SAPIENZA
UNIVERSITÀ DI ROMA

Facoltà di Ingegneria

XIX ciclo di Dottorato di Ricerca in Ingegneria Elettrica
Dipartimento di Ingegneria Elettrica

Safety-Oriented Discrete Event Model for Airport A-SMGCS Reliability Assessment

Tesi di Dottorato

PhD Student:	Ing.	**Bernardino Ricci**
Tutor:	Ch.mo Prof. Ing.	**Umberto Grasselli**
Coordinator:	Ch.mo Prof. Ing.	**Carlo Mazzetti di Pietralata**

Anno Accademico 2006-2007

If there is an amateur reader still left in the world – or anybody who just reads and runs – I ask him or her, with untellable affection and gratitude, to split the dedication of this Thesis three ways with my Family, my Friends and my Tutor.

RINGRAZIAMENTI

PROF. UMBERTO GRASSELLI, PROF. LUCA PODESTÀ

PAPÀ, MAMMA, ELENA E ALESSANDRA, AIRALDO ED EDOARDO, FABIO
ANDREA G., CIRO, VITTORIO, CRISTIAN, MARCO, ALESSIO N.
FEDERICO, FIAMMA, MATILDE, BRA, TOES, MONICA, GIANFILIPPO
RAFAL ZBIKOWSKI, ANTONIOS TSOURDOS, STUART URBAN
IRENE, MARCELLO, HEINZ, GIULIO, FEDERICA
ANDREA Z., LILLO, ANTONIO, ANTONELLA, ALFREDO E L'ING. DI NARDO

E TUTTI COLORO CHE HANNO VOLUTO CREDERE INSIEME A ME A QUESTA AVVENTURA

INDEX

INDEX

ACRONYMS

EXPLANATION OF TERMS

CHAPTER 1 – INTRODUCTION

CHAPTER 2 – AIR TRAFFIC CONTROL

CHAPTER 3 – FUNDAMENTALS OF NAVIGATION SYSTEMS

CHAPTER 4 – AIRPORT OPERATIONS

CHAPTER 5 – WEATHER ADVERSELY AFFECTING AIRPORT OPERATIONS

CHAPTER 6 – A-SMGCS

CHAPTER 7 – RANDOM NUMBERS GENERATION (RNG)

CHAPTER 8 – MONTE CARLO SIMULATION

CHAPTER 9 – PETRI NETS

CHAPTER 10 – SIMULATING COLLISION PROBABILITIES OF LANDING AIRPLANES

CHAPTER 11 – AIRPORT TRAFFIC FLOW USING TIMED COLOURED PETRI NETS

CHAPTER 12 – CONCLUSIONS AND FUTURE WORK

REFERENCES

CHAPTER 1

INTRODUCTION

1 CURRENT AIR TRAFFIC SITUATION PICTURE

The Italian National Airspace System is predominately a hub-and-spoke network, with only two main hub airports (Milan Malpensa and Rome Fiumicino) and lots (at least 15) of "minor" airports with an overall capacity of about 2 million operations per year. Current forecasts predict that demand for these airports will soon exceed capacity (e.g., ATAG 2000, Donohue and Shaver 2000, EUROCONTROL 2006-2007).

Closely coupled with the issue of *Capacity* is the issue of *Safety*. Flying airplanes closer together has the potential to increase the likelihood of an accident. Thus, any increase in capacity must be accompanied by a demonstration that such an increase will be safe. As capacity increases, the relative safety must decrease to achieve the same absolute accident rate over time. In fact, the commercial accident rate has remained relatively stable over the last two decades (Barnett and Higgins 1989, Machol 1995), but the absolute number of accidents has increased due to more operations.

In this thesis, we will try to evaluate the *Safety* of a proposed system where airplanes land at high volume at small airports with small amount of control from the control tower, but with presence of state of the art System Technologies on A-SMGCS (to be read 'smigs'), acronym standing for Advanced-Surface Movement Guidance and Control System, i.e. the ensemble of all the radio-electric equipment installed on-board and devices on the ground in order to perform Air Traffic Control over an Airport.

Evaluating the safety of a new system is difficult. First, historical data may not directly apply to the new system, so it is generally not possible to extrapolate existing data to get a safety estimate of the new system. Second, since events like collisions are so rare (Van Es 2001 estimates the rate of plane-to-plane collisions for commercial aviation between 1980 and 1999 to be about 0.3 collisions per million flights), simply observing the new system does not provide a safety estimate. Finally, new systems have new safety hazards that may have not existed before. Thus, a critical part of safety analysis is understanding what these hazards are, how they interact with and influence pilot/controller behaviour, airplane trajectories, etc., and how to numerically quantify their impact on the *Total System Level Of Safety*. The results will illustrate that this new methodology supports also safety-based ATM design.

By now, *Safety* is recognised [ATC-1] as a key quality on which to select/design advanced ATM concepts, even when capacity and efficiency are the drivers of the development. The Safety target is often described as 'equal or better' in comparison with existing practice, allowing a large freedom in how safety is expressed, let alone measured. In effect, new CNS (Communications Navigation and

Surveillance)/ATM concept developments are typically accomplished without the use of feedback from appropriate *Safety Assessments*.

ATM concept design teams (e.g. of Free Flight, or 4D-ATM) try to realise capacity-efficiency enhancements by exploiting new technology, changing human controller roles and introducing new procedures, while relying on the established safety-related indicators in ATM such as conflict rates and types, workload of human operators and failure rates and effects of technical systems.

ATM, however, is the result of complex interactions between multiple human operators, procedures and technical systems, all highly distributed. This yields that providing safety is more than making sure that each of the ATM elements functions properly safe; it is the complex interaction between them that determines safety. The assessment of isolated indicators falls short in covering the complex interactions between procedures, human operators and technical systems in safety-critical non-nominal situations. In order to improve this situation, this thesis outlines a novel probabilistic risk assessment methodology which has specifically been developed for application to Airport Traffic Control.

The demand for air transport has been increasing rapidly over the years and this trend is expected to continue in the future. Therefore, it is of great practical importance to develop, verify and validate simple, yet powerful dynamic models that help in design and retrofit air traffic management systems (ATM). ATM can be divided into two parts: air traffic flow management (ATFM), and air traffic services (ATS) including air traffic control (ATC) where the subject of this thesis belongs to.

One of the recent problems in ATC is to minimize the delays of arriving and departing flights, where the dynamic modelling of runways and scheduling strategies based thereon are the critical issues. There are different approaches to this problem: for instance, decision support systems based on a stochastic analytic model for the estimation of the capacity envelope of an airports runway system. Besides the analytical and mathematical programming techniques applied to ATM problems, combinatorial methods for scheduling various operations on airfields are also used. It has recently been accepted that the relatively recent formalization of concepts of discrete event systems (DES) may well lead to more powerful analysis techniques suited to ATC analysis.

Coloured Petri nets (CPNs) belong to the area of discrete event system methodology, where CPNs are well known for their capability in simulating and analyzing discrete event systems. There are useful extensions of the original CPN concept that includes its timed and stochastic versions. Dynamically Coloured Petri Nets (DCPN) have already been used in ATM but applications are strictly limited to safety purposes. The same applies for DCPN representation of runways.

The aim of this thesis is to propose a simple dynamic model that describes the traffic flow of a single runway Airport due to schedule (i.e. estimated times of arrivals and departures). The model is governed by elementary ATC principles (one aircraft at a time on RWY, arrivals priority on departures), and it is a subject of random disturbances (such as weather conditions, false alarms, low probability of detection, integrity over communications, etc.). Therefore, Timed CPNs have been selected as modelling and simulation tools to develop and analyze a dynamic model of a small Airport.

Current crises in the air traffic industry demonstrate that changes are required to present systems. The levels of delay and poor safety standards being experienced around the world dictate the need for improved Air Navigation Services (ANS). A multitude of reasons explain these problems. For instance, fragmentation of national systems prevents optimum use of the world's airspace. In addition, inherent limitations of present ANS technologies and procedures mean that it is usually not possible to separate numerous aircraft on random routings. Thus, aircraft must often plan their flights along routes and be channelled so that the necessary separation can be maintained. This results in fuel and time penalties, in addition to airspace capacity being consequently constrained. Accordingly, this thesis provides an analysis of changes that should occur to ANS during this decade by evaluating current and planned technologies and procedures.

2 CNS/ATM SYSTEMS

As already stated, air traffic congestion problems in many areas of the world are well known and have been highly publicised in recent years. This airspace dilemma, which results in delays and other undesirable knock-on effects, is escalating at a phenomenal rate and requires immediate attention. Correspondingly, there is concern about safety standards in some worldwide airspace regions. It should be noted that congestion and poor safety levels affect all involved with aviation and, indeed, society in general. It is evident, therefore, that present methods are inefficient and unable to guarantee punctual or safe services now or in the future. It is imperative that the significant projected growth in air transport movements during this decade is accommodated. Thus, there is an urgent need to solve the current airspace problems and plan to meet forecast demand in a responsible manner. Solutions to these predicaments have been developed and are encompassed under the auspices of the terms, 'future air navigation systems' or 'CNS/ATM systems'.

Future air navigation systems use Communications, Navigation & Surveillance (CNS) technologies to provide enhanced Air Traffic Management (ATM) through continuous information on aircraft positions and intentions so that reductions in separation are possible without compromising safety. The systems include technologies and procedures that merge to optimise the potential of airport and airspace resources so that the capacity, flexibility and safety of these resources are maximised, while delays and their operating costs are minimised.

However, there has been a significant lack of progress to date in implementing the new technologies and procedures. Indeed, the assured integration of many systems is currently uncertain and far from guaranteed. Therefore, we will try to summarize the evolution of Air Navigation Services (ANS) during this decade. Noting that the present systems will still fulfil many ANS roles in 2010, we will consider both current and future CNS/ATM, split into the Communications, Navigation & Surveillance (CNS) technologies and Air Traffic Management (ATM) procedures.

2.1 PRESENT CNS/ATM TECHNOLOGIES AND PROCEDURES

Although current ANS is based on CNS technologies and ATM procedures that were developed many years ago, their applications have been continually modified to suit the different types of airspace that exist around the world in an attempt to satisfy the various capacity requirements and specific operating environments [ATC-1]. However, the levels of delay now being experienced in many airspace sectors, and the lack of adequate safety standards in other regions, have necessitated further development of these systems. Thus, the present systems form the basis of many planned technologies and procedures. Indeed, improvements in the efficiency, flexibility and safety of the overall ANS will only occur if present systems evolve into more responsive elements. Accordingly, it should be noted that, as most new CNS/ATM systems will be phased in using an evolutionary approach, present technologies will continue to be employed in many situations at the end of this decade. In addition, a description of present CNS and ATM serves to identify the functional shortcomings of present systems and highlight the potential benefits of implementing new technologies. Therefore, apart from facilitating a contrasting analysis of ANS evolution, there is a need to describe present CNS/ATM methods.

Current Communications Technology. Given that the main objective of an aeronautical communication service is to ensure that telecommunications and radio aids necessary for the safety, regularity and efficiency of air navigation are continuously available and reliable, two categories of Air Traffic Control (ATC) communication are presently conducted.

Air-ground. Very High Frequency (VHF) transceivers are used to provide voice contact between ATC and pilots when within line-of-sight coverage, which is invariably near or over land in dense traffic areas. Due to the nature of VHF signal propagation along the curvature of the Earth, it can be received and transmitted over greater distances from higher altitudes. However, the availability of VHF spectrum channels in the long-term is a concern. Europe recently implemented a narrowing of bandwidth to 8±33 KHz, which will soon be necessary in other parts of the world. High Frequency (HF) is used over areas exceeding VHF's range, such as in oceanic or remote continental areas. The fact that HF requires its waves to be reflected from the ionospheric layers above the earth is often a drawback because destructive interference can occur, where the transmitted signals arrive at the receiver at different times because they were reflected at different ionospheric levels and consequently sent via different routes. Additionally, there is limited availability of frequencies that work reasonably on a given day. The need for pilots constantly to monitor their assigned HF frequency is removed if SELect CALling (SELCAL) is fitted on the aircraft. Since the early 1990s, a development of VHF technology, VHF DataLink (VDL), has enabled electronic messages such as ATC clearances to be sent from the ground to the cockpit, where they arrive in digital format. HF DataLink (HFDL) is also being used, albeit in fewer locations. Datalink applications are similar to Aircraft Communications, Addressing and Reporting Systems (ACARS), which are operated by four ground network providers, namely Air Canada in Canada, AVICOM in Japan, in addition to ARINC and SITA on a worldwide scale. Correspondingly, it should be noted that Inmarsat is an international organisation that provides satellite communications for the aviation, maritime and land shipping industries. ARINC and SITA use Inmarsat for the satellite portion of their services.

Ground-ground. Adjacent Air Traffic Service (ATS) units are linked by dedicated telephone lines and telex systems between controllers and control centres. Many airlines are linked with the ATS units via the Aeronautical Fixed Telecommunications Network (AFTN).

Current Navigation Technology. Navigation presently uses different aids in the following two categories of airspace types.

Near or Over Land with Dense Traffic. The Flight Management Computer (FMC) on board the aircraft, part of its Flight Management System (FMS), employs lateral navigation concepts such as aRea NAVigation (RNAV) using Non-Directional Beacons (NDB), VHF Omni-directional radio Range (VOR) and Distance Measuring Equipment (DME). These are ground-based navigational aids, many of which have been in use for decades. For instance, Basic-RNAV (B-RNAV) is in place over Europe, whereby flights must maintain a track keeping performance of within 5 nm of the centre-line trajectory for 95% of the time. Barometric altimetry is also employed, to provide vertical navigational guidance and separation. The terminal environment uses Instrument Landing Systems (ILS). A Microwave Landing System (MLS) has been operational for many years, but has had only limited introduction, even though it was previously envisaged that MLS would have replaced ILS at busy airports by 1998-2000. Reasons for the lack of comprehensive implementation to date include the continued international standardisation of ILS as the definitive landing system and the consequently poorly perceived cost-benefit scenarios particularly as landing systems based on satellite navigation are now being developed.

In Oceanic and Remote Areas. Long Range Navigation Systems (LRNS) are used, which include LORAN C and the self-contained Inertial Navigation System (INS). INS positioning errors increase with time and consequently limit the system's accuracy. Although Airborne Collision Avoidance Systems (ACAS) are primarily surveillance facilities, they are increasingly being developed for navigational purposes such as in-trail climbs and descents. The concept of Minimum Navigation Performance Specification Airspace (MNPSA), which requires that all flights in a designated region achieve high standards of navigational performance accuracy, is a method of navigational assurance frequently employed in oceanic and remote areas. Some carriers have started to introduce the Global Positioning System (GPS), the US constellation of navigation satellites, as a source of navigation reference in all airspace types. GPS is part of the Global Navigation Satellite System (GNSS), as envisaged by the International Civil Aviation Organisation (ICAO). Because of concerns over integrity, GPS has not yet been accepted as a sole means of navigation, but it can be employed virtually anywhere on the globe. However, GPS has been approved as a primary means of navigation in a number of airspace regions around the world. It should be noted that Inmarsat has placed a set of navigational transponders on their Inmarsat-3 satellites, although use to date has only been for testing purposes.

Current Surveillance Technology. Surveillance is the basic tool for the controller to monitor the maintenance of safe separation, manage the airspace efficiently and to assist pilots in navigating their aircraft safely. The following surveillance techniques are presently employed in the respective categories of airspace.

Near or Over Land with Dense Traffic. The main means of surveillance by ATC in such areas is the use of radar, which is independent of the pilots' cooperation. This is usually Secondary Surveillance Radar (SSR), which requires aircraft to be fitted with a (Mode A or C) transponder that automatically transmits information when 'interrogated' by ATC. Primary radar is also still in use. Airborne Collision Avoidance Systems (ACAS), such as Traffic alert and Collision Avoidance Systems (TCAS), are technologies that enable aircraft to avoid each other in the air. They are not dependent on any ground-based system: the equipment interrogates SSR transponders of other aircraft in its vicinity, analyses the replies by computer to see which aircraft represent potential collision hazards and provides appropriate advisory information to the flight crew. TCAS has been mandatory in US airspace for many years and became requisite for aircraft with Maximum Take-Off Weights (MTOW) greater than 15 tonnes flying in Europe at the beginning of 2000. Other regions around the world are being actively encouraged to enforce its implementation on flight decks.

In Oceanic and Remote Areas. Surveillance methods in these regions, where radar coverage is not available, are usually dependent on co-operation from pilots. Aircraft positioning is determined on board and is transmitted to ATC using VHF and/or HF radio contact, invariably through procedural voice position reporting, particularly in oceanic areas. This necessitates large separations between aircraft due to the slow nature of the process. Supplemental monitoring is now conducted by ATC using technologies such as Enhanced Traffic Management Systems (ETMS) and Flight Data Processing Systems (FDPS), which constantly work out estimates of aircraft locations and alert controllers of any impending conflicts. However, it should be noted that aircraft report their positions automatically using datalink technologies in some areas. For example, aircraft equipped with 'FANS-1/A' avionics packages, designed to enable aircraft to exploit early benefits from future systems, presently provide automatic reporting in the North Atlantic and Pacific oceanic regions. Additionally, aircraft users are encouraged to place some form of Ground-Proximity Warning System (GPWS) in their avionics for use in all airspace regions. GPWS monitors aircraft instruments and provides an audible warning of proximity to the ground. Enhanced GPWS (EGPWS) uses a world-wide digital terrain database to provide a visual display and advanced warning of threatening terrain to the pilots in a colour code to indicate the level of threat posed. EGPWS is being mandated in Europe and North America.

Current Air Traffic Management Procedures. Current Air Navigation Services (ANS) provide international Air Traffic Services (ATS) based on the availability of CNS technology. ICAO is the agency responsible for safe and orderly ATS operations at a worldwide level. ATS were formed to expedite the safe and orderly flow of air traffic. ATS usually consist of a flight information service, an alerting service and Air Traffic Control (ATC). The ATC service is further divided into aerodrome control at the airport, approach control in the vicinity of the airport and area control for en-route flights. Such services apply to both Instrument Flight Rules (IFR) and Visual Flight Rules (VFR) traffic. ATC control centres employ Air Space Management (ASM) and Air Traffic Flow Management (ATFM) to provide Air Traffic Management (ATM). Ultimately, ATM is conducted through strategic and tactical planning from months in advance, in addition to real time monitoring and control of flights. It should be noted that the world is divided into many Flight Information Regions (FIR) for the purpose of providing ATS. Some only occupy small volumes of airspace because FIR boundaries are usually defined by national borders or by agreed lines of demarcation over water. Air navigation facilities and procedures of each region are referred to in ICAO Air Navigation Plans (ANP). The FIRs can span the boundaries of more than one nation, but each country's Civil Aviation Authority (CAA) puts together its own ATS system. Controlled airspace areas are manifested as TerMinal control Areas (TMA), interconnecting airways and area-type control areas. The latter two are further divided into ATC sectors. Other than the aforementioned types of controlled airspace and ConTRol zones (CTR), which facilitate the separation of aircraft operating within the vicinity of busy airports, the airspace within FIRs is usually uncontrolled. Airspace is categorised into different classes, the exact composition being determined by each nation. The ATM procedures within an FIR are based on the types and densities of air traffic, the CNS technology and infrastructure that is available, in addition to the topography and economic conditions of the country involved. Thus, similar to the applications of the various types of CNS technology in the different categories of airspace, ATM is presently based on procedures and separation standards, which vary with operating environment.

Near or Over Land with Dense Traffic. ATM in these situations uses ATC strips, based on the aircraft's flight plan, which are colour-coded to denote the direction of the aircraft. Strips incorporate information about the flight such as its timings, desired route level and radar code. The controller places the strip on a flight progress board in geographical and time order. After taking the requirements of all other traffic in the sector airspace at that time and in the immediate future, the controller plans a safe Flight Level (FL) and route for the aircraft using an ATS route structure of Upper Airways (UA). The exact altitude of the FL is dependent on the aircraft track, whether between 0° and 179° or between 180° and 359°. Vertical separation between FLs is in the process of being reduced from 2,000 ft to 1,000 ft in some worldwide areas above FL290. The programme is termed Reduced Vertical Separation Minima (RVSM); vertical separation is already at 1,000 ft below FL290. Horizontal separation minima, which are split into longitudinal (along the track) and lateral (across the track) components, vary and are dependent on the availability of ground-based navigational aids and radar surveillance. Horizontal separation can be based on time or distance. ATC presently enables automated functions such as short/medium-term conflict alert, trajectory prediction and flight progress monitoring. Central Flow Management Units (CFMU), such as that operated by Eurocontrol in Brussels and the Central Flow Control Facility by the US FAA in Washington DC, facilitate co-ordination of flights. Efforts such as Eurocontrol's Free Route Airspace Project (FRAP), which offers users direct routes through the upper airspace of eight European nations, are currently adding capacity to ATM systems.

In Oceanic and Remote Areas. A lack of surveillance information due to absence of radar technology means that flights are often controlled using procedural ATC methods in such areas. These methods use very large separation criteria, which result in low system capacity and poor availability of efficient flight profiles. Flight tracks are based on route structures that are not necessarily the desired routings of aircraft. The aircraft are assigned slot entry times to a track and told to fly at a specific speed, to facilitate the application of the Mach Number Technique (MNT). Reduced Vertical Separation Minima (RVSM) were successfully introduced in the North Atlantic in 1997-1998 and are presently being implemented in the Pacific. The latter region facilitates Dynamic Airborne Re-routing Procedures (DARP), whilst the former is planning to implement direct routing and Reduced Horizontal Separation Minima (RHSM).

In the Terminal Area. Standard Instrument Departure (SID) and STandard ARrival (STAR) routes are established at busy aerodromes to ensure that traffic departs and arrives at an airport in a safe, orderly and expeditious flow. SID and STAR routes link with significant points of ATS en-route tracks and are designed to take account of noise abatement procedures. Separation is maintained between aircraft in such a manner that wake vortices from larger, preceding aircraft do not affect the safe flight of other aircraft. Guidance is sometimes given by ATC using radar vectors, but routes are created for navigation using the ground-based VOR/DME radio navigation facilities. At some cities with more than one airport, aerodromes share SIDs, STARs and holding patterns. Many ATM programmes, such as staggered, dual runway approaches, attempt to maximise the approach capacity of slot-restricted airports. Due to the fragmented nature of ATC between countries, different systems and standards have resulted in poor harmonisation between adjacent units. Correspondingly, the lack of common air to ground data interchange systems means that current ATM is still based on a multitude of traditional methods. This is manifested by inflexible fixed route structures, which result in demand exceeding capacity levels, air traffic delays and increased operating costs for users. Coupled with poor ATC infrastructure in many countries, there are many FIRs that are considered unsafe. The future CNS/ATM concepts aim to alleviate such problems by introducing an interoperable, seamless system, with full global coverage for safe ATM.

2.2 FUTURE CNS/ATM TECHNOLOGIES AND PROCEDURES

The key to optimisation of air traffic flow in the various types of airspace is [ATC-1] the introduction of automated Communications, Navigation and Surveillance (CNS) systems that can provide enhanced Air Traffic Management (ATM) with continuous information on aircraft position and intentions. Hence ICAO's insistence that Future Air Navigation Systems (FANS) be known as CNS/ATM, which aims to be a method of using technology to:
(a) Enhance communication links between aircraft and air traffic controllers with computers selecting the optimum method of transmission;
(b) Improve pilots' ability to navigate their aircraft safely; and
(c) Increase air traffic controllers' capacity to monitor and survey flights.
ICAO defined their envisaged CNS/ATM in terms of technologies and procedures of future ATC systems. Indeed, if their plans were adhered to, then ATC systems would evolve as listed below during this decade.

Future Communications Technology. Use of data transmission will introduce many changes in air-ground communication. Future communications systems will allow more direct and efficient linkages between ground and airborne automated systems, thereby offering the possibility of the most routine pilot-controller communication taking place via datalink. Such Controller Pilot DataLink Communications (CPDLC) use displays instead of voice and may be seen as the key to development of new ATM concepts. However, it is not currently expected that CPDLC will be used for urgent messages. It should be noted that CPDLC is already in use in the Pacific oceanic region. The transmission of voice will continue to take place over existing Very High Frequency (VHF) channels, certainly in 2010, with reduced channel spacing to 8±33 KHz in many areas. The continued use of VHF voice communications will be essential for safety-related communications and in high traffic density terminal areas, which both require high integrity and rapid response. Therefore, there is a need to safeguard aviation's current spectrum allocation and not lose frequencies to commercial, mobile telecommunications. Standards are also being developed for a time division multiple access digital radio as the medium-term solution to spectrum congestion and enhanced air-ground services. In addition to a new generation of Inmarsat satellites (Inmarsat-4), Low Earth Orbit (LEO) and Medium Earth Orbit (MEO) satellite constellations, such as Globalstar and New ICO, will provide additional communications facilities. The VHF channels will additionally be used to transmit digital data using enhanced modes of today's VHF DataLink (VDL), such as VDL Modes 2, 3 and 4. VDL is presently employed to send electronic messages from the ground to the cockpit, where they appear in paper format. VDL Mode 2 has replaced Aircraft Communications, Addressing and Reporting System (ACARS) since 2004. VDL Mode 3, which is particularly suited to digital voice communications, is being developed in the US and is expected to be operational before end of 2008. VDL Mode 4, which is being developed in Europe, uses a Self-organising Time Division Multiple Access (STDMA) technique that, in addition to providing its intended data communication functions, makes navigation and surveillance datalink capabilities available. The implementation date for VDL Mode 4 is uncertain. It should be noted that, although ICAO is developing standards for both Mode 3 and Mode 4, there is an industry-wide debate regarding which mode to adopt. In addition, HF DataLink (HFDL) will continue to be employed at the end of this decade.

Satellite data and voice communications, capable of global coverage, will be introduced using Aeronautical Mobile Satellite Services (AMSS). The introduction of satellite-based data and voice communication technology is not subject to the many limitations of today's other communications infrastructure, thereby ensuring greater, global availability and integrity of such services. For example, AMSS may well become the primary method of obtaining weather reports, in lieu of the current Automatic Terminal Information Service (ATIS). It will also be used on a more widespread basis for issuing clearances to aircraft, consequently removing the need for the present, lengthy methods. The benefits of satellite communications currently available should be maximised. Indeed, it is ironic that passengers are sitting in the cabin, talking in a crystal clear manner to their associates on the ground, while the cockpit crew is struggling to understand ATC instructions via poor quality HF communications. It is even more ironic that over 80% of the global wide-body fleet is now equipped with satellite communications technologies.

Secondary Surveillance Radar (SSR) Mode S, which will be used for surveillance in high-density airspace, will also have the capability of transmitting digital communications data. SSR Mode S will provide an air-ground datalink that is specifically suitable for limited data messaging in high-density airspace. Mode S should be integrated by 2010.

The Aeronautical Telecommunications Network (ATN) will provide for the interchange of digital data between end users over dissimilar air-ground and ground-ground communication links by adopting common interface services and protocols based on the International Organisation for Standardisation (ISO) Open Systems Interconnection (OSI) reference model. It may be compared with the Internet. Automated ATM interaction between ground-based and aircraft-located computer systems will be supported using On-Line Data-Interchange (OLDI). Recent trials suggest that the ATN will be available in the next 5 years in Europe and in the US. However, noting that aircraft have been benefiting from the FANS-1/A package and CPDLC technologies in some airspace regions such as the Pacific Ocean, there is an issue regarding the choice of datalink platform as the ATN or FANS-1/A.

Although the availability of several communication systems does provide a degree of flexibility to planning and implementation in the different types of airspace, the proliferation of sub-networks will undoubtedly add to the operational complexity of global communications. Thus, there is a need to translate all relevant operational requirements in a particular airspace into a series of communication performance parameters: the concept of Required Communication Performance (RCP) refers to a set of requirements such as capacity, availability, error-rate and transit delay. By 2010, the RCP will be specified by ICAO for operational scenarios in various airspace environments, indicating that any single communication system, or combination of systems, meeting the set parameters can be considered operationally acceptable.

Future Navigation Technology. Improvements in navigation will include the progressive introduction of aRea NAVigation (RNAV) capabilities and elements of the Global Navigation Satellite System (GNSS). These systems will provide worldwide en-route navigational coverage and, ultimately, precision approaches using a service with high integrity and accuracy. The implementation of these systems will eventually enable aircraft to navigate in all types of airspace around the world. For instance, more so-called 'FANS routes' will proliferate. By 2010, based on the current situation, it is inevitable that ground-based equipment will still be used as navigational aids, although many nations will have dismantled at least a portion of their ground-based infrastructures. Aids such as Distance Measuring Equipment (DME) and VHF Omni-directional radio Range (VOR) will continue to be used for RNAV in a manner similar to today's system or as a back-up to GNSS. In the future CNS/ATM system, however, RNAV will operate with a greater dependency on satellite-based systems to determine aircraft positions. This will facilitate a greater proportion of flights flying on direct routes, especially if RNAV is employed with Dynamic Aircraft Route Planning (DARP). The Flight Management Computer (FMC) will periodically download its 4D trajectory and expected routing to allow for automated negotiation of flight plans. It should be added that the B-RNAV presently in place over Europe has evolved in 2005 to become Precision-RNAV (P-RNAV), where the tolerance is reduced to 1 nm. Indeed, the airspace will be designated as 'RNP-1'. GNSS will be a worldwide position and time determination system that includes one or more satellite constellations, aircraft receivers, ground monitor stations and system integrity monitoring devices. Only three satellite constellations are currently available with navigational capabilities: the US Global Positioning System (GPS); GLONASS, which is provided by the Russian Federation and is similar to GPS; and Inmarsat-3 satellites which, as well as their comprehensive communications capability, are equipped with navigational transponders. In addition, it should be noted that Europe aims to declare its proposed constellation of 'Galileo' satellites operational by 2008. GNSS will be the key feature of the future navigation system and will evolve to be a sole means of navigation with the intent that it eventually replaces current systems. This will occur, in certain cases, by 2010. Integration of GPS and the Inertial Navigation System (INS) may be seen as a step towards full GNSS. To overcome inherent system limitations and to meet the performance accuracy, integrity, availability and continuity requirements for all phases of flight, GNSS will require varying degrees of augmentation to support the actual phase of operation. Such systems, which monitor signal reliability and enhance accuracy to make GNSS suitable for civilian use, are still being developed and may be broadly categorised as space-based (SBAS), aircraft-based (ABAS) and ground-based (GBAS). Based on GPS, Inmarsat and planned Japanese satellite systems, three SBAS should be operational by 2005: the Wide Area Augmentation System (WAAS) in the US, the European Geostationary Navigation Overlay Service (EGNOS) and the Multifunctional transport Satellite Augmentation System (MSAS) in Japan. ABAS either use Receiver Autonomous Integrity Monitoring (RAIM), where satellite signals over and above the four required to provide a three-dimensional fix are used to check the integrity of those used in this positioning solution, or the same integrity function is achieved using other aircraft sensors such as INS or VOR/DME. GBAS are being designed primarily to provide high integrity and accuracy for airfield approach systems as described below. The standard non-visual aids to precision approach and landing that will be used in 2010 include a continuation of Instrument Landing System (ILS) operations as long as they are operationally acceptable. Microwave Landing Systems (MLS) will be implemented in cases where they are operationally required and economically beneficial. Based on current progression of its development, Differential GPS (DGPS) should be available in the near future. DGPS uses GPS signals, augmented on a local basis using a Local Area Augmentation System (LAAS), to guide the aircraft on its approach to landing; these systems are becoming known as Satellite Landing Systems (SLS). Other technologies are being developed, such as a completely airborne system, Autonomous Precision Approach and Landing System (APALS), Head-Up Displays (HUD) and Multi-Mode Receivers (MMR) to enable the flexibility for aircraft to use ILS, MLS or SLS. Similar to the aforementioned RCP concept, Required Navigation Performance (RNP) recognises that aircraft navigation systems are capable of achieving predictable levels of performance accuracy within a defined airspace. RNP types for operations are identified by a single accuracy value, defined as the minimum navigation performance accuracy required within a specified containment level. Noting that success to date with RNP has been better than RCP, the navigational requirements for many airspace environments will be stated in RNP terms by 2010.

Future Surveillance Technology. Over the next 5 to 10 years, it is expected that both Primary and Secondary Surveillance Radar (PSR and SSR) will continue to be used, with the gradual introduction of Mode S (Selective) datalink in both terminal areas and high-density continental airspace started in 2003 in Europe; 2008 in the US. SSR Mode S will be used in high-density airspace regions to provide high accuracy, reliable surveillance capable of providing conflict alerts. Accordingly, more advanced versions of the current Airborne Collision Avoidance Systems (ACAS), such as TCAS 2, has become standard recently in 2005. In addition, TCAS 4 will be developed, which uses satellite-based position data to provide better conflict detection and resolution advisories. The major breakthrough, however, will be the more widespread implementation of Automatic Dependent Surveillance (ADS), which allows aircraft automatically to transmit their position and other data, such as heading, speed and intentions. This will be performed via satellite or some other air-ground communication link to an Air Traffic Control (ATC) unit, where the position of the aircraft will be displayed as on conventional radar screens. ADS is particularly appropriate to oceanic and remote regions. It may be seen as a beneficial merging of communication and navigation technology, which enables oceanic airspace to make use of the Aeronautical Mobile Satellite Services (AMSS). Software is being developed to allow ground computers to use ADS digital data to detect and resolve conflicts through conformance monitoring. Eventually, this could lead to clearances being negotiated between airborne and ground-based computers, with little or no human intervention.
ADS-B (Broadcast) is an application of ADS technology that enables an aircraft to broadcast its position, altitude and vector information for display by other aircraft and also by ground users, such as ATS providers. ADS-B has become available starting from 2003 in the US, and current interest remains high in widely implementing this technology before the end of the decade. ADS-B information can also be used as a basis for a Cockpit Display of Traffic Information (CDTI). In addition to CDTI, other Situational Awareness Systems (SAS) will provide flight crews with wake vortex hazard prediction and avoidance; synthetic vision with HUDs; and enhanced meteorological awareness. On the ground, ADS-B will be implemented with ground movement radar and airport surface detection equipment such as Surface Movement Guidance and Control Systems (SMGCS). Indeed, it should also be noted that parallel precision runway monitors will increase approach traffic capacity at airports with closely spaced parallel runways.

As with the communications and navigational elements of CNS/ATM, although the availability of a plethora of surveillance systems provides planning flexibility, it complicates the harmonisation of surveillance functions. To facilitate the planning, it is necessary to translate the relevant operational requirements into a series of surveillance performance parameters, termed Required Surveillance Performance (RSP). Once ICAO has specified the RSP for an operational scenario for a given airspace, any single system or combination of surveillance systems meeting the set parameters can be considered operationally acceptable. This will occur before 2010.

Future Air Traffic Management Procedures. The objectives to be reached through the envisaged, evolutionary ATM system include:
(a) To meet evolving air traffic demand;
(b) To support a safe and orderly growth of international civil aviation;
(c) To enhance safety, regularity and efficiency;
(d) To optimise benefits through global integration;
(e) To enhance economy of commercial air transport.

Specifically, it is how each country and worldwide region will achieve these aims that differentiates the concepts being developed and determines the ATM procedures that will be in place by 2010. Within this timeframe, the controller will still play a pivotal role using skills such as spatial perception, information processing, reason and decision making, which are presently employed. Indeed, many of today's ATM procedures will still be used in the latter half of this decade. Significant improvement will only be achieved through the development of powerful decision support software tools. The creation of automated ATM will provide an enhanced set of such tools that will assist the controller with conflict prediction, detection and resolution. The introduction of the future CNS technologies described in the previous sections will enable the evolution of more sophisticated ATM. This particularly applies to technologies that facilitate the automatic sending of aircraft position data. The combined benefits from the aforementioned advances in CNS technologies will serve as tools to support ATM. Indeed, according to ICAO, the integrated benefits and technical attributes of CNS technology will amalgamate to improve ATM, as portrayed in Figure 1.

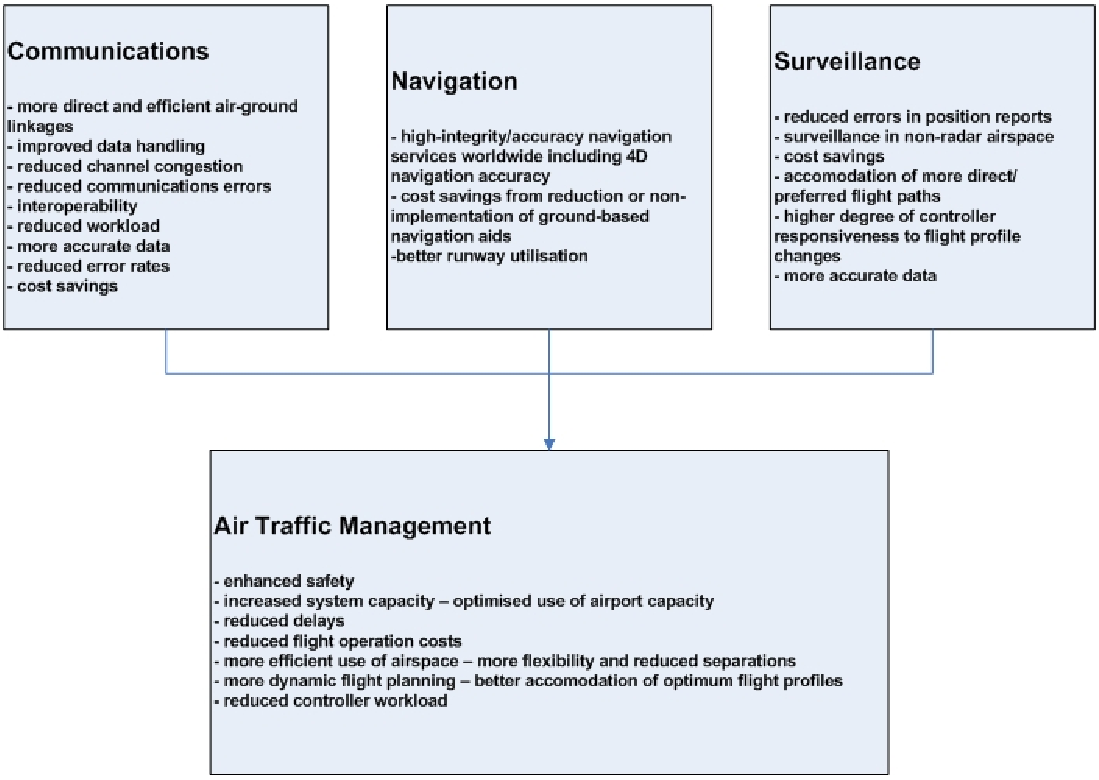


Figure 1 – Benefits of CNS technologies to improve ATM

The ultimate expectation is that accuracy could be improved through the rapid calculations associated with automation. Conflict prediction and detection, based on advanced computational methods, should allow more direct routings. These systems will be introduced in an evolutionary manner, and it is the rate of implementation that will determine the standard of ATM procedures by 2010. Nonetheless, some success is evident, such as the 'Direct-To' tool, which automatically searches for aircraft eligible for shorter trajectories to downstream fixes. Since software completion in 1999, it has been operating in shadow mode. The new Canadian Automated Air Traffic System (CAATS), which offers a paper-free environment with functions such as 4-D conflict probe and clearance validation, is also an example of advances in automated systems. Similarly, The Australian Advanced Air Traffic System (TAAATS), which became fully operational in 2000, incorporates new technology that links ground ATM automation systems with airborne avionics, thereby increasing flight path flexibility and providing more time for decision making.
Based on the current activity around the world regarding future ATM, the expected status of ATM by the end of the decade varies for different regions, but will ultimately be the result of many capacity enhancing programmes. For instance, regions that are capacity constrained will continue to see the

implementation of Reduced Vertical Separation Minima (RVSM) above FL290 and implementation of the Flexible Use of Airspace (FUA) concepts. Also, the fragmentation of airspace structures will be simplified, in order to provide seamless systems with greater capacity. For example, Europe's "Single Sky" must be achieved in this decade. Additionally, mandatory requirements to have Airborne Collision Avoidance Systems (ACAS) will complement future ATM.

ATM solutions will be brought about by ATM operational concepts, which are intended to assist and guide airspace planners with ATM design, in order to provide efficient and safe operations for all phases of flight. For instance, ICAO's operational concept in the "Global ATM plan" is still being developed. The concept will be complete when consensus has been obtained on issues such as autonomy of flight, separation assurance, situational awareness and collision avoidance. The concept will be functionally integrated with Air Traffic Services (ATS), Air Space Management (ASM) and Air Traffic Flow Management (ATFM) on the part of the providers, in addition to ATM aspects of flight operations from the airspace users. The concept will provide the means to quantify and assure performance in terms of safety, efficiency and regularity. There is an urgent need for ICAO to finish developing this concept.

The USA is implementing a programme for its operational concept of Free Flight, which aims to add levels of autonomous flight. The latest version of the programme, the Operational Evaluation Plan (OEP), was released recently. The US Federal Aviation Administration (FAA) completed Free Flight, Phase 1 in 2002. This is mostly ground-based and involves making existing, but not widely used, ATM capabilities quickly available to airspace users so that short-term benefits can be achieved in a desperate attempt to alleviate the country's air traffic congestion. Phase 2 aims to deploy decision-support systems for ATM between 2003 and 2007. Phase 3 will take place thereafter, when the FAA aims to complete the required infrastructure and integration of new automation to enable limited Free Flight operations. The FAA has found that the best method of achieving advanced ATM procedures is to work in conjunction with industry. Indeed, industry is conducting many trials itself, particularly in the area of new technologies.

Eurocontrol's gate-to-gate operational concept, "ATM Strategy for 2000", expects to meet capacity needs until 2015. In contrast with the US notion that future ATM will enable aircraft to navigate more cost-efficient routes, the European idea of future ATM is more concerned with being able to satisfy demand through improved airspace organisation and flow management. Noting the escalation of delays in recent times, Eurocontrol aims to provide greater levels of capacity by 2008. Indeed, there is need for serious short-term action in Europe. Thus, Eurocontrol is targeting the Area Control Centres (ACC) that were highlighted in recent performance review reports as having created many ATC delays over the last few years. Specific improvements in the ATM procedures of these centres should bring benefits. Accordingly, Eurocontrol's implementation programmes, which are discussed in previous sections of this paper, should increase capacity if they are conducted in a timely manner. They include, among others, RVSM, 8±33 KHz, SSR Mode S, PRNAV and FRAP. In addition, Collaborative Decision Making (CDM) will aid procedural ATM in terms of better slot co-ordination, airspace management and information management.

Another integral aspect of improving ATM in this region is the increased cooperation of member nations in sharing the ATM of their upper airspace, with a view to creating a uniform region. Therefore, ventures such as the Central European ATS (CEATS) centre in Vienna should also act as suitable solutions to the predicament.

The European Commission (EC) is involved, having created a High Level Group on ATM reform. In order that the ATM 2000- Strategy may be implemented, the EC believes that the European Union, not Eurocontrol, should be given greater powers based on new decision-making mechanisms and regulatory frameworks. Ultimately, attaining the goal of an integrated, global ATM system requires harmonisation and standardisation of regional and national system elements, which, in turn, must be based on plans that are specific to the different airspace types. It is the evolutionary implementation of the CNS elements and their orchestrated interaction that will form the backbone of the integrated ATM system in 2010.

To maximise use of CNS technologies for airspace planning purposes, given a region's ATM requirements, each of the three categories will be based on their required performance criteria, thus: Required Communications Performance (RCP); Required Navigation Performance (RNP); and Required Surveillance Performance (RSP).
The introduction of these performance criteria (RCP, RNP and RSP) will help simplify the basis of ATM in the various types of airspace in addition to enabling the appropriate technologies and procedures to be developed. Many limitation factors of the present CNS system will inhibit the development of ATM throughout the world unless the more advanced technologies described in this thesis are implemented.

Figure 2 – Relationships between CNS and RCP, RNP and RSP

Accordingly, Figure 2 summarises suitable CNS systems upon which ATM will most likely be based in 2010 for different airspace types. Hence, there is a need for the new technologies and ATM approaches that will lead to reductions in separation between aircraft and allow for increases in airspace capacity, cost-effectiveness, efficiency, flexibility and safety. Indeed, CNS/ATM systems provide an opportunity to overcome the present system's shortcomings and accommodate forecast traffic levels. However, there are many issues that affect their rate of introduction, which incorporate political, management, regulatory, technical, certification, liability, financial and institutional aspects. Therefore, in order to implement CNS/ATM successfully, these different issues must be addressed. For instance, institutional problems relating to ATS providers may often be solved through (part) privatisation or corporatisation, whereby commercialisation is seen as a means of improving their investment capability and institutional efficiency. It is thought that ATS organisations become more competitive with such status and that they do not possess the apparent hindrances associated with being government departments. Similarly, ATS providers are encouraged by ICAO to partake in "international co-operation in the provision and operation of air navigation services where this is beneficial for the providers and users concerned". Thus, changing the structure of ATS providers can reduce adverse institutional issues. Ultimately, mixed opinion exists about where Air Navigation Services (ANS) will be at the end of this decade in terms of CNS/ATM technologies and procedures. This particularly applies to when the satellite-based systems will replace ground-based aids, noting that there will be greater advances in some nations and regions than others. Indeed, concepts such as Free Flight per se will be more applicable to specific types of airspace. Nonetheless, it may be observed that the present dilemmas would not exist if the future air navigation systems had already been sufficiently applied around the world. However, many technologies and procedures are nearly ready for mainstream implementation, advances have been made with datalink applications and satellite-based communications facilities, GPS enhanced navigation procedures in all flight phases are becoming more mature, while the concept of RNP is aiding airspace planning and facilitating adherence to standards in many regions. Correspondingly, the success of surveillance systems, such as ADS and ADS-B, is encouraging, and enhanced ATM procedures, such as automated sequencing tools and dynamic aircraft re-routing, are currently operational. Therefore, it would appear that the evolution of ANS is at a crossroads and that this decade will witness profound changes, which will hopefully alleviate many of the existing air traffic problems.

Airspace	Function	Current Systems	Systems in 2010
Continental and oceanic en-route airspace with low-density traffic	Communications	VHF data & voice HF data & voice	ATN VHF data & voice AMSS data & voice HF data, particularly in Polar regions
	Navigation	LORAN-C NDB VOR/DME Barometric altimetry INS/IRS	RNAV GNSS (with SBAS) Barometric altimetry GNSS altitude INS/IRS
	Surveillance	Primary radar/SSR Voice position reports	ADS & ADS-B
Continental airspace with high-density traffic	Communications	VHF voice	VHF data & voice AMSS data & voice SSR Mode S datalink
	Navigation	LORAN-C NDB VOR/DME Barometric altimetry INS/IRS	RNAV GNSS (with SBAS) Barometric altimetry GNSS altitude INS/IRS
	Surveillance	Primary radar SSR Mode A/C	SSR (Mode A/C & S) ADS & ADS-B
Oceanic airspace with high-density traffic	Communications	HF data & voice	AMSS data & voice
	Navigation	MNPS LORAN-C Barometric altimetry INS/IRS	RNAV GNSS (with SBAS) Barometric altimetry GNSS altitude INS/IRS
	Surveillance	Voice position reports	ADS & ADS-B
Terminal areas with high-density traffic	Communications	VHF voice	ATN VHF voice/data AMSS data/voice SSR Mode S datalink
	Navigation	NDB VOR/DME ILS Barometric altimetry INS/IRS	RNAV GNSS (using LAAS) ILS/MLS DME Barometric altimetry INS/IRS
	Surveillance	Primary radar SSR Mode A/C	SSR (Mode A/C & S) ADS & ADS-B

Table 1 – Summary of Future Air Navigation Systems [ATC-1]

SUMMARY OF CHAPTER 1

CHAPTER 2

AIR TRAFFIC CONTROL

Every National Airspace System (NAS) is the network of: air navigation facilities, equipment, services, airports or landing areas, aeronautical charts, information/services, rules, regulations, procedures, technical information, manpower, and material. Included are system components shared jointly with the military. The system's present configuration is a reflection of the technological advances concerning the speed and altitude capability of jet aircraft, as well as the complexity of microchip and satellite-based navigation equipment. To conform to international aviation standards, every country adopted the primary elements of the classification system developed by the International Civil Aviation Organization (ICAO).

This chapter discusses also airspace classification; en route, terminal, approach procedures, and operations within every NAS. It will cover the communication equipment, communication procedures, and air traffic control (ATC) facilities and services available for a flight under instrument flight rules (IFR) in the National Airspace System (NAS).

Since no single procedure can be outlined that is applicable to the planning and preparation involved with all flights conducted under instrument flight rules (IFR). Once you'll have understood the overall operation of IFR flight, the many procedural details can be put into the appropriate sequence. Hence, this chapter explains also the sources for flight planning, the conditions associated with instrument flight, and the procedures used for each phase of IFR flight: departure, en route, and approach.

1 AIRSPACE CLASSIFICATION

Airspace is internationally designated as follows (Figure 1 and Table 1):

Class A—Generally, that airspace from 18,000 feet mean sea level (MSL) up to and including flight level (FL) 600, including the airspace overlying the waters within 12 nautical miles (NM) of the coast of the 48 contiguous states and Alaska. Unless otherwise authorized, all pilots must operate their aircraft under instrument flight rules (IFR).

Class B—Generally, that airspace from the surface to 10,000 feet MSL surrounding the nation's busiest airports in terms of airport operations or passenger enplanements. The configuration of each Class B airspace area is individually tailored and consists of a surface area and two or more layers (some Class B airspace areas resemble upside-down wedding cakes), and is designed to contain all published instrument procedures once an aircraft enters the airspace. An air traffic control (ATC) clearance is required for all aircraft to operate in the area, and all aircraft that are so cleared receive separation services within the airspace.

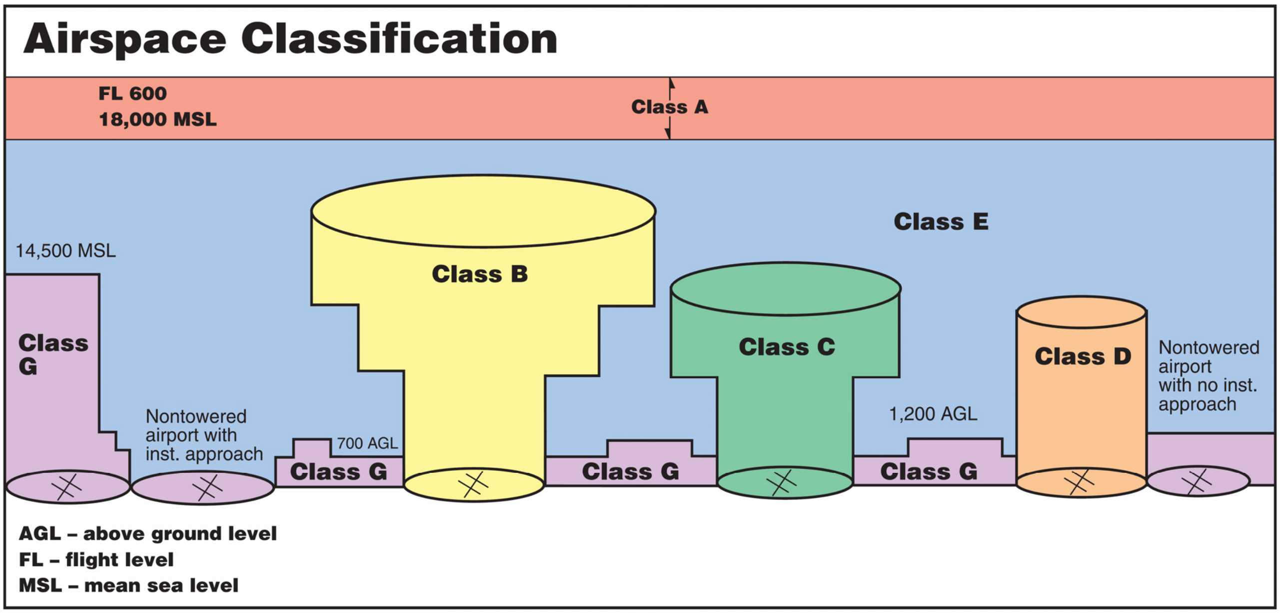


Figure 1 – Airspace Classification

Airspace	Class A	Class B	Class C	Class D	Class E	Class G
Entry Requirements	ATC clearance	ATC clearance	Prior two-way communications	Prior two-way communications	Prior two-way communications*	Prior two-way communications*
Minimum Pilot Qualifications	Instrument Rating	Private or Student certification. Local restric-tions apply	Student certificate	Student certificate	Student certificate	Student certificate
Two-Way Radio Communications	Yes	Yes	Yes	Yes	Yes, under IFR flight plan*	Yes*
Special VFR Allowed	No	Yes	Yes	Yes	Yes	N/A
VFR Visibility Minimum	N/A	3 statute miles	3 statute miles	3 statute miles	3 statute miles**	1 statute mile†
VFR Minimum Distance from Clouds	N/A	Clear of clouds	500' below, 1,000' above, 2,000' horizontal	500' below, 1,000' above, 2,000' horizontal	500' below,** 1,000' above, 2,000' horizontal	Clear of clouds†
VFR Aircraft Separation	N/A	All	IFR aircraft	Runway Operations	None	None
Traffic Advisories	Yes	Yes	Yes	Workload permitting	Workload permitting	Workload permitting
Airport Application	N/A	• Radar • Instrument Approaches • Weather • Control Tower • High Density	• Radar • Instrument Approaches • Weather • Control Tower	• Instrument Approaches • Weather • Control Tower	• Instrument Approaches • Weather	• Control Tower

*Only if a temporary tower or control tower is present is the exception.
**Only true below 10,000 feet.
†Only true during day at or below 1,200 feet AGL (see 14 CFR part 91).

Table 1 – Airspace Classification

Class C—Generally, that airspace from the surface to 4,000 feet above the airport elevation (charted in MSL) surrounding those airports that have an operational control tower, are serviced by a radar approach control, and have a certain number of IFR operations or passenger enplanements. Although the configuration of each Class C area is individually tailored, the airspace usually consists of a surface area with a 5 NM radius, an outer circle with a 10 NM radius that extends from 1,200 feet to 4,000 feet above the airport elevation and an outer area. Each person must establish two way radio communications with the ATC facility providing air traffic services prior to entering the airspace and thereafter maintain those communications while within the airspace.

Class D—Generally, that airspace from the surface to 2,500 feet above the airport elevation (charted in MSL) surrounding those airports that have an operational control tower. The configuration of each Class D airspace area is individually tailored and when instrument procedures are published, the airspace will normally be designed to contain the procedures. Arrival extensions for instrument approach procedures (IAPs) may be Class D or Class E airspace. Unless otherwise authorized, each person must establish two-way radio communications with the ATC facility providing air traffic services prior to entering the airspace and thereafter maintain those communications while in the airspace.

Class E—Generally, if the airspace is not Class A, Class B, Class C, or Class D, and it is controlled airspace, it is Class E airspace. Class E airspace extends upward from either the surface or a designated altitude to the overlying or adjacent controlled airspace. When designated as a surface area, the airspace will be configured to contain all instrument procedures. Also in this class are airways, airspace beginning at either 700 or 1,200 feet above ground level (AGL) used to transition to and from the terminal or en route environment, en route domestic, and offshore airspace areas designated below 18,000 feet MSL.

Class G—That airspace not designated as Class A, B, C, D, or E. Class G airspace is essentially uncontrolled by ATC except when associated with a temporary control tower.

1.1 SPECIAL USE AIRSPACE

Special use airspace is the designation for airspace in which certain activities must be confined, or where limitations may be imposed on aircraft operations that are not part of those activities. Certain special use airspace areas can create limitations on the mixed use of airspace. The special use airspace depicted on instrument charts includes the area name or number, effective altitude, time and weather conditions of operation, the controlling agency, and the chart panel location.

On National Aeronautical Charting Office (NACO) en route charts, this information is available on the panel opposite the air/ground (A/G) voice communications.

Prohibited areas contain airspace of defined dimensions within which the flight of aircraft is prohibited. Such areas are established for security or other reasons associated with the national welfare. These areas are published in the Federal Register and are depicted on aeronautical charts. The area is charted as a "P" with a number (e.g., "P-123"). As the name implies, flight through this airspace is not permitted.

Restricted areas are areas where operations are hazardous to non participating aircraft and contain airspace within which the flight of aircraft, while not wholly prohibited, is subject to restrictions. Activities within these areas must be confined because of their nature, or limitations imposed upon aircraft operations that are not a part of those activities, or both. Restricted areas denote the existence of unusual, often invisible, hazards to aircraft (e.g., artillery firing, aerial gunnery, or guided missiles). IFR flights may be authorized to transit the airspace and are routed accordingly. Penetration of restricted areas without authorization from the using or controlling agency may be extremely hazardous to the aircraft and its occupants. ATC facilities apply the following procedures when aircraft are operating on an IFR clearance (including those cleared by ATC to maintain visual flight rules (VFR)-On-Top) via a route that lies within joint-use restricted airspace:

1. If the restricted area is not active and has been released to the Federal Aviation Administration (FAA), the ATC facility will allow the aircraft to operate in the restricted airspace without issuing specific clearance for it to do so.
2. If the restricted area is active and has not been released to the FAA, the ATC facility will issue a clearance which will ensure the aircraft avoids the restricted airspace.

Restricted areas are charted with an "R" followed by a number (e.g., "R-5701") and are depicted on the en route chart appropriate for use at the altitude or FL being flown.

Warning areas are similar in nature to restricted areas; however, the government does not have sole jurisdiction over the airspace. The purpose of such areas is to warn non participating pilots of the potential danger. A warning area may be located over domestic or international waters or both. The airspace is designated with a "W" and a number (e.g., "W-123").

Military operations areas (MOAs) consist of airspace of defined vertical and lateral limits established for the purpose of separating certain military training activities from IFR traffic. Whenever an MOA is being used, non participating IFR traffic may be cleared through an MOA if IFR separation can be provided by ATC. Otherwise, ATC will reroute or restrict non participating IFR traffic. MOAs are depicted on sectional, VFR terminal area, and en route low altitude charts and are named rather than numbered (e.g., "Boardman MOA").

Alert areas are depicted on aeronautical charts with an "A" and a number (e.g., "A-123") to inform non participating pilots of areas that may contain a high volume of pilot training or an unusual type of aerial activity. Pilots should exercise caution in alert areas. All activity within an alert area shall be conducted in accordance with regulations, without waiver, and pilots of participating aircraft, as well as pilots transiting the area shall be equally responsible for collision avoidance.

Military Training Routes (MTRs) are routes used by military aircraft to maintain proficiency in tactical flying. These routes are usually established below 10,000 feet MSL for operations at speeds in excess of 250 knots. Some route segments may be defined at higher altitudes for purposes of route continuity. Routes are identified as IFR (IR), and VFR (VR), followed by a number. MTRs with no segment above 1,500 feet AGL are identified by four number characters (e.g., IR1206, VR1207, etc.). MTRs that include one or more segments above 1,500 feet AGL are identified by three number characters (e.g., IR206, VR207, etc.). IFR Low Altitude En Route Charts depict all IR routes and all VR routes that accommodate operations above 1,500 feet AGL. IR routes are conducted in accordance with IFR regardless of weather conditions.

Temporary flight restrictions (TFRs) are put into effect when traffic in the airspace would endanger or hamper air or ground activities in the designated area. For example, a forest fire, chemical accident, flood, or disaster-relief effort could warrant a TFR, which would be issued as a Notice to Airmen (NOTAM).

National Security Areas (NSAs) consist of airspace of defined vertical and lateral dimensions established at locations where there is a requirement for increased security and safety of ground facilities. Flight in NSAs may be temporarily prohibited by regulation under the provisions of Title 14 of the Code of Federal Regulations (14 CFR) part 99, and prohibitions will be disseminated via NOTAM.

1.2 AIRWAYS

The primary navigational aid (NAVAID) for routing aircraft operating under IFR is the airways system. Each airway is based on a centreline that extends from one NAVAID or intersection to another NAVAID specified for that airway. An airway includes the airspace within parallel boundary lines 4 NM to each side of the centreline. As in all instrument flight, courses are magnetic, and distances are in NM. The airspace of an airway has a floor of 1,200 feet AGL, unless otherwise specified. An airway does not include the airspace of a prohibited area.

Victor airways include the airspace extending from 1,200 feet AGL up to, but not including 18,000 feet MSL. The airways are designated on sectional and IFR low altitude en route charts with the letter "V" followed by a number (e.g., "V23"). Typically, Victor airways are given odd numbers when oriented north/south and even numbers when oriented east/west. If more than one airway coincides on a route
segment, the numbers are listed serially (e.g., "V287-495-500").

Jet routes exist only in Class A airspace, from 18,000 feet MSL to FL450, and are depicted on high-altitude en route charts. The letter "J" precedes a number to label the airway (e.g., J12).

Preferred IFR routes have been established between major terminals to guide pilots in planning their routes of flight, minimizing route changes and aiding in the orderly management of air traffic on airways. Low and high altitude preferred routes are listed in the Airport/Facility Directory (A/FD).

Tower En Route Control (TEC) is an ATC program that uses overlapping approach control radar services to provide IFR clearances. By using TEC, you are routed by airport control towers. Some advantages include abbreviated filing procedures, fewer delays, and reduced traffic separation requirements. TEC is dependent upon the ATC's workload and the procedure varies among locales. The latest version of Advisory Circular (AC) 90-91, National Route Program, provides guidance to users of the NAS for participation in the National Route Program (NRP). All flights operating at or above FL290 are eligible to participate in the NRP, the primary purpose of which is to allow operators to plan minimum time/ cost routes that may be off the prescribed route structure. NRP aircraft are not subject to route-limiting restrictions (e.g., published preferred IFR routes) beyond a 200 NM radius of their point of departure or destination.

2 AIR TRAFFIC PUBLICATIONS

2.1 IFR EN ROUTE CHARTS

The objective of IFR en route flight is to navigate within the lateral limits of a designated airway at an altitude consistent with the ATC clearance. Your ability to fly instruments in the system, safely and competently, is greatly enhanced by understanding the vast array of data available to the pilot within the instrument charts. The NACO maintains the database and produces the charts for the U.S. government. En route high-altitude charts provide aeronautical information for en route instrument navigation (IFR) at or above 18,000 feet MSL. Information includes the portrayal of jet routes, identification and frequencies of radio aids, selected airports, distances, time zones, special use airspace, and related information. Established routes from 18,000 feet MSL to FL450 use NAVAIDs not more than 260 NM apart. Scales vary from 1 inch = 45 NM to 1 inch = 18 NM. The charts are revised every 56 days.

To effectively depart from one airport and navigate en route under instrument conditions you need the appropriate IFR en route low-altitude chart(s). The IFR low altitude en route chart is the instrument equivalent of the sectional chart. When folded, the cover of the NACO en route chart displays a map of the U.S. showing the coverage areas. Cities near congested airspace are shown in black type and their associated area chart is listed in the box in the lower left-hand corner of the map coverage box. Also noted is the highest off-route obstruction clearance altitude. The effective date of the chart is printed on the other side of the folded chart. Information concerning MTRs are also included on the chart cover. Scales vary from 1 inch = 5 NM to 1 inch = 20 NM. The en route charts are revised every 56 days. When the NACO en route chart is unfolded, the legend is displayed and provides information concerning airports, NAVAIDs, air traffic services, and airspace.

Area navigation (RNAV) routes, including routes using global positioning system (GPS) for navigation, are not normally depicted on IFR en route charts. However, a number of RNAV routes have been established in the high-altitude structure and are depicted on the RNAV en route high altitude charts. RNAV instrument departure procedures (DPs) and standard terminal arrival routes (STARs) are contained in the U.S. Terminal Procedures booklets. The Graphic Notices and Supplemental Data also contains a tabulation of RNAV routes.

In addition to the published routes, one may fly a random RNAV route under IFR if it is approved by ATC. Random RNAV routes are direct routes, based on area navigation capability, between waypoints defined in terms of latitude/longitude coordinates, degree-distance fixes, or offsets from established routes/airways at a specified distance and direction.

Radar monitoring by ATC is required on all random RNAV routes. These routes can only be approved in a radar environment. Factors that will be considered by ATC in approving random RNAV routes include the capability to provide radar monitoring, and compatibility with traffic volume and flow. ATC will radar monitor each flight; however, navigation on the random RNAV route is the responsibility of the pilot.

Reliance on RNAV systems for instrument approach operations is becoming more commonplace as new systems, such as GPS and wide area augmentation system (WAAS) are developed and deployed. In order to foster and support full integration of RNAV into the NAS, the FAA has developed a charting format for RNAV approach charts.

2.2 AIRPORT INFORMATION

Airport information is provided in the legend, and the symbols used for the airport name, elevation, and runway length are similar to the sectional chart presentation. Instrument approaches can be found at airports with blue or green symbols, while the brown airport symbol denotes airports that do not have approved instrument approaches. Asterisks are used to indicate the part-time nature of tower operations, lighting facilities, and airspace classifications (consult the communications panel on the chart for primary radio frequencies and hours of operation). The asterisk could also indicate that approaches are not permitted during the non-operating hours, and/or filing as an alternate is not approved during specified hours. A box after an airport name with a “C” or “D” inside indicates Class C and D airspace, respectively.

2.3 TERMINAL PROCEDURES PUBLICATIONS

While the en route charts provide the information necessary to safely transit broad regions of airspace, the Terminal Procedures Publication (TPP) enables pilots to guide their aircraft into airports. Terminal routes feed aircraft to a point where IAPs can be flown to a minimum altitude for landing.

Whether for departing or arriving, these procedures exist to make the controllers' and pilots' jobs safer and more efficient. Available in booklets by region (published by the NACO), the TPP includes approach procedures, arrival and DPs, and airport diagrams.

2.4 DEPARTURE PROCEDURES (DPS)

Departure procedures (DPs) provide obstacle clearance protection to aircraft in instrument meteorological conditions (IMC), while reducing communications and departure delays. DPs are published in text and/or charted graphic form. Regardless of the format, all DPs provide a way to depart the airport and transition to the en route structure safely. When available, pilots are strongly encouraged to file and fly a DP at night, during marginal visual meteorological conditions (VMC), and IMC. All DPs provide obstacle clearance provided the aircraft crosses the end of the runway at least 35 feet AGL; climbs to 400 feet above airport elevation before turning; and climbs at least 200 feet per nautical mile (FPNM), unless a higher climb gradient is specified to the assigned altitude. ATC may vector an aircraft off a previously assigned DP; however, the 200 FPNM or the FPNM specified in the DP, is required. Textual DPs are listed by airport in the IFR Take-Off Minimums and Departure Procedures Section, Section C, of the TPP. Graphic DPs are depicted in the TPP following the approach procedures for the airport.

2.5 STANDARD TERMINAL ARRIVAL ROUTES (STARS)

Standard terminal arrival routes (STARs) depict prescribed routes to transition the instrument pilot from the en route structure to a fix in the terminal area from which an instrument approach can be conducted. If you do not have the appropriate STAR in your possession, you can write "No STAR" in the flight plan. However, if the controller is busy, you might be cleared along the same route and, if necessary, the controller will have you copy the entire text of the procedure.

Textual DPs and STARs are listed alphabetically at the beginning of the NACO booklet, and graphic DPs (charts) are included after the respective airport's IAP.

2.6 INSTRUMENT APPROACH PROCEDURES CHARTS (IAPS)

The IAPs chart provides the method to descend and land safely in low visibility conditions. The FAA has established the IAPs after thorough analyses of obstructions, terrain features, and navigational facilities. Manoeuvres, including altitude changes, course corrections, and other limitations, are prescribed in the IAPs. The approach charts reflect the criteria associated with the Standard for Terminal Instrument Approach Procedures (TERPs), which prescribes standardized methods for use in designing instrument flight procedures.

2.7 TERMINAL ARRIVAL AREA (TAA)

The design objective of the Terminal Arrival Area (TAA) procedure is to provide a transition method for arriving aircraft with GPS/RNAV equipment. TAAs will also eliminate or reduce the need for feeder routes, departure extensions, and procedure turns or course reversal. The TAA is controlled airspace established in conjunction with the standard or modified RNAV approach configurations.

The standard TAA has three areas: straight-in, left base, and right base. The arc boundaries of the three areas of the TAA are published portions of the approach and allow aircraft to transition from the en-route structure direct to the nearest IAF. When crossing the boundary of each of these areas or when released by ATC within the area, the pilot is expected to proceed direct to the appropriate waypoint IAF for the approach area being flown. A pilot has the option in all areas of proceeding directly to the holding pattern.

The TAA provides the pilot and air traffic controller with an efficient method for routing traffic from the en-route to the terminal structure. The basic "T" contained in the TAA normally aligns the procedure on runway centreline, with the missed approach point (MAP) located at the threshold, the FAF 5 NM from the threshold, and the intermediate fix (IF) 5 NM from the FAF.

In order to accommodate descent from a high en route altitude to the initial segment altitude, a hold in lieu of a procedure turn provides the aircraft with an extended distance for the necessary descent gradient. The holding pattern constructed for this purpose is always established on the centre IAF waypoint. Other modifications may be required for parallel runways, or due to operational requirements. When published, the RNAV chart will depict the TAA through the use of "icons" representing each TAA associated with the RNAV procedure. These icons will be depicted in the plan view of the approach plate, generally arranged on the chart in accordance with their position relative to the aircraft's arrival from the en route structure.

2.8 AIRPORT DIAGRAM

The airport diagram, located on the bottom right side of the chart, includes many helpful features. IAPs for some of the larger airports devote an entire page to an airport diagram. Information concerning runway orientation, lighting, final approach bearings, airport beacon, and obstacles all serve to guide the pilot in the final phases of flight. The diagram shows the runway configuration in solid black, while the taxiways and aprons are shaded gray. Other runway environment features are shown, such as the runway identification, dimensions, magnetic heading, displaced threshold, arresting gear, usable length, and slope. The airport elevation is indicated in a separate box at the top of the airport diagram box. The touch down zone elevation (TDZE), which is the highest elevation within the first 3,000 feet of the runway, is designated at the approach end of the procedure's runway. Beneath the airport diagram is the time and speed table. The table provides the distance and the amount of time required to transit the distance from the FAF to the MAP for selected groundspeeds. The approach lighting systems and the visual approach lights are depicted on the approach chart. White on black symbols are used for identifying pilot-controlled lighting (PCL). Runway lighting aids are also noted (e.g., REIL, HIRL), as is the runway centreline lighting (RCL).

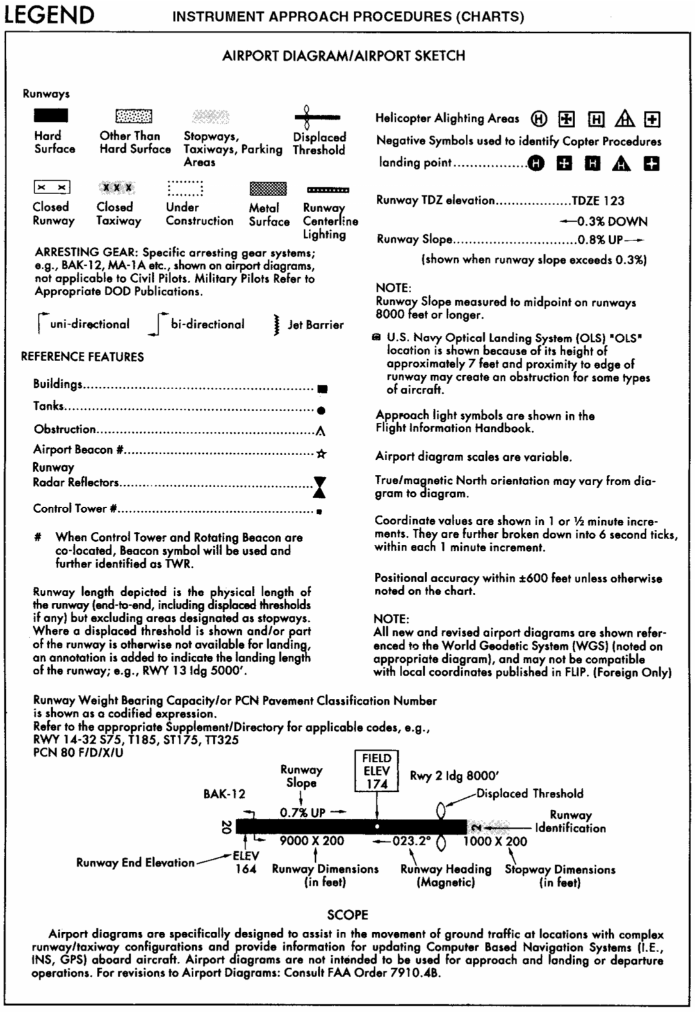


Figure 2 – Airport diagram legend

2.9 RNAV INSTRUMENT APPROACH CHARTS

Instrument approach charts are being converted to a charting format similar to the format developed for RNAV IAP. This format avoids unnecessary duplication and proliferation of instrument approach charts. The approach minimums for unaugmented GPS, Wide Area Augmentation System (WAAS), Local Area Augmentation System (LAAS), will be published on the same approach chart as lateral navigation/vertical navigation (LNAV/VNAV). Other types of equipment may be authorized to conduct the approach based on the minima notes in the front of the TPP approach chart books. Approach charts titled "RNAV RWY XX" may be used by aircraft with navigation systems that meet the required navigational performance (RNP) values for each segment of the approach. The chart may contain as many as four lines of approach minimums: Global landing system (GLS); WAAS and LAAS; LNAV/VNAV; LNAV; and circling. LNAV/VNAV is an instrument approach with lateral and vertical guidance with integrity limits similar to barometric vertical navigation (BARO VNAV).

RNAV procedures that incorporate a final approach step-down fix may be published without vertical navigation, on a separate chart, also titled RNAV. During a transition period when GPS procedures are undergoing revision to a new title, both RNAV and GPS approach charts and formats will be published. ATC clearance for the RNAV procedure will authorize a properly-certificated pilot to utilize any landing minimums for which the aircraft is certified. The RNAV chart will include formatted information required for quick pilot or flight crew reference located at the top of the chart. This portion of the chart was developed based on a study by the Department of Transportation (DOT), Volpe National Transportation Systems Centre.

Chart terminology will change slightly to support the new procedure types:

1. DA replaces the term DH. DA conforms to the international convention where altitudes relate to MSL and heights relate to AGL. DA will eventually be published for other types of IAPs with vertical guidance, as well. DA indicates to the pilot that the published descent profile is flown to the DA (MSL), where a missed approach will be initiated if visual references for landing are not established. Obstacle clearance is provided to allow a momentary descent below DA while transitioning from the final approach to the missed approach. The aircraft is expected to follow the missed approach instructions while continuing along the published final approach course to at least the published runway threshold waypoint or MAP (if not at the threshold) before executing any turns.
2. MDA will continue to be used for the LNAV-only and circling procedures.
3. Threshold crossing height (TCH) has been traditionally used in precision approaches as the height of the glide slope above threshold. With publication of LNAV/VNAV minimums and RNAV descent angles, including graphically depicted descent profiles, TCH also applies to the height of the "descent angle," or glidepath, at the threshold. Unless otherwise required for larger type aircraft which may be using the IAP, the typical TCH will be 30 to 50 feet.

The minima format changes slightly:

1. Each line of minima on the RNAV IAP will be titled to reflect the RNAV system applicable (e.g., GLS, LNAV/VNAV, and LNAV.) Circling minima will also be provided.
2. The minima title box will also indicate the nature of the minimum altitude for the IAP. For example: DA will be published next to the minima line title for minimums supporting vertical guidance, and MDA will be published where the minima line supports only lateral guidance. During an approach where an MDA is used, descent below MDA is not authorized.
3. Where two or more systems share the same minima, each line of minima will be displayed separately.

3 NAVIGATION/COMMUNICATION (NAV/COM) EQUIPMENT

Civilian pilots communicate with ATC on frequencies in the very high frequency (VHF) range between 118.000 and 136.975 MHz. To derive full benefit from the ATC system, radios capable of 25 kHz spacing are required (e.g., 134.500, 134.575, 134.600, etc.). If ATC assigns a frequency that cannot be selected on your radio, one can ask for an alternative frequency. Many radios allow the pilot to have one or more frequencies stored in memory and one frequency active for transmitting and receiving (called simplex operation). It is possible to communicate with some automated flight service stations (AFSS) by transmitting on 122.1 MHz (selected on the communication radio) and receiving on a VHF omnidirectional range (VOR) frequency (selected on the navigation radio). This is called duplex operation. It should be necessary to select a receiver, communication or navigation, only when you want to monitor one communications frequency while communicating on another. One example is listening to automatic terminal information service (ATIS) on one receiver while communicating with ATC on the other. Monitoring a navigation receiver to check for proper identification is another reason to use the switch panel. Most audio switch panels also include a marker beacon receiver; all marker beacons transmit on 75 MHz, so there is no frequency selector.

3.1 RADAR AND TRANSPONDERS

ATC radars have a limited ability to display primary returns, which is energy reflected from an aircraft's metallic structure. Their ability to display secondary returns (transponder replies to ground interrogation signals) makes possible the many advantages of automation. A transponder is a radar beacon transmitter/receiver installed in the instrument panel. ATC beacon transmitters send out interrogation signals continuously as the radar antenna rotates. When an interrogation is received by a transponder, a coded reply is sent to the ground station where it is displayed on the controller's scope. A "Reply" light on the transponder panel flickers every time it receives and replies to a radar interrogation. Transponder codes are assigned by ATC. When a controller asks the pilot to "ident" and you push the ident button, pilot's return on the controller's scope is intensified for precise identification of the flight. When requested, briefly push the ident button to activate this feature. It is good practice to verbally confirm you have changed codes or pushed the ident button.

3.2 MODE C (ALTITUDE REPORTING)

Primary radar returns indicate only range and bearing from the radar antenna to the target; secondary radar returns can display altitude Mode C on the control scope if the aircraft is equipped with an encoding altimeter or blind encoder. In either case, when the transponder's function switch is in the ALT position the aircraft's pressure altitude is sent to the controller. Adjusting the altimeter's Kollsman window has no effect on the altitude read by the controller. Transponders must be ON at all times when operating in controlled airspace; altitude reporting is required by regulation in Class B and Class C airspace and inside of a 30-mile circle surrounding the primary airport in Class B airspace. Altitude reporting should also be ON at all times.

3.3 COMMUNICATION PROCEDURES

Clarity in communication is essential for a safe instrument flight. This requires pilots and controllers to use terms that are understood by both—the Pilot/Controller Glossary in the Aeronautical Information Manual (AIM) is the best source of terms and definitions. The AIM is revised twice a year and new definitions are added, so the Glossary should be reviewed frequently. Because clearances and instructions are comprised largely of letters and numbers, a phonetic pronunciation guide has been developed for both. Air traffic controllers must follow the guidance of the Air Traffic Control Manual when communicating with pilots. The manual presents the controller with different situations and prescribes precise terminology that must be used. This is advantageous for pilots, because once they have recognized a pattern or format they can expect future controller transmissions to follow that format. Controllers are faced with a wide variety of communication styles based on pilot experience, proficiency, and professionalism.

3.4 COMMUNICATION FACILITIES

The controller's primary responsibility is separation of aircraft operating under IFR. This is accomplished with ATC facilities which include the AFSS, airport traffic control tower (ATCT), terminal radar approach control (TRACON), and air route traffic control centre (ARTCC).

3.4.1 AUTOMATED FLIGHT SERVICE STATIONS (AFSS)

A pilot's first contact with ATC will probably be through AFSS, either by radio or telephone. AFSS's provide pilot briefings, receives and processes flight plans, relays ATC clearances, originates Notices to Airmen (NOTAMs), and broadcasts aviation weather. Some facilities provide En Route Flight Advisory Service (EFAS), take weather observations, and so on.

3.4.2 AIR TRAFFIC CONTROL TOWERS

Several controllers in the tower cab will be involved in handling an instrument flight. Where there is a dedicated clearance delivery position, that frequency will be found in the A/FD and on the instrument approach chart for the departure airport. Where there is no clearance delivery position, the ground controller will perform this function. At the busiest airports, pre-taxi clearance is required; the frequency for pre-taxi clearance can be found in the A/FD. Taxi clearance should be requested not more than 10 minutes before proposed taxi time. Instrument clearances can be overwhelming if you try to copy them verbatim, but they follow a format that allows you to be prepared when you say "Ready to copy." The format is: Clearance limit (usually the destination airport); Route, including any departure procedure; initial Altitude; Frequency (for departure control); and Transponder code. With the exception of the transponder code, pilots will know most of these items before engine start.
One technique for clearance copying is writing C-R-A-F-T:

C learance limit
R oute (including DP, if any)
A ltitude
F requency
T ransponder code.

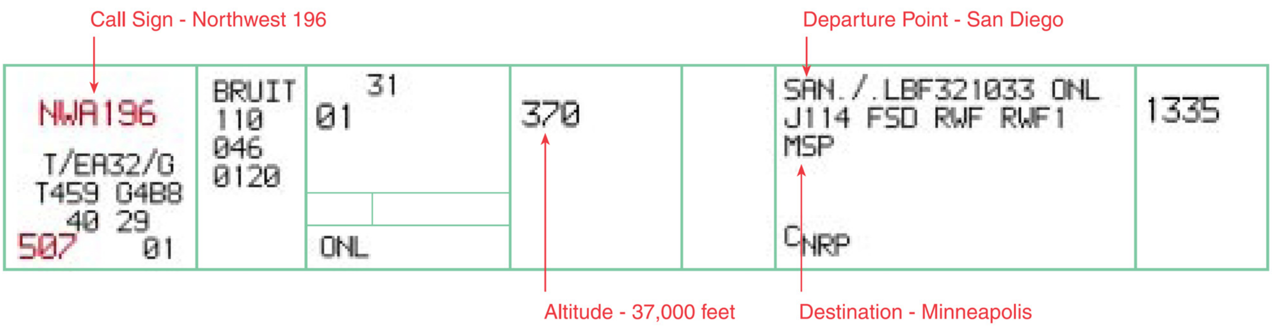


Figure 3 – Flight strip

Assume you are the pilot and you have filed an IFR flight plan from Seattle, Washington to Sacramento, California via V-23 at 7,000 feet. You note traffic is taking off to the north from Seattle-Tacoma (Sea-Tac) airport and, by monitoring the clearance delivery frequency, you note the departure procedure being assigned to southbound flights. Your clearance limit will be the destination airport, so you can write "SAC" after the letter C. Write "SEATTLE TWO – V23" after R for Route, because you heard departure control issue this departure to other flights (you could also call the tower on the telephone to ask what departure is in use). Write "7" after the A, the departure control frequency printed on the approach charts for Sea-Tac after F, and leave the space after T blank—the transponder code is generated by computer and can seldom be determined in advance. Now call clearance delivery and report ready to copy.

The last clearance received supersedes all previous clearances. For instance, if the DP says "Climb and maintain 2,000 feet, expect higher in 6 miles" and upon contacting the departure controller you hear "Climb and maintain 8,000 feet," the 2,000-foot restriction has been cancelled. This rule applies in both terminal and Centre airspace.

The "local" controller is responsible for operations in the Class D airspace and on the active runways. At some towers, designated as IFR towers, the local controller has vectoring authority. At visual flight rules (VFR) towers, the local controller accepts inbound IFR flights from the terminal radar facility and cannot provide vectors. The local controller also coordinates flights in the local area with radar controllers. Although Class D airspace normally extends 2,500 feet above field elevation, towers frequently release the top 500 feet to the radar controllers to facilitate overflights. Accordingly, when your flight is vectored over an airport at an altitude that appears to enter the tower controller's airspace, there is no need for you to contact the tower controller—all coordination is handled by ATC.

The departure radar controller may be in the same building as the control tower, but it is more likely that the departure radar position is remotely located. The tower controller will not issue a takeoff clearance until the departure controller issues a release.

3.4.3 TERMINAL RADAR APPROACH CONTROL (TRACON)

TRACONs are considered terminal facilities because they provide the link between the departure airport and the en route structure of the NAS. Terminal airspace normally extends 30 nautical miles (NM) from the facility, with a vertical extent of 10,000 feet; however, dimensions vary widely. Class B and Class C airspace dimensions are provided on aeronautical charts. At terminal radar facilities the airspace is divided into sectors, each with one or more controllers, and each sector is assigned a discrete radio frequency. All terminal facilities are approach controls, and should be addressed as "Approach" except when directed to do otherwise ("Contact departure on 120.4").

Terminal radar antennas are located on or adjacent to the airport. Figure 4 shows a typical configuration. Terminal controllers can assign altitudes lower than published procedural altitudes called minimum vectoring altitudes (MVAs). These altitudes are not published and accessible to pilots, but are displayed at the controller's position, as shown in Figure 5.

After receiving and accepting a clearance and reporting ready for takeoff, a controller in the tower contacts the TRACON for a release—one will not be released until the departure controller can fit the flight into the departure flow. One may have to hold for release. When you receive takeoff clearance, the departure controller is aware of your flight and is waiting for your call. All of the information the controller needs is on the departure strip or the computer screen, so you need not repeat any portion of your clearance to that controller; simply establish contact with the facility when instructed to do so by the tower controller. The terminal facility computer will pick up your transponder and initiate tracking as soon as it detects the assigned code; for this reason, the transponder should remain on standby until takeoff clearance has been received.

Your aircraft will appear on the controller's radar as a target with an associated data block that moves as your aircraft moves through the airspace. The data block includes aircraft identification, aircraft type, altitude, and airspeed.

Figure 4 – Combined radar and beacon antenna

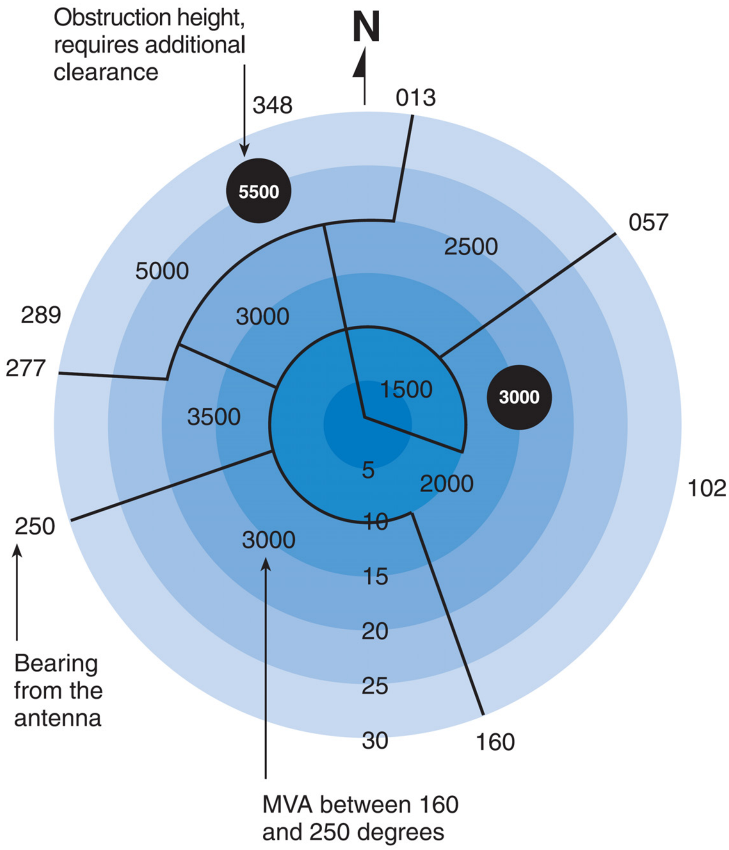


Figure 5 – Minimum Vectoring Altitude Chart

A TRACON controller uses Airport Surveillance Radar (ASR) to detect primary targets and Automated Radar Terminal Systems (ARTS) to receive transponder signals; the two are combined on the controller's scope.

Figure 6 – What the approach controller sees, from the A-SMGCS console in tower

At facilities with ASR-3 equipment, radar returns from precipitation are not displayed as varying levels of intensity, and controllers must rely on pilot reports and experience to provide weather avoidance information. With ASR-9 equipment, the controller can select up to six levels of intensity. Level 1 precipitation does not require avoidance tactics, but the presence of levels 2 or 3 should cause pilots to investigate further. The returns from higher levels of intensity may obscure aircraft data blocks, and controllers may select the higher levels only on pilot request. When you are uncertain about the weather ahead, ask the controller if the facility can display intensity levels—pilots of small aircraft should avoid intensity levels 3 or higher.

3.4.4 TOWER EN ROUTE CONTROL (TEC)

At many locations, instrument flights can be conducted entirely in terminal airspace. These TEC routes are generally for aircraft operating below 10,000 feet, and they can be found in the A/FD. Pilots desiring to use TEC should include that designation in the remarks section of the flight plan. Pilots are not limited to the major airports at the city pairs listed in the A/FD. A valuable service provided by the automated radar equipment at terminal radar facilities is the Minimum Safe Altitude Warnings (MSAW). This equipment predicts your aircraft's position in 2 minutes based on present path of flight—the controller will issue a safety alert if the projected path will encounter terrain or an obstruction. An unusually rapid descent rate on a nonprecision approach can trigger such an alert.

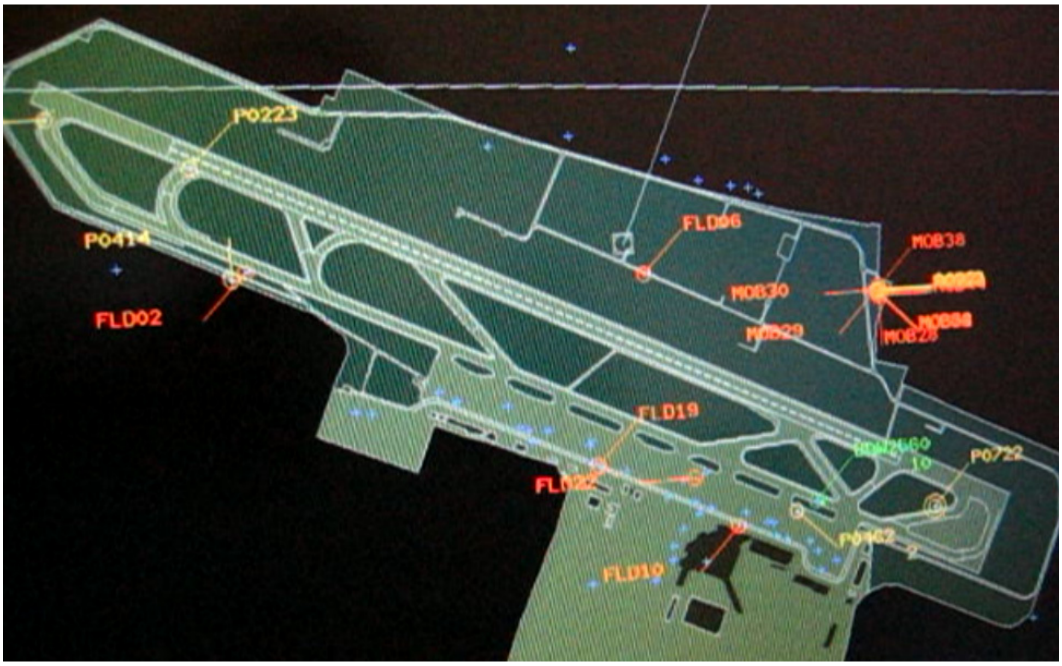


Figure 7 – Surface Movement Radar Display, Integrated with Approach and Multilateration

3.4.5 AIR ROUTE TRAFFIC CONTROL CENTRES (ARTCC)

Air route traffic control centre facilities are responsible for maintaining separation between IFR flights in the en route structure. Centre radars (Air Route Surveillance Radar) acquire and track transponder returns using the same basic technology as terminal radars.

Earlier Centre radars display weather as an area of slashes (light precipitation) and H's (moderate rainfall). Because the controller cannot detect higher levels of precipitation, pilots should be wary of areas showing moderate rainfall. Newer radar displays show weather as three levels of blue. Controllers can select the level of weather to be displayed. Weather displays of higher levels of intensity can make it difficult for controllers to see aircraft data blocks, so pilots should not expect ATC to keep weather displayed continuously.

Centre airspace is divided into sectors in the same manner as terminal airspace; additionally, most Centre airspace is divided by altitudes into high and low sectors. Each sector has a dedicated team of controllers and a selection of radio frequencies, because each Centre has a network of remote transmitter/receiver sites. You will find all Centre frequencies in the back of the A/FD; they are also found on en route charts.

Each ARTCC's area of responsibility covers several states; as you fly from the vicinity of one remote communication site toward another, expect the same controller to talk to you on different frequencies.

3.4.6 CENTRE APPROACH/DEPARTURE CONTROL

The majority of airports with instrument approaches do not lie within terminal radar airspace, and when operating to or from these airports pilots will communicate directly with the Centre controller. Imagine to be a pilot. If you are departing a tower-controlled airport, the tower controller will provide instructions for contacting the appropriate Centre controller. When you depart an airport without an operating control tower, your clearance will include instructions such as "Upon entering controlled airspace, contact Houston Centre on 126.5." Pilots are responsible for terrain clearance until they reach the controller's MVA. Simply hearing "Radar contact" is not sufficient to relieve pilots of this responsibility.

If obstacles in the departure path require a steeper-than standard climb gradient (200 feet per NM), pilots should be so advised by the controller. However, they should check the departure airport listing in the A/FD to determine if there are trees or wires in the departure path just to be sure; when in doubt, ask the controller for the required climb gradient.

A common clearance in these situations is "When able, proceed direct to the Astoria VOR…" The words "when able" mean to proceed when you can do so while maintaining terrain and obstruction clearance—they do not mean to proceed as soon as a signal suitable for navigation is received from the NAVAID. Using the standard climb gradient, you will be 2 miles from the departure end of the runway before it is safe to turn (400 feet above ground level (AGL)). When a Centre controller issues a heading, a direct route, or says "direct when able," the controller becomes responsible for terrain and obstruction clearance.

Another common Centre clearance is "Leaving (altitude) fly (heading) or proceed direct when able." This keeps the terrain/obstruction clearance responsibility in the cockpit until above the minimum IFR altitude. A controller cannot issue an IFR clearance until you are above the minimum IFR altitude unless you are able to climb in VFR conditions.

On a Centre controller's scope, 1 NM is about 1/28 of an inch; when a Centre controller is providing Approach/Departure control services at an airport many miles from the radar antenna, estimating headings and distances is very difficult. Controllers providing vectors to final must set the range on their scopes to not more than 125 NM; this is to provide the greatest possible accuracy for intercept headings. Accordingly, at locations more distant from a Centre radar antenna, pilots should expect a minimum of vectoring.

4 CONTROL SEQUENCE

The IFR system is flexible and accommodating if you have done your homework, have as many frequencies as possible written down before they are needed, and have an alternate in mind if your flight cannot be completed as planned. Pilots always have to know where the nearest VFR conditions can be found, and be prepared to head in that direction if your situation deteriorates.

A typical IFR flight, with departure and arrival at airports with control towers, would use the ATC facilities and services in the following sequence:

1. AFSS: Obtain a weather briefing for departure, destination and alternate airports, and en route conditions, then pilots have to file their flight plan.
2. ATIS: Pre-flight complete, listen for present conditions and the approach in use.
3. Clearance Delivery: Prior to taxiing, obtain departure clearance.
4. Ground Control: Noting that you are IFR, receive taxi instructions.
5. Tower: Pre-takeoff checks complete, receive clearance to takeoff.
6. Departure Control: Once the transponder "tags up" with the ARTS, the tower controller will instruct to contact Departure to establish radar contact.
7. ARTCC: After departing the departure controller's airspace, aircrafts will be handed off to Centre who will coordinate the flight while en route. You may be in contact with multiple ARTCC facilities; they will coordinate the hand-offs.
8. EFAS/HIWAS: Coordinate with ATC before leaving their frequency to obtain in-flight weather information.

9. ATIS: Coordinate with ATC before leaving their frequency to obtain ATIS information.
10. Approach Control: Centre will hand you off to approach control where you will receive additional information and clearances.
11. Tower: Once cleared for the approach, pilots will be instructed to contact tower control; the flight plan will be cancelled by the tower controller upon landing.

4.1 IFR FLIGHT

In addition to current IFR en route charts, area charts, and United States (U.S.) Terminal Procedures Publications (TPP) published by the National Aeronautical Charting Office (NACO), the Federal Aviation Administration (FAA) publishes the Aeronautical Information Manual (AIM), the Airport/Facility Directory (A/FD), and the Notices to Airmen Publication (NTAP) for flight planning in the National Airspace System (NAS). Pilots should also consult the Pilot's Operating Handbook/Airplane Flight Manual (POH/AFM) for flight planning information pertinent to the aircraft to be flown.

4.1.1 AERONAUTICAL INFORMATION MANUAL (AIM)

The AIM provides the aviation community with basic flight information and air traffic control (ATC) procedures used in the U.S. NAS. An international version called the Aeronautical Information Publication contains parallel information, as well as specific information on the international airports used by the international community.

4.1.2 AIRPORT/FACILITY DIRECTORY (A/FD)

The A/FD contains information on airports, communications, and navigation aids pertinent to IFR flight. It also includes very-high frequency omnidirectional range (VOR) receiver checkpoints, automated flight service station (AFSS), weather service telephone numbers, and air route traffic control centre (ARTCC) frequencies. Various special notices essential to IFR flight are also included, such as land and hold short (LAHSO) data, the civil use of military fields, continuous power facilities, and special flight procedures.

In the major terminal and en route environments, preferred routes have been established to guide pilots in planning their routes of flight, to minimize route changes, and to aid in the orderly management of air traffic using the federal airways. The A/FD lists both high and low altitude preferred routes.

Communications Facility	Description	Frequency
Airport Advisory Area "[AFSS name] RADIO"	AFSS personnel provide traffic advisories to pilots operating within 10 miles of the airport.	123.6 MHz.
UNICOM "[airport name] UNICOM"	Airport advisories from an airport without an operating control tower or AFSS.	Listed in A/FD under the city name; also on sectional charts in airport data block.
Air Route Traffic Control Center (ARTCC) "CENTER"	En route radar facilities that maintain separation between IFR flights, and between IFR flights and known VFR flights. Centers will provide VFR traffic advisories on a workload permitting basis.	Listed in A/FD, and on instrument en route charts.
Approach/Departure Control "[airport name] APPROACH" (unless otherwise advised)	Positions at a terminal radar facility responsible for handling of IFR flights to and from the primary airport (where Class B airspace exists).	Listed in A/FD; also on sectional charts in the communications panel, and on terminal area charts.
Automatic Terminal Information Service (ATIS)	Continuous broadcast of audio tape prepared by ATC controller containing wind direction and velocity, temperature, altimeter setting, runway and approach in use, and other information of interest to pilots.	Listed in A/FD under the city name; also on sectional charts in airport data block and in the communications panel, and on terminal area charts.
Clearance Delivery "[airport name] CLEARANCE"	Control tower position responsible for transmitting departure clearances to IFR flights.	Listed on instrument approach procedure charts.
Common Traffic Advisory Frequency (CTAF)	CTAF provides a single frequency for pilots in the area to use for contacting the facility and/or broadcasting their position and intentions to other pilots.	Listed in A/FD; also on sectional charts in the airport data block (followed by a white C on a blue or magenta background). At airports with no tower, CTAF is 122.9, the "MULTICOM" frequency.
Automated Flight Service Station (AFSS) "[facility name] RADIO"	Provides information and services to pilots, using remote communications outlets (RCOs) and ground communications outlets (GCOs).	Listed in A/FD and sectional charts, both under city name and in a separate listing of AFSS frequencies. On sectional charts, listed above the VOR boxes, or in separate boxes when remote.
Ground Control "[airport name] GROUND"	At tower-controlled airports, a position in the tower responsible for controlling aircraft taxiing to and from the runways.	Listed in A/FD under city name.
Hazardous Inflight Weather Advisory Service (HIWAS)	Continuous broadcast of forecast hazardous weather conditions on selected NAVAIDs. No communication capability.	Black circle with white "H" in VOR frequency box; notation in A/FD airport listing under "Radio Aids to Navigation."
MULTICOM "[airport name] TRAFFIC"	Intended for use by pilots at airports with no radio facilities. Pilots should use self-announce procedures given in the AIM.	122.9 MHz. A/FD shows 122.9 as CTAF; also on sectional charts 122.9 is followed by a white C on a dark background, indicating CTAF.
Tower "[airport name] TOWER"	"Local" controller responsible for operations on the runways and in Class B, C, or D airspace surrounding the airport.	Listed in A/FD under city name; also on sectional and terminal control area charts in the airport data block and communications panel.
En Route Flight Advisory Service (EFAS) "FLIGHT WATCH"	For in-flight weather information.	122.0 MHz (0600-2200 local time)

Table 2 – ATC facilities, services, and radio call signs

4.1.3 IFR FLIGHT PLAN

As specified in Title 14 of the Code of Federal Regulations (14 CFR) part 91, no person may operate an aircraft in controlled airspace under IFR unless that person has filed an IFR flight plan. Flight plans may be submitted to the nearest AFSS or air traffic control tower (ATCT) either in person, by telephone, by computer (using the direct user access terminal system (DUATS)), or by radio if no other means are available. Pilots should file IFR flight plans at least 30 minutes prior to estimated time of departure to preclude possible delay in receiving a departure clearance from ATC. The AIM provides guidance for completing and filing FAA Form 7233-1, Flight Plan. These forms are available at flight service stations (FSS's), and are generally found in flight planning rooms at airport terminal buildings. Operating on an IFR flight plan to an airport with an operating control tower, your flight plan is cancelled automatically upon landing.

4.1.4 CLEARANCES

An ATC clearance allows an aircraft to proceed under specified traffic conditions within controlled airspace for the purpose of providing separation between known aircraft. Once the clearance is accepted, you are required to comply with ATC instructions. You may request a clearance different from that issued if you consider another course of action more practicable or if your aircraft equipment limitations or other considerations make acceptance of the clearance inadvisable.
Pilots should also request clarification or amendment, as appropriate, any time a clearance is not fully understood or considered unacceptable because of safety of flight. The pilot is responsible for requesting an amended clearance if ATC issues a clearance that would cause a pilot to deviate from a rule or regulation or would place the aircraft in jeopardy.

Clearance Separations: ATC will provide the pilot on an IFR clearance with separation from other IFR traffic. This separation is provided:
1. Vertically—by assignment of different altitudes.
2. Longitudinally—by controlling time separation between aircraft on the same course.
3. Laterally—by assignment of different flight paths.
4. By radar—including all of the above.

ATC does not provide separation for an aircraft operating:
1. Outside controlled airspace;
2. On an IFR clearance:
a. With "VFR-On-Top" authorized instead of a specific assigned altitude.
b. Specifying climb or descent in "VFR conditions."
c. At any time in VFR conditions, since uncontrolled VFR flights may be operating in the same airspace.

In addition to heading and altitude assignments, ATC will occasionally issue speed adjustments to maintain the required separations. At times, ATC may also employ visual separation techniques to keep aircraft safely separated. A pilot who obtains visual contact with another aircraft may be asked to maintain visual separation or to follow the aircraft.

In the absence of radar contact, ATC will rely on position reports to assist in maintaining proper separation. Using the data transmitted by the pilot, the controller follows the progress of your flight. ATC must correlate your reports with all the others to provide separation; therefore, the accuracy of your reports can affect the progress and safety of every other aircraft operating in the area on an IFR flight plan.

4.1.5 Departure Procedures (DPs)

DPs are designed to expedite clearance delivery, to facilitate transition between takeoff and en route operations, and to ensure adequate obstacle clearance. They furnish pilots' departure routing clearance information in both graphic and textual form. To simplify clearances, DPs have been established for the most frequently used departure routes in areas of high traffic activity. A DP will normally be used where such departures are available, since this is advantageous to both users and ATC.

4.1.6 Radar Controlled Departures

On IFR departures from airports in congested areas, one will normally receive navigational guidance from departure control by radar vector. When departure is to be vectored immediately following takeoff, pilots will be advised before takeoff of the initial heading to be flown. This information is vital in the event one may experience a loss of two way radio communications during departure. The radar departure is normally simple. Following takeoff, you contact departure control on the assigned frequency when advised to do so by the control tower. At this time departure control verifies radar contact, and gives headings, altitude, and climb instructions to move you quickly and safely out of the terminal area. Fly the assigned headings and altitudes until the controller tells you your position with respect to the route given in your clearance, whom to contact next, and to "resume your own navigation". Departure control will vector you to either a navigation facility or an en route position appropriate to your departure clearance, or you will be transferred to another controller with further radar surveillance capabilities.

A radar controlled departure does not relieve you of your responsibilities as pilot in command. You should be prepared before takeoff to conduct your own navigation according to your ATC clearance, with navigation receivers checked and properly tuned. While under radar control, monitor your instruments to ensure that you are continuously oriented to the route specified in your clearance, and record the time over designated checkpoints.

4.1.7 En Route Procedures

Procedures en route will vary according to the proposed route, the traffic environment, and the ATC facilities controlling the flight. Some IFR flights are under radar surveillance and controlled from departure to arrival and others rely entirely on pilot navigation.

Where ATC has no jurisdiction, it does not issue an IFR clearance. It has no control over the flight; nor does the pilot have any assurance of separation from other traffic.

4.1.8 ATC REPORTS

All pilots are required to report unforecast weather conditions or other information related to safety of flight to ATC.

4.1.9 POSITION REPORTS

Unless in radar contact with ATC, pilots are required to furnish a position report over certain reporting points. Position reports are required over each compulsory reporting point (shown on the chart as solid triangle figures) along the route being flown regardless of altitude, including those with a VFR-On-Top clearance. Along direct routes, reports are required of all IFR flights over each point used to define the route of flight. Reports at reporting points (shown as open triangle figures) are made only when requested by ATC.

Position reports should include the following items:

1. Identification.
2. Position.
3. Type of flight plan, if your report is made to an AFSS.
4. The estimated time of arrival (ETA) over next reporting point.
5. The name only of the next succeeding (required) reporting point along the route of flight.
6. Pertinent remarks.

En route position reports are submitted normally to the ARTCC controllers via direct controller-to-pilot communications channels, using the appropriate ARTCC frequencies listed on the en route chart.

Whenever an initial Centre contact is to be followed by a position report, the name of the reporting point should be included in the call-up. This alerts the controller that such information is forthcoming.

4.1.10 PLANNING THE DESCENT AND APPROACH

ATC arrival procedures and cockpit workload are affected by weather conditions, traffic density, aircraft equipment, and radar availability.

When landing at airports with approach control services and where two or more IAPs are published, you will be provided in advance of arrival with information on the type of approach to expect or if you will be vectored for a visual approach. This information will be broadcast either on automated terminal information service (ATIS) or by a controller. It will not be furnished when the visibility is 3 miles or better and the ceiling is at or above the highest initial approach altitude established for any low altitude IAP for the airport.

The purpose of this information is to help you in planning arrival actions; however, it is not an ATC clearance or commitment and is subject to change. Fluctuating weather, shifting winds, blocked runway, etc., are conditions that may result in changes to the approach information you previously received. It is important to advise ATC immediately if you are unable to execute the approach, or if you prefer, another type of approach.

4.1.11 STANDARD TERMINAL ARRIVAL ROUTES (STARS)

Standard terminal arrival routes have been established to simplify clearance delivery procedures for arriving aircraft at certain areas having high density traffic. A STAR serves a purpose parallel to that of a DP for departing traffic.

The following points regarding STARs are important to remember:

1. All STARs are contained in the TPP, along with the IAP charts for the destination airport. The AIM also describes STAR procedures.
2. If the destination is a location for which STARs have been published, pilots may be issued a clearance containing
a STAR whenever ATC deems it appropriate. Pilots must possess at least the approved textual description.
3. It is pilot's responsibility to either accept or refuse an issued STAR. If you do not wish to use a STAR, you should advise ATC by placing "NO STAR" in the remarks section of your filed flight plan or by advising ATC.
4. If you accept a STAR in your clearance, you must comply with it.

4.1.12 APPROACHES

Compliance with Published Standard Instrument Approach Procedures shown on the approach charts provides necessary navigation guidance information for alignment with the final approach courses, as well as obstruction clearance. Under certain conditions, a course reversal manoeuvre or procedure turn may be necessary. However, this procedure is not authorized when:

1. The symbol "NoPT" appears on the approach course on the plan view of the approach chart.
2. Radar vectoring is provided to the final approach course.
3. A holding pattern is published in lieu of a procedure turn.
4. Executing a timed approach from a holding fix.
5. Otherwise directed by ATC.

4.1.13 INSTRUMENT APPROACHES TO CIVIL AIRPORTS

Unless otherwise authorized, when an instrument letdown to an airport is necessary, pilots should use a standard IAP prescribed for that airport. IAPs are depicted on IAP charts and are found in the TPP.

ATC approach procedures depend upon the facilities available at the terminal area, the type of instrument approach executed, and the existing weather conditions. The ATC facilities, navigation aids (NAVAIDs), and associated frequencies appropriate to each standard instrument approach are given on the approach chart. Individual charts are published for standard approach procedures associated with the following types of facilities:

1. Non-directional beacon (NDB)
2. Very-high frequency omni-directional range (VOR)
3. Very-high frequency omni-directional range with distance measuring equipment (VORTAC or VOR/DME)

4. Localizer (LOC)
5. Instrument landing system (ILS)
6. Localizer-type directional aid (LDA)
7. Simplified directional facility (SDF)
8. Area navigation (RNAV)
9. Global positioning system (GPS)

An IAP can be flown in one of two ways: as a full approach or with the assistance of radar vectors. When the IAP is flown as a full approach, pilots conduct their own navigation using the routes and altitudes depicted on the instrument approach chart. A full approach allows the pilot to transition from the en route phase, to the instrument approach, and then to a landing with minimal assistance from ATC. This type of procedure may be requested by the pilot but is most often used in areas without radar coverage. A full approach also provides the pilot with a means of completing an instrument approach in the event of a communications failure.

When an approach is flown with the assistance of radar vectors, ATC provides guidance in the form of headings and altitudes which positions the aircraft to intercept the final approach. From this point, the pilot resumes navigation, intercepts the final approach course, and completes the approach using the IAP chart. This is often a more expedient method of flying the approach, as opposed to the full approach, and allows ATC to sequence arriving traffic. A pilot operating in radar contact can generally expect the assistance of radar vectors to the final approach course.

4.1.14 Approach to an Airport With an Operating Tower, With an Approach Control

Where radar is approved for approach control service, it is used to provide vectors in conjunction with published IAPs. Radar vectors can provide course guidance and expedite traffic to the final approach course of any established IAP.

Approach control facilities that provide this radar service operate in the following manner:

1. Arriving aircraft are either cleared to an outer fix most appropriate to the route being flown with vertical separation and, if required, given holding information; or,
2. When radar hand-offs are effected between ARTCC and approach control, or between two approach control facilities, aircraft are cleared to the airport, or to a fix so located that the hand-off will be completed prior to the time the aircraft reaches the fix.
a. When the radar hand-offs are utilized, successive arriving flights may be handed-off to approach control with radar separation in lieu of vertical separation.
b. After hand-off to approach control, aircraft are vectored to the appropriate final approach course.
3. Radar vectors and altitude/flight levels will be issued as required for spacing and separating aircraft; therefore, you must not deviate from the headings issued by approach control.
4. You will normally be informed when it becomes necessary to vector you across the final approach course for spacing or other reasons. If you determine that approach course crossing is imminent and you have not been informed that you will be vectored across it, you should question the controller. You should not turn inbound on the final approach course unless you have received an approach clearance. Approach control will normally issue this clearance with the final vector for interception of the final approach course, and the vector will be such as to enable you to establish your aircraft on the final approach course prior to reaching the final approach fix. In the event you are already inbound on the final approach course, you will be issued approach clearance prior to reaching the final approach fix.

5. After you are established inbound on the final approach course, radar separation will be maintained between you and other aircraft, and you will be expected to complete the approach using the NAVAID designated in the clearance (ILS, VOR, NDB, GPS, etc.) as the primary means of navigation.
6. After passing the final approach fix inbound, you are expected to proceed direct to the airport and complete the approach, or to execute the published missed approach procedure.
7. Radar service is automatically terminated when the landing is completed or the tower controller has your aircraft in sight, whichever occurs first.

4.1.15 RADAR APPROACHES

With a radar approach, the pilot is "talked down" while a controller monitors the progress of the flight with radar. This is an option should the pilot experience an emergency or distress situation. These approaches require a radar facility and a functioning airborne radio.

Initial radar contact for either a surveillance or precision approach radar (PAR) is made with approach control. Pilots must comply promptly with all instructions when conducting either type of procedure. They can determine the radar approach facilities (surveillance and/or precision) available at a specific airport by referring to the appropriate En route Low Altitude Chart and IAP chart. Surveillance and precision radar minimums are listed alphabetically by airport on pages with the heading, "Radar Instrument Approach Minimums", in each TPP. Note that both straight-in and circling minimums are listed.

When an instrument approach is being radar monitored, the radar advisories serve only as a secondary aid. Since pilots would have selected a NAVAID such as the ILS or VOR as the primary aid for the approach, the minimums listed on the approach chart apply.

4.1.16 SURVEILLANCE APPROACH

On an airport surveillance radar approach (ASR), the controller will vector the pilots to a point where they can begin a descent to the airport or to a specific runway. During the initial part of the approach, they will be given communications failure/missed approach instructions. Before they begin their descent, the controller will give them the published straight-in minimum descent altitude (MDA). Pilots will not be given the circling MDA unless they request it and tell the controller their aircraft category.

During the final approach, the controller will provide navigational guidance in azimuth only. Guidance in elevation is not possible, but you will be advised when to begin descent to the MDA, or if appropriate, to the intermediate "step-down fix" MDA and subsequently to the prescribed MDA. In addition, you will be advised of the location of the missed approach point (MAP) and your position each mile from the runway, airport, or MAP as appropriate. If you so request, the controller will issue recommended altitudes each mile, based on the descent gradient established for the procedure, down to the last mile that is at or above the MDA.

Pilots will normally be provided navigational guidance until they reach the MAP. The controller will terminate guidance and instruct you to execute a missed approach at the MAP, if at that point they do not have the runway or airport in sight, or if they are on a point-in-space approach in a helicopter, the prescribed visual reference with the surface is not established. If at any time during the approach the controller considers that safe guidance cannot be provided for the remainder of the approach, the approach will be terminated, and pilots will be instructed to execute a missed approach. Guidance termination and missed approach will be effected upon pilot request, and the controller may terminate guidance when the pilot reports the runway, airport/heliport, or visual surface route (point-in-space approach) in sight or otherwise indicates that continued guidance is not required. Radar service is automatically terminated at the completion of the radar approach.

4.1.17 PRECISION APPROACH

The installations that have PAR are joint civil/military airports and usually provide service to civilian pilots flying IFR only with prior permission, except in an emergency.

A PAR serves the same purpose as an ILS, except that guidance information is presented to the pilot through aural rather than visual means. If a PAR is available, it is normally aligned with an ILS. During a PAR approach, pilots are provided highly accurate guidance in both azimuth and elevation.

The precision approach begins when the aircraft is within range of the precision radar and contact has been established with the PAR controller. Normally this occurs approximately 8 miles from touchdown, a point to which aircrafts are vectored by surveillance radar or are positioned by a non-radar approach procedure. Pilots will be given headings to fly, to direct them to, and to keep their aircraft aligned with, the extended centreline of the landing runway.

Before intercepting the glidepath, you pilots be advised of communications failure/missed approach procedures and told not to acknowledge further transmissions. Pilots will be told to anticipate glidepath interception approximately 15 to 30 seconds before it occurs and when to start their descent. The published decision altitude/decision height (DA/DH) will be given only if you request it.

During the final approach, the controller will give elevation information as "slightly/well above" or " slightly/well below" glidepath, and course information as "slightly/well right" or "slightly/well left" of course. Extreme accuracy in maintaining and correcting headings and rate of descent is essential. The controller will assume the last assigned heading is being maintained and will base further corrections on this assumption. Range from touchdown is given at least once each mile. If the aircraft is observed by the controller to proceed outside of specified safety zone limits in azimuth and/or elevation and continue to operate outside these prescribed limits, the pilots will be directed to execute a missed approach or to fly a specified course unless they have the runway environment in sight. Pilots will be provided navigational guidance in azimuth and elevation to the DA/DH. Advisory course and glidepath information will be furnished by the controller until the aircraft passes over the runway threshold, at which point they will be advised of any deviation from the runway centreline. Radar service is automatically terminated at the completion of the approach.

4.1.18 APPROACHES TO PARALLEL RUNWAYS

Procedures permit ILS instrument approach operations to dual or triple parallel runway configurations. Parallel approaches are an ATC procedure that permits parallel ILS approaches to airports with parallel runways separated by at least 2,500 feet between centrelines. Wherever parallel approaches are in progress, pilots are informed that approaches to both runways are in use.
Simultaneous approaches are permitted to runways:
1. With centrelines separated by 4,300 to 9,000 feet;
2. That are equipped with final monitor controllers;
3. That require radar monitoring to ensure separation between aircraft on the adjacent parallel approach course.

The approach procedure chart will include the note i.e. "simultaneous approaches authorized RWYS 14L and 14R", identifying the appropriate runways. When advised that simultaneous parallel approaches are in progress, pilots must advise approach control immediately of malfunctioning or inoperative components.

Parallel approach operations demand heightened pilot situational awareness. The close proximity of adjacent aircraft conducting simultaneous parallel approaches mandates strict compliance with all ATC clearances and approach procedures. Pilots should pay particular attention to the following approach chart information: name and number of the approach, localizer frequency, inbound course, glide-slope intercept altitude, DA/DH, missed approach instructions, special notes/procedures, and the assigned runway location and proximity to adjacent runways. Pilots also need to exercise strict radio discipline, which includes continuous monitoring of communications and the avoidance of lengthy, unnecessary radio transmissions.

4.1.19 SIDE-STEP MANOEUVRE

ATC may authorize a side-step manoeuvre to either one of two parallel runways that are separated by 1,200 feet or less, followed by a straight-in landing on the adjacent runway. Aircraft executing a side-step manoeuvre will be cleared for a specified nonprecision approach and landing on the adjacent parallel runway. For example, "Cleared ILS runway 7 left approach, side-step to runway 7 right." Pilots are expected to commence the side-step manoeuvre as soon as possible after the runway or runway environment is in sight. Landing minimums to the adjacent runway will be based on nonprecision criteria and therefore higher than the precision minimums to the primary runway, but will normally be lower than the published circling minimums.

4.1.20 CIRCLING APPROACHES

Landing minimums are listed on the approach chart under "CIRCLING." Circling minimums apply when it is necessary to circle the airport or manoeuvre for landing, or when no straight-in minimums are specified on the approach chart like in Figure 8.

The circling minimums published on the instrument approach chart provide a minimum of 300 feet of obstacle clearance in the circling area. During a circling approach, pilots should maintain visual contact with the runway of intended landing and fly no lower than the circling minimums until they are in position to make a final descent for a landing. Remember—circling minimums are just that—minimums. If the ceiling allows it, it is better to fly at an altitude that more nearly approximates the VFR traffic pattern altitude. This will make any manoeuvring safer and bring the view of the landing runway into a more normal perspective.

Figure 9 shows patterns that can be used for circling approaches. Pattern "A" can be flown when your final approach course intersects the runway centreline at less than a 90° angle, and you sight the runway early enough to establish a base leg. If you sight the runway too late to fly pattern "A," you can circle as shown in "B." You can fly pattern "C" if it is desirable to land opposite the direction of the final approach, and the runway is sighted in time for a turn to downwind leg. If the runway is sighted too late for a turn to downwind, you can fly pattern "D." Regardless of the pattern flown, the pilot must manoeuvre the aircraft so as to remain within the designated circling area.

Sound judgment and knowledge of pilots' capabilities and the performance of their aircraft are the criteria for determining the pattern to be flown in each instance, since they must consider all factors: airport design, ceiling and visibility, wind direction and velocity, final approach course alignment, distance from the final approach fix to the runway, and ATC instructions.

Approach Category	Radius (miles)
A	1.3
B	1.5
C	1.7
D	2.3
E	4.5

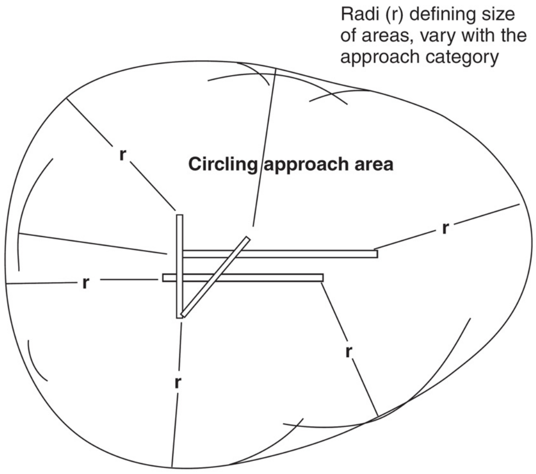


Figure 8 – Circling approach area radii

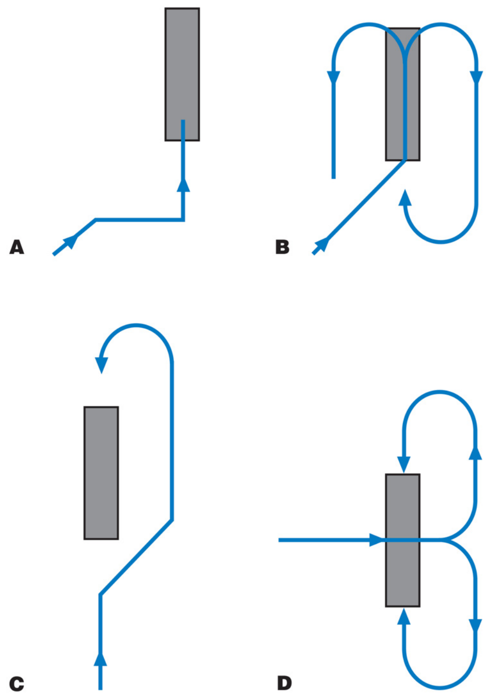


Figure 9 – Circling approaches

4.1.21 IAP MINIMUMS

Pilots may not operate an aircraft at any airport below the authorized MDA or continue an approach below the authorized DA/DH unless:

1. The aircraft is continuously in a position from which a descent to a landing on the intended runway can be made at a normal descent rate using normal manoeuvres;

2. The flight visibility is not less than that prescribed for the approach procedure being used; and
3. At least one of the following visual references for the intended runway is visible and identifiable to the pilot:
a. Approach light system
b. Threshold
c. Threshold markings
d. Threshold lights
e. Runway end identifier lights (REIL)
f. Visual approach slope indicator (VASI)
g. Touchdown zone or touchdown zone markings
h. Touchdown zone lights
i. Runway or runway markings
j. Runway lights

4.1.22 MISSED APPROACHES

A missed approach procedure is formulated for each published instrument approach and allows the pilot to return to the airway structure while remaining clear of obstacles. The procedure is shown on the approach chart in text and graphic form. Since the execution of a missed approach occurs when your cockpit workload is at a maximum, the procedure should be studied and mastered before beginning the approach. When a MAP is initiated, a climb pitch attitude should be established while setting climb power. Pilots should configure the aircraft for climb, turn to the appropriate heading, advise ATC that they are executing a missed approach, and request further clearances. If the missed approach is initiated prior to reaching the MAP, unless otherwise cleared by ATC, pilots should continue to fly the IAP as specified on the approach plate to the MAP at or above the MDA or DA/DH before beginning a turn.

If visual reference is lost while circling-to-land from an instrument approach, pilots should execute the appropriate MAP. They should make the initial climbing turn toward the landing runway and then manoeuvre to intercept and fly the missed approach course.

Pilots should immediately execute the missed approach procedure:
1. Whenever the requirements for operating below DA/DH or MDA are not met when the aircraft is below MDA, or upon arrival at the MAP and at any time after that until touchdown;
2. Whenever an identifiable part of the airport is not visible to the pilot during a circling manoeuvre at or above MDA;
3. When so directed by ATC.

The missed approach procedures are related to the location of the FAF. When the FAF is not located on the field, the missed approach procedure will specify the distance from the facility to the MAP. The airport diagram on the IAP shows the time from the facility to the missed approach at various groundspeeds, which pilots must determine from airspeed, wind, and distance values. This time determines when they report and execute a missed approach if they do not have applicable minimums. Missed approach instructions will be provided prior to starting the final approach of either an ASR or PAR approach.

4.1.23 LANDING

According to part 91, no pilot may land when the flight visibility is less than the visibility prescribed in the standard IAP being used. ATC will provide the pilot with the current visibility reports appropriate to the runway in use. This may be in the form of prevailing visibility, runway visual value (RVV), or runway visual range (RVR). However, only the pilot can determine if the flight visibility meets the landing requirements indicated on the approach chart. If the flight visibility meets the minimum prescribed for the approach, then the approach may be continued to a landing. If the flight visibility is less than that prescribed for the approach, then the pilot must execute a missed approach, regardless of the reported visibility.

The landing minimums published on IAP charts are based on full operation of all components and visual aids associated with the instrument approach chart being used. Higher minimums are required with inoperative components or visual aids. For example, if the approach lighting system were inoperative, the visibility minimums for an ILS must be increased by one-quarter mile. If more than one component is inoperative, each minimum is raised to the highest minimum required by any single component that is inoperative. ILS glide-slope inoperative minimums are published on instrument approach charts as localizer minimums.

SUMMARY OF CHAPTER 2

CHAPTER 3

FUNDAMENTALS OF NAVIGATION SYSTEMS

This chapter provides first of all an introduction to Navigation. Hence the basic radio principles applicable to navigation equipment, as well as an operational knowledge of how to use these systems in instrument flight. This information provides the framework for all instrument procedures, including departure procedures (DPs), holding patterns, and approaches, because each of these manoeuvres consist mainly of accurate attitude instrument flying and accurate tracking using navigation systems. They will be the basis of understanding navigation procedures on the ground. Navigation is the process of piloting an airplane from one geographic position to another while monitoring one's position as the flight progresses. It introduces the need for planning, which includes plotting the course on an aeronautical chart, selecting checkpoints, measuring distances, obtaining pertinent weather information, and computing flight time, headings, and fuel requirements. The methods used in this chapter include pilotage—navigating by reference to visible landmarks, dead reckoning—computations of direction and distance from a known position, and radio navigation—by use of radio aids.

1 POSITION IN SPACE AND TIME

1.1 LATITUDE AND LONGITUDE (MERIDIANS ANS PARALLELS)

The Equator is an imaginary circle equidistant from the poles of the Earth. Circles parallel to the Equator (lines running east and west) are parallels of latitude. They are used to measure degrees of latitude north or south of the Equator. The angular distance from the Equator to the pole is one-fourth of a circle or 90°. The arrows in Figure 1 labelled "LATITUDE" point to lines of latitude. Meridians of longitude are drawn from the North Pole to the South Pole and are at right angles to the Equator. The "Prime Meridian" which passes through Greenwich, England, is used as the zero line from which measurements are made in degrees east and west to 180°. The arrows in Figure 1 labelled "LONGITUDE" point to lines of longitude. Any specific geographical point can thus be located by reference to its longitude and latitude. Washington, DC for example, is approximately 39° N. latitude, 77° W. longitude. Chicago is approximately 42° N. latitude, 88°W. longitude.

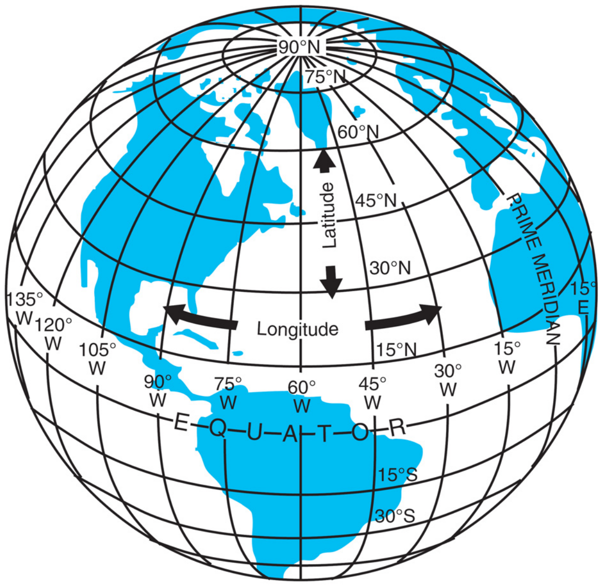


Figure 1 – Meridians and parallels—the basis of measuring time, distance, and direction

1.2 TIME ZONES

The meridians are also useful for designating time zones. A day is defined as the time required for the Earth to make one complete rotation of 360°. Since the day is divided into 24 hours, the Earth revolves at the rate of 15° an hour. Noon is the time when the Sun is directly above a meridian; to the west of that meridian is morning, to the east is afternoon. In most aviation operations, time is expressed in terms of the 24-hour clock. Air traffic control instructions, weather reports and broadcasts, and estimated times of arrival are all based on this system. For example: 9 a.m. is expressed as 0900, 1 p.m. is 1300, and 10 p.m. is 2200. Because a pilot may cross several time zones during a flight, a standard time system has been adopted. It is called Universal Coordinated Time (UTC) and is often referred to as Zulu time. UTC is the time at the 0° line of longitude which passes through Greenwich, England. All of the time zones around the world are based on this reference. For daylight saving time, 1 hour should be subtracted from the calculated times.

1.3 MEASUREMENT OF DIRECTION

By using the meridians, direction from one point to another can be measured in degrees, in a clockwise direction from true north. To indicate a course to be followed in flight, draw a line on the chart from the point of departure to the destination and measure the angle which this line forms with a meridian. Direction is expressed in degrees, as shown by the compass rose in following Figure 2. Because meridians converge toward the poles, course measurement should be taken at a meridian near the midpoint of the course rather than at the point of departure. The course measured on the chart is known as the true course. This is the direction measured by reference to a meridian or true north. It is the direction of intended flight as measured in degrees clockwise from true north.

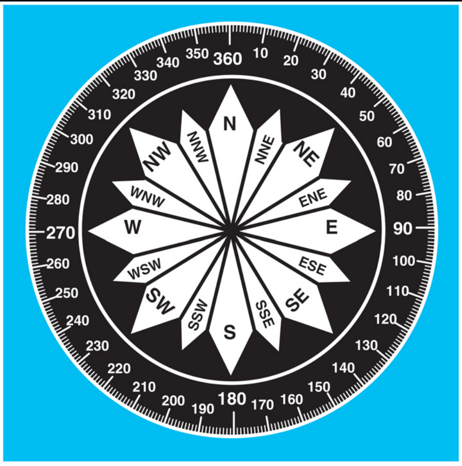


Figure 2 – Compass rose

1.3.1 VARIATION

Variation is the angle between true north and magnetic north [ATC-3]. It is expressed as east variation or west variation depending upon whether magnetic north (MN) is to the east or west of true north (TN). The north magnetic pole is located close to 71° N. latitude, 96° W. longitude and is about 1,300 miles from the geographic or true north pole, as indicated in Figure 3.

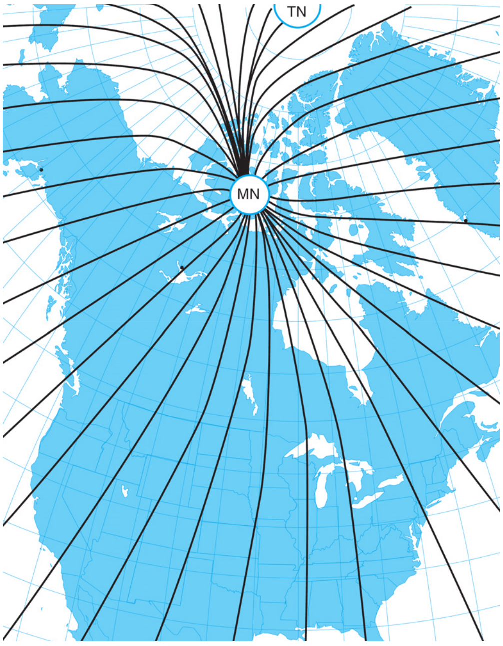


Figure 3 – Isogonic chart

In Figure 3, magnetic meridians are in black; geographic meridians and parallels are in blue. Variation is the angle between a magnetic and geographic meridian. If the Earth were uniformly magnetized, the compass needle would point toward the magnetic pole, in which case the variation between true north (as shown by the geographical meridians) and magnetic north (as shown by the magnetic meridians) could be measured at any intersection of the meridians. Actually, the Earth is not uniformly magnetized. In the United States, the needle usually points in the general direction of the magnetic pole, but it may vary in certain geographical localities by many degrees. Consequently, the exact amount of variation at thousands of selected locations in the United States has been carefully determined. The amount and the direction of variation, which change slightly from time to time, are shown on most aeronautical charts as broken magenta lines, called isogonic lines, which connect points of equal magnetic variation. (The line connecting points at which there is no variation between true north and magnetic north is the agonic line.) An isogonic chart is shown in Figure 4. Minor bends and turns in the isogonic and agonic lines are caused by unusual geological conditions affecting magnetic forces in these areas. Because courses are measured in reference to geographical meridians which point toward true north, and these courses are maintained by reference to the compass which points along a magnetic meridian in the general direction of magnetic north, the true direction must be converted into magnetic direction for the purpose of flight. This conversion is made by adding or subtracting the variation which is indicated by the nearest isogonic line on the chart. The true heading, when corrected for variation, is known as magnetic heading.

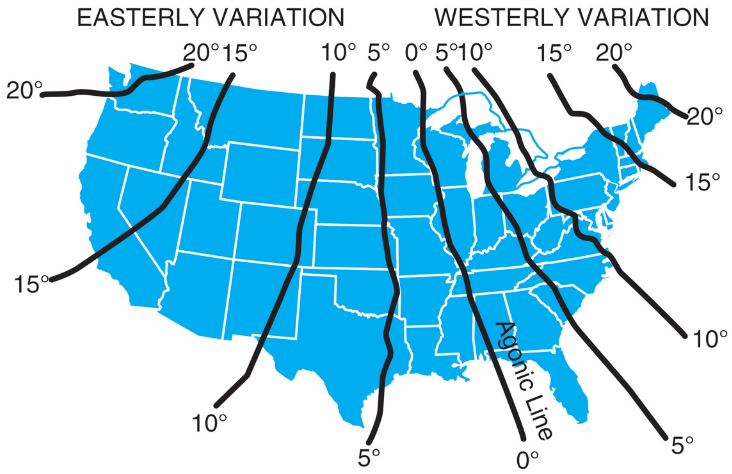


Figure 4 – A typical isogonic chart

In Figure 4, the black lines are isogonic lines which connect geographic points with identical magnetic variation. Remember, to convert true course or heading to magnetic course or heading, note the variation shown by the nearest isogonic line. If variation is west, add; if east, subtract. One method for remembering whether to add or subtract variation is the phrase "east is least (subtract) and west is best (add)."

1.3.2 DEVIATION

Determining the magnetic heading is an intermediate step necessary to obtain the correct compass heading for the flight. To determine compass heading, a correction for deviation [ATC-3] must be made. Because of magnetic influences within the airplane such as electrical circuits, radio, lights, tools, engine, and magnetized metal parts, the compass needle is frequently deflected from its normal reading. This deflection is deviation. The deviation is different for each airplane, and it also may vary for different headings in the same airplane. For instance, if magnetism in the engine attracts the north end of the compass, there would be no effect when the plane is on a heading of magnetic north. On easterly or westerly headings, however, the compass indications would be in error, as shown in Figure 5.

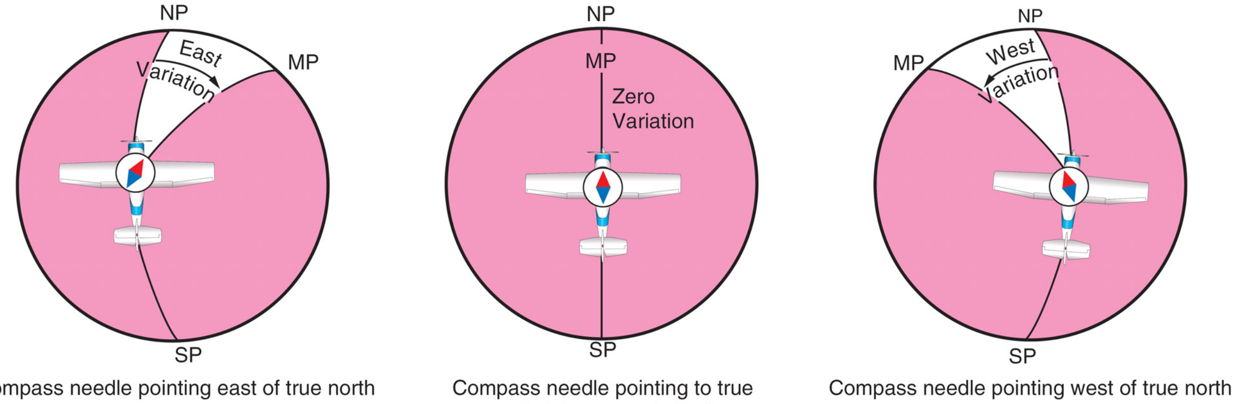


Figure 5 – Effect of variation on the compass

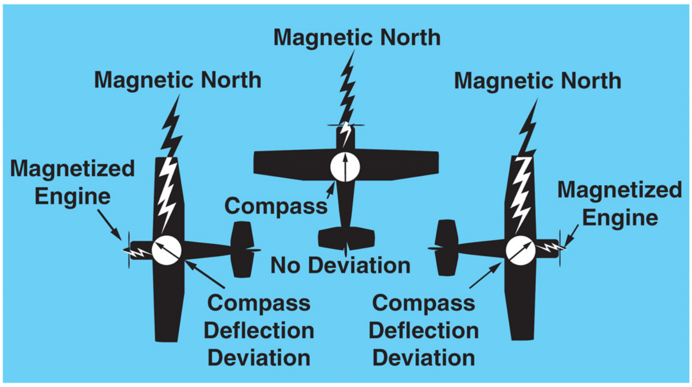


Figure 6 – Magnetized portions of the airplane cause the compass to deviate from its normal indications

Magnetic attraction can come from many other parts of the airplane; the assumption of attraction in the engine is merely used for the purpose of illustration. Some adjustment of the compass, referred to as compensation, can be made to reduce this error, but the remaining correction must be applied by the pilot. Proper compensation of the compass is best performed by a competent technician. Since the magnetic forces within the airplane change, because of landing shocks, vibration, mechanical work, or changes in equipment, the pilot should occasionally have the deviation of the compass checked. The procedure used to check the deviation (called "swinging the compass") is briefly outlined. The airplane is placed on a magnetic compass rose, the engine started, and electrical devices normally used (such as radio) are turned on. Tail wheel-type airplanes should be jacked up into flying position. The airplane is aligned with magnetic north indicated on the compass rose and the reading shown on the compass is recorded on a deviation card. The airplane is then aligned at 30° intervals and each reading is recorded. If the airplane is to be flown at night, the lights are turned on and any significant changes in the readings are noted. If so, additional entries are made for use at night. The accuracy of the compass can also be checked by comparing the compass reading with the known runway headings.

The following method is used by many pilots to determine compass heading: After the true course (TC) is measured, and wind correction applied resulting in a true heading (TH), the sequence TH ± variation (V) = MH ± deviation (D) = compass heading (CH) is followed to arrive at compass heading.

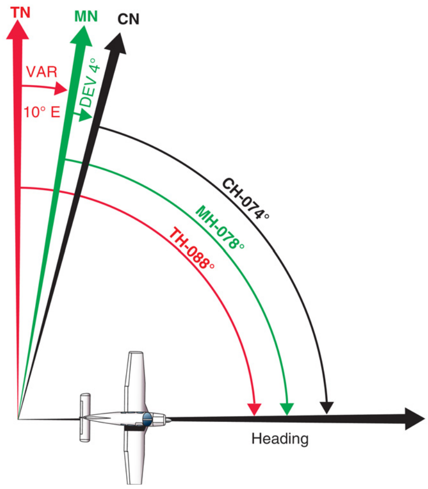


Figure 7 – Relationship between true, magnetic, and compass headings for a particular instance

2 RADIO WAVES

2.1 GROUND RADIO WAVE

The ground wave travels across the surface of the Earth. You can best imagine the ground wave's path as being in a tunnel or alley bounded by the surface of the Earth and by the ionosphere, which keeps it from going out into space. Generally, the lower the frequency, the farther the signal will travel.

Ground waves are usable for navigation purposes because they reliably and predictably travel the same route, day after day, and are not influenced by too many outside factors. The ground wave frequency range is generally from the lowest frequencies in the radio range (perhaps as low as 100 Hz) up to approximately 1,000 kHz (1 MHz). Although there is a ground wave component to frequencies above this, even to 30 MHz, the ground wave at these higher frequencies loses strength over very short distances.

2.2 SKY RADIO WAVE

The sky wave, at frequencies of 1 to 30 MHz, is good for long distances because these frequencies are refracted or "bent" by the ionosphere, causing the signal to be sent back to Earth from high in the sky and received great distances away, as shown in Figure 8.

Used by high frequency (HF) radios in aircraft, messages can be sent across oceans using only 50 to 100 watts of power. Frequencies that produce a sky wave are not used for navigation because the pathway of the signal from transmitter to receiver is highly variable. The wave is "bounced" off of the ionosphere, which is always changing due to the varying amount of the sun's radiation reaching it (night/day and seasonal variations, sunspot activity, etc.). The sky wave is not reliable for navigation purposes.

For aeronautical communication purposes, the sky wave (HF) is about 80 to 90 percent reliable. HF is being gradually replaced by satellite communication which is more reliable.

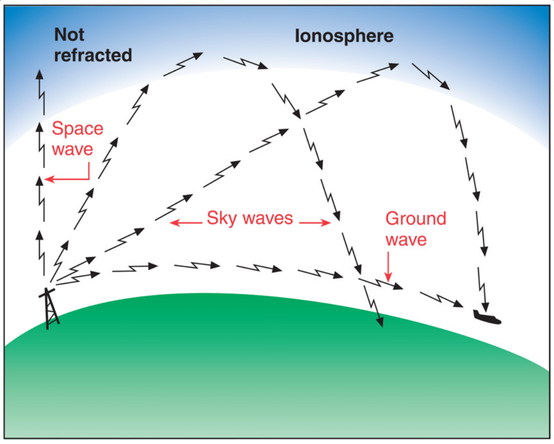


Figure 8 – Ground, space, and sky wave propagation [ATC-3]

2.3 SPACE WAVE

Radio waves of 15 MHz and above (all the way up to many GHz), when able to pass through the ionosphere, are considered space waves. Most navigation systems operate with their signals propagating as space waves. Frequencies above 100 MHz have nearly no ground or sky wave components. They are space waves, but (except for global positioning system (GPS)) the navigation signal is used before it reaches the ionosphere so the effect of the ionosphere, which can cause some propagation errors, is minimal. GPS errors caused by passage through the ionosphere are significant and are corrected for by the GPS receiver system. Space waves have another characteristic of concern to users. Space waves reflect off hard objects and may be blocked if the object is between the transmitter and the receiver. Site and terrain error, as well as propeller/rotor modulation error in very-high omnidirectional range (VOR) systems is caused by this bounce. Instrument landing system (ILS) course distortion is also the result of this phenomenon, which led to the need for establishment of ILS critical areas.

Generally, space waves are “line of sight” receivable, but those of lower frequencies will “bend” over the horizon somewhat. Since the VOR signal at 108 to 118 MHz is a lower frequency than distance measuring equipment (DME) at 962 to 1213 MHz, when an aircraft is flown “over the horizon” from a VOR/DME station, the DME will normally be the first to stop functioning.

2.4 DISTURBANCES TO RADIO WAVE RECEPTION

Static distorts the radio wave and interferes with normal reception of communications and navigation signals. Low frequency airborne equipment such as automatic direction finder (ADF) and long range navigation (LORAN) are particularly subject to static disturbance. Using very-high frequency (VHF) and ultra-high frequency (UHF) frequencies avoids many of the discharge noise effects. Static noise heard on navigation or communication radio frequencies may be a warning of interference with navigation instrument displays. Some [ATC-3] of the problems caused by precipitation static (P-static) are:

• Complete loss of VHF communications.
• Erroneous magnetic compass readings.
• Aircraft flies with one wing low while using the autopilot.
• High-pitched squeal on audio.
• Motorboat sound on audio.
• Loss of all avionics.

- Very-low frequency (VLF) navigation system inoperative.
- Erratic instrument readouts.
- Weak transmissions and poor radio reception.
- St. Elmo's Fire.

3 RADIO ELECTRIC NAVIGATION FACILITIES

3.1 NON-DIRECTIONAL RADIO BEACON (NDB) AND ADF

The non-directional beacon (NDB) is a ground-based radio transmitter [ATC-4] that transmits radio energy in all directions. The ADF, when used with an NDB, determines the bearing from the aircraft to the transmitting station. The indicator may be mounted in a separate instrument in the aircraft panel, as shown in Figure 9.

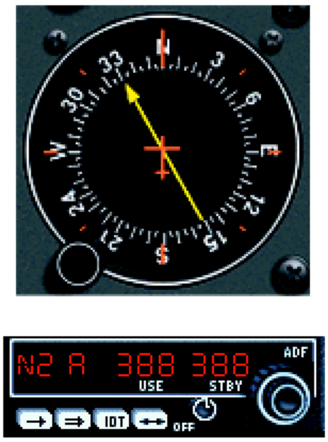


Figure 9 – ADF indicator instrument and receiver [ATC-4]

The ADF needle points to the NDB ground station to determine the relative bearing (RB) to the transmitting station. Magnetic heading (MH) plus RB equals the magnetic bearing (MB) to the station.

The ground equipment, the NDB, transmits in the frequency range of 190 to 535 kHz. Most ADFs will also tune the AM broadcast band frequencies above the NDB band (550 to 1650 kHz). However, these frequencies are not approved for navigation because stations do not continuously identify themselves, and they are much more susceptible to sky wave propagation especially from dusk to dawn. NDB stations are capable of voice transmission and are often used for transmitting the automated weather observing system (AWOS). The aircraft must be in operational range of the NDB. Coverage depends on the strength of the transmitting station. Before relying on ADF indications, identify the station by listening to the Morse code identifier. NDB stations are usually two letters or an alpha-numeric combination.

The airborne equipment includes two antennas, a receiver, and the indicator instrument. The "sense" antenna (non-directional) receives signals with nearly equal efficiency from all directions. The "loop" antenna receives signals better from two directions (bidirectional). When the loop and sense antenna inputs are processed together in the ADF radio, the result is the ability to receive a radio signal well in all directions but one, thus resolving all directional ambiguity. The indicator instrument can be one of three kinds: the fixed card ADF, movable-card ADF, or the radio magnetic indicator (RMI). The fixed-card ADF (also known as the relative bearing indicator (RBI)), always indicates zero at the top of the instrument, and the needle indicates the RB to the station. The movable-card ADF allows the pilot to rotate the aircraft's present heading to the top of the instrument so that the head of the needle indicates MB to the station, and the tail indicates MB from the station.

The ADF can be used to plot your position, track inbound and outbound, and intercept a bearing. These procedures are used to execute holding patterns and non-precision instrument approaches.

3.2 VERY-HIGH FREQUENCY OMNIDIRECTIONAL RANGE (VOR)

VOR is the primary navigational aid (NAVAID) used by civil aviation in the National Airspace System (NAS). The VOR ground station [ATC-4] is oriented to magnetic north and transmits azimuth information to the aircraft, providing 360 courses TO or FROM the VOR station. When DME is installed with the VOR, it is referred to as a VOR/DME and provides both azimuth and distance information. When military equipment is installed with the VOR, it is known as a VORTAC and provides both azimuth and distance information. The courses oriented FROM the station are called radials. The VOR information received by an aircraft is not influenced by aircraft attitude or heading. (Figure 10). For example, aircraft A (heading 180°) is inbound on the 360° radial; after crossing the station, the aircraft is outbound on the 180° radial at A-1. Aircraft B is shown crossing the 225° radial. Similarly, at any point around the station, an aircraft can be located somewhere on a VOR radial.

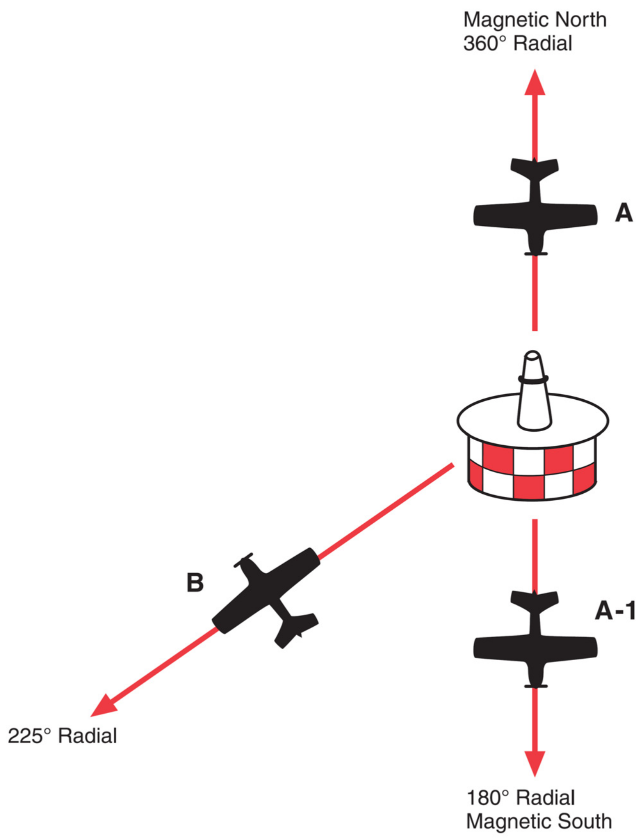


Figure 10 – VOR radials [ATC-4]

The VOR receiver measures and presents information to indicate bearing TO or FROM the station. In addition to the navigation signals transmitted by the VOR, a Morse code signal is transmitted concurrently to identify the facility, as well as voice transmissions for communication and relay of weather and other information. VORs are classified according to their operational uses. The standard VOR facility has a power output of approximately 200 watts, with a maximum usable range depending upon the aircraft altitude, class of facility, location and siting of the facility, terrain conditions within the usable area of the facility, and other factors. Above and beyond certain altitude and distance limits, signal interference from other VOR facilities and a weak signal make it unreliable. Coverage is typically at least 40 miles at normal minimum instrument flight rules (IFR) altitudes. VORs with accuracy problems in parts of their service volume are listed in Notices to Airmen (NOTAMs) and in the Airport/Facility Directory (A/FD) under the name of the NAVAID.

The ground equipment consists of a VOR ground station, which is a small, low building topped with a flat white disc, upon which are located the VOR antennas and a fibreglass cone-shaped tower, as shown in Figure 11.

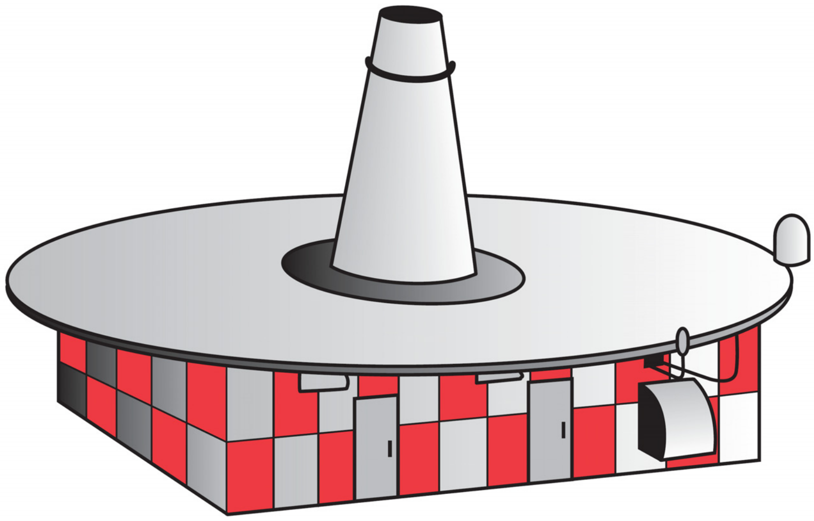


Figure 11 – VOR transmitter (ground station) [ATC-4]

The station includes an automatic monitoring system. The monitor automatically turns off defective equipment and turns on the standby transmitter. Generally, the accuracy of the signal from the ground station is within 1°. VOR facilities are aurally identified by Morse code, or voice, or both. The VOR can be used for ground-to-air communication without interference with the navigation signal. VOR facilities operate within the 108.0 to 117.95 MHz frequency band and assignment between 108.0 and 112.0 MHz is in even-tenth decimals to preclude any conflict with ILS localizer frequency assignment, which uses the odd tenths in this range.

The airborne equipment includes an antenna, a receiver, and the indicator instrument. The receiver has a frequency knob to select any of the frequencies between 108.0 to 117.95 MHz. The ON/OFF/volume control turns on the navigation receiver and controls the audio volume. The volume has no effect on the operation of the receiver. One should listen to the station identifier before relying on the instrument for navigation.

The effectiveness of the VOR depends upon proper use and adjustment of both ground and airborne equipment. The accuracy of course alignment of the VOR is generally plus or minus 1°. On some VORs, minor course roughness may be observed, evidenced by course needle or brief flag alarm. At a few stations, usually in mountainous terrain, the pilot may occasionally observe a brief course needle oscillation, similar to the indication of "approaching station." Pilots flying over unfamiliar routes are cautioned to be on the alert for these vagaries, and in particular, to use the TO/FROM indicator to determine positive station passage. Certain propeller revolutions per minute (RPM) settings or helicopter rotor speeds can cause the VOR CDI to fluctuate as much as plus or minus 6°. Slight changes to the RPM setting will normally smooth out this roughness. Pilots are urged to check for this modulation phenomenon prior to reporting a VOR station or aircraft equipment for unsatisfactory operation.

Figure 12 – VOR Indicator [ATC-4]

3.3 DISTANCE MEASURING EQUIPMENT (DME)

When used in conjunction with the VOR system, DME makes it possible for pilots to determine an accurate geographic position of the aircraft [ATC-4], including the bearing and distance TO or FROM the station. The aircraft DME transmits interrogating radio frequency (RF) pulses, which are received by the DME antenna at the ground facility. The signal triggers ground receiver equipment to respond back to the interrogating aircraft. The airborne DME equipment measures the elapsed time between the interrogation signal sent by the aircraft and reception of the reply pulses from the ground station. This time measurement is converted into nautical miles (NMs) distance from the station.
Some DME receivers provide a groundspeed in knots by monitoring the rate of change of the aircraft's position relative to the ground station. Groundspeed values are only accurate when tracking directly to or from the station.

VOR/DME, VORTAC, ILS/DME, and LOC/DME navigation facilities established by the FAA provide course and distance information from collocated components under a frequency pairing plan. DME operates on frequencies in the UHF spectrum between 962 MHz and 1213 MHz. Aircraft receiving equipment which provides for automatic DME selection assures reception of azimuth and distance information from a common source when designated VOR/DME, VORTAC, ILS/DME, and LOC/DME are selected. Some aircraft have separate VOR and DME receivers each of which must be tuned to the appropriate navigation facility.

The airborne equipment includes an antenna and a receiver.

A DME is used for determining the distance from a ground DME transmitter. Compared to other VHF/UHF NAVAIDs, a DME is very accurate. The distance information can be used to determine the aircraft position or flying a track that is a constant distance from the station. This is referred to as a DME arc.

DME/DME fixes (a location based on two DME lines of position from two DME stations) provide a more accurate aircraft location than using a VOR and a DME fix.
DME signals are line-of-sight; the mileage readout is the straight line distance from the aircraft to the DME ground facility and is commonly referred to as slant range distance. Slant range refers to the straight line distance from the aircraft antenna to the ground station, which differs somewhat from the distance from the station to the point on the ground beneath the aircraft. This error is smallest at low altitude and long range. It is greatest when the aircraft is over the ground facility, at which time the DME receiver will display altitude (in NM) above the facility. Slant-range error is negligible if the

aircraft is 1 mile or more from the ground facility for each 1,000 feet of altitude above the elevation of the facility.

3.4 AREA NAVIGATION (RNAV) EQUIPMENT

Area navigation (RNAV) equipment includes VOR/DME, LORAN, GPS, and inertial navigation systems (INS). RNAV equipment is capable of computing the aircraft position, actual track, groundspeed, and then presenting meaningful information to the pilot [ATC-3]. This information may be in the form of distance, cross track error, and time estimates relative to the selected track or waypoint. In addition, the RNAV equipment installations must be approved for use under IFR. The Pilot's Operating Handbook/Airplane Flight Manual (POH/AFM) will report what equipment is installed, the operations that are approved, and the details of how to use the equipment. Some aircraft may have equipment that allows input from more than one RNAV source thereby providing a very accurate and reliable navigation source.

3.4.1 VOR/DME RNAV

VOR RNAV is based on information generated by the present VORTAC or VOR/DME systems to create a waypoint using an airborne computer [ATC-3]. As shown in the next Figure 13, the value of side A is the measured DME distance to the VOR/DME. Side B, the distance from the VOR/DME to the waypoint, angle 1 (VOR radial or the bearing from the VORTAC to the waypoint), are values set in the cockpit control. The bearing from the VOR/DME to the aircraft, angle 2, is measured by the VOR receiver. The airborne computer continuously compares angles 1 and 2 and determines angle 3 and side C, which is the distance in NMs and magnetic course from the aircraft to the waypoint. This is presented as guidance information on the cockpit display.

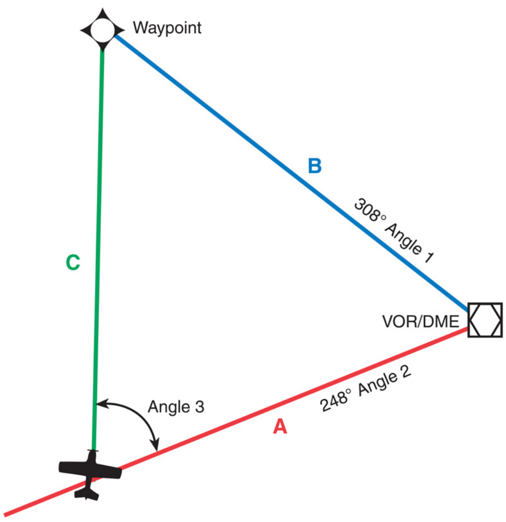


Figure 13 – RNAV computation [ATC-3]

Although RNAV cockpit instrument displays vary among manufacturers, most are connected to the aircraft CDI with a switch or knob to select VOR or RNAV guidance. There is usually a light or indicator to inform the pilot whether VOR or RNAV is selected, like in Figure 14. The display includes the waypoint, frequency, mode in use, waypoint radial and distance, DME distance, groundspeed, and time to station.

Most VOR/DME RNAV systems have the following airborne controls:
1. Off/On/Volume control to select the frequency of the VOR/DME station to be used.
2. MODE select switch used to select VOR/DME mode, with:
a. Angular course width deviation (standard VOR operation); or
b. Linear cross track deviation as standard (±5 NM full scale CDI).
3. RNAV mode, with direct to waypoint with linear cross track deviation of ±5 NM.
4. RNAV/APPR (approach mode) with linear deviation of ±1.25 NM as full scale CDI deflection.
5. Waypoint select control. Some units allow the storage of more than one waypoint; this control allows selection of any waypoint in storage.
6. Data input controls. These controls allow user input of waypoint number or ident, VOR or LOC frequency, waypoint radial and distance.

While DME groundspeed readout is accurate only when tracking directly to or from the station in VOR/DME mode, in RNAV mode the DME groundspeed readout is accurate on any track.

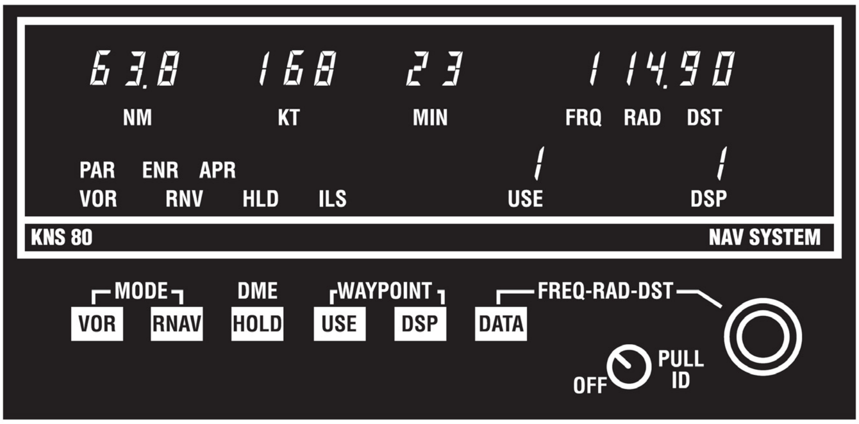


Figure 14 – Typical RNAV display [ATC-4]

The advantages of the VOR/DME RNAV system stem from the ability of the airborne computer to locate a waypoint wherever it is convenient, as long as the aircraft is within reception range of both a nearby VOR and DME facility. A series of these waypoints make up an RNAV route. In addition to the published routes, a random RNAV route may be flown under IFR if it is approved by air traffic control (ATC). RNAV DPs and standard terminal arrival routes (STARs) are contained in the DP and STAR booklets. VOR/DME RNAV approach procedure charts are also available. Note that in the VOR/DME RNAV chart, that the waypoint identification boxes contain the following information: waypoint name, coordinates, frequency, identifier, radial distance (facility to waypoint), and reference facility elevation. The initial approach fix (IAF), final approach fix (FAF), and missed approach point (MAP) are labelled. To fly either a route or to execute an approach under IFR, the RNAV equipment installed in the aircraft must be approved for the appropriate IFR operations. In Vertical Nav mode, vertical, as well as horizontal guidance is provided in some installations. A waypoint is selected at a point where the descent begins, and another waypoint is selected where the descent ends. The RNAV equipment computes the rate of descent relative to the groundspeed, and on some installations, displays vertical guidance information on the glide-slope indicator. When using this type of equipment during an instrument approach, the pilot must keep in mind the vertical guidance information provided is not part of the non-precision approach. Published non-precision approach altitudes must be observed and complied with, unless otherwise directed by ATC.

The limitation of this system is the reception volume. Published approaches have been tested to ensure this is not a problem. Descents/approaches to airports distant from the VOR/DME facility may not be possible because during the approach, the aircraft may descend below the reception altitude of the facility at that distance.

3.4.2 LONG RANGE NAVIGATION (LORAN)

LORAN uses a network of land-based transmitters to provide an accurate long range navigation system [ATC-3]. The FAA and the United States (U.S.) Coast Guard (USCG) arranged the stations into chains. The signals from these stations are a carefully structured sequence of brief RF pulses centred at 100 kHz. At that frequency, signals travel considerable distances as ground waves, from which accurate navigation information is available. The airborne receiver monitors all of the stations within the selected chain, then measures the arrival time difference (TD) between the signals. All of the points having the same TD from a station pair create a line of position (LOP). The aircraft position is determined at the intersection of two or more LOPs. Then the computer converts the known location to latitude and longitude coordinates. While continually computing latitude/longitude fixes, the computer is able to determine and display:

1. Track over the ground since last computation;
2. Groundspeed by dividing distance covered since last computation by the time since last computation (and averaging several of these);
3. Distance to destination;
4. Destination time of arrival; and
5. Cross track error.

The Aeronautical Information Manual (AIM) provides a detailed explanation of how LORAN works. LORAN is a very accurate navigation system if adequate signals are received. There are two types of accuracy that must be addressed in any discussion of LORAN accuracy.

Repeatable accuracy is the accuracy measured when a user notes the LORAN position, moves away from that location, then uses the LORAN to return to that initial LORAN position. Distance from that initial position is the error. Propagation and terrain errors will be essentially the same as when the first position was taken, so those errors are factored out by using the initial position. Typical repeatable accuracy for LORAN can be as good as 0.01 NM, or 60 feet, if the second position is determined during the day and within a short period of time (a few days).

Absolute accuracy refers to the ability to determine present position in space independently, and is most often used by pilots. When the LORAN receiver is turned on and position is determined, absolute accuracy applies. Typical LORAN absolute accuracy will vary from about 0.1 NM to as much as 2.5 NM depending on distance from the station, geometry of the TD LOP crossing angles, terrain and environmental conditions, signal-to-noise ratio (signal strength), and some design choices made by the receiver manufacturer.

The LORAN receiver incorporates a radio receiver, signal processor, navigation computer, control/display, and antenna. When turned on, the receivers go through an initialization or warm-up period, then inform the user they are ready to be programmed. LORAN receivers vary widely in their appearance, how they are programmed by the user, and how they display navigation information. Therefore, it is necessary to become familiar with the unit, including how to program it, and how to interpret output from it. The LORAN operating manual should be in the aircraft at all times and available to the pilot. IFR-approved LORAN units require that the manual be aboard and that the pilot is familiar with the unit's functions, before flight.

After initialization, you select for the present location waypoint (the airport), and select GO TO in order to determine if the LORAN is functioning properly. Proper operation is indicated by a low distance reading (0 to 0.5 NM). The simplest mode of navigation is referred to as GO TO: you select a waypoint from one of the databases and choose the GO TO mode. Before use in flight, you should verify that the latitude and longitude of the chosen waypoint is correct by reference to another approved information source. An updatable LORAN database that supports the appropriate operations (e.g., en route, terminal, and instrument approaches), is required when operating under IFR. In addition to displaying bearing, distance, time to the waypoint, and track and speed over the ground, the LORAN receiver may have other features such as flight planning (waypoint sequential storage), emergency location of several nearest airports, vertical navigation capabilities, and more.

LORAN is subject to interference from many external sources, which can cause distortion of, or interference with LORAN signals. LORAN receiver manufacturers install "notch filters" to reduce or eliminate interference. Proximity to 60 Hz alternating current power lines, static discharge, precipitation static, electrical noise from generators, alternators, strobes, and other onboard electronics may decrease the signal-to-noise ratio to the point where the LORAN receiver's performance is degraded. Proper installation of the antenna, good electrical bonding, and an effective static discharge system are the minimum requirements for LORAN receiver operation. Most receivers have internal tests that verify the timing alignment of the receiver clock with the LORAN pulse, and measure and display signal-to-noise ratio. A signal will be activated to alert the pilot if any of the parameters for reliable navigation are exceeded on LORAN sets certified for IFR operations.
LORAN is most accurate when the signal travels over sea water during the day, and least accurate when the signal comes over land and large bodies of fresh water or ice at night; furthermore, the accuracy degrades as distance from the station increases. However, LORAN accuracy is generally better than VOR accuracy.

3.4.3 GLOBAL NAVIGATION SATELLITE SYSTEM (GNSS)

GNSS refers to the use of one or more satellite navigation systems. The U.S. GPS (Global Positioning System) is one such system, as are the Russian GLONASS (Global Navigation Satellite System) and the proposed GALILEO system, a project funded by the EU (European Union). Recently developed augmentation satellite systems such as the FAA (Federal Aviation Administration) WAAS (Wide Area Augmentation System) and EGNOS (European Geostationary Navigation Overlay System) also provide supplementary capability. Several of these systems are described, beginning with GPS since it is the core of the system developed in this research, as are some of the factors that affect satellite positioning performance [AIRP-16]. Some information about clock and height augmentations of normal satellite positioning are also described.

GPS is a satellite based radio-navigation system designed by the U.S. DoD (Department of Defence). It is a spread spectrum CDMA (Code Division Multiple Access) system that broadcasts bi-phase modulated RHCP (Right Hand Circularly Polarized) signals centred at about 1.2 (L2) and 1.6 GHz (L1). This makes it an all weather system since there is little atmospheric absorption at these frequencies, and because it is spread spectrum, it is also resistant to jamming (Spilker 1996). Due mainly to geometrical spreading, the signal strength at the Earth's surface is only around –160 dbW, so the system operates basically through LOS (Line Of Sight).

The satellite constellation nominally consists of 24 satellites on 6 orbital planes with an inclination of 55° to the equator (U.S. JPO 1995). The satellites are in near circular orbit at a height of about 20,000 km above the surface, with a nominal period of just under 12 hours. The specifications call for a minimum of 21 operational satellites and three spares to provide a minimum of four visible satellites at any place and any time on the planet. At present, the constellation has 27 satellites and unobstructed visibility is always seven or more satellites in the European region. Each satellite broadcasts a unique C/A (Coarse Acquisition) PRN (Pseudo-Random Noise) code modulated on to the L1 carrier. SA (Selective Availability) was used until May 1, 2000, to degrade the performance of civilian users in single-point mode through dithering of the satellite clock offset. A 50 bps navigation message, which gives the time of transmission of the PRN sequence and other necessary information, and a second code (P) are modulated on both the L1 and L2 carriers. These P code is intended primarily for military users and provides better resolution (less receiver noise) and immunity from SA, but is encrypted through a procedure known as AS (Anti-Spoofing). Civilians can still make use of the

P code, but a loss is suffered through correlation techniques used to sidestep the encryption, and SA if not removed.

The MCS (Master Control Station) processes range measurements taken at the five monitoring stations and develops predictions for the orbits and satellite clock behavior. The MCS then sends this data to the monitor stations for upload of the navigation message to the satellites. The navigation message also includes the health of the satellites and an almanac that can be used to predict the visibility of satellites for any time and place.

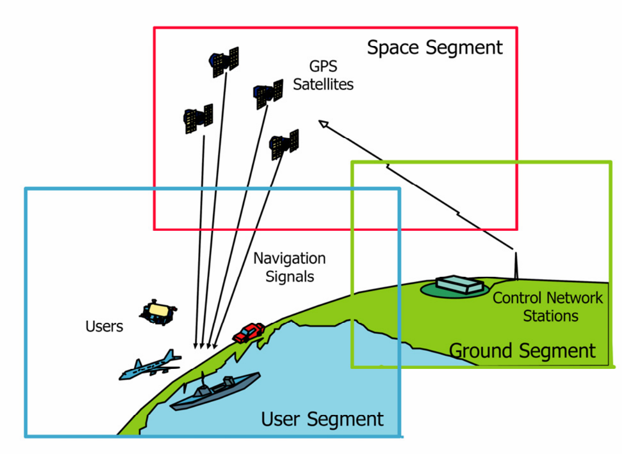


Figure 15 – GPS Segments [AIRP-16]

The GPS Space Segment consists of 24 Navstar satellites in semi-synchronous (approximately 12hour) orbits. The satellites are arranged in six orbital planes with four satellites in each plane. The orbital planes have an inclination angle of 55 degrees relative to the earth's equator. The satellites have an average orbit altitude of 20200 kilometres (10900 nautical miles) above the surface of the earth. Figure 16 illustrates the GPS satellite constellation.

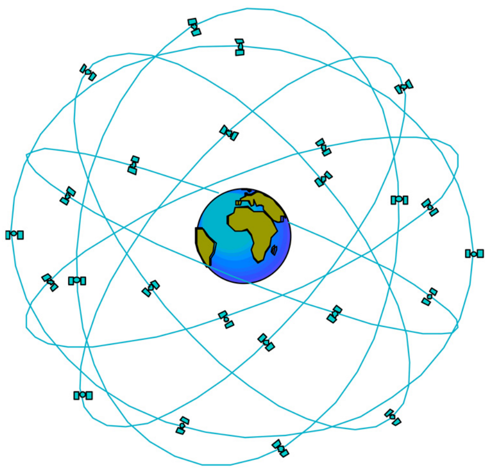


Figure 16 – GPS Satellites Constellation [AIRP-16]

The Control Segment primarily consists of a Master Control Station (MCS), at Falcon Air Force Base (AFB) in Colorado Springs, USA, plus monitor stations (MS) and ground antennas (GA) at various locations around the world. The monitor stations are located at Falcon AFB, Hawaii, Kwajalein, Diego Garcia, and Ascension. All monitor stations except Hawaii and Falcon AFB are also equipped with ground antennas. The Control Segment includes a Prelaunch Compatibility Station (PCS) located at Cape Canaveral, USA, and a back-up MCS capability.

The user segment is the largest and the most widely influenced of the three segments. It consists mainly of GPS receivers belonging to a wide spectrum of user requirements. All these receivers use the satellite ranging signals to determine their position and time with the accuracy specified by the Joint Program Office (JPO). The user segment can be broadly classified into military users and civilian users. Military users have access to the high accuracy Precise Positioning Service (PPS), whereas the civilian users have access to the less precise Coarse Acquisition (C/A) code. Civilian users constitute a major portion of the user segment market, the estimated users in the civilian segment is approximately a few million and is growing every year. The GPS market in the user segment is currently a $2 billion/year market and is expected to grow to $30 billion/year market by 2008. GPS receivers compute the range by measuring the transit time of the signal from the satellite to the receiver antenna. To compute position and time, the receivers perform triangulation using range measurements from at least 3 satellites. However, since the receiver local clock is not in synchronism with satellite clocks, an additional measurement is required to solve for the clock offset.

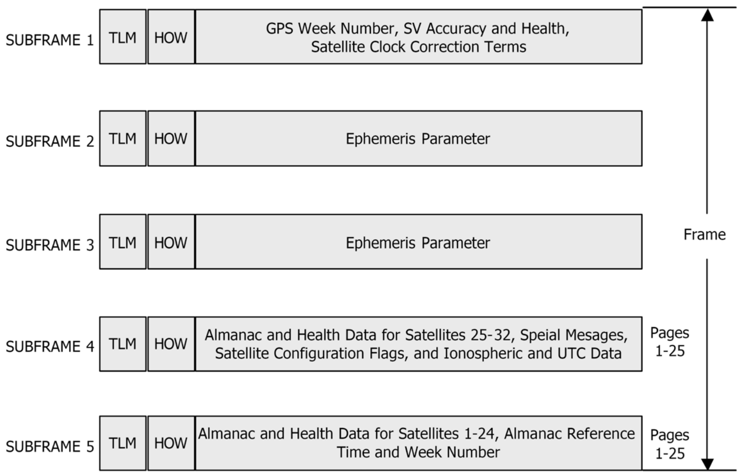


Figure 17 – GPS Navigation Message [AIRP-16]

The ranging codes broadcast by the satellites enable a GPS receiver to measure the transit time of the signals and thereby determine the range between a satellite and the user. The navigation message provides data to calculate the position of each satellite at the time of signal transmission. From this information, the user position coordinates and the user clock offset are calculated using simultaneous equations. Four satellites are normally required to be simultaneously "in view" of the receiver for 3-D positioning purposes.

Two levels of service are provided by the GPS, the *Precise Positioning Service* (PPS) and the *Standard Positioning Service* (SPS).

Precise Positioning Service. The PPS is an accurate positioning velocity and timing service which is available only to authorized users. The PPS is primarily intended for military purposes. Authorization to use the PPS is determined by the U.S. Department of Defence (DoD), based on internal U.S. defence requirements or international defence commitments. Authorized users of the PPS include U.S. military users, NATO military users, and other selected military and civilian users such as the Australian Defence Forces and the U.S. Defence Mapping Agency. The PPS is specified to provide 16 metres Spherical Error Probable (SEP) (3-D, 50%) positioning accuracy and 100 nanosecond (one sigma) Universal Coordinated Time (UTC) time transfer accuracy to authorized users. This is approximately equal to 37 metres (3-D, 95%) and 197 nanoseconds (95%) under typical system operating conditions. PPS receivers can achieve 0.2 metres per second 3-D velocity accuracy, but this is somewhat dependent on receiver design.

Standard Positioning Service. The SPS is a less accurate positioning and timing service which is available to all GPS users. In peacetime before May 2001, the level of SA is controlled to provide 100 metre (95%)

horizontal accuracy which is approximately equal to 156 metres 3D (95%). SPS receivers can achieve approximately 337 nanosecond (95%) UTC time transfer accuracy. Selective Availability (SA) was the intentional degradation of the signal by DoD. SA could be implemented in two different ways 1) introducing error into the satellite broadcast orbit (known as the e-process), and 2) dithering the satellite clock frequency (known as the d-process). SA was switched off on May 1, 2000 by a Presidential directive. Actual performances of the SPS is 8.1 m (1s), considering all errors contained in the broadcasting of the GPS message. The SPS is primarily intended for civilian purposes, although it has potential peacetime military use.

3.4.4 DIFFERENTIAL GPS (D-GPS)

Differential GPS (D-GPS) was developed to meet the needs of positioning and distance measuring applications that required higher accuracies than stand-alone GPS could deliver [AIRP-16]. A typical differential GPS architecture (see Figure 18, where an installation for a precise landing approach station is depicted) consists of one or more reference receivers located at a surveyed, known location, and one or more D-GPS user receivers. The user receivers are often called "mobile" receivers because they are not confined to a fixed location like the reference receiver. Both sets of receivers either collect and store the necessary data for later processing (typical geodetic or general surveying application), or send them to the desired location in real time via the data link.

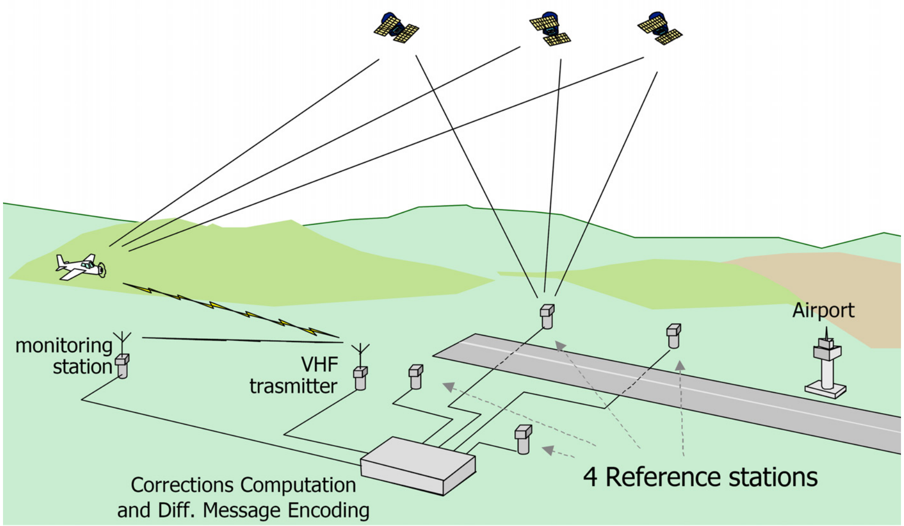


Figure 18 – Typical Differential System Architecture [AIRP-16]

This overview outlines some of the fundamental issues of DGPS. These issues should be considered by any user considering the need for a positioning system that can give accuracies better than the absolute PPS or SPS performance.

DGPS is based on the principle that receivers in the same vicinity will simultaneously experience common errors on a particular satellite ranging signal. In general, the user (mobile) receivers use measurements from the reference receiver to remove the common errors. In order to accomplish this, the user (mobile) receivers must simultaneously use a subset or the same set of satellites as the reference station. The DGPS positioning equations are formulated so that the common errors cancel. The common errors include signal path delays through the atmosphere, and satellite clock and ephemeris errors. For PPS users, the common satellite errors are residual system errors that are normally present in the PVT solution. For SPS users, the common satellite errors also include the intentionally added errors from SA. Errors that are unique to each receiver, such as receiver measurement noise and multipath, cannot be removed without additional recursive processing (by the reference receiver, user receiver, or both) to provide an averaged, smoothed, or filtered solution. Various DGPS techniques are employed depending on the accuracy desired, where the data processing is to be performed, and whether real-time results are required. If real-time results are required then a data link is also required. For applications without a real-time requirement, the data can be collected and processed later. The accuracy requirements usually dictate which measurements are used and what algorithms are employed. Under normal conditions, DGPS accuracy is independent of whether SPS or PPS is being used, although real-time PPS DGPS can have a lower data rate than SPS DGPS because the rate of change of the nominal system errors is slower than the rate of change of SA. However, the user and the Reference Station must be using the same service (either PPS or SPS).

The accepted standard for SPS DGPS was developed by the Radio Technical Commission for Maritime Services (RTCM) Special Committee-104 (SC-104). The RTCM developed standards for the use of differential corrections, and defined the data format to be used between the reference station and the user. The standards are primarily intended for real time operational use and cover a wide range of DGPS measurement types. Most SPS DGPS receivers are compatible with the RTCM SC-104 differential message formats. DGPS standards have also been developed by the Radio Technical Commission for Aeronautics (RTCA) for special Category I precision approach using ranging-code differential. The standards are contained in RTCA document DO-217. This document is intended only for limited use until an international standard can be developed for precision approach.

3.4.5 INERTIAL NAVIGATION SYSTEM (INS)

INS is a system that navigates precisely by dead reckoning, without any input from outside of the aircraft [ATC-4]. It is fully self-contained. The INS is initialized by the pilot, who enters into the system its exact location while the aircraft is on the ground before the flight. The INS is also programmed with waypoints along the desired route of flight.

INS is considered a stand-alone navigation system, especially when more than one independent unit is onboard. The airborne equipment consists of an accelerometer to measure acceleration—which, when integrated with time, gives velocity—and gyros to measure direction. Later versions of the INS, called IRS (inertial reference systems) utilize laser gyros and more powerful computers; therefore, the accelerometer mountings no longer need to be kept level and aligned with true north. The computer system can handle the added workload of dealing with the computations necessary to correct for gravitational and directional errors. Consequently, these newer systems are sometimes called strap down systems, as the accelerometers and gyros are strapped down to the airframe, rather than being mounted on a structure that stays fixed with respect to the horizon and true north.

The principal error associated with INS is degradation of position with time. INS computes position by starting with an accurate position input which is changed continuously as accelerometers and gyros provide speed and direction inputs. Both the accelerometers and the gyros are subject to very small errors; as time passes, those errors likely will accumulate. While the best INS/IRS display errors of 0.1 to 0.4 NM after flights across the North Atlantic of 4 to 6 hours, smaller and less expensive systems are being built that show errors of 1 to 2 NM per hour. This accuracy is more than sufficient for a navigation system that can be combined with and updated by GPS. The synergy of a navigation system consisting of an INS/IRS unit in combination with a GPS resolves the errors and weaknesses of both systems. The GPS is accurate all the time it is working but may be subject to short and periodic outages. The INS is made more accurate because it is continually updated and will continue to function with good accuracy if the GPS has moments of lost signal.

3.5 INSTRUMENT APPROACH SYSTEMS

Most navigation systems approved for en route and terminal operations under IFR, such as VOR, NDB, and GPS, may also be approved to conduct IAPs. The most common systems in use in the world are the ILS, simplified landing facility (SDF), localizer directional aid (LDA), and microwave landing system (MLS). These systems operate independently of other navigation systems. There are new systems being developed, such as wide area augmentation system (WAAS), local area augmentation system (LAAS), and other systems have been developed for special use [ATC-4].

3.6 INSTRUMENT LANDING SYSTEMS (ILS)

The ILS system provides both course and altitude guidance to a specific runway [ATC-4]. The ILS system is used to execute a precision instrument approach procedure or precision approach (Figure 19). The system consists of the following components:

1. A localizer provides horizontal (left/right) guidance along the extended centreline of the runway.
2. A glide slope provides vertical (up/down) guidance toward the runway touchdown point, usually at a 3° slope.
3. Marker beacons provide range information along the approach path.
4. Approach lights assist in the transition from instrument to visual flight.

The following supplementary elements, though not specific components of the system, may be incorporated to increase safety and utility:

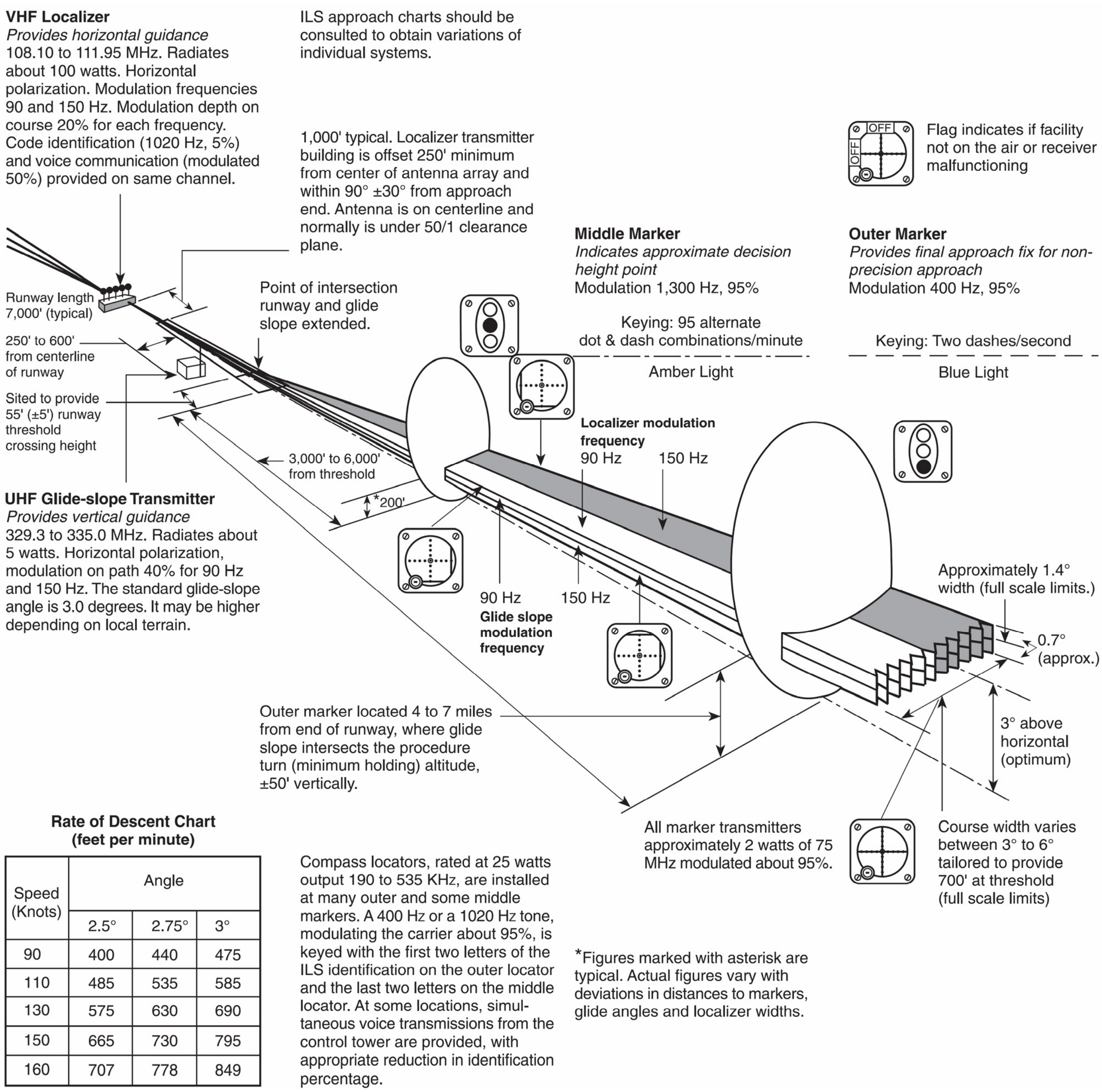
1. Compass locators provide transition from en route NAVAIDs to the ILS system; they assist in holding procedures, tracking the localizer course, identifying the marker beacon sites, and providing a FAF for ADF approaches.
2. DME colocated with the glide-slope transmitter provide positive distance-to-touchdown information or DME associated with another nearby facility (VOR or standalone), if specified in the approach procedure.

ILS approaches are categorized into three different types of approaches, based on the equipment at the airport and the experience level of the pilot. Category I approaches provide for approach height above touchdown of not less than 200 feet. Category II approaches provide for approach to a height above touchdown of not less than 100 feet. Category III approaches provide lower minimums for approaches without a decision height minimum. While pilots must only be instrument rated and the aircraft be equipped with the appropriate airborne equipment to execute Category I approaches, Category II and III approaches require special certification for the pilots, ground equipment, and airborne equipment.

The ILS uses a number of different ground facilities. These facilities may be used as a part of the ILS system, as well as part of another approach. For example, the compass locator may be used with NDB approaches.

3.6.1 LOCALIZER

The localizer (LOC) ground antenna array is located on the extended centreline of the instrument runway of an airport, remote enough from the opposite (approach) end of the runway to prevent it from being a collision hazard.



Speed (Knots)	Angle		
	2.5°	2.75°	3°
90	400	440	475
110	485	535	585
130	575	630	690
150	665	730	795
160	707	778	849

Figure 19 – Instrument Landing Systems [ATC-4]

This unit radiates a field pattern, which develops a course down the centreline of the runway toward the middle markers (MMs) and outer markers (OMs), and a similar course along the runway centreline in the opposite direction. These are called the front and back courses, respectively. The localizer provides course guidance, transmitted at 108.1 to 111.95 MHz (odd tenths only), throughout

the descent path to the runway threshold from a distance of 18 NM from the antenna to an altitude of 4,500 feet above the elevation of the antenna site.
The localizer course width is defined as the angular displacement at any point along the course between a full "fly-left" (CDI needle fully deflected to the left) and a full "fly-right" indication (CDI needle fully deflected to the right.) Each localizer facility is audibly identified by a three-letter designator, transmitted at frequent regular intervals. The ILS identification is preceded by the letter "I" (two dots). For example, the ILS localizer at Springfield, Missouri transmits the identifier ISGF. The localizer includes a voice feature on its frequency for use by the associated ATC facility in issuing approach and landing instructions. The localizer course is very narrow, normally 5°. This results in high needle sensitivity. With this course width, a full-scale deflection shows when the aircraft is 2.5° to either side of the centreline. This sensitivity permits accurate orientation to the landing runway. With no more than one-quarter scale deflection maintained, the aircraft will be aligned with the runway (Figure 21).

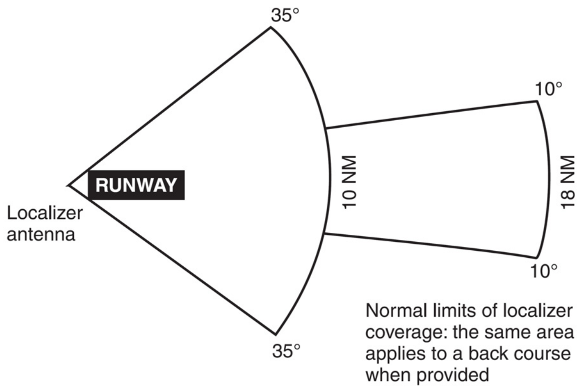


Figure 20 – Localizer Coverage Limits [ATC-4]

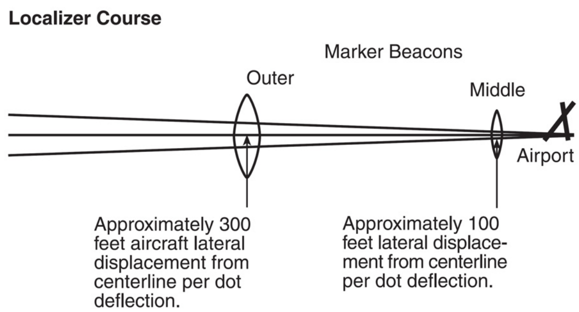


Figure 21 – Localizer Receiver Indications and Aircraft Displacement [ATC-4]

3.6.2 GLIDE SLOPE

Glide slope (GS) describes the systems that generate, receive, and indicate the ground facility radiation pattern [ATC-4]. The glidepath is the straight, sloped line the aircraft should fly in its descent from where the glide slope intersects the altitude used for approaching the FAF, to the runway touchdown zone. The glide-slope equipment is housed in a building approximately 750 to 1,250 feet down the runway from the approach end of the runway, and between 400 and 600 feet to one side of the centreline. The course projected by the glide-slope equipment is essentially the same as would be generated by a localizer operating on its side. The glide-slope projection angle is normally adjusted to 2.5° to 3.5° above horizontal, so it intersects the MM at about 200 feet and the OM at about 1,400 feet above the runway elevation. At locations where standard minimum obstruction clearance cannot be obtained with the normal maximum glide-slope angle, the glide-slope equipment is displaced farther

from the approach end of the runway if the length of the runway permits; or, the glideslope angle may be increased up to 4°.
Unlike the localizer, the glide-slope transmitter radiates signals only in the direction of the final approach on the front course. The system provides no vertical guidance for approaches on the back course. The glidepath is normally 1.4° thick. At 10 NM from the point of touchdown, this represents a vertical distance of approximately 1,500 feet, narrowing to a few feet at touchdown.

3.6.3 MARKER BEACONS

Two VHF marker beacons, outer and middle, are normally used in the ILS system [ATC-4]. A third beacon, the inner, is used where Category II operations are certified. A marker beacon may also be installed to indicate the FAF on the ILS back course. The OM is located on the localizer front course 4 to 7 miles from the airport to indicate a position at which an aircraft, at the appropriate altitude on the localizer course, will intercept the glidepath. The MM is located approximately 3,500 feet from the landing threshold on the centreline of the localizer front course at a position where the glide-slope centreline is about 200 feet above the touchdown zone elevation. The inner marker (IM), where installed, is located on the front course between the MM and the landing threshold. It indicates the point at which an aircraft is at the decision height on the glidepath during a Category II ILS approach. The back-course marker, where installed, indicates the back-course FAF.

3.6.4 ILS AIRBORNE COMPONENTS

Airborne equipment for the ILS system includes receivers for the localizer, glide slope, marker beacons, ADF, DME, and the respective indicator instruments [ATC-4]. The typical VOR receiver is also a localizer receiver with common tuning and indicating equipment. Some receivers have separate function selector switches, but most switch between VOR and LOC automatically by sensing if odd tenths between 108 and 111.95 MHz have been selected. Otherwise, tuning of VOR and localizer frequencies is accomplished with the same knobs and switches, and the CDI indicates "on course" as it does on a VOR radial.

Though some glide-slope receivers are tuned separately, in a typical installation the glide slope is tuned automatically to the proper frequency when the localizer is tuned. Each of the 40 localizer channels in the 108.10 to 111.95 MHz band is paired with a corresponding glide-slope frequency.

When the localizer indicator also includes a glide-slope needle, the instrument is often called a cross-pointer indicator. The crossed horizontal (glide slope) and vertical (localizer) needles are free to move through standard five-dot deflections to indicate position on the localizer course and glidepath.

When the aircraft is on the glidepath, the needle is horizontal, overlying the reference dots. Since the glidepath is much narrower than the localizer course (approximately 1.4° from full up to full down deflection), the needle is very sensitive to displacement of the aircraft from on-path alignment. With the proper rate of descent established upon glide-slope interception, very small corrections keep the aircraft aligned.

The localizer and glide-slope warning flags disappear from view on the indicator when sufficient voltage is received to actuate the needles. The flags show when an unstable signal or receiver malfunction occurs.

The OM is identified by a low-pitched tone, continuous dashes at the rate of two per second, and a purple/blue marker beacon light. The MM is identified by an intermediate tone, alternate dots and dashes at the rate of 95 dot/dash combinations per minute, and an amber marker beacon light. The IM, where installed, is identified by a high-pitched tone, continuous dots at the rate of six per second, and a white marker beacon light. The back-course marker (BCM), where installed, is identified by a high-pitched tone with two dots at a rate of 72 to 75 two-dot combinations per minute, and a white marker beacon light. Marker beacon receiver sensitivity is selectable as high or low on many units. The low-sensitivity position gives the sharpest indication of position and should be used during an approach. The high-sensitivity position provides an earlier warning that the aircraft is approaching the marker beacon site.

The ILS and its components are subject to certain errors, which are listed below.
Localizer and glide-slope signals are subject to the same type of bounce from hard objects as space waves.

1. *Reflection*. Surface vehicles and even other aircraft flying below 5,000 feet above ground level (AGL) may disturb the signal for aircraft on the approach.
2. *False courses*. In addition to the desired course, glideslope facilities inherently produce additional courses at higher vertical angles. The angle of the lowest of these false courses will occur at approximately 9°–12°. An aircraft flying the LOC/glide-slope course at a constant altitude would observe gyrations of both the glide-slope needle and glide-slope warning flag as the aircraft passed through the various false courses. Getting established on one of these false courses will result in either confusion (reversed glide-slope needle indications), or result in the need for a very high descent rate. However, if the approach is conducted at the altitudes specified on the appropriate approach chart, these false courses will not be encountered.

The very low power and directional antenna of the marker beacon transmitter ensures that the signal will not be received any distance from the transmitter site. Problems with signal reception are usually caused by the airborne receiver not being turned on, or by incorrect receiver sensitivity. Some marker beacon receivers, to decrease weight and cost, are designed without their own power supply. These units utilize a power source from another radio in the avionics stack, often the ADF. In some aircraft, this requires the ADF to be turned on in order for the marker beacon receiver to function, yet no warning placard is required. Another source of trouble may be the "High/Low/Off" three-position switch, which both activates the receiver and selects receiver sensitivity. Usually, the "test" feature only tests to see if the light bulbs in the marker beacon lights are working. Therefore, in some installations, there is no functional way for the pilot to ascertain the marker beacon receiver is actually on except to fly over a marker beacon transmitter, and see if a signal is received and indicated (e.g., audibly and marker beacon lights).

3.6.5 Simplified Directional Facility (SDF)

The SDF provides a final approach course similar to the ILS localizer. The SDF course may or may not be aligned with the runway and the course may be wider than a standard ILS localizer, resulting in less precision [ATC-4]. Usable off-course indications are limited to 35° either side of the course centreline. Instrument indications in the area between 35° and 90° from the course centreline are not controlled and should be disregarded. The SDF antenna may be offset from the runway centreline. Because of this, the angle of convergence between the final approach course and the runway bearing should be determined by reference to the instrument approach chart. This angle is usually not more than 3°. You should note this angle since the approach course originates at the antenna site, and an approach continued beyond the runway threshold would lead the aircraft to the SDF offset position rather than along the runway centreline.

The course width of the SDF signal emitted from the transmitter is fixed at either 6° or 12°, as necessary, to provide maximum flyability and optimum approach course quality. A three-letter identifier is transmitted in code on the SDF frequency; there is no letter "I" (two dots) transmitted before the station identifier, as there is with the LOC. For example, the identifier for Lebanon, Missouri, SDF is LBO.

3.6.6 LOCALIZER TYPE DIRECTIONAL AID (LDA)

The LDA is of comparable utility and accuracy to a localizer but is not part of a complete ILS. The LDA course width is between 3° and 6° and thus provides a more precise approach course than an SDF installation [ATC-4]. Some LDAs are equipped with a glide slope. The LDA course is not aligned with the
runway, but straight-in minimums may be published where the angle between the runway centreline and the LDA course does not exceed 30°. If this angle exceeds 30°, only circling minimums are published. The identifier is three letters preceded by "I" transmitted in code on the LDA frequency. For example, the identifier for Van Nuys, California, LDA is I-BUR.

3.7 COMPASS LOCATOR

Compass locators are low-powered NDBs and are received and indicated by the ADF receiver [ATC-4]. When used in conjunction with an ILS front course, the compass locator facilities are colocated with the outer and/or MM facilities. The coding identification of the outer locator consists of the first two letters of the three-letter identifier of the associated LOC. For example, the outer locator at Dallas/Love Field (DAL) is identified as "DA." The middle locator at DAL is identified by the last two letters "AL."

3.8 APPROACH LIGHTING SYSTEMS (ALS)

Normal approach and letdown on the ILS is divided into two distinct stages: the instrument approach stage using only radio guidance, and the visual stage, when visual contact with the ground runway environment is necessary for accuracy and safety [ATC-3]. The most critical period of an instrument approach, particularly during low ceiling/visibility conditions, is the point at which the pilot must decide whether to land or execute a missed approach. As the runway threshold is approached, the visual glidepath will separate into individual lights. At this point, the approach should be continued by reference to the runway touchdown zone markers. The ALS provides lights that will penetrate the atmosphere far enough from touchdown to give directional, distance, and glidepath information for safe visual transition. Visual identification of the ALS by the pilot must be instantaneous, so it is important to know the type of ALS before the approach is started. Check the instrument approach chart and the A/FD for the particular type of lighting facilities at the destination airport before any instrument flight. With reduced visibility, rapid orientation to a strange runway can be difficult, especially during a circling approach to an airport with minimum lighting facilities, or to a large terminal airport located in the midst of distracting city and ground facility lights. Some of the most common ALS systems are shown in Figure 22.

A high-intensity flasher system, often referred to as "the rabbit," is installed at many large airports. The flashers consist of a series of brilliant blue-white bursts of light flashing in sequence along the approach lights, giving the effect of a ball of light travelling towards the runway. Typically, "the rabbit" makes two trips toward the runway per second.

Runway end identifier lights (REIL) are installed for rapid and positive identification of the approach end of an instrument runway. The system consists of a pair of synchronized flashing lights placed laterally on each side of the runway threshold facing the approach area.

The visual approach slope indicator (VASI) gives visual descent guidance information during the approach to a runway. The standard VASI consists of light bars that project a visual glidepath, which provides safe obstruction clearance within the approach zone. The normal glide-slope angle is 3°; however, the angle may be as high as 4.5° for proper obstacle clearance. On runways served by ILS, the VASI angle normally coincides with the electronic glide-slope angle. Visual left/right course guidance is obtained by alignment with the runway lights. The standard VASI installation consists of either 2-, 3-, 4-, 6-, 12-, or 16-light units arranged in downwind and upwind light bars. Some airports serving long-bodied aircraft have three-bar VASIs which provide two visual glidepaths to the same runway. The first glidepath encountered is the same as provided by the standard VASI. The second glidepath is about 25 percent higher than the first and is designed for the use of pilots of long-bodied aircraft.

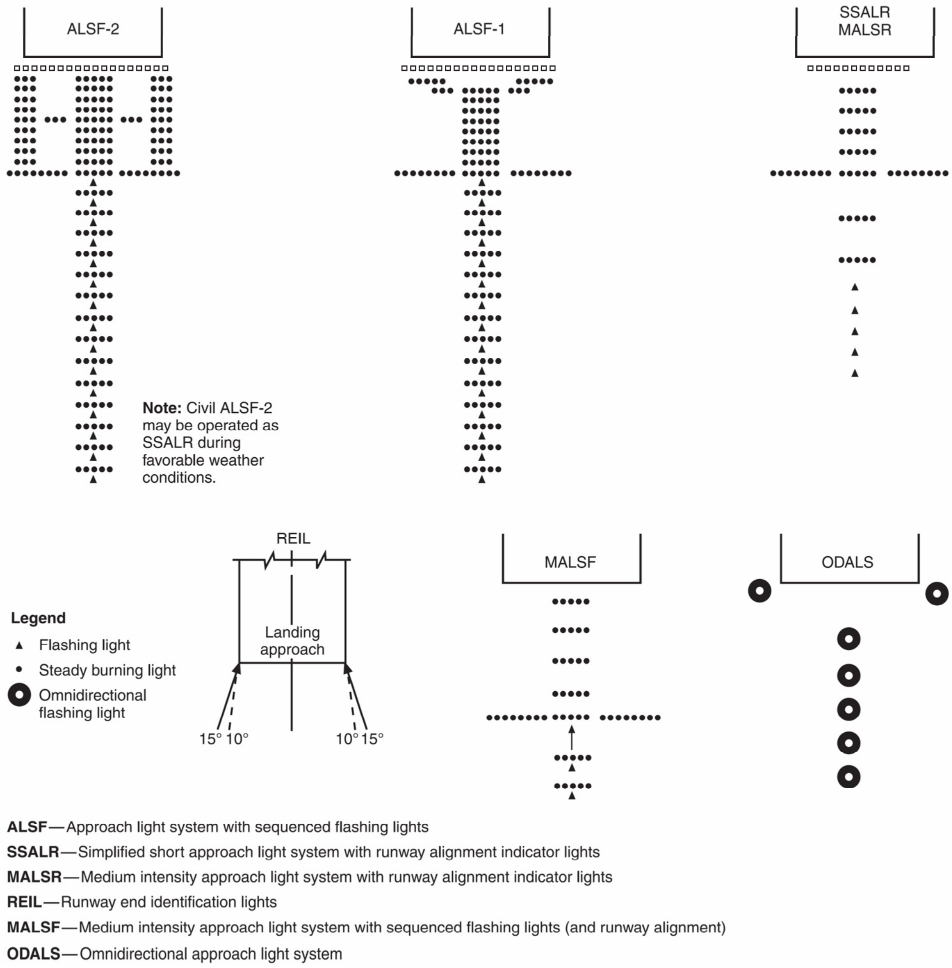


Figure 22 – Precision and nonprecision ALS configuration [ATC-3]

The basic principle of VASI is that of colour differentiation between red and white. Each light projects a beam having a white segment in the upper part and a red segment in the lower part of the beam. From a position above the glidepath the pilot sees both bars as white. Lowering the aircraft with respect to the glidepath, the colour of the upwind bars changes from white to pink to red. When on the proper glidepath, the landing aircraft will overshoot the downwind bars and undershoot the upwind bars. Thus the downwind (closer) bars are seen as white and the upwind bars as red. From a position below the glidepath both light bars are seen as red. Moving up to the glidepath, the colour of the downwind bars changes from red to pink to white. When below the glidepath, as indicated by a distinct all-red signal, a safe obstruction clearance might not exist. A standard two-bar VASI is illustrated in Figure 23.

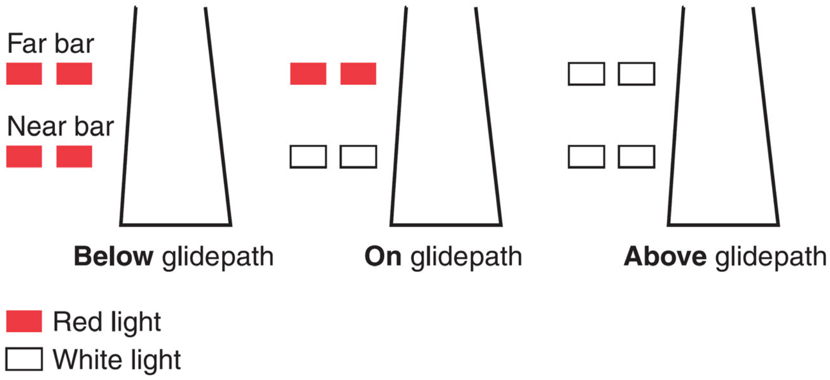


Figure 23 – Standard 2-bar VASI [ATC-3]

3.9 MICROWAVE LANDING SYSTEM (MLS)

The Microwave Landing System (MLS) provides precision approach navigation guidance. Transmitting in the frequency range of 5031 to 5091 MHz, it provides azimuth (left/right) and elevation (glide slope) information, displayed either on conventional CDIs or with multifunction cockpit displays. Range information is also provided.

MLS requires separate airborne equipment to receive and process the signals from what is normally installed in general aviation aircraft today. It has data communications capability, and can provide audible information about the condition of the transmitting system and other pertinent data such as weather, runway status, etc. The MLS transmits an audible identifier consisting of four letters beginning with the letter M, in Morse code at a rate of at least six per minute. The MLS system monitors itself and transmits ground-to-air data messages about the system's operational condition. During periods of routine or emergency maintenance, the coded identification is missing from the transmissions.

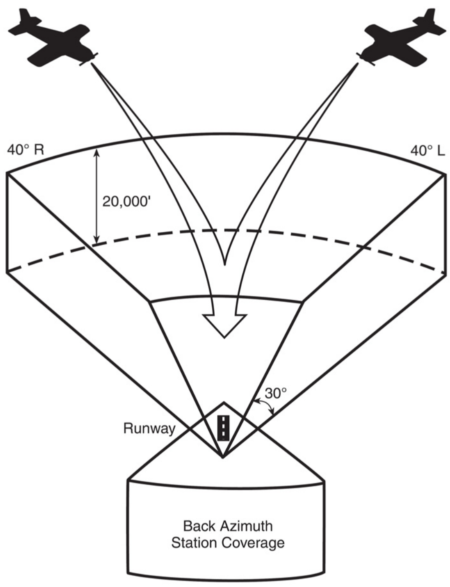


Figure 24 – MLS coverage volumes, 3-D representation [ATC-3]

The MLS is made up of an approach azimuth station (transmitter/antenna) with data transmission capability, an elevation station, a range station, and sometimes a back azimuth station.

3.9.1 Approach Azimuth Station

The approach azimuth station, unlike ILS, is able to provide approach guidance along any path within its ±40° (to runway alignment) range, like in Figure 24. Therefore, curved and segmented approaches are possible. This facility also provides the data communications capability of the system. This station is normally located about 1,000 feet beyond the stop end of the runway, but beyond this limitation there is considerable latitude in actual station location. A back azimuth station may be operating in conjunction with the approach azimuth station. If so, lateral guidance is available for missed approach and departure navigation.

3.9.2 Elevation Guidance Station

Like the approach azimuth station, the elevation guidance station has considerably more capability than the ILS glideslope system. Approach glidepath angles are selectable over a wide range up to at least 15°, with coverage to a maximum of 30°. This provides considerable flexibility for developing multipath approaches.

3.9.3 Range Guidance Station

The range guidance station transmits both normal and precision DME (DME/P) signals that function the same as normal DME (DME/N), with some technical and accuracy differences. Accuracy is improved to be consistent with the accuracy provided by the MLS azimuth and elevation stations.

4 Flight Management System (FMS)

The Flight Management System (FMS) is not a navigation system in itself. Rather, it is a system that automates the tasks of managing the onboard navigation systems. FMS may perform other onboard management tasks, but this discussion is limited to its navigation function [ATC-4].

FMS is an interface between flight crews and flight-deck systems. FMS can be thought of as a computer with a large database of airport and NAVAID locations and associated data, aircraft performance data, airways, intersections, DPs, and STARs. FMS also has the ability to accept and store numerous user-defined waypoints, flight routes consisting of departures, waypoints, arrivals, approaches, alternates, etc. FMS can quickly define a desired route from the aircraft's current position to any point in the world, perform flight plan computations, and display the total picture of the flight route to the crew.

FMS also has the capability of controlling (selecting) VOR, DME, and LOC NAVAIDs, and then receiving navigational data from them. INS, LORAN, and GPS navigational data may also be accepted by the FMS computer. The FMS may act as the input/output device for the onboard navigation systems, so that it becomes the "go-between" for the crew and the navigation systems.

At startup, the crew programs the aircraft location, departure runway, DP (if applicable), waypoints defining the route, approach procedure, approach to be used, and routing to alternate. This may be entered manually, be in the form of a stored flight plan, or be a flight plan developed in another computer and transferred by disk or electronically to the FMS computer. The crew enters this basic information in the control/display unit (CDU) shown in Figure 26.

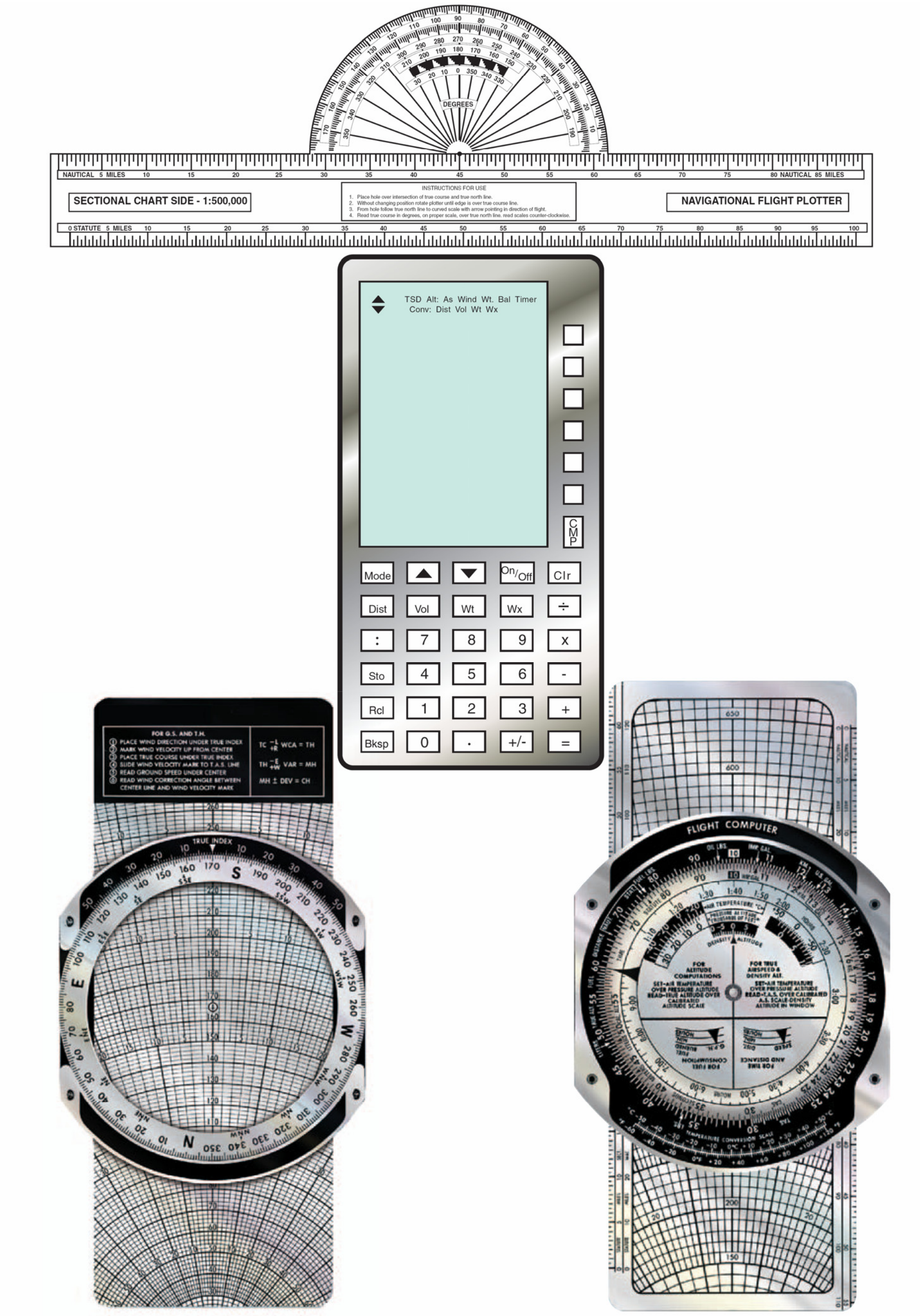


Figure 25 – A picture of the computational and wind side of a common mechanical computer, an electronic computer, and plotter [ATC-4]

Once airborne, the FMS computer channels the appropriate NAVAIDs and takes radial/distance information, or channels two NAVAIDs, taking the more accurate distance information. FMS then indicates position, track, desired heading, groundspeed and position relative to desired track. Position information from the FMS updates the INS. In more sophisticated aircraft, the FMS provides inputs to the HSI, RMI, glass cockpit navigation displays, head-up display (HUD), autopilot, and auto-throttle systems.

Figure 26 – FMS CDU in flight plan mode [ATC-4]

5 HEAD-UP DISPLAY (HUD)

The HUD is a display system that provides a projection of navigation and air data (airspeed in relation to approach reference speed, altitude, left/right and up/down glide slope) on a transparent screen between the pilot and the windshield [ATC-4].

Other information may be displayed, including a runway target in relation to the nose of the aircraft. This allows the pilot to see the information necessary to make the approach while also being able to see out the windshield, which diminishes the need to shift between looking at the panel to looking outside. Virtually any information desired can be displayed on the HUD if it is available in the aircraft's flight computer, and if the display is user-definable.

6 RADAR NAVIGATION (GROUND BASED)

Radar works by transmitting a pulse of RF energy in a specific direction. The return of the echo or bounce of that pulse from a target is precisely timed. From this, the distance travelled by the pulse and its echo is determined and displayed on a radar screen in such a manner that the distance and bearing to this target can be instantly determined. The radar transmitter must be capable of delivering extremely high power levels toward the airspace under surveillance, and the associated radar receiver must be able to detect extremely small signal levels of the returning echoes.

The radar display system provides the controller with a maplike presentation upon which appear all the radar echoes of aircraft within detection range of the radar facility. By means of electronically-generated range marks and azimuth indicating devices, the controller can locate each radar target with respect to the radar facility, or can locate one radar target with respect to another.

Another device, a video-mapping unit, generates an actual airway or airport map and presents it on the radar display equipment. Using the video-mapping feature, the air traffic controller not only can view the aircraft targets, but will see these targets in relation to runways, navigation aids, and hazardous ground obstructions in the area. Therefore, radar becomes a navigational aid, as well as the most significant means of traffic separation.

In a display presenting perhaps a dozen or more targets, a primary surveillance radar system cannot identify one specific radar target, and it may have difficulty "seeing" a small target at considerable distance—especially if there is a rain shower or thunderstorm between the radar site and the aircraft. This problem is solved with the Air Traffic Control Radar Beacon System (ATCRBS), sometimes called secondary surveillance radar (SSR), which utilizes a transponder in the aircraft. The ground equipment is an interrogating unit, with the beacon antenna mounted so it rotates with the surveillance antenna. The interrogating unit transmits a coded pulse sequence that actuates the aircraft transponder. The transponder answers the coded sequence by transmitting a preselected coded sequence back to the ground equipment, providing a strong return signal and positive aircraft identification, as well as other special data such as the aircraft's altitude. A transponder code consists of four numbers from zero to seven (4,096 possible codes). There are some standard codes, or ATC may issue a four-digit code to an aircraft. When a controller requests a code or function on the transponder, the word "squawk" may be used.

The radar systems used by ATC are air route surveillance radar (ARSR), airport surveillance radar (ASR), precision approach radar (PAR) and airport surface detection equipment (ASDE). Surveillance radars scan through 360° of azimuth and present target information on a radar display located in a tower or centre [ATC-3]. This information is used independently or in conjunction with other navigational aids in the control of air traffic.

ARSR is a long-range radar system designed primarily to cover large areas and provide a display of aircraft while en route between terminal areas. The ARSR enables air route traffic control centre (ARTCC) controllers to provide radar service when the aircraft are within the ARSR coverage. In some instances, ARSR may enable ARTCC to provide terminal radar services similar to but usually more limited than those provided by a radar approach control.

ASR is designed to provide relatively short-range coverage in the general vicinity of an airport and to serve as an expeditious means of handling terminal area traffic through observation of precise aircraft locations on a radarscope [AIRP-9]. Nonprecision instrument approaches are available at airports that have an approved surveillance radar approach procedure. ASR provides radar vectors to the final approach course, and then azimuth information to the pilot during the approach. Along with range (distance) from the runway, the pilot is advised of MDA, when to begin descent, and when at the MDA. If requested, recommended altitudes will be furnished each mile while on final.

PAR is designed to be used as a landing aid displaying range, azimuth, and elevation information rather than an aid for sequencing and spacing aircraft. PAR equipment may be used as a primary landing aid, or it may be used to monitor other types of approaches. Two antennas are used in the PAR array, one scanning a vertical plane, and the other scanning horizontally. Since the range is limited to 10 miles, azimuth to 20°, and elevation to 7°, only the final approach area is covered. The controller's scope is divided into two parts. The upper half presents altitude and distance information, and the lower half presents azimuth and distance.

The PAR is one in which a controller provides highly accurate navigational guidance in azimuth and elevation to a pilot. Pilots are given headings to fly, to direct them to, and keep their aircraft aligned with the extended centreline of the landing runway. They are told to anticipate glidepath interception approximately 10 to 30 seconds before it occurs and when to start descent. The published decision height (DH) will be given only if the pilot requests it. If the aircraft is observed to deviate above or below the glidepath, the pilot is given the relative amount of deviation by use of terms "slightly" or "well" and is expected to adjust the aircraft's rate of descent/ascent to return to the glidepath. Trend information is also issued with respect to the elevation of the aircraft and may be modified by the terms "rapidly" and "slowly"; e.g., "well above glidepath, coming down rapidly." Range from touchdown is given at least once each mile. If an aircraft is observed by the controller to proceed outside of specified safety zone limits in azimuth and/or elevation and continue to operate outside these prescribed limits, the pilot will be directed to execute a missed approach or to fly a specified course unless the pilot has the runway environment (runway, approach lights, etc.) in sight. Navigational guidance in azimuth and elevation is provided to the pilot until the aircraft reaches the published decision altitude (DA)/DH.

Advisory course and glidepath information is furnished by the controller until the aircraft passes over the landing threshold, at which point the pilot is advised of any deviation from the runway centreline. Radar service is automatically terminated upon completion of the approach.

Figure 27 – PAR Radar at Milan Malpensa Airport

The Airport Surface Detection Equipment SMR (Surface Movement Radar) is specifically designed to detect all principal features on the surface of an airport, including aircraft and vehicular traffic, and to present the entire image on a radar indicator console in the control tower. It is used to augment visual observation by tower personnel of aircraft and/or vehicular movements on runways and taxiways.

Radar Limitations [ATC-3]:

1. It is very important for the aviation community to recognize the fact that there are limitations to radar service and that ATC controllers may not always be able to issue traffic advisories concerning aircraft which are not under ATC control and cannot be seen on radar.
2. The characteristics of radio waves are such that they normally travel in a continuous straight line unless they are "bent" by abnormal atmospheric phenomena such as temperature inversions; reflected or attenuated by dense objects such as heavy clouds, precipitation, ground obstacles, mountains, etc.; or screened by high terrain features.
3. Primary radar energy that strikes dense objects will be reflected and displayed on the operator's scope thereby blocking out aircraft at the same range and greatly weakening or completely eliminating the display of targets at a greater range.
4. Relatively low altitude aircraft will not be seen if they are screened by mountains or are below the radar beam due to curvature of the Earth.
5. The amount of reflective surface of an aircraft will determine the size of the radar return. Therefore, a small light airplane or a sleek jet fighter will be more difficult to see on primary radar than a large commercial jet or military bomber.
6. All ARTCC radar in the conterminous U.S. and many airport surveillance radar have the capability to interrogate Mode C and display altitude information to the controller from appropriately equipped aircraft. However, a number of airport surveillance radar do not have Mode C display capability; therefore, altitude information must be obtained from the pilot.

7 AUTOMATIC DEPENDENT SURVEILLANCE (ADS)

The classic original idea of ADS was recommended to ICAO by the FANS/I Committee. Subsequently the developments of appropriate standards and guidance material has been entrusted to the ADS Panel, whose work provided the basis for the publication of an ICAO circular [AIRP-12]. This "classic" ADS is a surveillance technique for use by air traffic services in which aircraft automatically provide, via data-link, data derived from on-board position-fixing and navigation systems. ADS will allow controllers to obtain position data and other information from ADS equipped aircraft in a timely manner in accordance with their requirements, and will allow the aircraft to be tracked even in non-radar airspace.

The ADS application allows the implementation of reporting agreements, which, with the exception of an aircraft in an emergency situation, are established exclusively by the ground. An ADS agreement is an ADS reporting plan which establishes the conditions of ADS data reporting (i.e. data required by the ATC system and the frequency of the ADS reports which have to be agreed prior to the provision of the ADS-ATS services). The terms of an ADS agreement will be exchanged between the ground system and the aircraft by means of a contract, or a series of contracts. An ADS contract specifies under what conditions an ADS report would be initiated, and what data groups will be included in the reports. There are three types of contract:

- on demand, which provides an immediate report;
- periodic, which provides a report at a regular periodic intervals determined by the ground system;
- event, which provides a report when or if a specified event or events take place.

ADS contracts necessary for the control of the aircraft will be established with each aircraft by the relevant ground system, for at least the portion of the aircraft flight over which that ground system has jurisdiction. The contract may include the provision of a basic ADS reports at a periodic interval defined by the ground system with, optionally, one or more additional blocks containing specific information, which may or may be not be sent with each periodic report. The agreement may also provide for ADS reports at geographically-defined points such as way-points and intermediate points, in addition to other specific event-driven reports. The primary objective of the ADS application is to provide automated aircraft position data for ATC. The ADS application may also be useful in air traffic flow management (ATFM) and airspace management (ASM). ATM benefits from the use of the ADS application may include separation minima reduction, and more efficient use of the airspace. The first operational application of the "classic" ADS was implemented in the Pacific Ocean in 1995, taking advantage of character oriented protocols (not compatible with the ICAO standardised ATN) and satellite communications.

Although the application of ADS does not specifically encompass ATC communications, automation or procedures, all of these elements must be tailored to support the ADS application and to make meaningful use of the data. Thus, it is critical to consider the ATC automation and communication system as the foundation upon which an ADS-based ATC system is built. The ADS application and associated communications will have to be supported by advanced airborne and ground facilities and data-link communications with proven end-to end- integrity, reliability and availability, within the boundaries of acceptable cost and complexity. The implementation of the "classic" ADS, through reliable data-link communications and accurate aircraft navigation systems, therefore provides surveillance services in oceanic airspace and other areas where non-radar air traffic control services are currently provided.

In non-radar airspace, the effective use of ADS in air traffic services will facilitate the reduction of separation minima, enhance flight safety and better accommodate user preferred profiles, but the proper enhanced implementation of ADS may also provide benefits in en-route continental, terminal areas and on the airport surface. The automatic transmission of the aircraft position through ADS will replace present pilot position reports or complement radar information. The content and the frequency of reporting will be determined by the controlling ATC unit.

ADS basic information report contains the following information:

- the 3-D position of the aircraft (latitude, longitude, and altitude);
- the time;
- an indication of the accuracy of the position data information (figure of merit).

An ADS report may contain any (or all) of the following information too:

- aircraft identification;
- ground vector (*Track, Ground speed, Rate of climb or descent*);
- air vector (*Heading, Mach or IAS, Rate of climb or descent*);
- projected profile (*Next and Next+1 waypoint altitude and time*);
- meteorological information (*Wind vector, Temperature and Turbulence*);
- short term intent (*Lat, Lon, Alt and Time at projected intent point*);
- intermediate intent;
- extended projected profile.

If an altitude, track or speed change is predicted to occur between the aircraft's current position and the projected intent point (indicated above), additional information to the short term intent data would be provided as intermediate intent (repeated as necessary).

7.1 ADS BROADCAST (ADS-B)

ADS-B is a surveillance application transmitting parameters, such as position, ground track and ground speed, via a broadcast mode data link for utilisation by any air and/or ground user requiring it. This capability will permit enhanced airborne and ground situational awareness to provide for specific surveillance functions and co-operative pilot-controller and pilot-pilot ATM. The ADS-B application will not be limited to the traditional roles associated with ground based radar system. ADS-B will provide opportunities for new functionality both onboard the aircraft and within the ground ATC automation systems. Depending on the implementation, ADS-B may encompass both air-ground and air-air surveillance functionality, as well as applications between and among aircraft on the ground and ground vehicles. ADS-B will have many benefits in extending the range beyond that of secondary surveillance radar, particularly in airport surface and low altitude airspace, and air-to-air situational awareness.

Each ADS-B capable aircraft will periodically broadcast its position and other required data provided by the onboard navigation system. Any user, either airborne or ground based, within range of this broadcast may choose to receive and process this information. The aircraft originating the broadcast need have no knowledge of what systems are receiving its broadcast. Because broadcast data might be received by the ground station at a rate in excess of the requirements of the ATC system, some filtering and/or tracking may be necessary.

The ADS-B application supports improved use of airspace, reduced ceiling/visibility restrictions, improved surface surveillance, and enhanced safety. ADS-B equipage may be extended to vehicles on the airport surface movement area, uncharted obstacles not identified by a current NOTAM, and non-powered airborne vehicles or obstacles if a satisfactory cost/benefit ratio is achieved. ADS-B will support several services, including those designed for both air-ground and air-air use:

- ATC surveillance;
- Airborne situational awareness (i.e. cockpit display of traffic information, CDTI);
- Conflict detection (both airborne and ground based);
- Airborne Separation Assurance System;
- ATC conformance monitoring.

ADS-B will enhance ATC surveillance in the following ways:

- in a mixed ADS-B/radar surveillance environment, ADS-B data will complement or supplement radar data ADS-B can be used as back-up for SSR introducing a cheaper and dissimilar redundancy since it can be seen as an additional independent information channel;
- ADS-B will extend surveillance services into non-radar airspace such as low altitude airspace, remote airspace and coastal waters using economically acceptable line of sight communication means.

Airborne ADS-B is performed by airborne systems and applications that utilize surveillance information pertaining to other nearby aircraft. The users of ADS-B reports are the flight crews and normally independent of the ground ATC infrastructure.

7.2 INTEGRATION OF ADS AND SSR DATA

The safe operation of aircraft at close proximity requires an increase in the availability of very accurate positional data in order to apply separation closer to the minima and increase the airspace capacity. An enhancement of tracking algorithms is obviously necessary in order to take advantage of all available surveillance sources as well as processing new parameters related to aircraft motion. The primary objective of the ADS/SSR integration technique is to take advantage of the generalised ADS concept for implementation within areas covered by radar surveillance as well as transition areas between radar and ADS only coverage. Complete radar coverage in ADS/SSR airspace is not required, although outer horizontal limits should normally be coincident. In addition an ADS transition buffer is advisable. In areas where duplicate radar coverage is currently mandatory the integration of ADS might lead to a mitigation of that requirement, as well as for the provision of single radar coverage in homogeneous areas where the installation of radar system is not feasible or economically justifiable. The use of the ADS/SSR integration in areas already having multiple radar coverage will take advantage of additional data providing the system with the capability of making as uniform as possible track quality within radar covered airspace, thus overcoming residual radar shortcomings. ADS/SSR integration will result in the augmentation of surveillance performance in existing radar environments, as well as beyond radar coverage. The ADS/SSR integration will result also in a more reliable data availability for conflict detection and conformance monitoring function thus reducing the probability of false alarms of this function. This will be essentially due to kinematics data measured on-board and availability of aircraft intent.

ADS/SSR data integration will use different technique for “classic” ADS or ADS-Broadcast. Since the ADS technique relies upon the capability of an ADS facility to set up a contract with the aircraft to send reports with appropriate content and periodicity, the contract management function will play a key role in defining the most appropriate periodicity and content to optimise the ADS/SSR integration. The strategy to define the best contract for this function should take into account constraints on airspace and traffic scenario, as well as aircraft flight plan and communication infrastructure performances ADS-B relies upon the capability of an aircraft to transmit periodically a set of flight related information via a broadcast data link. ICAO indicates only a minimum set of information transmitted and maximum update period for various operational domain.

The ADS/SSR data integration can provide the following improvements to the surveillance function:

- automatic acquisition of certain airborne data containing parameters such as true track, speed, etc. (i.e. in general vector information), which will improve the ground tracking of aircraft.

- availability of surveillance data also when the radar limitations occur. These limitations are:
  - mechanical rotation of the radar antenna,
  - garbling, fruiting and splitting.
- coding of the altitude data in smaller altitude unit increments (about 3 meters or 10 feet) and the availability of the vertical rate, as provided by Ground Vector or Air Vector, which will improve the ability of ATC to monitor and make high quality predictions of aircraft trajectories in the vertical plane, thus improving the Short-Term Conflict Alert (STCA) function to significantly reduce the number of potential false alarms;
- automatic acquisition of aircraft call signs by ATC system, thus overcoming current problems connected with SSR code-call sign correlation and with radar identification and transfer procedures;
- acquisition of surveillance data, when satellite data link is used to support the ADS function, also when radar shortcomings such as line of sight propagation limitations (e.g., shadowing by orography, earth curvature, low level flight) become apparent, from properly equipped aircraft;
- minimisation of the number of SSRs required to supply mono-radar coverage, since ADS fills in the small areas not covered by them ("gap filler");
- increase of the level of availability using the ADS as one more level of dissimilar redundancy;
- availability of a means for a cross-check of ADS/Navigation data or radar integrity (NIM, Navigational Integrity Monitoring);
- possibility of adapting the degree of surveillance redundancy for each aircraft according to instantaneous ATC needs, thus providing redundancy in a very cost effective manner, and paving the way towards the Required Surveillance Performance (RSP) which may emerge in ICAO after the Required Navigation Performance (RNP) and the Required Communication Performance (RCP).

In summary the improvements above are applicable to integration between ADS and Mode A/C and Monopulse SSR.
Systems developed to support ASD/SSR data integration will be capable of meeting the communication performance appropriate to generate a reliable and effective data integration process.
When considering ADS/SSR integration, the following should be taken into account:

- performance requirements for ADS, including availability and integrity;
- accuracy of both Radar and ADS position reports;
- use of ADS data for example, as part of a data fusion and not just as back-up;
- trajectory prediction requirements;
- the development of a common surveillance processing system, where both the ADS and Radar tracks may be amalgamated to generate a single system track; and
- the synchronisation of both radar and ADS update rates.

7.3 GROUND MOVEMENTS

When the aircraft are moving on the airport surface useful data will be basic information with very high reporting rate, to insure that the required tracking accuracy is met.
ADS-B will also support ground conflict detection function. Airport surface conflict detection function shall detect conflicts including:

- potential collision with static aircraft;
- potential collision with moving aircraft;
- potential collision with known static obstacles;
- potential incursion into a restricted area
- potential incursion into a controlled area.

Basic ADS-B information, ground vector block, flight identification and/or airplane identification block and aircraft type or category will be also used for conflict detection function in airport surface

8 MODE S MULTILATERATION

The ICAO Manual of Surface Movement Guidance and Control Systems (SMGCS) describes how traffic should be controlled on the surface of an airport, based upon the principle of "see and be seen". It is recognised that the current SMGCS is not capable of delivering the required sophistication and capacity, particularly on complex airports and under conditions of reduced visibility. As a result of this A-SMGCS is being developed (A-stands for Advanced).

Multilateration, or hyperbolic positioning, is the process of locating an object based on the Time Difference of Arrival (TDOA) of a signal emitted from that object to three or more sensors. When a signal is transmitted from an object, it will be received by two spatially separate sensors at different times. The time difference is then used to calculate the objects position. By using three or more sensors, a complete position analysis can be attained.

For ATC applications, multilateration provides the same level of fleet coverage as traditional SSR (ie. all aircraft or vehicles equipped with an operational Mode A, Mode C or Mode S transponder). Multilateration will generally provide higher accuracy, greater update rate, better coverage and improved reliability when compared to traditional SSR, and will do so at a much lower initial cost and with lower annual maintenance costs.

Eurocontrol is focusing on the introduction of improved surveillance and procedures for air traffic controllers. Mode S Multilateration technology provides accurate surveillance and identification of all aircraft and transponder equipped vehicles on the airport surface.

Controllers see the result on a dedicated display screen, with each aircraft and vehicle securely tagged with its identification and position. This is also useful for the controllers in good visibility, particularly when multiple aircraft with similar colour schemes are manoeuvring in close proximity to one another (e.g. in a holding bay).

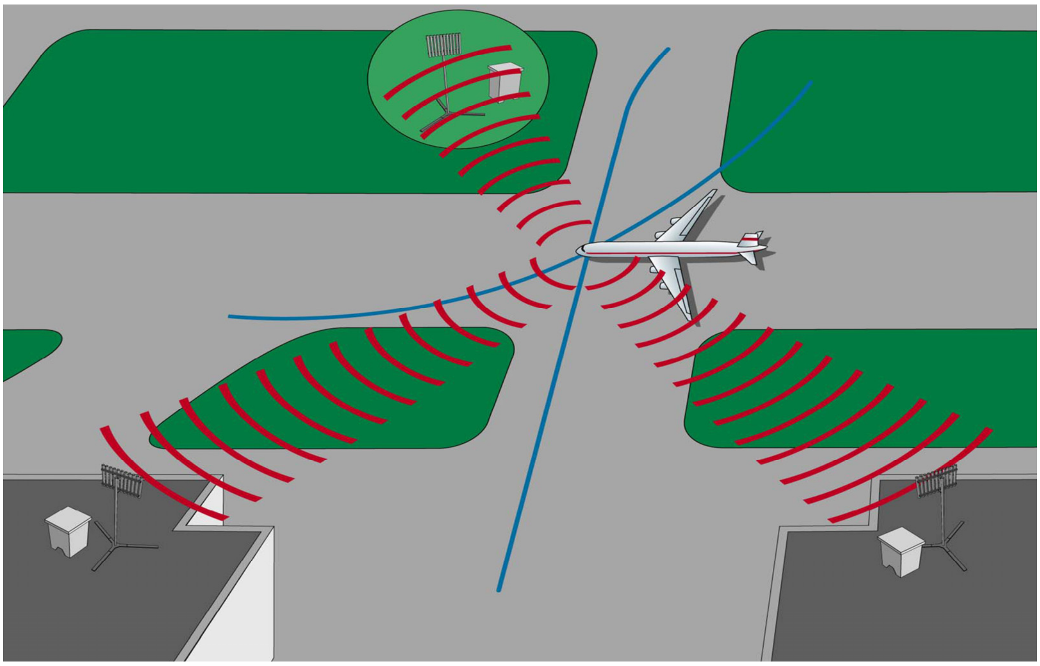


Figure 28 – The system normally uses three or more receivers to calculate the position of the aircraft or transponder equipped vehicle

The ground systems consists of a number of Receiver units, Receiver/Transmitter units, a Reference Transponder (all non-rotating sensors) and a Central Processing Station. The number of

Receiver and Receiver/transmitter units vary depending of the size and layout of the aerodrome. The Multilateration system uses multiple receivers to capture the "squitter" transmitted from the Mode S transponder. Then, by comparing the time difference, the system calculates the position. For aircraft the system will get the identity by selectively interrogating the transponder to receive the assigned Mode A code or the Aircraft Identification (i.e. The ICAO call sign used in flight inserted in the FMS or Transponder Control Panel). For transponder equipped vehicles the system will get the identity by the unique Mode S address transmitted by the transponder.

Mode S Multilateration systems have been already initially introduced at Amsterdam, Brussels, Copenhagen, Frankfurt, Geneva, London Heathrow, Milan (Malpensa), Milan (Linate), Paris Charles de Gaulle, Prague, Rome (Fiumicino), Vienna and Zurich Airports. Further implementation is foreseen at airports such as Palma De Majorca and Madrid. Widespread implementation within the USA is also planned.

Before Push back/Taxi the pilot will be requested to enter a Mode A code at start up (i.e. assigned Mode A code). This code will be either a discrete code or the non-discrete code 1000. Whenever the aircraft is capable of reporting Aircraft Identification, the Aircraft's Identification should also be entered through the FMS or the Transponder Control Panel. Flight crew must use the 3-letter ICAO designator of the operator, followed by flight identification number (e.g. BAW123, AFR456, SAS945…).

The ATC system will make the correlation with the flight plan either from the discrete code or, when the non-discrete code 1000 is entered, from the Aircraft Identification entered through the FMS or the Transponder Control Panel.

Pilots should ensure that the transponder is operating (i.e. XPNDR or the equivalent according to specific installation, AUTO if available, not OFF or STBY) and the assigned Mode A code selected from the request for push back or taxi whichever is earlier.

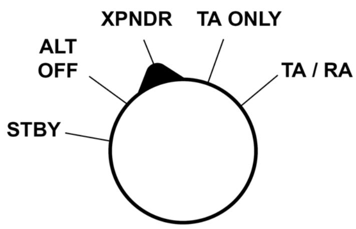


Figure 29 – TCAS should normally be selected at the holding position/point. After Landing and vacating the runway, TCAS should normally be deselected

After landing Pilots should ensure that the transponder is operating (i.e. XPNDR or the equivalent according to specific installation, AUTO if available, not OFF or STBY) after landing continuously until the aircraft is fully parked on stand.

Most aircraft are already equipped with Mode S transponders for the purposes of ACAS. Also from 2005 all aircraft flying IFR in the core area need to be suitably equipped in accordance with the planned implementation of Mode S Surveillance. Requirements relating to the ground operation of Mode S transponders were included in ICAO Annex 10, Vol. IV more than ten years ago. The latest version (Amendment 77 applicable from 28th of November 2002) states that when the transponder is switched on and not in the stand-by mode and in the on-the-ground status, only replies to all call transmissions shall be inhibited (used for acquisition by Mode S Radars). This status requirement is normally fulfilled automatically through a nose wheel weight switch. There aren't any new requirements, except that the transponder needs to remain operating on the ground with on-the-ground status, otherwise the Multilateration system cannot determine the position and identity of the aircraft. The Mode S transponder must be compliant with JAA Technical Standard Order JTSO-2C112a, or an equivalent standard that is compliant with the relevant ICAO SARPS and which is acceptable to the certification authority. For the purpose of IFR/GAT flights, existing SSR mode-A/C transponders must be replaced by Mode S transponders with effect from 31 March 2005.

In its surveillance role, multilateration does away with the need for expensive rotating radar antennas, replacing SSRs with several small and inexpensive stations strategically located to cover the same or greater airspace volume. Each unmanned station acts as a passive "listening post", instantaneously receiving every aircraft transponder transmission within line-of-sight range, out to the highest jet altitudes. One or more multilateration stations can be combined active/passive units, both transmitting transponder interrogation signals identical to an SSR and then listening to their responses. Multilateration stations can receive all transponder responses (i.e. basic Mode-A/C, Mode-S, military IFF and ADSB).

When incoming transponder signals are received by these stations, each immediately transfers its data to a centrally located processing unit the size of a filing cabinet, where advanced signal time-of-arrival and triangulation techniques are applied to determine the precise position of each aircraft These are then passed to the air traffic control (ATC) centre using standard protocols, including all the data normally provided by an SSR. However, the multilateration process is much faster, allowing controllers to track traffic every second, compared to viewing targets with each sweep of an SSR's rotating antenna.

The frequent tracking produces a very smooth trace on the controller's display screen rather than the progressive "jumps" characteristic of SSR targets. What's more, strategic location of the listening stations allows reception of aircraft signals in areas that are below SSR coverage or that are blocked by intervening structures. Position accuracy is also a strong point: over many evaluations, multilateration has been shown to be at least as accurate, and usually more so, than conventional SSR.

Multilateration installations are now operating at a large number of locations around the world. These applications range from long-range, high-altitude airspace surveillance over very large areas, to terminal area traffic monitoring, to precise tracking and display of aircraft and vehicle movements on the airport surface. Such installations are even being used to automate the collection of aircraft user fees.

Among countries that have adopted wide area multilateration is the Czech Republic, which has chosen this solution over SSR to accurately track, maintain separation and record the increasing number of high-altitude aircraft transiting its airspace to destinations beyond. The Czech Republic's ANSP has established what is arguably one of the largest airspace areas in the world to be covered by multilateration traffic surveillance, at least with respect to those areas where the provision of conventional SSR is considered too costly. In the terminal area, the tracking and separation of aircraft has been traditionally an exclusive SSR function. While unquestionably safe and efficient, SSR coverage of key areas can be hampered by local high terrain. This limitation led to the earliest certified use of multilateration for terminal airspace control at Ostrava, in the Czech Republic. Located within a wide horseshoe of mountains, Ostrava and its surrounding terminal airspace posed a difficult challenge for the operation of SSR. The ANSP consequently opted in 2001 for a multilateration network solution, which was commissioned in 2002. By the following year, sufficient data had been obtained on the system's performance to receive formal approval from the Civil Aviation Authority to reduce terminal area aircraft separation to three nautical miles from the typical five miles. As well, the low level coverage of the multilateration system permitted its exclusive use below 3,000 feet in the Ostrava terminal area, well below the surveillance coverage of regional radars.

Yet one of the most demanding applications of the multilateration technique is in its use for monitoring landing approaches and airport surface movements. To achieve this, a number of the system's small receivers are strategically placed around the airport and its runway approaches to produce clear and unobstructed line-of-sight displays of all air craft from the commencement of their multifinal approaches to their landings and subsequent taxiing to their terminals.

In the approach area, multilateration's very high one-second update rate provides controllers with a virtually continuous and extremely accurate picture of the approach stream, which is especially valuable in monitoring parallel runway operations. As a result, multilateration systems have recently been chosen by the authorities at Beijing's Capital International and Madrid's Barajas airports for this purpose - an application that previously had been the exclusive domain of radar.

Perhaps equally important is the system's application to monitoring movements on the aerodrome's surface, which paradoxically remains one of aviation's most hazardous operating environments. With the current emphasis on avoiding runway incursions at aerodromes around the world, surface surveillance has understandably increased in importance. In this respect, multilateration offers significant advantages. Inevitably, even local radars are "blanked" in some sectors by airport buildings and other obstructions, sometimes leaving significant areas of the airport surface uncovered and therefore invisible to controllers. In contrast, appropriately positioned multilateration stations can cover the total aerodrome surface without being affected by adverse weather conditions.

Aside from their installation in aircrafts, vehicles tracking units can be installed on all vehicles which use the operational areas. Small, affordable and quickly installed, these units can readily be seen on the controller's display, with their unique vehicular identification tags allowing quick differentiation from taxiing aircraft, and thereby allowing appropriate communication messages to be sent. In January 2007, the Dutch ANSP selected multilateration-based vehicles tracking units for over 300 service vehicles at Amsterdam Schiphol, following similar introductions at the Copenhagen, Prague, Santiago and Cape Town airports. At Schiphol, vehicles movement monitoring will be an integral part of the airport's advanced surface movement guidance and control system (A-SMGCS), as it will be in the multilateration-supported A-SMGCS at Beijing.

One unique application of the technology is in its very precise height measurement of aircraft overflying at high altitudes. The worldwide introduction of reduced vertical separation minima (RVSM) is unquestionably having a very beneficial effect on airspace capacity. Throughout most of the world, aircraft may now operate with 1000 feet vertical separation between 29000 and 44000 feet, compared to the previous 2000 feet of separation necessary because of the lower accuracy of earlier aircraft altimeters. Nevertheless, the importance that aviation attaches to redundancy dictates that independent checks of aircraft flying in RVSM airspace is a prudent step. Consequently, a growing number of RVSM monitoring stations have been established at key points along major traffic routes. At Linz, Austria, a purpose-built five-station multilateration system routinely measures the altitude of aircraft operating in RVSM airspace to an accuracy of 50 feet and passes this information – along with each aircraft's individual identification – to the Eurocontrol Centre in Bretigny.

Multilateration has also found use in tasks not directly concerned with air traffic control, but which provide unusually valuable services to airport administrations. For example, the system lends itself well to noise and curfew monitoring, and several small airports have adopted it for that purpose. More importantly, it provides very accurate details of the arrival and departure times of all aircraft - information that facilitates the automation of airport billing systems.

Airport operators typically employ monitors that observe aircraft movements and enter the data manually. Estimates are hard to come by, but it appears certain that a substantial amount of an airport's revenue may be lost under such circumstances. In December 2006, the Port Authority of New York and New Jersey issued a contract for provision of a specifically tailored system dedicated to the fully automatic tracking and billing of all aircraft movements at its Kennedy, LaGuardia, Newark and Teterboro airports. Combined, these airports handle over 1.4 million aircraft operations and over 94 million passengers per year, making it one of the largest airport systems in the world, with significant revenue generated from the accurate billing of users.

9 MULTILATERATION AND ADS-B WORKING TOGETHER.

It is not surprising that automatic dependent surveillance-broadcast (ADS-B) is poised to become a key element of the world's future air traffic management (ATM) system [AIRP-15]. Its benefits are now well understood, both by operators and air navigation service providers (ANSPs). Not only does ADS-B enhance safety while increasing capacity and efficiency, it also promises substantial cost savings. In announcing its ADS-B programme last year, for example, the U.S. Federal Aviation Administration (FAA) stated that it would permit the eventual decommissioning of much of the country's secondary surveillance radar (SSR) network, thereby saving about U.S. $1 billion. Airservices Australia launched a similar programme in 2005, and it too expects significant savings when use of ADS-B reaches the point where its SSR network can be safely retired.

Perhaps less well known is the fact that a number of ANSPs are already moving to a new surveillance technology which provides equivalent and often better performance than the traditional SSR, at a much lower acquisition and maintenance cost. But even more important is the fact that this technology, known as multilateration, can provide ANSPs with an economical foundation for their eventual transition to a full ADSB environment.

ADS-B is now entering service and will gradually spread to worldwide adoption over the next 10 to 15 years. In ADS-B each appropriately equipped aircraft automatically transmits bursts of data which include the aircraft's identification, altitude, track, speed, intent (i.e., climbing, descending or flying level), plus other information. These transmissions are received by ATC and also by all other ADS-B equipped aircraft within reception range, where they are presented on cockpit displays in a similar fashion to that of an ATC screen. The cockpit displays usually restrict the presentation to show other aircraft within a crew selectable altitude band of up to 3,000 feet above and below their aircraft, thereby providing pilots with exceptional situational awareness of the traffic of interest, with obvious safety benefits. Eventually, it may be possible, under certain circumstances for the system to be used by pilots for maintaining separation.

Today, relatively few aircrafts carry complete ADS-B installations, but a growing number are transmitting ADS-B data from transponders that are tracked by multilateration installations. In the future worldwide ADS-B environment, networks of unmanned, strategically located groundbased transceivers will be established. These will receive the ADS-B transmissions from all aircrafts within reception range and instantly retransmit this information to controllers' screens at the nearest ATC centre. In certain systems, the ATC centre can uplink weather, NOTAMs, and other important flight information to the aircrafts via same transceivers. During the lengthy transition to a complete ADS-B environment, the most important uplinked data will undoubtedly be that describing the flight paths of aircraft not yet equipped with ADS-B. These would be visible to controllers using radar or multilateration surveillance, but without the uplink would not be visible to the crew of an aircraft equipped for ADS-B. Details of unequipped traffic are provided to ADS-B equipped aircraft via the traffic information service-broadcast (TISB), one of the ADS-B system's two supporting features. The other, known as the flight information system-broadcast (FIS-B), will carry weather, NOTAMs and other priority information. Both uplink transmissions are specific to the ground-based transceiver's coverage areas.

For ANSP the critical questions are therefore those that concern the timing of their transition to a full ADS-B control environment and the associated investment in a network of dedicated ground-based transceivers to cover their airspace. The date by which ground-based transceivers will be needed in a specific area is hard to determine, being dependent on the rate at which aircraft operators install equipment in their aircraft. *One economical solution to this transition dilemma is a scheme in which the basic multilateration ground station incorporates the full functions of a ground-based transceiver for a future time when aircraft equipage is extensive*. During the interim period, *basic multilateration provides high performances coupled with low-cost acquisition and reduced maintenance expense*, either in covering additional airspace surveillance requirements or in replacing legacy SSRs should that be necessary before full ADS-B service is required.

The benefits of multilateration are being increasingly realized by ANSPs around the world, especially now that system's flexibility, breadth of application and economic advantages are being demonstrated daily. As in *communications*, where data links in certain applications are gradually superseding routine voice messages, and likewise in *navigation* – a field in which terrestrial aids are being supplanted by satellite positioning – technological evolution also characterizes surveillance, with multilateration and ADS-B expected to eventually replace radar in most, if not all, aviation applications.

But such transformations are in fact the normal state in air traffic management, and one can anticipate still more fundamental changes in the way aviation is conducted as the 21st century unfolds. Elements that were formerly regarded as being exclusively government-owned and operated, such as the traditional flag carrier, have already been widely privatized, and corporatization and privatization are spreading to other areas. Air Traffic Control, for example, is being transformed in a similar manner, as demonstrated by the growing number of privatized or partially privatized ANSPs. Along with this development, the provision of ATC related supporting services is beginning to change. In a significant step in this direction, the U.S. FAA recently announced that its nationwide ADS-B surveillance service – which it describes as the critical "backbone" of its Next Generation Air Transportation System (NGATS) – will be provided by a private industry contractor. Under the FAA contract, the selected organization will be responsible for the design, production, installation, support and ongoing maintenance of over 500 ADS-B ground-based transceivers across the United States. In a break with the past, the FAA will not own and operate the surveillance system; its sole commitment, along with the operators that will benefit from the system, will be to pay for the service. Yet another example of this trend was the FAA's transfer of the staffing and operation of many of its airport control towers and its general aviation flight service briefing activity to the industry.

Taking process a step further, one might expect that in the future, many of the world's air traffic support services will be provided by private organizations under exclusive agreements with national or regional ATM authorities. The aforementioned system provides a scalable network backbone to enable these types of surveillance service models, with clear benefits to all stakeholders of air navigation services.

SUMMARY OF CHAPTER 3

CHAPTER 4

AIRPORT OPERATIONS

Each time a pilot operates an airplane, the flight normally begins and ends at an airport. An airport may be a small sod field or a large complex utilized by air carriers. This chapter discusses airport operations and identifies features of a complex airport, as well as provides information on operating on or in the vicinity of an airport.

1 TYPES OF AIRPORTS

There are two types of airports [AIRP-3]:
• Controlled Airport
• Uncontrolled Airport

1.1 CONTROLLED AIRPORT

A controlled airport has an operating control tower. Air traffic control (ATC) is responsible for providing for the safe, orderly, and expeditious flow of air traffic at airports where the type of operations and/or volume of traffic requires such a service. Pilots operating from a controlled airport are required to maintain two-way radio communication with air traffic controllers, and to acknowledge and comply with their instructions. Pilots must advise ATC if they cannot comply with the instructions issued and request amended instructions. A pilot may deviate from an air traffic instruction in an emergency, but must advise ATC of the deviation as soon as possible.

1.2 UNCONTROLLED AIRPORT

An uncontrolled airport does not have an operating control tower. Two-way radio communications are not required, although it is a good operating practice for pilots to transmit their intentions on the specified frequency for the benefit of other traffic in the area. Table 1 lists recommended communication procedures. More information on radio communications will be discussed later in this chapter.

2 SOURCES FOR AIRPORT DATA

When a pilot flies into a different airport, it is important to review the current data for that airport. This data can provide the pilot with information, such as communication frequencies, services available, closed runways, or airport construction [ATC-3]. Three common sources of information are:

- Aeronautical Charts
- Airport/Facility Directory (A/FD)
- Notices to Airmen (NOTAMs)

2.1 AERONAUTICAL CHARTS

Aeronautical charts provide specific information on airports. An aeronautical chart is provided together with an aeronautical chart legend, which provides guidance on interpreting the information on the chart.

2.2 AIRPORT/FACILITY DIRECTORY

The Airport/Facility Directory (A/FD) provides the most comprehensive information on a given airport. It contains information on airports, heliports, and seaplane bases that are open to the public. The A/FDs are contained in seven books, which are organized by regions. These A/FDs are revised every 8 weeks. For a complete listing of information provided in an A/FD and how the information may be decoded, one should refer to the "Directory Legend Sample" located in the front of each A/FD. In the back of each A/FD, there is information such as special notices, parachute jumping areas, and facility telephone numbers. It would be helpful to review an A/FD to become familiar with the information it contains.

FACILITY AT AIRPORT	FREQUENCY USE	COMMUNICATION/BROADCAST PROCEDURES		
		OUTBOUND	INBOUND	PRACTICE INSTRUMENT APPROACH
UNICOM (No Tower or FSS)	Communicate with UNICOM station on published CTAF frequency (122.7, 122.8, 122.725, 122.975, or 123.0). If unable to contact UNICOM station, use self-announce procedures on CTAF.	Before taxiing and before taxiing on the runway for departure.	10 miles out. Entering downwind, base, and final. Leaving the runway.	
No Tower, FSS, or UNICOM	Self-announce on MULTICOM frequency 122.9	Before taxiing and before taxiing on the runway for departure.	10 miles out. Entering downwind, base, and final. Leaving the runway.	Departing final approach fix (name) or on final approach segment inbound.
No Tower in operation, FSS open	Communicate with FSS on CTAF frequency.	Before taxiing and before taxiing on the runway for departure.	10 miles out. Entering downwind, base, and final. Leaving the runway.	Approach completed/ terminated.
FSS closed (No Tower)	Self-announce on CTAF.	Before taxiing and before taxiing on the runway for departure.	10 miles out. Entering downwind, base, and final. Leaving the runway.	
Tower or FSS not in operation	Self-announce on CTAF.	Before taxiing and before taxiing on the runway for departure.	10 miles out. Entering downwind, base, and final. Leaving the runway.	

Table 1 – Recommended communication procedures [ATC-3]

2.3 NOTICES TO AIRMEN

Notices to Airmen (NOTAMs) provide the most current information available [AIRP-3]. They provide time-critical information on airports and changes that affect the national airspace system and are of concern to instrument flight rule (IFR) operations. NOTAM information is classified into three categories. These are NOTAM-D or distant, NOTAM-L or local, and flight data centre (FDC) NOTAMs. NOTAM-Ds are attached to hourly weather reports and are available at flight service stations (AFSS/FSS). NOTAM-Ls include items of a local nature, such as taxiway closures or construction near a runway. These NOTAMs are maintained at the FSS nearest the airport affected. NOTAM-Ls must be requested from an FSS other than the one nearest the local airport for which the NOTAM was issued. FDC NOTAMs are issued by the National Flight Data Center and contain regulatory information, such as temporary flight restrictions or an amendment to instrument approach procedures. The NOTAM-Ds and FDC NOTAMs are contained in the Notices to Airmen publication, which is issued every 28 days. Prior to any flight, pilots should check for any NOTAMs that could affect their intended flight.

3 AIRPORT MARKINGS AND SIGNS

There are markings and signs used at airports, which provide directions and assist pilots in airport operations [AIRP-3]. Some of the most common markings and signs will be discussed. Additional information may be found in the Aeronautical Information Manual (AIM).

3.1 RUNWAY MARKINGS

Runway markings vary depending on the type of operations conducted at the airport. Figure 1 shows a runway that is approved as a precision instrument approach runway and also shows some other common runway markings. A basic VFR runway may only have centreline markings and runway numbers. Since aircraft are affected by the wind during takeoffs and landings, runways are laid out according to the local prevailing winds. Runway numbers are in reference to magnetic north. Certain airports have two or even three runways laid out in the same direction. These are referred to as parallel runways and are distinguished by a letter being added to the runway number.
Examples are runway 36L (left), 36C (centre), and 36R (right).

Another feature of some runways is a displaced threshold. A threshold may be displaced because of an obstruction near the end of the runway. Although this portion of the runway is not to be used for landing, it may be available for taxiing, takeoff, or landing rollout.

Some airports may have a blast pad/stop-way area. The blast pad is an area where a propeller or jet blast can dissipate without creating a hazard. The stop-way area is paved in order to provide space for an airplane to decelerate and stop in the event of an aborted takeoff. These areas cannot be used for takeoff or landing.

3.2 TAXIWAY MARKINGS

Airplanes use taxiways to transition from parking areas to the runway. Taxiways are identified by a continuous yellow centreline stripe. A taxiway may include edge markings to define the edge of the taxiway. This is usually done when the taxiway edge does not correspond with the edge of the pavement. If an edge marking is a continuous line, the paved shoulder is not intended to be used by an airplane. If it is a dashed marking, an airplane may use that portion of the pavement. Where a taxiway approaches a runway, there may be a holding position marker. These consist of four yellow lines (two solid and two dashed). The solid lines are where the airplane is to hold. At some controlled airports, holding position markings may be found on a runway. They are used when there are intersecting runways, and air traffic control issues instructions such as "cleared to land—hold short of runway 30."

3.3 OTHER MARKINGS

Some of the other markings found on the airport include vehicle roadway markings, VOR receiver checkpoint markings, and non-movement area boundary markings. Vehicle roadway markings are used when necessary to define a pathway for vehicle crossing areas that are also intended for aircraft. These markings usually consist of a solid white line to delineate each edge of the roadway and a dashed line to separate lanes within the edges of the roadway. A VOR receiver checkpoint marking consists of a painted circle with an arrow in the middle. The arrow is aligned in the direction of the checkpoint azimuth. This allows pilots to check aircraft instruments with navigational aid signals.

A non-movement area boundary marking delineates a movement area under air traffic control. These markings are yellow and located on the boundary between the movement and non-movement area. They normally consist of two yellow lines (one solid and one dashed).

3.4 AIRPORT SIGNS

There are six types of signs that may be found at airports. The more complex the layout of an airport, the more important the signs become to pilots. Figure 2 shows examples of signs, their purpose, and appropriate pilot action. The six types of signs are:

• *Mandatory Instruction Signs*—have a red background with a white inscription. These signs denote an entrance to a runway, a critical area, or a prohibited area.

• *Location Signs*—are black with yellow inscription and a yellow border and do not have arrows. They are used to identify a taxiway or runway location, to identify the boundary of the runway, or identify an instrument landing system (ILS) critical area.

• *Direction Signs*—have a yellow background with black inscription. The inscription identifies the designation of the intersecting taxiway(s) leading out of an intersection.

• *Destination Signs*—have a yellow background with black inscription and also contain arrows. These signs provide information on locating things, such as runways, terminals, cargo areas, and civil aviation areas.

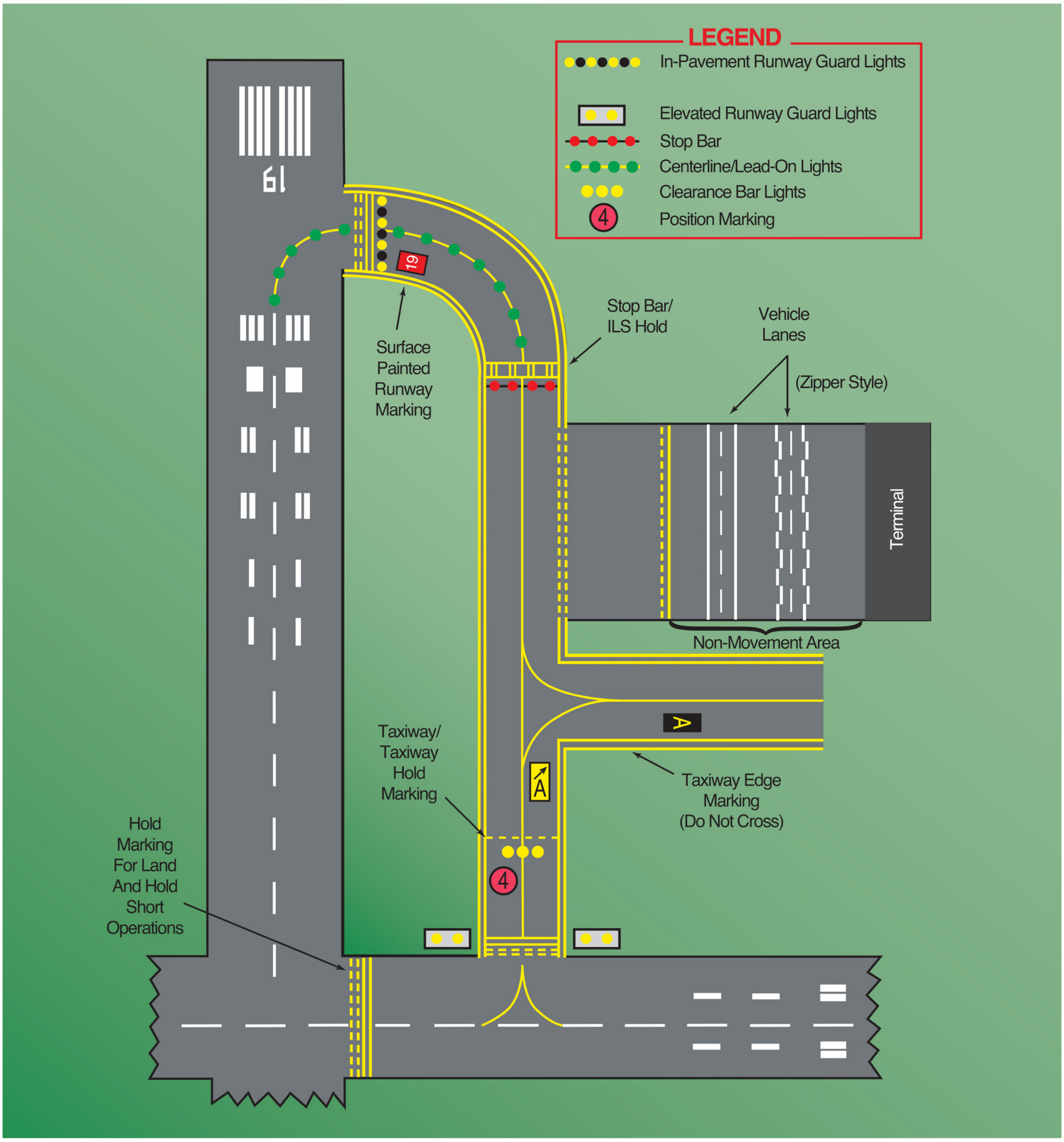


Figure 1 – Selected airport markings and surface lighting [ATC-3]

• *Information Signs*—have a yellow background with black inscription. These signs are used to provide the pilot with information on such things as areas that cannot be seen from the control tower, applicable radio frequencies, and noise abatement procedures. The airport operator determines the need, size, and location of these signs.

• *Runway Distance Remaining Signs*—have a black background with white numbers. The numbers indicate the distance of the remaining runway in thousands of feet.

AIRPORT SIGN SYSTEMS			
TYPE OF SIGN AND ACTION OR PURPOSE		*TYPE OF SIGN AND ACTION OR PURPOSE*	
4-22	**Taxiway/Runway Hold Position:** Hold short of runway on taxiway		**Runway Safety Area/Obstacle Free Zone Boundary:** Exit boundary of runway protected areas
26-8	**Runway/Runway Hold Position:** Hold short of intersecting runway		**ILS Critical Area Boundary:** Exit boundary of ILS critical area
8-APCH	**Runway Approach Hold Position:** Hold short of aircraft on approach	J →	**Taxiway Direction:** Defines direction & designation of intersecting taxiway(s)
ILS	**ILS Critical Area Hold Position:** Hold short of ILS approach critical area	↙ L	**Runway Exit:.** Defines direction & designation of exit taxiway from runway
⊖	**No Entry:** Identifies paved areas where aircraft entry is prohibited	22 ↑	**Outbound Destination:** Defines directions to takeoff runways
B	**Taxiway Location:** Identifies taxiway on which aircraft is located	↖ MIL	**Inbound Destination:** Defines directions for arriving aircraft
22	**Runway Location:** Identifies runway on which aircraft is located		**Taxiway Ending Marker** Indicates taxiway does not continue
4	**Runway Distance Remaining** Provides remaining runway length in 1,000 feet increments	↙A G L →	**Direction Sign Array:** Identifies location in conjunction with multiple intersecting taxiways

Figure 2 – Airport signs [ATC-3]

4 AIRPORT LIGHTING

The majority of airports have some type of lighting for night operations. The variety and type of lighting systems depends on the volume and complexity of operations at a given airport. Airport lighting is standardized [AIRP-3] so that airports use the same light colours for runways and taxiways.

4.1 AIRPORT BEACON

Airport beacons help a pilot identify an airport at night. The beacons are operated from dusk till dawn and sometimes they are turned on if the ceiling is less than 1,000 feet and/or the ground visibility is less than 3 statute miles (visual flight rules minimums). However, there is no requirement for this, so a pilot has the responsibility of determining if the weather is VFR.

The beacon has a vertical light distribution to make it most effective from 1-10° above the horizon, although it can be seen well above or below this spread. The beacon may be an omnidirectional capacitor-discharge device, or it may rotate at a constant speed, which produces the visual effect of flashes at regular intervals.

The combination of light colours from an airport beacon indicates the type of airport (see Figure 3).

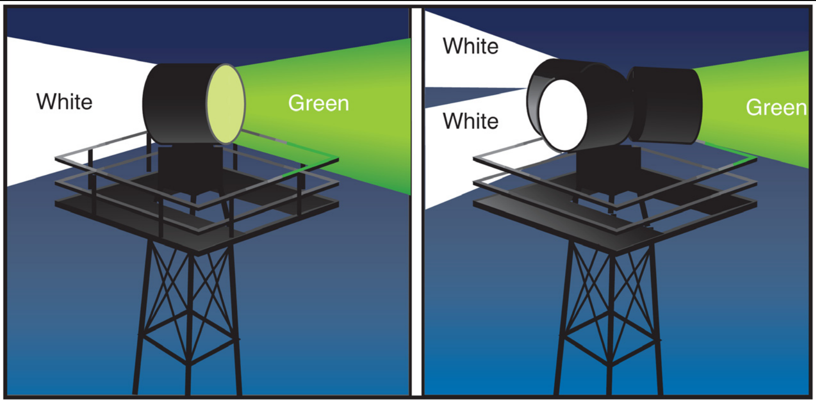


Figure 3 – Airport rotating beacons

Some of the most common beacons are:
• Flashing white and green for civilian land airports.
• Flashing white and yellow for a water airport.
• Flashing white, yellow, and green for a heliport.
• Two quick white flashes followed by a green flash identifies a military airport.

4.2 APPROACH LIGHT SYSTEMS

Approach light systems are primarily intended to provide a means to transition from instrument flight to visual flight for landing. The system configuration depends on whether the runway is a precision or non-precision instrument runway. Some systems include sequenced flashing lights, which appear to the pilot as a ball of light travelling towards the runway at high speed. Approach lights can also aid pilots operating under VFR at night.

4.3 VISUAL GLIDESLOPE INDICATORS

Visual glideslope indicators provide the pilot with glidepath information that can be used for day or night approaches. By maintaining the proper glidepath as provided by the system, a pilot should have adequate obstacle clearance and should touch down within a specified portion of the runway.

4.4 VISUAL APPROACH SLOPE INDICATOR

Visual approach slope indicator (VASI) installations are the most common visual glidepath systems in use. The VASI provides obstruction clearance within 10° of the runway extended runway centreline, and to 4 nautical miles (NM) from the runway threshold.
AVASI consists of light units arranged in bars. There are 2-bar and 3-bar VASIs. The 2-bar VASI has near and far light bars and the 3-bar VASI has near, middle, and far light bars. Two-bar VASI installations provide one visual glidepath which is normally set at 3°. The 3-bar system provides two glidepaths with the lower glidepath normally set at 3° and the upper glidepath one-fourth degree above the lower glidepath.
The basic principle of the VASI is that of colour differentiation between red and white. Each light unit projects a beam of light having a white segment in the upper part of the beam and a red segment in the

lower part of the beam. The lights are arranged so the pilot will see the combination of lights shown in Figure 4 to indicate below, on, or above the glidepath.

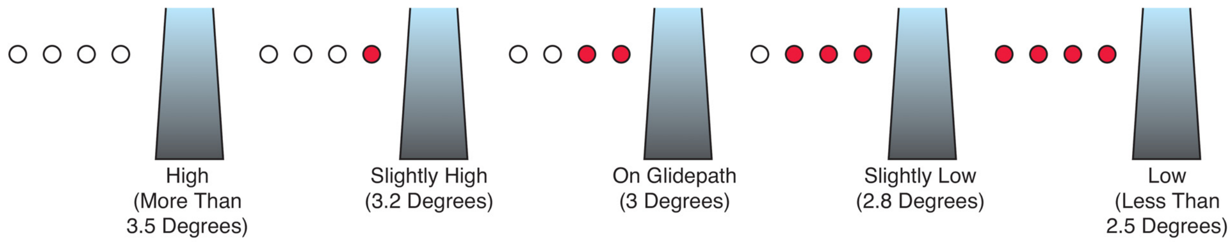


Figure 4 – Precision approach path indicator

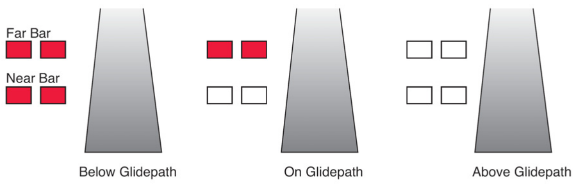


Figure 5 – 2-Bar VASI system

4.5 OTHER GLIDEPATH SYSTEMS

A precision approach path indicator (PAPI) uses lights similar to the VASI system except they are installed in a single row, normally on the left side of the runway, like in Figure 5.
A tri-colour system consists of a single light unit projecting a three-color visual approach path. A below the glidepath indication is red, on the glidepath colour is green, and above the glidepath is indicated by amber. When descending below the glidepath, there is a small area of dark amber. Pilots should not mistake this area for an "above the glidepath" indication as per Figure 6. There are also pulsating systems, which consist of a single light unit projecting a two-colour visual approach path. A below the glidepath indication is shown by a steady red light, slightly below is indicated by pulsating red, on the glidepath is indicated by a steady white light, and a pulsating white light indicates above the glidepath, as per Figure 7.

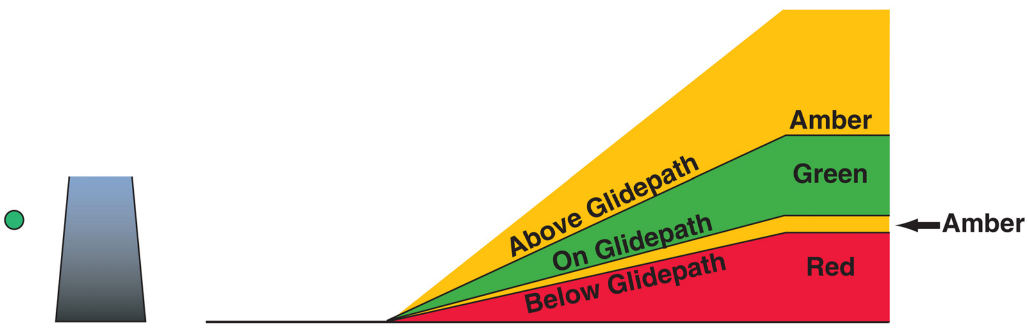


Figure 6 – Tri-color visual approach slope indicator [ATC-3]

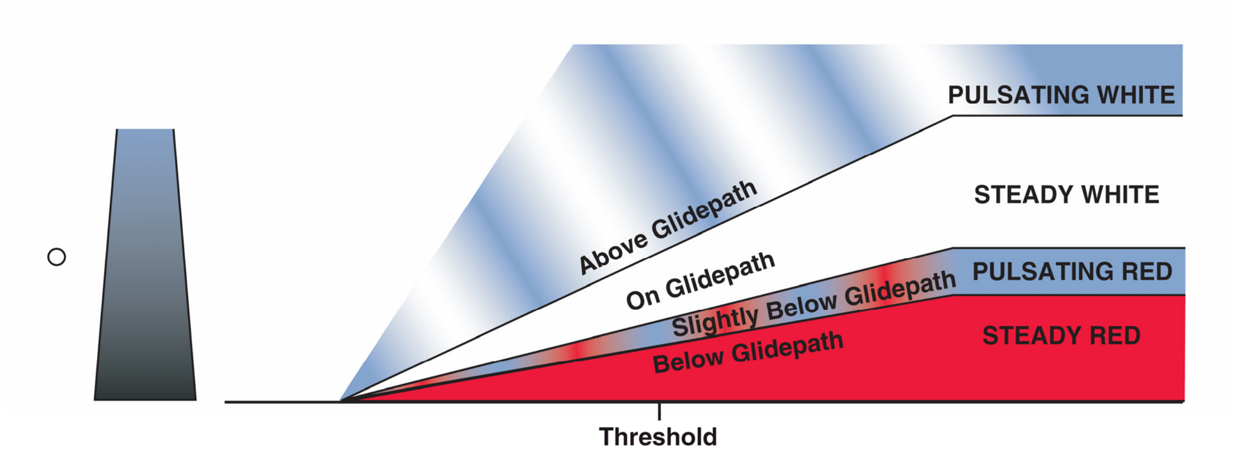


Figure 7 – Pulsating visual approach slope indicator [ATC-3]

4.6 Runway Lighting

There are various lights that identify parts of the runway complex. These assist a pilot in safely making a takeoff or landing during night operations.

4.6.1 Runway End Identifier Lights

Runway end identifier lights (REIL) are installed at many airfields to provide rapid and positive identification of the approach end of a particular runway. The system consists of a pair of synchronized flashing lights located laterally on each side of the runway threshold. REILs may be either omnidirectional or unidirectional facing the approach area.

4.6.2 Runway Edge Lights

Runway edge lights are used to outline the edges of runways at night or during low visibility conditions. These lights are classified according to the intensity they are capable of producing. They are classified as high intensity runway lights (HIRL), medium intensity runway lights (MIRL), or low intensity runway lights (LIRL). The HIRL and MIRL have variable intensity settings. These lights are white, except on instrument runways, where amber lights are used on the last 2,000 feet or half the length of the runway, whichever is less. The lights marking the end of the runway are red.

4.6.3 In-Runway Lighting

Touchdown zone lights (TDZL), runway centreline lights (RCLS), and taxiway turnoff lights are installed on some precision runways to facilitate landing under adverse visibility conditions. TDZLs are two rows of transverse light bars disposed symmetrically about the runway centreline in the runway touchdown zone. RCLS consists of flush centreline lights spaced at 50 foot intervals beginning 75 feet from the landing threshold. Taxiway turnoff lights are flush lights, which emit a steady green colour.

4.7 CONTROL OF AIRPORT LIGHTING

Airport lighting is controlled by air traffic controllers at controlled airports. At uncontrolled airports, the lights may be on a timer, or where an FSS is located at an airport, the FSS personnel may control the lighting. A pilot may request various light systems be turned on or off and also request a specified intensity, if available, from ATC or FSS personnel. At selected uncontrolled airports, the pilot may control the lighting by using the radio. This is done by selecting a specified frequency and clicking the radio microphone. For information on pilot controlled lighting at various airports, refer to the Airport/Facility Directory, as in Table 2.

KEY MIKE	FUNCTION
7 times within 5 seconds	Highest intensity available
5 times within 5 seconds	Medium or lower intensity (Lower REIL or REIL off)
3 times within 5 seconds	Lowest intensity available (Lower REIL or REIL off)

Table 2 – Radio control runway lighting [ATC-4]

4.8 TAXIWAY LIGHTS

Omnidirectional taxiway lights outline the edges of the taxiway and are blue in colour. At many airports, these edge lights may have variable intensity settings that may be adjusted by an air traffic controller when deemed necessary or when requested by the pilot.
Some airports also have taxiway centreline lights that are green in colour.

4.9 OBSTRUCTION LIGHTS

Obstructions are marked or lighted to warn pilots of their presence during daytime and night-time conditions. Obstruction lighting can be found both on and off an airport to identify obstructions. They may be marked or lighted in any of the following conditions.

• Red Obstruction Lights—either flash or emit a steady red colour during night-time operations, and the obstructions are painted orange and white for daytime operations.

• High Intensity White Obstruction Light—flashes high intensity white lights during the daytime with the intensity reduced for night-time.

• Dual Lighting—is a combination of flashing red beacons and steady red lights for night-time operation, and high intensity white lights for daytime operations.

5 WIND DIRECTION INDICATORS

It is important for a pilot to know the direction of the wind. At facilities with an operating control tower, this information is provided by ATC. Information may also be provided by FSS personnel located at a particular airport or by requesting information on a common traffic advisory frequency (CTAF) at airports that have the capacity to receive and broadcast on this frequency.
When none of these services is available, it is possible to determine wind direction and runway in use by visual wind indicators. A pilot should check these wind indicators even when information is provided on the CTAF at a given airport because there is no assurance that the information provided is accurate.

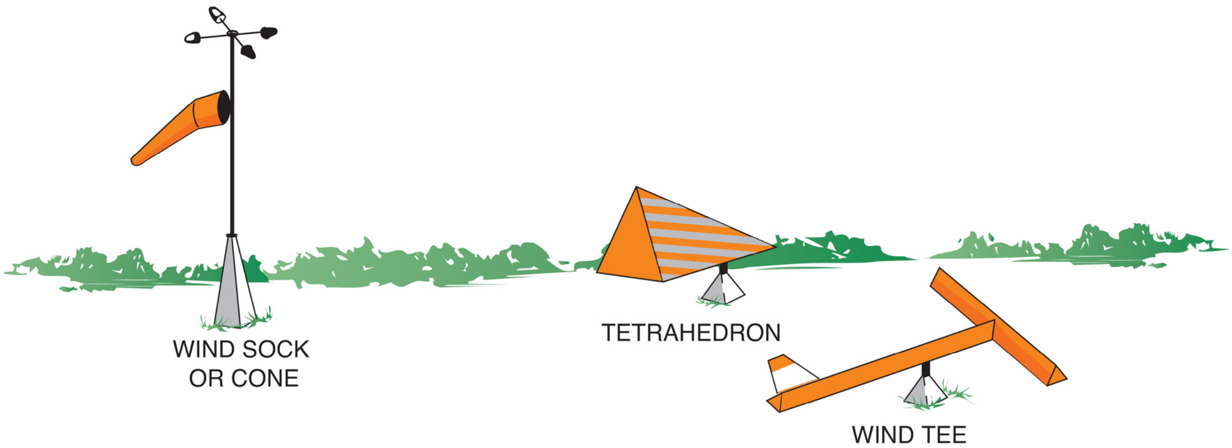


Figure 8 – Wind direction indicators [ATC-3]

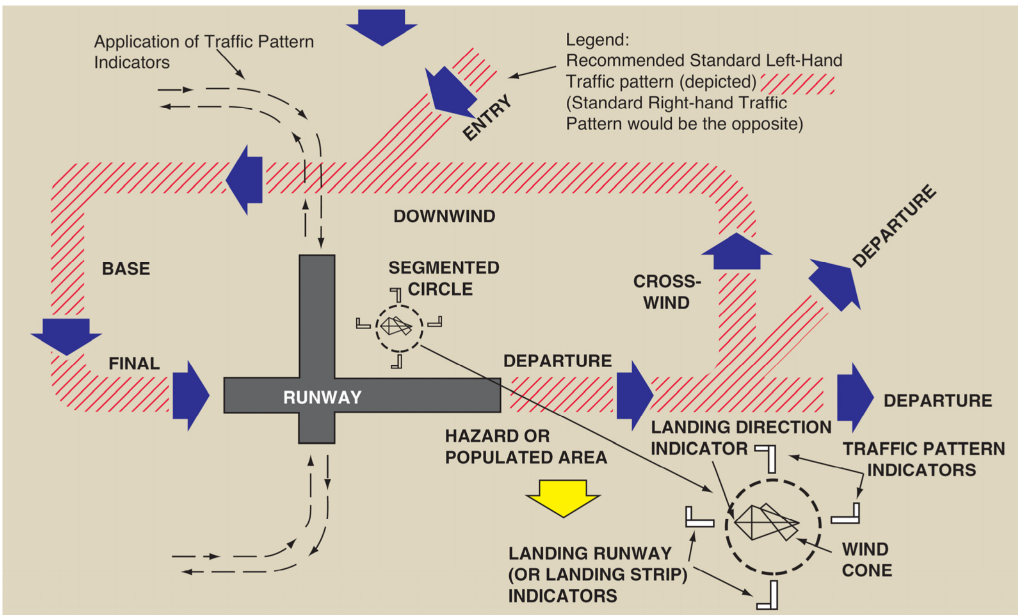


Figure 9 – Segmented circle and airport traffic pattern [ATC-3]

Wind direction indicators include a wind sock, wind tee, or tetrahedron. These are usually located in a central location near the runway and may be placed in the centre of a segmented circle, which will identify the traffic pattern direction, if it is other than the standard left-hand pattern, as per Figure 9. The wind sock is a good source of information since it not only indicates wind direction, but allows the pilot to estimate the wind velocity and gusts or factor. The wind sock extends out straighter in strong winds and will tend to move back and forth when the wind is gusty. Wind tees and tetrahedrons can swing freely, and will align themselves with the wind direction. The wind tee and tetrahedron can also be manually set to align with the runway in use; therefore, a pilot should also look at the wind sock, if available.

6 RADIO COMMUNICATIONS

Operating in and out of a controlled airport, as well as in a good portion of the airspace system, requires that an aircraft have two-way radio communication capability. For this reason, a pilot should be knowledgeable of radio station license requirements and radio communications equipment and procedures.

6.1 RADIO EQUIPMENT

In general aviation, the most common types of radios are VHF. A VHF radio operates on frequencies between 118.0 and 136.975 and is classified as 720 or 760 depending on the number of channels it can accommodate. The 720 and 760 uses .025 spacing (118.025, 118.050) with the 720 having a frequency range up to 135.975 and the 760 going up to 136.975. VHF radios are limited to line of sight transmissions; therefore, aircraft at higher altitudes are able to transmit and receive at greater distances. Using proper radio phraseology and procedures will contribute to a pilot's ability to operate safely and efficiently in the airspace system. A review of the Pilot/Controller Glossary contained in the Aeronautical Information Manual (AIM) will assist a pilot in the use and understanding of standard terminology. The AIM also contains many examples of radio communications, which should be helpful.

The International Civil Aviation Organization (ICAO) has adopted a phonetic alphabet, which should be used in radio communications. When communicating with ATC, pilots should use this alphabet to identify their aircraft, like in Table 3.

6.2 LOST COMMUNICATION PROCEDURES

It is possible that a pilot might experience a malfunction of the radio. This might cause the transmitter, receiver, or both to become inoperative. If a receiver becomes inoperative and a pilot needs to land at a controlled airport, it is advisable to remain outside or above Class D airspace until the direction and flow of traffic is determined. A pilot should then advise the tower of the aircraft type, position, altitude, and intention to land. The pilot should continue, enter the pattern, report a position as appropriate, and watch for light signals from the tower. Light signal colours and their meanings are contained in Table 4.

If the transmitter becomes inoperative, a pilot should follow the previously stated procedures and also monitor the appropriate air traffic control frequency. During daylight hours air traffic control transmissions may be acknowledged by rocking the wings, and at night by blinking the landing light.

When both receiver and transmitter are inoperative, the pilot should remain outside of Class D airspace until the flow of traffic has been determined and then enter the pattern and watch for light signals.

If a radio malfunctions prior to departure, it is advisable to have it repaired, if possible. If this is not possible, a call should be made to air traffic control and the pilot should request authorization to depart without two-way radio communications. If authorization is given to depart, the pilot will be advised to monitor the appropriate frequency and/or watch for light signals as appropriate.

CHARACTER	MORSE CODE	TELEPHONY	PHONIC (PRONUNCIATION)
A	•-	Alfa	(AL-FAH)
B	-•••	Bravo	(BRAH-VOH)
C	-•-•	Charlie	(CHAR-LEE) OR (SHAR-LEE)
D	-••	Delta	(DELL-TAH)
E	•	Echo	(ECK-OH)
F	••-•	Foxtrot	(FOKS-TROT)
G	--•	Golf	(GOLF)
H	••••	Hotel	(HOH-TEL)
I	••	India	(IN-DEE-AH)
J	•---	Juliet	(JEW-LEE-ETT)
K	-•-	Kilo	(KEY-LOH)
L	•-••	Lima	(LEE-MAH)
M	--	Mike	(MIKE)
N	-•	November	(NO-VEM-BER)
O	---	Oscar	(OSS-CAH)
P	•--•	Papa	(PAH-PAH)
Q	--•-	Quebec	(KEH-BECK)
R	•-•	Romeo	(ROW-ME-OH)
S	•••	Sierra	(SEE-AIR-RAH)
T	-	Tango	(TANG-GO)
U	••-	Uniform	(YOU-NEE-FORM) OR (OO-NEE-FORM)
V	•••-	Victor	(VIK-TAH)
W	•--	Whiskey	(WISS-KEY)
X	-••-	Xray	(ECKS-RAY)
Y	-•--	Yankee	(YANG-KEY)
Z	--••	Zulu	(ZOO-LOO)
1	•----	One	(WUN)
2	••---	Two	(TOO)
3	•••--	Three	(TREE)
4	••••-	Four	(FOW-ER)
5	•••••	Five	(FIFE)
6	-••••	Six	(SIX)
7	--•••	Seven	(SEV-EN)
8	---••	Eight	(AIT)
9	----•	Nine	(NIN-ER)
0	-----	Zero	(ZEE-RO)

Table 3 – Phonetic alphabet

LIGHT GUN SIGNALS			
COLOR AND TYPE OF SIGNAL	MOVEMENT OF VEHICLES, EQUIPMENT AND PERSONNEL	AIRCRAFT ON THE GROUND	AIRCRAFT IN FLIGHT
STEADY GREEN	Cleared to cross, proceed or go	Cleared for takeoff	Cleared to land
FLASHING GREEN	Not applicable	Cleared for taxi	Return for landing (to be followed by steady green at the proper time)
STEADY RED	STOP	STOP	Give way to other aircraft and continue circling
FLASHING RED	Clear the taxiway/runway	Taxi clear of the runway in use	Airport unsafe, do not land
FLASHING WHITE	Return to starting point on airport	Return to starting point on airport	Not applicable
ALTERNATING RED AND GREEN	Exercise Extreme Caution!!!!	Exercise Extreme Caution!!!!	Exercise Extreme Caution!!!!

Table 4 – Light gun signals [ATC-3]

7 WAKE TURBULENCE

All aircraft generate a wake while in flight. This disturbance is caused by a pair of counter-rotating vortices trailing from the wingtips. The vortices from larger aircraft pose problems to encountering aircraft. The wake of these aircraft can impose rolling moments exceeding the roll-control authority of the encountering aircraft. Also, the turbulence generated within the vortices can damage aircraft components and equipment if encountered at close range. For this reason, a pilot must envision the location of the vortex wake and adjust the flight-path accordingly.

During ground operations and during takeoff, jet-engine blast (thrust stream turbulence) can cause damage and upsets at close range. For this reason, pilots of small aircraft should consider the effects of jet-engine blast and maintain adequate separation.

Also, pilots of larger aircraft should consider the effects of their aircraft's jet-engine blast on other aircraft and equipment on the ground.

7.1 VORTEX GENERATION

Lift is generated by the creation of a pressure differential over the wing surface. The lowest pressure occurs over the upper wing surface, and the highest pressure under the wing. This pressure differential triggers the rollup of the airflow aft of the wing resulting in swirling air masses trailing downstream of the wingtips. After the rollup is completed, the wake consists of two counter rotating cylindrical vortices. Most of the energy is within a few feet of the centre of each vortex, but pilots should avoid a region within about 100 feet of the vortex core (see Figure 10).

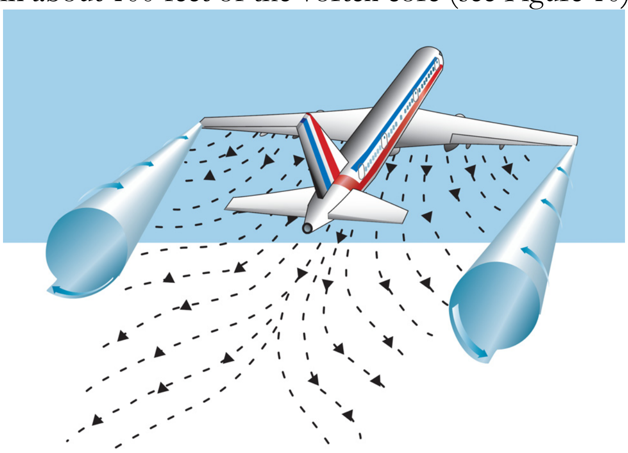


Figure 10 – Figure Vortex generation

7.2 VORTEX STRENGTH

The strength of the vortex is governed by the weight, speed, and shape of the wing of the generating aircraft. The vortex characteristics of any given aircraft can also be changed by the extension of flaps or other wing configuration devices as well as by a change in speed. The greatest vortex strength occurs when the generating aircraft is heavy, clean, and slow.

7.3 VORTEX BEHAVIOUR

Trailing vortices have certain behavioural characteristics that can help a pilot visualize the wake location and take avoidance precautions. Vortices are generated from the moment an aircraft leaves the ground, since trailing vortices are the by-product of wing lift. The vortex circulation is outward, upward, and around the wingtips when viewed from either ahead or behind the aircraft. Tests have shown that vortices remain spaced a bit less than a wingspan apart, drifting with the wind, at altitudes greater than a wingspan from the ground. Tests have also shown that the vortices sink at a rate of several hundred feet per minute, slowing their descent and diminishing in strength with time and distance behind the generating aircraft (see Figure 11).

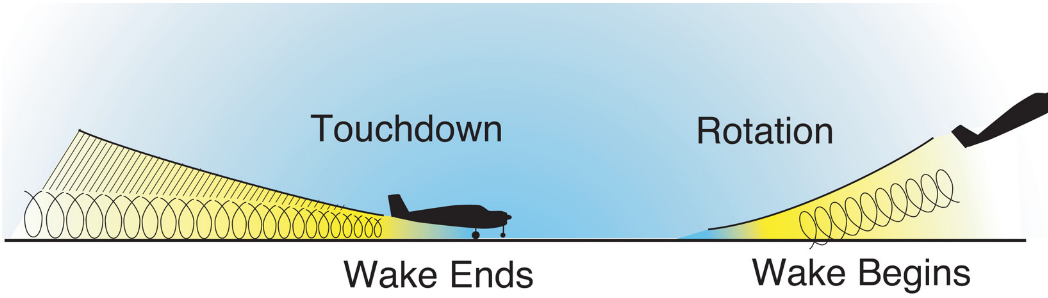


Figure 11 – Vortex behaviour

When the vortices of larger aircraft sink close to the ground (within 100 to 200 feet), they tend to move laterally over the ground at a speed of 2 or 3 knots. A crosswind will decrease the lateral movement of the upwind vortex and increase the movement of the downwind vortex. A tailwind condition can move the vortices of the preceding aircraft forward into the touchdown zone.

7.4 VORTEX AVOIDANCE PROCEDURES

• Landing behind a larger aircraft on the same runway—stay at or above the larger aircraft's approach flight path and land beyond its touchdown point.
• Landing behind a larger aircraft on a parallel runway closer than 2,500 feet—consider the possibility of drift and stay at or above the larger aircraft's final approach flight path and note its touchdown point.
• Landing behind a larger aircraft on crossing runway— cross above the larger aircraft's flight path.
• Landing behind a departing aircraft on the same runway—land prior to the departing aircraft's rotating point.
• Landing behind a larger aircraft on a crossing runway—note the aircraft's rotation point and if past the intersection, continue and land prior to the intersection. If the larger aircraft rotates prior to the intersection, avoid flight below its flight path. Abandon the approach unless a landing is ensured well before reaching the intersection.
• Departing behind a large aircraft, rotate prior to the large aircraft's rotation point and climb above its climb path until turning clear of the wake.
• For intersection takeoffs on the same runway, be alert to adjacent larger aircraft operations, particularly upwind of the runway of intended use. If an intersection takeoff clearance is received, avoid headings that will cross below the larger aircraft's path.
• If departing or landing after a large aircraft executing a low approach, missed approach, or touch and go landing (since vortices settle and move laterally near the ground, the vortex hazard may exist along

the runway and in the flight path, particularly in a quartering tailwind), it is prudent to wait 2 minutes prior to a takeoff or landing.
• En route it is advisable to avoid a path below and behind a large aircraft, and if a large aircraft is observed above on the same track, change the aircraft position laterally and preferably upwind.

8 COLLISION AVOIDANCE

Title 14 of the Code of US Federal Regulations (14 CFR) part 91 has established right-of-way rules, minimum safe altitudes, and VFR cruising altitudes to enhance flight safety. The pilot can contribute to collision avoidance by being alert and scanning for other aircraft. This is particularly important in the vicinity of an airport.
Effective scanning is accomplished with a series of short, regularly spaced eye movements that bring successive areas of the sky into the central visual field. Each movement should not exceed 10°, and each should be observed for at least 1 second to enable detection. Although back and forth eye movements seem preferred by most pilots, each pilot should develop a scanning pattern that is most comfortable and then adhere to it to assure optimum scanning. Even if entitled to the right-of-way, a pilot should give way if it is felt another aircraft is too close.

9 CLEARING PROCEDURES

The following procedures and considerations should assist a pilot in collision avoidance under various situations [AIRP-2].
• Before Takeoff—Prior to taxiing onto a runway or landing area in preparation for takeoff, pilots should scan the approach area for possible landing traffic, executing appropriate manoeuvres to provide a clear view of the approach areas.
• Climbs and Descents—During climbs and descents in flight conditions which permit visual detection of other traffic, pilots should execute gentle banks left and right at a frequency which permits continuous visual scanning of the airspace.
• Straight and Level—During sustained periods of straight-and-level flight, a pilot should execute appropriate clearing procedures at periodic intervals.
• Traffic Patterns—Entries into traffic patterns while descending should be avoided.
• Traffic at VOR Sites—Due to converging traffic, sustained vigilance should be maintained in the vicinity of VORs and intersections.
• Training Operations—Vigilance should be maintained and clearing turns should be made prior to a practice manoeuvre. During instruction, the pilot should be asked to verbalize the clearing procedures (call out “clear left, right, above, and below”).

High-wing and low-wing aircraft have their respective blind spots. High-wing aircraft should momentarily raise their wing in the direction of the intended turn and look for traffic prior to commencing the turn. Low-wing aircraft should momentarily lower the wing.

10 RUNWAY INCURSION AVOIDANCE

It is important to give the same attention to operating on the surface as in other phases of flights. Proper planning can prevent runway incursions and the possibility of a ground collision. A pilot should be aware of the airplane's position on the surface at all times and be aware of other aircraft and vehicle operations on the airport. At times controlled airports can be busy and taxi instructions complex. In this situation it may be advisable to write down taxi instructions. The following are some practices to help prevent a runway incursion [AIRP-2].

- Read back all runway crossing and/or hold instructions.
- Review airport layouts as part of pre-flight planning and before descending to land, and while taxiing as needed.
- Know airport signage.
- Review Notices to Airmen (NOTAM) for information on runway/taxiway closures and construction areas.
- Request progressive taxi instructions from ATC when unsure of the taxi route.
- Check for traffic before crossing any Runway Hold Line and before entering a taxiway.
- Turn on aircraft lights and the rotating beacon or strobe lights while taxing.
- When landing, clear the active runway as soon as possible, then wait for taxi instructions before further movement.
- Study and use proper phraseology in order to understand and respond to ground control instructions.
- Write down complex taxi instructions at unfamiliar airports.

SUMMARY OF CHAPTER 4

CHAPTER 5

WEATHER ADVERSELY AFFECTING AIRPORT OPERATIONS

1 FOG

Instrument pilots and Air Traffic Controllers must learn to anticipate conditions leading to the formation of fog and take appropriate action early in the progress of the flight [ATC-3]. Before a flight, close examination of current and forecast weather should alert the pilot to the possibility of fog formation. When fog is a consideration, pilots should plan adequate fuel reserves and alternate landing sites. En route, the pilot must stay alert for fog formation through weather updates from EFAS, ATIS, and ASOS/AWOS sites.

Figure 1 – Typical Fog Visual Range Reduction at Bologna Airport, Italy

Two conditions will lead to the formation of fog [ATC-2]. Either the air is cooled to saturation, or sufficient moisture is added to the air until saturation occurs. In either case, fog can form when the temperature/dew point spread is 5° or less. Pilots planning to arrive at their destination near dusk with decreasing temperatures should be particularly concerned about the possibility of fog formation.

2 VOLCANIC ASH

Volcanic eruptions create volcanic ash clouds containing an abrasive dust that poses a serious safety threat to flight operations. Adding to the danger is the fact that these ash clouds are not easily discernible from ordinary clouds when encountered at some distance from the volcanic eruption.

When an aircraft enters a volcanic ash cloud, dust particles and smoke may become evident in the cabin, often along with the odour of an electrical fire. Inside the volcanic ash cloud, the aircraft may also experience lightning and St. Elmo's fire on the windscreen. The abrasive nature of the volcanic ash can pit the windscreens, thus reducing or eliminating forward visibility. The pitot-static system may become clogged, causing instrument failure. Severe engine damage is probable in both piston and jet-powered aircraft.

Every effort must be made to avoid volcanic ash [ATC-3]. Since volcanic ash clouds are carried by the wind, pilots should plan their flights to remain upwind of the ash-producing volcano. Visual detection and airborne radar are not considered a reliable means of avoiding volcanic ash clouds. Pilots witnessing volcanic eruptions or encountering volcanic ash should immediately pass this information along in the form of a pilot report. The National Weather Service monitors volcanic eruptions and estimates ash trajectories. This information is passed along to pilots in the form of SIGMETs.

Like many other hazards to flight, the best source of volcanic information comes from PIREPs. Pilots who witness a volcanic eruption or encounter volcanic ash in flight should immediately inform the nearest agency. Volcanic Ash Forecast Transport and Dispersion (VAFTAD) charts are also available; these depict volcanic ash cloud locations in the atmosphere following an eruption, and also forecast dispersion of the ash concentrations over 6- and 12-hour time intervals. See AC 00-45, Aviation Weather Services.

3 THUNDERSTORMS

A thunderstorm packs just about every weather hazard known to aviation into one vicious bundle. Turbulence, hail, rain, snow, lightning, sustained updrafts and downdrafts, and icing conditions are all present in thunderstorms. Do not take off in the face of an approaching thunderstorm or fly an aircraft that is not equipped with thunderstorm detection in clouds or at night in areas of suspected thunderstorm activity [ATC-3].

There is no useful correlation between the external visual appearance of thunderstorms and the severity or amount of turbulence or hail within them. All thunderstorms should be considered hazardous, and thunderstorms with tops above 35,000 feet should be considered extremely hazardous.

Weather radar, airborne or ground based, will normally reflect the areas of moderate to heavy precipitation (radar does not detect turbulence). The frequency and severity of turbulence generally increases with the radar reflectivity closely associated with the areas of highest liquid water content of the storm. A flight path through an area of strong or very strong radar echoes separated by 20 to 30 miles or less may not be considered free of severe turbulence.

Figure 2 – A thunderstorm packs just about every weather hazard known to aviation into one vicious bundle

The probability of lightning strikes occurring to aircraft is greatest when operating at altitudes where temperatures are between -5 °C and +5 °C. In addition, an aircraft flying in the clear air near a thunderstorm is also susceptible to lightning strikes. Thunderstorm avoidance is always the best policy.

4 WIND SHEAR

Wind shear can be defined as a change in wind speed and/or wind direction in a short distance. It can exist in a horizontal or vertical direction and occasionally in both [ATC-3]. Wind shear can occur at all levels of the atmosphere but is of greatest concern during takeoffs and landings. It is typically associated with thunderstorms and low-level temperature inversions; however, the jet stream and weather fronts are also sources of wind shear.

While an aircraft is on an instrument approach, a shear from a tailwind to a headwind will cause the airspeed to increase and the nose to pitch up with a corresponding balloon above the glidepath. A shear from a headwind to a tailwind will have the opposite effect and the aircraft will sink below the glidepath.

A headwind shear followed by a tailwind/downdraft shear is particularly dangerous because the pilot has reduced power and lowered the nose in response to the headwind shear. This leaves the aircraft in a nose-low, power-low configuration when the tailwind shear occurs, which makes recovery more difficult, particularly near the ground. This type of wind shear scenario is likely while making an approach in the face of an oncoming thunderstorm. Pilots should be alert for indications of wind shear early in the approach phase and be ready to initiate a missed approach at the first indication. It may be impossible to recover from a wind shear encounter at low altitude.

To inform pilots of hazardous wind shear activity, some airports have installed a Low-Level Wind Shear Alert System (LLWAS) consisting of a centerfield wind indicator and several surrounding boundary-wind indicators. With this system, controllers are alerted of wind discrepancies (an indicator of wind shear possibility) and provide this information to pilots.

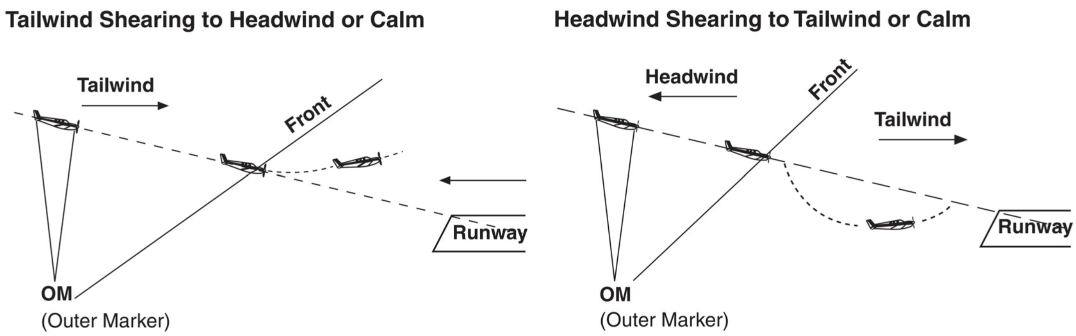


Figure 3 – Glide-Slope Deviations due to wind shear encounter

5 EFFECT OF OBSTRUCTIONS ON WIND

Another atmospheric hazard exists that can create problems for pilots. Obstructions on the ground affect the flow of wind and can be an unseen danger [ATC-3]. Ground topography and large buildings can break up the flow of the wind and create wind gusts that change rapidly in direction and speed. These obstructions range from manmade structures like hangars to large natural obstructions, such as mountains, bluffs, or canyons. It is especially important to be vigilant when flying in or out of airports that have large buildings or natural obstructions located near the runway.

The intensity of the turbulence associated with ground obstructions depends on the size of the obstacle and the primary velocity of the wind. This can affect the takeoff and landing performance of any aircraft and can present a very serious hazard. During the landing phase of flight, an aircraft may "drop in" due to the turbulent air and be too low to clear obstacles during the approach.

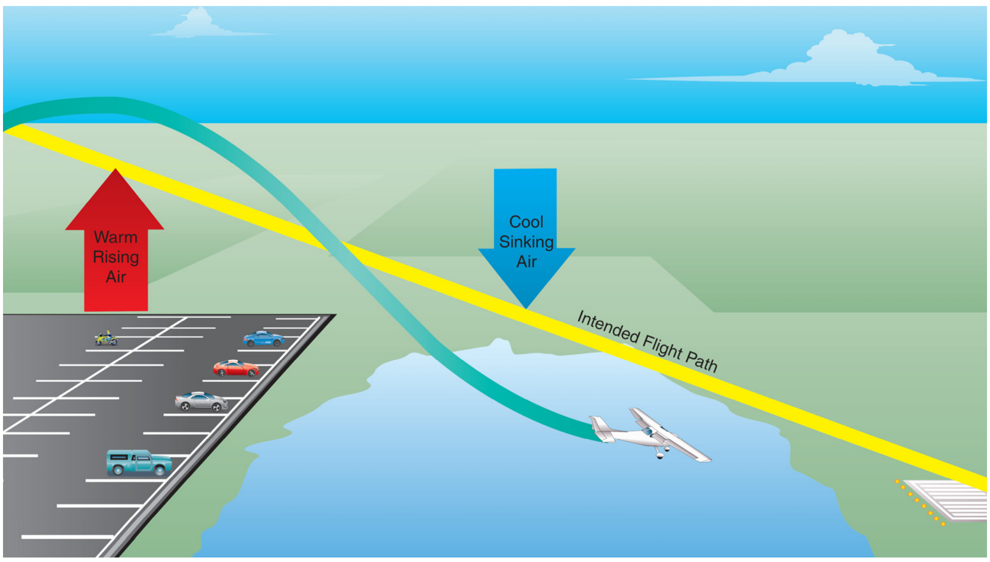


Figure 4 – Currents generated by varying surface conditions [ATC-3]

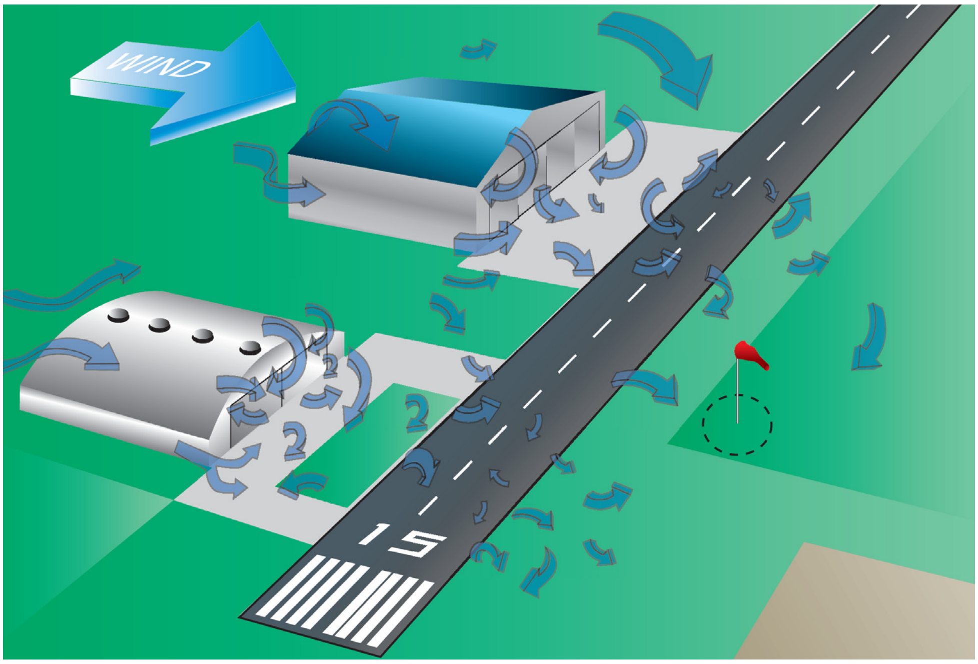


Figure 5 – Turbulence caused by manmade obstructions [ATC-3]

6 LOW-LEVEL WIND SHEAR

While wind shear can occur at any altitude, low-level wind shear [ATC-3] is especially hazardous due to the proximity of an aircraft to the ground. Directional wind changes of 180° and speed changes of 50 knots or more are associated with low-level wind shear. Low-level wind shear is commonly associated with passing frontal systems, thunderstorms, and temperature inversions with strong upper level winds (greater than 25 knots).

Wind shear is dangerous to an aircraft for several reasons. The rapid changes in wind direction and velocity changes the wind's relation to the aircraft disrupting the normal flight attitude and performance of the aircraft. During a wind shear situation, the effects can be subtle or very dramatic depending on wind speed and direction of change. For example, a tailwind that quickly changes to a headwind will cause an increase in airspeed and performance. Conversely, when a headwind changes to a tailwind, the airspeed will rapidly decrease and there will be a corresponding decrease in performance. In either case, a pilot must be prepared to react immediately to the changes to maintain control of the aircraft.

In general, the most severe type of low-level wind shear is associated with convective precipitation or rain from thunderstorms. One critical type of shear associated with convective precipitation is known as a microburst. A typical microburst occurs in a space of less than 1 mile horizontally and within 1,000 feet vertically. The lifespan of a microburst is about 15 minutes during which it can produce downdrafts of up to 6,000 feet per minute. It can also produce a hazardous wind direction change of 45 knots or more, in a matter of seconds. When encountered close to the ground, these excessive downdrafts and rapid changes in wind direction can produce a situation in which it is difficult to control the aircraft (see Figure 6). During an inadvertent takeoff into a microburst, the plane first experiences a performance-increasing headwind (1), followed by performance-decreasing downdrafts (2). Then the wind rapidly shears to a tailwind (3), and can result in terrain impact or flight dangerously close to the ground (4).

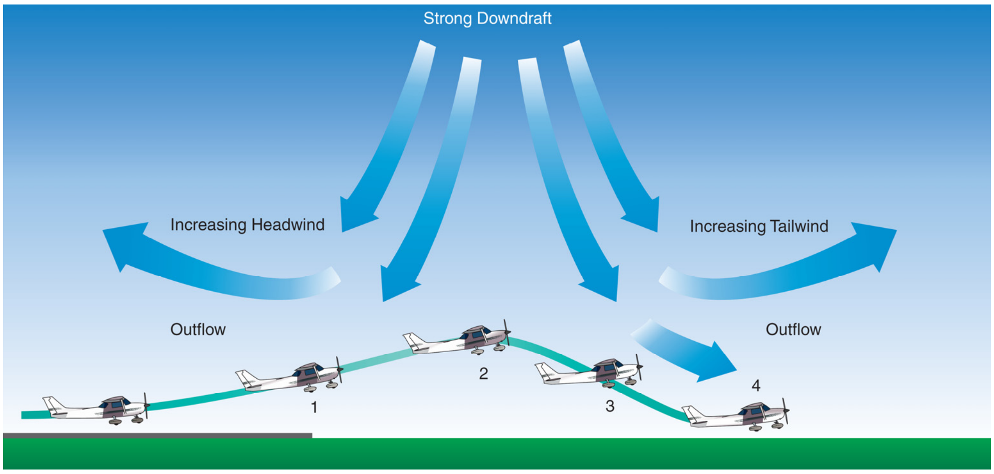


Figure 6 – Effect of a microburst wind [ATC-3]

Microbursts are often difficult to detect because they occur in a relatively confined area. In an effort to warn pilots of low-level wind shear, alert systems have been installed at several airports around the country. A series of anemometers, placed around the airport, form a net to detect changes in wind speeds. When wind speeds differ by more than 15 knots, a warning for wind shear is given to pilots. This system is known as the low-level wind shear alert system, or LLWAS.

It is important to remember that wind shear can affect any flight and any pilot at any altitude. While wind shear may be reported, it often remains undetected and is a silent danger to aviation. Always be alert to the possibility of wind shear, especially when flying in and around thunderstorms and frontal systems.

7 VISIBILITY

Closely related to cloud cover and reported ceilings is visibility information. Visibility [ATC-3] refers to the greatest horizontal distance at which prominent objects can be viewed with the naked eye. Current visibility is also reported in METAR and other aviation weather reports, as well as automated weather stations. Visibility information, as predicted by meteorologists, is available during a pre flight weather briefing.

8 PRECIPITATION

Precipitation refers to any form of water particles that form in the atmosphere and fall to the ground. It has a profound impact on flight safety. Depending on the form of precipitation, it can reduce visibility, create icing situations, and affect landing and takeoff performance of an aircraft.

Precipitation occurs because water or ice particles in clouds grow in size until the atmosphere can no longer support them. It can occur in several forms as it falls toward the Earth, including drizzle, rain, ice pellets, hail, and ice.

Drizzle is classified [ATC-3] as very small water droplets, smaller than 0.02 inches in diameter. Drizzle usually accompanies fog or low stratus clouds. Water droplets of larger size are referred to as rain. Rain that falls through the atmosphere but evaporates prior to striking the ground is known as virga. Freezing rain and freezing drizzle occur when the temperature of the surface is below freezing; the rain freezes on contact with the cooler surface.

If rain falls through a temperature inversion, it may freeze as it passes through the underlying cold air and fall to the ground in the form of ice pellets. Ice pellets are an indication of a temperature inversion and that freezing rain exists at a higher altitude. In the case of hail, freezing water droplets are carried up and down by drafts inside clouds, growing larger in size as they come in contact with more moisture. Once the updrafts can no longer hold the freezing water, it falls to the Earth in the form of hail. Hail can be pea-sized, or it can grow as large as 5 inches in diameter, larger than a softball.

Snow is precipitation in the form of ice crystals that falls at a steady rate or in snow showers that begin, change in intensity, and end rapidly. Falling snow also varies in size, being very small grains or large flakes. Snow grains are the equivalent of drizzle in size.

Precipitation in any form poses a threat to safety of flight. Often, precipitation is accompanied by low ceilings and reduced visibility. Aircraft that have ice, snow, or frost on their surfaces must be carefully cleaned prior to beginning a flight because of the possible airflow disruption and loss of lift. Rain can contribute to water in the fuel tanks. Precipitation can create hazards on the runway surface itself, making takeoffs and landings difficult, if not impossible, due to snow, ice, or pooling water and very slick surfaces.

9 AVIATION WEATHER FORECASTS

Observed weather condition reports are often [ATC-3] used in the creation of forecasts for the same area. A variety of different forecast products are produced and designed to be used in the pre-flight planning stage. The printed forecasts that pilots need to be familiar with are the terminal aerodrome forecast (TAF), aviation area forecast (FA), in-flight weather advisories (SIGMET, AIRMET), and the winds and temperatures aloft forecast (FD).

9.1 TERMINAL AERODROME FORECASTS (TAF)

A terminal aerodrome forecast is a report established for the 5 statute mile radius around an airport. TAF reports are usually given for larger airports. Each TAF is valid for a 24-hour time period, and is updated four times a day at 0000Z, 0600Z, 1200Z, and 1800Z. The TAF utilizes the same descriptors and abbreviations as used in the METAR report.

The terminal forecast includes the following information in sequential order:
1. Type of Report—A TAF can be either a routine forecast (TAF) or an amended forecast (TAF AMD).
2. ICAO Station Identifier—The station identifier is the same as that used in a METAR.
3. Date and Time of Origin—Time and date of TAF origination is given in the six-number code with the first two being the date, the last four being the time. Time is always given in UTC as denoted by the Z following the number group.
4. Valid Period Date and Time—The valid forecast time period is given by a six-digit number group. The first two numbers indicate the date, followed by the two-digit beginning time for the valid period, and the last two digits are the ending time.
5. Forecast Wind—The wind direction and speed forecast are given in a five-digit number group. The first three indicate the direction of the wind in reference to true north. The last two digits state the wind-speed in knots as denoted by the letters "KT." Like the METAR, winds greater than 99 knots are given in three digits.
6. Forecast Visibility—The forecast visibility is given in statute miles and may be in whole numbers or fractions. If the forecast is greater than 6 miles, it will be coded as "P6SM."

7. Forecast Significant Weather—Weather phenomenon is coded in the TAF reports in the same format as the METAR. If no significant weather is expected during the forecast time period, the denotation "NSW" will be included in the "becoming" or "temporary" weather groups.
8. Forecast Sky Condition—Forecast sky conditions are given in the same manner as the METAR. Only cumulonimbus (CB) clouds are forecast in this portion of the TAF report as opposed to CBs and towering cumulus in the METAR.
9. Forecast Change Group—For any significant weather change forecast to occur during the TAF time period, the expected conditions and time period are included in this group. This information may be shown as From (FM), Becoming (BECMG), and Temporary (TEMPO). "From" is used when a rapid and significant change, usually within an hour, is expected. "Becoming" is used when a gradual change in the weather is expected over a period of no more than 2 hours. "Temporary" is used for temporary fluctuations of weather, expected to last for less than an hour.
10. Probability Forecast—The probability forecast is given percentage that describes the probability of thunderstorms and precipitation occurring in the coming hours. This forecast is not used for the first 6 hours of the 24-hour forecast.

9.2 PRACTICAL EXAMPLES

The following examples describe surface problem situations in which weather in the terminal area adversely affects the capacity of an airport [ATC-2]. In each example, we describe not only the weather event, but also the impact that it has on conditions in the NAS.
Each of the problem situations described below can have purely local (isolated) effects or both local and global (propagating) effects.

9.2.1 SURFACE SITUATION 1: SHIFTING WIND DIRECTION CHANGES THE RUNWAY CONFIGURATION

A sudden change in runway configuration due to unpredicted or poorly predicted changes in wind direction (and/or the presence of wind shear or microburst warnings) results in high workload and added delays while aircraft are re-routed to new arrival fixes and runways. This may also result in a runway configuration selection that is non-optimal for the current conditions and traffic complexity, in which case, additional arrival and departure delays may occur.

Such a situation can occur when the ATC tower tries to predict likely future conditions, picking a runway configuration that is less sensitive to wind shifts – potentially at the expense of throughput in the event that the weather event doesn't actually materialize.
Generally, the impacts of such events on NAS operations include:
- Wind direction is such that it requires a configuration change
- If timing of wind shift is not predicted properly, then aircraft currently in queue for a now inactive runway need to be taxied to an active runway
- Aircraft still at gates will need to be assigned to different runways
- Aircraft on final approach may need to execute a missed approach procedure
- Aircraft outside of arrival metering fixes must be rerouted to new fixes
- Departures using the new configuration must wait until terminal airspace has "stabilized" (e.g., remaining arrivals clear of departure corridors).

The impact on capacity varies depending on the condition. In general, a runway configuration change adds delay to both arrivals and departures, reducing capacity during the transition. However, the steady state effect of a change in configuration may be to increase, decrease, or leave unchanged the achievable arrival and departure rates at a given airport.

9.2.2 SURFACE SITUATION 2: LOW VISIBILITY

Reductions in visibility and Runway Visual Range (RVR) due to fog, haze, snow, etc. negatively impact surface operations in the NAS in many ways, including:

- Ground and Local Controllers unable to discern exact positions of aircraft (see fog in Figure 1)
- Controllers cannot discern the order of flights in a queue at the runway or spot location
- Increased separations are required between aircraft to maintain safety as relative distances are hard to monitor
- Runway crossing is more difficult as it relies on controller visual judgment of gaps
- Pilots ability to see airport signage and pavement markings is significantly reduced, leading to a reduction in awareness of actual surface position
- Pilots have more difficulty seeing other aircraft or knowing their exact position in a queue
- Controllers must often rely on pilot-reported positions, which may be in error
- Inability to conduct closely spaced (or parallel) approaches.

The impacts on capacity are as follows. Pilot uncertainties regarding position lead to a reduced taxi speed; simultaneously, controller uncertainty regarding aircraft position results in increased runway crossing times and thus taxi delays. Surface movement inefficiency "backs up" to the runways, reducing both AAR and ADR (Airport Departure Rate). Elimination of closely-spaced parallel approaches, common during low visibility, further lowers the AAR.

9.2.3 SURFACE SITUATION 3: AIRCRAFT REQUIRING DE-ICING

Figure 7 – De-icing operations increase the time required between flights and reduce airport capacity

De-icing (e.g. Figure 7) is a cumbersome procedure requiring time, equipment, de-icing fluid, and personnel that impacts capacity in terms of departure rates. Ice accumulation on aircraft wings and control surfaces must be removed. The extent of the capacity reduction depends, in a large part, on the availability of equipment and personnel and the location of the de-icing pads relative to the gates and runways (some airports de-ice at the gates, while others use a remote de-icing pad). Once aircraft are de-iced, they must takeoff within a given time period (typically 15 minutes) or be re-treated. The impact on capacity is primarily related to the time consumed by the de-icing process, namely any additional taxi time to/from remote de-icing pads plus the service time required to apply the de-icing treatment. This impact manifests itself as departure delays – the extent of which is determined by the efficiency of the process. Note that since passengers on commercial flights expect flights to be boarded as scheduled, there is little opportunity for airlines to board aircraft early to "make up" for the time needed for de-icing (assuming it could be predicted).

An indirect capacity impact is related to the need to expedite the departures for aircraft which have been de-iced to avoid the need for re-treatment – which can impact the required inter-arrival spacing on mixed use runways.

9.2.4 Surface Situation 4: Snow, Ice, Slush, Water on Runway

Slick conditions on the airport surfaces can reduce aircraft braking and directional control. Complicating the situation is the fact that braking conditions are not necessarily the same on all parts of a runway, due to its length. Impacts of this problem on NAS surface operations include:

- Increased runway occupancy time as aircraft must rollout to the last runway exit
- Ceasing of Land and Hold Short Operations (LAHSO)
- Some shorter runways may not be usable
- Temporary runway closure due to the need for removal of accumulated snow, and
- Impaired visibility of surface pavement markings and lighting from the flight deck.

Figure 8 – Snow on the airport surface

The impacts on capacity are as follows. A reduction in options of runway exits impacts taxi routing flexibility, potentially leading to arrival taxi delays (or at least increased taxi times) and possible surface congestion.

With an increase in runway occupancy time, there will be a corresponding AAR reduction due to the need for increased inter-arrival spacing. This reduction is exacerbated by the closure of runways for snow removal and when certain runways are unusable due to poor braking action. The inability to utilize LAHSO procedures further impacts capacity by limiting the ability of controllers to coordinate operations between dependent runways.

SUMMARY OF CHAPTER 5

CHAPTER 6
A-SMGCS

Over the past decade or so there has been increasing interest and recognition of the ground movement environment as part of the overall air traffic control problem. There are a number of reasons for this, both on safely and airport efficiency grounds. One recognition of this was the adoption by EUROCONTROL of a gate to gate air traffic control concept. There is also considerable activity by organisations such as, FAA, ICAO, the EU and EUROCAE. As an example of this, the FAA is currently procuring a large number of ground surveillance systems for use at US airports. Equally many European airports and aviation authorities are or have recently purchased surface movement systems.

Nowadays, because of the ever-growing number of aircrafts in the areas to be controlled; the raising demand for safety of passengers and cargoes; and the greater care about the environment, the problem of improving the performances of traditional Airport Control systems must be faced.
Research trend in future years is oriented toward the development of integrated systems, which collect information coming from various sensors and distribute it both to ground air-traffic controllers and to moving vehicles. The realisation of a system for the automatic control of airport surface traffic can play an important role in this context. The systems and procedures that ensure the above-mentioned operations can be included in the " Advanced-Surface Movement Guidance and Control System" (A-SMGCS).

The Air Traffic Management (ATM) concept envisages an air traffic service that works at optimum efficiency, capacity, and safety. This could be achieved through the integration and harmonisation of all the systems, services, and procedures required for the air traffic services.
The objective of the A-SMGCS is to optimise the efficiency, capacity, and safety of operations at an aerodrome.

Traditionally, radar has been the main sensor for surveillance of the airport surface. While this is still the case, the overall concept of ground movement control is considerably more complex than that of the radar alone, and will involve multiple types of sensors data fused to provide data output.

Following the EUROCONTROL Performance Review Report of 2005 airport delays are a growing proportion of the total ATM delays. Nearly all European hubs and already some mid-size airports are on the list of the 15 most penalising airports in Europe, which together generate 77% of all airport ATFM delays.

Extending existing airport infrastructure, e.g. by building new runways, is a very difficult and complex process associated with many restrictions. Therefore, the optimal usage of existing infrastructure more and more becomes a necessity. Despite the importance of optimal resource usage, flight deck operations on the ground are still not very sophisticated nowadays. Implementation of modern cockpit technology for surface operations lags behind the developments for other flight phases. "Seen and be seen" is still the most common practice on ground. After landing pilots have to navigate using paper maps and look out of the window to avoid other traffic. Above that ATCOs are performing the surveillance task mainly visually. Frequently, ATCOs are supported by surface movement radar (SMR) only giving them poor analogue radar plots with a lot of clutter and nuisance targets. As soon as the visual reference is impaired all surface operations are severely impacted by an increasing workload and a decreasing situation awareness of all participants, compromising safety and airport capacity and increasing delays. This leads to negative consequences for the approach areas and finally to unfavourable network effects in the overall air transport system.

An A-SMGCS helps to overcome this poor situation. In its basic level 1 it provides the ATCOs with a display showing the complete traffic situation that includes the position of all aircraft and vehicle movements and their identification [AIRP-11]. Since it is assumed that each day in the US and Europe at least one runway incursion is occurring, which may lead to severe accidents, such as the Milan-Linate accident in 2001, in its level 2 the A-SMGCS provides the ATCO with an Automatic Runway Incursion Monitoring And Alerting Function [AIRP-11].

Figure 1 – Milan-Linate accident in 2001

1 A-SMGCS PURPOSE

A-SMGCS is a modular system defined in the ICAO Manual on Advanced Surface Movement Guidance and Control Systems (A-SMGCS). Such systems aim to "maintain the declared surface movement rate under all weather conditions within the aerodrome visibility operational level (AVOL) while maintaining the required level of safety".

With the complete concept of an A-SMGCS, air traffic controllers (ATCO), flight crews, and vehicles drivers are assisted with surface operations in terms of surveillance, control, routing/planning and guidance tasks.

The EUROCAE MASPS document gives a useful overall system description. The overall system is subdivided into a number of elements. These are:

1. *Surveillance Element.* The task of this element is to identify (or if not possible classify) and position all targets of operational significance at the time of each update. The system update time is at least once a second. This element provides a fundamental input to the rest of the system
2. *Monitoring and alerting.* Having provided a suitable surveillance system , the ASMGCS must use this information to monitor the situation on the aerodrome surface and provide alerts when particular situations are encountered. Such an alert could be if two aircraft are on the same runway at the same

time going in different directions, or if one aircraft as entered onto the runway before the runway is clear etc.

3. *Guidance.* It is proposed, at least in more complex airports, that guidance will be provided to aircraft automatically or semi-automatically. The use of airport lighting systems (taxiway lights and stop bars) is one such guidance system- cockpit data displays are a longer term possibility

4. *Route planning.* Automatic or semi-automatic allocation of routes by the system to deal with both strategic (or long term) and tactical (short term)route planning , for example in emergency situations.

The standard also notes that different airport needs will be met by different levels of implementation of the system- from just surveillance right through to full A-SMGCS operation.

1.1 SURVEILLANCE FUNCTION

The surveillance element as defined above requires as a minimum detection and identification of targets of operational interest [AIRP-6]. Operational interest will vary according to size of airport, number of movements, level of implementation, etc. In this context targets will certainly include large and small aircraft both at high and low speeds but will also include ground vehicles(baggage handlers, service and emergency vehicles) and could include people, animals such as dogs and even single pieces of luggage. The surveillance element must provide situation awareness of all these. No one sensor will be capable of providing this information with sufficiently high probability of detection and sufficiently low false alarm rate to be acceptable in a safety critical system. So it is anticipated the surveillance element will comprise a number of sensors and a data fusion processor. It is further commented here that clearly a very high probability of detection is required in a complex airport environment, however a low rate of false detection is as important otherwise false alerts may be generated by the Monitoring and Alerting function which will yield the system not usable in a busy ground movement control situation. Crucially, to be used for decision making by downstream system element such as monitoring and alerting the information must be timely- that is the surveillance element must provide at it's output the situation on the airport surface at the time it is reported, not necessarily the time when the information was gathered from the sensors. This means the data fusion must include a high integrity tracking process to overcome different sensor interrogation rates, system latency issues etc.

The sensors are categorised in two broad groups- co-operative and non-cooperative:

Co-operative sensors require a transponder or similar on the target. The most common type is some form of multilateration technique using existing SSR mode S or ADS-B transponders . There are also a number of proprietary ground vehicle tracking schemes. These systems have the advantage of being able to provide the identity of the target in addition to position. They have the fundamental disadvantage of requiring a operational transponder on the target. Not all targets have such devices, or can have them (consider the dog!) - also the system must be able to handle the situation of a defective transponder.

Non co-operative sensors have no need for equipment on the target, but while some classification may be possible (large, medium or small for example) non cooperative sensors cannot in general provide reliable information regarding the identity of the target. While there are a number of non cooperative sensors proposed (for example magnetic loop sensors) by far the most common and widely used is radar.

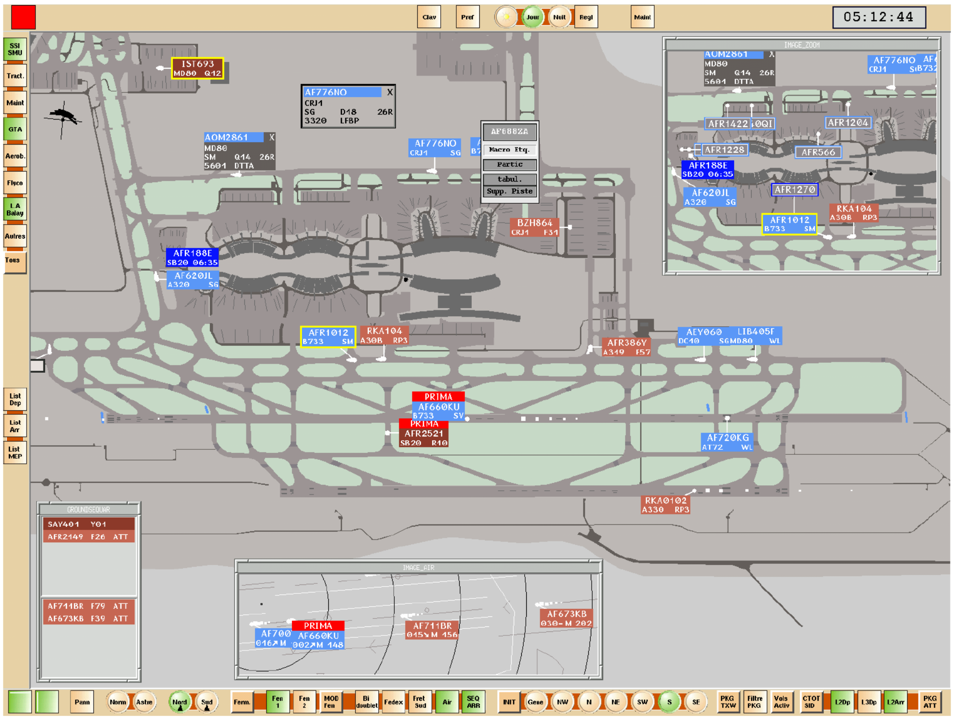


Figure 2 – Ground ATCO A-SMGCS HMI at Paris Charles De Gaulle

1.1.1 THE RADAR SENSOR

Before discussing the details of radar sensors used in the airport surface movement application it is worthwhile to highlight some of the difficulties faced by the radar design:

1. Wide range of RCS of targets, with small targets close to large ones
2. Extended and point source targets
3. High and low velocity targets (including stationary targets)
4. High resolution (range and azimuth)
5. High clutter both ground and often rain
6. Complex multipath and shadowing scenarios
7. High system update rate
8. High system integrity
9. Cost constraints

When working as a A-SMGCS sensor, as opposed to a stand alone radar, the system must use target extracted output- the option of using raw video is clearly not possible- and the challenge of generating a even better (signal) to (clutter plus noise) ratio is, if anything, more demanding [AIRP-9].

Within the overall market interest for surface movement systems high detection capability and reliability coupled with low cost are essential. History has proven that high cost systems can only be implemented at a few airports- what is needed is a range of solutions to meet different operational needs.

One major factor in the design of the radar system is the assessment of it’s detection performance. The major difficulty here is the impact of rain clutter on the system.

Work undertaken during the development of such radars shows why the situation is much more complex than the standard air surveillance scenario. When the target is at some height, the radar performance in rain is dominated by backscatter from the rain itself, as is well known, and is commonly reduced by use of circular polarisation in the radar. However in the ground movement scenario the

presence of the ground greatly complicates the situation. In essence the rain clutter contribution occurs not only from direct backscatter but also from reflections from the ground.
There are four main clutter paths - direct reflection, then a path reflecting from the clutter to the ground and back to the radar, also the path from ground to clutter and back to the radar and finally a path from the ground to the rain clutter and back via the ground. Now, these extra paths greatly affect the amount of rain clutter entering the system, particularly at short ranges.
This type of model is very important as the system is clutter limited- hence only by adoption of a more complex model will allow the designer to correctly optimise the system. It is not the purpose of this thesis to discuss competing radar predictions in detail- but an incorrect model may yield a system highly susceptible to high rainfall clutter- the sort of conditions when the radar is really needed.

Traditionally, Ku-band (approximately 16GHz) has been used for high performance ground movement use, principally as this offers a good angular resolution from a practical sized antenna . This frequency band was adopted in the 1970s and 1980s .for example. by the ASDE-3 system in the USA and many authorities worldwide. Millimetric radar has been used experimentally since 1960s, and continues to be offered. It provides high resolution from small antennas, but suffers from limitations at high rainfall rates. Therefore networks of such radars , possibly used in conjunction with lower frequency systems have been proposed t cover the airport surface

The largest growth however has been at X-band (around 9 GHz). Here the development of lower cost, high performance components, the inherent lower sensitivity to rain clutter coupled with recent developments in both transceiver and antenna technology have yielded high performance lower cost systems. Recently many authorities has adopted X-band radars including, the USA, Canada. and many European countries. Arguably, X-band is now the frequency of choice for most new systems. Turning to transmitter technology, most installed systems (both Ku and X-band) use magnetron based sources. However there has been considerable development of X-band transceiver technology. Newer designs offer sophisticated control of the transceiver and features such as frequency diversity (normally using two magnetrons of different frequencies) needed for clutter reduction. Commercial receiver front end noise figures of 4dB or better are considered normal. The older (but still sophisticated) ASDE-3 system utilises a TiVT approach with multiple frequencies available. The number of frequencies used (up to 14) is needed to minimise rain clutter (which is inherently higher at Ku than X-band). This approach has lead to a excellent RF performance but is high cost and has a high cost of ownership.
Most recently, the ASDE-X system being provided to the FAA in its most recent contract uses a innovative all solid state design of transceiver.

Other solutions being offered to the market place include electronically scanned phased array- high cost, and only limited scan angle per face, but with the obvious benefits of being non rotating.
Research into sensors is also yielding other possible solutions such as the Near Radar Network - a type of multilateration using fixed, non-rotating, X-band sensors – currently being developed in Germany.

1.2 MONITORING FUNCTION

Once the route has been assigned to an aircraft the Monitoring/Alerting function must check that the aircraft is maintaining the correct route, without deviation (i.e., the aircraft should reach each waypoint within the allotted time slot, and take the correct leg when leaving the waypoint). This is called "Conformance Monitoring." validation service on the proposed route. They range from very simple to more complicated methods.
There are several ways to implement the conflict validation service on the proposed route. They range from very simple to more complicated methods. There are at least three possible methods to implement the conflict validation service on the proposed route. They all validate the Route sampling each Route in the Waypoints.
There are several ways to implement the conflict

1.2.1 TIME AT WAYPOINT

The Time at Waypoint method takes into account the time each target should be at given Waypoints. The method takes a Route and for each Waypoint checks if there are any other Assigned Routes where a target should transit by the same Waypoint at nearly the same time. It must be defined a Time Separation.

1.2.2 FREE LEG

The Free Leg method, shown on previous page, takes into account the Layout of the Airport and the position of each Target. No other target should be in any Leg connected to the Waypoint. The method takes a Route and for each Waypoint checks if there are any other Assigned Routes where a target should transit by any Leg connected to the Waypoint at the same time.

1.2.3 DISTANCE SEPARATION

The Distance Separation method takes into account the position of each target. No other target should be at a distance less then a given spatial separation. The method takes a route and for each Waypoint checks if there are any other Assigned Route where a target at a distance less then a given Spatial Separation at the same time. An improvement for any of the above mentioned methods could be to take into consideration the relative motion at that time (intersecting Routes, parallel Routes, same Route, same direction, opposite direction, and so on...).

1.3 GUIDANCE FUNCTION

The main objective of the Guidance function [AIRP-8] is the provision of clear indication of assigned routes to pilots in the movement area such that they can follow the assigned route. Guidance aids indicate where on the runway, taxiway, or apron the aircraft or vehicle can be manoeuvred safely. The A-SMGCS have the ability to automatically control taxiway lights, stop bars, and other guidance aids to prevent conflicts and provide optimum routings
Guidance results from a combination of the following actions:

- a controlling authority (e.g., ATC) issues a clearance that gives a pilot permission to move an aircraft along the assigned route;
- the A-SMGCS provides the processing and signalling functionality to provide the necessary guidance associated with the clearance;
- the aircraft moves in accordance with a pilot's interpretation of the control and guidance given.

The Guidance function of the A-SMGCS:

- provides necessary guidance for any authorized movement and available for all possible route selections;
- provides clear indication to pilots to allow them to follow their assigned route;
- is capable of accepting a change of route at any time; and

- allows monitoring of the operational status of all guidance aids.

It should be noted that pilots will need to retain the right to refuse clearances and to stop the aircraft if, in their judgement, a continuation of the movement would cause a hazard to the aircraft. The Guidance function is fed with target position information from the Surveillance function. On the basis of this, the target position with respect to the Airport Layout is known. Guidance requires that the taxiway centreline lighting for a given route to be followed is lit up in front of a moving target and extinguished when the target has passed.
The overall structure of the Guidance Function, considers a main loop which receives the Target's information from the Surveillance Function. The Guidance Function main loop should know all the Routes already assigned to Targets by the Routing Function and collected in the Assigned Routes Table.
Finally the Airport Layout, especially about the Runways, is supposed to be coded in several databases:

- Light Segment Data Base Collects information on Taxiway centreline segments and Stopbar Segments; mainly position and status. The Database of the Light System is made up by segments, one segment for each straight leg, curve, and stopbar.
- Runway Data Base Collects information about the Runway structure; mainly position, list of exits and a list of stopbars.

1.4 ROUTING FUNCTION

The main objective of the Routing function [AIRP-8] is to generate routes in order to make full use of aerodrome capacity. This requires the provision of a tactical planning tool which reacts to the dynamics of the traffic situation providing variations to former assigned routes to maintain the required flow despite unplanned occurrences.
The Routing function must have information about the current traffic situation on the aerodrome, as well as aerodrome layout. from one location to another. Routes are composed of a number of waypoints selected from a predefined routing grid (Airport Layout).
Routes clearly define how the aircraft is to proceed
The Routing function of an A-SMGCS:

- enable a route to be designated for each aircraft within the movement area;
- allow for a change of destination at any time; and
- allow for a change of route to the same destination.

The Routing function provides the control authority with advisory information on designated routes. The assignment of routes is carried out by the control authority. When designating routes, the A-SMGCS:

- minimise taxiing times in accordance with the most efficient operational configurations;
- be interactive with the Monitoring & Control function to minimise junction conflicts;
- be responsive to operational changes (e.g., runway changes, taxiways closed for maintenance);
- provide means of validating routes.

The route depends solely on the ground movement situation that exists at the time when the route is issued. It is critical to the efficient and flexible operation of any aerodrome that planning elements be tactically adjusted to meet changing circumstances.

1.4.1 AIRPORT LAYOUT

The aerodrome layout is the most mandatory external system. It consists of a digitised airport map, representing runways, taxiways, aprons, manoeuvring areas, roads, buildings, restricted areas. It shall also contain sensor position, restricted areas, parking positions and shelters. It must define precisely an airport reference point (generally the control tower or the centre of airport bounds). All other co-ordinates are expressed in reference to this point. The "interface with external databases" partly responsible of access to the aerodrome layout has the responsibility of converting co-ordinates in the airport referential to WGS-84.

1.4.2 WAYPOINT

Each Waypoint is typically a point at the crossing of two Taxiways or one Taxiway and one Runway. It follows that each point has a position on the Airport Map which could be expressed either in metres with reference to a reference point on the Airport Map or with absolute geographical co-ordinate. In this latter case, the WGS84 system must be used. If the relative system is used, then it should be provided a means to convert the position (x, y, z) of each Waypoint into the WGS84 coordinate system. For each route the time at which the aircraft is expected at each Waypoint is estimated by adding the time at which the aircraft is expected at the previous Waypoint and the time needed to run the Leg that links the two Waypoints. The times needed to run the segment is simply: time = length/speed; and the length is the Euclidean distance between P1 and P2. For the speed is used the maximum speed allowed in the Leg.

1.4.3 CONFLICT VALIDATION

When planning a 3D (x, y, t) route it is important to ensure that the route is conflict-free at the time the route plan is generated. This is the task of the Route Planning function. Clearly, the best way to ensure conflict-free routes is to avoid any intersections, but this may not always be possible. Normally, there will be no conflicts if aircraft follow behind each other with a reasonable separation.
Except in low visibility conditions, pilots can see the aircraft in front and maintain separation. Furthermore, conflicts between outbound and inbound aircraft are normally avoided by assigning non-intersecting routes.

2 EUROPEAN AIRPORT MOVEMENT MANAGEMENT BY A-SMGCS (EMMA)

To harmonise the implementation of the first two levels of A-SMGCS, which focus on surveillance and conflict monitoring, and to further mature the necessary technology and operating procedures, the European Commission funded the project EMMA [AIRP-10] (European airport Movement Management by A-SMGCS) within the sixth framework programme. Within EMMA, A-SMGCS level 1&2 systems were installed at three European mid-size airports: Milan-Malpensa, Prague-Ruzyně, and Toulouse-Blagnac [AIRP-7]. Technical and operational trials were conducted at all three sites to verify the technical performance against the requirements and to prove operational feasibility. Additionally, real-time simulations were performed in order to tune parameters of the monitoring and alerting function and to also assess operational improvements under experimental conditions.

Here are the most important figures of merit of EMMA:

- Probability of Identification PID ≥ 99.9% for identifiable targets
- Probability of False Identification PFID < 10E-3 per Reported Target
- Target Report Update Rate TRUR ≤ 1 s
- Probability of Detection of an Alert Situation PDAS ≥ 99.9%
- Probability of False Alert PFA < 10E-3 per Alert
- Alert Response Time ART ≤ 0.5 s
- Routing Process Time RPT < 10 s

High-level Objective	Low-level Objective	Indicator
Safety	Reduced number of incidents and accidents	Number of incidents and accidents
	Faster identification and mitigation of safety hazards	Time for conflict detection, identification and resolution
Efficiency/Capacity	Lower Taxi Time for in and outbound traffic	Taxi Time
	Lower duration of radio communications	Duration of radio communications (R/T load)
Human Factors	Higher Situation Awareness	Situational Awareness
	Convenient level of workload	Workload

Table 1 – EMMA's objectives and relative indicators

Parameter	Definition	Source
Alert Response Time (ART)	The time delay between an alert situation occurring at the input to the Alert Situation Detection Element and the corresponding alert report being generated at its output.	*[EUROCAE MASPS]*
Display Resolution (DR)	The number of individually addressed picture elements (pixels) along each axis of the display screen. (For a raster-scan display, the resolution is normally expressed in terms of the number of raster lines and the number of pixels per line.)	*[EUROCAE MASPS]*
Identification Renewal Time-Out Period (IRTOP)	The elapsed time after which the output of valid identification data in target reports for a specific track will be terminated due to a lack of renewed identification data from any sensor system.	*[EUROCAE MASPS]*
Information Display Latency (IDL)	The maximum time delay between a report, other than a target report, being received by the A-SMGCS HMI and the corresponding presentation on the HMI display of the information contained in the report.	*[EUROCAE MASPS]*
Position Registration Accuracy (PRA)	The difference between the co-ordinates contained in the dynamic input data to the HMI and the corresponding geographical position represented on the HMI display.	*[EUROCAE MASPS]*
Position Renewal Time-Out Period (PRTOP)	The elapsed time after which the output of target reports for a specific track will be terminated due to a lack of new position information from any sensor system.	*[EUROCAE MASPS]*
Probability of Detection (PD)	The probability that an actual target is reported at the output of the Surveillance Element of an A-SMGCS.	*[EUROCAE MASPS]*
Probability of Detection of an Alert Situation (PDAS)	The probability that the Monitoring/Alerting Element correctly reports an alert situation.	*[EUROCAE MASPS]*
Probability of False Alert (PFA)	The probability that the Control service reports anything other than actual alert situations.	*[EUROCAE MASPS]*
Probability(PFD) of False Detection	The probability that the Surveillance Element of an A-SMGCS reports anything other than actual targets.	*[EUROCAE MASPS]*
Probability of False Identification (PFID)	The probability that the identity reported at the output of the Surveillance Element of an A-SMGCS is not the correct identity of the actual target.	*[EUROCAE MASPS]*
Probability(PID) of Identification	The probability that the correct identity of a co-operative target is reported at the output of the Surveillance Element.	*[EUROCAE MASPS]*
Reported Position Accuracy (RPA)	The difference, at a specified confidence level, between the reported position of the target and the actual position of the target at the time of the report.	*[EUROCAE MASPS]*
Reported Velocity Accuracy (RVA)	The difference, at a specified confidence level, between the reported target velocity and the actual target velocity at the time of the report.	*[EUROCAE MASPS]*
Response Time to Operator Input (RTOI)	The maximum time delay between the operator making an input on a data entry device of an A-SMGCS HMI and the corresponding action being completed or acknowledged on the HMI display.	*[EUROCAE MASPS]*

Surveillance Capacity	The number of target reports in a given period, which the Surveillance Element is able to process and output without degradation below the minimum performance requirements.	*[EUROCAE MASPS]*
System Accuracy	The term accuracy generally describes the degree of conformance between a platform's true position and/or velocity and its estimated position and/or velocity.	*[ICAO A-SMGCS]*
System Availability	Availability is the ability of an A-SMGCS to perform a required function at the initiation of the intended operation within an A-SMGCS area.	*[ICAO A-SMGCS]*
System Capacity	The maximum number of simultaneous movements of aircraft and vehicles that the system can safely support within an acceptable delay commensurate with the runway and taxiway capacity at a particular airport.	*[ICAO A-SMGCS]*
System Continuity	Continuity is the ability of an A-SMGCS to perform its required function without non-scheduled interruption during the intended operation in an A-SMGCS area.	*[ICAO A-SMGCS]*
System Integrity	Integrity relates to the trust, which can be placed in the correctness of the information provided by an A-SMGCS. Integrity includes the ability of an A-SMGCS to provide timely and valid alerts to the user(s) when an A-SMGCS must not be used for the intended operation.	*[ICAO A-SMGCS]*
System Reliability	Reliability is defined as the ability of an A-SMGCS to perform a required function under given conditions for a given time interval.	*[ICAO A-SMGCS]*
Target Display Latency (TDL)	The maximum time delay between a target report being received by the A-SMGCS HMI and the corresponding presentation on the HMI display of the target position contained in the report.	*[EUROCAE MASPS]*
Target(TRUR) Report Update Rate	The frequency with which target reports are output from the Surveillance Element of the A-SMGCS.	*[EUROCAE MASPS]*
Transaction Expiration Time (ETRCP)	Maximum time for completion of a transaction after which peer parties should revert to an alternative procedure. The rate at which a transaction expiration time can be exceeded is determined by the continuity parameter.	[ED-78A]
95% Transaction Time (TT95)	Time before which 95% of the transactions are completed. This is the time at which controllers and pilots can nominally accept the system performance and represents normal operating performance.	[ED-78A]
Continuity (CRCP)	Probability that the transaction will be completed before the transaction expiration time, assuming that the communication system is available when the transaction is initiated	[ED-78A]
Availability (ARCP)	Probability that the communication system between the two parties is in service when it is needed	[ED-78A]
Availability (AProvision)	Probability that communication with all aircraft in the area is in service.	[ED-78A]
Communication Integrity (IRCP)	Acceptable rate of transactions completed with error undetected.	[ED-78A]

Table 2 – A-SMGCS main identification parameters

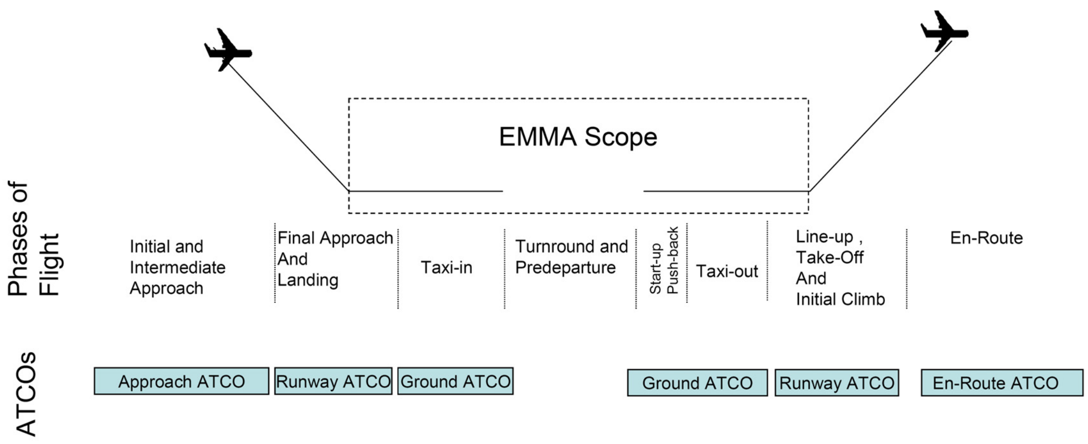


Figure 3 – EMMA's domain

3 AVAILABLE STANDARDS, MANUALS, AND DOCUMENTATIONS

3.1 ICAO

One of ICAO's chief activities is standardisation, the establishment of international standards, recommended practices and procedures covering the technical fields of aviation: licensing of personnel, rules of the air, aeronautical meteorology, aeronautical charts, units of measurement, operation of aircraft, nationality and registration marks, airworthiness, aeronautical telecommunications, air traffic services, search and rescue, aircraft accident investigation, aerodromes, aeronautical information services, aircraft noise and engine missions, security and the safe transport of dangerous goods. After a standard is adopted it is put into effect by each ICAO contracting State in its own territories. As aviation technology continues to develop rapidly, the Standards are kept under constant review and amended as necessary.

In keeping pace with the rapid development of international civil aviation, ICAO is conscious of the need to adopt in its specifications modern systems and techniques. Guidance Material is produced to supplement the SARPS and PANS and to facilitate their implementation. Guidance material is issued as Attachments to Annexes or in separate documents such manuals, circulars and lists of designators/addresses. Usually it is approved at the same time as the related SARPS are adopted.
Manuals provide information to supplement and/or amplify the Standards and Recommended Practices and Procedures for Air Navigation Services. They are specifically designed to facilitate implementation and are amended periodically to ensure their contents reflect current practices and procedures.

Advanced Surface Movement Guidance and Control Systems Manual (A-SMGCS), Doc 9830 AN/452, first edition 2004.
The systems described in the ICAO Manual of Surface Movement Guidance and Control Systems (SMGCS), (Doc 9476) are not always capable of providing the necessary support to aircraft operations

in order to maintain required capacity and safety levels, especially under low visibility conditions. An advanced SMGCS (ASMGCS) therefore, is expected to provide adequate capacity and safety in relation to specific weather conditions, traffic density and aerodrome layout by making use of modern technologies and a high level of integration between the various functionalities. The manual was produced to enable manufacturers and operators as well as certifying authorities to develop and introduce A-SMGCS depending on local circumstances and taking into account global interoperability requirements for international civil aviation operations.

The Manual presents the common basic operational concept and system design for a universal A-SMGCS adaptable to the specific local aerodrome needs. The concept describes the necessary basic A-SMGCS functions Surveillance, Control, Routing and Guidance together with their communication and their evolution towards higher automation.

A-SMGCS provides full service under a wide range of operational conditions:

• all visibility conditions until AVOL,
• growing traffic, and
• complex traffic flows,

to aircraft and affected vehicles on the movement area between runways and stands in order to maintain:

• safety - the required high level of safety, and
• capacity - maximum utilisation of flow rates, given by infrastructure.

To help airport operators to decide on the level of automation they need, ICAO has defined five levels of implementation for particular aerodromes. All four basic ASMGCS functions (i.e. surveillance, control, routing and guidance) are provided at all levels, but the part played by automation and avionics increases progressively through the levels. The international agreed modular design, open architecture and foreseen standardisation of modules/interfaces enable cost effective A-SMGCS solutions for local demands by implementing only those modules for the identified necessary ASMGCS functions. A future upgrade of A-SMGCS is possible in an economic way by adding the appropriate necessary modules without replacing the whole existing system.

Manual of the Secondary Surveillance Radar (SSR) Systems (Doc 9684, 2nd edition, 1998).

This document provides guidance material on characteristics of the ground stations and airborne transponders of SSR systems which are defined in the Standards and Recommended Practices (SARPs) of Annex 10, Volume IV. It also describes the contribution of SSR as a major system for surveillance purposes in most air traffic control (ATC) systems and the data link capability of the Mode S component to be utilized as part of the aeronautical telecommunication network (ATN).

Air Traffic Management - Procedures for Air Navigation Services (PANS-ATM), Doc 4444, 14th edition, 2001.

These procedures are complementary to the Standards and Recommended Practices contained in Annex 2 and Annex 11 and specify, in greater detail than in the Standards and Recommended Practices, the actual procedures to be applied by air traffic services units in providing the various air traffic services to air traffic.

Manual of Surface Movement Guidance and Control Systems (SMGCS) (Doc 9476, reprinted March 2003).

This manual has been developed to facilitate the implementation of specifications relating to SMGC systems found in various Annexes and the PANS-RAC. It contains information on: designing an SMGC system for an aerodrome; the functions and responsibilities of personnel, procedures; low visibility operations; high traffic volume operations; runway protection measures and apron management service.

ANNEX 10 Volume III (Part I - Digital Data Communication Systems; Part II - Voice Communication Systems).

Volume III of Annex 10 contains Standards and Recommended Practices (SARPs) and guidance material for various air-ground and ground-ground voice and data communication systems, including aeronautical telecommunication network (ATN), aeronautical mobile-satellite service (AMSS), secondary surveillance radar (SSR) Mode S air-ground data link, very high frequency (VHF) air-ground digital link (VDL), aeronautical fixed telecommunication network (AFTN), aircraft addressing system, high frequency data link (HFDL), aeronautical mobile service, selective calling system (SELCAL), aeronautical speech circuits and emergency locator transmitter (ELT).

Annex 10 - Volume IV (Surveillance Radar and Collision Avoidance Systems).

Volume IV of Annex 10 contains Standards and Recommended Practices (SARPs) and guidance material for secondary surveillance radar (SSR) and airborne collision avoidance systems (ACAS), including SARPs for SSR Mode A, Mode C and Mode S; and the technical characteristics of ACAS.

3.2 EUROCONTROL/AIRPORT OPERATIONS-PROGRAMME

The Airport Operations Programme forms a part of the EUROCONTROL activity in European Air Traffic Management (EATM) and hosts four projects that will enhance airside safety and capacity. This will be achieved by improving airside efficiency and harmonising the introduction of new technologies.

The A-SMGCS is one of the four above mentioned projects. Its main objectives are set out below:

- To fully develop A-SMGCS Levels 1 & 2, ensuring that all issues relevant to operational implementation are identified and addressed
- Identify, develop & validate appropriate procedures
- Verify performance requirements
- Build Safety & Human Factors Cases
- Address training & licensing
- Support harmonised implementation within ECAC
- Support global implementation (through ICAO)
- Ensure coordination with ICAO, EC, FAA, NAV CANADA etc.
- Provide baseline for further developments (A-SMGCS Levels 3 & 4)

To date a considerable amount of work has been performed by ICAO (AOPG PT/2), EUROCAE (WG41) and the European Commission (DGTREN). The need for A-SMGCS is also recognised through the ATM Strategy for the Years 2000+. This Strategy contains a number of 'Directions for Change' and complementary Operational Improvements, which contribute towards a realisation of the overall concept for the ATM Network within ECAC. The Direction for Change applicable to A-SMGCS is entitled "Improved Traffic Management on the Movement Area" and the associated Operational Improvements are:

1. Improvement of Aerodrome Control Service on the Manoeuvring Area;
2. Improvement of Conflict Detection and Alert for all Traffic on the Movement Area;
3. Improvement of Planning and Routing on the Movement Area;
4. Improvement of Guidance and Control on the Movement Area.

These Operational Improvements are to be met through the A-SMGCS Project of the EATM Airport Operations Programme, which aims to facilitate the implementation of A-SMGCS Levels 1 & 2 through the development of appropriate operational concepts, requirements and procedures. It also aims at addressing related operational issues such as safety, human factors and licensing of controllers.

The first phase of the project has involved the development of agreed requirements for A-SMGCS Levels 1 & 2, namely:

- Agreed User Requirements
- A Concept of Operations
- Operational Requirements
- Functional Requirements

This phase of the A-SMGCS project is now complete and the following documents are available:

A-SMGCS Project Strategy.
The A-SMGCS subjects and specifications have already been tackled and extensively investigated by several organizations such as ICAO, EUROCAE, FAA and EUROCONTROL. The aim of this document is to propose a strategy for A-SMGCS implementation on the basis of the work that has already been performed by these organizations. The Strategy presents an operational vision on how ATS and the relationships amongst airport stakeholders are expected to evolve through the evolutionary implementation of ASMGCS. Such a vision encompasses airspace users, ATM stakeholders, airport operators (i.e. pilots, airlines, airport managers, handling operators, apron vehicle drivers). The document also phases the A-SMGCS implementation in compliance with the context of the gate-to-gate ATM network and the related EATMP Programs as well as the availability of ECAC airport projects and technology such as CDM, AMAN, DMAN, ADS, GNSS etc.

Definition of A-SMGCS Implementation Levels.
This document aims at defining the A-SMGCS implementation levels corresponding to the A-SMGCS project strategy. These Implementation Levels form a coherent series that:
- Recognizes operational needs;
- Reflects the evolution of technologies and procedures;
- Enables airports to equip according to local requirements.

Operational Concept & Requirements for A-SMGCS Implementation Levels I & II.
The EUROCONTROL A-SMGCS project aims at defining pragmatic implementation steps for A-SMGCS. The first step named A-SMGCS Level I focuses on the implementation of automated surveillance. A-SMGCS Level II aims at complementing surveillance service with a control service that provides a runway safety net and prevents incursions into restricted areas. This documents aim at defining the operational concept and the requirements for ASMGCS implementation Levels I & II, i.e. how ATS is expected to evolve through the introduction of the integrated Surveillance technology/functions and how ATS is expected to evolve through the introduction of the A-SMGCS automated control functions, respectively.

Functional Specifications for A-SMGCS Implementation Levels I & II.
On the basis of the analysis of the users needs presented in the Operational Concept & Requirements for A-SMGCS Implementation Levels I & II, this documents define the functional specifications for A-SMGCS Implementation Level I & II. These documents focus on the operational and functional requirements. The operational requirements are already presented in the OC & R documents and they are recalled and listed in these documents for readability purpose.

Validation Master Plan for A-SMGCS Implementation Levels I & II.
This documents aim at defining the Validation Master Plan for A-SMGCS implementation Levels I & II. The Validation Master Plan identifies the objectives and the steps of the validation process. It provides for each step a full description (resources, timeframe, training etc.) and identifies its prerequisites. This documents also identifies the techniques of evaluation (fast time and real time simulations, pre-operational trials at representative airports,…) to assess, demonstrate and confirm that

A-SMGCS fulfil the Operational Concept with respect to the airport manoeuvring area, for all visibility conditions, times of the day and traffic densities. A particular emphasis is placed upon the validation of A-SMGCS related procedures, with the view to providing the data necessary to support their submission to ICAO. To develop The Validation Master Plans for A-SMGCS Levels I & II have been defined following the steps proposed by the MAEVA methodology [MAEVA], which has been especially designed for this kind of exercise by the Master ATM European Validation Plan (MAEVA) project.

A-SMGCS Safety Plan.
In connection with the EATM Airport Operations Programme (APR) - maintained by the Airport Operations Domain - a safety plan shall be developed to define the safety activities to be undertaken within the Advanced Surface Movement Guidance and Control System (A-SMGCS) project of the APR. The safety plan shall be addressing separately, ASMGCS Implementation Levels I & II. The safety activities will be linked with the validation process and will result in two safety cases (for Implementation Levels I & II) proving that the project as it is defined is safe for introduction. Thus, the safety plan shall help to ensure that the safety–related data including safety objectives and safety requirements are validated during the validation process and can be used to develop the appropriate safety cases for A-SMGCS Implementation Levels I & II.

A-SMGCS Safety Policy.
This document, the APR A-SMGCS Safety Policy Document, has been developed by the Airport Throughput Business Unit of EUROCONTROL Advanced Surface Movement Guidance and Control System (A-SMGCS) Project. The APR A-SMGCS Safety Policy Document sets out the Safety Policy, the Safety Objectives and describes the safety-related tasks and actions to ensure a safe development, implementation and continued operation of the A-SMGCS. The APR A-SMGCS Safety Policy Document is intended to provide a framework to facilitate the safety regulation process of the APR A-SMGCS project. The document also represents the initiation of a co-ordination dialogue between the APR A-SMGCS Project and the Safety Regulation Commission (SRC). This co-ordination process will continue throughout the APR A-SMGCS Project and will cover the development, implementation and continued operation of the A-SMGCS. The APR A-SMGCS Safety Policy Document includes description of the deliverables of the APR A-SMGCS Project. The document is closely related to the APR A-SMGCS Safety Plan.

A-SMGCS Human Factors Plan.
The document describes a Human Factors Plan supporting the Validation master Plan for the A-SMGCS. A principle aim of the validation process is to assess the Human Factors impact of A-SMGCS Level 1 and/or 2 implementation, with data obtained during the validation process being used to develop appropriate Human Factor Cases (i.e. Human Factor Cases for Implementation Levels 1 & 2). A 'Human Factor Plan' has to be developed to help identify and ensure that the appropriate data are available for the validation This plan shall define the Human Factor activities to be undertaken within the A-SMGCS project (particularly during the validation process) following EUROCONTROL's guidelines for Human Factors Integration.

In parallel, work has progressed on the identification and development of the procedures, necessary to support the implementation of A-SMGCS. These procedures focus upon enabling the controller, when appropriate, to issue ATC instructions and clearances to aerodrome traffic on the basis of surveillance data alone. Harmonised procedures are being developed to ensure that as A-SMGCS becomes widespread, pilots, vehicle drivers and controllers will be working to the same rules and standards throughout the European region. Their description represents the main content of the "A-SMGCS Operating Procedures" document produced inside the A-SMGCS project, as well.

3.3 EUROCAE'S MISSION AND WG'S 41 ACTIVITIES

EUROCAE is an international non-profit making organisation. Membership is open to European manufacturers of equipment for aeronautics, trade associations, national civil aviation administrations, users, and non-European organisations. Its work programme is principally directed to the preparation of performance specifications and guidance documents for civil aviation equipment, for adoption and use at European and worldwide levels.
The findings of EUROCAE are resolved after discussion among its members and in cooperation with RTCA Inc., Washington DC, USA and/or the Society of Automotive Engineers (SAE), Warrendale PA, USA through their appropriate committees.
The following EUROCAE documents are available:

ED-87 MASPS for Advanced Surface Movement Guidance and Control Systems (ASMGCS).
This document contains the Minimum Aviation System Performance Specification (MASPS) for A-SMGCS at the current level of maturity. This document specifies system and equipment characteristics that should be useful to designers, installers, manufacturers, service providers and users of systems intended for operational use at aerodromes. Functional requirements are used wherever possible to allow flexibility in the design of sub-system equipment. This specification was produced from existing and proposed products, methods and requirements that support the A-SMGCS concept. Adherence to this specification is intended to enable early operational implementation of a system at an aerodrome in accordance with the aerodrome's operational requirements.
Chapter 1 of this document outlines the A-SMGCS role within the future air traffic management system. It describes the rationale for the system concept, the expected operational goals of the A-SMGCS and outlines the fundamental design concepts. It provides definitions of terms used in the document, lists of abbreviations and references.
Chapter 2 outlines the enabling functions of an A-SMGCS and defines an evolutionary approach to configuration levels which have been identified by EUROCAE as the steps for building the A-SMGCS. Advice on interfaces is provided.
Chapter 3 specifies performance requirements for the principal enabling functions of an ASMGCS.
Chapter 4 describes methodologies for the verification of A-SMGCS performance requirements specified in Chapter 3 and contains descriptions of the recommended generic test procedures needed to verify compliance at both functional element and overall system level.
Supplementary information on test, variation and verification methods is provided within appendices A and B.

ED 78A GUIDELINES for approval of the provision and use of Air Traffic Services supported by Data Communications.
This guidance document was jointly prepared by Special Committee 189 (SC-189) and the European Organization for Civil Aviation Equipment (EUROCAE) Working Group 53 (WG-53) and approved by the RTCA Program Management Committee (PMC) on December 14, 2000. This guidance material recommends minimum acceptable criteria for approving the provision and use of an ATS supported by data communications when approvals are required to show compliance to civil regulations. The criteria are in the form of process objectives and guidance for evidence. As used throughout this document, evidence is data produced during the accomplishment of the process objectives. Applicants can use the evidence to show an approval authority that the objectives have been satisfied. For example, evidence may take the form of standards such as the SPR and INTEROP standards, or plans such as the approval plan, or results of verification activities such as test results. This document provides means to establish the operational, safety, performance, and interoperability requirements for ATS supported by

data communications, to assess their validity, and to qualify the related CNS/ATM system. It is a single source document that provides guidance for approval of the CNS/ATM system and its operation where coordination is necessary across organizations. The guidance material considers the allocations of the operational, safety, performance, and interoperability requirements to the elements of the CNS/ATM system. These 4 include ground-based elements, operational procedures, including the human, and aircraft equipage.

SUMMARY OF CHAPTER 6

CHAPTER 7

RANDOM NUMBERS GENERATION (RNG)

This chapter describes algorithms for the generation of pseudorandom numbers with both uniform and normal distributions. Simulations in the following chapters are performed by generating realizations of the underlying stochastic processes, by using random numbers.

Regardless of whether or not one employs variance-reducing techniques, the statistical integrity of every Monte Carlo sampling experiment depends on the availability of an algorithm for generating random numbers or, more precisely, for generating sequences of numbers which by relatively stringent standards can be regarded as indistinguishable from sequences of truly random numbers. These computer-generated quantities are called pseudorandom numbers and their use leads to at least two sources of error. The first comes from the finite word-sizes of digital computers. This limitation induces a discretization error generic to all numerical evaluation methods. Insisting on extended precision in all arithmetic calculations makes the seriousness of this error source relatively minor. Regrettably, developers of pseudorandom number generating algorithms for microcomputers have occasionally ignored this precaution, casting suspicion on the scientific integrity of results garnered from Monte Carlo sampling experiments that used these algorithms (e.g., [RNG-9]).
Unfortunately, the need to produce pseudorandom numbers efficiently restricts the class of generating algorithms that one can realistically consider; and this restriction leads to the second potential source of error [RNG-1]. Ideally, for every positive integer k, one would like to draw statistically independent samples $U^{(1)} = (U_{11},...,U_{1k})$, $U^{(2)} = (U_{21},...,U_{2k}),\dots$ each having the uniform probability distribution over the k-dimensional unit hypercube $\Im^k = [0,1]^k$. However, the rapidly generating algorithms in common use produce sequences whose properties depart from this ideal with error that increases as k increases. Careful choice of a generating algorithm can substantially reduce the potential error from this source., with emphasis on the trade-off between speed and statistical error. However, a user of the Monte Carlo method must recognize that employing the best of methods rarely offers more than two- to three-digit accuracy; and pretending otherwise would misrepresent the accuracy of experimental results. This last observation prompts a more expansive comment about the disparity between theory and practice.

1 PSEUDORANDOM NUMBERS

Random numbers generators (RNGs) are needed for practically all kinds of computer applications, such as simulation of stochastic systems, numerical analysis, probabilistic algorithms, secure communications, computer games, and gambling machines, to name a few [RNG-6]. The so-called random numbers may come from a physical device, like thermal noise from electronic diodes, but are more often the output of a small computer program which, from a given initial value called the seed, produces a deterministic sequence of numbers that are supposed to imitate typical realizations of independent uniform random variables. The latter are sometimes called pseudo-random [RNG-4] number generators, or algorithmic RNGs. Since they have a deterministic and periodic output, it is clear a priori that they do not produce independent random variables in the mathematical sense and that they cannot pass all possible statistical tests of uniformity and independence. But some of them have huge period lengths and turn out to behave quite well in statistical tests that can be applied in reasonable time. On the other hand, several popular RNGs, some available in commercial software, fail very simple tests.

Good RNGs are not designed by trying some arbitrary algorithms and applying empirical tests to the output until all the tests are passed. Instead, their design should involve a rigorous mathematical analysis of their period lengths and of the uniformity of the vectors of successive values that they produce over their entire period length; that's how independence is assessed theoretically. However, once they have been selected and implemented, they must also be tested empirically. Statistical tests are also required for RNGs based (totally or partially) on physical devices. These RNGs are used to generate, for example, keys in cryptosystems and to generate numbers in lotteries and gambling machines.

A good definition for PseudoRandom Numbers was given in 1951 by Berkeley Professor D. H. Lehmer, a pioneer in computing and, especially, computational number theory:

"A random sequence is a vague notion ... in which each term is unpredictable to the uninitiated and whose digits pass a certain number of tests traditional with statisticians ..."

It may seem perverse to use a computer, that most precise and deterministic of all machines conceived by the human mind, to produce "random" numbers. More than perverse, it may seem to be a conceptual impossibility. Any program, after all, will produce output that is entirely predictable, hence not truly "random."

One sometimes hears computer-generated sequences termed *pseudorandom*, while the word random is reserved for the output of an intrinsically random physical process, like the elapsed time between clicks of a Geiger counter placed next to a sample of some radioactive element. We will not try to make such fine distinctions. A working, though imprecise, definition of randomness in the context of computer-generated sequences, is to say that the deterministic program that produces a random sequence should be different from, and — in all measurable respects — statistically uncorrelated with, the computer program that uses its output. In other words, any two different random number generators ought to produce statistically the same results when coupled to your particular applications program. If they don't, then at least one of them is not (from your point of view) a good generator.

The above definition may seem circular, comparing, as it does, one generator to another. However, there exists a body of random number generators which mutually do satisfy the definition over a very, very broad class of applications programs. And it is also found empirically that statistically identical results are obtained from random numbers produced by physical processes. So, because such generators are known to exist, we can leave to the philosophers the problem of defining them.

2 UNIFORM DISTRIBUTION

Uniform deviates are just random numbers that lie within a specified range (typically 0 to 1), with any one number in the range just as likely as any other. They are, in other words, what you probably think "random numbers" are. However, we want to distinguish uniform deviates from other sorts of random numbers, for example numbers drawn from a normal (Gaussian) distribution of specified mean and standard deviation. These other sorts of deviates are almost always generated by performing appropriate operations on one or more uniform deviates, as we will see in subsequent sections. So, a reliable source of random uniform deviates, the subject of this section, is an essential building block for any sort of stochastic modelling or Monte Carlo computer work.

Lehmer also invented the *multiplicative congruential algorithm*, which is the basis for many of the random number generators in use today. Lehmer's generators involve three integer parameters, a, c and m, and an initial value, x_0, called the seed. A sequence of integers is defined by:

$$x_{k+1} = ax_k + c \bmod m \qquad \textbf{Eq. 1}$$

The operation "$\bmod m$" means take the remainder after division by m. For example, with $a = 13, c = 0, m = 31,$ and $x_0 = 1$, the sequence begins with

$$1, 13, 14, 27, 10, 6, 16, 22, 7, 29, 5, 3, \ldots \qquad \textbf{Eq. 2}$$

What's the next value? Well, it looks pretty unpredictable, but you've been initiated. So you can compute $13 \cdot 3 \bmod 31$, which is 8. The first 30 terms in the sequence are a permutation of the integers from 1 to 30 and then the sequence repeats itself. It has a period equal to $m-1$.

If a pseudorandom integer sequence with values between 0 and m is scaled by dividing by m, the result is floating-point numbers uniformly distributed in the interval [0, 1]. Our simple example begins with

$$0.0323, 0.4194, 0.4516, 0.8710, 0.3226, 0.1935, 0.5161, \ldots \qquad \textbf{Eq. 3}$$

There are only a finite number of values, 30 in this case. The smallest value is 1/31; the largest is 30/31. Each one is equally probable in a long run of the sequence.

Although this general framework is powerful enough to provide quite decent random numbers, its implementation in many, if not most, ANSI C libraries is quite flawed; quite a number of implementations are in the category "totally botched." Blame should be apportioned about equally between the ANSI C committee and the implementers.

This can be disastrous in many circumstances: for a Monte Carlo integration, you might well want to evaluate 10^6 different points, but actually be evaluating the same 32767 points 30 times each, not at all the same thing! You should categorically reject any library random number routine with a two-byte returned value. Second, the ANSI committee's published rationale includes the following mischievous passage: "The committee decided that an implementation should be allowed to provide a rand function which generates the best random sequence possible in that implementation, and therefore mandated no standard algorithm. It recognized the value, however, of being able to generate the same pseudorandom sequence in different implementations, and so it has published an example. . . .[emphasis added]" The "example" is:

```
unsigned long next=1;
int rand(void) /* NOT RECOMMENDED (see text) */
{
next = next*1103515245 + 12345;
return (unsigned int)(next/65536) % 32768;
}
void srand(unsigned int seed)
{
next=seed;
}
```

This corresponds to having with $\mathrm{a} = 1103515245, \mathrm{c} = 12345$, and $\mathrm{m} = 2^{32}$ (since arithmetic done on `unsigned long` quantities is guaranteed to return the correct low-order bits). These are not particularly good choices for a and c (the period is only 230), though they are not gross embarrassments by themselves.

The real botches occur when implementers, taking the committee's statement above as license, try to "improve" on the published example. For example, one popular 32-bit PC-compatible compiler provides a long generator that uses the above congruence, but swaps the high-order and low-order 16 bits of the returned value.
Somebody probably thought that this extra flourish added randomness; in fact it ruins the generator. While these kinds of blunders can, of course, be fixed, there remains a fundamental flaw in simple linear congruential generators, which we now discuss. The linear congruential method has the advantage of being very fast, requiring only a few operations per call, hence its almost universal use. It has the disadvantage that it is not free of sequential correlation on successive calls.

Even worse, one might be using a generator whose choices of m, a, and c have been botched. One infamous such routine, `RANDU`. Invented in the 1960's the Scientific Subroutine Library on IBM mainframe computers included a random number generator named `RND` or `RANDU`. It was a multiplicative congruential with parameters $\mathrm{a} = 65539, \mathrm{c} = 0$, and $\mathrm{m} = 231$. With a 32-bit integer word size, arithmetic $\bmod 2^{31}$ can be done quickly. Furthermore, because $a = 2^{16} + 3$, the multiplication by a can be done with a shift and an addition.

Such considerations were important on the computers of that era, but they gave the resulting sequence a very undesirable property. The following relations are all taken $\bmod 2^{31}$

$$x_{k+2} = (2^{16}+3)x_{k+1} = (2^{16}+3)^2 x_k = (2^{32}+6\cdot 2^{16}+9)x_k = [6\cdot(2^{16}+3)-9]x_k \qquad \textbf{Eq. 4}$$

Hence:

$$x_{k+2} = 6x_{k+1} - 9x_k \text{ for all } k \qquad \textbf{Eq. 5}$$

As a result, there is an extremely high correlation among three successive random integers of the sequence generated by `RANDU`.

Correlation in k-space is not the only weakness of linear congruential generators. Such generators often have their low-order (least significant) bits much less random than their high-order bits.

Park and Miller have surveyed a large number of random number generators that have been used over the last 30 years or more. Along with a good theoretical review, they present an anecdotal sampling of a number of inadequate generators that have come into widespread use. The historical record is nothing if not appalling. For many years, the Matlab uniform random number function, rand, was also a multiplicative congruential generator. The parameters were:

$$\begin{cases} a = 7^5 = 16807 \\ c = 0 \\ m = 2^{31} - 1 = 2147483647 \end{cases} \quad \textbf{Eq. 6}$$

These values are recommended in a 1988 paper by Park and Miller ("Minimal Standard" generator [RNG-9]). First proposed by Lewis, Goodman, and Miller in 1969, this generator has in subsequent years passed all new theoretical tests, and (perhaps more importantly) has accumulated a large amount of successful use. Park and Miller [RNG-9] do not claim that the generator is "perfect" (we will see below that it is not), but only that it is a good minimal standard against which other generators should be judged.

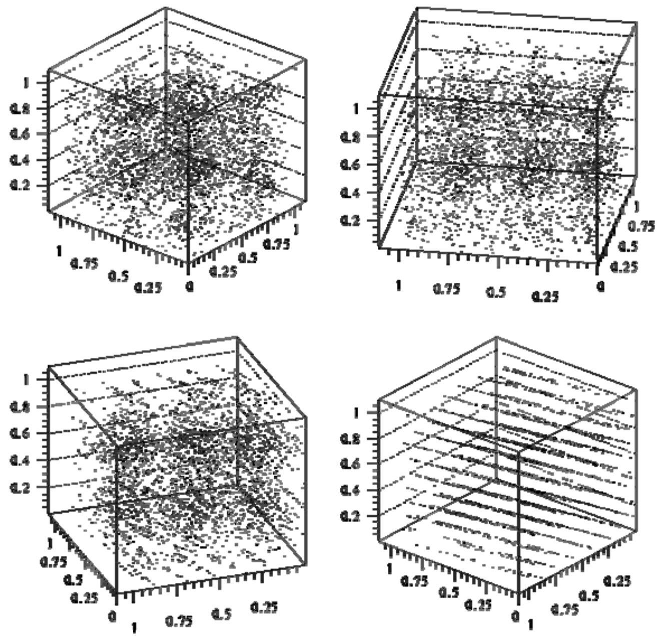


Figure 1 – 3D Scatter Plot

Figure 1 shows that the points do not suffer the correlation of the SSP generator. They generate a much better "random" cloud within the cube.

Like our toy generator, `randmcg` and the old version of the Matlab function `rand` generate all real numbers of the form $k/m \quad k = 1 \ldots m-1$. The smallest and largest are 0.00000000046566 and 0.999999999534340. The sequence repeats itself after $m-1$ values, which is a little over 2 billion numbers. A few years ago, that was regarded as plenty. But today, an 800 MHz Pentium laptop can exhaust the period in less than half an hour. Of course, to do anything useful with 2 billion numbers takes more time, but we would still like to have a longer period.

In 1995, Version 5 of Matlab introduced a completely different kind of random number generator. The algorithm is based on work of George Marsaglia [RNG-8], a professor at Florida State University and author of the classic analysis of random number generators, "Random numbers fall mainly in the planes". Marsaglia's generator [RNG-8] does not use Lehmer's congruential algorithm. In fact, there are no multiplications or divisions at all. It is specifically designed to produce floating-point values. The results are not just scaled integers. In place of a single seed, the new generator has 35 words of internal memory or *state*. Thirty-two of these words form a cache of floating-point numbers, z, between 0 and 1. The remaining three words contain an integer index i, which varies between 0 and 31, a single random integer j, and a "borrow" *flag* b. This entire state vector is built up a bit at a time during an initialization phase. Different values of j yield different initial states.

The generation of the ith floating-point number in the sequence involves a "subtract with borrow" step, where one number in the cache is replaced by the difference of two others:

$$z_i = z_{i+20} - z_{i+5} - b \qquad \textbf{Eq. 7}$$

The three indices, $i, i+20$, and $i+5$, are all interpreted $\bmod 32$ (by using just their last five bits). The quantity b is left over from the previous step; it is either zero or a small positive value. If the computed z_i is positive, b is set to zero for the next step. But if the computed z_i would be negative, it is made positive by adding 1.0 before it is saved and b is set to 2^{-53} for the next step. The quantity 2^{-53}, which is half of the Matlab [RNG-7] constant `eps`, is called one *ulp* because it is one *unit in the last place* for floating-point numbers slightly less than 1.

By itself, this generator would be almost completely satisfactory. Marsaglia [RNG-8] has shown that it has a huge period - almost 2^{1430} values would be generated before it would repeat itself. But it has one slight defect. All the numbers are the results of floating-point additions and subtractions of numbers in the initial cache, so they are all integer multiples of 2^{-53}. Consequently, many of the floating-point numbers in the interval [0, 1] are not represented.

The floating-point numbers between 1/2 and 1 are equally spaced with a spacing of one *ulp*, and our subtract-with-borrow generator will eventually generate all of them. But numbers less than 1/2 are more closely spaced and the generator would miss most of them. It would generate only half of the possible numbers in the interval [1/4, 1/2], only a quarter of the numbers in [1/8,1/4] and so on. This is where the quantity j in the state vector comes in. It is the result of a separate, independent, random number generator based on bitwise logical operations. The floating-point fraction of each z_i is XORed with j to produce the result returned by the generator. This breaks up the even spacing of the numbers less than 1/2. It is now theoretically possible to generate all the floating-point numbers between 2^{-53} and $1-2^{-53}$. We're not sure if they are all actually generated, but we don't know of any that can't be.

Figure 2 shows what the new generator is trying to accomplish. For this graph, one ulp is equal to 2^{-4} instead of 2^{-53}.

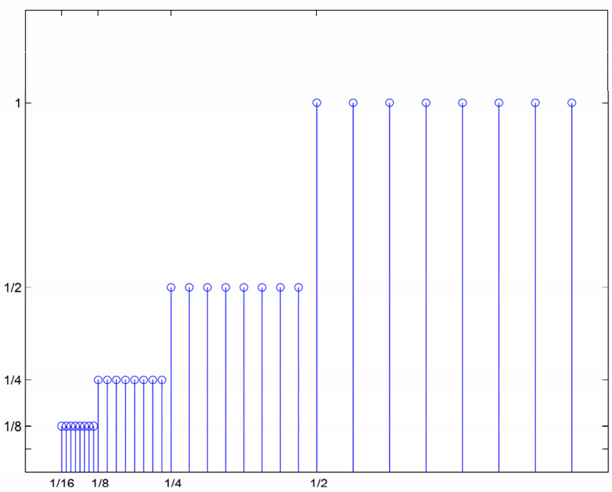


Figure 2 – Uniform distribution of floating-point numbers [RNG-8]

The graph in Figure 1 depicts the relative frequency of each of the floating-point numbers. A total of 32 floating-point numbers is shown. Eight of them are between 1/2 and 1 and they are all equally like to occur. There are also eight numbers between 1/4 and 1/2, but, because this interval is only half as wide, each of them should occur only half as often. As we move to the left, each subinterval is half as wide as the previous one, but it still contains the same number of floating-point numbers, so their relative frequencies must be cut in half. Imagine this picture with 253 numbers in each of 232 smaller intervals and you will see what the new random number generator is doing.

With the additional bit fiddling, the period of the new generator becomes something like 2^{1492}. Maybe we should call it the Christopher Columbus generator. In any case, it will run for a very long time before it repeats itself.

3 TRANSFORMATION METHOD: EXPONENTIAL AND NORMAL DEVIATES

In the previous section, we learned how to generate random deviates with a uniform probability distribution, so that the probability of generating a number between x and $x+dx$, denoted $p(x)dx$, is given by

$$p(x)dx = \begin{cases} dx & 0 < x < 1 \\ 0 & otherwise \end{cases} \qquad \textbf{Eq. 8}$$

The probability distribution $p(x)$ is of course normalized, so that

$$\int_{-\infty}^{+\infty} p(x)dx = 1 \qquad \textbf{Eq. 9}$$

Now suppose that we generate a uniform deviate x and then take some prescribed function of it, $y(x)$. The probability distribution of y, denoted $p(y)dy$, is determined by the fundamental transformation law of probabilities, which is simply

$$|p(y)dy| = |p(x)dx| \qquad \textbf{Eq. 10}$$

or better:

$$p(y) = p(x)\left|\frac{dx}{dy}\right| \qquad \textbf{Eq. 11}$$

3.1 EXPONENTIAL DEVIATES

As an example, suppose that $y(x) = -\ln(x)$, and that $p(x)$ is as given by the equation for a uniform deviate. Then:

$$p(y)dy = \left|\frac{dx}{dy}\right|dy = e^{-y}dy \qquad \text{Eq. 12}$$

which is distributed exponentially. This exponential distribution occurs frequently in real problems, usually as the distribution of waiting times between independent Poisson-random events, for example the radioactive decay of nuclei. You can also easily see that the quantity y/λ has the probability distribution $\lambda e^{-\lambda y}$.

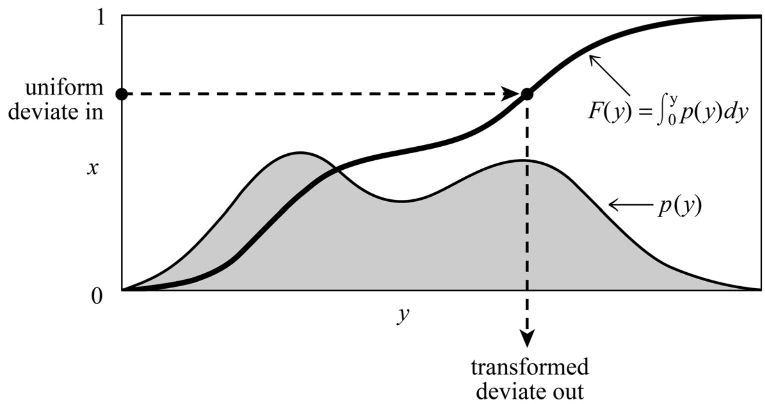


Figure 3 – Transformation method for generating a random deviate y from a known probability distribution p(y). The indefinite integral of p(y) must be known and invertible. A uniform deviate x is chosen between 0 and 1. Its corresponding y on the definite-integral curve is the desired deviate [RNG-6]

Let's see what is involved in using the above *transformation method* to generate some arbitrary desired distribution of y's, say one with $p(y) = f(y)$ for some positive function f whose integral is 1. (See Figure 3) According to what stated above, we need to solve the differential equation

$$\frac{dx}{dy} = f(y) \qquad \text{Eq. 13}$$

But the solution of this is just $x = F(y)$, where $F(y)$ is the indefinite integral of $f(y)$. The desired transformation which takes a uniform deviate into one distributed as $f(y)$ is therefore

$$y(x) = F^{-1}(x) \qquad \text{Eq. 14}$$

where F^{-1} is the inverse function to F. Whether is feasible to implement depends on whether the *inverse function of the integral* of $f(y)$ is itself feasible to compute, either analytically or numerically. Sometimes it is, and sometimes it isn't.

Incidentally, this has an immediate geometric interpretation: Since $F(y)$ is the area under the probability curve to the left of y, this is just the prescription: choose a uniform random x, then find the value y that has that fraction x of probability area to its left, and return the value y.

3.2 REJECTION METHOD: GAMMA, POISSON, BINOMIAL DEVIATES

The rejection method is a powerful, general technique for generating random deviates whose distribution function $p(x)dx$ (probability of a value occurring between x and $x+dx$) is known and

computable. The rejection method does *not* require that the cumulative distribution function [indefinite integral of $p(x)$] be readily computable, much less the inverse of that function — which was required for the transformation method in the previous section.
The rejection method is based on a simple geometrical argument: Draw a graph of the probability distribution $p(x)$ that you wish to generate, so that the area under the curve in any range of x corresponds to the desired probability of generating an x in that range. If we had some way of choosing a random point in two dimensions, with uniform probability in the area under your curve, then the x value of that random point would have the desired distribution.
Now, on the same graph, draw any other curve $f(x)$ which has finite (not infinite) area and lies everywhere above your original probability distribution. (This is always possible, because your original curve encloses only unit area, by definition of probability.) We will call this $f(x)$ the comparison function. Imagine now that you have some way of choosing a random point in two dimensions that is uniform in the area under the comparison function. Whenever that point lies outside the area under the original probability distribution, we will reject it and choose another random point. Whenever it lies inside the area under the original probability distribution, we will accept it. It should be obvious that the accepted points are uniform in the accepted area, so that their x values have the desired distribution.

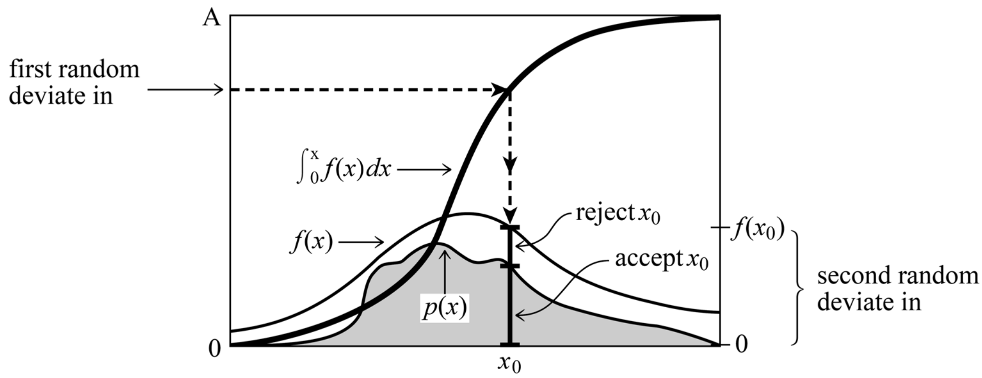


Figure 4 – Rejection method for generating a random deviate x from a known probability distribution p(x) that is everywhere less than some other function f(x). The transformation method is first used to generate a random deviate x of the distribution f (compare Figure 3). A second uniform deviate is used to decide whether to accept or reject that x. If it is rejected, a new deviate of f is found; and so on. The ratio of accepted to rejected points is the ratio of the area under p to the area between p and f [RNG-6]

It should also be obvious that the fraction of points rejected just depends on the ratio of the area of the comparison function to the area of the probability distribution function, not on the details of shape of either function. For example, a comparison function whose area is less than 2 will reject fewer than half the points, even if it approximates the probability function very badly at some values of x, e.g., remains finite in some region where $p(x)$ is zero.

It remains only to suggest how to choose a uniform random point in two dimensions under the comparison function $f(x)$. A variant of the transformation method does nicely: be sure to have chosen a comparison function whose indefinite integral is known analytically, and is also analytically invertible to give x as a function of "area under the comparison function to the left of x." Now pick a uniform deviate between 0 and A, where A is the total area under $f(x)$, and use it to get a corresponding x. Then pick a uniform deviate between 0 and $f(x)$ as the y value for the two-dimensional point. You should be able to convince yourself that the point (x, y) is uniformly distributed in the area under the comparison function $f(x)$. An equivalent procedure is to pick the second uniform deviate between zero and one, and accept or reject according to whether it is respectively less than or greater than the ratio $p(x)/f(x)$.

So, to summarize, the rejection method for some given $p(x)$ requires that one find, once and for all, some reasonably good comparison function $f(x)$. Thereafter, each deviate generated requires two uniform random deviates, one evaluation of f (to get the coordinate y), and one evaluation of p (to decide whether to accept or reject the point x, y). Figure 4 illustrates the procedure. Then, of course, this procedure must be repeated, on the average, A times before the final deviate is obtained.

3.3 POISSON DEVIATES

The Poisson distribution is conceptually related to the gamma distribution. It gives the probability of a certain integer number m of unit rate Poisson random events occurring in a given interval of time x, while the gamma distribution was the probability of waiting time between x and $x + dx$ to the m th event. Note that m takes on only integer values ≥ 0, so that the Poisson distribution, viewed as a continuous distribution function $p_x(m)dm$, is zero everywhere except where m is an integer ≥ 0. At such places, it is infinite, such that the integrated probability over a region containing the integer is some finite number. The total probability at an integer j is

$$\text{Prob}(j) = \int_{j-\varepsilon}^{j+\varepsilon} p_x(m)dm = \frac{x^j e^{-x}}{j!} \qquad \textbf{Eq. 15}$$

At first sight this might seem an unlikely candidate distribution for the rejection method, since no continuous comparison function can be larger than the infinitely tall, but infinitely narrow, *Dirac delta functions* in $p_x(m)$. However, there is a trick that we can do: spread the finite area in the spike at j uniformly into the interval between j and $j+1$. This defines a continuous distribution $q_x(m)dm$ given by

$$q_x(m)dm = \frac{x^{[m]} e^{-x}}{[m]!} dm \qquad \textbf{Eq. 16}$$

where $[m]$ represents the largest integer less than m. If we now use the rejection method to generate a (noninteger) deviate from $q_x(m)dm$, and then take the integer part of that deviate, it will be as if drawn from the desired distribution $\text{Prob}(j)$. (See Figure 5) This trick is general for any integer-valued probability distribution. For x large enough, the distribution $q_x(m)dm$ is qualitatively bell-shaped (albeit with a bell made out of small, square steps), and we can use the same kind of Lorentzian comparison function as was already used above. For small x, we can generate independent exponential deviates (waiting times between events); when the sum of these first exceeds x, then the number of events that would have occurred in waiting time x becomes known and is one less than the number of terms in the sum.
These ideas produce the following routine:

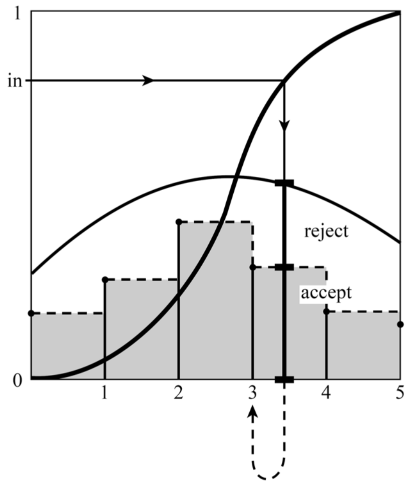


Figure 5 – Rejection method as applied to an integer-valued distribution. The method is performed on the step function shown as a dashed line, yielding a real-valued deviate. This deviate is rounded down to the next lower integer, which is output [RNG-6]

```
#include <math.h>
#define PI 3.141592654
float poidev(float xm, long *idum)
Returns as a floating-point number an integer value that is a random deviate drawn from a Poisson
distribution of mean xm, using ran1(idum) as a source of uniform random deviates.
{
float gammln(float xx);
float ran1(long *idum);
static float sq,alxm,g,oldm=(-1.0); oldm is a flag for whether xm has changed since last call.
float em,t,y;
if (xm < 12.0) { Use direct method.
if (xm != oldm) {
oldm=xm;
g=exp(-xm); If xm is new, compute the exponential.
}
em = -1;
t=1.0;
do { Instead of adding exponential deviates it is equivalent to multiply uniform deviates. We never
actually have to take the log, merely compare to the pre-computed exponential.
++em;
t *= ran1(idum);
} while (t > g);
} else { Use rejection method.
if (xm != oldm) { If xm has changed since the last call, then
preoldm=xm; compute some functions that occur below.
sq=sqrt(2.0*xm);
alxm=log(xm);
g=xm*alxm-gammln(xm+1.0);
The function gammln is the natural log of the gamma function
}
do {
do { y is a deviate from a Lorentzian comparison function.
y=tan(PI*ran1(idum));
em=sq*y+xm; em is y, shifted and scaled.
} while (em < 0.0); Reject if in regime of zero probability.
em=floor(em); The trick for integer-valued distributions.
t=0.9*(1.0+y*y)*exp(em*alxm-gammln(em+1.0)-g);
The ratio of the desired distribution to the comparison function; we accept or reject by comparing
it to another uniform deviate. The factor 0.9 is chosen so that t never exceeds 1.
} while (ran1(idum) > t);
}
return em;
}
```

4 QUALITY CRITERIA FOR RANDOM NUMBER GENERATORS

The default generators of many popular software programs fail several tests miserably. The survivors are the long-period MRGs with good structure, the multiplicative lagged-Fibonacci generators, some nonlinear generators designed for cryptology, and some combined generators with components from different families. SWB and additive lagged-Fibonacci generators also pass when using appropriate decimation, but the decimation slows them down significantly. It seems that combined generators with components from different families should be given better attention because theoretical guarantees about their uniformity can be proved, their period can easily be made very long, splitting their period into long disjoint sub-streams is easy to do if one can do it for the components and it is not hard to select the components with this in mind.

RNGs for all types of applications are designed so that their output sequence is a good imitation of a sequence of independent uniform random variables, usually over the real interval $(0,1)$ or over the binary set $\{0,1\}$. In the first case, the relevant hypothesis H_0^A to be tested is that the successive output values of the RNG, say $u_0, u_1, u_2, \ldots$, are independent random variables from the uniform distribution over the interval $(0,1)$, i.e, i.i.d. $U(0,1)$.

In the second case, H_0^B says that we have a sequence of independent random bits, each taking the value 0 or 1 with equal probabilities independently of the others.

These two situations are strongly related, because under the i.i.d. $U(0,1)$ hypothesis, any pre-specified sequence of bits (e.g., the bit sequence formed by taking all successive bits of u_0, or every second bit, or the first five bits of each u_i, etc.) must be a sequence of independent random bits. So statistical tests for bit sequences can be used as well (indirectly) for testing the null hypothesis H_0^A.

In the $U(0,1)$ case, H_0^A is equivalent to saying that for each integer $t > 0$, the vector $(u_0, \ldots, u_{t-1})$ is uniformly distributed over the t-dimensional unit cube $(0,1)^t$.

This cannot be true for algorithmic RNGs, because these vectors always take their values only from the finite set Ψ_t of all t-dimensional vectors of t successive values that can be produced by the generator, from all its possible initial states (or seeds).

The cardinality of this set cannot exceed the number of admissible seeds for the RNG. Assuming that the seed is chosen at random, vectors are actually generated in Ψ_t to approximate the uniform distribution over $(0,1)^t$. This suggests that Ψ_t should be very evenly distributed over the unit cube. Theoretical figures of merit for measuring this uniformity are used for the design of good RNGs. These criteria are much easier to compute for linear generators.

This is one of the main reasons for the popularity of generators based on linear recurrences; e.g., linear congruential generators (LCGs), multiple recursive generators (MRGs), linear feedback shift-register (LFSR) generators, and generalized feedback shift-register (GFSR) generators.

For a sequence of bits, the null hypothesis H_0^B cannot be formally true as soon as the length t of the sequence exceeds the number b of bits in the generator's state, because the number of distinct sequences of bits that can be produced cannot exceed 2^b. For $t > b$, the fraction of all 2^t sequences of t bits that can be visited is at most 2^{b-t}. The goal, then, is to make sure that those sequences that can be visited are "uniformly scattered" in the set of all 2^t possible sequences, and perhaps hard to distinguish.

Different quality criteria are used for RNGs in cryptology-related applications and for gambling machines in casinos. In these settings, an additional concern is unpredictability of the forthcoming numbers. The theoretical analysis of RNGs in cryptology is usually asymptotic, in the framework of computational complexity theory. Nonlinear recurrences and/or output functions are used, which prevents one from measuring the uniformity of the set Ψ_t. As a result, empirical testing is even more necessary.

For a more detailed discussion on empirical testing of RNGs and a discussion on the links and similarities between different tests, identification of specific types of regularities or defects that they are likely to detect, please refer to "TestU01: A C Library for Empirical Testing of Random Number Generators", L'Ecuyer, Simard, Université de Montréal, ACM Transactions on Mathematical Software [RNG-5].

SUMMARY OF CHAPTER 7

CHAPTER 8
MONTE CARLO SIMULATION

We can define Monte Carlo Analysis as a computer-based method of analysis that uses statistical sampling techniques in obtaining a probabilistic approximation to the solution of a mathematical equation or model [MC-2].

The actual use of Monte Carlo methods as a research tool stems from work on the atomic bomb during the Second World War. This work involved a direct simulation of the probabilistic problems concerned with random neutron diffusion in fissile material; but even at an early stage of these investigations, von Neumann and Ulam refined this particular "Russian roulette" and "splitting" methods. However, the systematic development of these ideas had to await for the work of Harris and Herman Kahn in 1948. About 1948 Fermi, Metropolis, and Ulam obtained Monte Carlo estimates for the eigenvalues of Schrodinger equation.

The name "Monte Carlo" was coined by Metropolis (inspired by Ulam's interest in poker) during the Manhattan Project, because of the similarity of statistical simulation to games of chance, and because Monte Carlo, the capital of Monaco was a centre for gambling.

The generally accepted birth date of the Monte Carlo method is 1949, when an article entitled "The Monte Carlo method" by Metropolis and Ulam appeared [MC-1]. The American mathematicians John von Neumann and Stanislav Ulam are considered its main originators. In the Soviet Union, the first papers on the Monte Carlo method were published in 1955 and 1956 by V. V. Chavchanidze, Yu. A. Shreider and V. S. Vladimirov.

Curiously enough, the theoretical foundation of the method had been known long before the von Neumann-Ulam article was published. Furthermore, well before 1949 certain problems in statistics were sometimes solved by means of random sampling - that is, in fact, by the Monte Carlo method. However, because simulation of random variables by hand is a laborious process, use of the Monte Carlo method as a universal numerical technique became practical only with the advent of computers.
As for the name "Monte Carlo", it is derived from that city in the Principality of Monaco famous for its casinos. The point is that one of the simplest mechanical devices for generating random numbers is the roulette wheel.

In about 1970, the newly developing theory of computational complexity began to provide a more precise and persuasive rationale for employing the Monte Carlo method. The theory identified a class of problems for which the time to evaluate the exact solution to a problem within the class grows, at least, exponentially. The question to be resolved was whether or not the Monte Carlo method could estimate the solution to a problem in this intractable class to within a specified statistical accuracy in time bounded.

Monte Carlo now refers to any method that utilizes sequences of random numbers to perform statistical simulation. The main requirement to use Monte Carlo method for simulation of a physical system is that it must be possible to describe the system in terms of probability density function (PDF), also called partition function (Z). Once the PDF or Z for a system is known, then the simulation begins by random "sampling" from the PDF, and subsequently determining the desired properties of the sample by conducting some kind of a "trial". There must be a rule available, based on some reasonable mathematical and/or physical theory, to decide the outcome of such a trial. Many trials are conducted and outcomes of all of these trials are recorded. The final step in the Monte Carlo method is that the behaviour of the overall system is obtained by computing the average of outcomes of the trails conducted [MC-3].

1 PRINCIPLES OF MONTE CARLO METHOD

The Monte Carlo method provides approximate solutions to a variety of mathematical problems by performing statistical sampling experiments on a computer. Remarkably, the method applies to problems with absolutely no probabilistic content as well as to those with inherent probabilistic structure [MC-4]. This alone does not give the Monte Carlo method an advantage over other methods of approximation. However, among all numerical methods that rely on n-point evaluations in m-dimensional space to produce an approximate solution, the Monte Carlo method has absolute error of estimate that decreases as $n^{-1/2}$ whereas, in the absence of exploitable special structure, all others have errors that decrease as $n^{-1/m}$ at best. This property gives the Monte Carlo method a considerable edge in computational efficiency as m, the size of the problem, increases. Combinatorial settings illustrate this property especially well. Whereas the exact solution to a combinatorial problem with m elements often has computational cost that increases exponentially or super-exponentially with m, the Monte Carlo method frequently provides an estimated solution with tolerable error at a cost that increases no faster than as a polynomial in m.

Deceptively simple in concept, the Monte Carlo method incorporates three distinct, but related, historical developments in the mathematical sciences. First, games of chance motivated seventeenth- and eighteenth-century mathematicians to regard outcomes on successive trials as forming a sequence of random events. Observing that the mean of a function of continuous random variables took the form of an integral, nineteenth- and early-twentieth-century statisticians subsequently recognized that, in principle, one could randomly draw numbers, transform them according to prescribed rules, and derive an approximate solution to an integral in a problem that intrinsically contained no probabilistic content whatever (e.g., National Bureau of Standards 1951, p. 40). In the late nineteenth century, a second line of inquiry developed when Lord Rayleigh (1899) showed that a one-dimensional random walk without absorbing barriers could provide an approximate solution to a parabolic differential equation. In the course of demonstrating that under appropriate conditions a particular finite difference equation could produce an approximate solution to the Dirichlet boundary-value problem of partial differential equations, Courant et al. (1928) showed that the recursive form of the solution to a two-dimensional random walk on a square grid within a closed region whose boundary points were absorbing states produced the identical difference equation. Shortly thereafter, Kolmogorov (1931) showed the relationship between Markov stochastic processes and certain integro-differential equations. Petrowsky (1933) considerably generalized the result of Courant et al. by showing the asymptotic connection between a random walk whose sequence of locations formed a Markov chain and the solution to an elliptic partial differential equation. He called this formulation the generalized Dirichlet problem. At that time, these revelations derived their significance from the observation that solutions to problems encountered in stochastic processes often corresponded to solutions that arose in the study of partial differential equations, and, therefore, one could employ the difference equation methods applicable for solving the differential equations to provide approximate solutions to the original stochastic problem. During the development of atomic energy in the post-World War I era, a third line of inquiry evolved. Scientists needed to solve problems of neutron diffusion or transport through an isotropic medium. These multidimensional problems proved too formidable for the difference equation approach. Since it had already been established that the resulting partial differential and integral equations all had analogous in stochastic processes, John von Neumann and Stanislaw Ulam suggested that sampling experiments using random walk models and executed on the newly developed digital computer could provide readily usable approximations to the desired solutions. This proposal reversed the direction of reasoning. Instead of using the cumbersome difference equation approach to provide solutions to probabilistic problems, one conducted sampling experiments to provide solutions to the integro-differential equations, which did not necessarily have a probabilistic basis themselves. Moreover, the availability of the digital computer provided the means for putting this suggestion into practice. Later, the concept of employing sampling experiments on a computer came to prevail in many scientific disciplines, most notably, in chemistry, engineering, operations research, physics, and statistics. In operations research, studies of all but the most elementary queuing models rely on generating random tours on computers to provide approximate solutions. In physics, the Monte Carlo method has come to be recognized as the only approach capable of providing useful insights to many neutron transport and lattice structure problems. In about 1970, the newly developing theory of computational complexity began to provide a more precise and persuasive rationale for employing the Monte Carlo method. The theory identified a class of problems for which the time to evaluate the exact solution to a problem within the class grows, at least, exponentially with m. Many commonly encountered problems belong to this class and, although special structure when it exists often can lead to an algorithm for exact evaluation in time bounded above by a polynomial in m, the absence of special structure leaves intact the formidable cost of exact evaluation. The existence of this substantial class of problems gave new currency to the idea that the Monte Carlo method might offer a competitive solution. The question to be resolved was whether or not the Monte Carlo method could estimate the solution to a problem in this intractable class to within a specified statistical accuracy in time bounded above by a polynomial in m. Numerous examples now support this contention. Karp (1985) shows this property for estimating reliability in a planar multi-terminal network with randomly failing edges. Dyer et al. (1989) establish it for estimating the volume of a convex body in m-dimensional Euclidean space (Sec. 5.32). Broder (1986) and Jerrum and Sinclair (1988) establish the property for estimating the permanent of a matrix or, equivalently, the number of perfect matchings in a bipartite graph. The simplicity of the concept belies the mathematical sophistication with which one can imbue the Monte Carlo method. In particular, von Neumann, Ulam, and others recognized that one could modify the standard Monte Carlo method in a way that produced a solution to the original problem with specified error bound at considerably less cost, in terms of computing time, than directly generating the random tour that corresponded to the original problem. Some of these variance reducing techniques were already known to statisticians working in the area of sample survey, but others owe their origin to the Monte Carlo method. Collectively, these procedures have come to represent the central focus of the Monte Carlo method. The improved computing efficiency to which they lead follows from their exploiting available structural information about a problem that the elementary or standard Monte Carlo method ignores. At the beginning of the computer age, the relative slowness of first-generation computers made the application of variance-reducing techniques an essential ingredient of any successful application of the Monte Carlo method. However, now as then, an analyst must tailor a variance-reducing technique for each problem, and this tailoring takes the analyst's time. In an age in which personal microcomputers are readily available and workstations are becoming increasingly available, one may choose to conserve the amount of exceedingly valuable tailoring time in favour of allowing a computer to grind away for long, but exceedingly inexpensive, hours to provide an approximate solution using the standard Monte Carlo method. Regarded as methodological blasphemy until as recently as 1970, when all Monte Carlo work was performed on mainframe computers, following this philosophy in some circumstances for problems that can fit comfortably time-wise and space-wise into a microcomputer or workstation now seems eminently reasonable. Nevertheless, problems of substantial size remain and, regardless of whether one uses a microcomputer, workstation, or a mainframe computer, the feasibility of achieving estimated solutions with moderately high statistical accuracy depends critically on how effectively one exploits available information.

2 CONTROLLING THE ERROR IN MONTE CARLO ANALYSIS

While controlling error prevails as an underlying theme, limits exist on our ability to do just that. Justification for employing the Monte Carlo method comes from two basic tenets of statistics. Let $X_1, X_2, ..., X_n$ denote a sequence of mutually independent, identically distributed random variables. Let $S_n = X_1 + ... + X_n$. If the expectation $\mu = EX_i$ exists, the Weak Law of Large Numbers states that, for every $\varepsilon > 0$,

$$\lim_{n\to\infty} pr\left(\left|\frac{S_n}{n} - \mu\right| > \varepsilon\right) = 0 \qquad \textbf{Eq. 1}$$

If, additionally, the expectation $\sigma^2 = E(X_i - \mu)^2$ exists, then the *Central Limit Theorem* asserts that, for every fixed a:

$$\lim_{n\to\infty} pr\left(\frac{S_n - n\mu}{\sigma\sqrt{n}} < a\right) = (2\pi)^{-1/2} \int_{-\infty}^{a} e^{-z^2/2} dz \qquad \textbf{Eq. 2}$$

The Law of Large Numbers leads one to believe that, as the sample size n increases, the error of approximation in estimating μ by S_n / n becomes vanishingly small and, in principle, the Central Limit Theorem provides a way of assessing the extent of that statistical error for large n. In practice, however, the Monte Carlo method relies on techniques that contradict both of these statistical laws. Since all numerical-valued samples generated in a Monte Carlo experiment arise from transforming numbers that a pseudorandom number generator produces, and since sequences of these numbers repeat themselves after a finite number of steps P, sampling without limit in a Monte Carlo experiment does not make statistical error vanish. For experiments executed on conventional mainframes, personal computers, and workstations, these realities limit achievable accuracy. Although most experiments demand levels of accuracy well within these limits, this sobering fact of life should be clearly understood before employing the Monte Carlo method. Although the Central Limit Theorem provides a simple way of deriving an approximation to the statistical error in a point estimate, the goodness of the approximation depends on the rate at which the distribution of $\frac{(S_n - n\mu)}{\sigma\sqrt{n}}$ converges to the standard normal distribution. Inevitably, convergence rates differ among problems and, therefore, in the absence of any information on convergence one must regard the normal approximation as an additional source of error for any finite sample size n. However, this last error can be eliminated by resorting to alternative statistical techniques. This is especially true for random variables with known lower and upper bounds. To achieve a specified bound on error, these alternative techniques generally predict a larger sample size than the approximating normal approach does. This is the penalty for eliminating the additional error that assuming normality would induce.

3 SEQUENTIAL AND NON-SEQUENTIAL TECHNIQUES

The basic goal of our Monte Carlo analysis in this thesis is to characterize, quantitatively, the uncertainty and variability in estimates of risk of collision among aircrafts. A secondary goal is to identify key sources of variability and uncertainty and to quantify the relative contribution of these sources to the overall variance and range of model results.

Analytical techniques represent every system by analytical models and evaluate the indices from these models using mathematical solutions. Monte Carlo simulation methods, however, estimate the indices by simulating the actual process and random behaviour of the system. When complex operating conditions are involved (such as an airport), Monte Carlo methods are often preferable.

There are two basic approaches when Monte Carlo methods are applied to reliability evaluation. These methods are known as the *sequential* and *non-sequential* techniques.

The analytical approach and the non-sequential Monte Carlo methods are usually restricted to the evaluation of expected values and, sometimes, to a limited range of system parameters. There is frequently a need to know the likely range of the reliability indices, the likelihood of a certain value being exceeded, and similar parameters. These can only be assessed from a knowledge of the probability distribution associated with the index in question, and this is rarely achievable using an analytical approach. In such cases, sequential simulation can be utilised.

In sequential simulation, each subsequent system state sample is related to the previous set of system states. A sequential time evolution of system behaviour is created which enables a wide range of reliability indices to be evaluated. Sequential simulation is very useful when the system to be analysed is past-dependent, i.e. the state of the system at any given time is partially determined by the historical evolution of the system. Sequential simulation becomes a necessity when the operating system is history-dependent or time correlated. At the present time, sequential simulation is the only realistic option available to develop the distributions associated with the system index mean values [MC-6]. This approach can be used to represent most of the contingencies and operating characteristics inherent in the system and, also, to provide the most comprehensive range of reliability indices. The sequential Monte Carlo simulation is the only method which can comprehensively deal with the time-varying aspect of the loads at each bus.

The probability distributions associated with the times-to-failure and times-to-repair of each system component are often not known and only the meantime-to-failure (MTTF) and the mean-time-to-repair (MTTR) are available. In these cases, it is generally assumed that the underlying distributions are exponential and the expected values are calculated on that basis. The sequential simulation method can, however, be used to sample the failure and repair times from any statistical distribution, such as Weibull, normal etc..

In this thesis we are using a *sequential* Monte Carlo approach for composite system reliability/safety assessment [MC-7]. The reliability indices of an actual physical system can be estimated by collecting data on the occurrence of failures and the Monte Carlo method mimics the failure and repair history of the components and the system by using the probability distributions of the component state durations. Statistics are then collected and indices estimated using statistical inference [MC-8].

3.1 FIXED INTERVAL METHOD AND ASYNCHRONOUS TIMING

We have already seen that there are two basic approaches for Monte Carlo simulation, sequential simulation, and random sampling. The sequential simulation proceeds by generating a sequence of events using random numbers and probability distributions of random variables representing component state durations. In random sampling, states are drawn based on the probability distributions of component states and random numbers. The sampling method is generally faster than the sequential technique, but is suitable when component failures and repairs are independent. In this thesis we will use the sequential method for reliability analysis.

Further, there are two methods for representing the passage of time in sequential simulation: (1) the *fixed interval method*, also called *synchronous timing*, and (2) the next event or *asynchronous timing* method. In the fixed interval method, time is advanced in steps of fixed length and the system state is updated. In the next event method, time is advanced to the occurrence of the next event. In actual implementations, it is likely that combinations of the timing controls may be used.

The flowchart for this method is shown in Figure 1. The whole procedure consists of the following steps.

3.1.1 DATA INPUT AND INITIALIZATION

The input data consists of the failure rate (λ) and duration (r) of every component. The failure rate is the reciprocal of the mean up time. The failure duration or mean down time is the reciprocal of the repair rate (μ). The failure and repair rates, λ and μ, of a component will be used to determine how long the component will remain in the "UP" state and the "DOWN" state.
Simulation could be started from any system state, but it is customary to begin simulation with all the components in the "UP" state.

3.1.2 RANDOM NUMBER GENERATION

Random numbers constitute a sequence in which each number has an equal probability of assuming any one of the possible values, and is statistically independent of the other numbers in the sequence. Random numbers, therefore, basically constitute a uniform distribution over a suitably selected range of values. This distribution may be constructed using any suitable means. For a more detailed description of the Random Number Generation process, please refer to Chapter 7.

3.1.3 COMPUTATION OF TIME TO THE NEXT EVENT

The time to the next event is generated by using the inverse of probability distribution method. This method can be understood by considering the probability mass function of a random variable. The first step is to convert this mass function into the corresponding distribution function. Now a random number z between 0 and 1 is generated and $F(x)$ is set equal to z. The corresponding value of X gives the value of the random variable. An example is shown in Figure 2, with $z = 0.55$, for which $X = 2$.

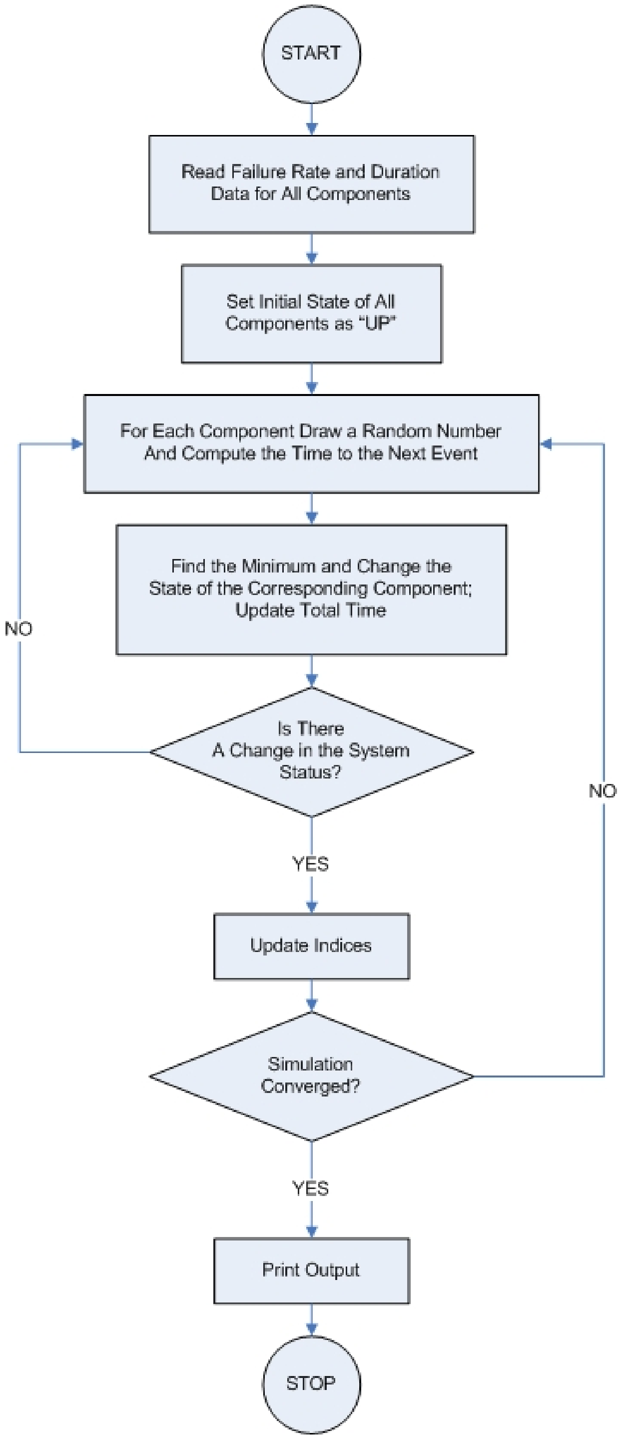


Figure 1 – Flowchart for Next Event Simulation

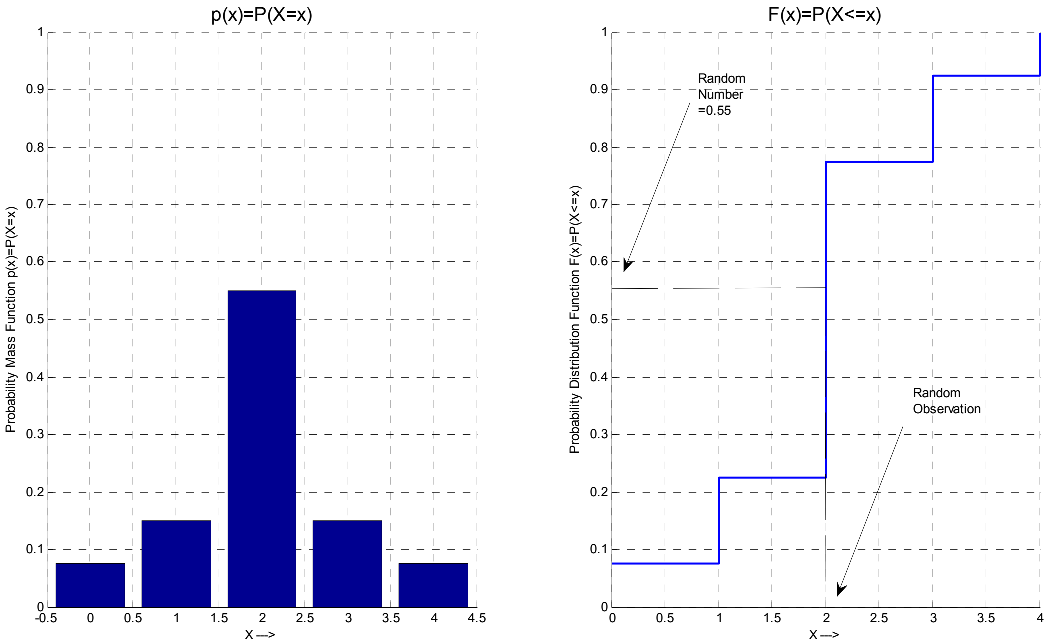


Figure 2 – PMF and PDF of a Random Variable

It should be noticed that $F(x_i)-F(x_{i-1})$ is equal to $P(X=x_i)$, and if the random number falls in the interval $[F(x_i),F(x_{i-1})]$, the value of $X=x_i$ will be selected. The procedure therefore essentially allocates the random numbers to the random variables in the proportion of their probabilities of occurrence.

This procedure can also be used for continuous distributions. Continuous distributions are approximated by discrete distributions whose irregularly spaced points have equal probabilities. The accuracy can be increased by increasing the number of intervals into which $[0,1]$ is divided. This requires additional data in the form of tables. Although the method is quite general, its disadvantages are the great amount of work required to develop tables and possible computer storage problems. The following analytic inversion approach is simpler.

Let z be a random number in the range 0 to 1 with a uniform probability density function, i.e., a triangular distribution function:

$$f(z)=\begin{cases}0 & Z<0\\ 1 & 0\le Z\le 1\\ 0 & Z>1\end{cases} \qquad \textbf{Eq. 3}$$

Similarly

$$F(z)=\begin{cases}0 & Z<0\\ z & 0\le Z\le 1\\ 1 & Z>1\end{cases} \qquad \textbf{Eq. 4}$$

Let $F(x)$ be the distribution function from which the random observations are to be generated. Let

$$z=F(x) \qquad \textbf{Eq. 5}$$

Solving the equation for x gives a random observation of X. That the observations so generated do have $F(x)$ as the probability distribution can be shown as follows.
Let ϕ be the inverse of F; then

$$x = \phi(z) \quad \textbf{Eq. 6}$$

Now x is the random observation generated. We determine its probability distribution as follows:

$$P(x \le X) = P(F(x) \le F(X)) = P(z \le F(X)) = F(X) \quad \textbf{Eq. 7}$$

Therefore the distribution function of x is $F(X)$, as required.
In the case of several important distributions, special techniques have been developed for efficient random sampling.
For our purposes, the distributions assumed for up and down times are exponential. The exponential distribution has the following probability distribution:

$$P(X \le x) = 1 - e^{-\rho x} \quad \textbf{Eq. 8}$$

where $1/\rho$ is the mean of the random variable X. Setting this function equal to a random decimal number between 0 and 1,

$$z = 1 - e^{-\rho x} \quad \textbf{Eq. 9}$$

Since the complement of such a random number is also a random number, the above equation can as well be written as

$$z = e^{-\rho x} \quad \textbf{Eq. 10}$$

Taking the natural logarithm of both sides and simplifying, we get:

$$x = -\frac{\ln(z)}{\rho} \quad \textbf{Eq. 11}$$

which is the desired random observation from the exponential distribution having $1/\rho$ as the mean.
This method is used to determine the time to the next transition for every component, using λ, or μ for ρ, depending on whether the component is UP or DOWN. The smallest of these times indicates the most imminent event, and the corresponding component is assigned a change of state. If this event also results in a change of status, (i.e., failure or restoration) of the system, then the corresponding system indices are updated.

3.1.4 THE INDICES

At any time t, the mean failure frequency is given by

$$\lambda_t = \frac{1}{t}(\text{number of failures till time } t) \qquad \textbf{Eq. 12}$$

And the mean down time is given by:

$$r_t = \frac{1}{t}(\text{total time spent in failed state}) \qquad \textbf{Eq. 13}$$

3.1.5 CONVERGENCE

The simulation is said to have converged when the indices attain stable values [MC-6]. This "stabilization" of the value of an index i is measured by its standard error, defined as:

$$\eta = \frac{\sigma_i}{\sqrt{n_c}} \qquad \textbf{Eq. 14}$$

where σ_i is the standard deviation of the index i and n_c the number of cycles simulated.
Convergence is said to occur when the standard error drops below a pre-specified fraction, ε, of the index i, i.e., when

$$\eta \le \varepsilon \cdot i \qquad \textbf{Eq. 15}$$

If, for instance, the mean down time r is chosen as the index to converge upon, then, after every system restoration simulated, the following relation is tested for validity:

$$\frac{\sigma_i}{\sqrt{n_c}} \le \varepsilon \cdot r \qquad \textbf{Eq. 16}$$

If this criterion is satisfied, the simulation is said to have converged.

3.1.6 STATISTICS OBTAINED FROM SIMULATION

Simulation is advantageous in that it not only allows the computation of indices at various points in the system, but also permits the accumulation of data pertaining to the distribution of these indices, thereby affording a better understanding of the system behaviour.

For an emergency power system, for instance, statistics may be collected for failure frequency and duration at various points in the system, the annual incidence rates for failures, as well as for the variances of these indices.

The approximation in the Monte Carlo method comes from the fact that in this approach, the statistics are estimates of the true values, and therefore cannot be exact. It can be seen that if the error is to be reduced to half, the number of samples has to increase four times. Thus for systems with high reliability (or low probability of failure), the Monte Carlo simulation can take a very long time to converge.

The main advantage of simulation is that it is very flexible for incorporating dependent failures and is very suitable for large systems. Also, it yields the probability distribution of indices in addition to estimating mean values. These probability distributions are useful for performing sensibility analysis.

SUMMARY OF CHAPTER 8

CHAPTER 9
PETRI NETS

This thesis, for the purpose of airport traffic simulation, mainly deals with Petri nets, one of the existing formalism for DES modelling and control.

Petri Nets [PN-7] are a graphical and mathematical modelling tool applicable to many systems. They are a promising tool for describing and studying information processing systems that are characterized as being concurrent, asynchronous, distributed, parallel, nondeterministic, and/or stochastic.

As a graphical tool, Petri Nets [PN-8] can be used as a visual-communication aid similar to flow charts, block diagrams, and networks. In addition, tokens are used in these nets to simulate the dynamic and concurrent activities of systems. As a mathematical tool, it is possible to set up state equations, algebraic equations, and other mathematical models governing the behaviour of systems. Petri nets can be used by both practitioners and theoreticians. Thus, they provide a powerful medium of communication between them: practitioners can learn from theoreticians how to make their models more methodical, and theoreticians can learn from practitioners how to make their models more realistic.

1 INTRODUCTION TO PETRI NETS

Historically speaking, the concept of the Petri net has its origin in Carl Adam Petri's dissertation submitted in 1962 to the faculty of Mathematics and Physics at the Technical University of Darmstadt, West Germany. The dissertation was prepared while C. A. Petri worked as a scientist at the University of Bonn. Petri's work came to the attention of A. W. Holt, who later led the Information System Theory Project of Applied Data Research, Inc., in the United States. The early developments and applications of Petri nets (or their predecessor) are found in the reports associated with this project, and in the Record of the 1970 Project MAC Conference on Concurrent Systems and Parallel Computation. From 1970 to 1975, the Computation Structure Group at MIT was most active in conducting Petri-net related research, and produced many reports and theses on Petri nets. In July 1975, there was a conference on Petri Nets and Related Methods at MIT, but no conference proceedings were published. Most of the Petri-net related papers written in English before 1980 are listed in the annotated bibliography of the first book on Petri nets. More recent papers up until 1984 and those works done in Germany and other European countries are annotated in the appendix of another book. Three tutorial articles provide a complementally, easy-to-read introduction to Petri nets.

Since the late-1970's, the Europeans have been very active in organizing workshops and publishing conference proceedings on Petri nets. In October 1979, about 135 researchers mostly from European countries assembled in Hamburg, West Germany, for a two-week advanced course on General Net Theory of Processes and Systems. The 17 lectures given in this course were published in its proceedings, which is currently out of print. The second advanced course was held in Bad Honnef, West Germany, in September 1986. The proceedings of this course contain 34 articles, including two recent articles by C. A. Petri; one is concerned with his axioms of concurrency theory and the other with his suggestions for further research. The first European Workshop on Applications and Theory of Petri Nets was held in 1980 at Strasbourg, France. Since then, this series of workshops has been held every year at different locations in Europe: 1981, Bad Honnef, West Germany; 1982, Varenna, Italy; 1983, Toulouse, France; 1984, Aarhus, Denmark; 1985, Espoo, Finland; 1986, Oxford, Great Britain; 1987, Zaragoza, Spain; 1988, Venice, Italy; and 1989, Bad Honnef, West Germany (planned). The distribution of the proceedings of these workshops is limited to mostly the workshop participants. However, selected papers from these workshops and other articles have been published by Springer-Verlag as Advances in Petri Nets. The 1987 volume contains the most comprehensive bibliography of Petri nets listing 2074 entries published from 1962 to early 1987. The "recent publications" section of Petri Net Newsletter lists short abstracts of recent publications three times a year, and is a good source of information about the most recent Petri net literature. In July 1985, another series of international workshops was initiated. This series places emphasis on timed and stochastic nets and their applications to performance evaluation. The first international workshop on timed Petri nets was held in Torino, Italy, in July 1985; the second was held in Madison, Wisconsin, in August 1987; the third was held in Kyoto, Japan, in December 1989; and the fourth was planned in Australia in 1991. The proceedings of the first two workshops, are available from the IEEE Computer Society Press.

The above is a brief history of Petri nets. Now, we look at some application areas considered in the literature. Petri nets have been proposed for a very wide variety of applications. This is due to the generality and permissiveness inherent in Petri nets. They can be applied informally to any area or system that can be described graphically like flow charts and that needs some means of representing parallel or concurrent activities. However, careful attention must be paid to a trade-off between modelling generality and analysis capability. That is, the more general the model, the less amenable it is to analysis. In fact, a major weakness of Petri nets is the complexity problem, i.e., Petri-net-based models tend to become too large for analysis even for a modest-size system. In applying Petri nets, it is often necessary to add special modifications or restrictions suited to the particular application. Two successful application areas are performance evaluation and communication protocols. Promising areas of applications include modelling and analysis of distributed-software systems, distributed-database systems, concurrent and parallel programs, flexible manufacturing/industrial control systems, discrete-event systems, multiprocessor memory systems, dataflow computing systems, fault-tolerant systems, programmable logic and VLSI arrays, asynchronous circuits and structures, compiler and operating systems, office-information systems, formal languages, and logic programs. Other interesting applications considered in the literature are local-area networks, legal systems, human factors, neural networks, digital filters, and decision models.

The use of computer-aided tools is a necessity for practical applications of Petri nets. Most Petri-net research groups have their own software packages and tools to assist the drawing, analysis, and/or simulation of various applications. A recent article provides a good overview of typical Petri-net tools existing as of 1986.

2 FUNDAMENTALS

We assume here that the reader already knows Discrete Event Systems (DES), Petri Net (PN) concepts and the terminology introduced in [PN-2], [PN-3], [PN-4] and [PN-5] as for Coloured Petri Nets (CPNs). This chapter recalls only the main characteristics about DESs, PNs and CPNs used in this thesis.

3 DISCRETE-EVENT CONTROL SYSTEMS

Transport systems can be seen as having several interacting concurrent processes, showing features such as conflict situations, mutual exclusion states and non-determinism. The interaction among processes occurs according to the abrupt occurrence of events and asynchronously (event-driven instead of time-driven).

Because of these characteristics transport systems can be classified as discrete-event dynamic systems since their behaviour is very difficult to describe using the traditional control theory, which deals with systems of continuous or asynchronous discrete variables modelled by differential or difference equations.

In general, however, "real world" systems either do not conform to some assumptions made in order to simplify a model, or they are just too complex to yield analytical solutions. The model may still be valid but there are no tools to solve the equations which make up such a model. Simulation is a process through which a system model is numerically evaluated, and the data from this process are used to estimate various quantities of interest. Analytical solutions are very hard to obtain while, on the other hand, simulations represent a very attractive tool for their study.

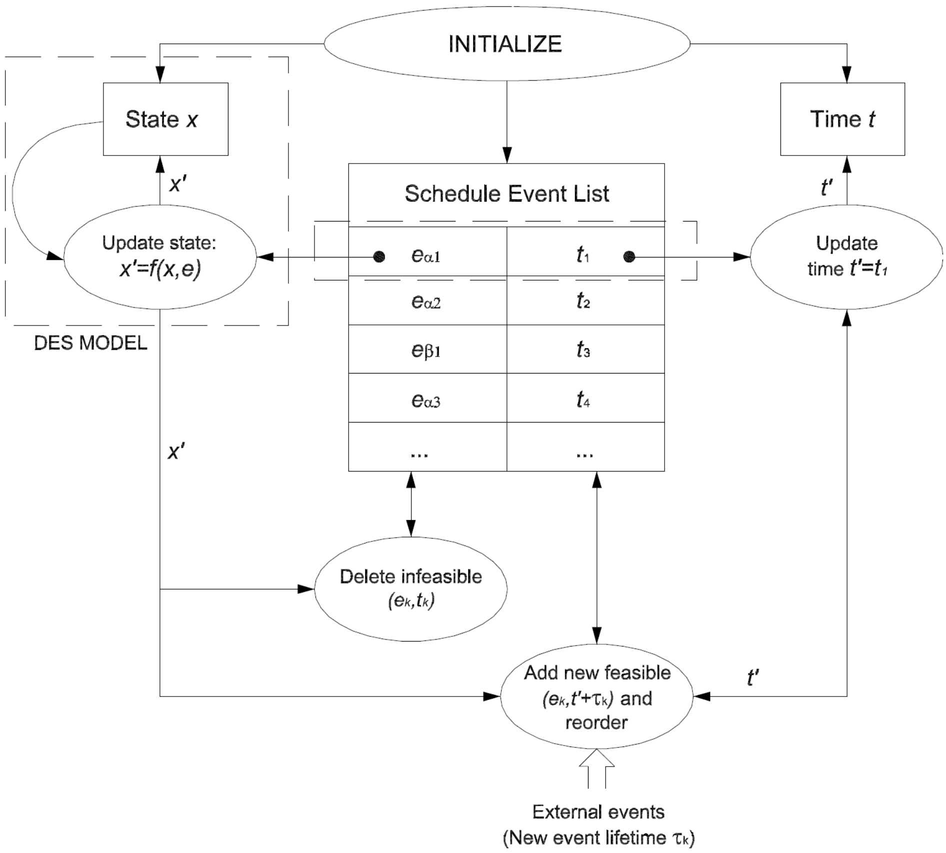


Figure 1 – The event scheduling scheme in computer simulation

DES simulation can be realized by a very general scheme, shown in Figure 1 and named event scheduling scheme. The event scheduling scheme should be thought as a procedure for generating sample paths based on the used DES model and driven by a given clock structure. The scheduled event list (SEL) contains all feasible events at the current state, sorted on a smallest-scheduled-time-first. The next event is always $e_{\alpha 1}$ and it occurs at time t_1. It causes updates to state x and time t. Then, based on the new value of the state, some events are activated and some are deactivated. Events that are activated are entered in the SEL with their scheduled occurrence times, maintaining the right sorting order.

4 TRANSITION ENABLING AND FIRING

In this section, we give the only rule one has to learn about Petri-net theory: the rule for transition enabling and firing. Although this rule appears very simple, its implication in Petri-net theory is very deep and complex. A Petri Net is a particular kind of directed graph, together with an initial state called the initial marking, M_0. The underlying graph N of a Petri net is a directed, weighted, bipartite graph consisting of two kinds of nodes, called places and transitions, where arcs are either from a place to a transition or from a transition to a place. In graphical representation, places are drawn as circles, transitions as bars or boxes. Arcs are labelled with their weights (positive integers), where a k-weighted arc can be interpreted as the set of k parallel arcs. Labels for unity weight are usually omitted. A marking (state) assigns to each place a non negative integer. If a marking assigns to place p a nonnegative integer k, we say that p is marked with k tokens. Pictorially, we place k black dots (tokens) in place p. A marking is denoted by M, an m-vector, where m is the total number of places. The p th component of M, denoted by $M(p)$, is the number of tokens in place p.

In modelling, using the concept of conditions and events, places represent conditions, and transitions represent events. A transition (an event) has a certain number of input and output places representing the pre-conditions and post-conditions of the event, respectively. The presence of a token in a place is interpreted as holding the truth of the condition associated with the place. In another interpretation, k tokens are put in a place to indicate that k data items or resources are available. Some typical interpretations of transitions and their input places and output places are shown in Table 1. A formal definition of a Petri net is given in Table 2.

Input Places	Transition	Output Places
Preconditions	Event	Postconditions
Input data	Computation step	Output data
Input signals	Signal processor	Output signals
Resources needed	Task or job	Resources released
Conditions	Clause in logic	Conclusion(s)
Buffers	Processor	Buffers

Table 1 – Some Typical Interpretations of Transitions and Places

Formal Definition of a Petri Net
A Petri net is a 5-tuple, $PN = (P, T, F, W, M_0)$ where:
$P = \{p_1, p_2, \ldots, p_m\}$ is a finite set of places, $T = \{t_1, t_2, \ldots, t_n\}$ is a finite set of transitions, $F \subseteq (P \times T) \cup (T \times P)$ is a set of arcs (flow relation), $W : F \rightarrow \{1,2,3,\ldots\}$ is a weight function, $M_0 : P \rightarrow \{0,1,2,3,\ldots\}$ is the initial marking, $P \cap T = 0$ and $P \cup T = 0$.
A Petri net structure $N = (P, T, F, W)$ without any specific initial marking is denoted by N.
A Petri net with the given initial marking is denoted by (N, M_0).

Table 2 – Formal Definition of a Petri Net

The behaviour of many systems can be described in terms of system states and their changes. In order to simulate the dynamic behaviour of a system, a state or marking in a Petri nets is changed according to the following *transition* (*firing*) rule:

1) A transition t is said to be enabled if each input place p of t is marked with at least $w(p,t)$ tokens, where $w(p,t)$ is the weight of the arc from p to t.
2) An enabled transition may or may not fire (depending on whether or not the event actually takes place).
3) A firing of an enabled transition t removes $w(p,t)$ tokens from each input place p of t, and adds $w(p,t)$ tokens to each output place p of t, where $w(p,t)$ is the weight of the arc from t to p.

A transition without any input place is called a source transition, and one without any output place is called a sink transition. Note that a source transition is unconditionally enabled, and that the firing of a sink transition consumes tokens, but does not produce any.
A pair of a place p and a transition t is called a self-loop if p is both an input and output place of t.
A Petri net is said to be pure if it has no self-loops. A Petri net is said to be ordinary if all of its arc weights are 1's.

5 BACKGROUND ON PETRI NETS

An ordinary net is a structure $N = \langle P, T, Pre, Post \rangle$ where: P is a set of m *places* represented by circles; T is a set of n *transitions* represented by bars; $P \cap T = 0$, $P \cup T \neq 0$ $Pre(Post)$ is the $|P| \times |T|$ sized, binary valued, pre-(post-)incidence matrix. For instance, $Pre(p,t) = 1(Post(p,t) = 1)$ means that there is an arc from $p(t)$ to $t(p)$. The *preset* and *postset* of a node $X \in P \cup T$ are denoted ${}^\bullet X$ and $X^\bullet$. The incidence matrix C of the net is defined as $C = Post - Pre$. An ordinary net N is a Marked Graph (MG) if ${}^\bullet p = p^\bullet = 1, \forall p \in P$. An ordinary net N is a Free Choice Net (FCN) if $\forall p \in P, |p^\bullet| \leq 1$ or ${}^\bullet\{p^\bullet\} = \{p\}$.

A *marking* is an $m \times 1$ vector $m : P \rightarrow \mathbb{N}$ that assigns to each place of a P/T net a non-negative integer number of tokens. A P/T system or net system $\langle N, m_0 \rangle$ is a P/T net N with an initial marking

m_0. A transition $t \in T$ is enabled at a marking m if $m \geq Pre(\cdot,t)$. If t is enabled, then it may fire yielding a new marking $m' = m + Post(\cdot,t) - Pre(\cdot,t) = m + C(\cdot,t)$. The notation $m[t > m'$ means that an enabled transition t may fire at m yielding m'. A *firing sequence* from m_0 is a (possibly empty) sequence of transitions $\sigma = t_1 \ldots t_k$ such that $m_0[t_1 > m_1[t_2 > m_2 \ldots [t_k > m_k$. A marking m is reachable in $\langle N, m_0 \rangle$ if there exists a firing sequence σ such that $m_0[\sigma > m$. Given a net system $\langle N, m_0 \rangle$ the set of reachable markings is denoted $R(N, m_0)$.

The function $\sigma : T \to \mathbb{N}$, where $\sigma(t)$ represents the number of occurrences of t t in σ, is called firing count vector of the fireable sequence σ. If $m_0[\sigma > m$, then it is possible to write in vector form $m = m_0 + C(\cdot,t)\cdot\sigma$, known as state equation of the system.

Right (Left) annuller vectors of C are called *T-flows* *(P-flows)*, i.e. $x : T \to \mathbb{Q}, x \neq 0 \mid Cx = 0$ $(x{:}P \to \mathbb{Q}, x \neq 0 | x^T C = 0)$. P-flows (T-flows) form a linear space (it is also possible to consider real-valued solutions; however since incidence matrices only have integer entries, every real-valued solution is the product of a real scalar and a rational-valued solution). When positive integer solutions are considered, right (left) annuller vectors of C are called *T-invariants (P-invariants)*, i.e. $x : T \to \mathbb{N}, x \neq 0 \mid Cx = 0$ $(x : P \to \mathbb{N}, x \neq 0 \mid x^T C = 0)$; P-invariants (T-invariants) do not form a linear space. The support of a T-invariant (P-invariant) x is defined as $\|x\| = \{t \in T \mid x(t) > 0\}$ $(\|x\| = \{p \in P \mid x(p) > 0\})$. A T-invariant (P-invariant) x has a *minimal support* if there exists no other invariant x' such that $\|x'\| \subset \|x\|$. A T-invariant (P-invariant) is *canonical* if the greatest common divisor of its components is 1. A T-invariant (P-invariant) is said to be *minimal* if it is canonical and has a minimal support. A T-invariant (P-invariant) x is said to be positive if $x > 0$. N is consistent if $\exists y \in (\mathbb{N}^+)^n$ such that $Cy = 0$.

A P/T system is *live* when, from every reachable marking, every transition can ultimately occur. N is *structurally live* if $\exists m_0$ such that $\langle N, m_0 \rangle$ is live.

A place $p \in P$ is said to be *k-bounded* if $\forall m \in R(N, m_0), m(p) \leq k$. A net system $\langle N, m_0 \rangle$ is said to be k-bounded if each one of its places is k-bounded, and it is bounded if it is bounded for some $k \in \mathbb{N}$. A net N is *structurally bounded* if $\forall m_0$ the net system $\langle N, m_0 \rangle$ is bounded. N is structurally bounded if $\exists x \in (\mathbb{N}^+)^m$ such that $x^T C \leq 0$. N is *conservative* if $\exists x \in (\mathbb{N}^+)^m$ such that $x^T C = 0$.

A nonempty subset of places S in a ordinary net N is called a *siphon* if ${}^\bullet S \subseteq S^\bullet$, i.e., every transition having an output place in S has an input place in S. A nonempty subset of places Q in an ordinary net N is called a *trap* if ${}^\bullet Q \subseteq Q^\bullet$, i.e., every transition having an input place in Q has an output place in Q. Given a net system $\langle N, m_0 \rangle$, it is possible to obtain from the initial marking m_0 as many new markings as the number of the enabled transitions. From each new marking, it is possible to reach again more markings. This process results in a tree representation of the markings. Nodes represent markings generated from m_0 (the root) and its successors, and each arc represents a transition firing, which transforms one marking to another. The obtained tree representation, however, grows infinitely if the net is unbounded. To keep the tree finite, it is possible to introduce a special symbol ω, which can be thought as "infinity". This tree, with the presence of ω is called *coverability tree.* For a bounded Petri net, the coverability tree is called *reachability tree* since it contains all possible reachable markings.

6 SUPERVISORY CONTROL OF P/T NETS

In the *supervisory con*trol PN theory it is assumed that the set of transitions T of a net is partitioned into two disjoint subsets: T_{uc}, the set of uncontrollable transitions, and T_c, the set of controllable transitions.

Let's consider a PN system $\langle N, m_0 \rangle$ with m places, whose set of reachable markings is $R(N, m_0) \subset \mathbb{N}^m$. Let $L \subseteq \mathbb{N}^m$ be a set of legal markings, and consider the basic control problem of designing a supervisor that restricts the reachability set of the plant in closed-loop to $L \cap R(N, m_0)$. When controllable transitions are present, also legal markings from which it is possible to exit from L by firing only uncontrollable transitions have to be forbidden.

Of particular interest are those PN state-based control problems where the set of legal markings L is expressed by a set of n_c linear inequality constraints called Generalized Mutual Exclusion Constraints (GMECs). A single GMEC is a couple (l, k) where $l : P \to \mathbb{Z}$ is an $m \times 1$ weight vector and $k \in \mathbb{Z}$. Given the net system $\langle N, m_0 \rangle$, a GMEC defines a set of markings called *legal markings*: $M(l, k) = \{m \in \mathbb{N}^m \mid l^T m \le k\}$. The markings that are not legal are called *forbidden markings*. A controllable transition may be disabled by the *supervisor* - a controlling agent which ensures that the behaviour of the system is a legal one, i.e. it must ensure that the forbidden markings are not reached. T_{uc}^* denotes the set of all possible sequences of uncontrollable transitions.

An approach to solve the supervisory control problem - when the plant system is modelled as a Petri net - has been presented in [PN-3]; it has been shown that if the critical subnet is acyclic, a maximally permissive control law has the following form:

a transition $t \in T_c$ has to be enabled under the net marking m if $m[t > m'$ and

$$l^T m' + l^T C_{uc} \sigma_{uc}^* \le k \qquad \textbf{Eq. 1}$$

Where σ_{uc}^* is the solution of the following ILP

$$\max_{\sigma_{uc}} l^T C_{uc} \sigma_{uc}$$
$$s.t. \begin{cases} \sigma_{uc} \ge 0 \\ C_{uc} \sigma_{uc} \ge -m' \end{cases} \qquad \textbf{Eq. 2}$$

where C_{uc} and σ_{uc} are respectively the incidence matrix and the firing count vector of the uncontrollable subnet (i.e. obtained from the plant net by removing uncontrollable transitions).

7 COLOURED TIMED PETRI NETS

Coloured Timed Petri Nets (CTPNs) extends the framework of PNs by adding colour, time and modular attributes to the net. The colour attribute is developed to deal with systems that have similar or redundant logical structures. The time attribute allows various time-based performance measures to be added in the system model.

A multi-set m is a set which may contain multiple occurrences of elements of a non-empty set S. Formally, a multi-set is a function $m: S \rightarrow \mathbb{N}$, where $\mathbb{N}$ is the set of non-negative integers, represented as a formal sum

$$\sum_{s \in S} m(s)s. \qquad \textbf{Eq. 3}$$

One denotes by: S_{MS} the set of all multi-sets over S; $Type(v)$ the type of a variable v; *Type(expr)* the type of an expression $expr$;
Var(expr) the set of variables in an expression $expr$. In addition the binding of a set of variables V is defined as the function b associating to each variable $v \in V$ an element $b(v) \in Type(v)$ and $expr < b >$ as the value obtained by evaluating an expression $expr$ in a binding b, i.e. substituting for each variable $v \in Var(expr)$ the value $b(v) \in Type(v)$ determined by the binding.
Now, the following definition is recalled [PN-2]:

Definition 1. A CPN is a 9-tuple $CPN = (S, P, T, A, N, C, G, E, I)$ where:

S is a set of non-empty types, called colour sets;
P is a finite set of places drawn by circles;
T is a finite set of transitions including immediate transitions drawn by black bars and timed transitions drawn by empty boxes;
A is a finite set of arcs such that $P \cap T = P \cap A = T \cap A = 0$;
N is a node function defined from A to $P \times T \cup T \times P$;
C is a colour function defined from P into S;
G is a guard function defined from T into expressions such that $\forall t \in T : [Type(G(t)) = B \wedge Type(Var(G(t))) \subseteq S]$, where B is the boolean type;
E is an arc expression function defined from A into expressions such that $\forall a \in A : [Type(E(a)) = C(p)_{MS} \wedge Type(Var(E(a))) \subseteq S]$, where p is a place of $N(a)$;
I is an initialization function defined from P into closed expressions such that $\forall p \in P : [Type(I(p)) = C(p)_{MS}]$.

In order to define the binding of a transition t, the following notation is introduced:

$$A(t) = \{a \in A \mid N(a) \in P \times \{t\} \cup \{t\} \times P\} \qquad \textbf{Eq. 4}$$

$$Var(t) = \{v \mid Var(G(t)) \vee \exists a \in A(t) : v \in Var(E(a))\} \qquad \textbf{Eq. 5}$$

A *binding* of a transition t is a function b defined on $Var(t)$ such that

$$v \in Var(t) : b(v) \in Type(v) \text{ and } G(t) < b >= 1$$ **Eq. 6**

i.e. the guard function is true. $B(t)$ denotes the set of all bindings for t.

A token is a pair (p,c) where $p \in P$ and $c \in C(p)$, while a binding element is a pair (t,b) where $t \in T$ and $b \in B(t)$. The set of all token elements is denoted by TE while the set of all binding elements is denoted by BE. A *marking* is a multi-set over TE while a *step* is a non-empty and finite multi-set over BE. The sets of all markings and steps are denoted by M and Y respectively. As assumption, $S = S_1 \times S_2 \times \ldots \times S_n$ and a function between each set S_i and the integer number set $\mathbb{Z}$ is defined; this leads to represent a token in a place p by a n-tuple $< t >= (c_1, c_2, \ldots, c_n)$, where c_j is an integer. The black token - i.e. the token having no colour - is represented by (1). A step Y is enabled in a marking M if the following property is satisfied

$$\forall p \in P : \sum_{(t,b) \in Y} E(p,t) < b > \leq M(p)$$ **Eq. 7**

When a step is enabled in a marking M_1 it may occur and, if the transition t fires, the net marking changes according to

$$M_2(p) = M_1(p) - \sum_{(t,b) \in Y} E(p,t) < b > + \sum_{(t,b) \in Y} E(t,p) < b > \quad \forall p \in P$$ **Eq. 8**

In this case it is said that M_2 is reachable by M_1. This is written as $M_1[Y > M_2$.

The introduction of a timing specification is essential in order to use Petri net models for performance evaluation of distributed systems. This thesis considers nets with deterministic timing and one phase firing rule, i.e. a timed enabling (called the service time of the transition) followed by an atomic firing. Service times of transitions are supposed to be mutually independent and time independent.
A time function $f : B(t) \to \mathbb{R}$ is introduced; it is the time required by the timed transition t, associated with a binding b, to complete the firing. After time function has been introduced, CPN models take the name of Coloured Timed Petri Nets (CTPNs).

In a CTPN model two kinds of *conflicts*[1] may arise: conflicts among transitions and conflicts among bindings. In this work the following solution has been adopted:

- In order to avoid the coupling between resolution of conflicts and duration of activities[2], transitions in conflict are supposed to be immediate except for timed transitions modelling a watchdog timer or immediate transitions modelling a failure detection; only in these special cases it is allowed that a timed transition, modelling a watchdog timer, can interrupt another timed transition or that an immediate transition, modelling an error detection, can be in conflict with a timed transition.
- In general, in a PN model there is no need to remove tokens from a place in the same order as they were added. However, in this case, when tokens contained in a place represent Stock Units (as explained afterwards), it is important to preserve the incoming order for their removing.

[1] Two or more transitions are in conflict if they have common input places
[2] When a timed activity is associated with transitions, a conflict resolution policy may be a race between conflicting transitions, that has no sense in the presence of a physical plant

Therefore each place is assumed as being a queuing place with a FIFO policy. Thus, no conflict among bindings may arise.

A net place is denoted by $P\{x,y\}$, where x denotes the subsystem of the model and y denotes the index of the place inside the sub-model. *Fusion places*, *source places* and *sink places* are denoted by $Pf\{x\}$, $Psr\{x\}$, $Psk\{x\}$ respectively and are all drawn by shaded circles. The following expression functions, used in the arc expressions and guard functions, are defined:

$pr(p,c_1,c_2,\ldots,c_k)$: for each token $<\tau>$ in the place p the function builds a new token formed only by the specified components $c_1,c_2,\ldots,c_k$ of the token $<\tau>$;

$conc(<\tau_1>,<\tau_2>,\ldots,<\tau_k>)$: the function builds the new token $(<\tau_1>,<\tau_2>,\ldots,<\tau_k>)$ formed by the concatenation of the specified components;

id: this function selects all the components of a token

SUMMARY OF CHAPTER 9

CHAPTER 10

SIMULATING COLLISION PROBABILITIES OF LANDING AIRPLANES

The first assumption that we will make here, in order to make our model the most simple as possible, is that the airport that we are going to analyze does not have a control tower: thus severely limiting its capacity during poor visibility conditions.

We will consider a system in which airplanes self-separate, so they are able to land at higher capacities without a Control Tower. A further step in our analysis (in the following chapter) will be, of course, putting some ATC control inside the system and see how it works, especially if and how safety increases.

What we would like to try to evaluate with this simulation is safety of such a system. Safety is a difficult metric to measure and predict, because accidents are very rare. Even computer simulation can be slow because of the long time to observe any accidents. In this chapter, we apply one methodology that has already shown being successful in assessing aviation safety through simulation. In particular, we will estimate the probability of collisions on the runway in poor visibility.

The methodology [PN-11] that we are going to use here has been already successfully used in safety analysis and is called TOPAZ (Traffic Organizer and Perturbation AnalyZer) modelling methodology, developed at the National Aerospace Laboratory NLR, the Netherlands (Blom, Bakker, et al. 2001).

This methodology provides a two-step framework for assessing safety:

- The first step qualitatively assesses safety by identifying hazards relevant to the scenario in question.
- The second step quantitatively estimates safety through simulation.

By identifying critical hazards in the first step, it is possible to create a simulation model which only samples the operational space where collisions are likely. Analytical models can, of course, also be used to further improve the efficiency of simulation.

Without a Control Tower, capacities in *Instrument Meteorological Conditions* (IMC) can be as low as three landings per hour, depending on the proximity of the airport to nearby radar coverage. We consider a proposed system where a nearby, supporting controller is responsible for the initial separation of airplanes entering the airspace near the airport (that is, leaving radar coverage). Of course, the pilots are still responsible for self-separation after that.

We specifically investigate the probability of a collision on the runway in this system [PN-12]. This analysis is a step in determining the potential capacity at such non-towered airports and, at the same time we want to understand which kind of level of safety is respected.

Among some existing models for estimating collision risk, we will review, in particular, the Reich collision model, which has been used extensively in literature for collision estimation purposes and, of course, into the TOPAZ modelling methodology.

1 COLLISION PREDICTION MODELS

This section provides background on analytical, quantitative models that have been used to estimate collision and conflict probabilities. We particularly focus on the Reich collision model, since it has been widely used all over Collision Risk models already present in literature and is part of the TOPAZ methodology that we will use it in our simulation analysis.

The simplest class of collision models are *intersection-type* models. In these models, one assumes that planes fly along pre-determined, crossing routes, generally at constant velocities. Under these assumptions, the probability of a collision at the crossing point can be computed from the arrival rates of airplanes along each path, their velocities, and the airplane geometries. For examples of such models, see Siddiqee (1973), Geisinger (1985), and Barnett (2000).

A similar class of models are *geometric conflict* models (e.g., Paielli and Erzberger 1997, 1999, Irving 2002). In these models, the velocities of the airplanes are fixed as before, but their initial positions (in three dimensions) are random. Then, based on extrapolating forward in time, it is possible to geometrically describe the set of initial locations that eventually lead to a conflict between two airplanes. (A conflict occurs when two airplanes are within, say, 5 NM of each other). Integrating the probability density of the initial positions over the conflict region gives the probability of a conflict. These models generally assume level flight with constant velocities (see Paielli and Erzberger 1999 for a generalization to non-level flight).

1.1 THE REICH COLLISION MODEL

A slightly more complex model is the Reich model. This model assumes random deviations in both the position and velocity components. It was originally developed to estimate collision risk for oceanic travel over the North Atlantic and to determine the appropriate spacing of flight paths (Reich 1966 [CRM-1], [CRM-2] and [CRM-3]).

To describe the model, we first start with some notation. Let $\vec{r}_1(t)$ and $\vec{v}_1(t)$ be the position and velocity vectors of one airplane at time t; let $\vec{r}_2(t)$ and $\vec{v}_2$ be similarly defined for a second airplane. To simplify notation, we drop the subscript t, keeping in mind that these and related quantities have an implicit dependence on time.

Let $\vec{r} \equiv (r_x, r_y, r_z) \equiv \vec{r}_1 - \vec{r}_2$ and $\vec{v} \equiv (v_x, v_y, v_z) \equiv \vec{v}_1 - \vec{v}_2$ be the relative position and velocity vectors of the two airplanes.

Now, $\vec{r}$ traces out a path in time. If $\vec{r}$ gets too small, there is a collision between the airplanes. Therefore, mathematically, if $\vec{r} = (0,0,0)$, each airplane's centre of mass is at the same point in space.

The Reich model assumes that each airplane is shaped like a box with dimensions:

- s_x (along-track length),
- s_y (across-track width),
- s_z (vertical height).

Under these assumptions, two airplanes are touching when, for example, one airplane is in front of the other by a distance s_x, or when one airplane is behind the other by a distance s_x (that is, whenever $-s_x \leq r_x \leq s_x$). More generally, therefore, a collision occurs along any direction whenever $\vec{r}$ passes through the $2s_x \times 2s_y \times 2s_z$ box, centred at the origin.
Such an event can be called an incrossing; mathematically, there may be multiple incrossings. In reality, this would correspond to only one collision. Thus, the probability of an incrossing is an upper bound on the probability of a collision.

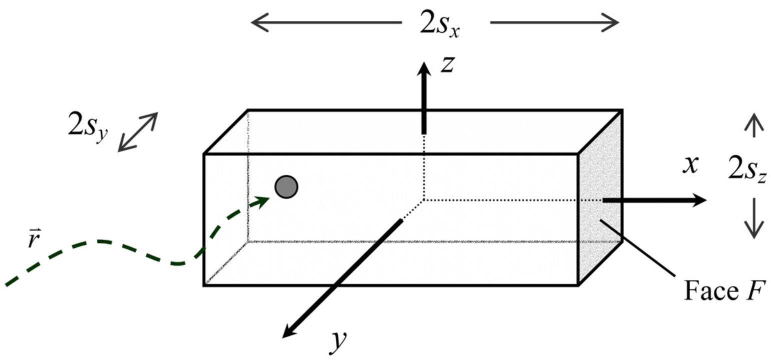


Figure 1 – Geometric representation of the Reich collision model [PN-11]

Hence, the probability of an incrossing depends on the joint probability density function (PDF):

$$f(\vec{r},\vec{v}) \equiv f(r_x, r_y, r_z, v_x, v_y, v_z) \qquad \text{Eq. 1}$$

The Reich model makes the following assumptions on this PDF:

- The density is independent in the x, y, and z dimensions. That is:

$$f(\vec{r},\vec{v}) \equiv f(r_x, r_y, r_z, v_x, v_y, v_z) = f_x(r_x, v_x) f_y(r_y, v_y) f_z(r_z, v_z) \qquad \text{Eq. 2}$$

where f_x, f_y and f_z are marginal densities of $f(\vec{r},\vec{v})$.

- The density is constant in $\vec{r}$ over the dimensions of the aircraft. That is:

$$f_x(r_x, v_x) = f_x(0, v_x) \text{ when } |r_x \leq s_x| \qquad \text{Eq. 3}$$

and similarly for the other dimensions. This is reasonable since the density is not expected to vary much over a distance as small as the dimensions of an airplane. However, if one changes the interpretation of s_x, s_y and s_z to represent a conflict box (instead of a collision box) with dimensions of several nautical miles, then this assumption won't be valid anymore.

• The density is independent on the position and velocity components (this is true for all the three dimensions). That is:

$$\begin{cases} f_x(r_x, v_x) = f_{r,x}(r_x) f_{v,x}(v_x) \\ f_y(r_y, v_y) = f_{r,y}(r_y) f_{v,y}(v_y) \\ f_z(r_z, v_z) = f_{r,z}(r_z) f_{v,z}(v_z) \end{cases} \quad \textbf{Eq. 4}$$

This last assumption is not in Reich's original assumptions; however, the expression of $\varphi(t)$ below , which is frequently quoted from Reich's paper (e.g., Hazelrigg and Busch 1986, Bakker and Blom 1993) requires this assumption.
As a matter of fact, the three assumptions above mentioned imply:

$$f(\vec{r}, \vec{v}) \equiv f(r_x, r_y, r_z, v_x, v_y, v_z) = f_{r,x}(0) f_{r,y}(0) f_{r,z}(0) f_{v,x}(v_x) f_{v,y}(v_y) f_{v,z}(v_z) \quad \textbf{Eq. 5}$$

whenever $\vec{r}$ is on or within the boundary of the collision box.

In addition, Reich (1966) assumes:
• Planes travel along parallel tracks without making turns. Thus, the orientation of the collision box does not change.

• All planes have the same geometric shape.

• There is no evasive manoeuvring by the pilot or intervention by the controller.

Under the above assumptions, the total incrossing rate through all sides of the box is:

$$\varphi(t) = f_{\vec{r}}(0,0,0) \sum_{i=1}^{3} A_i E|v_i| \quad \textbf{Eq. 6}$$

where the subscript i denotes the three dimensions x, y, and z, $f_{\vec{r}}$ is the marginal density of $f(\vec{r}, \vec{v})$, and A_i is the area of the face perpendicular to dimension i. Also, $f_{\vec{r}}$ and $E|v_i|$ are implicit functions of time. Since the previous equation gives the incrossing rate at time t, the total expected number of incrossings over the time interval $[a, b]$ is:

$$E = \int_a^b \varphi(t) dt \quad \textbf{Eq. 7}$$

If there are more than two airplanes, then the previous equation must be evaluated for every possible airplane pair and then summed to get the total number of expected incrossings among all airplanes.

An advantage of the Reich collision model is that it accounts for all possible directions of the aircraft. Some of the other models, by discounting the vertical dimension, only account for collisions through the four sides of the box.

1.2 GENERALIZED REICH MODEL

Some of the assumptions in the Reich model are quite restrictive – in particular assumptions the first and the third, which state that all components of $(\vec{r},\vec{v})$ are mutually independent.

In particular, velocity in one direction usually depends on velocity in another direction (for example, the ascent rate generally depends on the along-track rate).

Removing these two assumptions (first and third), Bakker and Blom (1993) derived a generalized Reich collision model. In particular, the incrossing rate through a single face (face F in Figure 1) and its opposing face is:

$$\varphi_x(t) = \int_{-s_y}^{s_y}\int_{-s_z}^{s_z}\int_{-\infty}^{0} -v_x f(v_x, r_x = s_x, r_y, r_z)\, dv_x dr_z dr_y + \\ + \int_{-s_y}^{s_y}\int_{-s_z}^{s_z}\int_{0}^{+\infty} v_x f(v_x, r_x = -s_x, r_y, r_z)\, dv_x dr_z dr_y \qquad \textbf{Eq. 8}$$

where $f(v_x, r_x, r_y, r_z)$ is a marginal distribution of $f(\vec{r},\vec{v})$. Thus, the total incrossing rate through all faces is:

$$\varphi(t) = \varphi_x(t) + \varphi_y(t) + \varphi_z(t) \qquad \textbf{Eq. 9}$$

where $\varphi_y(t)$ and $\varphi_z(t)$ are defined similarly to $\varphi_x(t)$. Although $\varphi_x(t)$ is difficult to evaluate numerically, Blom and Bakker (2002) show that if $f(\vec{r},\vec{v})$ is a mixture of Gaussian distributions, then evaluation of the integral is much easier.

Blom and Bakker also argue that the fourth assumption can be removed by assuming that $s_x = s_y$.

In other words, we can state that, if the length of a plane is approximately the same as its wing span, the bounding box does not change when the plane turns. That is, the collision box keeps a fixed orientation regardless of the orientation of the two airplanes.

2 SIMULATION METHODS: DYNAMICALLY COLOURED PETRI NETS

In the previous paragraph, we discussed several analytical models to estimate collision risk. A central problem with these models is that they only apply to simple scenarios. For example, they generally apply to level flight and do not consider corrective actions by pilots or controllers.

The purpose of this chapter is to consider the scenario of predicting plane-to-plane collisions on the runway [PN-11] and [PN-12]. While relatively simple, this scenario has several complications which make the previous models by themselves insufficient.

In particular, the analytical models do not account for corrective actions by the pilots to avoid a collision, nor do they account for equipment failures. But, the models are still quite useful as an enhancement to simulation, as we are going to discuss below.

We are now discussing the simulation method used in several TOPAZ models to evaluate collision risk (e.g., Blom, Klompstra, and Bakker 2003). The method uses *Dynamically Coloured Petri Nets* (DCPNs) as a framework for simulation.

To improve efficiency, the method also makes use of the generalized Reich collision model, discussed in the previous section.

The next Figure 2 shows the basic idea. Instead of returning the number of collisions, the simulation returns the probability density $f(\vec{r},\vec{v})$ (the distribution for the relative position and velocity of an airplane pair). Given $f(\vec{r},\vec{v})$, the probability of a collision can be calculated using the generalized Reich model, (also see Blom and Bakker 2002).

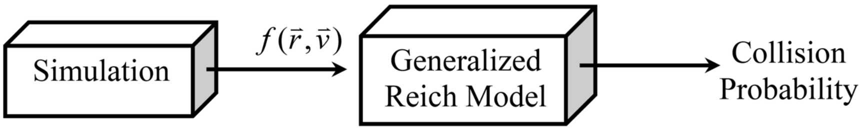


Figure 2 – Basic mechanism for combining simulation with an analytical model [PN-11]

DCPNs are a very versatile instrument: they can simulate a wide variety of system dynamics including flight dynamics and controller-pilot interactions. To illustrate the use of DCPNs in the context of aviation safety, we give here a simple example.

Figure 3 gives the DCPN used to simulate the small airport application discussed in the next section. For further details on DCPNs, see the chapter of this thesis regarding Petri Nets.

For an example of DCPNs applied to simultaneous approaches of two airplanes on two converging runways, see in particular Blom, Klompstra, and Bakker (2003).

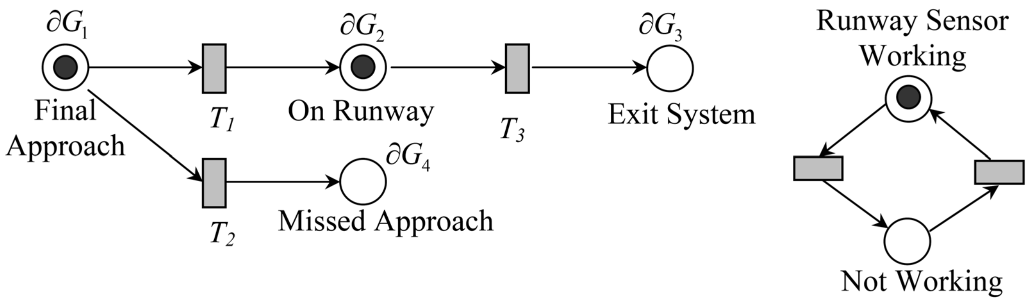


Figure 3 – Simple Petri net with possible missed approach [PN-11]

The previous Figure 1 shows two separate (but not independent) Petri nets:

- The *right Petri Net* models whether or not a runway sensor is working. The token (solid dot) indicates the current state. The transitions (gray boxes) represent events which trigger the token to move from one state to another. In this Petri net, we suppose that times between transitions are independent random variables. For example (and this is a very nice feature of Petri Nets), if they follow an exponential distribution, then the right Petri net is a continuous time Markov chain. We can also use similar Petri nets to give a very simplified model of human cognitive states - for example, to model when a pilot is relaxed, busy, or frantic. For an example of such an application in air traffic management, see Blom, Daams, and Nijhuis (2001). Also see Hollnagel (1993) for a more complete discussion of human reliability in the context of human internal states.

- The *left Petri Net* is more mathematically complex. The tokens correspond to airplanes and the places (open circles) correspond to phases of flight. In the Figure 3, there is one airplane on the runway and one airplane in the final approach. Associated with each token is a six-dimensional vector (not drawn) giving the position and velocity of that airplane in three dimensional space – from here the net is called coloured, since the token brings within itself some "coloured" information. Associated with each phase of flight there is therefore a set of six differential equations which govern each airplane's trajectory in time when in that phase.

In the Figure 3, ∂G_1 represents *a set of six differential equations* that govern airplane trajectories in the final approach. The differential equations can be stochastic, allowing for random perturbations due to wind or pilot/navigational error.

For example, two of the differential equations used in the final approach phase of the complete model (place p12 in the next Figure 5) are:

$$\begin{cases} dp_x = v_x dt \\ dv_x = (-bv_x - kp_x)dt + \sigma d\omega \end{cases}$$ **Eq. 10**

where p_x and v_x are the position and velocity of the airplane in the x (across-track) dimension; $d\omega$ represents Brownian motion; k, b, and σ are constants.

The previous equations represent a pilot who is trying to keep the airplane centred along the runway centreline (at $x = 0$), in the presence of some disturbances such as, for example, the wind. The first two terms in the second equation represent the pilot controls and the last term represents wind (that is, in a small time interval of length Δt, the perturbation due to wind is a normal random variable with mean 0 and variance $\sigma^2 \Delta t$).

There would also be four other equations corresponding to the y and z directions. For an introduction to stochastic differential equations, see Oksendal (1992).

We can also link the two Petri nets shown in the previous Figure 3. For example, to model the functionality of the runway sensor, we define the transition T1 in the left Petri net to trigger when:

- The airplane that is currently in final approach has just crossed the runway threshold (based on its position vector), ***and***
- - The runway is not occupied ***or***
  - The runway is occupied, but the runway sensor is not working.

Thus, the state of the *right Petri Net* affects the trigger events of the *left Petri Net*. This logic can also be drawn using standard Petri net notation. However, we do not do this to avoid clutter.

We can also create a second type of link between the two Petri nets: making the differential equations on the left a function of the Petri net state on the right.
An example would be using the right Petri net to model the state of an Instrument Landing System (ILS for a detailed description on that, see chapter regarding fundamentals of Navigation Systems).

When e.g. the ILS is not working, pilots deviate more from the glide path. Thus, two sets of differential equations are needed to model the airplane trajectory, depending on whether or not the ILS is working, and so on.

3 APPLICATION: NON-TOWERED AIRPORT (I.E. NO ATCO INTERVENTION)

This section describes a scenario involving a small airport with no Control Tower or let's say, where the Control Tower is not working at all. The goal is of double nature:

- From one side, is to increase the capacity of such airports in IMC, keeping constant the level of safety,
- From another, is to keep the actual capacity improving the level of safety.

Currently, *procedural separation rules* (see the chapters of this thesis regarding Airport Operations and Air Traffic Control) dictate that only one airplane is allowed into the airspace near the airport at one time (in IMC). This should "guarantee" that two airplanes are not simultaneously flying near the airport outside of radar coverage.

However, this can yield capacities as low as three operations per hour depending on nearby terrain and proximity to radar coverage.

One solution that has been proposed is to equip airplanes with a self-separation capability. Nearby controllers have responsibility for the initial separation of airplanes into the local airspace, but after that, airplanes must separate themselves. Referring to the next Figure 4, we will consider the following concept of operations:

- Since the local airport does not have a Control Tower and is outside radar coverage, en-route arrival-departure traffic is controlled by a supporting air traffic controller (ATCo) at a TRACON or ARTCC.
- The supporting ATCo meters airplanes into the local airspace through one of two approach legs. The two approach paths combine to form a "T". The ATCo is responsible for the initial separation of the airplanes.
- Once an airplane enters the airspace, the pilot is responsible for maintaining separation among other airplanes. At this point, the airplane is outside radar control and therefore is outside the control of the ATCo.
- A small terminal sensor located on the ground at the airport provides radar-like coverage for the local airspace. The sensor fusion system transmits airplane positions to all airplanes in the local airspace via a ground-to-air data link.
- A Cockpit Display of Traffic Information (CDTI) displays the locations of these airplanes to the pilots. Pilots use the display to maintain separation.
- As a deterrent to a collision on the runway, an infrared sensor on the ground detects the presence of an airplane on the runway. If the runway is occupied, the landing pilot receives a warning in the cockpit.

Three locations where plane-to-plane collisions are most likely to occur are (see next Figure 4):

1. Intersection of the "T". The controller has not properly separated incoming airplanes.
2. Final approach. A faster airplane overtakes and collides with a slower airplane.
3. Runway. One airplane fails to exit the runway before the approaching airplane lands.

The last collision type is probably the most common. This is because there are only two degrees of freedom on the runway, but three degrees of freedom in the air. Since the last collision type presents the largest safety issue, this chapter will concentrate only the focus on that scenario.

We will also discuss in the next section how the model can be used to evaluate the risk of the other collision types.

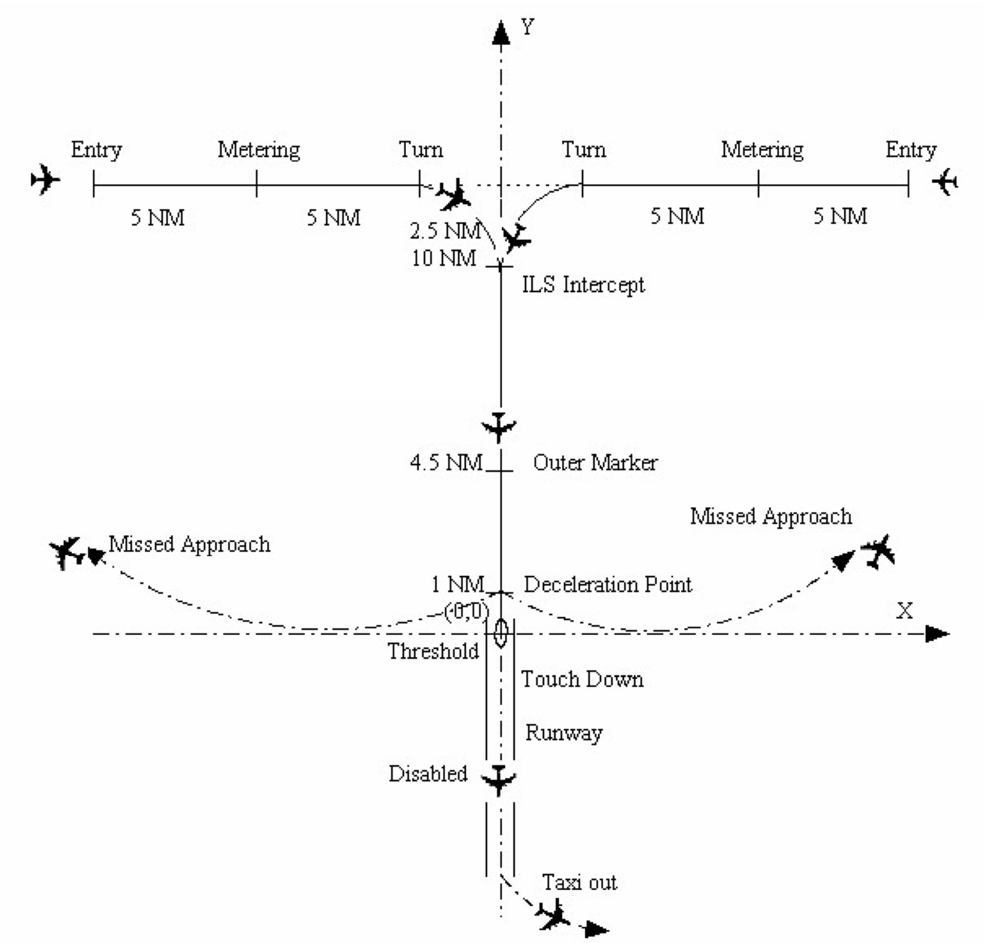


Figure 4 – Landing Airplanes at small, non-controlled airport [PN-11]

4 ANALYSIS AND RESULTS FOR THE CASE OF AIRPLANES LANDING AT SMALL, NON-CONTROLLED AIRPORT.

This section describes the analysis to estimate the probability of a collision on the runway. A first step in safety assessment is an identification of hazards (see Blom, Bakker, et al. 2001, and also "FAA System Safety Handbook, Appendix F" and, of course, MIL-STD-1629A).

Since the focus of this work is on simulation, we do not elaborate on this step, but simply list examples of hazards which can lead to a collision. Of course, the following is not an exhaustive list.

1. An airplane lands and becomes disabled, so it cannot exit the runway (blown tire, partial crash landing, etc.).
2. An airplane lands, and the pilot becomes disoriented; instead of exiting to the taxi-way, the pilot stays on the runway while going to the gate.
3. The pilot stops on the runway and does not immediately pull off.
4. The runway sensor fails.
5. The communication link between this sensor and other airborne airplanes fails.
6. The pilot fails to notice a warning from the runway sensor.
7. The pilot notices the warning, but chooses to ignore it.

8. The pilot is distracted and does not see another airplane on the runway. This could be due to poor visibility or because the pilot is concentrating on landing his or her own airplane.
9. The runway is slick, so a landing airplane cannot decelerate as quickly as normal.

Figure 5 gives the Petri net structure which models the scenario described previously. The model incorporates all of the hazards listed above. In some cases, we have grouped several hazards into a single model element. For example, the sub-net in the lower left-hand corner of the Figure 5 models hazards 1-3. The "Runway Sensor" sub-net models hazards 4 and 5. The "Pilot's Awareness" sub-net models hazards 6-8.

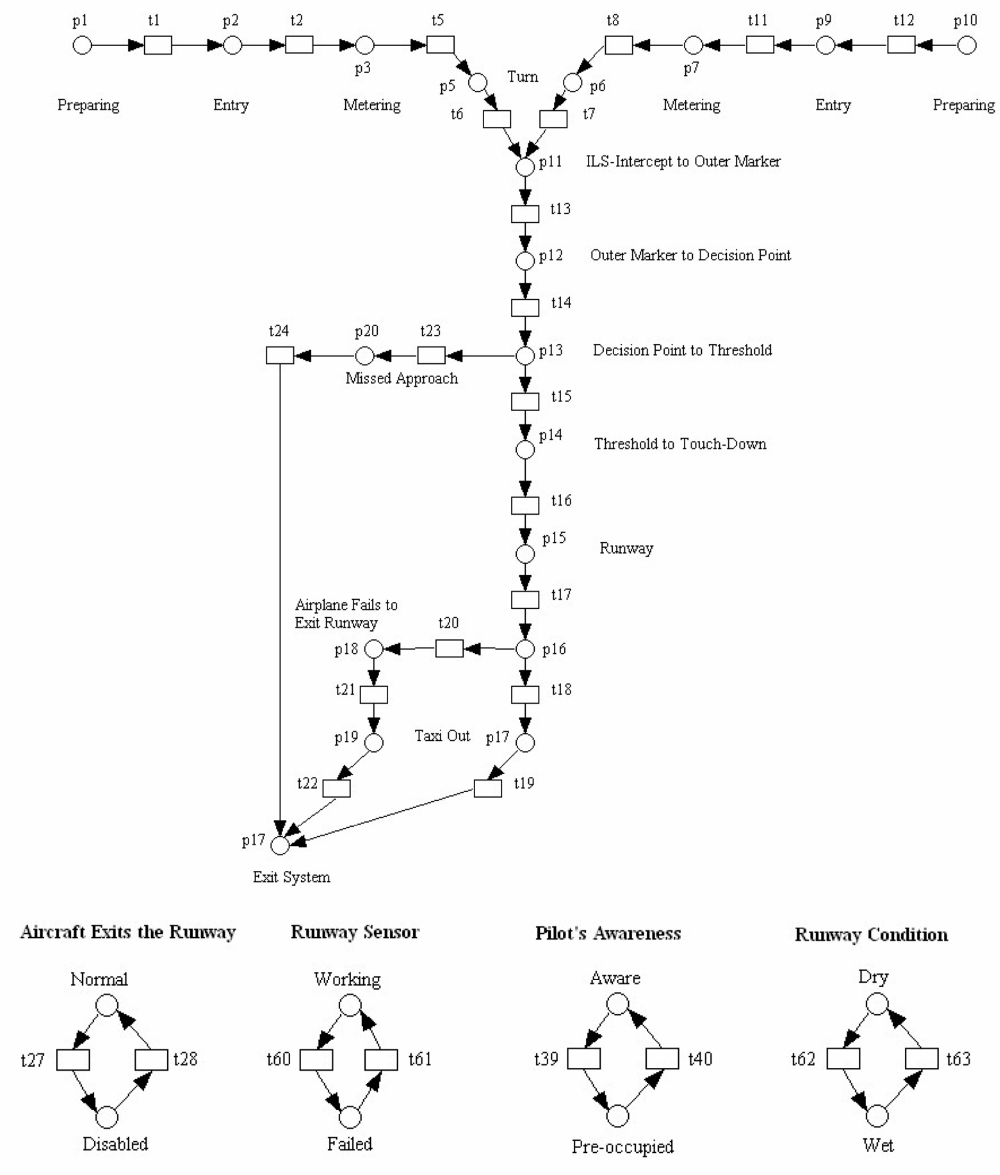


Figure 5 – Petri net diagram of non-towered airport [PN-11]

The following Table 1 gives probability estimates for hazards in the model. Since this is a proposed system, little data are available for these parameters, so at this point, they are estimates.
Later in this section we give sensitivity analysis on these parameters to understand which parameters most affect the system safety.

A	Landing airplane does not exit runway (Hazards 1-3)	$Pr(A) = 5 \times 10^{-3}$
B	Runway sensor is not working (Hazards 4-5)	$Pr(B) = 1 \times 10^{-4}$
C	Pilot fails to notice warning from sensor (Hazards 6-7)	$Pr(C) = 1 \times 10^{-3}$
D	Pilot fails to see airplane on runway (Hazard 8)	$Pr(D) = 1 \times 10^{-2}$
E	Runway is slick (Hazard 9)	$Pr(E) = 5 \times 10^{-2}$

Table 1 – Parameters used in the model [PN-11]

Just to further clarify the simulation logic, if airplane i reaches the decision point (indicated as A in the next Figure 6) and airplane j is disabled on the runway, then airplane i will land if the following happens (otherwise i will fly a missed approach):

- Pilot i does not receive or react to warning from runway sensor (due to any one of hazards 4-7), and
- Pilot i fails to visually see airplane j while between points A and B (hazard 8).

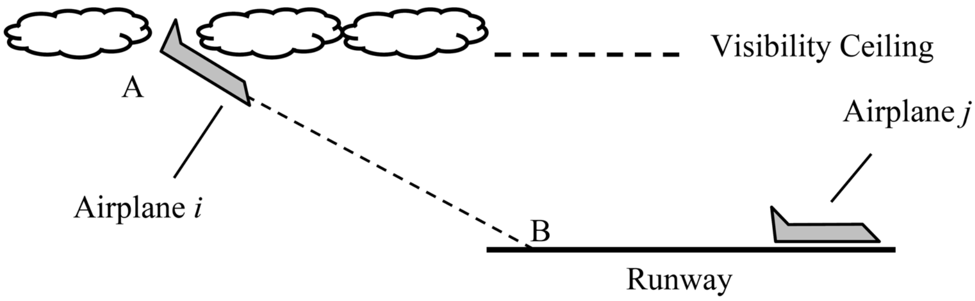


Figure 6 – Airplanes landing on runway [PN-11]

Also in this case, of course, Stochastic Differential equations govern the airplane trajectory for each phase of flight.

In general, these are second order response models where the pilot tries to maintain a target speed (different for each airplane) along a constant heading or glide slope, subject to random perturbations due to wind or pilot control.

We will also make the following assumptions:

- An airplane which fails to exit the runway remains on the runway for a random amount of time (following an exponential distribution).
- At touchdown, the pilot immediately sees a disabled airplane on the runway. At this point, the pilot decelerates the plane at the maximum possible rate.
- The runway may be wet (hazard 9) or dry. If the runway is wet, the maximum deceleration rate is less than the normal deceleration rate.

In addition, we will make the following assumptions regarding weather, approach paths, and airplane types:

- Weather conditions are IMC.
- The visibility ceiling is 250 feet.
- The Runway Visual Range (RVR) is ¾ mile.

- Pilots fly a 3 degree approach path.
- All airplanes have the same flight characteristics and onboard equipment.
- The airport has a single runway with a separate taxi-way. The runway is 5,000 feet in length.

Following the methods suggested by Blom, Klompstra, et. al (2003), we define τ_{ij} to be a time such that there is zero probability of a collision between airplane i and j prior to time τ_{ij}.

Since we are only interested in collisions on the runway, we can define:

$$\tau_{ij} = \text{time airplane } i \text{ lands (at B in Figure 6) while airplane } j \text{ is on the runway} \qquad \textbf{Eq. 11}$$

We also define $\tau_{ij} = \infty$ if j exits the runway before i lands, or i flies a missed approach, or i lands before j. We will define too:

$$B_{ij} = \begin{cases} 1 & \text{if airplane i collides with airplane j} \\ 0 & \text{otherwise} \end{cases} \qquad \textbf{Eq. 12}$$

Without loss of generality, we assume that airplanes are indexed in order of their arrival, and we only consider collisions when $i > j$ (so, $B_{ij} = 0$ for $i \leq j$). Since (by assumption) there is zero probability that airplane i collides with airplane j prior to τ_{ij}, the total probability that airplane i collides with airplane j is:

$$\Pr(B_{ij} = 1) = \Pr(B_{ij} = 1 \mid \tau_{ij} < \infty)\Pr(\tau_{ij} < \infty) \qquad \textbf{Eq. 13}$$

To compute the collision rate, we compute the right-hand side of the previous equation. The total expected number of collisions $E(N)$ is then:

$$E(N) = \sum_{i>j} \Pr(B_{ij} = 1) \qquad \textbf{Eq. 14}$$

First, to evaluate $\Pr(\tau_{ij} < \infty)$, we run the simulation (Petri net in Figure 5) and count the number of occurrences that $\tau_{ij} < \infty$ (that is, the number of times that some airplane i lands while another airplane j is disabled on the runway).

To evaluate $\Pr(B_{ij} = 1 \mid \tau_{ij} < \infty)$, we first observe:

$$\Pr(B_{ij} = 1 \mid \tau_{ij} < \infty) = \int_{\tau_{ij}}^{\infty} \varphi_{ij}(t \mid \tau_{ij} < \infty)dt \qquad \textbf{Eq. 15}$$

We assume here that there is a one-to-one correspondence between incrossings and collisions. This is reasonable, since airplane j is stationary. Since airplane i is always moving forwards, there is no possibility of more than one incrossing.
To compute this, we have run the simulation with parameters adjusted to achieve an artificially high number of instances where $\tau_{ij} < \infty$. In other words, we suppose that there is a very high probability that:

(a) airplanes become disabled on the runway, and
(b) that pilots and sensor systems fail to observe the disabled airplane.

Each time that $\tau_{ij} < \infty$, the simulation collects data on $f_t(\vec{r},\vec{v})$, the relative position and velocity of the landing airplane with respect to the disabled airplane. The subscript t denotes the dependence of this density on time, where t represents the time after τ_{ij}. (In other words, $t = 0$ refers to the time when airplane i lands, while j is still on the runway). Using the generalized Reich model, we compute $\varphi_{ij}(t \mid \tau_{ij} < \infty)$ using the equations described before. Then, we integrate as in the previous equations. Next Figure 7 shows the results of the preceding analysis and simulation. The figure shows the collision probability on the runway as a function of the arrival rate. The confidence intervals (95%) show potential errors due to the finite number of simulation replications [PN-11].

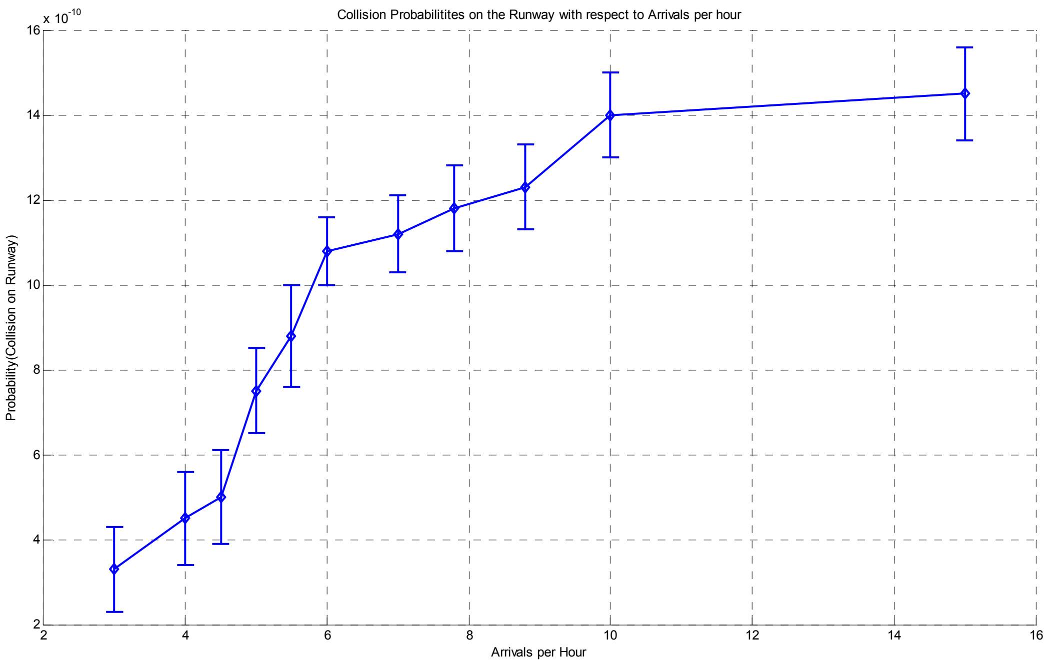


Figure 7 – Collision Probabilities on the Runway [PN-11]

Other sources of errors in the safety estimates include inaccuracies in parameter estimates, omissions in modelling safety hazards, and simplifications in modelling system dynamics. Since this is a proposed system and little data are available to populate some key parameters, the estimates in Figure 7, while calculated in an absolute sense, should be viewed in a relative sense. That is, the trend is more reliable than the absolute numbers.

The shape of the curve in Figure 7 is explained from the terms in $\Pr(B_{ij} = 1) = \Pr(B_{ij} = 1 \mid \tau_{ij} < \infty)\Pr(\tau_{ij} < \infty)$. First, $\Pr(B_{ij} = 1) = \Pr(B_{ij} = 1 \mid \tau_{ij} < \infty)$ is relatively constant as a function of the arrival rate. In other words, given airplane i lands while airplane j is disabled on the runway, the probability of a collision does not depend much on the background arrival rate. $\Pr(\tau_{ij} < \infty)$, on the other hand, does depend on the arrival rate. As the arrival rate gets large, the probability of a trailing airplane to arrive (gets to point A in Figure 6) while another airplane is on the runway levels off and goes to 1, which contributes to the basic shape in Figure 7.

While in this chapter we evaluated collisions on the runway, the same model can also be used to evaluate collisions at the top of the "T" and during the approach, and on the airport surface. The key output of the simulation is the relative separation and relative velocity PDF $f(\vec{r},\vec{v})$ of adjacent airplanes as a function of time. For example, to estimate the probability of a collision during the approach, we can collect data from the simulation to estimate $f(\vec{r},\vec{v})$ when airplanes are in this phase of flight. We then enter the resulting distribution into the generalized Reich collision model. A collision at the intersection presents a new challenge, since the airplanes are turning. Although the Petri net model in this chapter provides equations of motion for the turn, we have observed that the distribution of $f(\vec{r},\vec{v})$ in this area is not well described by a normal distribution. Since the efficient methods for integrating the equations in Blom and Bakker (2002) require a mixture of normals, there are further numerical issues with doing this computation.

In [PN-11] two techniques to increase the efficiency of simulation (similar to techniques used in Blom, Klompstra, and Bakker 2003) were used:
(a) Conditioning on hazardous events (so the probability of a simulated collision is higher) and
(b) the generalized Reich collision model to enhance the results of simulation.

Performance using these techniques was good, since we were able to get statistically significant results without excessive computer time (several hours on a PC).

Of course, simulation output can only be as good as the model generating the output. In safety modelling, it is impossible to account for all hazards that might lead to a collision. For example, in this analysis we won't consider hazards related to improper initial separations due to controller errors. This omission biases the collision probability estimate to be lower than actual. On the other hand, some parameter estimates are likely conservative, biasing the estimate in the other direction. Thus, conclusions based on the absolute results are tentative.

Nevertheless, we feel that such analysis has the following benefits. First, the spirit is to reveal general trends. In particular, the analysis reveals a levelling off of the runway collision risk as a function of arrival rate, contrary to an expected quadratic or exponential growth. Second, the analysis identifies critical parameters that have the highest impact on safety. In this case, pilot awareness is more critical than the reliability of the runway sensing device.
Although initially there are little data to populate the model, we can use such a model to intelligently guide the design of data and flight tests. That is, a new system must be proven to be safe, and a critical step in this process is a demonstration of the system with real flight tests. The model can help to determine what data are most important to collect and how much data should be collected in such tests. A general iterative procedure is: First, build a preliminary model to identify parameters that most affect safety. Then, design experiments to collect an appropriate amount of data on those parameters. Then, revise the model using the new data. These steps can be repeated until a sufficiently accurate safety estimate is achieved.

SUMMARY OF CHAPTER 10

CHAPTER 11

AIRPORT TRAFFIC FLOW USING TIMED COLOURED PETRI NETS

This chapter is the core of the thesis work. It describes a *Timed deterministic Coloured Petri Net* (TCPN) model of a single runway airport, that is capable of analyzing the effects of failures of A-SMGCS equipment on the capacity of the runway and on the timing of a given scheduled traffic, with respect to the Safety Level of the airport.

The developed TCPN model is governed by elementary Air Traffic Control (ATC) principles (one aircraft at a time on runway, arrivals priority on departures, etc.), and takes into consideration the effect of different Aircraft Speeds, Low Visibility Procedures, and so on.

1 STATECHARTS AND STATEFLOW

The literature [SC-1], [SC-2]on software and systems engineering is almost unanimous in recognizing the existence of a major problem in the specification and design of large and complex reactive systems. A *reactive system* in contrast with a *transformational system*, is characterized by being, to a large extent, event-driven, continuously having to react to external and internal stimuli. Examples include telephones, automobiles, communication networks, computer operating systems, missile and avionics systems, and the man-machine interface of many kinds of ordinary software. The problem is rooted in the difficulty of describing *reactive behaviour* in ways that are clear and realistic, and at the same time formal and rigorous, sufficiently so to be amenable to detailed computerized simulation. The behaviour of a reactive system is really the set of allowed sequences of input and output events, conditions, and actions, perhaps with some additional information such as timing constraints. What makes the problem especially acute is the fact that a set of sequences (usually a very large and complex one) does not seem to lend itself naturally to 'friendly' gradual, level-by-level descriptions, that would fit nicely into a human being's frame of mind.

For *transformational systems* (e.g., many kinds of data-processing systems) one really has to specify a transformation, or function, so that an input/output relation is usually sufficient. While transformational systems can also be highly complex, there are several excellent methods that allow one to decompose the system's transformational behaviour into ever-smaller parts in ways that are both coherent and rigorous. Many of these approaches are supported by languages and implemented tools that perform very well in practice.

Several people [SC-4] are of the opinion that for *reactive systems*, which present the more difficult cases, this problem has not yet been satisfactorily solved. Several important and promising approaches have been proposed. However, the general feeling is that many more improvements and developments are necessary.

Much of the literature [SC-5] also seems to be in agreement that states and events are a priori a rather natural medium for describing the dynamic behaviour of a complex system. A basic fragment of such a description is a state transition, which takes the general form "when event (Y occurs in state A, if condition C is true at the time, the system transfers to state B". Indeed, many of the informal exchanges concerning the dynamics of systems are of this nature; e.g., "when the plane is in cruise mode and switch x is thrown it enters navigate mode", or "when displaying the time, if button y is pressed the watch starts displaying the date". Finite state machines and their corresponding state-transition diagrams (or state diagrams for short) are the formal mechanism for collecting such fragments into a whole. State diagrams are simply directed graphs, with nodes denoting states, and arrows (labelled with the triggering events and guarding conditions) denoting transitions. Figure 1 shows a simple self-explanatory state diagram.

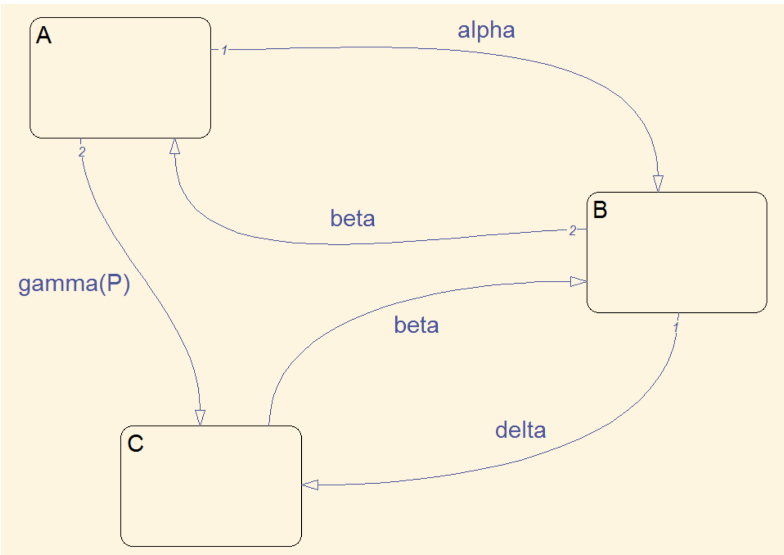


Figure 1 – Simple self-explanatory state diagram

However, it is also generally agreed that a complex system cannot be beneficially described in this naive fashion, because of the unmanageable, exponentially growing multitude of states, all of which have to be arranged in a 'flat' unstratified fashion, resulting in an unstructured, unrealistic, and chaotic state diagram. To be useful, a state/event approach must be *modular*, *hierarchical* and *well-structured*. It must also solve the exponential blow-up problem by somehow relaxing the requirement that all combinations of states have to be represented explicitly. A good state/event approach should also cater naturally for more general and flexible statements, such as:

(1) "in all airborne states, when yellow handle is pulled seat will be ejected",
(2) "gearbox change of state is independent of braking system",

(3) "when selection button is pressed enter selected mode",
(4) "display-mode consists of time-display, date-display and stopwatch-display".

Clause (1) calls for the ability to cluster states into a superstate, (2) introduces independence, or orthogonality, (3) hints at the need for more general transitions than the single event-labelled arrow, and (4) captures the refinement of states.

Statecharts constitute a visual formalism for describing states and transitions in a modular fashion, enabling *clustering*, orthogonality (i.e., *concurrency*) and refinement, and encouraging 'zoom' capabilities for moving easily back and forth between levels of abstraction. Technically speaking, the kernel of the approach is the extension of conventional state diagrams by AND/OR decomposition of states together with inter-level transitions, and a broadcast mechanism for communication between concurrent components. The two essential ideas enabling this extension are the provision for 'deep' descriptions and the notion of orthogonality. The approach is described solely in its diagrammatic terms, although the reader should be able to provide a textual, language-theoretic or algebraic equivalent if so desired. In a nutshell, one can say:

$$statecharts = state / diagrams + depth + orthogonality + broadcast / communication \quad \textbf{Eq. 1}$$

Stateflow is an interactive graphical design tool that works with Matlab Simulink to model and simulate *event-driven systems*, also called *reactive systems*. Event-driven systems transition from one operating mode to another in response to events and conditions. These systems are often used to model logic for dynamically controlling a physical device such as a fan, motor, or pump: in our case an entire Airport. Event-driven systems can be modelled as *finite-state machines*.

Finite-state machines represent operating modes as states. For example, a house fan can have states such as High, Medium, Low, and Off. To construct finite-state machines, Stateflow provides graphical objects that one can drag and drop from a design palette to create state-transition charts in which a series of transitions directs a flow of logic from one state to another. Stateflow also allows to add:

- Input and output data.
- Events for triggering Stateflow charts
- Actions and conditions, which can be attached to states and transitions to further define the behaviour of the Stateflow chart.

Stateflow allows to extend the capabilities of traditional state charts by:

- Adding *hierarchy* to charts
- Modelling *parallel* states
- Defining *functions* graphically, using flow diagrams; procedurally, using the Matlab language; and in tabular form, with truth tables
- Using *temporal logic* to schedule events
- Defining vector, matrix, and fixed-point data types

Furthermore, Stateflow performs simulation by generating a *C code implementation* of the Stateflow chart. The simulation code is generated from a simulation target. One can also generate portable C code from Stateflow charts automatically using Stateflow Coder. Stateflow Coder also works with Real-Time Workshop to generate C code for Simulink models that include Stateflow charts.

The fact that in this work we will make extensive use of Statecharts, implemented in Stateflow is based upon a number of theses:

- *Reactive* systems differ from *transformational* systems, and require different approaches to their specification.
- An essential element in the specification of reactive systems is the need for a clear and rigorous *behavioural description*, to serve as the backbone of development from requirements specification all the way to user documentation.
- *Statecharts* provide one possible fitting formalism for specifying *reactive behaviour*.
- The future lies in visual languages and methodologies that, with appropriate structuring elements, can exploit all the obvious advantages of graphical man-machine interaction.

As a matter of fact, we are convinced that people working with complex systems have for a long time appreciated the simplicity and appropriateness of the state/event approach but have lacked a formalism for it that possesses certain elementary properties (such as depth and modularity) that are provided by most programming languages and by many conventional approaches to the physical and functional aspects of system description. The lack of these, as well as the exponential blow-up syndrome and the inherent sequentiality of conventional state machines, seem to have hindered serious use of states and events in the design of really large systems.

We also believe that before long scientists and engineers will be sitting in front of graphical workstations with large (blackboard size?) displays of fantastic resolution, carrying out their everyday technical and scientific chores. It is quite fair to say that most existing visual description methods in computer science are predominantly intended as aids. The 'real' description of the object is usually given in some textual, algebraic form, and the picture is there only to help see things better, and to assist in comprehending the complexity involved. Here we are suggesting that visual formalism should be the name of the game; one uses statecharts as the formal description itself, with each graphical construct given a precise meaning. The language does not consist of linear combinations of icons or of one-level graphs, but of complex multi-level diagrams constructed in nontrivial ways from a few simple constructs. Textual representations of these visual objects can be given, but they are the aids (e.g., for users lacking graphical equipment, or for applications requiring textual reports), and not the other way around.

1.1 STATECHARTING PETRI NETS WITH STATEFLOW

Petri nets and statecharts are two popular visual formalisms for modelling systems that exhibit *concurrency*. As already stated in the related chapter, Petri nets were introduced in 1962 by C.A. Petri. Since the 1980's Petri nets found their way in practical applications like manufacturing, workflow modelling and performance analysis. On the other hand, Statecharts were introduced in 1987 by D. Harel [SC-3], [SC-4], for use in the structured analysis method Statemate. Soon after their appearance, statecharts were adopted in several object-oriented methods, and their successor Unified Modelling Language. Thus, both Petri nets and statecharts have found widespread use in both academia and industry. Both formalisms are supported by various tools, like CPN tools and GreatSPN for Petri nets, and Statemate and Stateflow for statecharts. UML tools are a special case, since they support not only statecharts but also activity diagrams, a Petri-net like notation. Tools supporting Petri nets are strongly focused on analysis of both functional and stochastic properties, while tools supporting statecharts are usually more focused on the software design process, for example offering the facility to generate code from a statechart.

Given this wide tool support, it is interesting to have well-defined translations between Petri nets and statecharts. Such translations could facilitate the exchange of models in different notations between tools, thus allowing designers to use the best of both worlds. Ideally, such translations are correct and structure preserving. A translation is correct if the original and translated model are equivalent.

concept	Petri Net	Statechart
state	place	node (state)
transition	transition	hyperedge (full compound transition)
active state	place containing a token	active node (active state)
global state	marking	configuration

Table 1 – Relating Petri Nets and Statecharts terminology

Correctness guarantees that operations done with a tool on a particular model are meaningful, even if the original model was expressed in another formalism. Preservation of structure ensures that the syntax of the original and translated model are alike, making it easier for designers to understand the translated model.

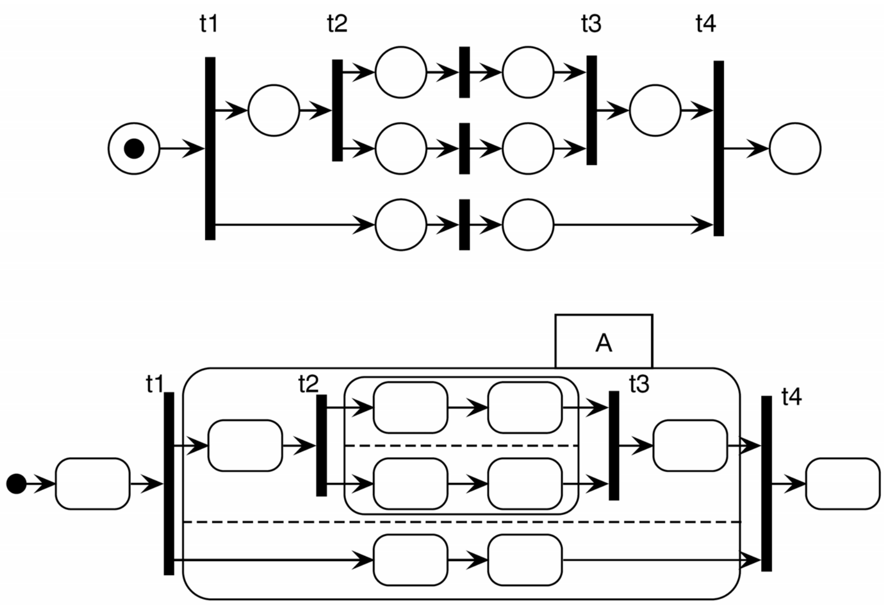


Figure 2 – Petri net with balanced forks and joins and equivalent statechart

The goal here was to define a translation from *Petri Nets* to *Statecharts*, (and not necessarily vice versa), that is correct and structure preserving. Before giving any details on this translation, it is useful to discuss the similarities and differences between Statecharts and Petri nets. Both formalisms are generalisations of *Finite State Machines*. Finite state machines model sequential processes as states connected by transitions. Mathematically, finite state machines are represented as directed graphs, consisting of nodes (representing states) and directed edges (representing transitions). The global state of such a machine at a certain point in time is always one single state, i.e., only one node of the graph is active at a time.

Both Petri nets and statecharts generalise finite state machines by using transitions that can enter and leave multiple states. Sources and targets of such transitions are concurrent: a transition can only be taken if all its sources are active in parallel, and if a transition is taken, all its targets become active in parallel.

Mathematically, a transition is represented as a hyperedge. Consequently, the underlying mathematical model is a *hypergraph* rather than a graph. The global state of a Petri net or Statechart is therefore distributed and typically consists of multiple states, i.e. multiple nodes of the hypergraph can be simultaneously active.

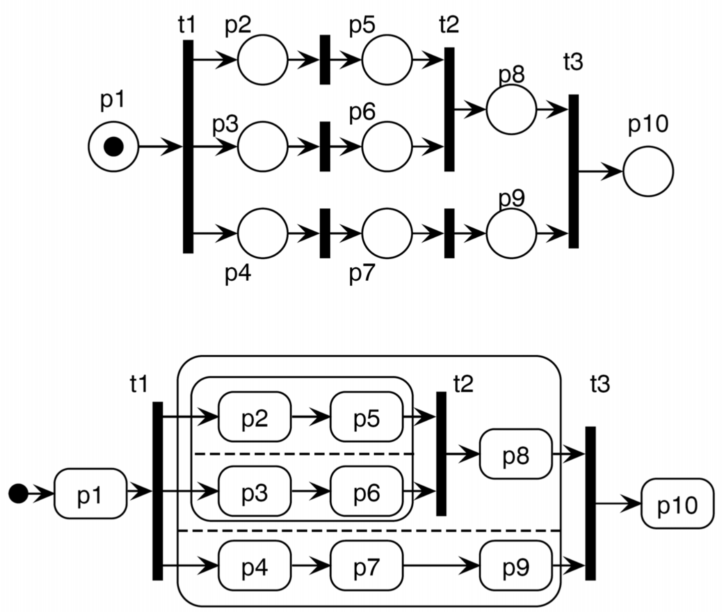


Figure 3 – Petri net with unbalanced forks and joins and equivalent statechart

Even though Petri nets and Statecharts share these concepts, the terminology is somewhat different (see Table 1). To clearly distinguish between the two formalisms, for statecharts we use the terms 'node' and 'hyperedge' rather than the more common 'state' and 'transition'.

Though both Petri nets and statecharts are based on hypergraphs, statecharts have as additional feature an AND/OR hierarchy on nodes to explicitly model concurrency. This hierarchy is visualised by node containment, where nodes that contain other nodes are called composite.

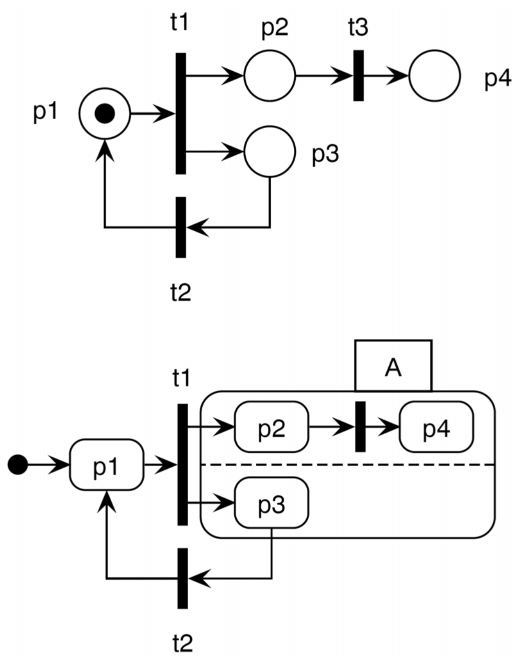


Figure 4 – Unbounded Petri net and non-equivalent statechart

There are two kinds of composite nodes: AND and OR, each imposing its own constraint on the global state of the statechart. An AND node denotes concurrency: If an AND node is active, all its immediate subnodes must be active as well. An OR node denotes exclusiveness: If an OR node is active, one of its immediate subnodes must be active as well. (So OR is actually XOR.)

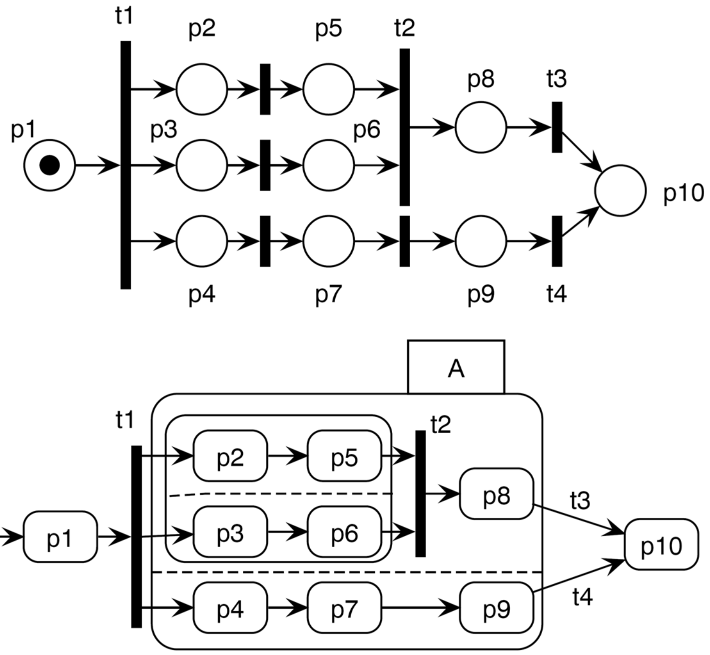


Figure 5 – Unsafe Petri net and inequivalent statechart

The constraints imposed by the Statechart hierarchy imply that leaf nodes of the hierarchy, *basic* nodes, are not active more than once in the same global state. This seems to suggest that statecharts correspond to *safe Petri nets* (nets in which in each marking every place has at most one token; see Figure 5 and Chapter 9 regarding Petri Nets). However, there is no known structure-preserving translation from safe Petri nets to statecharts. The difficulty in defining such a translation lies in constructing an appropriate AND/OR hierarchy. On the other hand, to translate a statechart into a Petri net, simply dropping composite nodes from the statechart seems to suffice.

In this thesis we have adopted a *structure-preserving algorithm* that translates a Petri net into an equivalent (bisimilar) statechart. The algorithm is structure-preserving in the sense that it maps each place to a *basic* node and each transition to a *hyperedge*. Thus, loosely speaking, the algorithm imposes an AND/OR hierarchy of nodes on the Petri net structure. Key property of the algorithm is that it is structural and does not use any Petri net analysis techniques, like *place invariants* or *reachability graphs*. The correctness of the algorithm has been already formally proven. However, it is not complete: it fails for some nets that do have a statechart equivalent. But statecharts which it fails to construct are not likely to be drawn by statechart designers, so this non-completeness does not appear to be a severe limitation in practice.

For the reverse direction, not every statechart can be translated into an equivalent Petri net, so omitting composite nodes from a statechart does not always yield an equivalent net.

2 AIRPORT LAYOUT

The airport that we have chosen to study has a very simple, but sufficiently flexible layout, letting us to explore the most important issues related to many Airport Air-Side Operations:

- One *Runway* with two opposite headings, that for our purposes will be called Heading One and Heading Two: these will be obviously used for take offs and landings, according to wind directions. The Runway will be split in five different *Places* (we will see later that these *Places* will coincide with Petri Nets Places, assuming that aircrafts will be considered as *Tokens*), according to the possible exits ad junctions that feed the Runway itself.
- Four *Junctions*, two of them are *Rapid Exits*: this is a very realistic assumption for the kind of airport that we are going to study: each junction will be considered as a single Place.
- One *High Speed-Taxiway*, for aircraft high-speed taxiing purposes. According to what has been already done with the Runway, the High Speed-Taxiway will be split too into five different Places, according to all the junctions available.
- Two normal *Taxiways*, for taxiing to and from the Apron (each one, as per the junctions, will be considered as a single Place).
- One *Apron* for aircraft docking purposes onto Stands. Apron will be described in details later, according to the definition of stands and of Time Slots on the Land-Side.

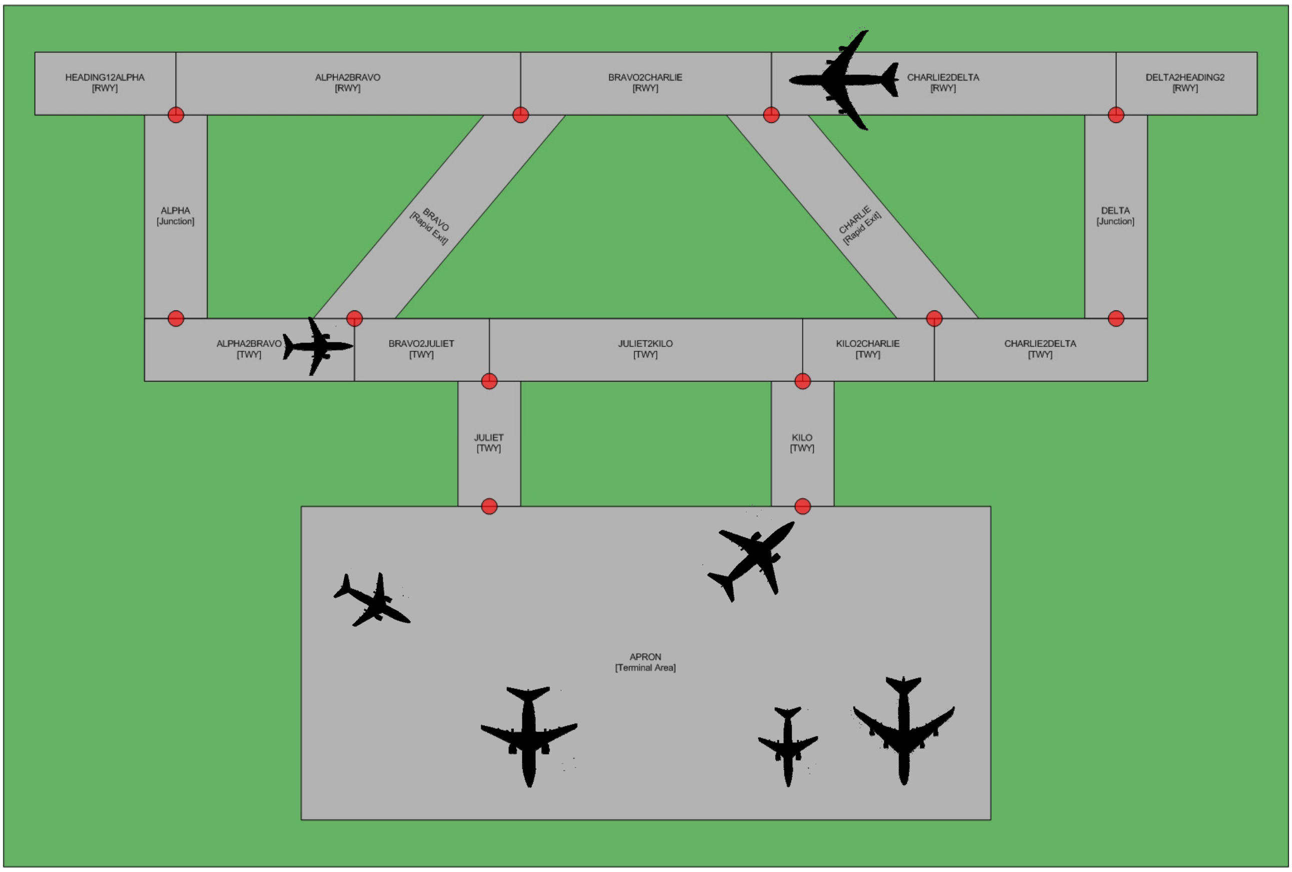


Figure 6 – Airport Layout

This kind of Airport layout is very similar to most of all medium size Italian airports, such as Bologna "Guglielmo Marconi" or Torino "Sandro Pertini". Of course, all the considerations on the following pages can be extended to major airports with more complex layouts, such as Milan Malpensa or Rome Fiumicino hubs, by building a more complex net of possible places for aircrafts.

In the following Table 2 and Table 3, all the possible aircraft paths along the "sample" airport that we have decided to study, will be presented, according to wind directions, aircraft sizes and scheduled departures/arrivals.

Wind Coming from Heading One	**Arrivals**				**Departures**			
	Light-Medium		**Heavy**		**Light-Medium**		**Heavy**	
	A	**B**	**C**	**D**	**A**	**B**	**C**	**D**
Runway								
HeadingOne2Alpha	-	-	-	-	7	8	9	10
Alpha2Bravo	-	-	4	4	6	7	8	9
Bravo2Charlie	3	3	3	3	5	6	7	8
Charlie2Delta	2	2	2	2	-	-	6	7
Delta2HeadingTwo	1	1	1	1	-	-	-	-
Junctions and Rapid Exits								
Alpha	-	-	5	5	-	-	-	-
Bravo	4	4	-	-	-	-	-	-
Charlie	-	-	-	-	4	5	-	-
Delta	-	-	-	-	-	-	5	6
High Speed Taxiway								
Alpha2Bravo	-	-	6	6	-	-	-	-
Bravo2Juliet	5	5	7	7	-	-	-	-
Juliet2Kilo	-	6	-	8	-	3	-	3
Kilo2Charlie	-	-	-	-	3	4	3	4
Charlie2Delta	-	-	-	-	-	-	4	5
Taxiways								
Juliet	6	-	8	-	-	2	-	2
Kilo	-	7	-	9	2	-	2	-
Apron								
Apron/Stands	7	8	9	10	1	1	1	1

Table 2 – Possible Sequences for places occupation, wind coming from Heading One

As one can notice from Table 2 and Table 3, descending from the choice of the layout of the airport, only four possible path are allowed for each operation, depending on the wind, thus letting us consider simple airport traffic control management. The shortest paths will be realistically assigned to light/medium weight aircrafts, since they will have the possibility to take off and land with shortest distance available. Heavy aircrafts will be assigned to longest path, which coincide with longest runway distance available.

All of the paths described in Table 2 and Table 3 are shown in detail also in Figure 7 and Figure 8. It is interesting to underline some clever considerations coming from Figure 7 and Figure 8. As a matter of fact one can notice that:

- for this kind of layout, *Leading Circulation* of the traffic along the airport (e.g. clockwise for wind coming from Heading Two) is mainly dictated by the prevailing wind. That's the reason why prevailing wind direction will be the first parameter that we will take into consideration inside our model.

Wind Coming from Heading Two	Arrivals				Departures			
	Light-Medium		Heavy		Light-Medium		Heavy	
	A	B	C	D	A	B	C	D
Runway								
HeadingOne2Alpha	1	1	1	1	-	-	-	-
Alpha2Bravo	2	2	2	2	-	-	6	7
Bravo2Charlie	3	3	3	3	5	6	7	8
Charlie2Delta	-	-	4	4	6	7	8	9
Delta2HeadingTwo	-	-	-	-	7	8	9	10
Junctions and Rapid Exits								
Alpha	-	-	-	-	-	-	5	6
Bravo	-	-	-	-	4	5	-	-
Charlie	4	4	-	-	-	-	-	-
Delta	-	-	5	5	-	-	-	-
High Speed Taxiway								
Alpha2Bravo	-	-	-	-	-	-	4	5
Bravo2Juliet	-	-	-	-	3	4	3	4
Juliet2Kilo	-	6	-	8	-	3	-	3
Kilo2Charlie	5	5	7	7	-	-	-	-
Charlie2Delta	-	-	6	6	-	-	-	-
Taxiways								
Juliet	-	7	-	9	2	-	2	-
Kilo	6	-	8	-	-	2	-	2
Apron								
Apron/Stands	7	8	9	10	1	1	1	1

Table 3 – Possible Sequences for places occupation, wind coming from Heading Two

- given the layout and the prevailing wind direction, there are only two Places among all of the Places of the air-side where the Circulation has two possible ways, in particular *Taxiways Juliet and Kilo*. Hence, if one would like to improve safety, of course reducing capacity, can decide to use Taxiways Juliet e.g for Arrivals and Kilo for Departures flights if Leading Wind is coming from Heading One and vice versa. In this way also Juliet and Kilo are one way Places, thus segregating in-bound and out-bound traffic and leaving the possibility of collision very remote. This choice corresponds to eliminating all B and D paths from Table 2 and Table 3. This policy may be safer, but not necessarily the most efficient in terms of capacity.

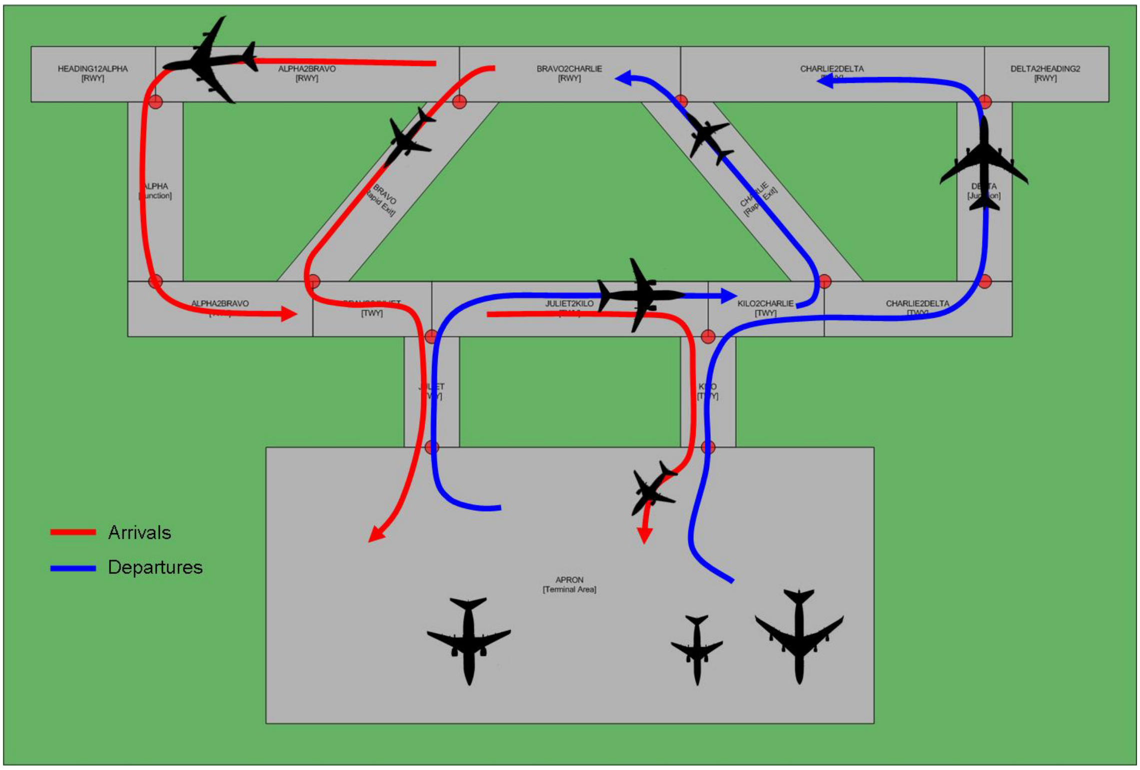


Figure 7 – Wind coming from Heading One

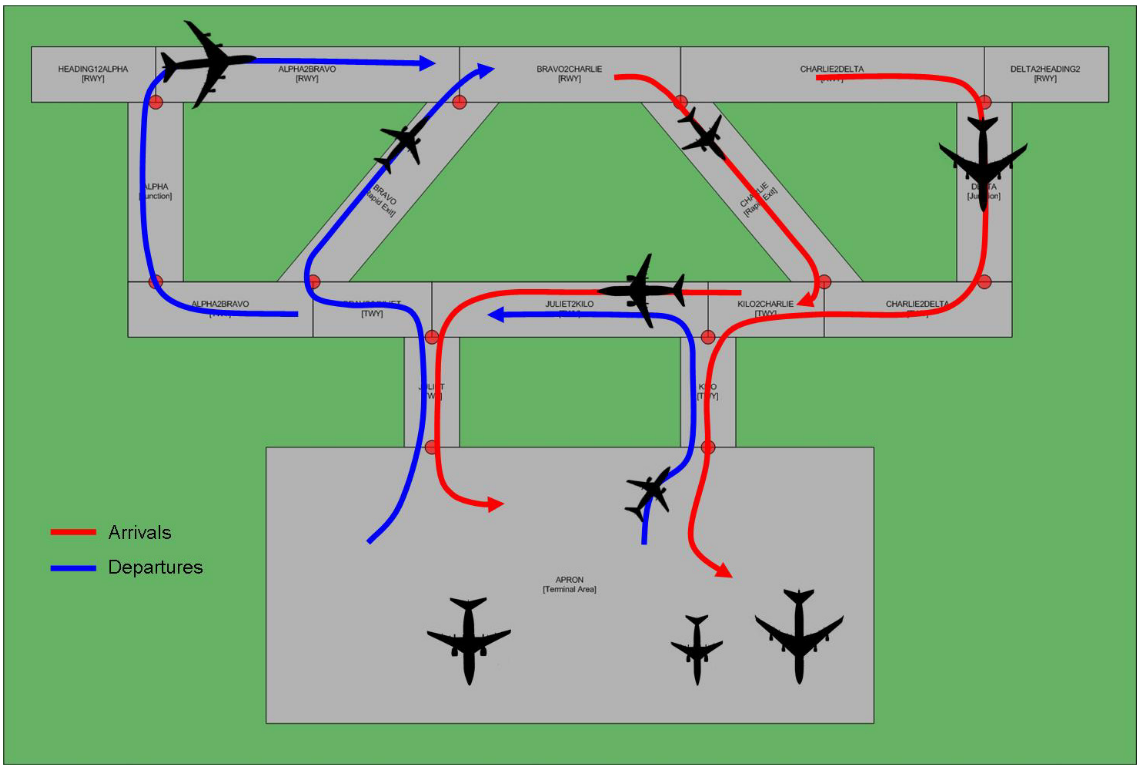


Figure 8 – Wind coming from Heading Two

3 PETRI NET MODEL OF THE AIRPORT

The Airport Layout chosen leads to a well defined Petri Net. Because of the timed discrete event nature of the problem, Timed Coloured Petri Nets (TCPNs) have been chosen as a modelling tool. The structure of a Petri net is a *Bipartite Directed Graph* describing the structure of a *Discrete Event System*, while the dynamics of the system is described by the execution of the Petri Net. A Petri net is coloured if the tokens are distinguishable.

For real systems it is often important to describe the temporal behaviour of the system, i.e. we need to model durations and delays. The CPN extended by Time gives a possibility to describe the dynamic properties of a system in the time space. The time concept of TCPNs is based on the introduction of a *global clock*. The clock values represent the model time and they are discrete (e.g. integers). Each token carries a time value, also called a time stamp. The time stamp describes the earliest model time at which the token can be used, i.e., removed by the occurrence of a binding element.

The *inputs* of the model are:
- the desired original scheduling,
- the expected occupancy time of every place,
- every information connected to aircrafts paths.

The *outputs* of the simulation are:
- occupancy graphs,
- text file with simulation results.

The Timed CPN model of the airport that obeys our modelling assumptions has been built in Stateflow. The model is shown in Figure 9 as it appears in the program package. The model is replicated into 10 layers, each one reading the input scheduling from a Simulink interface. The main part of the model realizes the simulation of airport operations (i.e. simulates aircraft moving along the airport as function of time). The modelling time is handled by the TCPN simulator. One *unit in the model time* is set to be one second.

To model the traffic flow as a *Discrete Events System*, we need to define *Events* that are relevant, such as *Clearances* (for landing, take-off) and passing some positions by the aircraft.

The first assumptions that we made about this Airport Petri Net are:

- The Aircrafts will be considered as Coloured Tokens; the colour, given by the Air Traffic Controller, contains all the necessary information for the Aircraft to move along the Airport (ID, speed, waypoints, holding points, runway heading, and so on…)
- The Airport Layout will be split into 18 different Places (including 'idle')
- No more than 10 aircraft can populate the Net contemporarily
- Therefore the Stateflow implementation of the Petri Net will have 10 different Layers, one for each aircraft. The single Petri Net Layer implemented in StateFlow is shown in Figure 9.

Having considered the modelling assumptions and goals, a timed stochastic CPN model with the elements below have been constructed.

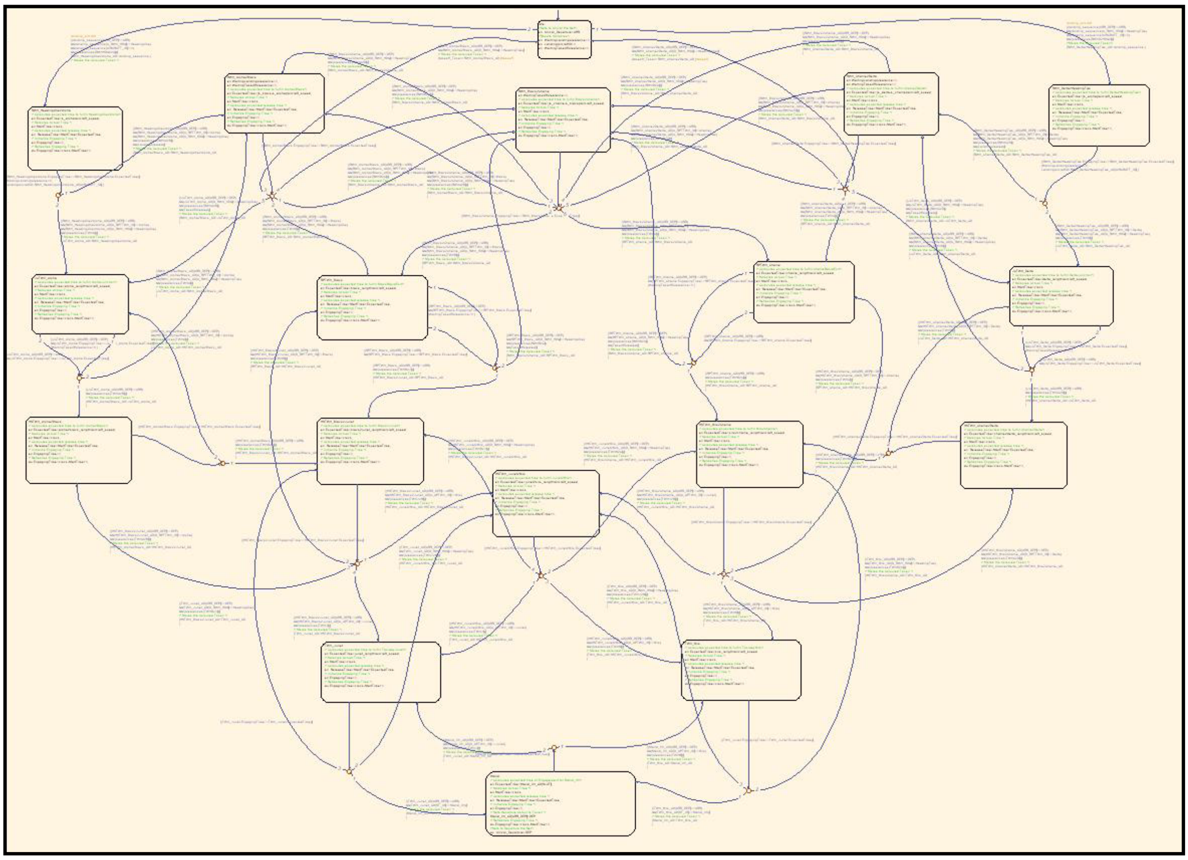


Figure 9 – Airport Petri Net

The chart in Figure 10 shows the detailed Stateflow implementation of the Petri Net Place representing the segment "Bravo2Charlie" inside the Runway. Firstly, StartTime is sampled from global clock, together with the calculation of ExpectedTime and forcasts on ReleaseTime. Then, Engaging time is calculated and updated at each sample time. The transition to the next place will be fired when local EngagingTime will be more than Expected Time. Note that the place is highlighted in blue by Stateflow since, when the simulation was running, that place was engaged by an aircraft. The reader should also note from Figure 10 that each time a transition is fired, the token is stored into the next local place variable, like in classical Petri Nets.

Stochastic disturbances have been taken into account to make more realistic the results. As a matter of fact, the aim of the simulation includes to handle the effect of uncertainty of occupancy time (due to weather, integrity of communications, etc.) on the airport capacity. So the developed model puts some uncertainty into the desired original scheduling. In addition, there are possibilities to change the aircrafts orders into the queues, and change the colour of any tokens during the simulation.

In order to model operations around an airport, we need a couple of elements to work with. In this case the most important elements are:

- Runway (RWY)
- Taxiways (TWY)
- Aircraft (A/C)
- Rules that govern the interaction between A/C and use of the RWY, of the Stands, and of the TWYs.

The characteristic properties of each of the model elements are as in the following paragraphs.

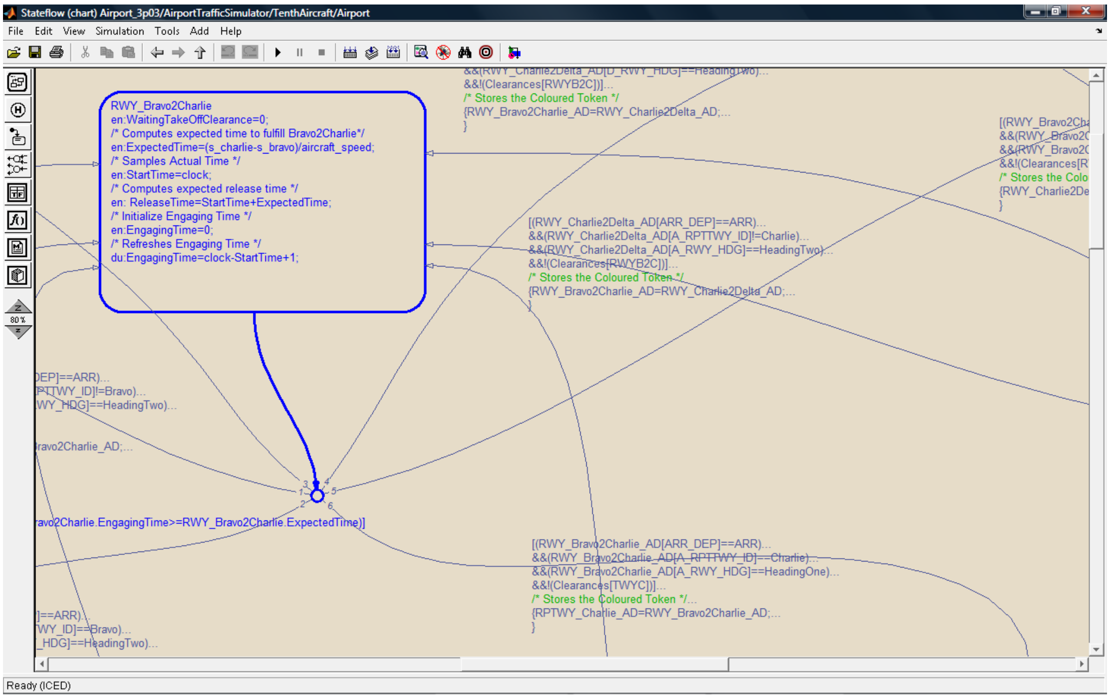


Figure 10 – Detail of Petri Net Place "RWY_Bravo2Charlie" StateFlow implementation

3.1 THE RUNWAY

The runway to be modelled is simple and yet has a potential containing most of the elements important in runway dynamics. In order to obtain a relatively simple model for simulation and dynamic analysis purposes, the following modelling assumptions have been made:

- It is assumed that the time value of reaching a definite position by an A/C can be pre-calculated.
- Only basic ATC laws are applied (no wake turbulence separation is taken into account, but the model is open to take into account also that; for numerical values of inter-arrival separations in terms of time see figure).
- Only visual flight rules (VFR) operations are considered (this condition is closely related to the previous assumption because in VFR pilots are able to maintain own separation).
- Pilots are assumed being familiar with airport configuration (this is important because runway occupancy times can be influenced by pilots but their performance depends on their awareness that they have to urgently exit the runway and on the knowledge where they can do it).
- The possibility of speed adjustment can be considered, but won't be implemented yet (this assumption ensures some stability into the planned sequences caused by the deviation in runway occupancy times).
- A single 4000 m Runway (RWY) is considered with two 90° TWY on both ends and two rapid exit taxiways, located at 1300 m from approach end threshold, thus leading to a completely symmetric configuration (see Figure 6).
- Aircraft Speed on the Runway is on average 150 knots.
- A dedicated Runway Monitoring System manages all clearances queues to landing and departing aircrafts, virtually solving all conflicts, but realistically relying on real (i.e. subject to failures) sensors.

3.2 THE TAXIWAYS

The assumptions are very similar to the ones already made for the runway. In particular

- It is assumed that the time value of reaching a definite position by an A/C can be pre-calculated.
- Only visual flight rules (VFR) operations are considered (this condition is closely related to the previous assumption because in VFR pilots are able to maintain own separation).
- Aircraft Speed on the Taxiway is on average 25 knots into taxiway links and 50 knots into high-speed taxiways.
- No specific monitoring is performed on the taxiways except from visual and SMR.

3.3 THE AIRCRAFTS (A/C) TOKENS

The difference between aircraft is based on ICAO threshold speed categories (A to E). Only aircraft with categories A, B and C are considered. The tokens are composed of 20 elements (some of them are spare fields):

```
% Arrival Information
AIRCRAFT_ID=0; %index of Aircraft Call-Sign
ETOA=1; %index of Estimated Time of Arrival field
ARR_DEP=2; %index of Arrival/Departure field
SPARE_0=3; %index of SPARE_0 field
SPARE_1=4; %index of SPARE_1 field
A_RWY_ID=5; %index of Runway ID field
A_RWY_HDG=6; %index of Runway Heading field
A_RPTTWY_ID=7; %index of Rapid Exit/Link ID field
A_HSTWY_ID=8; %index of High Speed Taxiway ID field
A_APTWY_ID=9; %index of Apron Taxiway ID field
ST_ID=10; %index of Stand ID field
SLOT=11; %index of Slot Time field
SPARE_2=12; %index of SPARE_2 field
% Departure Information
D_APTWY_ID=13; %index of Apron Taxiway ID field
D_HSTWY_ID=14; %index of High Speed Taxiway ID field
D_RPTTWY_ID=15; %index of Rapid Exit/Link ID field
D_RWY_HDG=16; %index of Runway Heading field
D_RWY_ID=17;%index of Runway ID field
SPARE_3=18; %index of SPARE_3 field
SPARE_4=19; %index of SPARE_4 field
```

Table 4 – Token Vector Fields

3.4 GOVERNING RULES

The following rules are used to control the interactions between A/C and the use of the airport places.

- Arrivals have priority on departures.
- A landing aircraft will not normally be permitted to cross the runway threshold on its final approach until the preceding departing A/C has crossed the end of the runway, or has started a turn, or until all preceding landing A/C are clear off the RWY.
- Communications Clearances queues are managed stochastically

3.5 SCHEDULING STRATEGIES

Due to uncertainty of arriving exactly on estimated time to the threshold, the input schedule of the model is calculated from the original schedule by adding or subtracting a random number to the times of estimated arrival on the threshold.

The input schedule contains A/C with estimated time of reaching the threshold (for arrivals) and scheduled time of reaching the runway on taxiways. The priority of arrivals law is applied with use of the philosophy *departures between arrivals not arrivals between departures*.

The model gives a clearance the latest time possible i.e. one clearance is valid at a time and a new clearance is given *after* the preceding A/C is off the previous place.

3.6 CLEARANCES

The scheduling method at a given model time is to select the next time an A/C arriving at the threshold and calculates, whether departing A/C scheduled to depart earlier are clear off the runway before this time. It is done by adding the time the A/C vacates the runway for departing (that depends on category) to the actual time or to the scheduled time of the aircraft reaching the runway on taxiway A, whichever is greater. When this value is smaller or equal to the estimated time the landing A/C reaching the threshold, the clearance is issued to the departing aircraft, otherwise not. This planning method can be considered as a *small scale window method* proposed in [PN-12].

3.7 SHORT-TERM CONFLICT ALERTS (STCA)

Time for conflict detection, identification, and resolution ranging from:

- Detection and identification up to 2 seconds
- Resolution ranging between 2 and 12 seconds

The reaction time is assessed by an ‘observer’ who measures the time between the initiation of a conflict and the reaction of the ATCO in charge. The reaction of an ATCO is defined by the time when the ATCO contacts the pilots to resolve the conflict.

4 SIMULATION RESULTS

The proposed TCPN model was verified against qualitative engineering expectations by using simulation experiments. The effects of varying reliability parameters of systems affecting safety such as SMR False Alarm Rates, ADS-B Probability of Detection and Clearances Communication Integrity have been studied extensively.

For the TCPN model simulation Simulink Realtime Workshop and Stateflow Coder program packages have been used.

The simulation outputs are Occupancy graphs and Output text files.

Some of the outputs of the simulation are graphs showing the taxiing occupancy vs time (see Figure 11, Figure 12 and Figure 13). It shows either low (i.e. runway is free in time interval on horizontal axis) or high level (i.e. runway is occupied). Airport and Stand occupancy times too have been investigated (Figure 14 and Figure 15).

The other output is a text file that includes the following data:

- type of the aircraft
- aircraft path followed
- scheduled operations time (e.g. scheduled time of departure)
- estimated PDs, CFAR and CI
- times of clearances (e.g. clearance was given at time 160 though scheduled time was 140)
- delays (i.e. the difference between the scheduled and the real clearance time)

4.1 TAXI TIMES

The taxi time is measured automatically for each aircraft starting from the gate (velocity > 0 kts) until the wheels leave the ground (take-off) for outbound movements. For inbound movements the time measurement starts when the wheels touch the ground (touch down) until the velocity is 0 at the gate or stand. Since the aircrafts always have a constant speed level, Taxi Time differences can only be interpreted as being caused by a more efficient control by the ATCOs using ASMGCS.

Some delays (see Figure 11 and Figure 12) are due to *reduced or simplified performances* (even if Safety is compliant) of some of the equipments (e.g. with reference to False Alarm Rates that were incremented from 1% to 10%): these delays can lead to previously unexpected economical consequences, thus requiring more accurate systems to be installed, in order to meet also Airport economical constraints. On the other hand, as one may expect easily, some other delays are due to the increase of Safety functions; one example is reported in Figure 13, where the RWY Monitoring function has been implemented: conflicts are drastically reduced (see also Figure 19), but delays increase further.

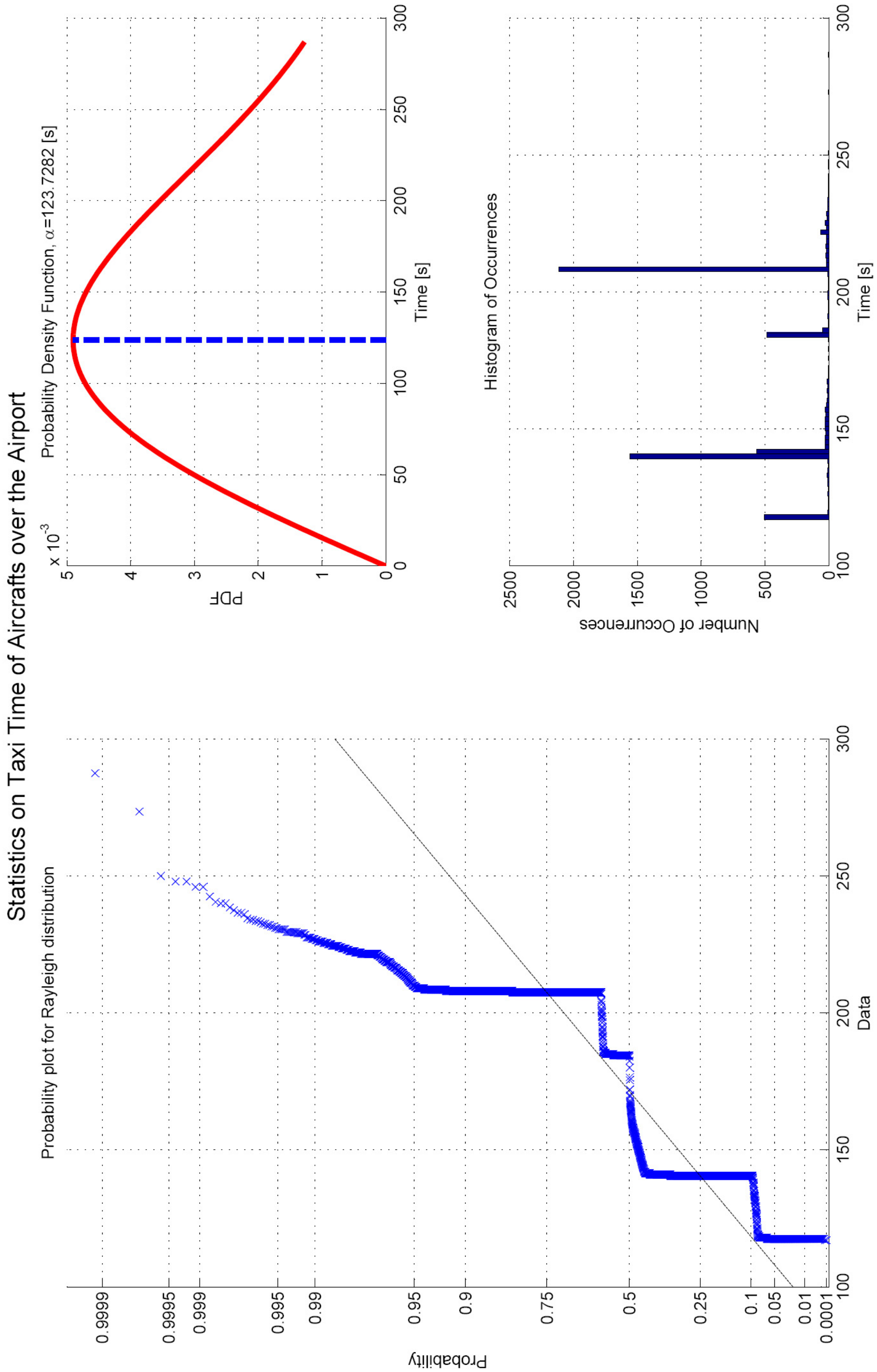


Figure 11 – Taxi Time Analysis (CFAR 1%, PD 95%, ICOMM 95%, No RWY Monitoring)

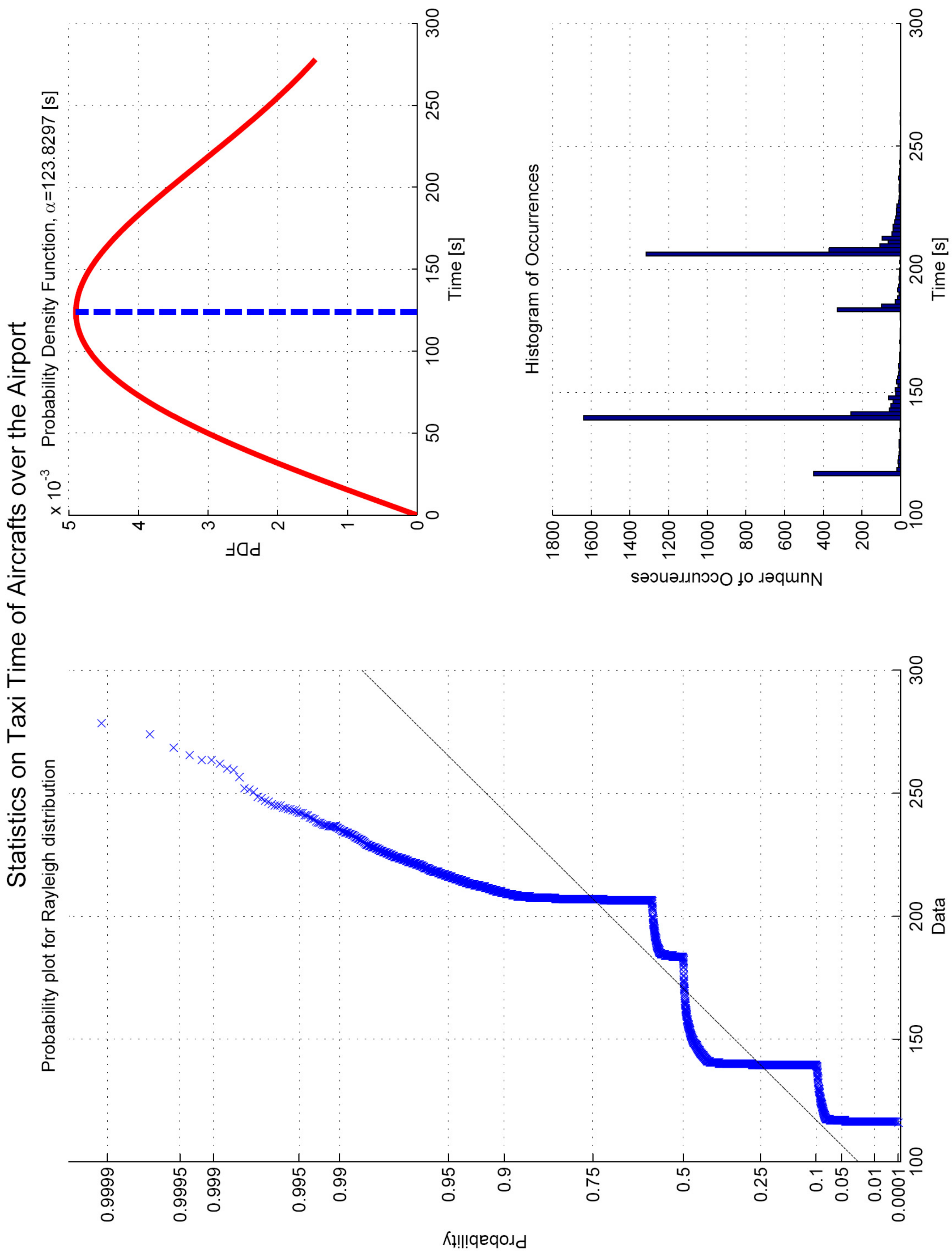


Figure 12 – Taxi Time Analysis (CFAR 10%, PD 95%, ICOMM 95%, No RWY Monitoring)

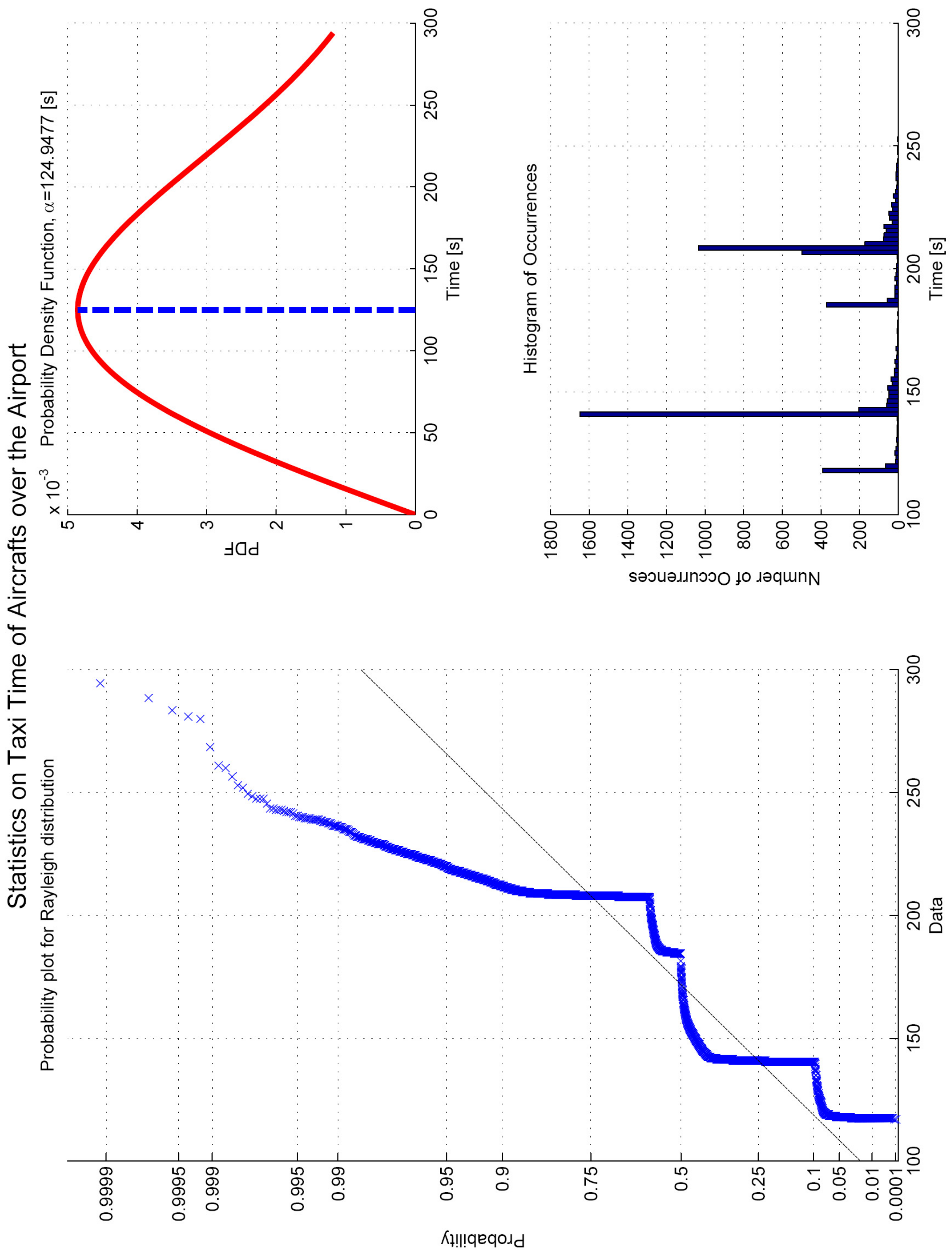


Figure 13 – Taxi Time Analysis (CFAR 10%, PD 95%, ICOMM 95%, With RWY Monitoring)

4.2 AIRPORT OCCUPANCY TIMES

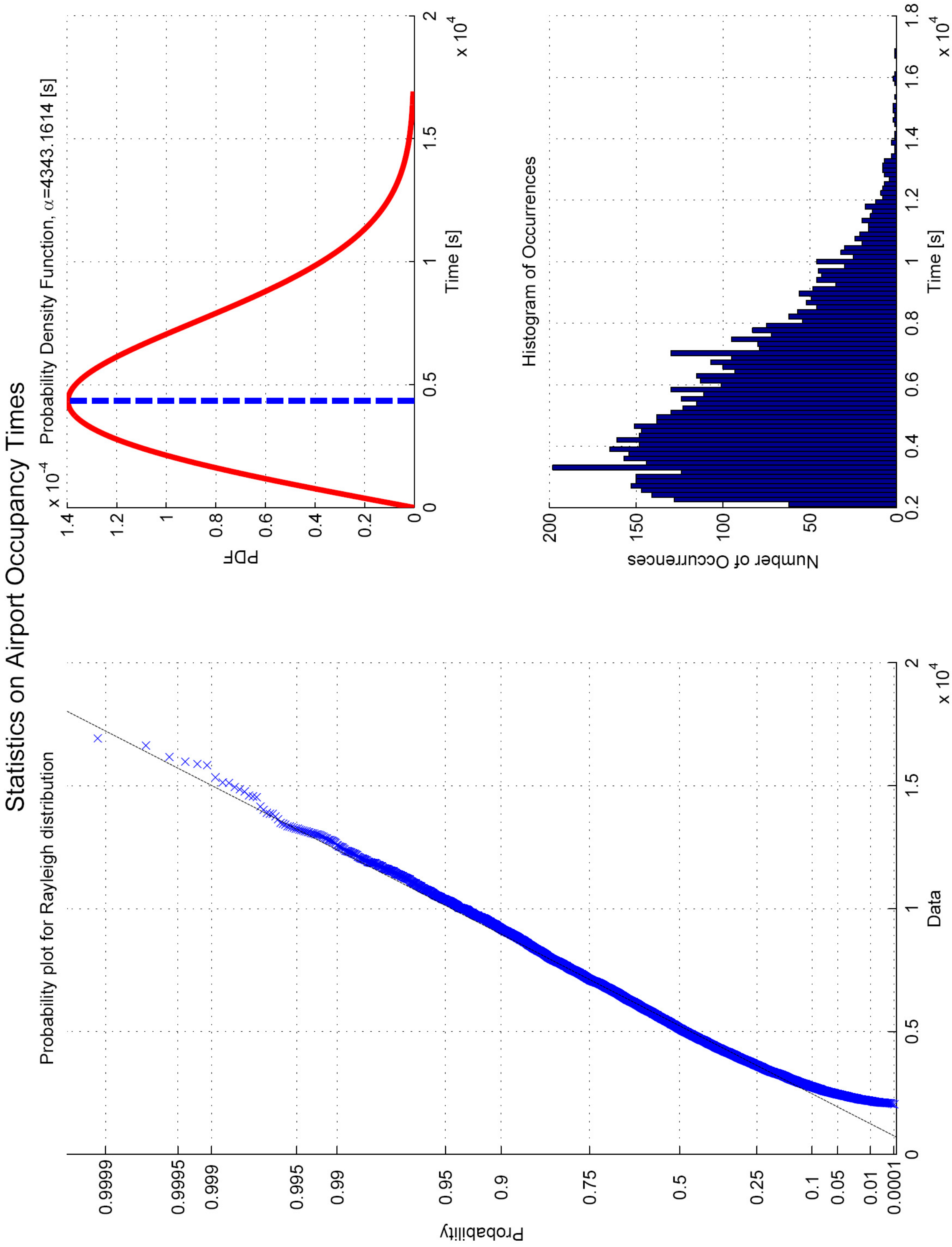


Figure 14 – Airport Occupancy Analysis (CFAR 1%, PD 95%, ICOMM 95%, With RWY Monitoring)

4.3 STAND PERMANENCE TIMES

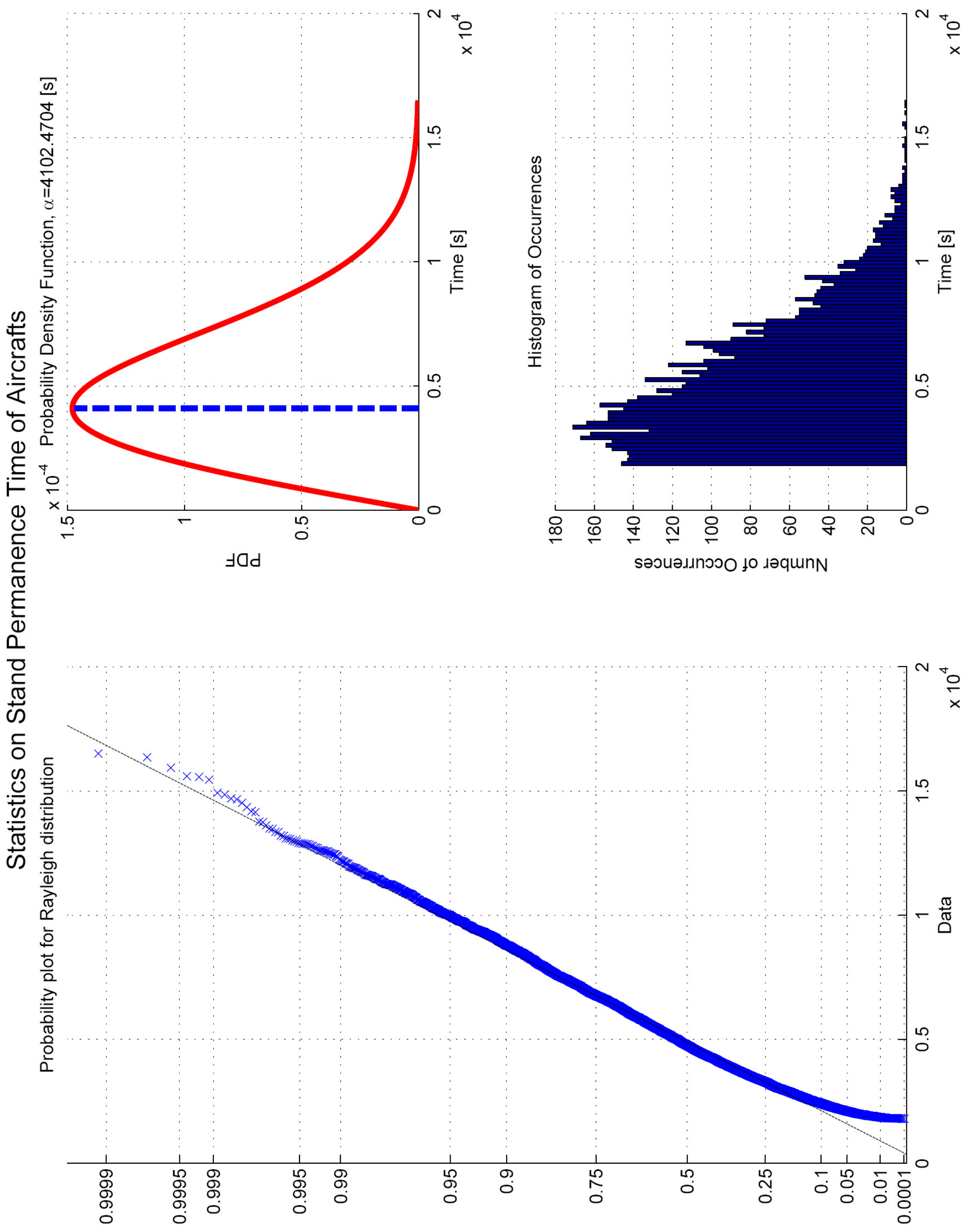


Figure 15 – Stand Permanence Analysis (CFAR 1%, PD 95%, ICOMM 95%, With RWY Monitoring)

4.4 RELIABILITY PERFORMANCES OF SMR, ADS-B AND COMMUNICATION INTEGRITY

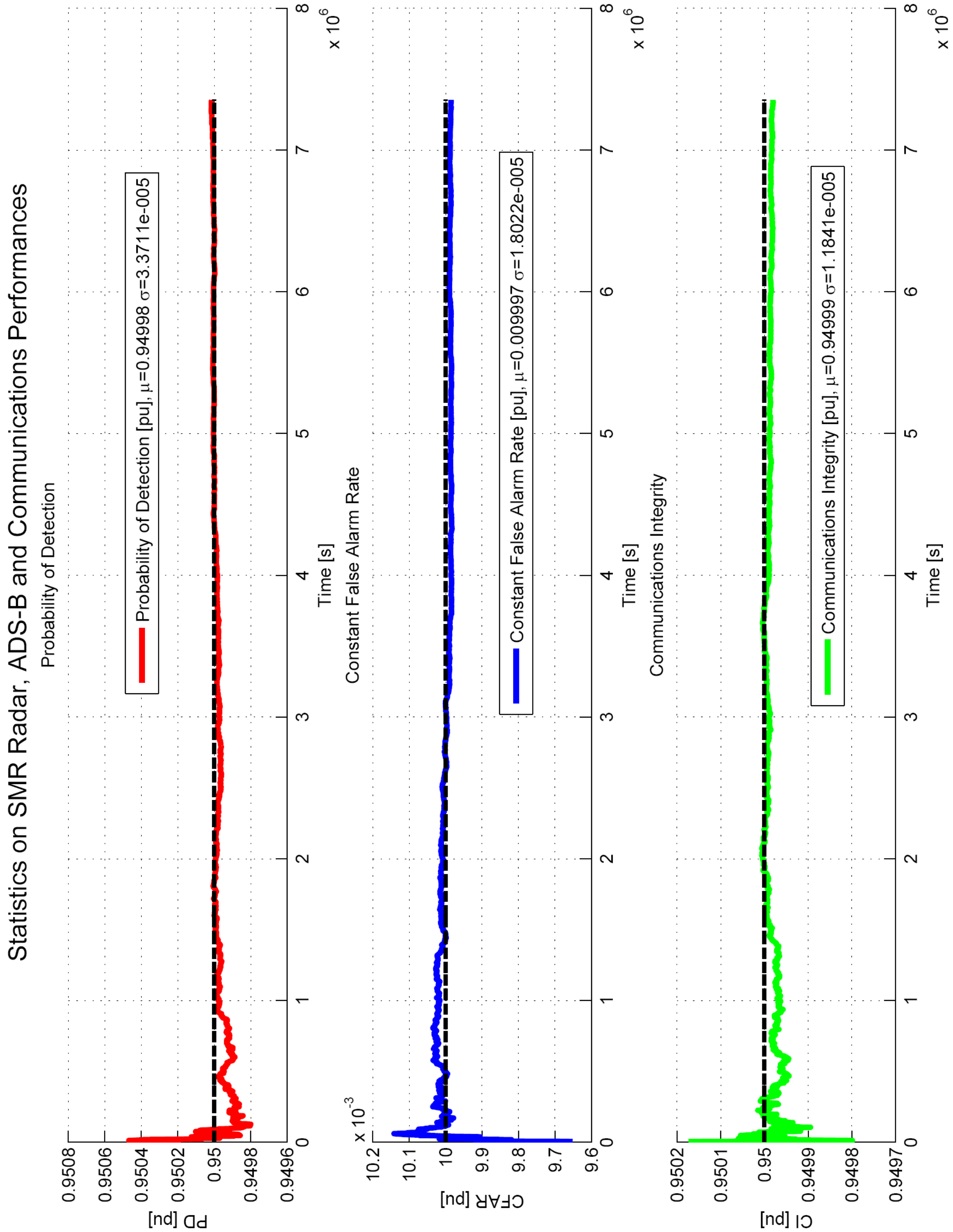


Figure 16 – Reliability Performance Analysis (SMR Radar, ADS-B, Communications Integrity)

4.5 A-SMGCS TRAFFIC VOLUME ANALYSIS

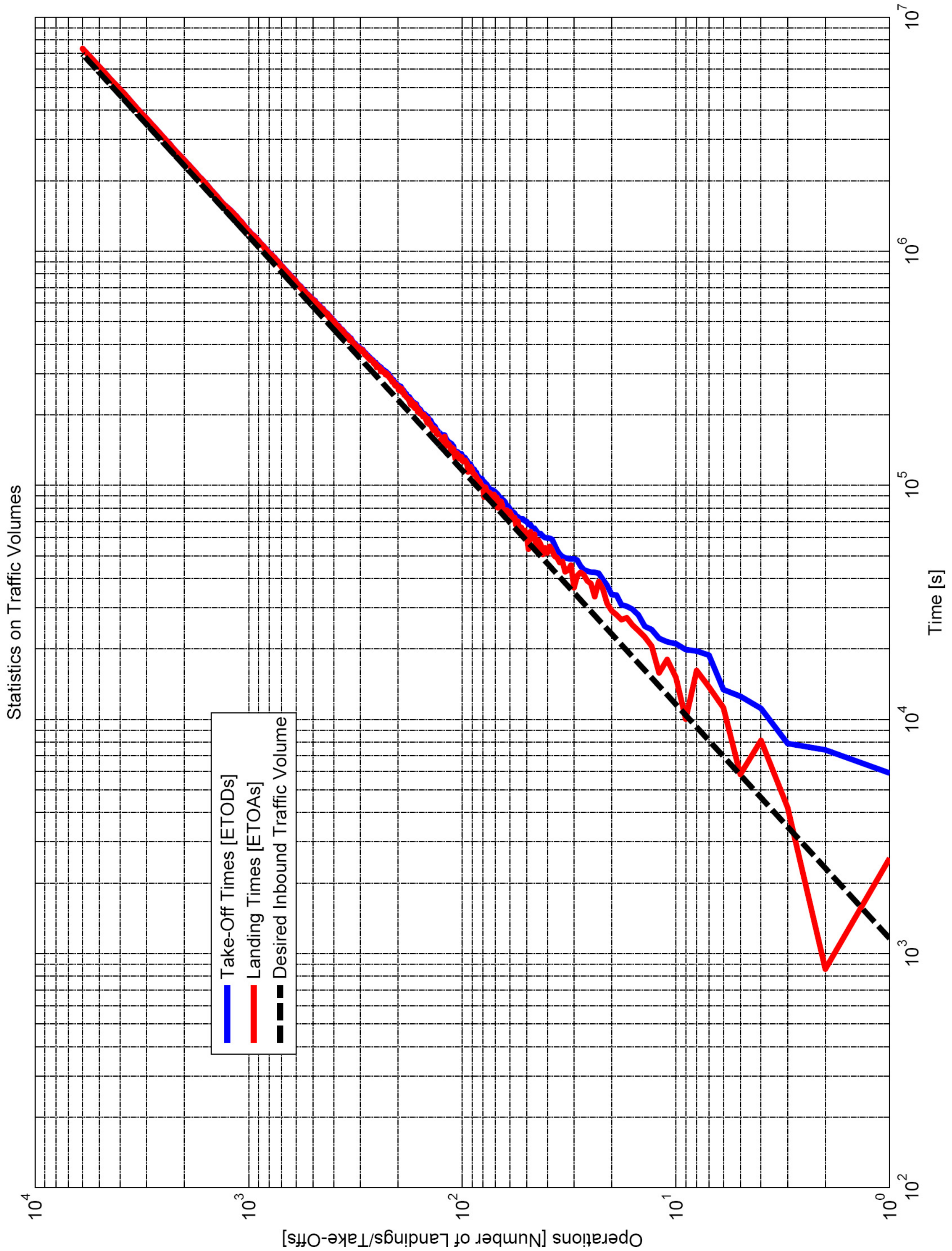


Figure 17 – Traffic Volume Analysis (CFAR 1%, PD 95%, ICOMM 95%, With RWY Monitoring)

4.6 A-SMGCS SAFETY ANALYSIS

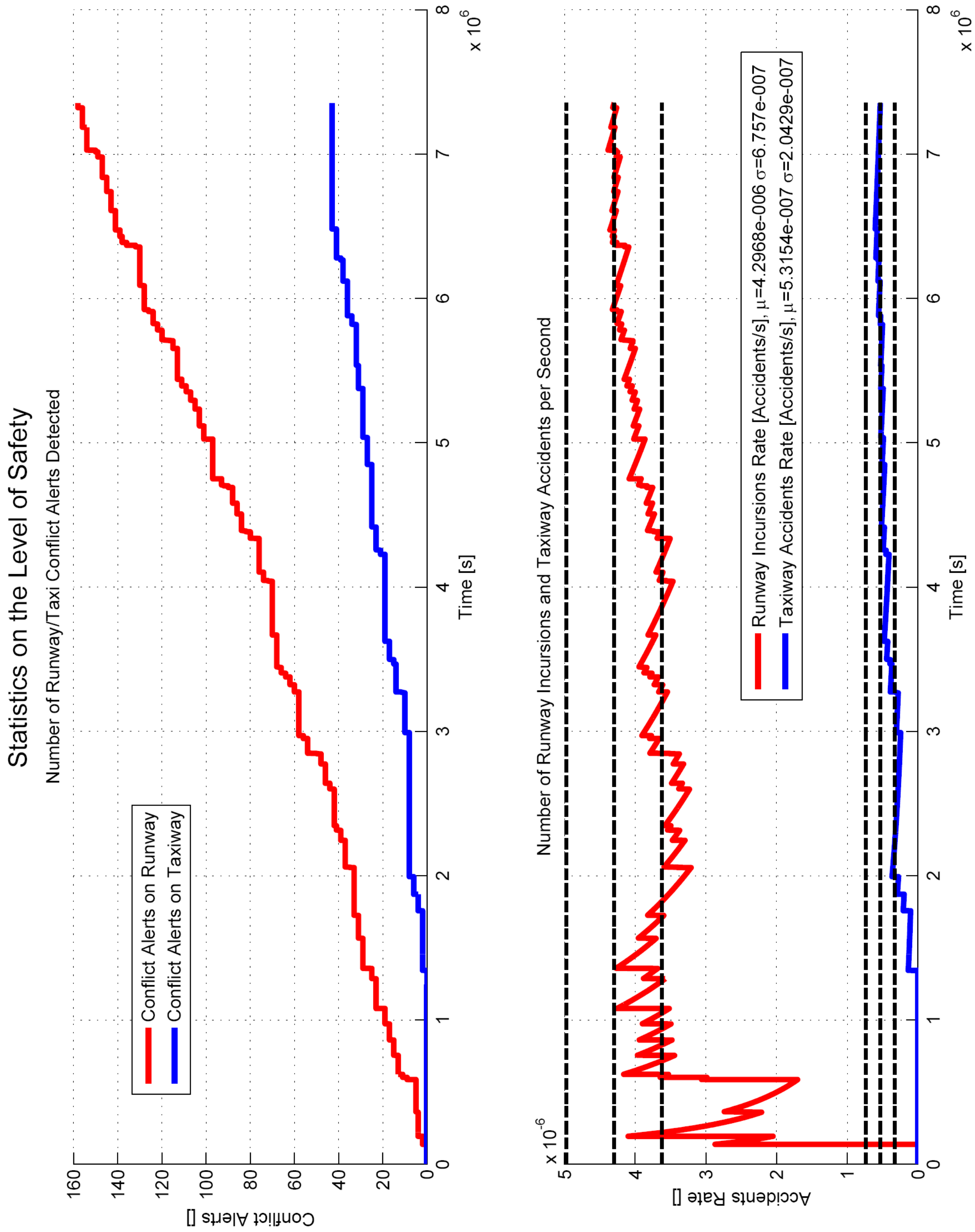


Figure 18 – A-SMGCS Safety Analysis (CFAR 10%, PD 95%, ICOMM 95%, With RWY Monitoring)

4.6.1 EFFECTS OF RUNWAY MONITORING SYSTEM ON THE LEVEL OF SAFETY

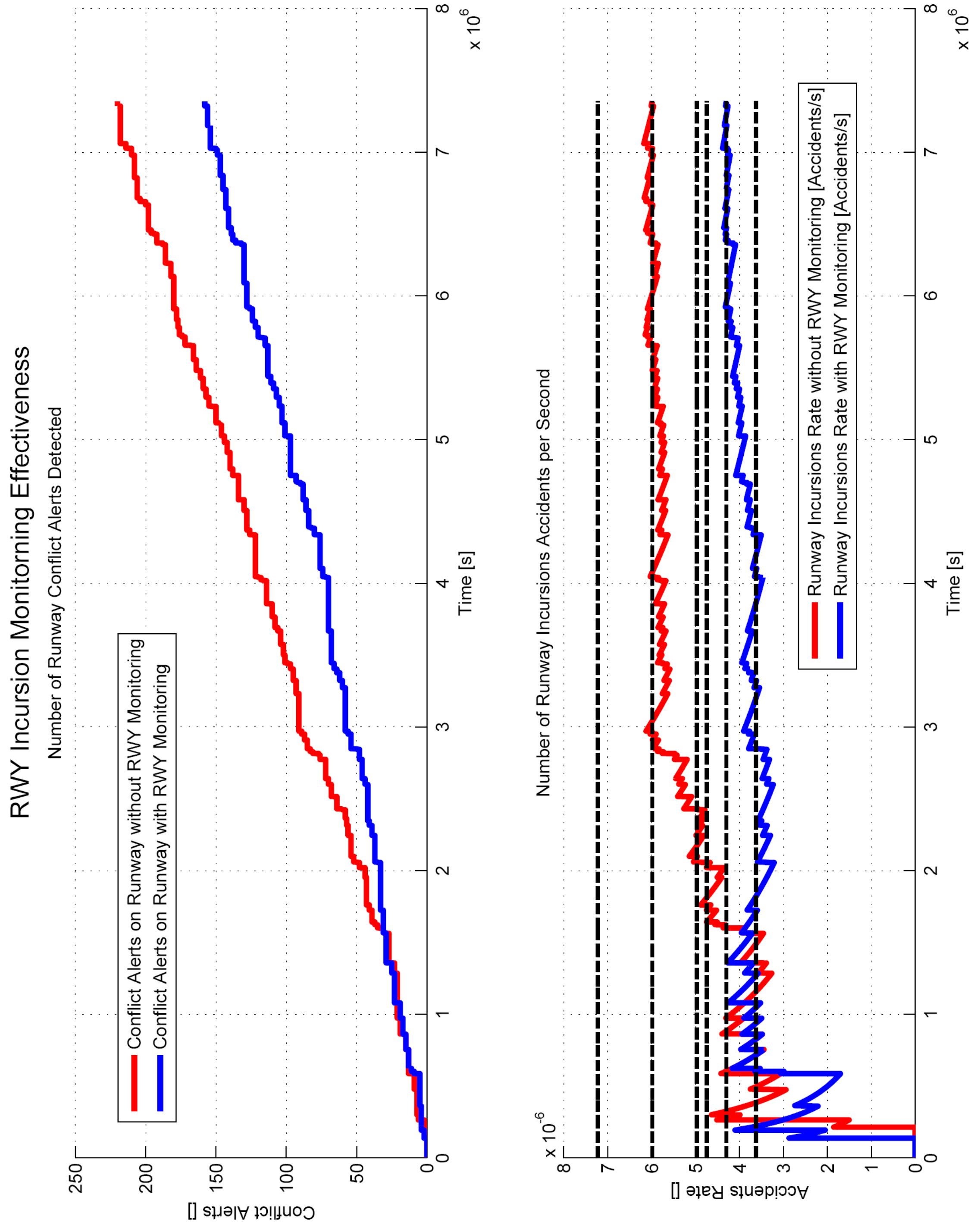


Figure 19 – Effectiveness of Runway Monitoring (CFAR 10%, PD 95%, ICOMM 95%)

SUMMARY OF CHAPTER 11

CHAPTER 12
CONCLUSIONS AND FUTURE WORK

A detailed analysis of State of the Art Technologies and Procedures into Airport *Advanced-Surface Movement Guidance and Control Systems* has been provided in this thesis, together with the review of Statistical Monte Carlo Analysis, Reliability Assessment and Petri Nets theories.

This practical and theoretical background has lead the author to the conclusion that there is a lack of linkage in between these fields. At the same of time the rapid increasing of Air Traffic all over the world, has brought in evidence the urgent need of practical instruments able to identify and quantify the risks connected with Aircraft operations on the ground, since the Airport has shown to be the actual 'bottle neck' of the entire Air Transport System.

Therefore, the only winning approach to such a critical matter has to be multi-disciplinary, sewing together apparently different subjects, coming from the most disparate areas of interest and trying to fulfil the gap.

The result of this thesis work has come to a start towards the end, when a *Timed Coloured Petri Net* (TCPN) model of a 'sample' Airport A-SMGCS has been developed, that is capable of taking into account different orders of questions arisen during these recent years and tries to give them some good answers.

The A-SMGCS Airport model is, in the end, a parametric tool relying on Discrete Event System theory, able to perform a *Reliability Analysis* of the system itself, that:

- uses a *Monte Carlo Analysis* applied to a *Timed Coloured Petri Net*, whose purpose is to evaluate the *Safety Level* of Surface Movements along an Airport
- lets the user to analyse the impact of *Procedures* and *Reliability Indexes* of Systems such as *Surface Movement Radars, Automatic Dependent Surveillance-Broadcast, Airport Lighting Systems, Microwave Sensors*, and so on… onto the *Safety Level* of Airport Aircraft Transport System
- not only is a valid instrument in the *Design Phase*, but it is useful also into the *Certifying Activities* an in monitoring the *Safety Level* of the above mentioned System with respect to changes to *Technologies* and different *Procedures*.

This TCPN model has been verified against qualitative engineering expectations by using simulation experiments and occupancy time schedules generated a priori.

Simulation times are good, and since the model has been written into Simulink/Stateflow programming language, it can be compiled to run real-time in C language (Real-time workshop and Stateflow Coder), thus relying on portable code, able to run virtually on any platform, giving even better performances in terms of execution time.

One of the most interesting applications of this work is the estimate, for an Airport, of the kind of A-SMGCS level of implementation needed (Technical/Economical convenience evaluation). As a matter of fact, starting from the Traffic Volume and choosing the kind of Ground Equipment to be installed, one can make *predictions* about the *Safety Level of the System*: if the value is compliant with the TLS required by ICAO, the A-SMGCS level of Implementation is sufficiently adequate. Nevertheless, even if the Level of Safety has been satisfied, some delays due to *reduced or simplified performances* (even if Safety is compliant) of some of the equipments (e.g. with reference to False Alarm Rates) can lead to previously unexpected economical consequences, thus requiring more accurate systems to be installed, in order to meet also Airport economical constraints.

Work in progress includes the analysis of the effect of weather conditions and re-sequencing of a given schedule. The effect of re-sequencing a given schedule is not yet enough realistic since the model does not apply inter arrival and departure separations. However, the model might show some effect on different sequences based on runway occupancy times. A further developed model containing wake turbulence separation conditions would be more sensitive for this case.
Hence, further work will be directed towards:

- The development of *On-Line Re-Scheduling* based on the available actual runway/taxiway configuration and weather conditions.
- The *Engineering Safety Assessment* of some small Italian Airport A-SMGCSs (Model validation with real data).
- The application of *Stochastic Differential Equations* systems in order to evaluate the collision risk on the ground inside the Place alone on the Petri Net, in the event of a *Short Term Conflict Alert* (STCA), by adopting *Reich Collision Risk Model.*
- *Optimal Air Traffic Control Algorithms Synthesis* (Adaptive look-ahead Optimization), by Dynamically Timed Coloured Petri Nets, together with the implementation of Error-Recovery Strategies and Diagnosis Functions.

REFERENCES

1 AIR TRAFFIC CONTROL [ATC]

[ATC-1] "Evolution of Air Navigation Services during this Decade", Dr. C. Whelan, The Journal Of Navigation

[ATC-2] "The Future National Airspace System: Design Requirements Imposed By Weather Constraints", J. Krozel, B., A. D. Andre, P. Smith, AIAA Guidance, Navigation, and Control Conference and Exhibit
11-14 August 2003, Austin, Texas.

[ATC-3] "Pilot's Handbook of Aeronautical Knowledge", U.S. Department of Transportation, Federal Aviation Administration Flight Standards Service, 2003, FAA-H-8083-25

[ATC-4] "Instrument Flying Handbook", U.S. Department of Transportation, Federal Aviation Administration Flight Standards Service, 2003, FAA-H-8083-15

2 AIRPORT TECHNOLOGIES AND A-SMGCS [AIRP]

[AIRP-1] "Enhancing Airside Capacity – The Complete Guide" EUROCONTROL, 15 September 2003, Edition 2.0

[AIRP-2] "Measurement of Pilot Reaction Times and Runway Occupancy Times – The Complete Guide" EUROCONTROL, 14 October 2003, Edition 1.0

[AIRP-3] "Aerodrome Standards – Aerodrome Design and Operations, Cooperative Development of Operational Safety and Continuing Airworthiness", COSCAP– South Asia, July 1999

[AIRP-4] "Development Of Required Navigation Performance (RNP) Requirements For Airport Surface Movement Guidance And Control", Cassell, Smith, AIAA 14th Digital Avionics Systems Conference, November 1995.

[AIRP-5] "Airport Surface RNP (Required Navigation Performance) - Implications for GNSS", Cassell, Bradfield, Smith, Institute of Navigation, January 1997

[AIRP-6] "Evaluation of Airport Surface Surveillance Technologies", Smith, Evers, Cassell, IEEE Radar 96 Conference, Beijing, China, October, 1996

[AIRP-7] "A-SMGCS Verification And Validation Results From The Project Emma (Level 1&2)", J. Jakobi, DLR, J. Teutsch, NLR, TREN/04/FP6AE/SI2.374991/503192

[AIRP-8] "A-SMGCS Routing Guidance Functions", E. Piazza, Navia Aviation AS, IEEE AES Systems Magazine, July 2000

[AIRP-9] "A Review of Radar as a sensor for Advanced Surface Movement Guidance and Control Systems (A-SMGCS)", A. K. Brown, 2002 The Institution of Electrical Engineers.

[AIRP-10] "EMMA - Air-Ground Functional Architecture (AGFA Update)", M. Moller, D1.4.1u, 2006, - All rights reserved - EMMA Project Partners

[AIRP-11] "A-SMGCS Simulations Level I & II - VALIDATION REPORT", EUROCONTROL, European Air Traffic Management Programme, 14 March 2005, Edition 1.0

[AIRP-12] "Automatic Dependent Surveillance (ADS) and Air Traffic Service (ATS) Data Link Applications", ICAO Circular 256-AN/152

[AIRP-13] "Manual of Air Traffic Services Data Link Applications" ICAO (draft version 0.3 29 March 96);

[AIRP-14] "FARAWAY – Report on User Requirements", FAW/ALE-REP-WP1-069-R4, Sept 30th, 1996

[AIRP-15] "Multilateration technology is well suited to a wide range of applications", Surveillance Technology, A. Smith, R. Hulstrom, ICAO Journal, Number 3, 2007

[AIRP-16] "Avionics System Technology - The GNSS/INS Navigation System", A GAFACS Course, Ing. Leonardo Daga

[AIRP-17] "Development of Airport Surface Required Navigation Performance (RNP)", June 1999 NASA/CR-1999-209109

3 RANDOM NUMBERS GENERATORS [RNG]

[RNG-1] "The Art of Computer Programming: Volume 2, Seminumerical Algorithms", D. E. Knuth, Addison-Wesley, Reading, MA, 1969.

[RNG-2] "Mersenne Twister: A 623-dimensionally equidistributed uniform pseudorandom number generator", Matsumoto, Nishimura, ACM Transactions on Modelling and Computer Simulations: Special Issue on Uniform Random Generation, 1998

[RNG-3] "A Nonempirical Test on the Weight of Pseudorandom Number Generators", Matsumoto and Nishimura

[RNG-4] "Dynamic Creation of Pseudorandom Number Generators", Matsumoto, Nishimura

[RNG-5] "TestU01: A C Library for Empirical Testing of Random Number Generators", L'Ecuyer, Simard, Université de Montréal, ACM Transactions on Mathematical Software

[RNG-6] "Numerical recipes in C", W.H. Press, S.A. Teukolsky, W.T. Vetterling e B.P. Flannery, Cambridge University Press, 1992

[RNG-7] "Numerical Computing with MATLAB". The MathWorks, Inc. (http://www.mathworks.com/moler)

[RNG-8] "The ziggurat method for generating random variables", G. Marsaglia and W.W. Tsang, Journal of Statistical Software, 5 (2000), pp. 1-7. (http://www.jstatsoft.org/v05/i08)

[RNG-9] "Random number generators: Good ones are hard to find", S. K. Park and K. W. Miller, Communications of the ACM (1988), pp. 1192-1201.

4 MONTE CARLO SIMULATION [MC]

[MC-1] "The Monte Carlo method", N. Metropolis S. Ulam , J. Am Stat Assoc., 44, N 247, 335-341, 1949

[MC-2] "Monte Carlo - Concepts, Algorithms, and Applications", G. S. Fishman, 1996 Springer-Verlag New York, Inc.

[MC-3] "A primer for the Monte Carlo method", I. M. Sobol', 1994 CRC Press

[MC-4] "Guiding Principles for Monte Carlo Analysis", EPA/630/R-97/001 March 1997. Risk Assessment Forum U.S. Environmental Protection Agency Washington, DC 20460

[MC-5] "Free flight collision risk estimation by sequential Monte Carlo simulation", H.A.P. Blom, J. Krystul, G.J. Bakker, M.B. Klompstra, B. Klein Obbink, NLR-TP-2006-288, National Aerospace Laboratory NLR, September 2006

[MC-6] "Monte Carlo Simulation for Reliability Analysis of Emergency and Standby Power Systems", C. Singh, J. Mitra 1995 IEEE

[MC-7] "Application Of Sequential Monte Carlo Simulation To Evaluation Of Distributions Of Composite System Indices", R. Billinton, A. Jonnavithula, IEE Proc.-Gener. Transm. Distrib., Vol. 144, No. 2, March 1997

[MC-8] "Sull'Analisi in Regime Permanente dei Sistemi Elettrici Dissimmetrici", Tesi di Dottorato di Ricerca in Ingegneria Elettrica E Dell'informazione, Università degli Studi di Cassino, Vittorio Di Vito, Anno Accademico 2003/2004.

5 STATISTICS [ST]

[ST-1] "Statistics: a guide to the use of statistical methods in the physical sciences", Barlow, J. Wiley & Sons, 1997

[ST-2] "Statistical data analysis", G. Cowan, Oxford University Press, 1998

[ST-3] "Mathematical methods of statistics", H. Cramér, Princeton University Press, 1946

[ST-4] "Teoria degli Errori e Fondamenti di Statistica - Introduzione alla Fisica Sperimentale", Loreti, Dipartimento di Fisica Universita` degli Studi di Padova, Maggio 2006

6 RELIABILITY, SAFETY, FMECA [RE]

[RE-1] "System Reliability Modelling and Evaluation", C. Singh, R. Billinton, Hutchinson of London, 1977

[RE-2] "FAA System Safety Handbook, Appendix F", December 30, 2000

[RE-3] MIL-STD-1629A – "Military Standard Procedures For Performing A Failure Mode, Effects And Criticality Analysis", 24 NOVEMBER 1980

[RE-4] "L'impatto della generazione distribuita sulle reti di distribuzione: metodi e strumenti per l'analisi di affidabilità", Tesi di Laurea in Ingegneria Elettrica, Università degli Studi di Cassino, Cristian Lai, Anno Accademico 2000-2001

[RE-5] "Reliability evaluation of power system", R. Billinton, R.N. Allan, Plenum Press. 1996

[RE-6] "Reliability evaluation of engineering system (concepts and techniques)", R. Billinton, R.N. Allan, Plenum Press 1996

[RE-7] "Probability, random variables and stochastic process", A. Papoulis, McGraw-Hill, 1965

[RE-8] "Design for Reliability" edited by D. Crowe & A. Feinberg, 2001 CRC Press

[RE-9] "Reliability Evaluation of Airport Power Systems Under Operative Constraints", published by Speedam 2006 – U. Grasselli, S. Terramano, B. Ricci

7 PETRI NETS [PN]

[PN-1] "FAA System Safety Handbook, Appendix J: Software Safety", December 30, 2000

[PN-2] "Petri Nets: Properties, Analysis and Applications", Tadao Murata, Proceedings Of The IEEE, vol. 77, no. 4, april 1989

[PN-3] "Petri Nets for Modeling of Dynamic Systems – A Survey", David, Alla, Automatica Vol 30 no. 2 pp 175-202, 1994

[PN-4] "Petri Nets 2000- Introductory Tutorial", Balbo, Desel, Jensen, Reisig, Rozenberg, Silva, 21st International Conference on Application and Theory of Petri Nets, Aarhus, Denmark, June 26-30, 2000

[PN-5] "Dealing with Global Guards in a Distributed Simulation of Coloured Petri Nets", Knoke, Hommel, Technische Universität Berlin, Proceedings of the 2005 Ninth IEEE International Symposium on Distributed Simulation and Real-Time Applications

[PN-6] "Modeling and Performance Evaluation of Supply Chains Using Batch Deterministic and Stochastic Petri Nets", Chen, Amodeo, Chu, Labadi, IEEE Transactions On Automation Science And Engineering, vol. 2, no. 2, April 2005

[PN-7] "Modellistica, Simulazione e Controllo Supervisivo di un Magazzino Automatico", Ciro Carbone, Tesi Di Laurea In Ingegneria E Tecnologie Dei Sistemi Di Controllo, Università degli Studi di Napoli Federico II, Anno Accademico 1999-2000

[PN-8] "Control-Oriented Discrete Event Models for Manufacturing and Transport Systems", Ciro Carbone, Tesi di Dottorato di Ricerca in Ingegneria dell'Informazione III Ciclo – Nuova Serie, Università degli Studi di Salerno, Anno Accademico 2003 – 2004

[PN-9] "Formal Verification of the NASA Runway Safety Monitor", Siminiceanu, Ciardo, Electronic Notes in Theoretical Computer Science, 2004

[PN-10] "On the Generation of a Variable Structure Airport Surface Traffic Control System", Demael, Levis, Laboratory for Information and Decision Systems Massachusetts Institute of Technology, IEEE 1989

[PN-11] "Simulating Collision Probabilities of Landing Airplanes at Non-towered Airports", J. F. Shortle, Yue Xie, C. H. Chen G. L. Donohue, Transactions of the Society for Computer Simulation, Vol. 79, No. 10, 2003

[PN-12] "Modeling and Optimization of Runway Traffic Flow Using Coloured Petri Nets", A. Kovatcs, E. Nemeth and K. M. Hangos, 2005 International Conference on Control and Automation (ICCA2005)

[PN-13] "Airport Capacity: Representation, Estimation, Optimization", E. P. Gilbo, IEEE Transactions On Control Systems Technology, vol. 1, no. 3, September 1993

8 COLLISION RISK MODELS [CRM]

[CRM-1] "Analysis of Long-Range Air Traffic Systems - Separation Standards I", Reich, P. G. (Royal Aircraft Establishment), Journal of the Institute of Navigation, Vol. XIX, 1966

[CRM-2] "Analysis of Long-Range Air Traffic Systems - Separation Standards II", Reich, P. G. (Royal Aircraft Establishment), Journal of the Institute of Navigation, Vol. XIX, 1966

[CRM-3] "Analysis of Long-Range Air Traffic Systems - Separation Standards III", Reich, P. G. (Royal Aircraft Establishment), Journal of the Institute of Navigation, Vol. XIX, 1966

[CRM-4] "Criterî di Separazione in Volo fra Aeromobili: Valutazione del Rischio di Collisione Attraverso Simulazione Numerica", B. Ricci, Tesi di Laurea in Ingegneria Aerospaziale, Marzo 2003

[CRM-5] "Funzioni di Distribuzione NON-Convenzionali nella Valutazione del Rischio di Collisione in Volo", ATTI dell'Istituto Italiano di Navigazione N177 – U. Grasselli, B. Ricci

[CRM-6] "Il metodo Monte Carlo per Simulare gli Errori di Navigazione nella Valutazione del Rischio di Collisione in Volo", to be published by previous journal – U. Grasselli, B. Ricci

9 STATECHARTS [SC]

[SC-1] "Statecharts: A Visual Formalism For Complex Systems", Harel, Department of Applied Mathematics, The Weizmann Institute of Science, Rehovot, Israel, 1987, Elsevier Science Publishers B.V. (North-Holland)

[SC-2] "On Visual Formalisms", Harel, Communications of the ACM, May 1988 Volume .31 Number 5

[SC-3] "The STATEMATE Semantics of Statecharts", D. Harel, A Aamad, ACM Transactions on Software Engineering and Methodology, Vol. 5, No. 4, October 1996, Pages 293–333

[SC-4] "Statecharting Petri Nets", Eshuis, Eindhoven University of Technology, Department of Technology Management, 2005

[SC-5] "Hierarchical Abilities Of Diagrammatic Representations Of Discrete Event Simulation Models", Ceric, Proceedings of the 1994 Winter Simulation Conference

10 PROGRAMMING [PR]

[PR-1] "Never Make the Same Mistake Twice- Using Configuration Control and Error Analysis to Improve Software Quality", James M. Haugh, IEEE AES Systems Magazine, January 1992

11 STOCHASTIC DIFFERENTIAL EQUATIONS [SDE]

[SDE-1] "Stochastic Differential Equations: An Introduction with Applications", B. Oksendal, 3rd Ed., Springer-Verlag, New York 1992

INTELLECTUAL PROPERTY (PATENTS)

"Sistema di Controllo del Traffico di Flotte di Veicoli Terrestri entro un'area di dimensioni adeguate, generalmente non servita da reti di telecomunicazioni a copertura commerciale, per applicazioni tipicamente 'stand-alone' (e.g. in un'area aeroportuale)", Ref. no.RM2007A000157, patented.