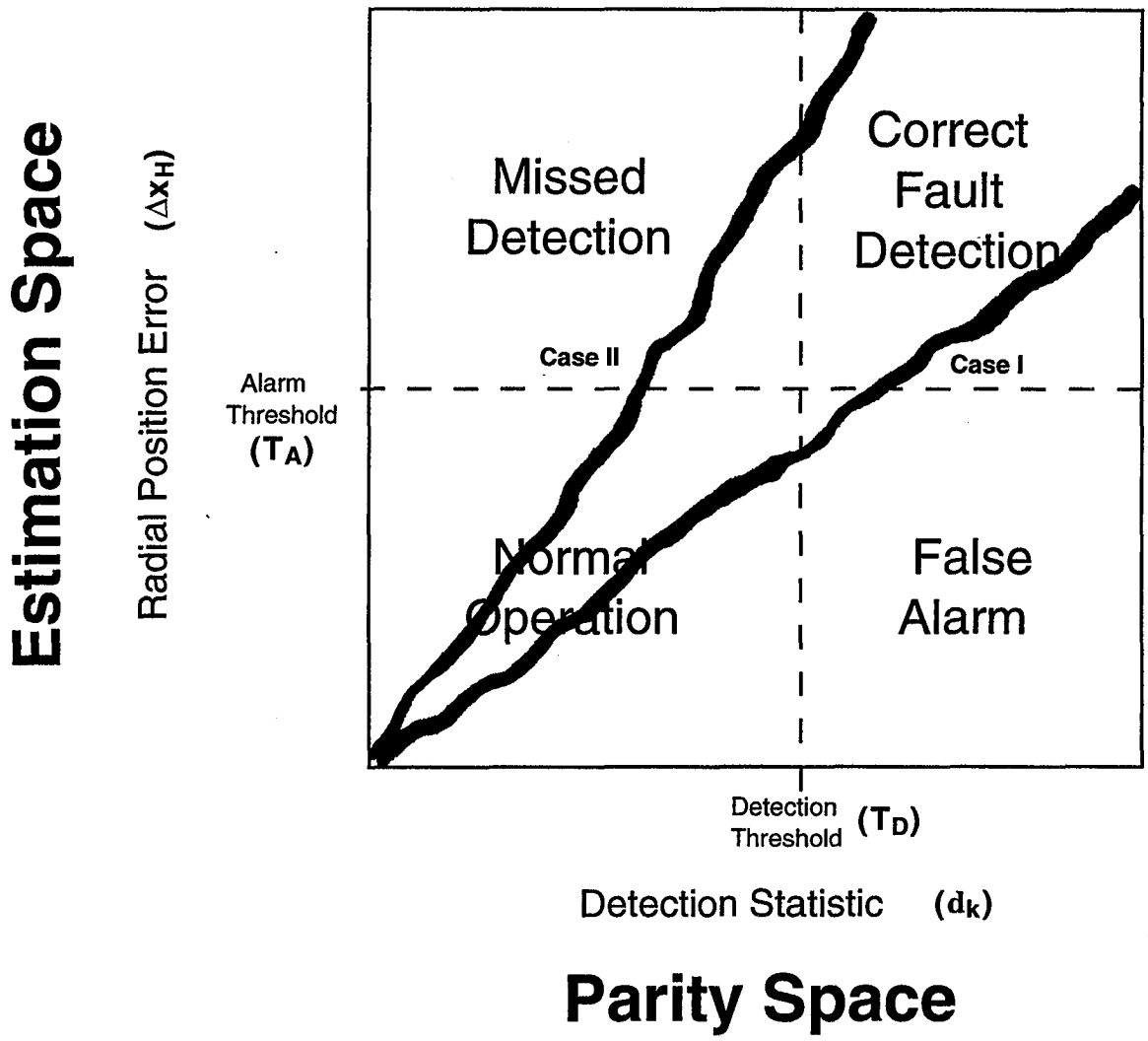


Figure 1. Two slowly growing measurement biases plotted in estimation space and parity space.

Figure 2. Probability Density Function for p - no measurement bias

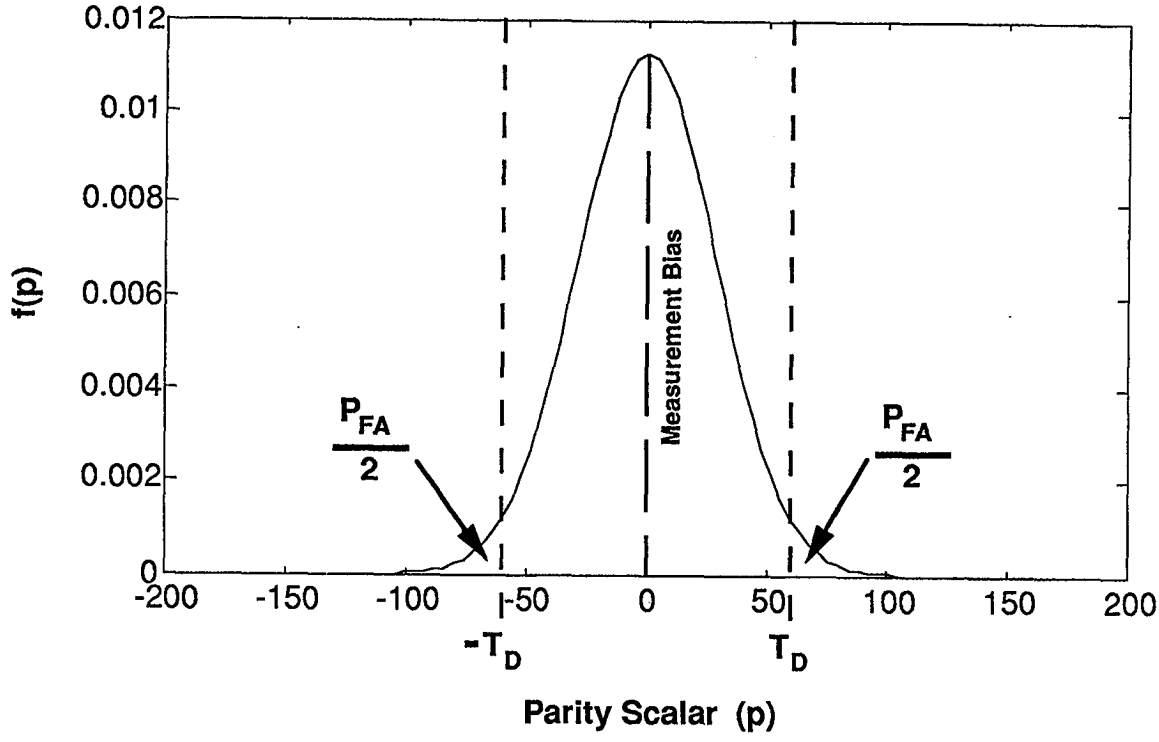
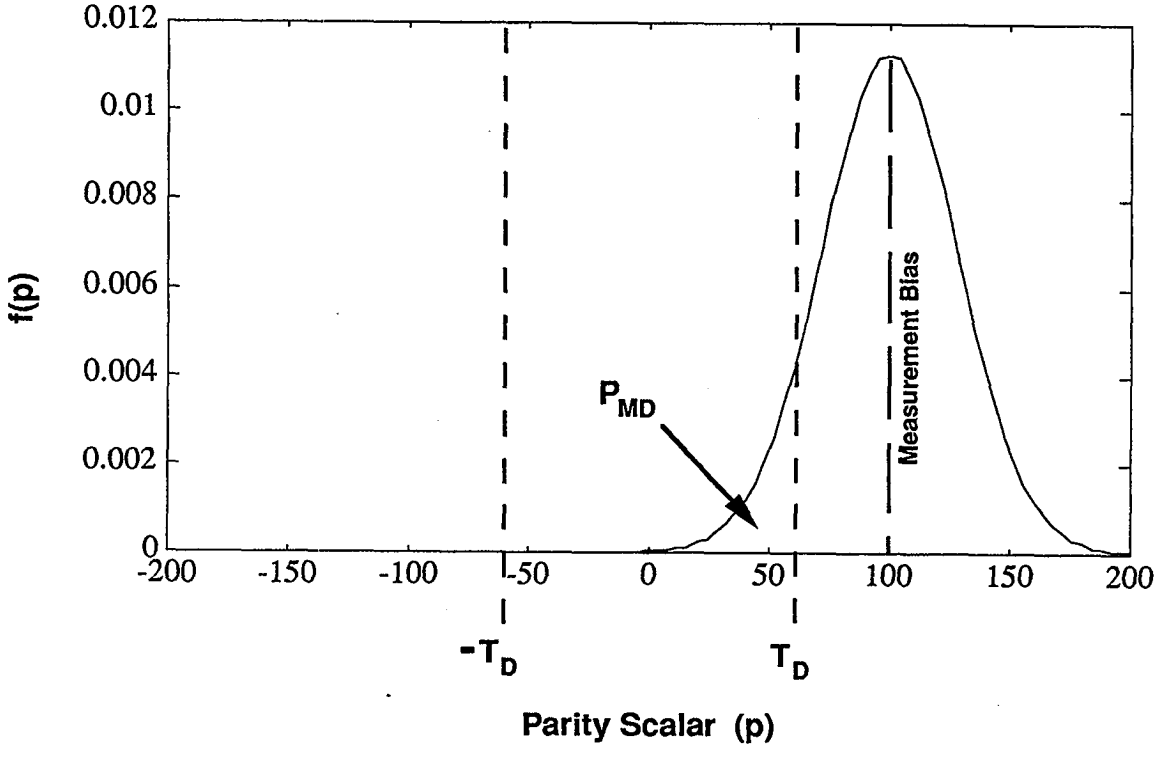


Figure 3. Probability Density Function for p - with measurement bias



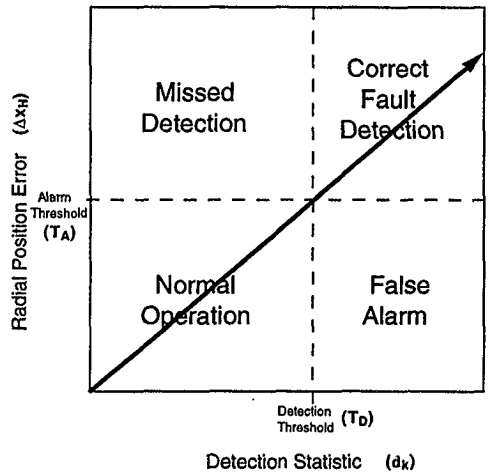


Figure 4. A growing measurement bias plotted in estimation space and parity space (ideal, noise-free case)

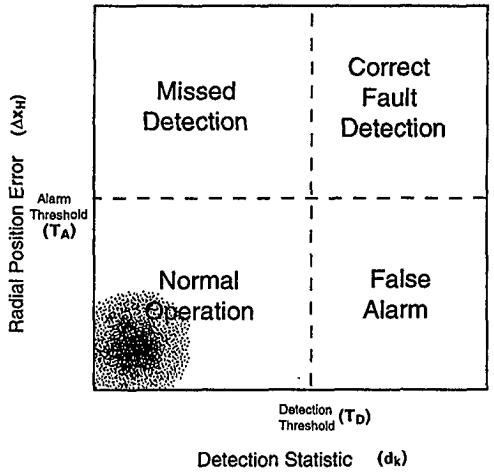


Figure 5. Estimation Space and Parity Space Zero Bias

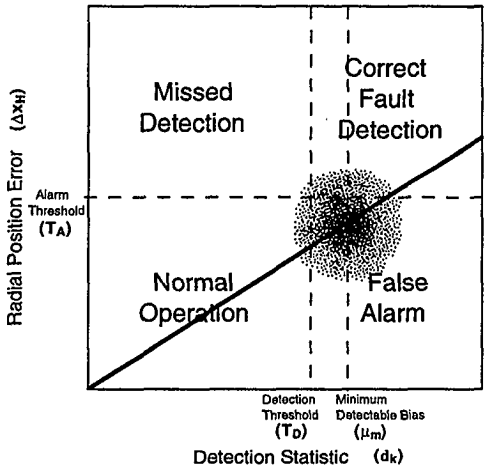


Figure 6. A slowly growing measurement bias with noise in estimation space and parity space, with minimum detectable bias set too low.

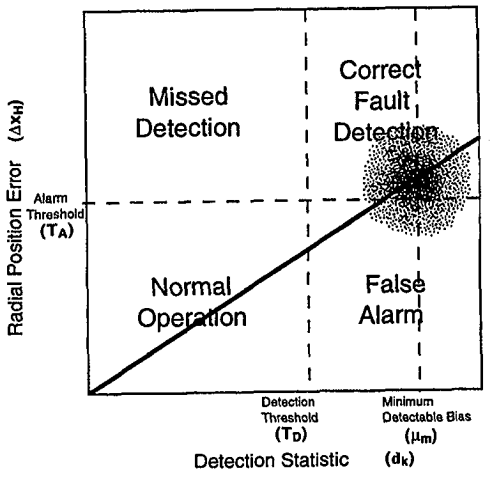


Figure 7. A slowly growing measurement bias with noise in estimation space and parity space, with minimum detectable bias set correctly.

Figure 8. Protection Radius (36N 140E)

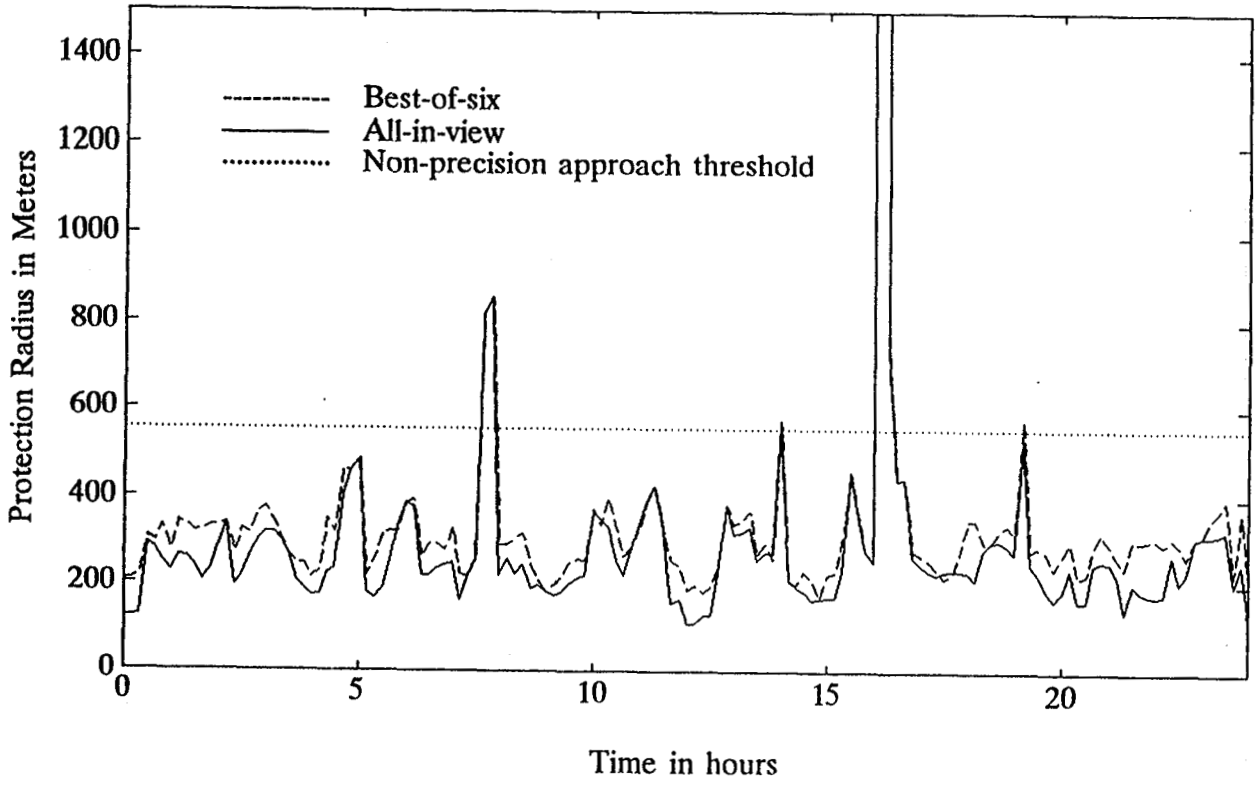
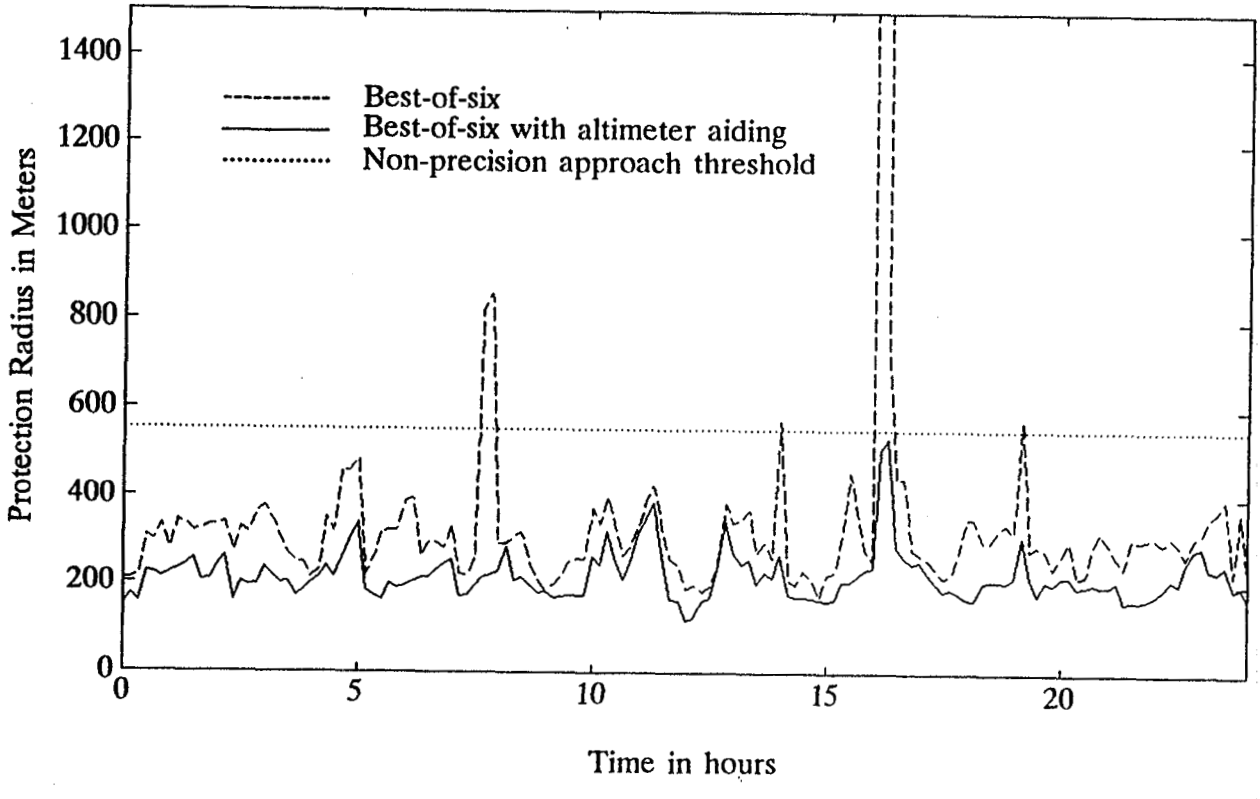


Figure 9. Protection Radius (36N 140E)



HYBRID SYSTEM
GMSK DIGITAL RECEIVER IMPLEMENTATION
IN
REAL TIME

Sanjiv Koshal
Ohio University
Athens, Ohio

SUMMARY

This paper is concerned with the design, simulation and implementation of a hybrid system using the GMSK type of signal format for phase modulation and demodulation. The performance of the designed transceiver structure is evaluated using the Bit Error Rate (BER) curves. The simulated system was also successfully implemented in real time.

INTRODUCTION

The transmission of real time weather data to the cockpit will greatly help in avoiding major accidents that have been taking place practically from the beginning of aviation history. These accidents caused by the unavailability of real time weather data to the pilot, are second rated only to pilot error as reported by the latest statistics given by the National Transportation Safety Board [1].

As airways become more and more crowded worldwide, improved weather information is a solution to maintaining flight safety and lowering operational costs. A group of organizations has come together to resolve this problem and have put forward several methods of displaying real time weather data both for civilian and military users [2]. They are proposing to use satellites, weather estimation software, cockpit avionics, etc. to accomplish this task. These organizations include: Advanced Satellite Telecommunications, Inc., Conwal, Inc., Garcia Consulting, ARINC, Lockheed, FAA, and NASA. The proposed system intends to use an onboard receiver that is simple and inexpensive. During landing, takeoff, and congested traffic conditions in the air space, the Air Traffic Control (ATC) is often unable to supply the pilot with the necessary weather data information that might be needed when approaching bad weather to determine a go/no-go situation [3] . In addition, this voice data information provided to the pilot is extremely limited therefore, and a better system needs to be developed. An ideal uplink would not affect the existing system and simultaneously provide the pilot with more comprehensive real time weather information.

HYBRID SYSTEM: POSSIBLE SIGNAL STRUCTURES FOR PHASE MODULATION

A hybrid modulation is where both amplitude modulation and phase modulation are applied to the same carrier. The existing AM signal can be modulated and demodulated using available equipment thus eliminating costly replacements. The additional phase modulation, representing the weather data, can be accomplished by a modification to the existing AM transmitter. In order to receive the signal, an additional demodulator would be necessary. This demodulator would cost far less than comparable airborne weather radar systems and would provide much more extensive information since the reflectivity patterns are generated by powerful ground-based systems.

Previous studies have looked into the possibility of using a coherent MSK signal structure for phase modulation and demodulation [4]. The proposed system resulted in a 2-4 dB loss in BER performance. MSK provides better results in terms of voice degradation when used in the hybrid signal than QPSK and BPSK because less spectral energy is removed from the phase modulation through the transmission process. This is due to the fact that MSK has a much more compact power spectrum than that of QPSK and BPSK.

Although MSK was thought to be the modulation of choice for bandwidth conservation, recently developed modulation techniques improve upon the properties of MSK. A class of these modulation methods exist in which the envelope of the signal remains constant, a requirement for the hybrid signal [5]. Recent developments include the smoothing of the phase path of these modulation methods by incorporating memory into bit transitions. Applying these techniques in a hybrid modulated signal will provide superior performance, for both the amplitude and phase modulations.

A study done by Dennis Akos at the Ohio University Avionics Engineering Center established that a new modulation scheme, termed as Gaussian Minimum Shift Keying (GMSK), would cause less interference in voice communications, when used for the hybrid system [6]. The GMSK type of implementation was proposed to be applied in fields of digital mobile radio telephony. The European Conference of Postal and Telecommunications Administration selected this type of modulation scheme for the Pan-European Cellular Radio System [7].

The latest simulation trials performed concluded that a higher data rate can be achieved provided the spectral efficiency of the phase modulation is greater. By using MSK as the phase modulation at 2400 b/s, a mean square error (MSE) of approximately 7.5 e-4 was introduced into the AM signal. For GMSK with a transmission filter bandwidth BT of 0.25, an acceptable MSE for data rates of up to 3900 b/s was achieved. It is evident from this study done by Dennis Akos that the use of GMSK for the phase modulation of the hybrid system would improve performance.

SIMULATION SET UP FOR BER EVALUATION OF PHASE MODULATION/DEMODULATION

The overall hybrid system BER simulation was done at baseband using the setup depicted in Figure 1. A sum of five tones, given by

$$f(t) = \sum_{i=1}^{5} a_i \cos(2\pi f_i t) \tag{1}$$

in which the coefficients a_i and f_i (in hertz) have the values indicated in Table 1 [4], is used to simulate a voice signal.

Amplitude (volts)		Frequencies (Hz)	
a_1	2/9	f_1	468.75
a_2	1/3	f_2	937.85
a_3	2/9	f_3	1406.25
a_4	1/9	f_4	1875.00
a_5	1/9	f_5	2343.75

Table 1: Values for a_i and f_i (in hertz) for equation (1)

The output of the voice generator is multiplied by the phase modulated signal structure at baseband. GMSK is used for phase modulation. The binary data used to phase modulate the signal was kept at a constant data rate of 2400 b/s. Since the simulation was performed at baseband; a lowpass equivalent of the channel filters were used with the following characteristics [4]:

(i) The transmitter filter was second order butterworth, with -3 dB bandwidth of 7.5 kHz.

(ii) The receiver filter was fourth order butterworth, with -3 dB bandwidth of 5 kHz.

The voice signal and the phase modulated signal are multiplied to form the overall hybrid signal. Complex Additive White Gaussian Noise (AWGN) is added to the transmitted signal. The amplitude envelope on the received signal is removed using a hard limiter followed by a 4th order butterworth filter having a -3dB bandwidth of 3 kHz. Two bit or

one bit non coherent detection schemes can be employed for the phase demodulation of the hybrid signal [8]. The GMSK demodulator block utilizes two bit differential detection which can also be replaced by one bit detection scheme. Both detection schemes were evaluated for phase demodulation.

The probability of bit error for the simulated system is evaluated as follows:

Let

P_s = Average power of the transmitted signal

P_n = Average noise power injected into the system

σ^2 = variance of the noise injected into the system

R = data rate of the transmitted digital signal

f_s = sampling frequency at which the system is simulated

N_o is a positive constant, representing the one-sided spectral noise density

Now

$$E_b = \frac{P_s}{R}; \tag{2}$$

The noise power is given as (Couch, 1989)

$$P_n = N_o \times f_s \tag{3}$$

For complex noise, the noise power is given by twice the variance of the noise

$$P_n = 2 \times \sigma^2; \tag{4}$$

and therefore combining equations (3) and (4), we get

$$2 \times \sigma^2 = N_o \times f_s \tag{5}$$

Thus

$$N_o = \frac{(2 \times \sigma^2)}{f_s} \tag{6}$$

Combining (2) and (6)

$$\frac{E_b}{N_o} = \frac{(P_s \times f_s)}{(2 \times \sigma^2)} \tag{7}$$

We can find P_s by adding a complex average power estimator at the point where the signal is being transmitted in the simulated system. Also the sampling frequency and the variance of the injected noise is calibrated for the simulation. Thus we can calculate the required E_b/N_o for our system.

The simulation setup was used to obtained the BER curves for both one bit and two bit differential detection schemes for phase demodulation of the hybrid system. Equation 7 was used to calculate the required E_b/N_o for the hybrid system. The resulting BER curve for the case of infinite-bandwidth and constant envelope data modulation using GMSK and that for hybrid system using GMSK is as shown in Figure 2. The computer simulations show a loss of about 2-4 dB with respect to the case of infinite-bandwidth and constant envelope data modulation using GMSK signal structure. This data is consistent with prior work using coherent detection methods.

A variation of the 3 dB bandwidth product, $B_t \cdot T$, of the transmitting Gaussian filter, results in a corresponding change in the spectrum of the transmitted signal. Variation of $B_t \cdot T$ were employed for the hybrid system and the respective BER curves were plotted in Figure 3. A value of 0.75 for B_t ,the 3 dB bandwidth product of the transmitting Gaussian filter, provides the best performance when compared with the other two values tested. This is due to the fact that there is less intersymbol interference in the transmitted signal and also an increase in the bandwidth of the transmitted signal.

A comparison of the BER curves for the hybrid system using a non coherent MSK receiver and a differential GMSK receiver, with a $B_t \cdot T$ of 0.75, is as shown in Figure 4. Although the non coherent MSK hybrid system does provide a slightly better performance then the GMSK hybrid system, the use of the GMSK modulation/demodulation is preferred due to restrictions imposed on us for the occupancy of the bandwidth by the transmitted signal and resulting degradations in the AM signal.

In an ideal situation the two bit detection scheme provides better performance when compared to the one bit detection scheme [8]. However, the simulation involves finite bandwidth and amplitude modulation. The one bit detection scheme was also evaluated for the hybrid system for phase demodulation and the BER curve is as plotted in Figure 5. From the plot it is apparent that the two bit detection scheme does provide better performance.

CONCLUSIONS

This paper describes the use of a GMSK type of signal structure for phase modulation/demodulation in the hybrid system. The use of a GMSK type of non coherent detection scheme has been discussed and has been proven to show an improvement in system performance as compared to the use of MSK signal structure in terms of voice degradation. The computer simulations for the phase modulation/demodulation of the digital data has

derived the performance BER curve. The use of GMSK thus allows the transmission of the hybrid signal in a much tighter bandlimited spectrum than MSK, which in turn introduces less voice degradation at equal data rates. The use of such a receiver structure can be used on a mobile platform, as Doppler shift has minimal effect on the receiver performance.

ACKNOWLEDGEMENTS

This work was supported by the Federal Aviation Administration and the National Aeronautics and Space Administration through the Joint University Program for Air Transportation Research (Grant NGR 36-009-17). The author wishes to express his thanks to Dr. Frank van Grass for his guidance and to Dennis Akos for his assistance.

REFERENCES

[1] National Transportation Safety Board, Annual Review of Aircraft Accident Data, U.S. General Aviation, Calendar Year 1988, NTSB/ARG-91/01, Washington D.C., March 21, 1991.

[2] Hart C. Damon, "Real-Time Weather Data in the Cockpit", Avionics, November 1993.

[3] National Transportation Safety Board, Annual Accident Report, Delta Airlines, Inc., Lockheed L-1011-385-1, N726DA, Dallas/Fort Worth International Airport, Texas, August 2, 1985, NTSB/AAR-86/05, Washington D.C., August 15, 1986.

[4] Benelli, G. , and Fantacci, R. , "An Integrated Voice-Data Communication System for VHF Links", IEEE Transactions on Communications, VOL. COM-31, NO. 12, December 1983.

[5] Sundberg, Carl-Erik, "Continuous Phase Modulation", IEEE Communications Magazine, April 1986.

[6] Akos, Dennis, A Hybrid Modulation for the VHF Aeronautical Channels, Master Thesis, Ohio University, 1992.

[7] McGrath and Burkley, C. J. , "Low Complexity GMSK Modulator and Demodulator for Integrated Circuit Implementation", Electronic Engineering Dept., University of Limerick, Ireland, 1990.

[8] Simon, K. ,Marvin and Wong, C. , Charles " Differential Detection of Gaussian MSK in a Mobile Radio Environment," IEEE Trans. Veh. Technol.,vol. VT-33, NO.4, pp. 307-320, November 1984.

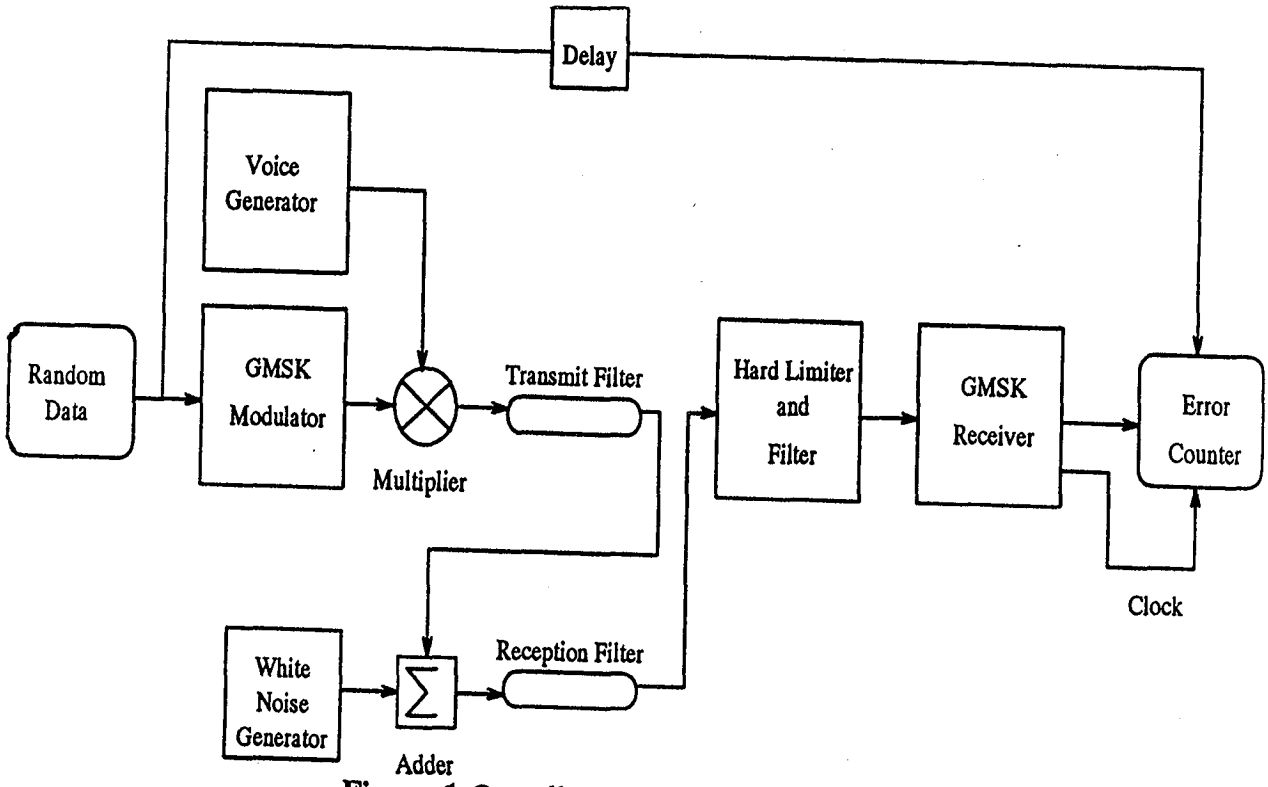


Figure 1 Overall Setup for BER Evaluation

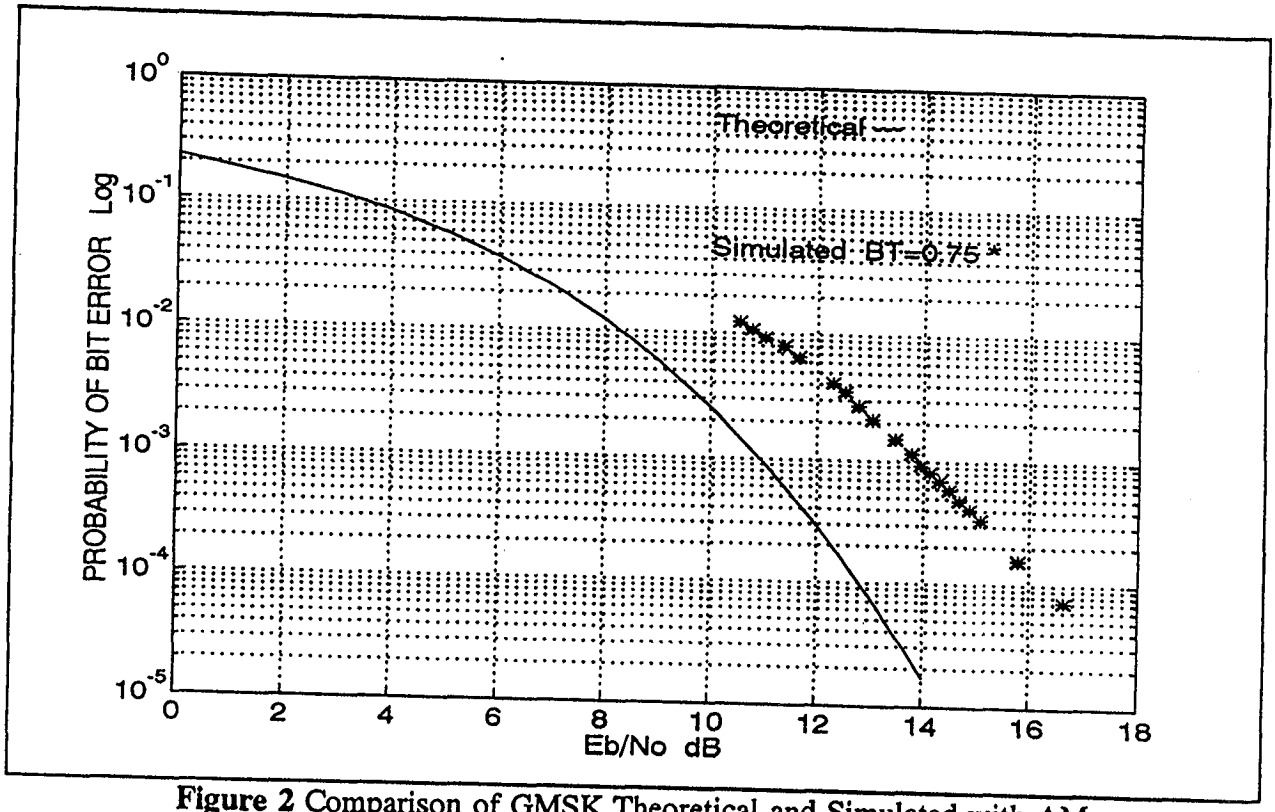


Figure 2 Comparison of GMSK Theoretical and Simulated with AM

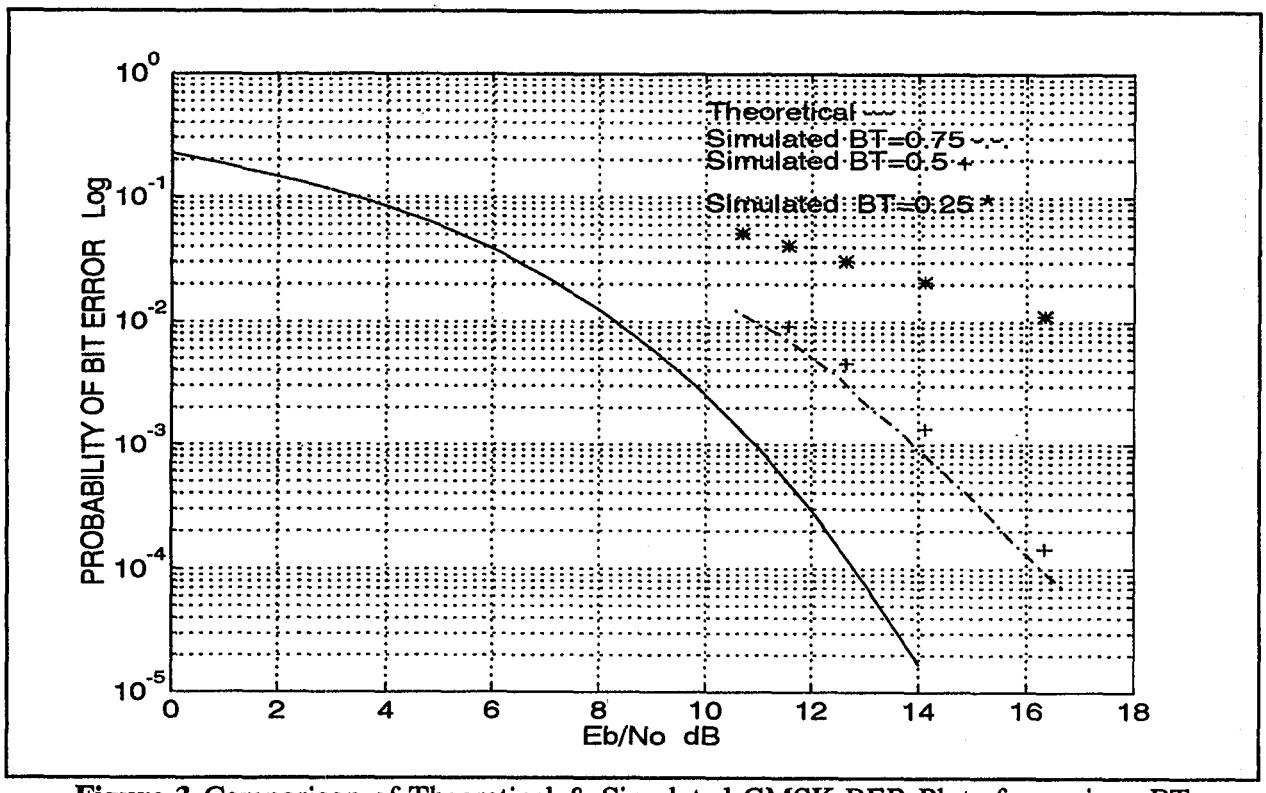


Figure 3 Comparison of Theoretical & Simulated GMSK BER Plots for various BT

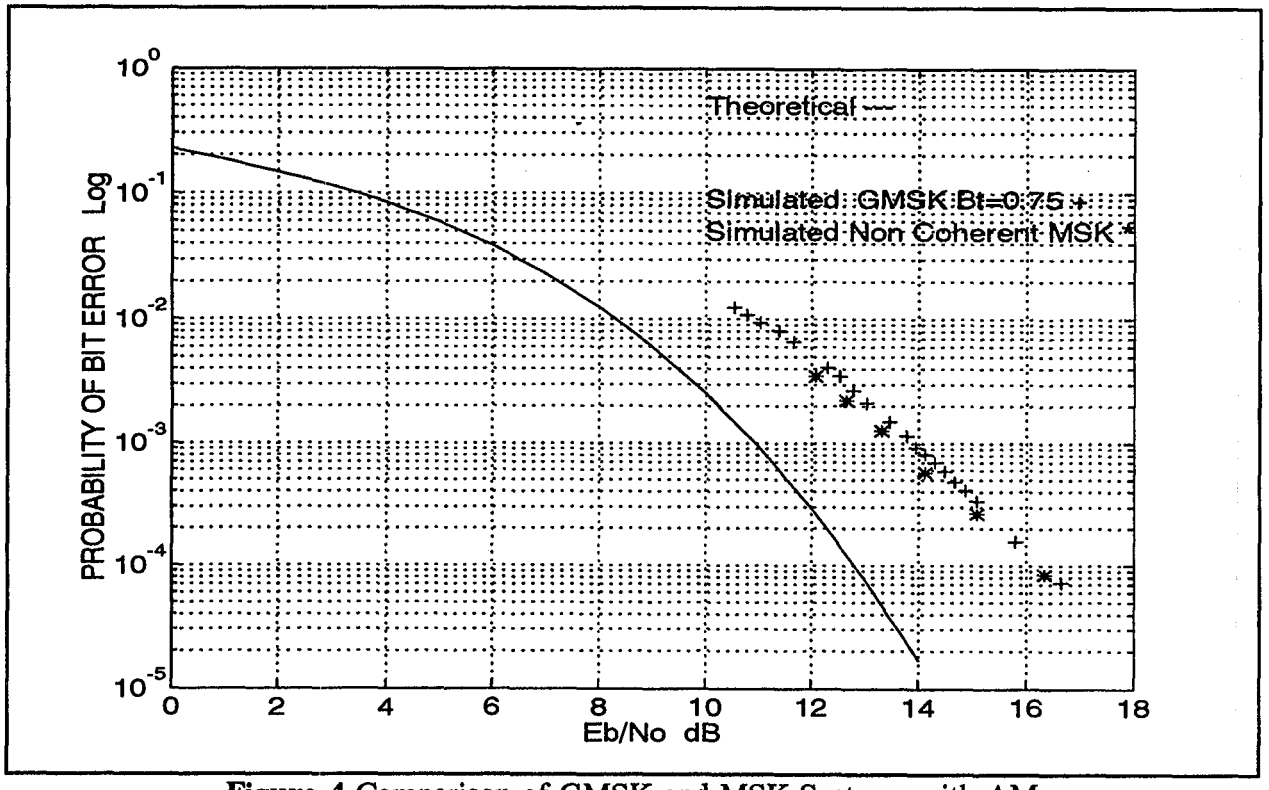


Figure 4 Comparison of GMSK and MSK Systems with AM

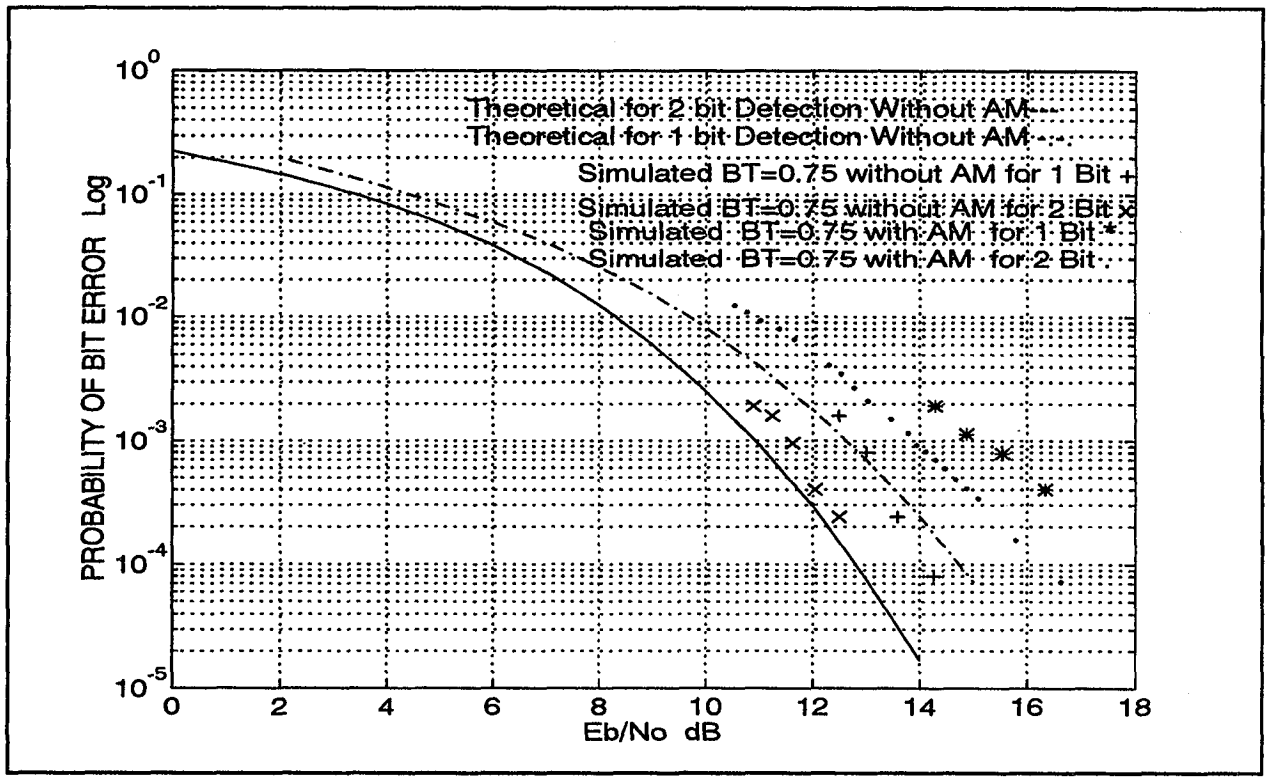


Figure 5 Comparison of 2-bit & 1-bit Detection Schemes for the Hybrid System

GPS INTEGRITY MONITORING AND MULTIPATH ERROR DISTRIBUTIONS

Trent A. Skidmore
Frank van Graas
Avionics Engineering Center
Ohio University

BIOGRAPHY

As a Research Engineer and newly-appointed Adjunct Assistant Professor at the Ohio University Avionics Engineering Center, Dr. Trent A. Skidmore has been involved in a variety of GPS-related projects. These have included the development of a coverage model for studying availability, integrity, and reliability issues associated with satellite-based navigation systems, a GPS radio-frequency interference model, and spectrum issues associated with the DGPS ground-to-aircraft datalink. He also participates in the FAA/NASA Joint University Program for Air Transportation Research and is the Co-Principal Investigator for the Aeronautical Hazard Alerts Program sponsored by the FAA. He is a graduate of Ohio University (Ph.D., 1991) and Michigan Technological University (M.S., 1988 and B.S. (with honors), 1986), all in electrical engineering.

Frank van Graas, Ph.D., is an Associate Professor of Electrical and Computer Engineering, and a Research Engineer with the Avionics Engineering Center of Ohio University. Since 1985 he has been involved with integrated navigation systems, Differential GPS, and Flight Reference/Autoland Systems.

ABSTRACT

The ability to perform integrity monitoring of a differential Global Positioning System (DGPS) precision approach system is dependent upon the form of monitor used and the various noise sources acting on the reference station, the monitor, and the approaching aircraft. This paper focuses on how multipath and assumptions about its probability density function (pdf) affect the availability of the integrity monitoring function. Also discussed is a technique specified as Code/Carrier Integrity Monitoring which can be used to greatly improve integrity monitoring availability over methods employing conventional C/A-Code operations.

1. INTRODUCTION AND BACKGROUND

The components of a typical differential Global Positioning System (DGPS) instrument approach system are illustrated in Figure 1. The reference station supplies the approaching aircraft with differential corrections in order to take advantage of its known location to remove errors common to both itself and the aircraft. These errors include ionospheric and tropospheric delays and those errors caused by Selective Availability. While the datalink is shown as a VHF channel, other channels, such as Mode-S, have been used experimentally.

The integrity monitor receives the differential corrections from the reference station and processes them, along with local measurements, to compute a location. If this location is outside a given alarm threshold the monitor alerts the reference station that its corrections are no longer valid. In general, integrity can be defined as the quality which relates to the trust that can be placed in the correctness of the information supplied by the total system. In this scenario, integrity is the ability of the system to provide a timely warning when the position error exceeds a specified threshold. The hangar depicted in Figure 1 is used to indicate that multipath is the dominant error source in any given DGPS operation. While it has generally been assumed, perhaps for mathematical convenience, that multipath errors exhibit Gaussian behavior, this premise has not been justified.

Previous work by Creamer et. al. [1] established a baseline integrity monitoring methodology for a Global Navigation Satellite System (GNSS) precision approach application consistent with the Instrument Landing System (ILS) requirements. In order to compute the availability of the integrity monitoring function several key assumptions were used. These are:

Presented at the Institute of Navigation 49th Annual Meeting, June 21-23, 1993

- Carrier smoothed C/A-Code techniques are used
- The dominant error source is multipath
- Multipath is Gaussianly distributed
- Vertical Error = VDOP x UERE
- Multipath correlation time is 200 seconds

Based on these assumptions and others specified in Reference [1], it was found that satisfactory levels of integrity could only be achieved with an average vertical dilution of precision (VDOP) of less than 1.5, yielding an integrity monitoring availability of approximately 0.6, equivalent to only 14.4 hours per day. This large unavailability is in part directly linked to the assumptions concerning multipath error distributions and the relationship between vertical position error and VDOP. The next section highlights the current research relevant to these assumptions.

2. MULTIPATH ERROR DISTRIBUTIONS

In order to study multipath error distributions it was necessary to devise an experiment that would be representative of a typical DGPS approach installation. For the test, data were collected simultaneously at two different sites at the Ohio University Airport. The sites were located approximately two miles apart to minimize the probability of common multipath. For each site, the antennas were placed in a cleared field on the ground. This was done to minimize multipath. While it could be argued that ground placement of antennas is impractical, previous results by Waid [2] have shown that careful antenna selection, even in multipath-rich environments, can yield results consistent with ground-placed antennas. At each site, two receivers (one employing a dual frequency architecture, the other a narrow-correlator design) were connected to a common antenna. Two twelve-hour static data collections were performed over a period of two days. The data were collected at a rate of one measurement per second.

The residual errors were obtained using the "code - integrated Doppler" technique as outlined in [2] and by Braasch [3]. The residual errors contain multipath, thermal noise, and receiver errors, of which it has been shown that multipath is the dominant component. An example plot of residual errors is shown in Figure 2. The plot contains one hour of data (3,600 points) from two consecutive days. The traces have been offset by +2 meters for Day 1 and by -2 meters for Day 2 for illustrative purposes. Note that while there is some degree of correlation due to multipath, the signals are predominantly noise-like. This is indicative of the natural multipath rejection mechanism provided by the ground-placed antenna.

In order to illustrate the distribution of the residual errors, and specifically the multipath errors, and how they compare with the simple Gaussian assumption, consider the following table. For this, two hours of Satellite 14 data from two consecutive days were used. The root-mean-square (RMS) of the error for the C/A-Code, P-Code, and Narrow Correlator (NC) C/A-Code receivers are shown. These RMS values are subsequently used for distribution comparisons.

Table of RMS Errors

Process	Day 1	Day 2
C/A-Code	0.76	0.76
P-Code	0.07	0.12
NC C/A-Code	0.15	0.13
	RMS Error (m)	

For the comparison, the cumulative distribution function (cdf) is used. The cdf is defined as the integral of the probability density function and is more suitable for comparing random variables than the pdf. This becomes apparent when the cumulative distribution is plotted using probabilistic scaling. Using probabilistic scaling, the cumulative distribution function of a Gaussian distribution appears as a straight line. Figures 3a - 3f compare the cumulative distribution functions of the residual errors to their corresponding Gaussian cdfs when the standard deviation of the Gaussian distributions are set equal to the RMS of the residual errors.

Notice in Figures 3a and 3b that the residual errors of the C/A-Code receiver appear Gaussian to within approximately $\pm 2\sigma$. Beyond $\pm 2\sigma$ the residuals' cdf appears truncated. Note also the similarity between Day 1 and Day 2. The cumulative distribution functions for the P-Code and the NC C/A-Code receivers are shown in Figures 3c - 3f. It is interesting to note that the errors here are much smaller than for the conventional C/A-Code receiver and that the mean of the residuals is no longer zero. This indicates that for these architectures, the residual errors are non-symmetric and/or non-stationary. Thus, while the residual errors for the conventional C/A-Code receiver can be approximated as a $\pm 2\sigma$ truncated Gaussian distribution, the P-Code and the Narrow Correlator C/A-Code do not appear strictly Gaussian. While this example is not intended to be the definitive work on residual error distributions, it does illustrate that the Gaussian assumption is not completely valid and will tend to overestimate the detrimental affect of multipath at the DGPS reference and monitor stations.

3. VERTICAL POSITION ERROR AND VDOP

The usual approximation for relating the RMS vertical position error to the vertical dilution of precision (VDOP) is given as:

$$\text{Vertical Position Error} = \text{VDOP} \times \text{UERE} \qquad (1)$$

where UERE is the RMS user equivalent range error from a given satellite.

A major factor affecting the validity of the above assumption is the elevation angle of the satellites used. It is recognized that elevation-dependent errors, such as multipath and tropospheric delay, decrease with increasing satellite elevation angle. This phenomena is illustrated in Figure 4 for tropospheric delays. In order to determine the extent to which Equation 1 overestimates the vertical position error, a simulation was conducted. For this, the average VDOP at three medium-latitude locations was found to be 1.6 (assuming the GPS 21 Primary Constellation and a mask angle of 7.5°). It was found that the RMS tropospheric delay is approximately 19.4 meters. Thus, if Equation 1 is used to compute vertical position error the result would be 31.0 meters. The true RMS vertical position error was computed to be 11.3 meters. This indicates that Equation 1 overestimates the effect of VDOP on vertical position error by a factor of approximately 2.6. While Equation 1 may indeed be useful for some applications, it is an inappropriate assumption for use in computing approach system integrity levels.

3. CODE/CARRIER INTEGRITY MONITORING

The Code/Carrier Integrity Monitoring (CCIM) technique uses the GPS carrier signal for the primary detection of signal failures that would cause the vertical containment warning level (CWL) to be exceeded (CWL = 8 meters for CAT I and 9.76 meters for the SCAT-I [4]). Figure 5 illustrates the basic concept of CCIM. The GPS carrier phase is integrated over time to obtain an accurate measure of the change in the range to the satellite. To find an accurate absolute range to the satellite, the offset between the true range to the satellite and the integrated carrier phase is estimated using the code phase measurements. Next, this offset is added to the integrated carrier phase to obtain the estimated true range. For example, the estimated true range (ETR) at time t could be calculated using the previous n measurements of code phase and integrated carrier phase:

$$ETR(t) = \phi(t) + \frac{\sum_{i=0}^{n-1} (\rho(t - iT) - \phi(t - iT))}{n} \qquad (2)$$

where: ρ is the measured code phase in meters
ϕ is the integrated carrier phase in meters
T is the time between measurements in seconds

The differential GPS correction is now the difference between the estimated true range and the calculated true range based on the satellite position and the a priori reference or monitor station coordinates. Note that the estimated true range is corrected for the known satellite clock offset from GPS time, but it still contains the effects of tropospheric and ionospheric propagation delays and Selective Availability.

The monitor station compares the difference between the differential correction as transmitted by the reference station and the differential correction as determined by the monitor station. If the absolute value of this difference is larger than the preset threshold, an alarm is raised.

Next, it is conservatively assumed that the code phase measurements are statistically independent and have a normal (Gaussian) probability distribution. Integrated carrier phase measurements error is assumed to be negligible (well below 1 cm RMS if integrated over 1 second). Note that if the direct signal is stronger than the reflected/diffracted signal, then the carrier phase multipath error is always less than or equal to one quarter wavelength (4.8 cm for the GPS L1 frequency). Based on these assumptions, the estimated true range will have errors that are also normally distributed with zero mean and a variance of σ^2/n, where n is the number of samples used to estimate the true range. Furthermore, with probability of missed detection (P_{MD}) the estimated true range has an offset error of less than $\pm x$ meters, where x is given by

$$x = \frac{\sigma\sqrt{2}}{\sqrt{n}} \text{erfc}^{-1}(P_{MD}) \qquad (3)$$

where: $\text{erfc} = \frac{2}{\sqrt{\pi}} \int_z^\infty e^{-\lambda^2} d\lambda$

Also, the probability of a false alarm (P_{FA}) is equal to the probability of a missed detection if the thresholds for detection are set at -x and x.

The position error is related to the range errors through the following relation:

$$\begin{bmatrix} \Delta E \\ \Delta N \\ \Delta u \\ \Delta t \end{bmatrix} = G \begin{bmatrix} \Delta R_1 \\ \Delta R_2 \\ \Delta R_3 \\ \vdots \\ \Delta R_m \end{bmatrix}, \quad G = (H^T H)^{-1} H^T \qquad (4)$$

where H is a data matrix. Each row of H consists of the partial derivatives of the range with respect to East position (E), North position (N), vertical position (u), and clock Offset (t), respectively.

The maximum possible vertical position error based on the offset error, x, is calculated from

$$\Delta u = x \sum_{k=1}^{n} |g_{3k}| \qquad (5)$$

where g_{3k} is the kth element of the third row of G.

If the maximum vertical position error is greater than the CWL, integrity monitoring cannot be achieved with the required probabilities of false alarm and missed detection.

In summary, the CCIM technique proceeds as follows:

1) Estimate the true range using the code phase and the integrated carrier phase (see Equation 2);

2) Form the differential correction using the estimated true range and the calculated true range based on the satellite position and the known reference location;

3) Similarly form the differential correction for the monitor station. Verify that the difference between the corrections from the reference and the monitor station are within the detection thresholds given by Equation 3. Calculate the maximum possible vertical position error.

4) Raise the alarm if the threshold has been exceeded, or if the CWL has been exceeded.

To illustrate the effectiveness of CCIM, a numerical example follows. Assume that the code phase measurement noise at the reference station and the monitor station have a normal probability distribution, with zero mean and a standard deviation of 1.2 meters. This number is a conservative estimate of the residual C/A-Code error (see Figures 3a- 3b). Also, the errors between the reference and the monitor station are assumed to be uncorrelated (proper siting of the monitor and the reference station should assure this). The noise level on the detection statistic is found by root-sum-squaring the reference station noise and the monitor station noise: $\sqrt{1.2^2 + 1.2^2} = 1.7$ meters. Next, assume that the correlation time of the samples is one second, such that the probability of a missed detection should be on the order of 10^{-9} on a per sample basis. (This value is more stringent than the ILS required probability of missed detection of 10^{-7} per approach and also satisfies the required alarm rate of less than 10^{-3} per hour). If a total of 300 samples (5 minutes) are used for the averaging process, then using Equation 3 with probability $(1 - 10^{-9})$, the true monitor statistic is within ±0.6 meters of the estimated statistic. Next, Equation 5 provides the maximum possible vertical position error. The effect of satellite geometry was simulated using three representative locations in the U.S., which resulted in the following relation:

$$\Delta u < 0.6 \times 2.4 \times VDOP_{AVG} \text{ (meters)} \qquad (6)$$

where $VDOP_{AVG}$ is the average value for the vertical dilution of precision. Equation 6 can be rewritten to obtain the final result based on a maximum CWL of 8 meters:

$$VDOP_{AVG} < 5.6 \qquad (7)$$

This is an important result. Although the derivation for the VDOP requirement using the CCIM is very conservative, the availability of CCIM is in fact higher than the availability of positioning for CAT I operations. Figure 6 is used to find the predicted availability of CCIM, which is greater than 99.8% [see Reference 5].

3.1 ADDITIONAL CCIM CONSIDERATIONS

The availability of CCIM can be further increased by:

- Longer averaging: the 300-second interval should not be much longer in order to avoid the effect of code/carrier divergence due to the ionosphere. However, if a dual-frequency receiver is used, this divergence effect can be measured and corrected for;

- Considering less than the worst case: it is unlikely that the maximum possible error acts on all measurements simultaneously;

- Use different errors for each of the measurements, and calculate a weighted solution. Also consider the propagation of these errors into the vertical position error.

Before the carrier phase data is used, it is imperative to detect and repair any cycle slips that may have occurred. For a static user, the detection of cycle slips is relatively straightforward. For instance, the change in the integrated carrier phase for each satellite between two successive observations could be examined. This change should be very smooth as a function of time. A rapid change indicates a cycle slip, while the change is proportional to the number of cycles that were slipped or advanced by the receiver tracking loop.

Dynamic users have three techniques to choose from for cycle-slip detection and correction;

1) Compare the GPS-derived velocities with inertially-derived velocities. A cycle slip over one second would cause a jump in the velocity of 0.19 m/s; or

2) Use redundant measurements and evaluate the residuals. Note that this technique is not always useable for cycle slip repair on multiple measurements; or

3) Compare the change in the L1 integrated carrier phase with the L2 integrated carrier phase (after conversion to a common unit such as meters). Typically, the noise on this difference is less than 1 cm. If an L1 cycle slip occurs, the difference would jump by 19 cm, while an L2 cycle slip would cause a 24-cm discontinuity. A simultaneous cycle slip on both L1 and L2 would cause a 5-cm discontinuity. For all possible combinations of up to 30 simultaneous L1 and L2 cycle slips, it follows that the most difficult to detect is the combination of 9 cycle slips on L1 and -7 cycle slips on L2. The residual for this case is 0.43 cm. Although this number is initially below the noise level, it is still large enough for timely detection. This method not only detects the cycle slip, but it also identifies the bad measurement, and in most cases is able to repair the cycle slip.

4. SUMMARY AND FUTURE WORK

This paper has summarized a baseline integrity monitoring technique and considered the validity of the assumptions concerning multipath error distributions and the propagation of VDOP into vertical position error. An alternative technique, designated Code/Carrier Integrity Monitoring, has been presented which can greatly improve the availability of DGPS approach system integrity. Several key areas require additional investigation. These are summarized below and will be the focus of ongoing research:

- Siting guidelines and criteria should be developed for DGPS reference and monitor stations.

- The analysis present in this paper is based on using all-in-view receivers at the user, the monitor station, and the reference station. The monitoring technique should be expanded to include different combinations of satellites.

- Implementation, long-term testing, and additional data collection are recommended for the Code/Carrier Integrity Monitoring.

ACKNOWLEDGEMENTS

This work was supported by the Federal Aviation Administration (ASE-300) under Research Grant Number 92-G-023 and the FAA/NASA Joint University Program for Air Transportation Research (Grant NAG1-1423). The dual-frequency GPS receivers were provided by Ashtech, Inc. and the narrow correlator receivers were provided Novatel, Inc.

REFERENCES

[1] Creamer, P.M., E.M. Geyer, J.J. Pisano, and D.P. Frank, "GNSS Integrity for Aircraft Precision Approach," Institute of Navigation: Proceedings of the 48th Annual Meeting, June 29 - July 1, 1992.

[2] Waid, J.D. "Ground Station Siting Considerations for DGPS," Institute of Navigation: Proceedings of the 1993 National Technical Meeting, January 20-22, 1993.

[3] Braasch, M.S., "On the Characterization of Multipath Errors in Satellite-Based Precision Approach and Landing Systems," Ph.D. Dissertation, Ohio University, Athens, Ohio, June, 1992.

[4] Minimum Aviation System Performance Standards, DGNSS Instrument Approach System: Special Category I (SCAT-I), RTCA Paper No. 200-93/SC-159-428, Third Draft, May 3, 1993.

[5] Skidmore, T.A., "The Satellite Coverage Research Analysis Model (SCRAM)", FAA/NASA Joint University Program for Air Transportation Research, Winter Quarterly Review, NASA Langley Research Center, January 28-29, 1993.

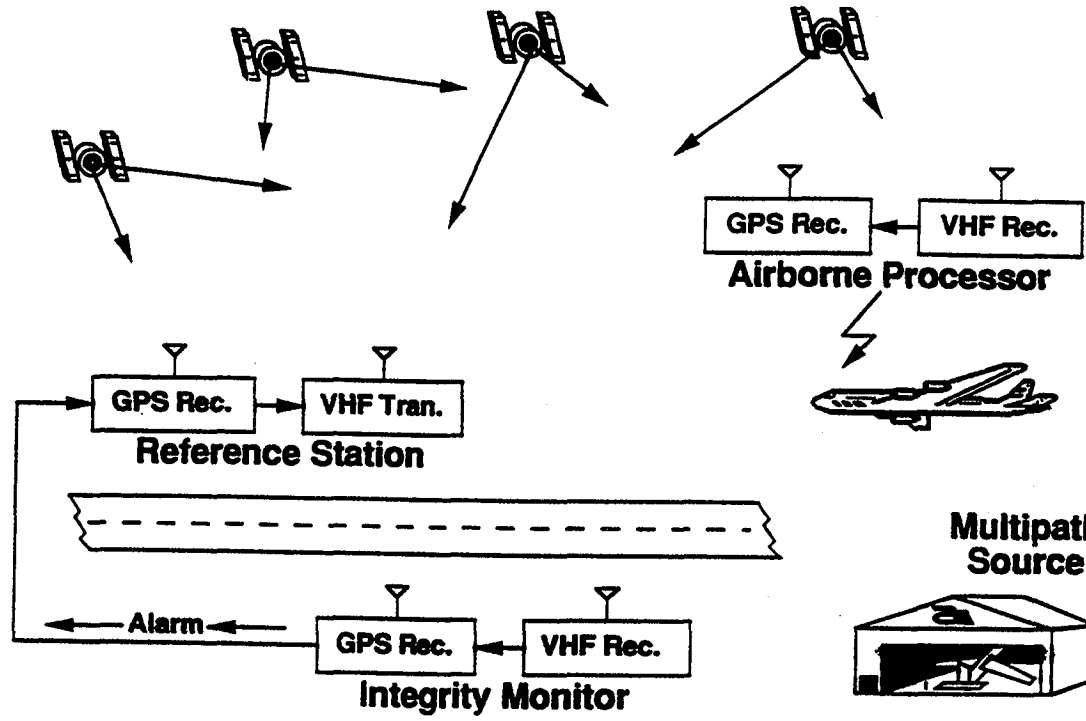


Figure 1. DGPS Instrument Approach System Integrity Monitoring

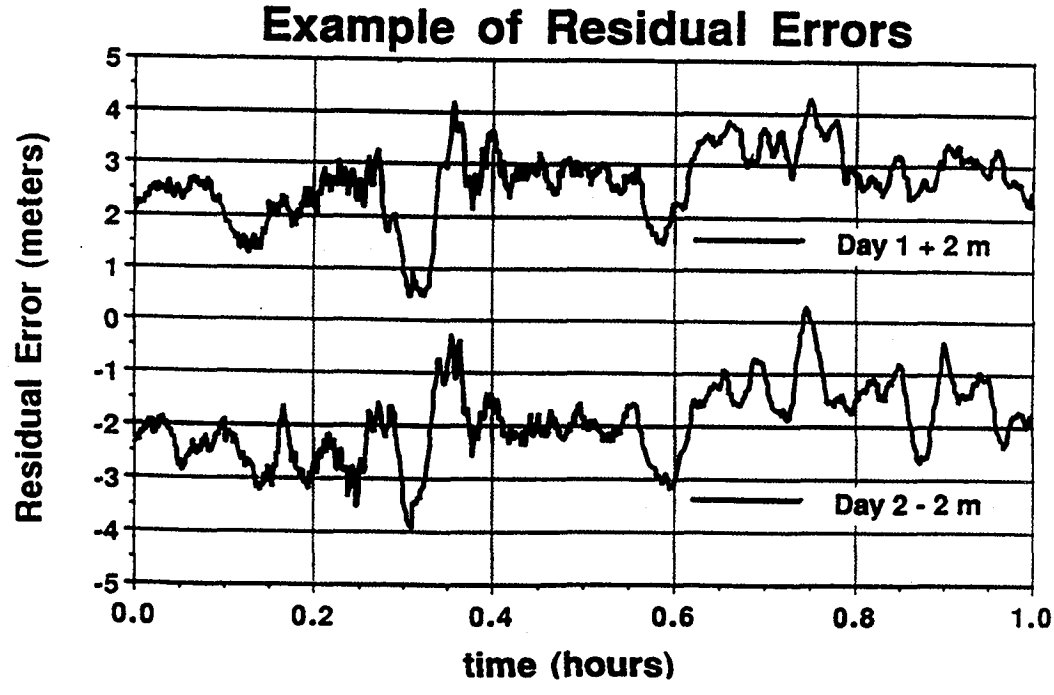
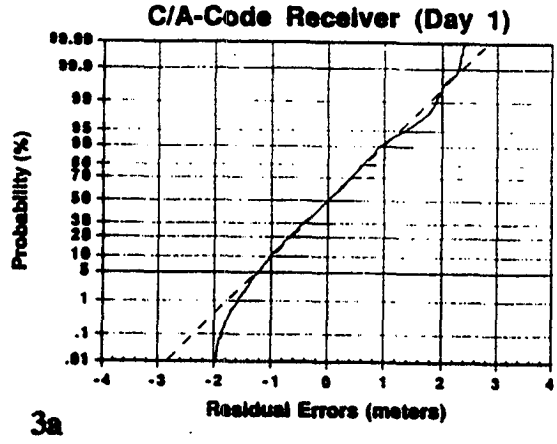
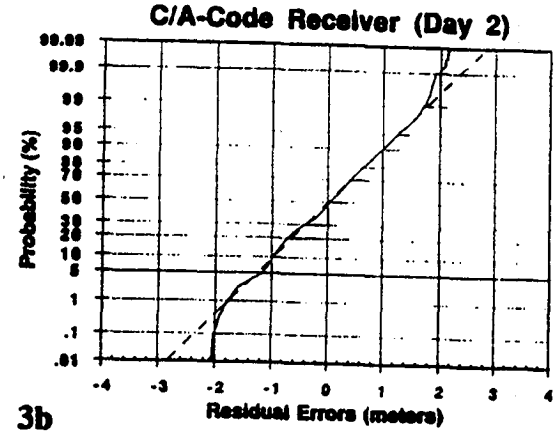


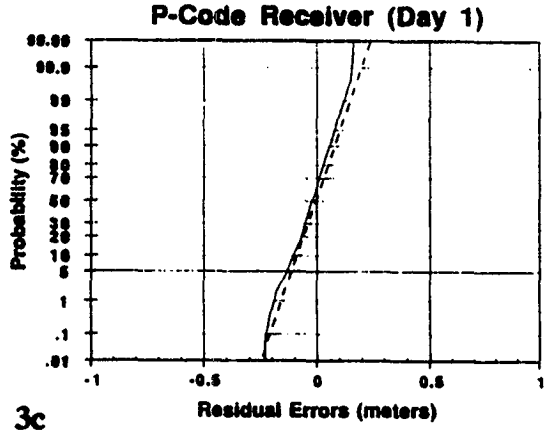
Figure 2. Sample Residual Errors using the "Code - Integrated Doppler" Technique



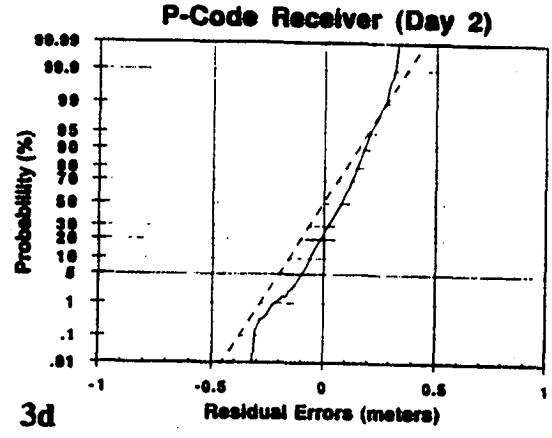
3a



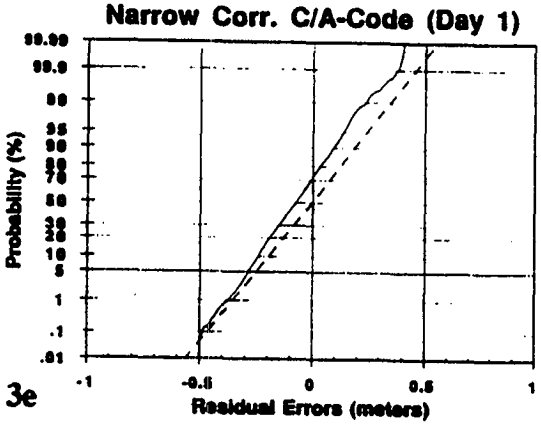
3b



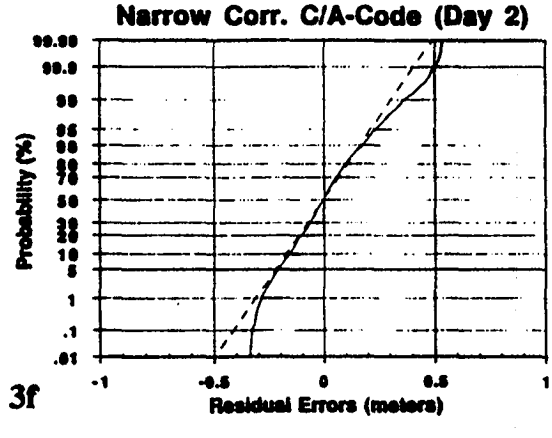
3c



3d



3e



3f

Figure 3. Residual-Error Cumulative Distribution Functions

DGPS GROUND STATION INTEGRITY MONITORING

Trent A. Skidmore
Frank van Graas
Avionics Engineering Center
Ohio University

BIOGRAPHY

Trent A. Skidmore, Ph.D. is a Research Scientist and Adjunct Assistant Professor at the Ohio University Avionics Engineering Center were he is involved in a variety of GPS-related projects, including the development of a coverage model for studying the availability and integrity of satellite-based navigation systems and spectrum engineering issues associated with the DGPS ground-to-aircraft data link. He is a graduate of Ohio University (Ph.D., 1991) and Michigan Technological University (M.S., 1988 and B.S. (with honors), 1986), all in electrical engineering.

Frank van Graas, Ph.D., is an Associate Professor of Electrical and Computer Engineering, and Principal Investigator with the Avionics Engineering Center of Ohio University. Since 1985 he has been involved with integrated navigation systems, Differential GPS, and Flight Reference/Autoland Systems.

ABSTRACT

This paper summarizes the development of a unique Differential Global Positioning System (DGPS) ground station integrity monitor which can offer improved availability over conventional code-differential monitoring systems. This monitoring technique, called Code/Carrier Integrity Monitoring (CCIM), uses the highly stable integrated Doppler measurement to smooth the relatively noisy code-phase measurements. The pseudorange correction is therefore comprised of the integrated Doppler measurement plus the CCIM offset. The design and operational results of a DGPS ground station integrity monitor are reported. A robust integrity monitor is realized which is optimized for applications such as the Special Category I (SCAT-I) defined in the RTCA Minimum Aviation System Performance Standards.

1. INTRODUCTION AND BACKGROUND

With the release of RTCA DO-217 (Minimum Aviation System Performance Standards (MASPS) DGNSS Instrument Approach System: Special Category I (SCAT-I) [1], and the subsequent non-Federal systems that are under development throughout the industry, there remains an abundance of both technical and political issues which must be addressed. This paper focuses on one such technical issue: Ground Reference Station Integrity Monitoring. Discussed are the parameters relevant to integrity monitoring and basic monitoring strategies, with emphasis on the SCAT-I system. With the exception of the paper concerning the Dallas/Fort Worth Trials [2], an overview of the current literature shows a lack of relevant information concerning ground station integrity monitoring for SCAT-I systems. However, previous work by the authors [3] demonstrated conceptually that a high availability integrity monitor is realizable using a technique designated as Code/Carrier Integrity Monitoring (CCIM). A detailed explanation and sample test results of CCIM will be given in Section 4.

In order to conceptualize a typical Differential Global Navigation Satellite System (DGNSS), of which the NAVSTAR Global Positioning System (GPS) is the key space-segment component, consider the precision-approach system depicted in Figure 1. The reference station supplies the approaching aircraft, via data link, with differential corrections. The aircraft uses these corrections to remove errors common to both itself and the reference station. These errors include ionospheric and tropospheric delays, satellite clock and ephemeris errors, and those errors caused by Selective Availability. While the data link is shown as a VHF channel, a Mode-S data link is also defined in the MASPS.

The integrity monitor receives the differential corrections from the reference station and processes

Presented at the Institute of Navigation 1994 National Technical Meeting, January 24-26, 1994

them, along with local measurements, in order to determine the validity of the uplinked corrections. If the monitor determines that some portion of the differential corrections are out of tolerance (beyond a given alarm threshold), it must alert the reference station that its corrections are not valid, thus forcing the reference station to discontinue operation. In general, integrity can be defined as the quality which relates to the trust that can be placed in the correctness of the information supplied by the total system [4]. In this scenario, integrity is the ability of the system to provide a timely warning when the position error exceeds a specified threshold. The hangar depicted in Figure 1 is used to indicate that multipath is the dominant error source in any given DGPS operation.

2. GROUND STATION INTEGRITY

The SCAT-I MASPS allocates the integrity requirements to the ground, air, and data link segments (see MASPS Appendix D, Figure D-1). Specifically, the SCAT-I system must be designed such that the probability of undetected violation of the outer tunnel containment surface, constructed around the flight path defined by the waypoints known to the airborne subsystem, is no greater than 10^{-7} per approach. This results in a probability of an undetected failure in the ground segment of 3×10^{-8} per approach. Although this allows for the design of the ground station integrity function, it should be recognized that some of the data transmitted by the ground station is used by the airborne segment to achieve the airborne integrity of 6×10^{-8} per approach. It should also be pointed out that waypoint integrity is not addressed in the MASPS.

To achieve ground station integrity, the SCAT-I MASPS calls for (at least) the following key monitor features to be implemented (Note: The following statements are taken almost directly from the MASPS and are included to help consolidate the relevant information).

1) Independent determination and verification of all data to be transmitted for compliance with the SCAT-I system integrity requirements.

2) Independent detection and monitoring of all DGNSS Instrument Approach System (DIAS) data transmissions to assure correct transmission to aircraft in the DIAS precision approach phase of flight.

3) Generate the appropriate actions to ensure aircraft are notified upon SCAT-I system integrity failure.

Specifically, the following must be implemented:

- The ground-based DGNSS Signal integrity Monitoring Function will be required to assess the quality of the DGNSS signal prior to and during its broadcast and provide an appropriate alert should that signal not meet its performance specifications.

- The ground-based integrity monitoring function shall be used to validate the accuracy of the data generated by the ground reference station and transmitted to airborne users. It shall be possible to demonstrate that the integrity monitoring function is functionally separate from the ground reference station function. The ground reference station and ground integrity monitor shall employ independent antennae, or otherwise address multipath errors. The monitoring function shall validate data prepared for transmission. This validation shall be performed prior to RF transmission. The ground integrity monitoring function shall validate continued acceptable performance of the SCAT-I data link.

- The system latency of an integrity warning shall not exceed 6 seconds.

- The total latency between the time of occurrence of a failure in the ground equipment and the time the last bit of the integrity alarm is transmitted by the data link shall be less than 3 seconds.

- SCAT-I systems shall satisfy the ground station pseudorange allocation of 1.1 meters (defined as the RMS value of the pseudorange correction errors).

The monitor station must verify and validate all parameter values generated by the reference station. These parameters are summarized in Table 1 for the required Type-1 Message (see MASPS Appendix A). In this paper we are mainly concerned with the pseudorange correction (PRC), the range-rate correction (RRC), and the user differential range error (UDRE). Two levels of accuracy must be monitored: the 95% level (inner tunnel) and the 10^{-7} level (outer tunnel). It is recognized that the 95% level of accuracy can be based on an average, while the 10^{-7} level must be set for each approach. Since the ground subsystem does not have specific knowledge of how the ground data is used by the airborne equipment, it is important that all parameter values are accurate. For example, the ground subsystem does not know which satellites are used in the airborne position solution, or how the satellites are weighted in the solution.

Table 1. Type-1 Message for SCAT-I {from [1]}

Parameter	Range
Modified Z-count	20 minutes
Acceleration Error Bound	0.015 m/s2
Satellite ID	1 - 64
Pseudorange Correction	+/- 655.34 m
Issue of Data	-
Range-Rate Correction	+/- 4.094 m/s
UDRE	12.4 m

(Note: See MASPS [1] for further details)

In addition to achieving accuracy and integrity, the ground station must also meet continuity and availability requirements. The performance of the SCAT-I system in these areas must also be optimized. Probability of loss-of-continuity must be less than 10^{-4} per approach. This requirement places a bound on the allowable alarm rate of the integrity function. Availability should be greater than 95%, which means that the availability of the integrity function should be at least as good as the availability of the accuracy performance.

3. PSEUDORANGE CORRECTIONS

The GPS pseudorange (PR) observation equation is given by:

$$PR = |\underline{s} - \underline{r}| + \Delta t_r + \varepsilon_{tropo} + \varepsilon_{iono} + \varepsilon_{mp} + \varepsilon_{rcvr} + \varepsilon_{tn} + \varepsilon_{\Delta t_{sv}} + \varepsilon_{SA} \quad (1)$$

where:

$\underline{s}$:true satellite coordinate vector
$\underline{r}$:true (surveyed) reference station antenna coordinate vector
Δt_r :reference receiver clock offset from GPS time
ε_{tropo} :tropospheric propagation delay
ε_{iono} :ionospheric propagation delay
ε_{mp} :multipath errors
ε_{rcvr} :receiver measurement errors
ε_{tn} :thermal noise
$\varepsilon_{\Delta t_{sv}}$:residual satellite clock offset from GPS time after applying the broadcast clock correction
ε_{SA} :Selective Availability errors

To obtain the pseudorange correction (PRC), the "true" range between the reference station antenna and the satellite is calculated using the satellite ephemeris data. Next, both the measured pseudorange (PR) and the estimated reference receiver clock offset from GPS time are subtracted from the "true" range:

$$PRC = |\underline{s}_c - \underline{r}| - PR - \hat{\Delta t_r} \quad (2)$$

where:

$\underline{s}_c$:calculated satellite coordinate vector
$\hat{\Delta t_r}$:estimated reference receiver clock offset from GPS time

which can also be written as:

$$PRC = -(\varepsilon_{\Delta t_r} + \varepsilon_r + \varepsilon_{tropo} + \varepsilon_{iono} + \varepsilon_{mp} + \varepsilon_{rcvr} + \varepsilon_{tn} + \varepsilon_{\Delta t_{sv}} + \varepsilon_{SA}) \quad (3)$$

where:

$\varepsilon_{\Delta t_r}$:residual reference receiver clock offset after subtracting the estimated ref. receiver clock offset

ε_r :residual satellite coordinate error ($\underline{s}_c - \underline{s}$)

Omitted from equation 3 are the spatial and temporal decorrelation errors, which will be addressed in the next section. Ideally, the PRC only contains those error sources that are in common between the reference station and the airborne user. These are: the tropospheric delay error, the ionospheric delay error, the difference between the true and calculated satellite clock offset, the line-of-sight component of the difference between the true and calculated satellite coordinates, and Selective Availability errors. The remaining components in equation 3 represent the error in the PRC. Any component that is common to the PRCs for all satellites, such as the error in the reference receiver clock offset estimation, propagates into an airborne clock error, but it does not affect the airborne position solution. Note, however, that this type of error would greatly affect a differential user who uses less than four satellites. Therefore, for this paper, it is assumed that the airborne user navigates with corrections for at least four satellites at all times.

Thus, the errors in the PRC are thermal measurement noise, reference station multipath, and receiver measurement errors, each of which are discussed below:

Thermal measurement noise: the thermal noise power in combination with the received GPS signal power determines the amount of noise on the code and carrier phase measurements [5]. Table 2 provides the RMS code and carrier tracking noise of a second-order tracking loop for the minimum GPS signal strength of -160 dBW (dB with respect to 1 Watt).

Table 2. RMS Errors Due To Thermal Noise {from [5]}

Code Tracking Loop (2nd-order)		Carrier Tracking Loop (2nd-order)	
4-Hz Bandwidth	0.5-Hz Bandwidth	0.5-Hz Bandwidth	16-Hz Bandwidth
4.2 meters	1.5 meters	0.2 mm	1.2 mm

From Table 2 it can be seen that the code tracking loop noise can be reduced by reducing the tracking loop bandwidth. For a stationary user it is generally not a problem to reduce the code tracking loop bandwidth to below one Hertz. It should be noted however that the effective data rate is also reduced. If one independent update per second is required, then the tracking loop bandwidth should not be reduced to much lower than 0.5 Hz.

The carrier tracking loop noise is small compared to the code noise. If the code tracking loop is carrier-aided, then the code tracking loop bandwidth could be set smaller than 0.5 Hz, as long as the independent updates are based on the carrier phase measurements. In that case, care must be taken to ensure that no cycle slips occur which would bias the code phase measurements. Also, the time period of the aiding process should not be longer than 1 to 2 minutes to avoid code-carrier divergence due to the ionosphere. The latter problem could be alleviated by making dual frequency measurements or through accurate modeling of the ionosphere. Most carrier-aided code tracking loops provide rms tracking noise on the order of 0.3 meters or better [6].

Multipath: Multipath error consists of two components: diffuse and specular. Diffuse multipath contains high frequencies (several Hertz) and originates from relatively small reflective or diffractive objects. Specular multipath has a low frequency with periods of up to several minutes. Specular multipath originates from larger objects, such as the ground or a large building. The effect of multipath on the code phase measurements is much larger than on the carrier phase measurement. Carrier phase multipath error is always less than a quarter wavelength (4.8 cm) as long as the direct signal is stronger that the reflected or diffracted signal. Although code phase multipath can be as large as 150 meters for a standard C/A-code receiver, typical measured values are on the order of 1-3 meters [6]. Code phase multipath is reduced by averaging against the integrated carrier phase. Care must be taken however that the period of integration is longer than the period of the code phase multipath error. Otherwise, a bias error will be introduced into the differential corrections. Ground station multipath is reduced by proper antenna design and ground station antenna siting.

Receiver measurement error: Measurement errors for current digital GPS receivers are small compared to errors introduced by thermal noise. The main error sources are receiver measurement resolution, dynamic tracking loop errors, and interchannel biases. Dynamic tracking loop errors are generally small for a stationary user. Digital receivers usually have very good measurement resolution (better than 0.1 m); however, older analog receiver architectures could have measurement resolutions on the order of one or two meters. Interchannel bias errors can be significant (greater than 0.1 m) for receivers that employ separate hardware channels for each satellite being tracked. Receiver measurement errors are easily identifiable by using a GPS signal generator or by performing a zero-baseline test against a good receiver (this is a test where two receivers are connected to the same antenna).

Spatial and Temporal Decorrelation Errors: The accuracy of the PRC will degrade as a function of both time (temporal decorrelation) and distance (spatial decorrelation). Temporal decorrelation must always be considered because it is independent of the location of the user relative to the reference station. This also means that temporal decorrelation errors can be fully observed by a monitor station. In contrast, spatial decorrelation errors cannot be fully observed by one monitor station, especially if the monitor station antenna is close to the reference station antenna. Spatial decorrelation is not an issue if the reference station is close to the airport runway (within a few kilometers). However, larger separation distances (30 to 100 miles) between the reference station and the user could potentially degrade the accuracy of the PRC beyond acceptable levels. Error sources that are susceptible to spatial and temporal decorrelation are ionospheric and tropospheric delays, satellite clock and ephemeris errors, and Selective Availability.

Temporal decorrelation errors are dominated by Selective Availability. Although these errors can be

measured by the monitor station, it is necessary that the rate of change of SA is slow enough such that the PRC and the RRC can correct for this error. If the nominal age of the differential correction is 3 seconds, then the maximum allowable unmodeled SA acceleration term should be less than 0.24 m/s^2 in order to maintain the 1.1 m PRC accuracy. Note however, that this same acceleration would cause a ranging error of 27 meters if the age of the correction is 15 seconds. If this occurs, then the Acceleration Error Bound in the correction message would be set to indicate unacceptable acceleration errors.

Spatial decorrelation errors can be separated into propagation delays and satellite orbit errors. The main problem is that these errors cannot be observed by a nearby monitor station. Although ionospheric decorrelation errors of up to 0.5 m over a 9 km distance have been reported [7], these errors are generally small. Similarly, tropospheric delays usually do not decorrelate very rapidly. Both ionospheric and tropospheric delays require additional study to verify the validity of differential corrections over longer distances. The main concern would be the 95% accuracy performance of the differentially corrected position solution.

Satellite orbit errors on the other hand can be significant. The range decorrelation is approximately 1 percent of the satellite Crosstrack (XTK) or Alongtrack (ATK) error and less than 0.3 percent of the satellite radial (RAD) error for a 100 nmi distance from the reference station. In the absence of SA, orbit errors are on the order of 3-6 meters for XTK and ATK errors, respectively and approximately 1 meter for RAD errors. This would result in PRC errors of less than 6 cm for a 100 nmi distance. Unfortunately, XTK and ATK errors could be increased to 300 meters while maintaining the 100 meters (95%) positioning accuracy [8]. This would result in PRC errors of 3 meters for a 100 nmi distance. This level of error would exceed the 1.1 meter level, and in addition, the error cannot be observed by a local monitor station (Note: SCAT-I systems are to meet specifications out to a range of 20 nmi (minimum) from the runway threshold for each supported approach).

4. CODE/CARRIER INTEGRITY MONITORING

Code/Carrier Integrity Monitoring (CCIM) is based on two fundamental principals: 1) averaging the code phase measurements against the integrated Doppler frequency shift (carrier) measurements to provide low-variance pseudorange corrections, and 2) using the highest elevation satellite(s) to detect multipath errors on the lower elevation satellites. These two principals can, in general, be considered as averaging and differencing techniques. For averaging, the goal is to base the pseudorange corrections on the carrier errors rather than the errors affecting the code. Consider Figure 2, which illustrates the true and measured parameters typically available from a C/A code, standard positioning service (SPS) GPS receiver. Given the true integrated Doppler measurement, ID_{true}, the following can be defined:

$$ID_{meas} = ID_{true} + error_{carr} \quad (4)$$

$$PR_{true} = ID_{true} + \Delta \quad (5)$$

$$\Delta = PR_{true}(0) \quad (6)$$

$$PRmeas = PR_{true} + error_{code} \quad (7)$$

Using eq. 2, the pseudorange correction (PRC) can be defined as the difference between the pseudorange calculated at the reference station and the estimated (or measured) PR:

$$PRC = PR_{Ref(calc)} - \widehat{PR} \quad (8)$$

Basing $\widehat{PR}$ on the the measured carrier plus the CCIM offset $\hat{\Delta}$

$$\widehat{PR} = ID_{meas} + \hat{\Delta} \quad (9)$$

$$\widehat{PR} = ID_{true} + error_{carr} + \hat{\Delta} \quad (10)$$

yielding

$$PRC = PR_{Ref(calc)} - (ID_{true} + error_{carr} + \hat{\Delta}) \quad (11)$$

Given that $\hat{\Delta}$ and $PR_{Ref(calculated)}$ can be found, the resulting pseudorange correction is given as

$$PRC = -(error_{carr} + \varepsilon) \quad (12)$$

where ε is the residual error.

The CCIM offset can be estimated in real time using the expression

$$\hat{\Delta} = \frac{1}{N}\sum_{n=1}^{N} PR_{meas} - ID_{meas} \quad (13)$$

where the initial value of the CCIM offset, $\hat{\Delta}$, is based

on using an N-point initialization period over a time of zero-mean multipath. Subsequent values of the CCIM offset are based on a sliding, "fading-memory" window which is updated with each new sample.

To illustrate how the windowing in CCIM addresses rapid noise fluctuations (multipath) in the pseudorange measurements, consider the results shown in Figures 3. For these, data were collected at once per second using a Novatel GPSCard™ with the antenna placed on the hangar roof at the Ohio University Airport (UNI). The plots on the left show the residual and satellite elevation angle as a function of time for three representative satellites. The residual is defined as follows:

$$\text{Residual} = \text{Code} - \text{Carrier} - \text{mean}\{\text{Code} - \text{Carrier}\} \quad (14)$$

From the graphs it can be seen that high elevation satellites exhibit, in general, only moderate errors, with the large fluctuations being due primarily to the multipath associated with the rooftop antenna placement. The graphs on the right in Figure 3 show the standard deviation of the CCIM offset (using a 300-point window) for the corresponding data. From these it can be seen that the CCIM offset smoothes the residuals and forms a metric for determining the overall error in the pseudorange measurements. Note that the standard deviation (essentially the RMS error) in the CCIM offset exceeds the alarm limit of 1.1 meters in several cases, illustrating the need for real-time monitoring of the corrections. It is important to point out that CCIM can also accounts for the ionospheric divergence inherent in the single-frequency residual calculation.

5. UDRE AND CCIM DIFFERENCING

An additional consideration to SCAT-I integrity monitoring is the user differential range error (UDRE) parameter. The MASPS defines UDRE as an estimate of the pseudorange correction accuracy as estimated by the reference station. The confidence in this estimate should be at least 99.5%. While the MASPS is clear in this definition, it leaves the determination and usage of UDRE up to the system designers. While it is apparent that UDRE can be based on satellite elevation angle and heuristic anomalies such as known multipath, it is important to point out that UDRE should also be based on real-time measurements. This is illustrated by again referring to Figure 3, where it can be seen that satellites at moderate elevations (20° - 50°) are prone to substantial error fluctuations. In the aircraft, the UDRE is used to calculate the tunnel incident alarm. Another significant application of the UDRE can be in weighting the position solution for the aircraft. For instance, to accommodate for satellite-dependent UDREs, a UDRE weighting matrix can be defined and applied in the aircraft.

The other facet of the CCIM technique relies on the relative elevation of the satellites to detect and isolate faulty satellite measurements. Consider a high (H) and low (L) elevation satellite pair at both the reference (Ref) and monitor (Mon) stations. (The two satellite case is for illustrative purposes only, with the recognition that all low elevation satellites would be tested against the highest satellite). The pseudorange corrections can be differenced as follow:

$$\begin{aligned} PRC_{H_{Ref\,-\,Mon}} &= PRC_{H_{Ref}} - PRC_{H_{Mon}} \\ &= \delta B + error_H \end{aligned} \quad (15)$$

and

$$\begin{aligned} PRC_{L_{Ref\,-\,Mon}} &= PRC_{L_{Ref}} - PRC_{L_{Mon}} \\ &= \delta B + error_L \end{aligned} \quad (16)$$

where it is noted that the residual clock bias (δB) is equal for both (all) satellites. Differencing the low elevation satellite against the high yields:

$$\begin{aligned} PRC_{H_{Ref\,-\,Mon}} &- PRC_{L_{Ref\,-\,Mon}} \\ &= error_L - error_H. \end{aligned} \quad (17)$$

Assuming that $error_H << error_L$ ($\sqrt{\sigma_H^2 + \sigma_L^2} = \sigma_L$) then it is clear that differencing shows the errors on the low elevation satellite(s), thus providing a mechanism for fault detection and isolation. It is assumed that prior to this differencing a full set of computations will be performed which calculates the position based on each set of four satellites. This will allow for a final integrity check on all the data.

6. SUMMARY AND FUTURE WORK

It has been shown that Code/Carrier Integrity Monitoring contains concepts useful in meeting the integrity requirements of a precision approach system such as the DGNSS SCAT-I realization. The paper has also highlighted the critical integrity monitoring design parameters which must be addressed in the development of any precision approach system. Future work will address quantifying the residual errors as a function of satellite elevation angle. Effort will also be focused

towards the development of a full system to serve as a test bed for various implementation strategies.

ACKNOWLEDGEMENTS

This work was supported by the Federal Aviation Administration (ASE-300) under Research Grant Number 92-G-023 and the FAA/NASA Joint University Program for Air Transportation Research (Grant NAG1-1423).

REFERENCES

1. Minimum Aviation System Performance Standards, DGNSS Instrument Approach System: Special Category I (SCAT-I)," RTCA-DO/217, August 27, 1993.

2. Johnson, D.E. and M. Euhus, "ARINC Veracious Differential Reference Station (AVDRS) Development for Dallas/Fort Worth Trials," Institute of Navigation: Proceedings of the ION GPS-93, Salt Lake City, Utah, September 22 - 24, 1993.

3. Skidmore, T.A., and F. van Graas, "GPS Integrity Monitoring and Multipath Error Distributions," Institute of Navigation: Proceedings of the 49th Annual Meeting, Cambridge Massachusetts, June 21 - 23, 1993.

4. Davis, J.M., and R. Kelly, "RNP Tunnel Concept for Precision Approach with GNSS Applications," Institute of Navigation: Proceedings of the 49th Annual Meeting, Cambridge Massachusetts, June 21 - 23, 1993.

5. Braasch, M.S. and F. van Graas, "Guidance Accuracy Considerations for Real Time GPS Interferometry," Institute of Navigation: Proceedings of the ION GPS-91, Albuquerque, NM, September 1991.

6. Waid, J., "Ground Station Siting Considerations for DGPS," Institute of Navigation: Proceedings of the 1993 National Technical Meeting, San Francisco, California, January 20 - 22, 1993.

7. Goad, C., "Optimal Filtering of Pseudoranges and Phases from Single-Frequency GPS Receivers," Navigation: Journal of The Institute of Navigation, Vol. 37, No. 3, Fall 1990.

8. Loomis, P., Sheynblatt, L., and T. Mueller, "Differential GPS Network Design," Institute of Navigation: Proceedings of the ION GPS-91, Albuquerque, NM, September 1991.

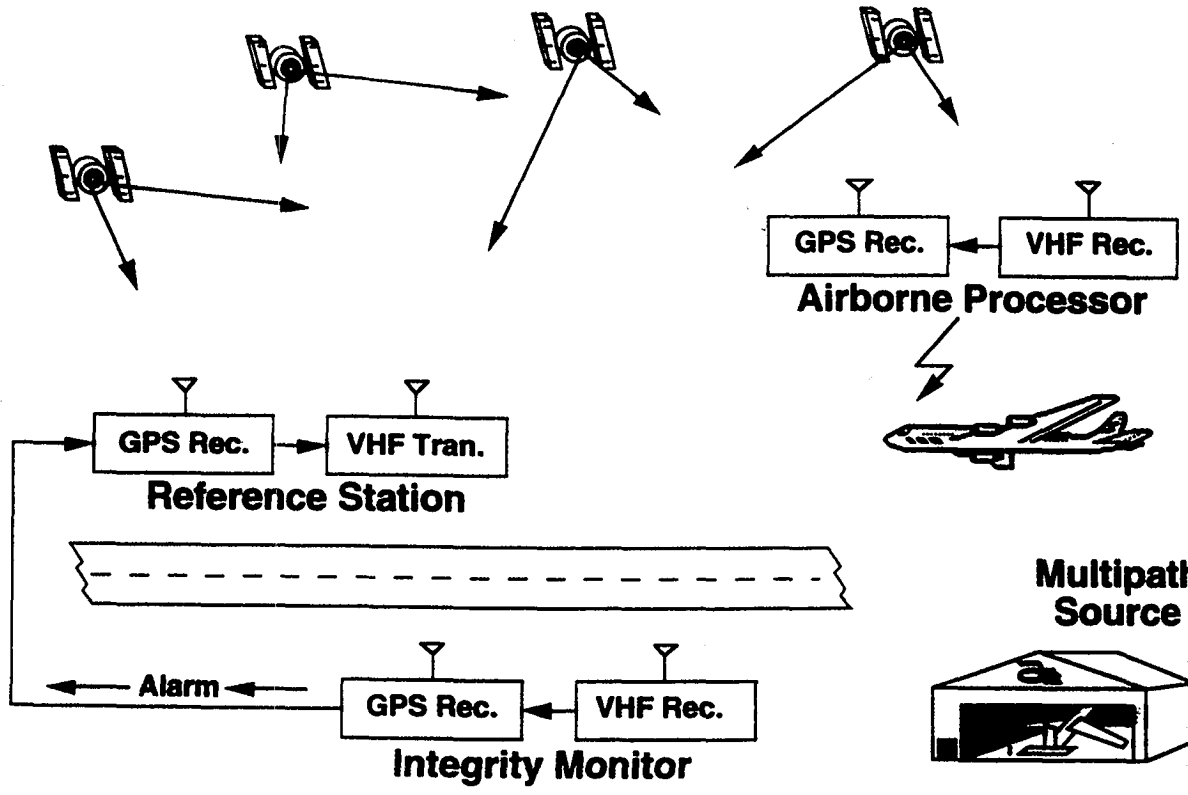


Figure 1. DGNSS Instrument Approach System

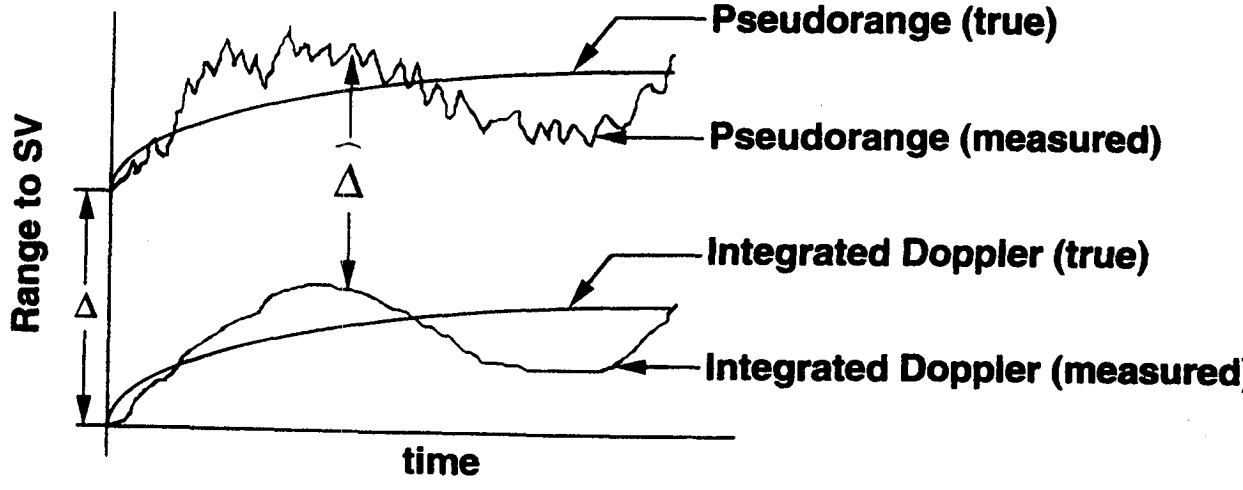


Figure 2. Code/Carrier Integrity Monitoring Parameters

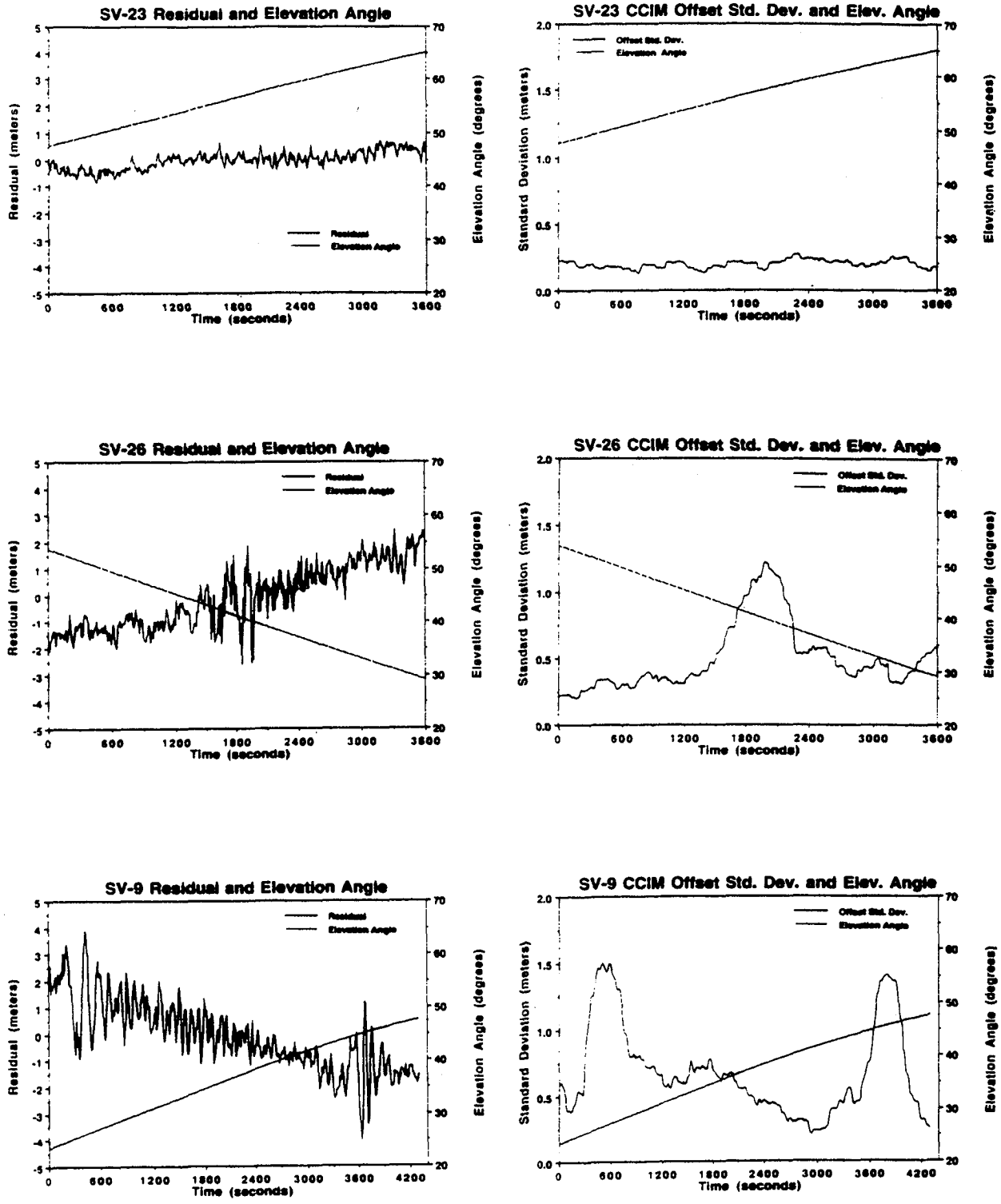


Figure 3. Sample Data: Residuals and CCIM Offset Standard Deviations

PRINCETON UNIVERSITY

INVESTIGATION OF AIR TRANSPORTATION TECHNOLOGY AT PRINCETON UNIVERSITY, 1993-1994

Professor Robert F. Stengel
Department of Mechanical and Aerospace Engineering
Princeton University
Princeton, New Jersey 08544

SUMMARY OF RESEARCH

The Air Transportation Research Program at Princeton University proceeded along four avenues during the past year:

- Microburst Hazards to Aircraft
- Wind Rotor Hazards to Aircraft
- Flight Control System Robustness
- Intelligent Aircraft/Airspace Systems

This research has resulted in a number of publications, including theses, archival papers, and conference papers. An annotated bibliography of publications that appeared between June 1993 and June 1994 appears at the end of this report. The research that these papers describe was supported in whole or in part by the Joint University Program, including work that was completed prior to the reporting period.

Severe downdrafts and resulting high velocity outflows caused by microbursts present a significant hazard to aircraft on takeoff and final approach. *Microbursts*, which are often associated with thunderstorm activity, also can occur in the vicinity of dissipating convective clouds that produce no rainfall at ground level. Microburst encounter is a rare but extremely dangerous phenomenon that accounts for one or two air carrier accidents and numerous general aviation accidents each year (on average). Conditions are such that an aircraft's performance envelope may be inadequate for safe penetration unless optimal control strategies are applied.

An expert system for wind shear avoidance that extends the FAA Microburst Windshear Guidelines to account for temporal and spatial variations in wind shear was developed in prior research. Measurements made by a look-ahead sensor (e.g., Doppler radar or lidar) are processed by extended

Kalman filters to develop a head-tailwind profile. An archival paper describing the wind-shear estimator was published during the period [1].

Real-time guidance for the case in which wind shear encounter has been encountered was being investigated. Our emphasis shifted from optimal strategies for abort and recovery [2, 3] to strategies based on nonlinear-inverse-dynamic controllers [4]. The former approach seeks to minimize a path-following cost function that implicitly maximizes the minimum altitude during an aborted approach to landing. The latter approach prescribes a desired rate of climb once an abort has been declared, then generates the necessary control commands by inverting the aircraft's dynamic model.

The dynamics and control of a twin-jet transport encountering an intense wind "rotor" were studied [5]. It was found that a physically realizable rotor could roll the aircraft to inverted attitude if left unopposed by lateral control. Similarly, unopposed full rudder deflection could invert the aircraft in its landing configuration. Conventional linear-quadratic flight control laws can maintain wings level through such encounters.

A new effort has begun to characterize the dynamics and control of following-aircraft response to leading-aircraft wake vortices. An aerodynamic model for a subsonic jet aircraft exposed to vortical flows will be developed, concentrating on the resultant forces and moments arising from rotating wind velocity distributions. Methods for designing feedback control logic that use available control power to minimize the disturbance to the following aircraft's flight path will be derived. This logic will combine features of optimization and nonlinear-inverse-dynamic control theory to synthesize practical digital control structures.

Control system robustness is defined as the ability to maintain satisfactory stability or performance characteristics in the presence of all conceivable system parameter variations. While assured robustness may be viewed as an alternative to gain adaptation or scheduling to accommodate known parameter variations, more often it is seen as protection against uncertainties in plant specification. Consequently, a statistical description of control system robustness is consistent with what may be known about the structure and parameters of the plant's dynamic model. Rarely will there be a single "most robust" controller, as design tradeoffs must inevitably be considered. For example, stability, settling time, and control usage all may be of concern; controllers that favor one criterion over the other two have dramatically different characteristics.

Our initial research focused on probabilistic analysis of the stability and performance robustness of given controllers, while more recent research has shifted to designing robust controllers [6]. Numerical search using a genetic algorithm produces robust controllers based on proportional-filter linear-quadratic regulators with implicit model-following. These controllers compare favorably to others designed by competing methods (e.g., those that minimize H_∞ cost functions). This research is proceeding with applications to nonlinear control system design and with the incorporation of parallel processors to speed computations.

Advanced concepts for air traffic management are being developed by modeling aircraft and air traffic centers as intelligent agents that engage in principled negotiation [7]. Each agent is characterized as a dynamic system that carries out declarative, procedural, and reflexive functions [8, 9]. Principled negotiation entails the proposal of alternative flight plans, evaluation of costs and constraints according to separate and shared interests, and conflict resolution. We are setting the groundwork for an *Intelligent Aircraft/Airspace System (IAAS)*. The goal is to identify means by which ground-based and airborne flight management systems can cooperate to produce a net gain in the efficiency and robustness of air transportation.

Prior research on modeling an aircraft's aerodynamic characteristics using neural networks was reported in an archival paper during the period [10]. It is shown that neural networks provide a promising means of representing aerodynamics for adaptive control and flight simulation.

Two doctoral theses were completed during the period. Sandeep Mulgund developed near-optimal nonlinear guidance logic for aircraft encountering microburst wind shear [11]. Dennis Linse developed the use of neural networks for aerodynamic modeling mentioned above [12].

ANNOTATED BIBLIOGRAPHY OF 1993-1994 PUBLICATIONS

1. D. A. Stratton and R. F. Stengel, "Robust Kalman Filter Design for Predictive Wind Shear Detection," *IEEE Trans. Aerospace and Electronic Systems*, Vol. 29, No. 4, Oct. 1993, pp. 1185-1194.

Severe, low-altitude wind shear is a threat to aviation safety. Airborne sensors under development measure the radial component of wind along a line directly in front of an aircraft. In this paper, optimal estimation theory is used to define a detection algorithm to warn of hazardous wind shear from these sensors. To achieve robustness, a wind shear detection algorithm must distinguish threatening wind shear from less hazardous gustiness, despite variations in wind shear structure. This paper presents statistical analysis methods to refine wind shear detection algorithm robustness. Computational methods predict the ability to warn of severe wind shear and avoid false warning. Comparative capability of the detection algorithm as a function of its design parameters is determined, identifying designs that provide robust detection of severe wind shear.

2. S. S. Mulgund and R. F. Stengel, Target Pitch Angle for the Microburst Escape Maneuver, *J. Aircraft*, Vol. 30, No. 6, Nov.-Dec. 1993, pp. 826-832.

Commuter and general aviation aircraft face no less a threat from microburst wind shear than do large jet transports, yet most studies of wind shear encounter have neglected them. The effects of microburst wind shear on a propeller-driven commuter aircraft are considered here. Recovery performance of a commuter-type aircraft in a microburst encounter is examined using a constant-pitch-attitude strategy and flight-path optimization. The goals are to identify a suitable target pitch angle for the escape maneuver and to determine the nature of an optimal escape maneuver for commuter aircraft. The results demonstrate that the pitch attitude that maximizes climb rate in a wind shear is strongly dependent on whether the shear is predominantly a downdraft or a horizontal shear. Simulated recoveries show that the optimal constant pitch angle depends on the altitude of encounter, the strength of the microburst, and the initial position of the aircraft relative to the microburst core. In severe wind shear encounters at very low altitude, best results are obtained at relatively low pitch angle. Excessively high target pitch angles subject the aircraft to prolonged periods near stall. Flight path optimization demonstrates that maximum ground clearance is obtained by maintaining a low pitch attitude early in the encounter followed by a gradual pitch-up that ceases when the aircraft exits the wind shear.

3. S. S. Mulgund and R. F. Stengel, Optimal Recovery from Microburst Wind Shear, *J. Guidance, Control, and Dynamics*, Vol. 16, No. 6, Nov.-Dec. 1993, pp. 1010-1017.

The flight path of a twin-jet transport aircraft is optimized in a microburst encounter during approach to landing. The objective is to execute an escape maneuver that maintains safe ground clearance and an adequate stall margin during the climb-out portion of the trajectory. A cost function penalizing rate of climb deviations from a nominal value and the rate of elevator deflection produces qualitatively good results in a variety of microburst encounters. The optimal maneuver is a gradual pitch-up that ceases near the core of the microburst, followed by a slight reduction in pitch attitude in the tailwind area of the microburst. A minimum airspeed constraint in the optimization prevents excessive airspeed loss in very severe microbursts. The aircraft equations of motion include short-period dynamics, so that the optimization solves directly for the control surface deflections required to achieve the optimal flight paths.

4. S. S. Mulgund and R. F. Stengel, "Aircraft Flight Control in Wind Shear Using Partial Dynamic Inversion," *Proceedings of the 1993 American Control Conference*, San Francisco, June 1993, pp. 400-404.

A flight control law based on partial inversion of the longitudinal dynamics of a twin-jet transport aircraft is presented. The controller is partitioned into a slow-time-scale and a fast-time scale to simplify its design. Three types of controllers are developed: airspeed/climb rate, groundspeed/climb rate, and throttle/climb rate. For microburst encounters during approach to landing, it is found that a combination of airspeed and groundspeed regulation is quite effective for controlling the flight path to touchdown. Regulation of groundspeed to a nominal value in the performance-increasing region of the microburst prevents an inadvertent reduction in thrust, while regulation of airspeed to a nominal value in the performance-decreasing area of the microburst prevents excessive airspeed loss. The throttle/climb rate controller is used for aborted-landing encounters. The combination of groundspeed and airspeed control is used until the decision is made to abort the landing, at which point maximum throttle and a specified positive climb rate are commanded.

5. D. R. Spilman and R. F. Stengel, "Jet Transport Response to a Horizontal Wind Vortex," *AIAA 31st Aerospace Sciences Meeting*, Reno, AIAA Paper No. 94-0811, Jan. 1994.

The dynamic response of a twin-jet transport aircraft encountering a single-axis wind vortex on final approach to landing is investigated. Severe performance degradation and possible ground impact may result from a low-altitude encounter with a wind rotor, which is formed by strong winds that flow over a mountain range and roll up on the leeward side of the mountain. The simulation makes use of the similarities between flow induced over the aircraft surfaces by angular rates and the flow induced by a wind gradient. A single-axis vortex model approximates the wind velocity field. Dynamic simulations illustrate the effects of vortex strength, vortex length, lateral entry position, vertical entry position, and encounter incidence angle on the aircraft roll response parameters. Results show that maximum roll rate and roll angle increase proportionally with vortex strength and vortex length, until a "saturation length" is reached. Roll response is highly dependent on entry location: changes in lateral entry position affect maximum roll angle while changes in vertical entry position affect maximum roll rate. Peak roll rate and roll angle obtain their largest values at near-zero incidence angles. The response is highly dependent on the initial conditions of the encounter -- even small variations cause significant changes in aircraft roll response.

6. C. I. Marrison and R. F. Stengel, "The Use of Random Search and Genetic Algorithms to Optimize Stochastic Robustness Functions," *Proc. 1994 American Control Conference*, Baltimore, Jun. 1994.

Stochastic robustness synthesis is a framework for designing practical control systems. It uses Monte Carlo simulation to evaluate the quality of candidate designs, and it searches a parameter space to find the best one. The global minimum of a probabilistic criterion function must be found, ideally with a minimum number of evaluations. This paper examines two approaches to minimizing the probabilistic function: random search and a genetic algorithm. The genetic algorithm is similar to previously published algorithms but has several modifications to improve its performance, most notably a clustering analysis at the beginning of each generation. Statistical tools are incorporated in the search algorithms, allowing intelligent decisions to be based on the "noisy" Monte Carlo estimates. Performance of the two methods is demonstrated by application to a 24-dimensional test function. The genetic algorithm is shown to be significantly better than the random search for this application. The genetic algorithms is then used to design compensators for a benchmark problem, producing control laws with excellent levels of stability and performance robustness.

7. R. F. Stengel and J. P. Wangermann, "Air Traffic Management as Principled Negotiation Between Intelligent Agents," in *Machine Intelligence in Air Traffic Management*, AGARD-CP-538, Neuilly-sur-Seine, Oct. 1993, pp. 5-1 to 5-10.

Air transportation provides the backbone for passenger transport over moderate to long distances in the U.S. and much of the world, and it is becoming an increasingly important mode for short-range travel and cargo transport as well. There is a growing demand for use of available airspace and a heightened concern for on-time performance. Demand frequently exceeds available capacity of the airspace system, causing flight delays, negative economic impact, and passenger inconvenience [1, 2]. New technologies are emerging that will make flight operations both simpler and more complex. On the one hand, advances hold promise for increasing the productivity, reliability, and safety of the air transportation system. On the other, advances in technology introduce uncertainty, increase human workload (if not properly implemented), increase the potential for dispute, and present new challenges for both certification and day-to-day operations. This paper presents a concept for an *Intelligent Aircraft/Airspace System (IAAS)* that could be a focal point for developing air traffic management in the coming decades. The *IAAS* would integrate the capabilities of all ground-based and airborne components of the system (identified as *Intelligent Agents*) in order to provide increased capacity and maintained or improved safety. *Principled Negotiation* is proposed as a framework for interactions between intelligent agents.

8. R. F. Stengel, "Toward Intelligent Flight Control,"*IEEE Trans. Systems, Man, and Cybernetics*, Vol. 23, No. 6, Nov.-Dec. 1993, pp. 1699-1717.

Flight control systems can benefit by being designed to emulate functions of natural intelligence. Intelligent control functions fall in three categories: declarative, procedural, and reflexive. Declarative actions involve decision-making, providing models for system monitoring, goal planning, and system/scenario identification. Procedural actions concern skilled behavior and have parallels in guidance, navigation, and adaptation. Reflexive actions are more-or-less spontaneous and are similar to inner-loop control and estimation. Intelligent flight control systems will contain a hierarchy of expert systems, procedural algorithms, and computational neural networks, each expanding on prior functions to improve mission capability, to increase the reliability and safety of flight, and to ease pilot workload.

9. R. F. Stengel, "Intelligent Flight Control Systems," in *Aerospace Vehicle Dynamics and Control*, Clarendon Press, Oxford, UK, 1994, pp. 33-80.

The capabilities of flight control systems can be enhanced by designing them to emulate functions of natural intelligence. Intelligent control functions fall in three categories. *Declarative* actions involve decision-making, providing models for system monitoring, goal planning, and system/scenario identification. *Procedural* actions concern skilled behavior and have parallels in guidance, navigation, and adaptation. *Reflexive* actions are spontaneous, inner-loop responses for control and estimation. Intelligent flight control systems learn knowledge of the aircraft and its mission and adapt to changes in the flight environment. Cognitive models form an efficient basis for integrating "outer-loop/inner-loop" control functions and for developing robust parallel-processing algorithms.

10. D. Linse and R. F. Stengel, "Identification of Aerodynamic Coefficients Using Computational Neural Networks," *J. Guidance, Control, and Dynamics*, Vol. 16, No. 6, Nov.-Dec. 1993, pp. 1018-1025.

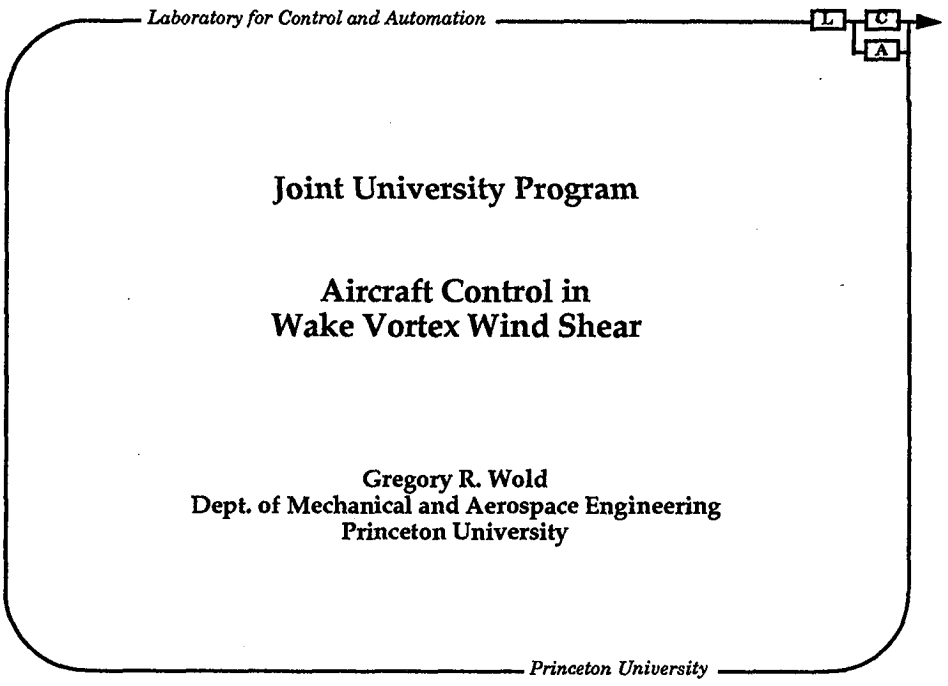
Precise, smooth aerodynamic models are required for implementing adaptive, nonlinear control strategies. Accurate representations of aerodynamic coefficients can be generated for the complete flight envelope by combining computational neural network models with an Estimation-Before-Modeling paradigm for on-line training. A novel method of incorporating first-partial-derivative information is employed to estimate the weights in individual feedforward networks for each coefficient. The method is demonstrated by generating a model of the normal force coefficient of a twin-jet transport aircraft from simulated flight data, and promising results are obtained.

11. S. S. Mulgund, *Longitudinal Aircraft Flight in Microburst Wind Shear*, Ph. D. Thesis, Princeton University, MAE-1995-T, June, 1994.

Severe low-altitude wind variability poses a significant hazard to aircraft in the terminal area. Longitudinal control strategies to improve flight safety in microburst wind shear are developed using deterministic trajectory optimization and feedback control based on the aircraft's nonlinear inverse dynamics. Optimal state estimates for feedback are computed using extended Kalman filtering.

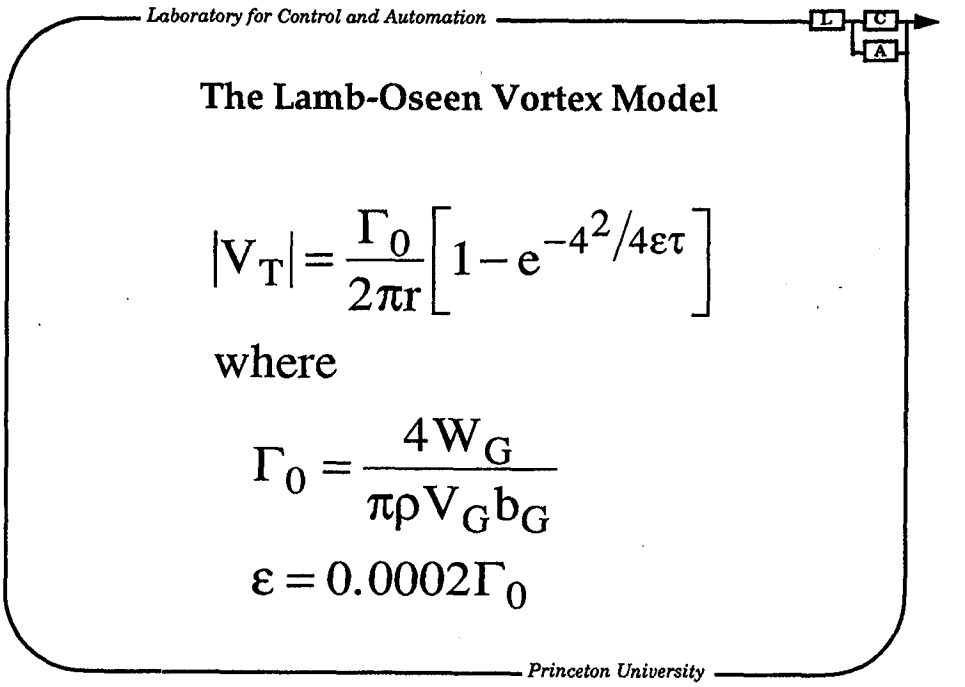
12. D. J. Linse, *System Identification for Adaptive Nonlinear Flight Control Using Neural Networks*, Ph. D. Thesis, Princeton University, MAE-1997-T, June 1994.

Nonlinear flight control strategies require precise, smooth representations of the airplane's aerodynamic model. Most nonlinear system identification techniques do not provide smooth models across a wide range of flight conditions. Using computational neural networks -- biologically inspired, massively parallel computational structures -- to perform generalized spline function approximation provides an excellent building block for identification. A novel training method incorporates first-partial-derivative information in the network's learning algorithm to constrain the network function space and to aid in selecting the weights in the network. An estimation-before-modeling structure provides a framework for generating online training information. Identification of a normal-force model for a twin-jet transport airplane demonstrates the technique. Independent maneuvers validate the network's performance in untrained regions of the flight envelope. They are adaptive, they learn from examples, and they can provide excellent approximation for a large class of functions.



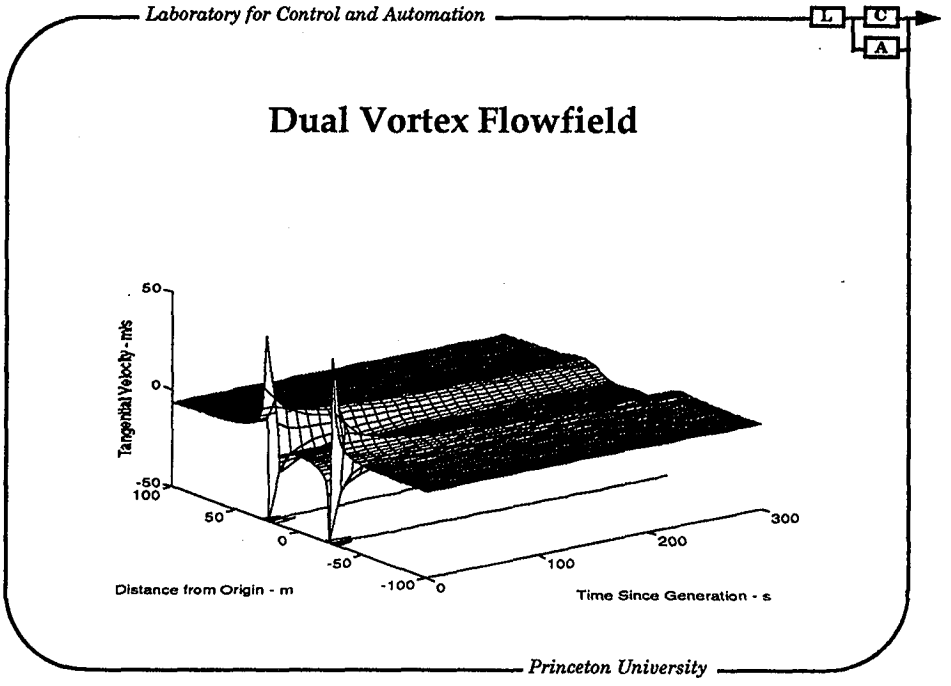
In the past, there have been a number of fatal incidents attributable to wake vortex encounters, involving both general aviation and commercial aircraft. In fact, the wake vortex hazard is considered to be the single dominant safety issue determining the aircraft spacing requirements at airports.

As the amount of air traffic increases, the number of dangerous encounters is likely only to increase. It is therefore imperative that a means be found to reduce the danger. That is the purpose of this research: to use nonlinear inverse dynamic (NID) control methods in the design of an aircraft control system which can improve the saftey margin in a wake vortex encounter.



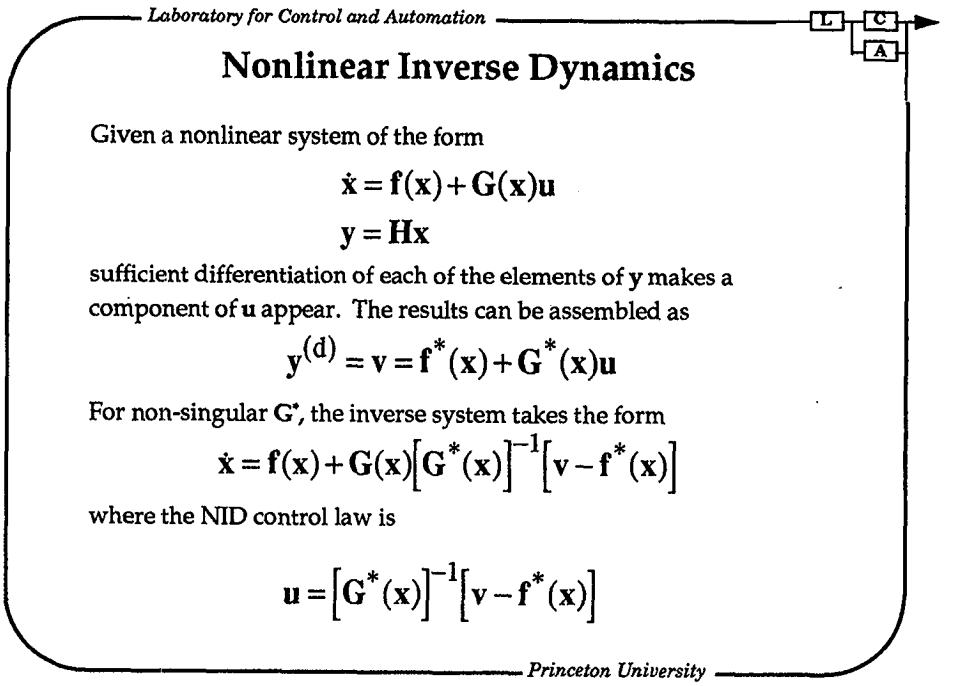
The Lamb-Oseen vortex model has been widely used to characterize the air flow within a wake vortex. The vortex strength is proportional to the weight of the generating craft, W_G, and inversely proportional to its speed, V_G. This highlights the fact that the danger is greatest for smaller craft following large jet transports near airports, either on takeoff or, more commonly, on landing.

Although the danger of a crash is greatest for general aviation aircraft following larger transports, even planes of nearly equal size to the wake-generating craft can experience hazardous excursions, according to pilots' subjective ratings.



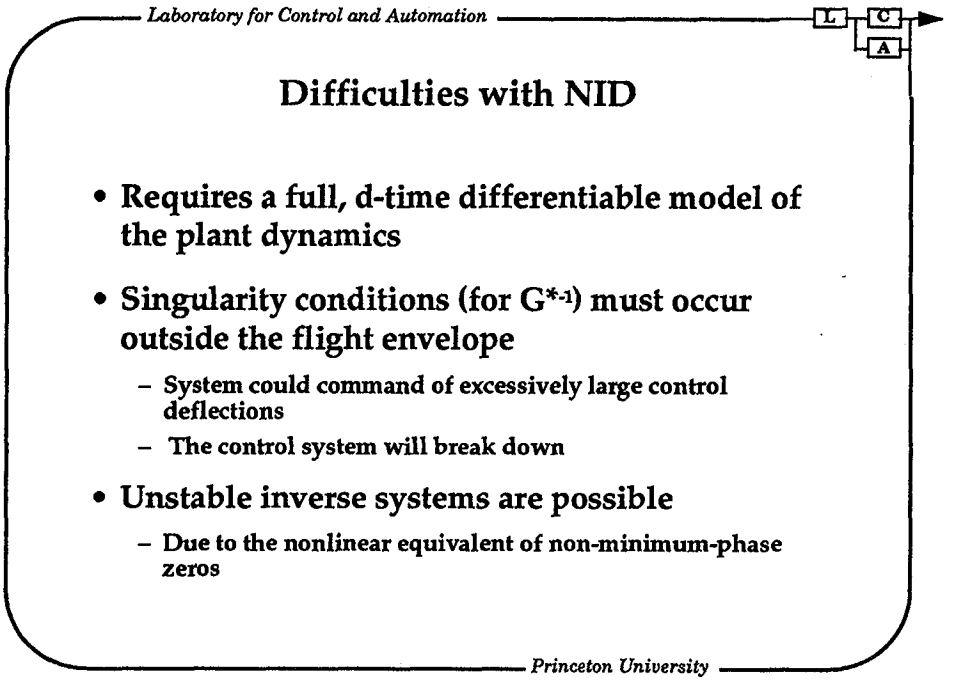
The wake vortex flowfield is characterized by two vortices with opposing rotations, one from each wing of the generating craft. A following airplane, entering the flowfield nearly parallel to the vortex axis, would initially roll away from the vortex. The pilot's response would be to counteract the roll to maintain his approach. This would carry the plane into the vortex core, where the vortex roll would immediately change direction, further rolling the plane, possibly flipping it, and driving it into the strong central downdraft. If the encounter occured too close to the ground, the pilot would be unable to recover the airplane, and a crash would ensue.

Vortex decay properties are currently under investigation. Measurements have indicated that the vortex strength does not begin to decrease immediately, as shown in the plot of dual Lamb-Oseen vortices above, but stays nearly constant as far as three to six nautical miles behind the generating craft.

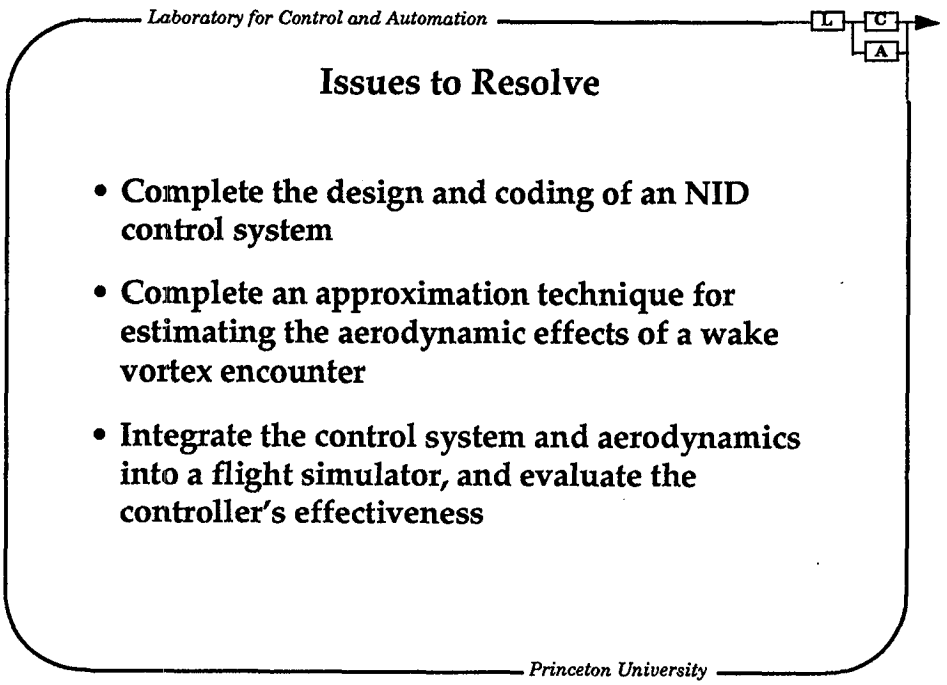


Conventional flight control systems are developed assuming the aircraft dynamics are linear and time invariant about some nominal flight condition. The performance of such linear control systems breaks down in extreme flight conditions, as the higher-order terms of linearization become non-negligible. A wake vortex encounter is just such an extreme condition.

Nonlinear inverse dynamics is one means of better accounting for these higher-order effects, providing better control over a greater portion of the flight envelope. An output vector y dependent on x is defined, and is differentiated with respect to time element by element until the control effect appears. The assembled derivative vector v becomes a new control input vector, which can be used to set the system dynamics.



Nonlinear Inverse Dynamic controllers are characterized by a significantly higher computational cost than are linear controllers. This is due to the need for the full, differentiable model of the plant dynamics within the controller. Such controllers, because of their dependence on the non-singularity of G^*, must also generally be paired with a warning system or active means of overriding commands in case singularity conditions (such as excessive angles of attack) are approached. Finally, the control system design is often based on a reduced plant model, as small effects, if fully accounted for, can result in system instability.



Current and future work will complete the design and testing of an NID controller, using a B-737 flight simulator. This will include code for estimating the effect of highly non-linear wind shear terms on the aerodynamic coefficients of the aircraft. As the NID control system is highly dependent on the chosen output vector, it is likely that several different designs will have to be tested before an appropriate one is found.

Research of Stochastic Robustness: Results, and Conclusions

Chris Marrison
Princeton University

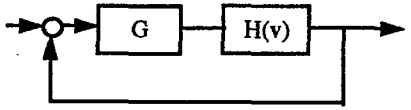
Joint Universities Programme
1994

With Stochastic Robustness we are creating tools to design robust compensators for practical systems.

During this year, the Stochastic Robustness research achieved the following results:

- Refined the search tools needed for synthesis.
- Successfully designed robust compensators for the American Controls Conference benchmark problem.
- Sucessfully designed robust compensators for a nonlinear Hypersonic aircraft model with uncertainties in 28 parameters.

- **The Problem**



Compensator, G
Plant Structure, H
Vector of Uncertain Parameters, v
Distribution of Uncertain Parameters, pr(v)

- **The Indicator Function**

Violation: $I[H(v),G]=1$, otherwise $I[H(v),G]=0$

- **The Robustness Metric**

$$P = \int_V I[H(v),G]\, pr(v)\, dv$$

- **The Estimate available from Monte Carlo evaluation**

$$\hat{P} = \frac{1}{N}\sum_{m=1}^{N} I[H(v_m), G] \quad v_m \text{ from } pr(v)$$

The Stochastic Robustness metric defines the robustness in terms of the *probability* of the closed-loop system allowing a metric to be violated. The binary metric is defined by the designer and is very flexible; it is able to describe stability, performance criteria, and actuator usage.

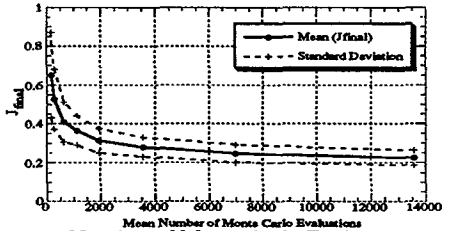
Random Search

Usual Random Search but with Statistical Tools:

- Reduce Variance by Stratification of V
- Compare Pairs of Sample Points by Bootstrapping with fixed set of v
- Estimate Optimal N

Performance on a Test Function

Mean of Result



Number of Monte Carlo Evaluations

Several statistical tools were developed to improve the efficiency of the synthesis methods. These tools were combined with random search to create a search algorithm that effectivly searches stochastic cost functions.

To quantify the performance of the random search it was used on a test function. This graph shows the expected result of the search as a function of the number of Monte Carlo evaluations.

Laboratory for Control and Automation — L C A

Genetic Algorithm

Initially a standard algorithm with decision mechanism for N

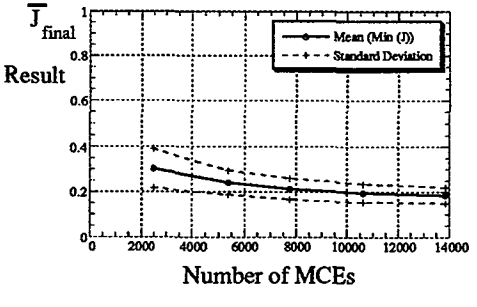
Number of Evaluations to Assess Next Generation:

$$N_{required} = N_{current}\left(\bar{\sigma}_{\hat{J}\ current} / \bar{\sigma}_{\hat{J}\ desired}\right)^2$$

$\bar{\sigma}_{\hat{J}\ desired}$ is based on the dispersion of the population

$\bar{\sigma}_{\hat{J}\ current}$ is based on Bootstrapping

Performance of the Baseline Algorithm

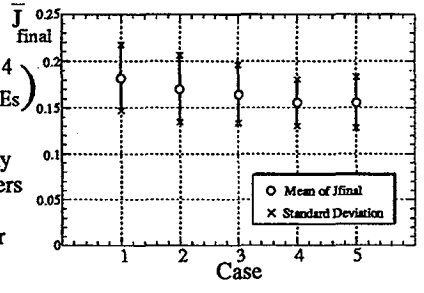


Princeton University 4

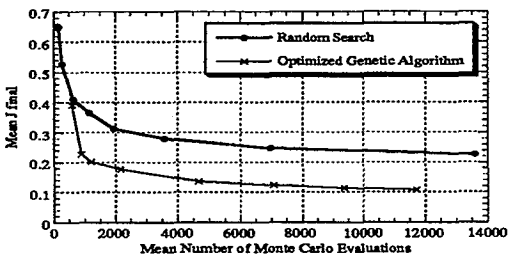
A genetic algorithm was also implemented and used to search the test function. The performance curve is shown here.

Addition of Clustering and Elite Selection

Case:
1) Baseline
2) Best Member only
3) Best Two Members
4) Best plus Cluster
5) Two from Cluster



Relative Performance of the Random Search and the Optimized Genetic Algorithm



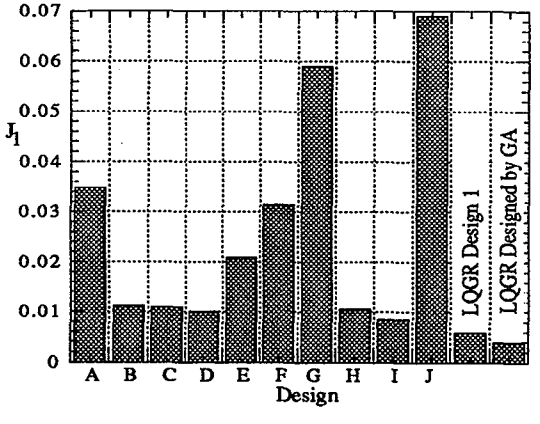
The genetic algorithm was improved by adding elite selection and a clustering algorithm.

Further improvements were made by tuning the parameters within the algorithm.

Here the performance of the optimized genetic algorithm is compared with that of the random search. The genetic algorithm is significantly better, taking 10 times fewer Monte Carlo evaluations to reach the same result.

Relative Robustness of Benchmark Compensator Designed by Genetic Algorithm

$$J = P_i^2 + 0.01\,P_{Ts}^2 + 0.01\,P_u^2$$



The genetic algorithm was used to design robust compensators for the American Controls Conference benchmark problem.

The cost function combines probability of instability, probability of settling time violation, and probability of actuator saturation.

The resulting compensator had a much lower robustness cost than 10 other designs prepared by other authors, and had a lower cost than the design found during earlier Stochastic Robustness research.

Hypersonic Aircraft Longitudinal Dynamics

Problem Statement

Unstable

Fifth Order

Nonlinear

28 Uncertain Parameters

39 Robustness Metrics

Compensator Structures

Linear Quadratic Regulators:

Proportional - Integral (10 design parameters)

Proportional - Integral - Filter (12 design parameters)

Robustness Cost Functions

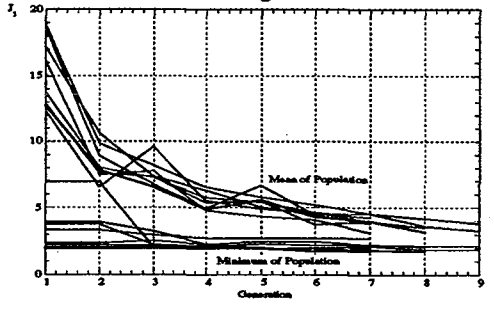
Weighted Quadratic Sums

The synthesis tools were proved on a complex problem by designing robust compensators for a hypersonic aircraft model (Based on the Hypersonic Vehicle Simulation Model: Winged-Cone Configuration, NASA TM 102610).

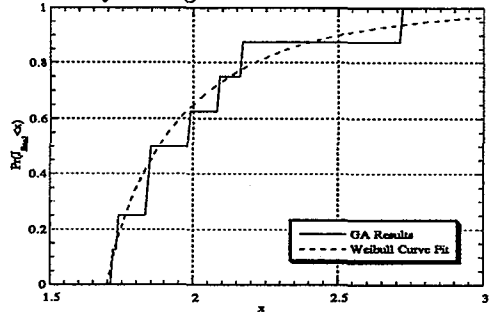
28 aerodynamic and physical parameters were taken to be uncertain and 39 aspects of performance were considered.

The compensator structures were full state feedback Linear Quadratic regulators

Results of Eight GA Runs



Estimation of Global Minimum by Fitting a Weibull Distribution



- Best Result is within 1% of Global Minimum
- Changing the Weights Suppressed Corresponding Probabilities
- PIF Eliminated Spikes

The algorithm was run eight times. The distribution of the cost of the eight results allows us to estimate the value of the global minimum.

We can conclude that the best of the runs gave a compensator with a robustness cost within one percent of the best robustness possible given this compensator structure.

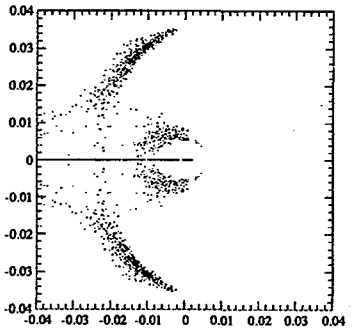
Robustness Profile

Probabilities	Open Loop	Base Compensator	Optimized Compensator
J	15.4832	2.7744	1.7405
P_i	0.8160	0.0145	0.0003
Pv, Ts25	0.9980	0.9665	0.5165
Pv, Ts50	0.9980	0.7065	0.1730
Pv,R25	0.9980	0.7475	0.4698
Pv,R50	0.9980	0.3540	0.1037
Pv,NMP	1.0000	0.3095	0.0297
Pv,D5	1.0000	0.0467	0.0127
Pv,D10	0.9960	0.0295	0.0115
Pv,OS10	0.0360	0.3463	0.0405
Pv,OS20	0.0360	0.2112	0.0135
Pv,Da0.5	0.0040	0.0982	0.0568
Pv,Da1	0.0040	0.0822	0.0413
Pv,g1	0.6800	0.1167	0.0650
Pv,g2	0.6580	0.1153	0.0635
Pv,Db0.25	0.8760	0.4490	0.3350
Pv,Db0.5	0.5900	0.1665	0.1065
Pv,dT50	0	0.1202	1.0000
Pv,dT100	0	0.1183	0.0658
Pv,dE5	0	0.1163	0.0645
Pv,dE10	0	0.1157	0.0640
Ph, Ts50	1.0000	1.0000	1.0000
Ph, Ts100	1.0000	0.8985	0.8692
Ph,R50	0.9960	0.9810	0.9888
Ph,R100	0.9960	0.4875	0.3645
Ph,NMP	0.0280	0.0152	0.0010
Ph,D10	0.9580	0.9935	0.9832
Ph,D20	0.9060	0.2145	0.1815
Ph,OS20	0.0300	0.0187	0.0003
Ph,OS40	0.0280	0.0148	0.0003
Ph,Da0.5	0.0140	0.0862	0.0410
Ph,Da1	0.0140	0.0742	0.0325
Ph,g1	0.8380	0.8858	0.9345
Ph,g2	0.7760	0.3110	0.4475
Ph,DV0.25	0.9280	0.9898	0.0522
Ph,DV0.5	0.8200	0.2118	0.0480
Ph,dT50	0	1.0000	0.0517
Ph,dT100	0	0.0997	0.0495
Ph,dE5	0	0.1008	0.0527
Ph,dE10	0	0.0973	0.0485

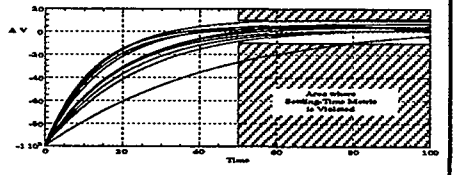
The final compensator was subjected to a full Monte Carlo analysis. The numerical results give a detailed picture of the Robustness profile for each aspect of performance.

Graphical Tools

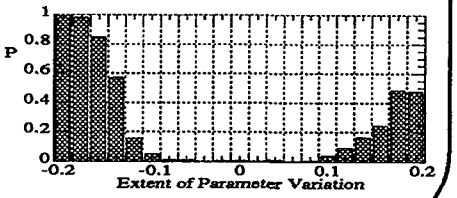
Stochastic Root Locus



Stochastic Response Envelope



Conditional Probability of Metric Violation



Graphical tools also give insight into the nature of the final compensator's robustness.

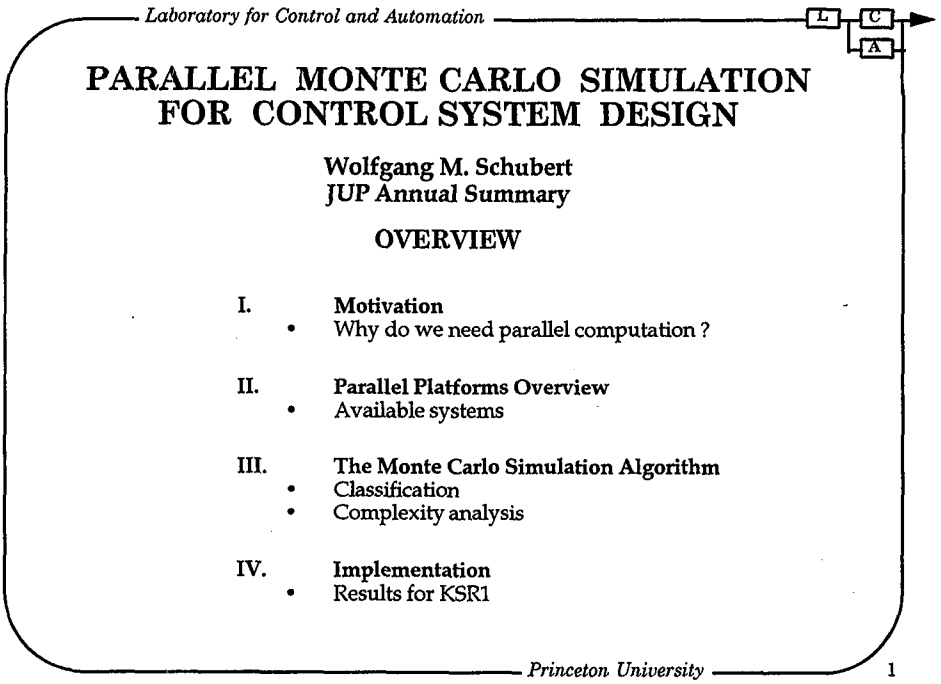
From the stochastic root locus we can see the types of instability that are possible.

The stochastic response envelopes show that in the cases when the performance metric is violated, the response is still benign.

The conditional probability graphs show which parameter variations cause metric violation.

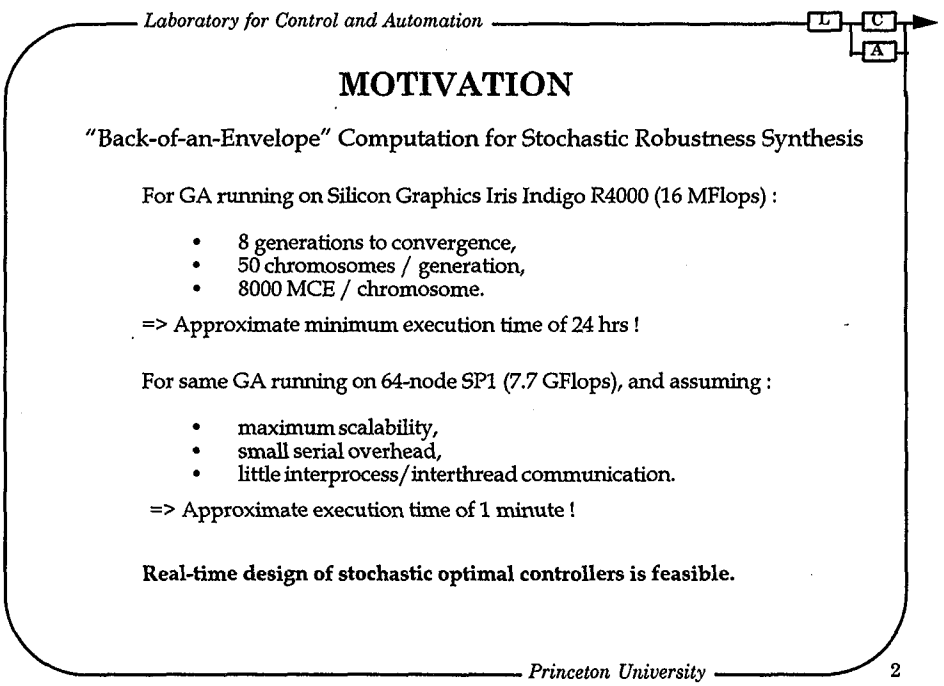
Conclusions

- LQGRs can be Robust
- Stochastic Robustness is excellent for the comparison of compensators
- Stochastic Robustness is excellent for synthesis of robust compensators
- Graphics give useful insights
- Stochastic Robustness synthesis was proven to be flexible
- Synthesis can be conducted efficiently and rigorously
- The tools are ready to be used

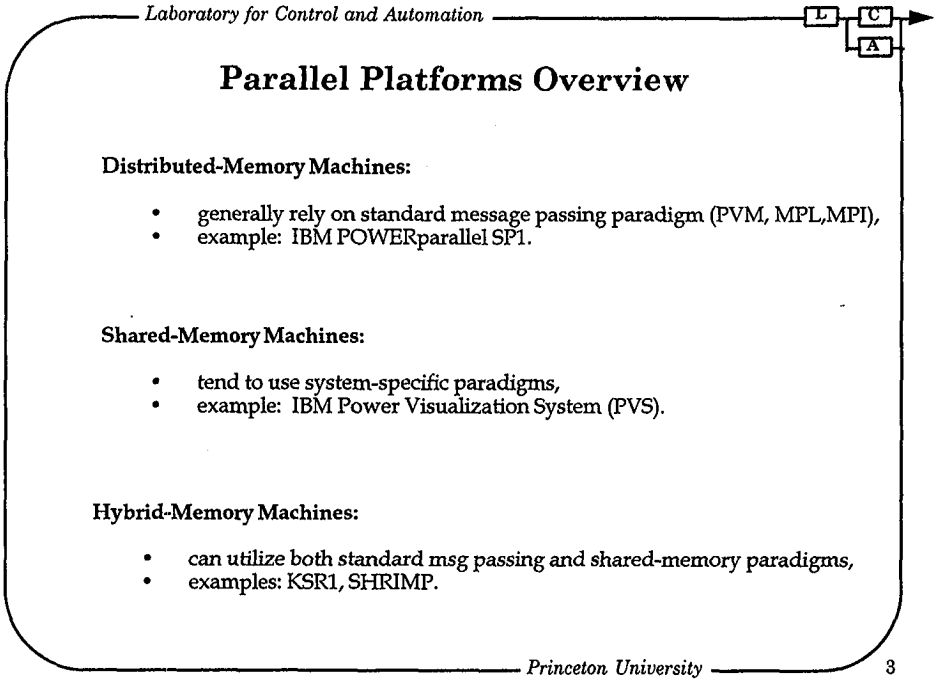


The research during the 1993/94 academic year addressed the design of parallel algorithms for Stochastic Robustness Synthesis (SRS). SRS uses Monte Carlo Simulation to compute probabilities of system instability and other design-metric violations. The probabilities form a cost function which is used by a genetic algorithm (GA). The GA searches for the stochastic optimal controller.

The existing sequential algorithm was analyzed and modified to execute in a distributed environment. For this, parallel approaches to Monte Carlo Simulation and Genetic Algorithms were investigated. Initial empirical results are available for the KSR1.

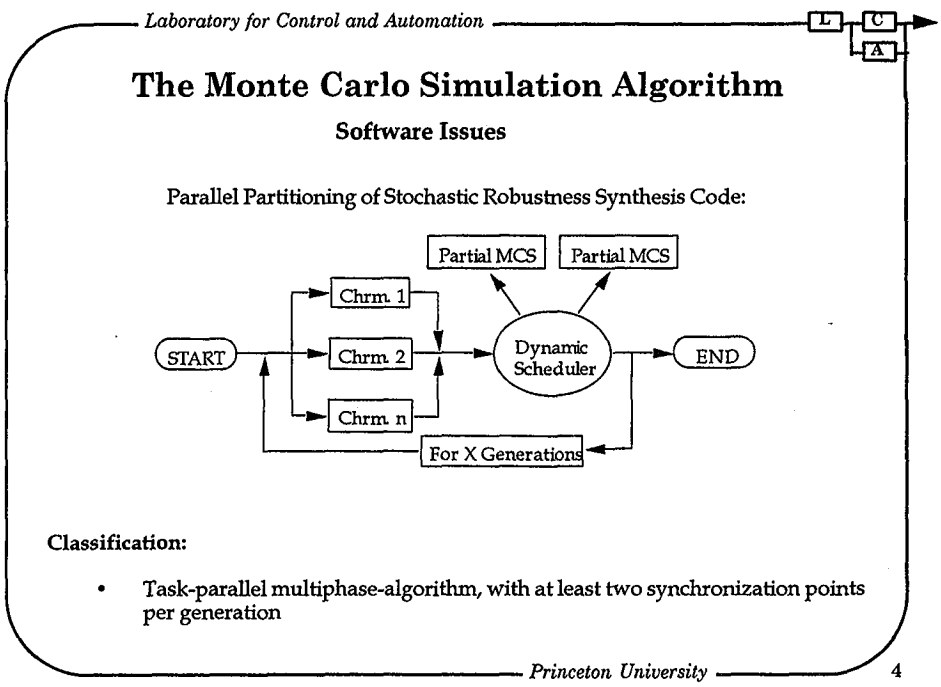


Stochastic Robustness Synthesis is computationally intensive. Execution times to complete one design using Matlab on an SGI Indigo R4000 have been on the order of several days. Using high-performance parallel hardware, such as the KSR1, and appropriate algorithms, controller designs could be computed within minutes. Additional advantages accrue as execution times decrease; for example real-time visualization tools for stochastic root loci could be developed.

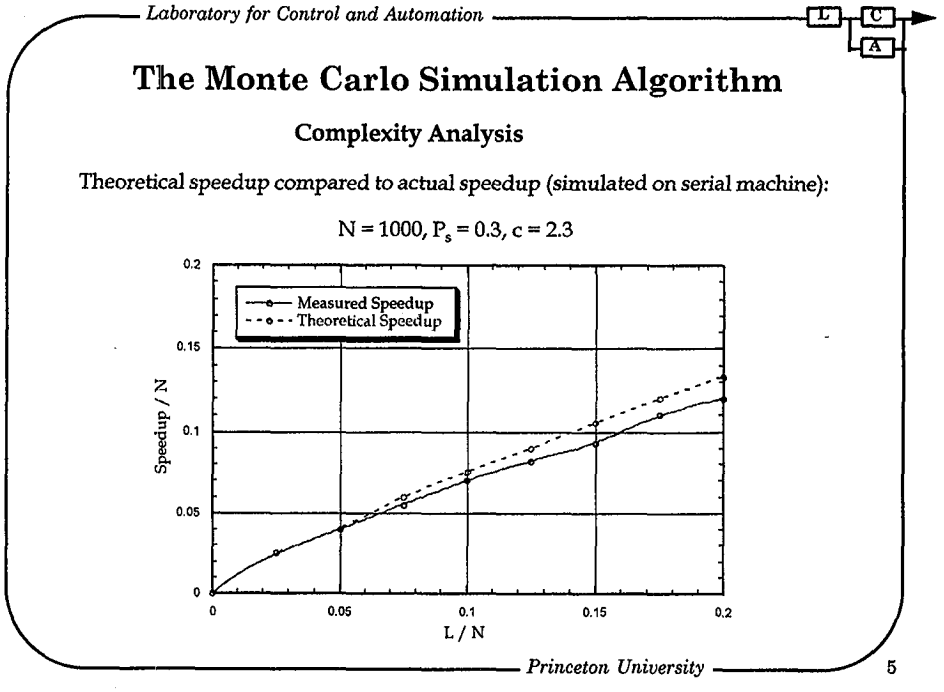


Most available parallel machines differ dramatically from each other in hardware and software architectures. The architectures may be separated into two main groups: shared-memory, with individual processes/threads sharing global data; and distributed-memory, with each process/thread owning its private data and message-passing being required to exchange the data.

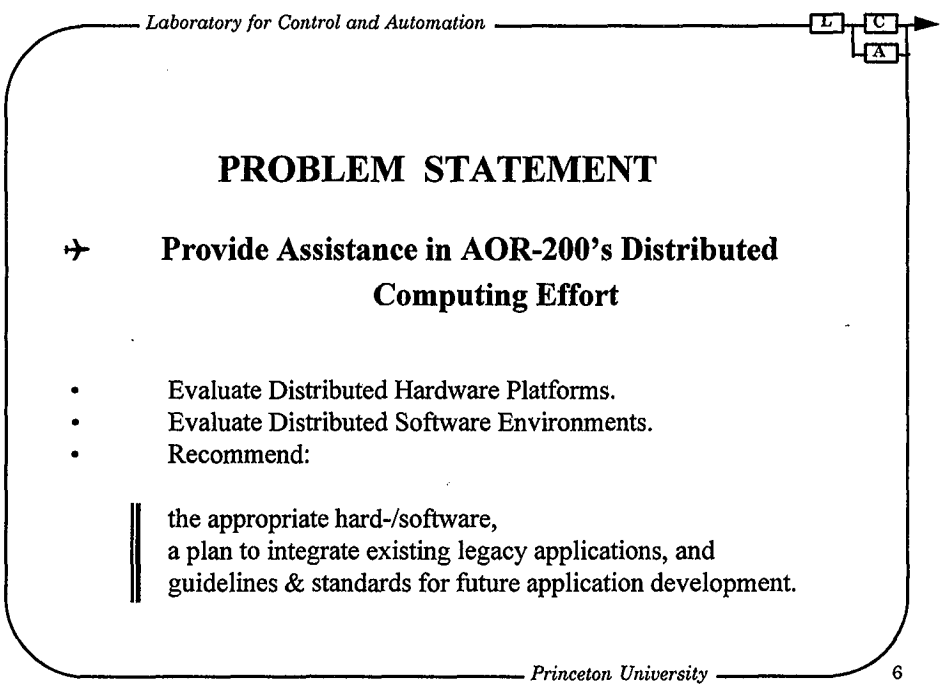
Because software (logical) and hardware (physical) architectures are separate, some machines are hybrids, i.e. they offer both shared and distributed memory at the software level. These hybrid machines are the most versatile and allow for both efficient task and data parallel algorithms. The hardware selected for this study includes the KSR1, the IBM POWERparallel SP1/2 and the Princeton University research computer called SHRIMP (Shared Really Inexpensive Multi-Processor).



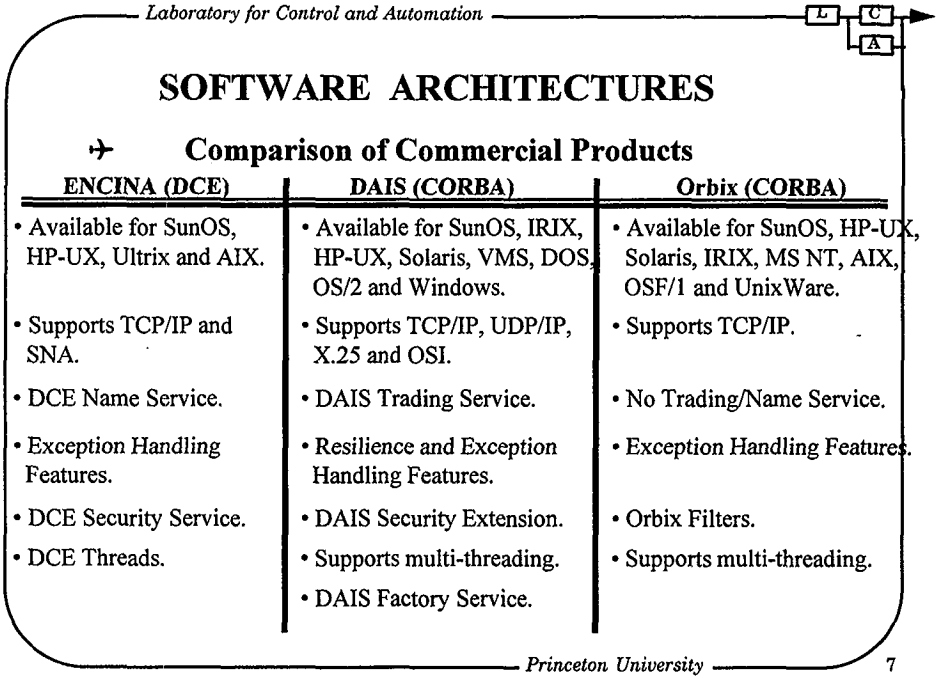
The SRS algorithm can be partitioned into two main phases. In the first phase, preliminary computations for each chromosome, or member, in the current GA generation are performed. During the second phase, a number of Monte Carlo evaluations are executed for each chromosome to determine the fitness. Computational tasks within each phase can be performed in parallel. For example, each Monte Carlo evaluation is data-independent from all other Monte Carlo evaluations in the current GA generation. Load imbalances occur if some processors are idle at some time during execution. Because the execution time for each Monte Carlo evaluation is stochastic, a dynamic scheduler is required during that phase of the algorithm. The scheduler will distribute work load such that all processors are used evenly. The algorithm can so far be classified as being task-parallel with the indicated two synchronization points.



A complexity analysis of the Monte Carlo Simulation phase of the SRS algorithm yields a theoretical "speed-up" curve. N is the number of Monte Carlo evaluations per chromosome, L is the number of processors used. Other parameters enter into the analysis, such as the probability of metric violation and the complexity and logical flow structure of an individual Monte Carlo evaluation. The ideal (100% load balancing, no communications) graph would be a 45-degree line (i.e. speed-up equals L). If stochastic load balancing techniques are neglected, the efficiency (theoretical speed-up divided by ideal speed-up) is less than 100%. In this particular example, the theoretical efficiency was considerably less than 100%. The theoretical results were validated using empirical results from the KSR1.



Research during the summer of 1994 at FAA headquaters in Washington, DC, addressed the Operation Research (AOR) Service's need for a distributed computing environment. Different commercially available software and hardware platforms for client/server and parallel computing were evaluated. The focus was on IBM and KSR for hardware and OSF/DCE and OMG/CORBA compliant products for software.



A comparison of hardware platforms established the IBM SP2 as the most economical commercially available option for its large software base and best price-performance ratio. The software architecture comparison was between the Open Software Foundation's Distributed Computing Environment (OSF/DCE) and the Object Management Group's Common Object Request Broker Architecture (OMG/CORBA). The table provides a detailed comparison between three commercially available products, one DCE-based and the other two CORBA-based. Overall, OMG/CORBA satisfies AOR's criteria better. It ideally allows for rapid prototyping and offers an object-oriented design.

The recommendations ask for testing ICL's DAIS (CORBA), Iona's Orbix (CORBA). A pilot coding project involving a couple of AOR's applications should be undertaken in the next few months.

Intelligent Aircraft/Airspace Systems

John P. Wangermann
Princeton University,
Princeton, New Jersey

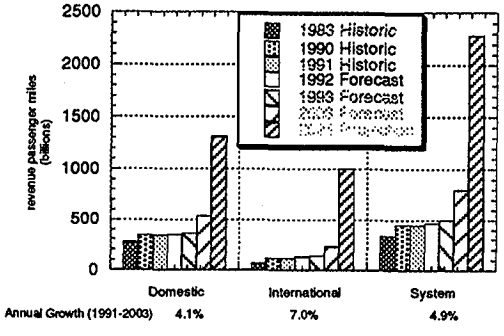
Projections of future air traffic predict at least a doubling of the number of revenue passenger miles flown by the year 2025. To meet this demand, an Intelligent Aircraft/Airspace System (IAAS) has been proposed. The IAAS operates on the basis of *principled negotiation* between *intelligent agents*.

Studies of the future air traffic continue to predict significant growth rates. The Federal Aviation Administration's 1992 report predicted an annual growth rate of 4.9% for the US aviation system to the year 2003. If this growth rate continues to the year 2025 the traffic will be more than three times today's levels.

Laboratory for Control and Automation

MOTIVATION

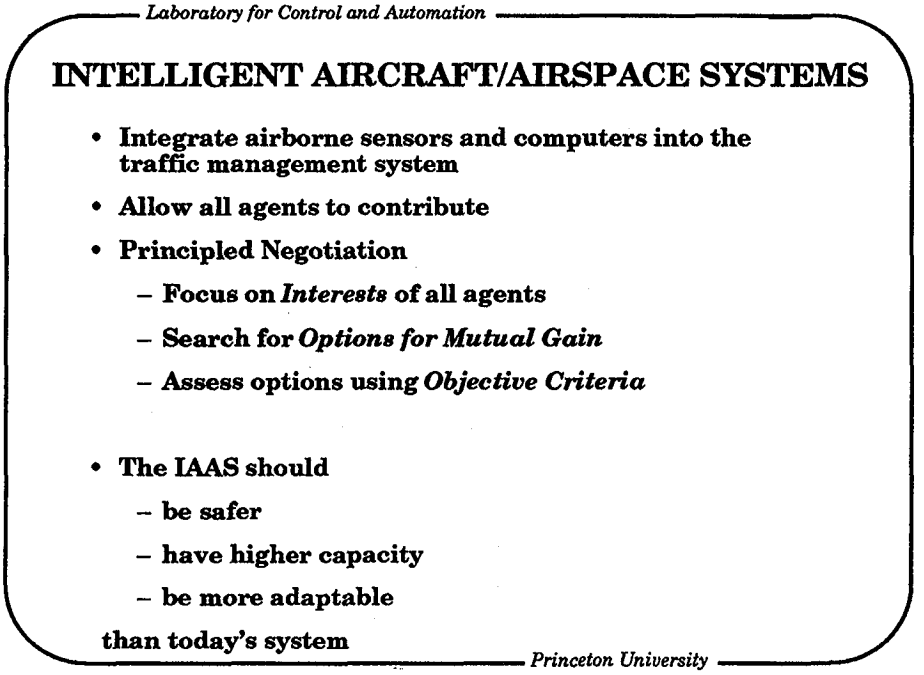
Growth In Revenue Passenger Miles



Princeton University

The aircraft/airspace system today consists of many *agents*, such as airlines, control facilities, and aircraft. All the agents are becoming increasingly capable as technology develops. These capabilities should be exploited to create an *Intelligent Aircraft/Airspace System* (IAAS) that would meet the predicted traffic levels of 2005.

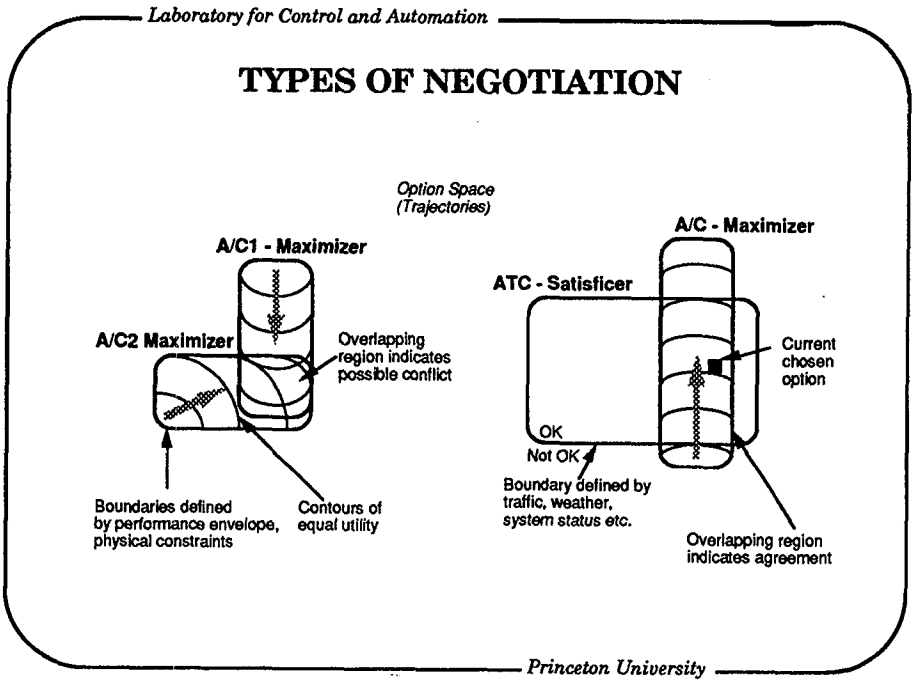
All agents could contribute to the IAAS through *principled negotiation*, a method that was first proposed to produce agreements that were beneficial to all parties in negotiation. In principled negotiation an agent would *focus on the interests* of all agents and could propose options for *mutual gain*. These options would be assessed using *objective criteria*. In this way any agent can propose changes to all aspects of a flight (departure runway, routing, altitude profile, etc.) that could increase airspace capacity. The options found by several agents with distinct capabilities are likely to be better than those found by a single agent. By applying these options the IAAS should have higher capacity, better safety levels, and more adaptability than today's system.



The agents in the IAAS can be classified according to their negotiating behavior. Agents such as aircraft and operators generally attempt to *maximize* some utility function. This function could reflect profit, fuel usage, deviation from planned arrival time, or a combination of these and other factors. Traffic management agents (TrMAs) exhibit *satisficing* behavior. They will accept a flight trajectory provided it meets certain parameters, such as minimum separation from other aircraft.

A maximizer will choose options from a set of *feasible* options (those options that the agents are capable of performing). If two maximizers are in negotiation, any option proposed must be in the feasible set of both agents forthe agents to be able to reach agreement. The option should also have a higher utility for both agents compared to the present plan.

If a maximizer is negotiating with a satisficer, it must propose options that lie within the satisficer's *satisfactory set* (those options that would meet the satisficer's criteria for acceptance). The agreements reached will be influenced by the utility function of the agent (shown as contours for the maximizers).



Maximizers will use many different factors in their assessment of an option. This behavior can be represented using *Multi-Attribute Utility Theory*. The values of the different factors are combined through a weighting matrix to calculate the utility of the option. The components of the weighting matrix are varied according to the state of the agent. For example, an aircraft that departed late may place higher weight on minimizing flight time, while an on-time aircraft weight fuel usage more heavily.

Laboratory for Control and Automation

MULTI-ATTRIBUTE UTILITY THEORY

- **Agents choose from a set of options $O=\{o_1,o_2, ..., o_n\}$**
- **Options assessed against attributes $A=\{a_1,a_2,.....,a_m\}$**
- **Value of each attribute calculated for each option.**
  - **Values for o_i are $\{a_1(o_i),a_2(o_i),.....,a_m(o_i)\}=A(o_i)$**
- **Assign weights to each attribute $W=\{w_1,w_2,.....,w_m\}$**
  - ***Weights are dynamic; they change according to the identified scenario and goals of the agent***
  - **Heuristic rules (expert system) set weights**
- **Calculate the utility of each option**

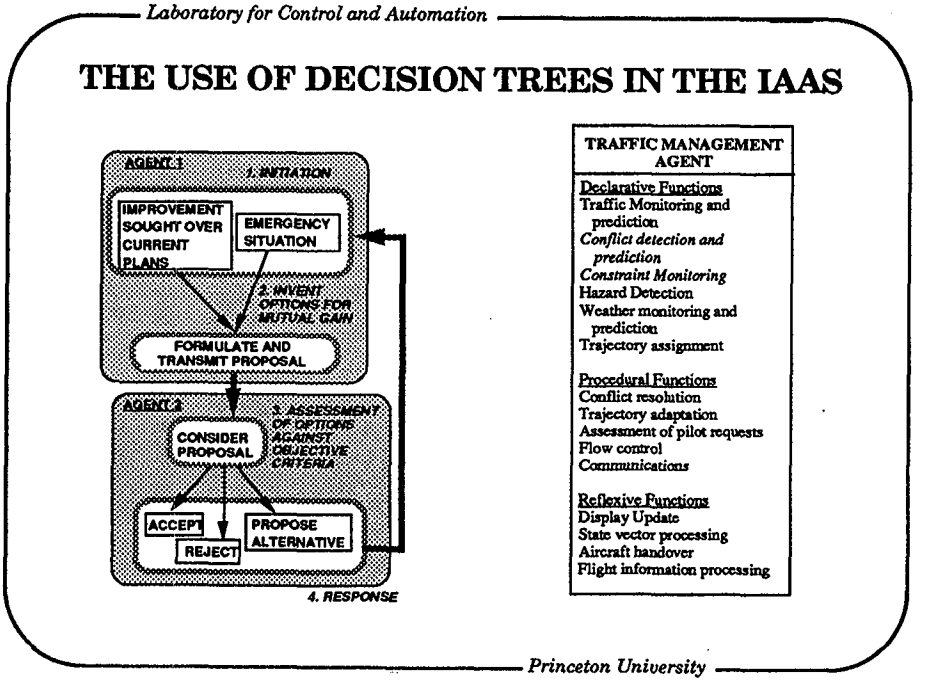
  $U(o_i)=A(o_i).W$
- **Choose the option with the maximum utility**

Princeton University

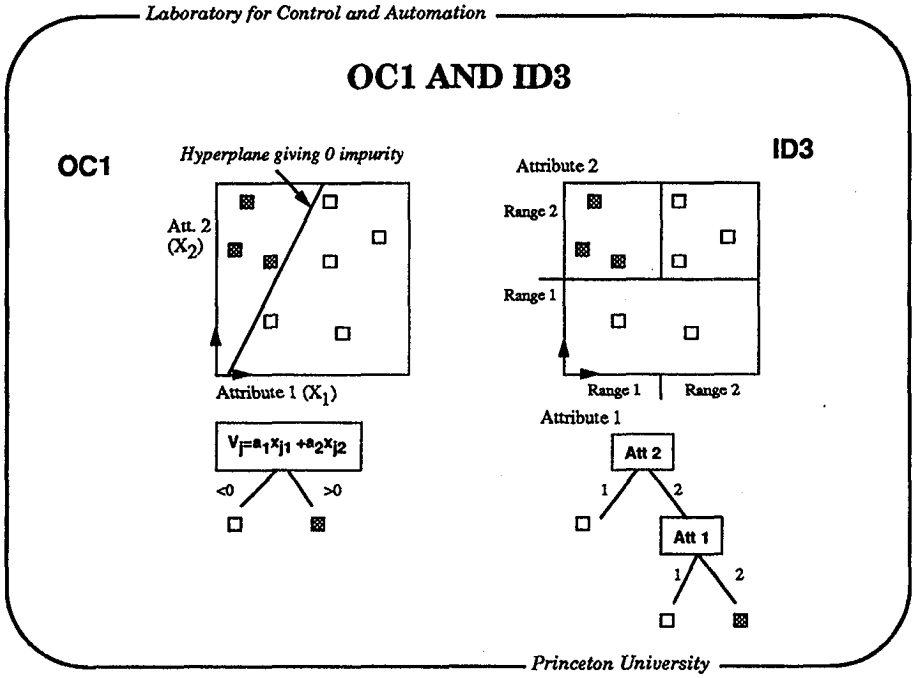
In the IAAS, TrMAs will still have to approve flights through controlled airspace. TrMAs are therefore likely to receive many proposals from both aircraft and operators for changes to all aspects of an aircraft's flight. TrMAs must also rapidly detect conflicts and other hazardous situations, which will impact negotiation. To meet these needs, a TrMA must be able to quickly assess present and future traffic situations and make decisions accordingly.

This type of situation can be tackled using *decision trees.* A decision tree describes a sequence of tests of various factors of a situation. The outcome of these tests determines the decision taken. The tree is constructed from a training set of examples and desired decisions. Decision trees are well suited to classification problems.

When a TrMA assesses a traffic situation it has to classify the situation as safe or unsafe. If a test set consisting of safe and unsafe situations and their attributes is supplied,a decision tree for a TrMA could be constructed.



Decision trees can be constructed using a number of different algorithms. Algorithms such as ID3 and C4.5 are based on univariate tests, i.e. only one attribute is tested at each node of the tree. Other algorithms such as OC1 have been developed that allow multivariate testing at each node. All the algorithms aim to produce trees that have the least number of nodes (tests) needed to correctly classify the set of test examples.



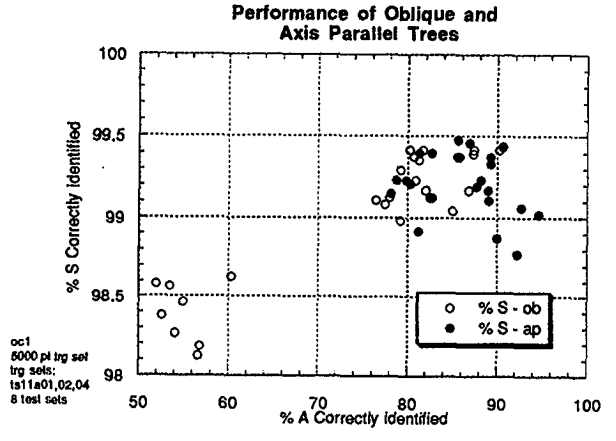
OC1 and ID3 were used to create decision trees that classified two-aircraft scenarios. The aircraft had random positions within a volume of airspace and velocities (within certain bounds). They could execute one maneuver (turn, climb, accelerate). The starting time, duration, and maneuver rate were all randomly distributed.

The decision-maker (the TrMA) had perfect information on the velocity, position, and maneuver of each aircraft, and an estimate of maneuver start time. It did not have information on the maneuver rate or duration. The trees were built using training sets of 5000 examples, of which 300 were class A, the rest class S. The trees were tested on test sets of 200 scenarios.

OC1 was significantly easier to use than ID3, as ID3 requires the user to categorize the attribute values of each scenario. OC1 can use the raw attribute values. OC1 can also produce trees that are either axis parallel (univariate) or oblique (multi-variate). Axis parallel trees were found to produce higher classification accuracies (see slide). The classification accuracy for class S cases was above 99% and betwen 75% and 90% for class A scenarios.

Laboratory for Control and Automation

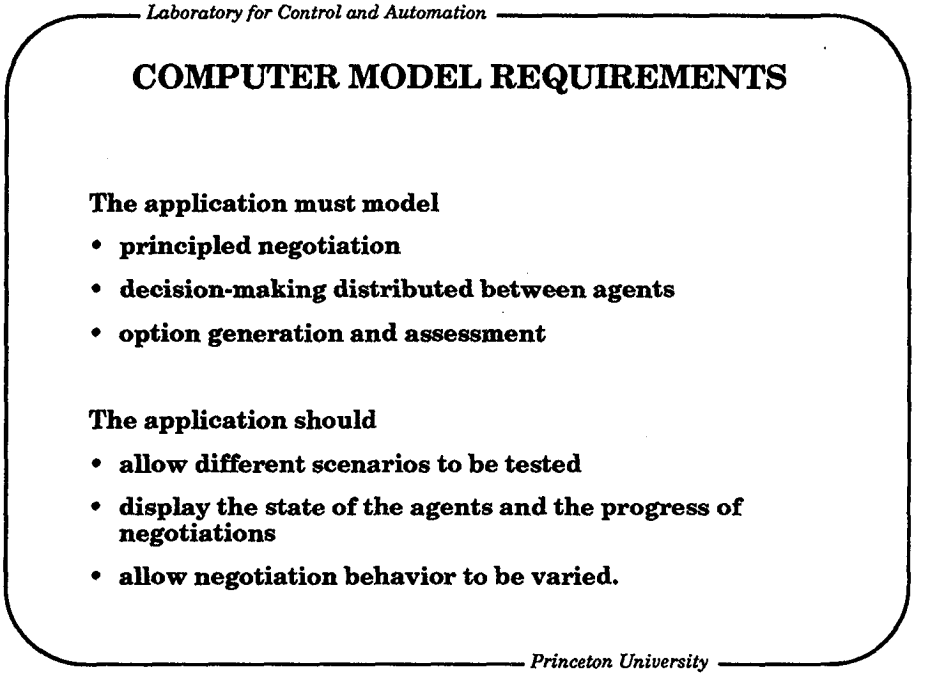
OBLIQUE vs. AXIS PARALLEL TREES



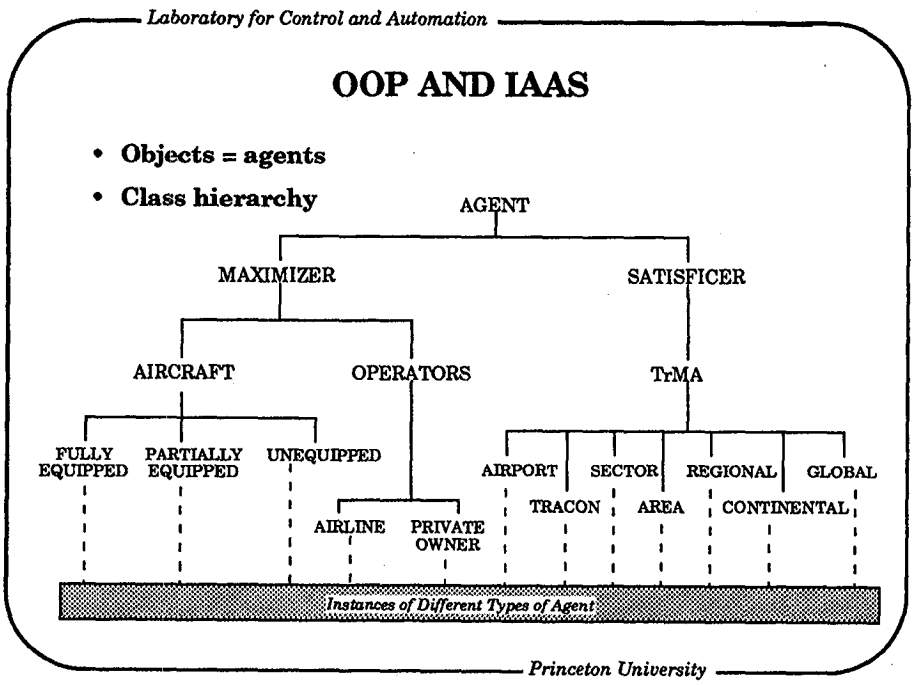
Princeton University

Experimental work to date has investigated different aspects of the IAAS, such as the A* algorithm, decision trees, etc. In order to test the IAAS concept more fully, a computer simulation is needed which represents accurately essential features such as principled negotiation between agents, invention of options for mutual gain, and assessment of options.

The simulation is being constructed using CLIPS, an expert system shell that provides object-oriented, procedural, and rule-based programming facilities. (CLIPS was developed at NASA JSC. It is written in C to provide portability between platforms and ease of interface with C and other languages.)

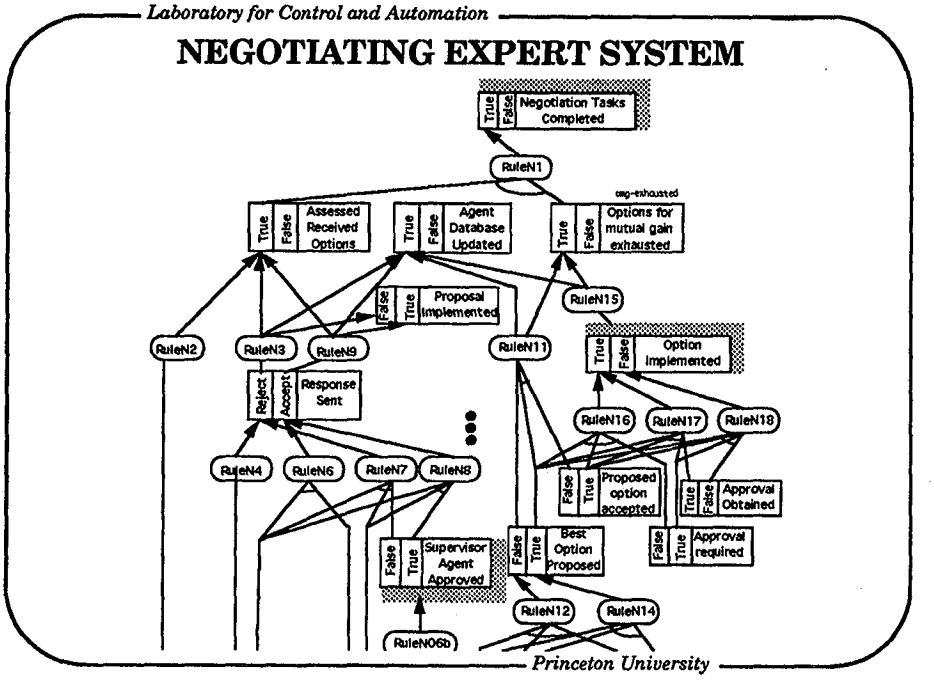


Principled negotiation between agents can be naturally represented as message passing between objects in an object-oriented program (OOP) environment. The class hierarchy for IAAS agents can be based on their type of negotiating behavior (see slide). An agent that is an instance of a fully-equipped aircraft inherits the same basic negotiating behavior as other maximizers such as an airline.



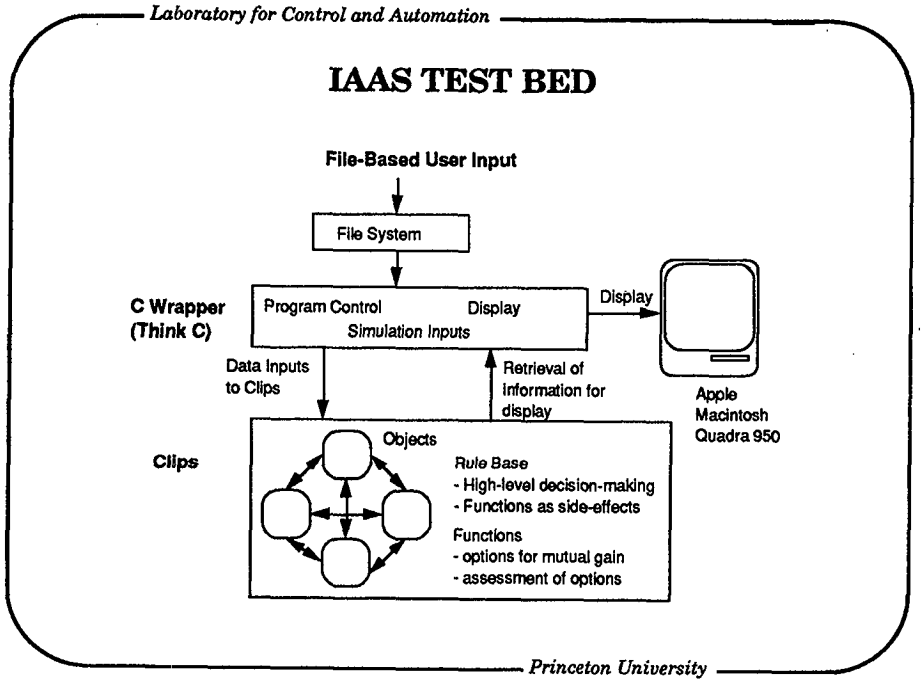
The declarative functions of an agents are well represented by an expert system. Procedural functions, such as the generation of options for mutual gain, are executed as a side-effect of the firing of rules in the expert system. A portion of the rule-base that governs negotiating behavior is shown in the slide. This rule base is used by both satisficers and maximizers, but the procedural functions called may be very different.

CLIPS allows rule bases to pattern match on the slot values of objects as well as facts. This enables an efficient program to be developed where a common rule-base can be used for all the agents without the need to program separate rule bases for each agent. The rule bases can then execute different procedural functions

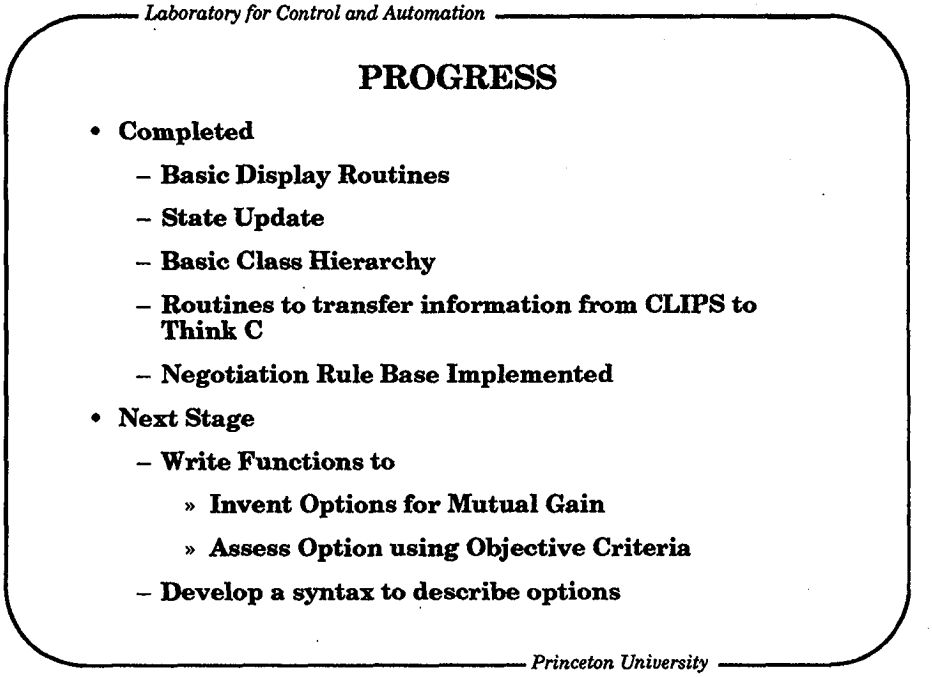


CLIPS is being embedded in a Think C program running on a Macintosh Quadra 950. The C wrapper produces the graphical display and controls the program execution. At this stage a file-based system will be used for user input, but a menu-based graphical user interface could be built at a later stage.

Another benefit of CLIPS is that a function written in C can be embedded into CLIPS. This may be required for more complex procedural functions.



The negotiation expert system, class hierarchy, state update routines, C-CLIPS transfer routines, and display routines have been implemented. The next phase of work will define a syntax for describing options, and procedural functions to invent and assess options will be implemented.



REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. AGENCY USE ONLY *(Leave blank)*	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	August 1995	Conference Publication

4. TITLE AND SUBTITLE
FAA/NASA Joint University Program for Air Transportation Research 1993–1994

5. FUNDING NUMBERS
505-64-52-01

6. AUTHOR(S)
Richard M. Hueschen, Compiler

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
NASA Langley Research Center
Hampton, VA 23681-0001

8. PERFORMING ORGANIZATION REPORT NUMBER
L-17507

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
Federal Aviation Administration
Washington, DC 20546

National Aeronautics and Space Administration
Washington, DC 20546-0001

10. SPONSORING/MONITORING AGENCY REPORT NUMBER
NASA CP-3305

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION/AVAILABILITY STATEMENT
Unclassified–Unlimited
Subject Category 01

12b. DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200 words)*

This report summarizes the research conducted during the academic year 1993–1994 under the NASA/FAA sponsored Joint University Program for Air Transportation Research. The year end review was held at Ohio University, Athens, Ohio, July 14–15, 1994. The Joint University Program is a coordinated set of three grants sponsored by NASA Langley Research Center and the Federal Aviation Administration, one each with the Massachusetts Institute of Technology (NGL-22-009-640), Ohio University (NGR-36-009-017), and Princeton University (NGL-31-001-252). Completed works, status reports, and annotated bibliographies are presented for research topics which include navigation, guidance and control theory and practice, aircraft performance, human factors, and expert systems concepts applied to aircraft and airport operations. An overview of the year's activities for each university is also presented.

14. SUBJECT TERMS	15. NUMBER OF PAGES
Aircraft guidance and control; Avionics; Aircraft navigation; Human factors; Air traffic management	119
	16. PRICE CODE A06

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
Unclassified	Unclassified	Unclassified	

DO NOT REMOVE SLIP FROM MATERIAL

Delete your name from this slip when returning material to the library.

NAME	DATE	MS
Paul Hyer	8/98	473

NASA Langley (Rev. Dec. 1991) RIAD N-75