AD-A014 513

A SURVEY OF MODERN AIR TRAFFIC CONTROL

VOLUME II

Advisory Group for Aerospace Research and Development

PREPARED FOR

North Atlantic Treaty Organization

July 1975

DISTRIBUTED BY:

NTIS

National Technical Information Service
U. S. DEPARTMENT OF COMMERCE

Best

Available

Copy

REPORT DOCUMENTATION PAGE

1. Recipient's Reference	2. Originator's Reference	3. Further Reference	4. Security Classification of Document
	AGARD-AG-209		UNCLASSIFIED

5. Originator

Advisory Group for Aerospace Research and Development
North Atlantic Treaty Organization
7 rue Ancelle, 92200 Neuilly sur Seine, France

6. Title

A Survey of Modern Air Traffic Control

7. Presented at

8. Author(s)	9. Date
Various Editor André Benoît	July 1975

10. Author's Address	11. Pages
Various	[illegible]

12. Distribution Statement

This document is distributed in accordance with AGARD policies and regulations, which are outlined on the Outside Back Covers of all AGARD publications.

13. Keywords/Descriptors			14. UDC
Human factors engineering	Collision avoidance		656.7.05:629.73.05
Ground support equipment	Automatic control	Reviews	
Artificial satellites	Navigational aids	Computers	
Air traffic control	Aircraft landings	Guidance	
	Air navigation		

15. Abstract

Thirty-four papers are collected in two volumes, covering the following main sections: (1) General Organization; (2) Human Aspects; (3) Automation of Control Procedures (A = Principles and Applications of Automation; B = On-Board and Ground Based Collision Avoidance Systems; C = Flow Control Techniques; D = Aircraft Trajectory Predictions; E = Centerline Spacing); (4) Technical Aids to Air Traffic Control (A = Ground Based Navigation Aids; B = Self-Contained Navigation Aids; C = Landing Guidance Systems; D = Surveillance; E = Visualization; F = The Computer and Processing Facilities; G = The Satellite); (5) Operational ATC Systems.

This AGARDograph was prepared at the request of the Guidance and Control Panel of AGARD-NATO.

AGARD-AG-209-VOL. II

AGARD-AG-209-VOL. II

AGARD

ADVISORY GROUP FOR AEROSPACE RESEARCH & DEVELOPMENT

7 RUE ANCELLE 92200 NEUILLY SUR SEINE FRANCE

AGARDograph No. 209

A Survey of Modern Air Traffic Control

Volume I

NORTH ATLANTIC TREATY ORGANIZATION

DISTRIBUTION AND AVAILABILITY
ON BACK COVER

AGARD-AG-209
Volume II

NORTH ATLANTIC TREATY ORGANIZATION

ADVISORY GROUP FOR AEROSPACE RESEARCH AND DEVELOPMENT

(ORGANISATION DU TRAITE DE L'ATLANTIQUE NORD)

AGARDograph No.209

A SURVEY OF MODERN AIR TRAFFIC CONTROL

André Benoît
Program Director and Editor

VOLUME II

This AGARDograph was prepared at the request of the Guidance and Control Panel of AGARD-NATO.

THE MISSION OF AGARD

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

- Exchanging of scientific and technical information;
- Continuously stimulating advances in the aerospace sciences relevant to strengthening the common defence posture;
- Improving the co-operation among member nations in aerospace research and development;
- Providing scientific and technical advice and assistance to the North Atlantic Military Committee in the field of aerospace research and development;
- Rendering scientific and technical assistance, as requested, to other NATO bodies and to member nations in connection with research and development problems in the aerospace field;
- Providing assistance to member nations for the purpose of increasing their scientific and technical potential;
- Recommending effective ways for the member nations to use their research and development capabilities for the common benefit of the NATO community.

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

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

A large part of the content of this publication has been reproduced directly from material supplied by AGARD or the authors; the remainder has been set by Technical Editing and Reproduction Ltd.

Published July 1975

656.7.05:629.73.05

Printed by Technical Editing and Reproduction Ltd
Harford House, 7–9 Charlotte St, London, W1P 1HD

CONTENTS

VOLUME I

Page

PART IV – TECHNICAL AIDS TO AIR TRAFFIC CONTROL

Preceding page blank 393

PART IV-A – GROUND BASED NAVIGATION AIDS

Preceding page blank

PRINCIPLES OF RADIOLOCATION

by

Dr W.Crone, Scientific Advisor
G.Peuker, Department Head
Standard Elektrik Lorenz AG
Stuttgart, Germany.

TABLE OF CONTENTS

PRINCIPLES OF RADIOLOCATION

by

Dr W.Crone, Scientific Advisor
G.Peuker, Department Head
Standard Elektrik Lorenz AG
Stuttgart, Germany

Summary

The chapter handles the radio location basic principles from the physical point of view, furthermore frequency and propagation problems as well as errors and error correction methods.

1. Introduction

Radio systems are particularly suited to deliver data required for navigational guidance of a vehicle. Numerous problems of position-fixing and navigation cannot be solved simpler and better by any other aid. Depending on the type and combination of the systems employed, various data can be obtained: distance between vehicle and a ground-based system; difference between two distances; angle between a reference surface and a surface or straight line intersecting the center of the ground system; difference from a selected angle. If two (or, in space, three) independent reference data are available, they can be used to compute the coordinates of the aircraft's position. In addition, the ground speed and the altitude above ground can be measured.

Radio systems used in communication are designed to supply the same information at any location within a given area. Radio systems used for position fixing, however, should provide information that differs from location to location and should provide data associated to the instantaneous position of the vehicle. The coverage of a radio location system (NAVAID) is therefore not determined by the range in the usual sense, but by the usefulness of the navigational information.

The quantities characterizing a radio field and capable of being measured are its amplitude, phase and frequency. The change of one or more of these quantities with the position in the radio field constitutes the basis of any radio navigation system. The distribution of these field quantities in space is called the amplitude-, phase-, or frequency-pattern. The properties of such patterns can be varied in wide limits and adapted to special requirements. Radio methods employing this basis can be conceived as transmitter- or receiver-navaid methods. In the first case, a transmitter produces a radio field that contains, at any point of the field, the location-fixing data related to this point. This data can be received and processed by any number of vehicles simultaneously. In the latter case, the vehicle transmits a signal identifying the vehicle, but containing no position information; this signal is received and processed by an equivalent ground system. The result of the evaluation can be transmitted to the vehicle. In this case, only one vehicle can be served at a given time. A typical example is the ground direction finder. The transmitter-navaid and receiver-navaid methods are on principle equivalent. The choice of the one or the other method depends on the general problem, operational requirements and technological feasibilities. There are cases, however, where a reversion is a priori precluded. An example are the long-distance methods where it is impossible to install antennas and transmitter power in an aircraft required to cover the large range. On the other hand at some navaids, which originally were designed as a pure transmitting system, the requirements on the airborne receiver became more and more stringent in the course of time. Hence more advantages may be seen in the conversion to a receiving system, where the groundstation is equipped with high-grade receiving and signal processing units, whereas the installation aboard the aircraft is reduced on a relatively simple transmitter.

For obvious reasons the trend in aviation will always be to keep the airborne equipment as compact as possible.

---

* The contributions "Principles of Radio Location; Omega, Loran; Very High Frequency Omnidirectional Ranges; Distance Measuring Systems, TACAN" come from the book "Funksysteme für Ortung und Navigation und ihre Anwendung in der Verkehrssicherung" (Radio Systems for Location and Navigation and their Application in Traffic Control) by Ernst Kramar, editor, translated into English with permission of the publishing house Verlag Berliner Union / W. Kohlhammer GmbH, Stuttgart.

The radio coordinates of a vehicle measured with a radio system should coincide with the true position coordinates referred to the system. The definition of the radio coordinates is based on the assumption that the electromagnetic waves propagate without disturbance under free-space conditions. This assumption is not always valid in the vicinity of the earth surface where the radio waves are subjected to various influences as diffraction, refraction, absorption, different ground constants, fluctuations of the earth magnetic field and of the state of the various ionospheric layers as well as reflections by obstacles. These phenomena highly depend on the frequency. Reduction of such interferences has been the objective in system development from the beginning of radio navigation.

The frequencies employed range from about 10 kHz to over 10 GHz, i.e. a span of over 6 powers of ten. The propagation properties of radio waves over such a wide range are in every respect so different that their knowledge is of fundamental importance in the design, planning and application of radio systems. Even the most sophisticated system cannot produce better results than the laws of propagation allow.

In the following the basic systems will be discussed assuming ideal free-space propagation for simplicity. It will be shown that the large number of radio systems that have come into operation in the course of many years are actually not more than modifications of a small number of basic principles.

Then the most important properties of electromagnetic waves will be summarized. Finally, methods for error reduction will be described, especially with reference to errors generated by multipath propagation and reflection by obstacles.

2. Basic Principles

Radio navigation employs a small number of basic methods that are used in numerous modifications. Their operating principle consists in measuring specific quantities that will now be described.

Travel Time

The signal (for instance, a pulse) radiated from a source at the time $t = 0$ propagates as a spherical wave with the velocity of light c and arrives at the destination after having travelled a path r during the travel time $t = r/c$. The distance is computed from

(a) measurement of the one-way travel time. This requires accurate knowledge of the instant when the signal is emitted. This is called a one-way method;

(b) measurement of the round-trip travel time of the signal. Employed for the return trip of the signal are either active devices (transponders with known delay; DME) or passive devices (reflecting objects; radar). These methods are called two-way methods.

(The value for the propagation velocity [13, 14, 15] in a vacuum concluded internationally is $c = 299792.50$ km/s.)

Propagation Time Difference

The travel time difference between two signals transmitted from two points A and B is either measured simultaneously or with a known delay. The locus of equal time difference thus obtained is the revolution hyperboloid with the focal points A and B.

The distance d is also called the base of the system. An example for a travel time difference method is LORAN.

Phase

The distance can also be measured in multiples or fractions of the wavelength λ. In this case, the travel time measurement can be substituted by phase measurement. One wavelength can be expressed in terms of 360° or 2π. This allows an extremely accurate distance measuring but the penalty is the ambiguity because the same phase angle is repeated after every wavelength λ. To eliminate ambiguity, sometimes the travel time is used for coarse measurement and the phase for fine measurement. Of course, the envelope of the pulse employed for travel time measurement should be locked to the carrier RF used for the fine measurement. An example is LORAN-C. Coarse/fine methods may also be employed to remove ambiguity by first obtaining an unambiguous measurement with a larger wavelength and then fine measurement with a shorter wavelength.

Phase Difference

The most important and most frequent application of the phase measurement is in the phase-difference method:

The in-phase RF signals emitted by two radiating sources A and B arrive at a destination having the distance r_A from A and the distance r_B from B with the phase difference.

As in the case of the travel time difference, the locus of equal phase difference (ϕ = const) is the revolution hyperboloid, both radiators being positioned in the focal points. The hyperbolas can be substituted by their asymptotes for large distances where r_A and r_B are large compared with the spacing d of the two radiators. The phase difference then obtained is where the solid angle formed by the base of both radiators and the straight line interconnecting the center of this base and the point of reception is. This term, or better its differential quotient illustrates that small changes of the bearing $d\eta$ result in large changes of the phase angle provided that the base referred to the wavelength d/λ is made large enough.

As phase measurements do not require much effort, very accurate navigational data can be obtained even with simple equipments when using wide-base systems. It should also be noted that the phase gradient $\frac{d\phi}{d\eta}$ is greatest in the direction perpendicular to the base ($\eta = 90^o$) and decreases with η by a sine law.

To measure the phase difference, it is necessary to separately receive the signals of A and B stations. To implement this separation, the signals are transmitted either in time sequence, the phase of the first signal being suitably stored (e.g. LORAN, OMEGA) or by different carrier frequencies that can be derived from a common fundamental frequency. In the latter case, phase measurement is obtained on a common reference frequency after frequency multiplication (example: DECCA). (The systems here mentioned in parantheses are explained in chapter 5.)

Phase-reference methods become ambiguous if the base exceeds one half of the wavelength of the frequency employed (reference frequency). This is so because the change of the propagation path by one full wavelength λ causes the phase angle to shift another full 360^o. This ambiguity can be resolved with a coarse measurement system and an additional measurement with a smaller base d or a larger reference wavelength (DECCA). The ambiguity can also be eliminated by combining travel time and phase measurement methods (e.g. LORAN-C).

Amplitude

The amplitude methods exploit the directivity of antennas or antenna arrays. The distribution of amplitudes in space results from the antenna radiation pattern of the individual antenna element and the group pattern of the antenna array. The properties of the group pattern are controlled by the geometric arrangement of the antenna elements and by the amplitude and phase of the current in each element.

Measuring of the amplitude alone does not give any indication on the bearing. Various methods are therefore used to derive the bearing data from the amplitude patterns:

(a) detection of the minimum or maximum of a directional pattern by mechanical or electrical rotation or swiveling of the pattern. This method can be employed at the receiver or transmitter side (direction finders, radar, modern landing systems). The sharper the minimum or the smaller the beam width in the maximum, the greater is the angular resolution of the measuring system.

(b) Physical or electrical rotation of a directional pattern with constant angular velocity.

This results in an amplitude modulation of the transmitted or received RF carrier; the bearing information is then contained in the phase of the amplitude-modulated signal. The phase angle is measured to a reference signal that is independent of the bearing. This is the principle of the rotating radio beacon (VOR, TACAN). In the case of a single-lobe pattern, the phase angle α of the modulation exactly corresponds to the bearing angle ϑ. In a multi-lobe directional pattern with m lobes, however, $\alpha = m \cdot \vartheta$. The accuracy generally increases by the factor m (fine bearing), but at the expense of unambiguity (see Fig. 1).

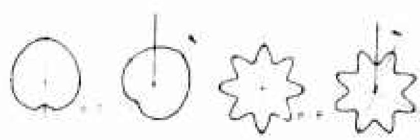


Fig. 1 Rotation of a directional pattern

(c) Intersection of two diagrams

A simple method that is unsurpassed to this day is the keying of two patterns so that the one supplies a short and the other one a long signal that interlace to form a continuous wave for a certain bearing. Any lag of perfect interlacing can be interpreted as a deviation to the left or to the right. This method is mainly used for guidance in a fixed direction (equisignal methods).

(d) **Generation of bearing-dependent modulation-depth diagrams**

Here a first, modulated carrier pattern is radiated in addition to a second, different directional pattern that contains only the sideband frequencies of the modulation. The superposition of both results in a bearing-dependent depth of modulation. If the mirror image of this method is employed for a second modulating frequency then the difference between the depths of modulation provides a useful measure for the left/right deviation from a fixed direction. This method is used, for instance, in the ILS or instrument-landing system.

(e) **Exploitation of a multi-lobe pattern generated from two radiators with large d/λ.**

Here the resulting field is called the interference field (Fig. 2).

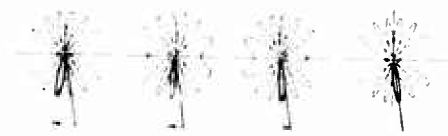


Fig. 2 Change in field distribution of an interference field

To resolve ambiguity, a coarse system with small d/λ is used in addition to the fine system, with intermediate stages (coarse/fine measurement) if required. Interferometers can be swiveled mechanically. Electrical swiveling is not feasible because the field distribution is hyperbolic. However, the field pattern can be changed. In this case, the positions of maxima and minima will be changed, but not the symmetry axis of the pattern (Fig. 2). An example is the CONSOL method. In the case of non-rotatable, stationary interferometers as used for surveying of satellite trajectories or in radio astronomy, the radiation from the satellite or the radio star is recorded and the minima or maxima encountered are analyzed.

Frequency

If a transmitter having the frequency f_o moves with the velocity v towards or away from an observer, the frequency f_o received suffers a doppler shift $f_D = \pm (f_o/c)v$. The doppler effect permits to derive the ground speed of an aircraft from the frequency shift between the radiation obliquely directed to the ground and that reflected from the ground.

The same principle may be used for bearing measurement: if a radiator is moved along a straight line with the velocity v, a stationary observer having the angle η to the direction of motion will receive a frequency that is shifted by the amount $\frac{f_o}{c} v \cos\eta$ against the frequency of the radiator; in other words, it is bearing-dependent. In actual practice the radiator periodically moves along a path of finite length and the motion is electrically simulated (DOPPLER-ILS). Depending on how the measurement is evaluated, either frequency/frequency-deviation patterns or phase/phase-deviation patterns are used.

If the radiator is allowed to move along a circular trajectory, a bearing-dependent frequency modulation is obtained. The bearing information is contained in the modulation phase (DOPPLER-VOR). All methods operating with directional frequency pattern have the important feature of remaining unambiguous even when d/λ is large because there is no splitting into many lobes.

Summary

The description of the basic radio navigation methods shows that the position coordinates can be obtained in various ways. The method employed for a special case will depend on technological feasibilities, operational and economic aspects, and user specifications. Involved are aspects as range coverage, number of simultaneous users (capacity), band-width, number of channels, information rate, monitoring and others.

Propagation Problems

An undisturbed propagation [1, 9] has been anticipated in the discussion of the basic principles. Undisturbed propagation means that the electromagnetic wave propagates from the source in all directions homogeneously, i.e. along straight lines with the velocity of light and constant polarisation. These conditions apply only to free space; on the surface of the earth and in the high-altitude atmosphere, however, disturbed conditions have to be expected. Almost any radio-navigation system has to take this fact into account. Not only the range, but also the accuracy is dictated by the laws and anomalies of the wave propagation. The large number of suggested, implemented and abandoned systems may be explained by the manifold efforts to provide useful position information and to improve them. All these systems have been modifications of the basic principles described.

Table 1 lists the frequency bands allocated for radio navigation and traffic control services. The most important properties of the various frequency groups will briefly be discussed in the following paragraphs in as much as they are significant for the performance of radio navigation systems.

Table 1 Allocated Frequency Bands

Frequency Band	Radio Navigation Systems
10-14 kHz	OMEGA
70-130 kHz	DECCA (DECTRA), LORAN-C/D
190-375 kHz	CONSOL
255-415 kHz	aeronautical and maritime radio beacons
1750-1950 kHz	LORAN-A
73.8-75.2 MHz	aeronautical marker beacons
108-118 MHz	ILS localizer, VOR, D-VOR
118-136 MHz	direction finders, civil (communication)
150/400 MHz	navigation satellite TRANSIT (NNSS)
225-400 MHz	direction finders, military (communication)
328-335 MHz	ILS glide path
960-1215 MHz	TACAN, DME, secondary radar
600 MHz, 1300 MHz, 2.8 GHz, 10 GHz, 15.5 GHz, 38 GHz	various radar systems as surveyance, airborne and shipborne radar, weather radar, airfield radar
1540/1650 MHz	Navigation satellite DIOSCURE
1558-1636	future collision-avoidance systems (CAS)
5.0-5.25 GHz	
15.4-15.6 GHz	future landing systems
440 MHz, 1630 MHz, 4.3 GHz	radio altimeters
8.75-8.85 GHz	
13.25-13.4 GHz	Doppler navigator

The free-space attenuation along a radio path r is $\alpha = \lg \frac{4\pi r}{\lambda}$ dB if a isotropic radiator is used with transmitter and receiver. The range therefore increases with the wavelength λ , assuming a defined ratio of transmitter-to-receiver power. For the radio link on the earth surface, the propagation laws are very complex. On the whole it may be said that the radio-location frequency employed should be the lower, the larger the area to be covered is, especially since the receiver sensitivity generally decreases with the increasing frequency. A global range can only be achieved with very long waves (disregarding the short waves which are hardly suitable for location). Range problems have been widely discussed in the existing literature; special reference is made to the CCIR propagation characteristics [1].

The use of radio waves for location presents less and less problems the more their characteristics resemble optical frequencies. Only the introduction of quasi-optical frequencies exceeding 30 MHz in the "thirties" has raised the classic methods to the level of modern radar technology.

Let us start with the very long waves.

Very-low Frequency (VLF, 3-30 kHz).

The propagation is conceived today as taking place in a waveguide having the shape of a spherical shell formed by the earth surface and the D layer of the ionosphere (height about 60 km). The attenuation and velocity of a wave propagating in a waveguide depends on the waveguide dimensions for a certain wavemode and a given frequency. The effective height of the D layer depends, however, not only on the frequency, but on numerous other factors: time of the day, season, latitude, sun activity, and direction of propagation referred to the terrestrial magnetic field. Significant is also the surface conductivity of the ground over land and over sea. The phase velocity of the wave is thus ruled by a large number of influences. The position-line criterion in this frequency range is usually the phase difference of two waves. These waves propagate around the earth along completely different paths. Obviously useful results can be expected only when the propagation phenomena are sufficiently known. This problem is solved by continuous monitoring and by using corrections for diurnal and seasonal forecasts of the propagation anomalies to be expected in the area concerned. Because atmospheric disturbances for very long waves are considerable, especially in tropical territories, extremely powerful transmitters are required.

Low Frequency (LF 30-300 kHz) and
Medium Frequency (MF 30-3000 kHz).

This frequency range is characterized by two wave components of quite different behaviour that may interact: the groundwave and the skywave.

The groundwave is the wave propagated along the surface of earth. The analysis starts from the limiting case of grazing incidence of a wave propagating in free atmosphere; the phase of this wave is influenced by the complex refractive index of the ground. The wave is thus additionally attenuated and the phase delay is increased. Both effects depend on the ground constants (ground conductivity and dielectric constant) as well as on polarization. The attenuation is a minimum over sea and a maximum above low-conductivity ground. Accordingly affected is the phase delay. The groundwave of this frequency range is actually highly suitable for accurate position measurements over medium distances (500-1000 km over ground, 1000-2000 km over sea, depending on the frequency); however, problems are encountered by the different propagation properties over changing ground and transition from land to sea, especially when the base (spacing between the ground stations) is large and propagation takes place over different topography ground. Though even less than in the case of VLF, the atmospherics are still considerable in some regions. The skywave is the radiation reflected by the ionosphere. Its field strength greatly fluctuates in daytime, but is only about one tenth of the strength at nighttime. Compared with the groundwave, the skywave is hardly noticeable in daytime up to distances of about 500 km as compared with the groundwave. In contrast, the groundwave is disturbed by the skywave in nighttime even at distances of less than 100 km. The skywaves have single-hop and multihop reflections. Their quantities (amplitude, phase, travel time, polarization) are subjected to fluctuations. Without the groundwave, the skywave can be used for location purposes only to a limited extent and at reduced accuracy. For radio-location systems in this frequency range, the appearance of the skywave will reduce the useful range unless special measures are employed to process the groundwave signals before the skywave signals are received (e.g. pulses in LORAN).

For these reasons (pure skywave), the short wave range (HF, 3-30 MHz) is not very useful for location purposes. Here no frequency is allocated.

Very-high Frequency (VHF, 30-300 MHz)
Ultra-high Frequency (UHF, 300-3000 MHz)
Super-high Frequency and higher (SHF, 3-30 GHz).

In this frequency range the propagation is quasi-optical and the more so, the higher the frequency is. The useful range along the earth surface is quasi-optical assuming 4/3 of the actual radius of earth in order to take into account the diffraction effect on the surface. The radiation will reach the area beyond the horizon by diffraction and refraction. The decrease of energy behind the horizon is the more rapid the higher the frequency is. For grazing incidence similar aspects apply as to the groundwave in the long wave range (effects of ground constants and polarization). No reflection by the ionosphere takes place.

The main field of application for these frequencies are in the space above the earth surface. The radiation from a source above ground has two components: a direct wave propagating just as in free space and a wave reflected by the ground. Both are combined vectorally. Amplitude and phase of the reflected wave depend on the ground reflection coefficient and on the path difference. The amplitude and phase of the reflection coefficient depend on the ground constants, the angle of incidence, and polarization [2]. The existence of the ground is thus the cause for the free-space radiation pattern of an antenna to be split into numerous lobes, the shapes of which depend on the height of the antenna and the other above mentioned factors. The field strength is therefore depending on the elevation angle; for a given elevation angle, however, it obeys the law of free-space propagation. An aircraft at great distance is generally located on the slope of the first lobe; here the field-strength increases with the flight altitude over ground. For a given transmitter power and receiver sensitivity, the range basically depends on the altitude of the observer. The statement of a distance alone is therefore insufficient to specify the range of a system. The ground reflection is used in some methods, e.g. the ILS glide path, to obtain a measurable quantity that depends on the angle of elevation. However, problems may be encountered through irregularities of the ground, variable with time. The higher the frequency used, the greater is the risk of disturbing reflections because even small obstacles may have the dimension of several wavelengths. On the other hand, it is easier to apply methods in the higher frequency ranges suitable to avoid or reduce reflection disturbances, namely high directional patterns, short pulses, and wide-base methods.

It should be noted that even these frequencies do not remain unaffected along their path through the atmosphere. Thus atmospheric layers may change in their dielectric constant and cause the beam direction to be bent. Finally, a rapid increase of absorption by fog, rain etc. [12, 13] has to be expected for frequencies exceeding 15 GHz.

A substantial advantage as compared with longer waves is the complete absence of atmospherics. Moreover, these frequencies have ranges highly suitable for radio communication with spacecraft because they penetrate the ionosphere. Hence, wide areas of the earth can be covered by satellites and large-area navigation can be performed at very high frequencies (navigation satellites).

An important factor for propagation conditions is the polarization of the electromagnetic wave. This polarization may be horizontal, vertical, or mixed-circular. The choice of polarization for a certain system will depend on the propagation conditions, the desired field distribution in space (antenna patterns), and problems in the implementation of ground and airborne antennas. Vertical polarization is primarily used for low frequencies, and various polarizations for high and very high frequencies.

3. Errors and Error Reduction Methods

The surfaces in space where the quantities used for location are of constant value, are called position surfaces or radio position surfaces.

When systems on the ground are used, the intersection of the earth surface with these position surfaces results in the position lines. Such position lines may be straight lines, circles, or hyperbolas. Often the position lines are also called the radio coordinates. A feature particularly important for the user of a radio location system is the magnitude of the error resulting from measuring on a position surface.

In the case of a stationary observer, two major error sources can be identified: equipment errors and errors caused by non-ideal propagation conditions (anomalies, reflections).

Equipment errors are generated by non-ideal technological properties of the system components (antenna, receiver, processors); they greatly depend on the state of the art. For a given magnitude of the equipment error, the error affecting the determination of a position surface will depend on the change of the field quantity measured in space (i.e. on the gradient of the field quantity). For a given gradient, the equipment errors determine the limiting accuracy aptly called the resolution. It defines the lowest limit where a system can discriminate between two position surfaces. If the equipment error is a minimum according to the state of the art, the resolution can be improved only by increasing the field quantity gradient, i.e. higher directivity, higher number of interference lines, shorter pulses and increased frequency shift.

Additional errors in the position surfaces determination are caused by the non-ideal propagation conditions. The magnitude of these errors greatly depends on the method employed. The user of a radio-location system is now interested in the overall error he has to consider in position surface determination. If radio-coordinates are derived from two or more position surface lines, the error of the position-fix can be determined.

The overall error should be determined with great care and observing certain prerequisites. All errors may be composed of systematical (i.e. constant) errors and statistical errors. If a spread is specified (generally the σ-value of the gaussian distribution) this can refer only to the statistical components. In the true sense of the laws of statistics it is impermissible to derive a σ-value from a single (or few) error diagrams (e.g. difference between theoretical and measured values along a test circle around the system). Furthermore, measurements have to be obtained from a sufficient number of installations, sitings and in various distances, flight altitudes, etc. Since such measurements are extremely expensive and time-consuming, the attempt is often made to compute the σ-value from many individual measurements on subunits. If this is not feasible, an error-statement at most can be made which is valid for only one installation and only under the prevailing conditions. Under these aspects all accuracy statements for a radio location system have to be subjected to a critical review. In most cases, a reliable value can be stated only after a lengthy system operation time period.

Systematical errors may be corrected to a certain degree with the aid of special tables or charts, provided that the correction values are secured by sufficiently extended observations throughout the time and area in question. Inasmuch as this is possible, the errors determined by a stationary monitor can be radioed to the users in this area.

A special source of errors is the multipath propagation. It appears at low frequencies when the skywave signals are superimposed over the groundwave signals. At higher frequencies, it is generated by reflections from the ground, mountains, buildings, vehicles, etc. The disturbing effects caused by multipath propagation are unfortunately conspicuous in this frequency range (which is highly suitable for navigation) because even small obstacles have the dimension of several wavelengths and will therefore act as intensive reflectors. The direct signal is then superimposed by spurious signals coming from various directions, signals which contain wrong information, other delay times, other amplitude and other phases. The resulting locating data can thus be distorted or even be unusable. The following measures can be applied to reduce errors of this type:

Avoid obstacle illumination

This method can be implemented by high directivity transmitter or receiver antenna patterns so that obstacles are not or only partly illuminated by the radiation. In addition, the installation site can be selected so that the number of obstacles in the vicinity of the system is low. Even the surrounding terrain can be prepared accordingly (e.g. ILS).

Use of short pulses

The reflected signals travel a detour path, i.e. they arrive at the point of destination later than the direct signal. If very short pulses or only their leading edges are evaluated the disturbing pulses do not affect the result. This principle may also be applied to all quantities; it is primarily employed for travel time or phase measurements.

Use of wide-base systems

Wide-base systems designate a system with a high ratio of aperture to wavelength d/λ. Under the usual condition that reflected signals are weaker than the direct signal, it can be shown that an extended base d/λ result in a substantial error reduction. The wide-base principle is applicable to all information carriers (amplitude, phase, frequency). A comparison of different systems is only possible on the basis of complex error formulas. It can be shown that all relevant formulas for the error analysis comprise always the term d/λ. Typical wide-base systems are CONSOL, DECCA, DOPPLER-VOR, direction finders [16].

Averaging

An error due to multipath propagation is in itself stationary, i.e. it does not change as long as the geometric constellation system/obstacle/vehicle remains unchanged. For a moving obstacle or vehicle, however, both the amount and the sense of the error will change. The speed of error fluctuation varies with the geometric constellation change referred to the wavelength. If the change is rapid enough, a significant error reduction can be achieved by averaging. Care should be taken to adapt the averaging period of time to the vehicle's speed.

These methods can be implemented with a particularly high effect for the higher frequencies. It is only in this range that high directional radiation patterns can be obtained, that wide-base systems have practicable geometric dimensions, that sufficient bandwidth is available for short pulses, and that the geometric configuration referred to the wavelength varies quickly enough. In some cases it is of advantage to combine several of the above mentioned methods, e.g. to use directional antennas for wide-base systems.

4. Development Trends

The importance of radio navigation is undisputable even for the future. The dedevelopment of radio location systems normally follows the development of the traffic carriers with a certain lag that is indispensable to identify the problems and to define the objectives. In aviation, for instance, the introduction of novel aircraft types in rapid succession, combined with an enormous increase of the traffic density have resulted, among others, in a gradual obsolescence of existing navigation systems. The primary objectives of development today are: radio guidance for fully automatic landing; more powerful radio systems for en route and area navigation and for air-traffic control (ATC); systems for an effective collision avoidance.

The wide field of application of radio systems for en route navigation employing navigation computers and a network of navigation systems will permit flying any course and high-precision area navigation.

The development of one-way distance-measuring methods on the basis of the transmission of pulses synchronized with extreme accuracy and individually associated to time will permit collision avoidance by measuring mutual distances and their changes for a large number of aircraft.

Finally, efforts will be needed to reduce the steadily increasing number of airborne units and to integrate as many functions as possible in a single equipment.

5. References

1 Meinke/Gundlach: Taschenbuch der Hochfrequenztechnik, Springer-Verlag, 1968, 3rd edition.

2 Hütte: Band IVB, Fernmeldetechnik, Verlag Ernst & Sohn, München, 1962.

3 Bärner, K.: Flugsicherungstechnik II, Fernmeldeanlagen, Hanns Reich Verlag, München, 1959.

4 Brandt, L.: Funkortungssysteme für Luft- und Seefahrt, Verkehrs- und Wirtschaftsverlag, Dortmund, 1962.

5 Freiesleben, H.C.: Navigation, Matthiesen Verlag, Lübeck and Hamburg, 1957.

6 Stanner, W.: Leitfaden der Funkortung, Deutsche Radar-Verlagsgesellschaft, Garmisch-Partenkirchen, 1957.

7 Kayton, M./Fried, W. (editors): Avionics Navigation Systems, John Wiley & Sons, New York, 1969.

8 Philippow, E.: Taschenbuch der Elektrotechnik, Vol. 3, Nachrichtentechnik, pp. 1288 ff. VEB Verlag Technik, Berlin, 1969

9 Fränz, K., Lassen, H.: Antennen und Ausbreitung, Springer-Verlag, Berlin-Göttingen-Heidelberg, 1956.

10 Bundesministerium für das Post- und Fernmeldewesen: Vollzugsordnung für den Funkdienst, Anlage zum internationalen Fernmeldevertrag, Geneva 1959, German translation, Bundesdruckerei 1961.

11 C.C.I.R.: Atlas of ground-wave propagation curves ..., (U.I.T.) Geneva, 1st edition 1955, expanded 2nd edition 1959.

12 Fengler, G.: Die Struktur der Troposphäre und ihr Einfluß auf die Wellenausbreitung, Vortrag auf der Internationalen Navigationstagung in Hamburg, Oct. 1969, (DGON).

13 McNish, A.G.: The speed of light, IRE Transactions on Instrumentation, J-11 (1962), pp. 139-148.

14 Großmann, W.: Berichte zur XI. Generalversammlung des IKGG, Assoz. für Geodäsie, in Sept. 1957 in Toronto (Canada), Sonderheft 7 der Zeitschr.f. Vermessungswesen (1958), pp. 11.

15 DIN 1357: Einheiten elektrischer Größen, August 1966 edition.

16 Rindfleisch, H.: Die Großbasis-Peilanlage "Wullenweber", Nachrichtentechnische Zeitschrift (NTZ) 1956, Heft 3.

DESCRIPTION OF SOME NAVAIDS

The radio navigation systems described in the following chapters deliver radio coordinates derived from special transmitting NAVAIDS which enable the aircraft to find its location by airborne receives and processors, independently of headings:

(A) – Long Distance Aids

(B) – Medium Distance Aids

Pages 408, 409 and 410 are Blank.

(A) LONG DISTANCE AIDS

(OMEGA, LORAN)

by

Walter Stanner, Scientific Advisor
Standard Elektrik Lorenz, AG
Stuttgart, Germany

Summary

The ground-based long distance aids Omega and Loran provide the user with position data by the hyperbolic principle. Desirable are ranges up to 10,000 km, i.e. one quarter of the circumference of earth. This objective determines the frequency, modulation methods and transmitting powers of the ground stations. Recently, additional to the described aids, a VLF Radio Navigation System appears which employs highly stable transmitters in the lo ... 25 kHz band for precise one-way distance measurements (Interavia 7/1974).

1. Introduction

It has been found that powerful transmitters in the VLF and LF ranges are best used considering the present state of the art and the international frequency allocation. These transmitters radiate position signals in such a way that the user can fix his position after the hyperbolic principle. In this procedure, the line of position is the locus of equal distance differences to a pair of transmitters, this locus being a hyperbola. The locations of the ground transmitters of such hyperbolic system are the focal points of a family of hyperbolas. In the cases of the hyperbolic systems Omega and Loran here described, not only one pair of transmitters, but up to 8 ground stations are linked in their functions for radiation of position signals in time and frequency multiplex. The user can then determine through his airborne receiver several lines of position, the intersection of which indicates the position wanted. Here airborne computers are becoming more and more important.

Although even more complex mathematical functions are involved in long-distance navigation because of the curvature of earth, the following paragraphs will be concerned with hyperbolic lines of position for the sake of simplicity. For position fixing, only a small number of the limitless quantity of loci are used in actual practice; the relationship between measurements and geographic data is provided in the form of tables. In addition, the competent authorities provide suitably prepared navigation charts. The sections of hyperbolas there printed are identified by numbers. It is these numbers (propagation delay differences) that the user reads from his airborne equipment. It should also be noted that the locus takes the shape of a straight line on the interconnecting base line between both transmitters and in the lateral extensions. When the base lines are very short or the distances are very large, the hyperbolas can frequently be replaced by their asymptotes. The phenomenon is utilized in the CONSOL method.

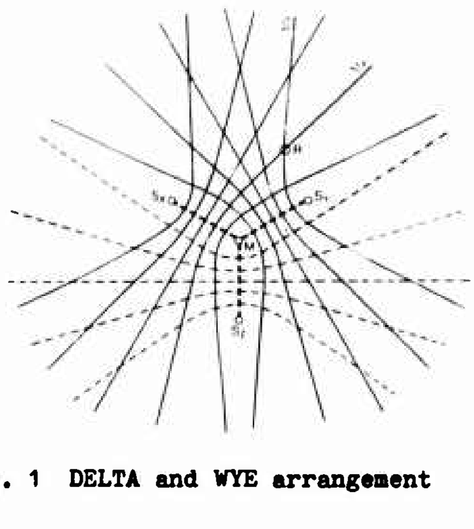


Fig. 1 DELTA and WYE arrangement

The solid hyperbolas correspond to the delta arrangement of the transmitters M, S_X and S_Y. The transmitter S_Z completes the wye arrangement.

Fig. 1 shows an example for the arrangement of a chain of transmitters in the hyperbolic system. The master transmitter M is in the center of the navigation area. It is surrounded in wye shape by the slave transmitters S_X, S_Y and S_Z. Because of the satisfactory area of coverage, the slave transmitters are usually arranged in the form of an equilateral triangle. However, certain objectives or geographic conditions may result in other transmitter configurations [1, 2, 3, 4].

In hyperbolic navigation, the user has a purely passive function. He does not transmit any signal for position fixing. This property is of importance especially for military missions because the user does not reveal his position to the enemy. This method is also never saturated, that is, the number of users is virtually unlimited. The costs for the range and the accuracy can largely be concentrated in the ground stations of the system. Here, however, considerable costs have to be faced for the sites, buildings, installations, maintenance and guarding personnel.

The airborne units are usually required in large quantities and standardized design. Therefore, their costs estimate is relatively modest. The latest models are primarily solid-state versions. They feature high reliability in operation (MTBF); compared with earlier designs, they display not only better navigational performance, but also lower volumes, weights and power consumption.

The airborne receiver analyzes the delay differences of the signals from the individual ground transmitters. This problem can be solved by pulse-delay measurements with the aid of delay networks (LORAN A), by phase measurements (OMEGA, DECCA) or by a combination of both measuring methods (LORAN C). Some sophistication is needed to properly associate measurements with position. In particular, the individual transmitters of a chain should be identifiable even though they are using the same carrier frequency.

2. Omega

The transmitters of the Omega system radiate their position data in the VLF range. Most of the users are ships; however, a sufficient field strength of the position signals may also be expected up to a diving depth of 15 m (50 ft).

The propagation of Omega signals is best described with the model of two spherical shells forming the boundaries of a wave guide [6, 7]. One of the boundaries is the surface of earth and the other is the D layer of the ionosphere about 70 km above ground. In the space enclosed by these two boundaries, a number of modes can be set up and these modes may interfere with each other. Careful investigations of propagation have shown that only a carrier frequency of about 10 kHz is useful for long-distance navigation. For this reason, the frequency of 10.2 kHz was selected for the Omega master carrier.

In 1966, the first transmitters of the Omega system started trial operations. In 1968 everybody was convinced that all expectations can be met and plans for the construction of a total of 8 ground transmitters were implemented for worldwide coverage. Table 2 lists the locations and states of the individual ground transmitters [5, 6, 7].

Table 1

Station chain	Country	Objectives planned or met
A	Norway	ERP increase from 2 to 10 kW (1974)
B	Trinidad	Equipment Substitution (Operable 1975)
C	Hawaii	Equipment Substitution (Operable 1974)
D	USA	Relocation of Transmitter, ERP 10 kW (1972)
E	Madagascar	Construction Phase (1975)
F	Argentine	In Planning Stage (Op. 1975)
G	Australia	(Op. 1975)
H	Japan	Under Construction (Op. 1974)

ERP = Effective Radiated Power

As may be seen from Table 1, only four ground transmitters were in operation on the northern hemisphere, partly with reduced radiation power, in 1974. The full operability of the Omega system with worldwide coverage could not be expected before 1975. Fig. 2 shows the locations of the 8 ground transmitters on the globe; shown are also the LORAN chains discussed later.

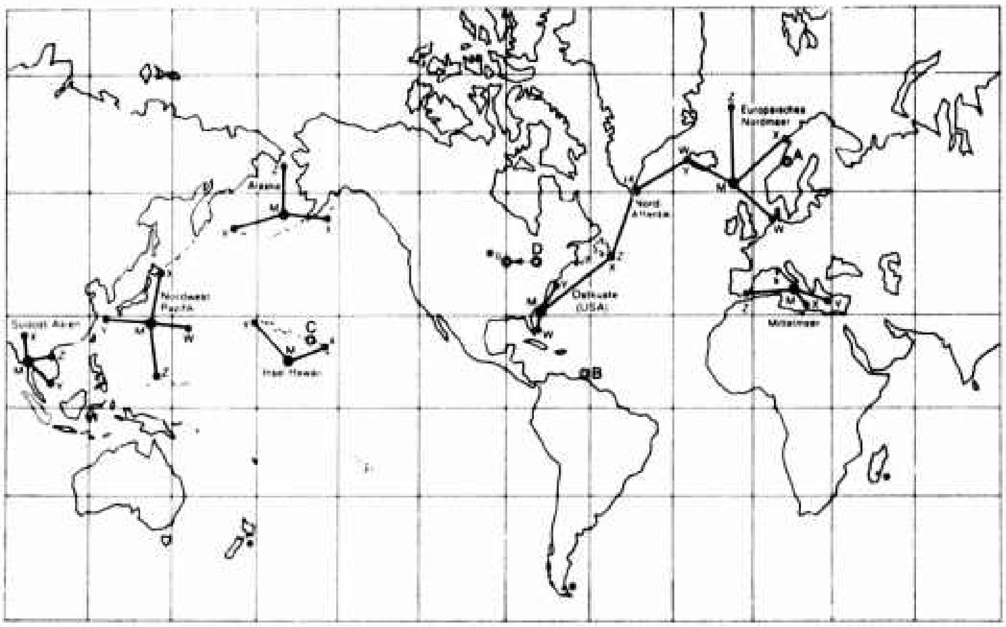


Fig. 2 Distribution of LORAN C and Omega stations on the earth. The existing LORAN C transmitters are identified as black dots, the existing Omega transmitters as double circles and the letters A, B, C and D. The asterisks indicate planned Omega stations.

The Omega signals permit the user to fix his position by the phase measurement principle: since the signals from the transmitters arrive in a sequence, the first signal has to be stored in the receiver for phase measuring against the second signal. This is usually accomplished by synchronizing a highly stable oscillator.

The master carrier for the phase comparison has the frequency f_1 = 10.20 kHz. This frequency corresponds to a lane width of 8 nautical miles (nm) or 14.7 km on the base line of two ground transmitters. The resulting ambiguity of Omega positioning can generally be eliminated by other navigational means, e.g. dead reckoning or astro-navigation. However, even the Omega system itself offers the user a possibility for coarse position fixing. For this purpose, the two subcarriers f_2 = 13.60 kHz and f_3 = 11.33 kHz are radiated. In the airborne receiver, the following difference frequencies are therefrom obtained:

$f_2 - f_1 = f_{1/3}$ = 3.40 kHz lane width 24 nm

$f_3 - f_1 = f_{1/9}$ = 1.13 kHz lane width 72 nm

Thus the lane width is always increased by the factor 3 and will in almost all cases permit unambiguous association of the lanes.

Under discussion are also suggestions to extend the information content of the position signals by additional frequencies [6].

The carriers f_1, f_2, and f_3 are identical for all ground transmitters of the Omega system. For unambiguous identification of the signals from these transmitters in the airborne unit, the signals have to be radiated in a unique sequence by time sharing. For this purpose, a signal scheme with a cycle of exactly 10.0 sec has been chosen. This cycle is broken down in working intervals of 0.9 to 1.2 sec. Provided between the working intervals are idle intervals of 0.2 sec each with a total of 1.6 sec. Table 3 shows the exact Omega signal format. It shows that the ground transmitter A in Norway (see Fig. 2) radiates the master carrier f_1 in the first working interval of 0.9 sec. During the same interval, the ground transmitter G in Australia will radiate the subcarrier f_3 and the ground transmitter H in Japan the subcarrier f_2. The user may make use, for phase measurements, of the hyperbolic lines of position associated with the stations A, H, and G. After this interval of 0.9 sec follows an idle interval of 0.2 sec (not shown in the Table 2) during which all position signals may decay and build-up. This scheme provides an information repetition rate of 10.0 sec which is fully satisfactory for marine use.

The radiation of the position signals is strictly linked to the universal time UTC-2. Thus synchronization is enforced in all 8 ground transmitters many thousand kilometers apart. Transmitter A starts with the carrier f_1 at 00.00 hours UTC-2 with the positive-going passage through zero of its oscillation, initiating the sequence of Table 2 that is then constantly recycled.

Table 2 Signal Scheme of the Omega Transmitters

Ground transmitters	Duration of the individual transmitter working intervals (in seconds)											
	0.9	1.0	1.1	1.2	1.1	0.9	1.2	1.0	0.9	1.0	1.1	1.2
ground transmitter A (Norway)	f_1	f_2	f_3						f_1	f_2	f_3	
ground transmitter B (Trinidad)		f_1	f_2	f_3						f_1	f_2	f_3
ground transmitter C (Hawaii)			f_1	f_2	f_3						f_1	f_2
ground transmitter D (USA)				f_1	f_2	f_3						f_1
ground transmitter E (Madagascar)					f_1	f_2	f_3					
ground transmitter F (Argentine)						f_1	f_2	f_3				
ground transmitter G (Australia)	f_3						f_1	f_2	f_3			
ground transmitter H (Japan)	f_2	f_3						f_1	f_2	f_3		
Radiated frequencies	**Ground transmitters employed for navigation**											
carrier f_1	A	B	C	D	E	F	G	H	A	B	C	D
carrier f_2	H	A	B	C	D	E	F	G	H	A	B	C
carrier f_3	G	H	A	B	C	D	E	F	G	H	A	B

Known is also a wide-range navigation system similar to Omega with transmitters in USSR. Here the carrier frequencies are

$$f_1 = 11.905 \text{ kHz}, \; f_2 = 12.649 \text{ kHz}, \; f_3 = 14.881 \text{ kHz}.$$

One cycle takes only 3.6 sec. This rapid data-repetition rate is valuable for aviation [11].

As to the range of the Omega system for worldwide coverage with position data, reliable reception for 3 lines of position can be expected at any point of the earth surface. Because of the magnetic field of earth, the maximum distances depend on the direction of propagation. The following values are known [1]: N and S: 8,100 nm; W-E: 11,300 nm; E-W: 4,900 nm.

The accuracy of the Omega signals is greatly influenced by the diurnal and seasonal variations of the phase velocity. These variations can be predicted rather accurately for a given user position. They are made available to the user in the form of correction tables. Table 3 examplifies the structure and content of this important aid [6, 7]. The unit of correction is one hundredth of one cycle (1 centicycle) of 10.2 kHz. The tables contain these sky wave corrections (SWC) for each full hour of the day and for every 15 days. Finally, a sequence of tables presents longitudes and latitudes spaced by 4°. The figures of the correction table are indicated with reversed signs by the Omega indicators for correction.

Table 3 Examples for correction tables after Swanson (SWC) to improve the Omega position accuracy

Position of the observer: 0.0 68.0 west
Transmitter received: ground transmitter D

Date	Time of the day (Greenwich Medium Time)																								
	00	01	02	03	04	05	06	07	08	09	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
Jan. 1-15	–56	–56	–56	–56	–56	–56	–56	–56	–56	–56	–56	–32	–6	–3	–1	1	2	2	2	1	–1	–4	–7	–54	–56
Mar. 1-15	–56	–56	–56	–56	–56	–56	–56	–56	–56	–56	–56	–13	–7	–3	0	2	4	4	4	2	–1	–4	–8	–17	–56
May 1-15	–50	–56	–56	–56	–56	–56	–56	–56	–56	–56	–19	–8	–4	–1	2	5	6	6	5	3	1	–3	–7	–12	–50
July 1-15	–36	–56	–56	–56	–56	–56	–56	–56	–56	–56	–14	–2	0	2	4	5	6	6	6	5	3	1	–1	–3	–36

The numerical values of the table in centicycles (about 10^{-6} sec) are true for the master carrier of 10.2 kHz.

In a limited area of navigation, the accuracy of the Omega position fixing can be improved with the aid of a monitor station. The latter will currently check the measurements at a known measuring location, determine the instantaneous optimum correction values, and pass these in a suitable way (usually through VHF) to the users in this area. This procedure is called Differential Omega. The correction values can be passed on in any feasible way, e.g. by radio, telephony, transmission of test tones, etc. [6].

The system accuracy thus depends especially on propagation of conditions. If sufficient time is available for position fixing as on ships, then an accuracy better than 1 nm can be achieved by correction values and combination of different measurements. Generally errors between 0.5 and 1 nm will have to be expected in daytime and doubled values in nighttime. The following rule of thumb applies to most cases [1].

$R_{CEP} = 0.12\,\delta\Omega$ where R_{CEP} is the probable error-circuit radius in nm and dt is the difference in time in μs (all hundredth of the lane width at 10 kHz). (CEP = Circular Error Probability.)

The Omega accuracy can be summarized [1] under three conditions (Table 4):

(a) As differential Omega; relative and reproducible.

(b) By returning of the starting point within one month and the same hour of the day; reproducible.

(c) Using the previously computed corrections for the hour of the day, season of the year, and propagation path; predictable.

Table 4 Summary of Omega accuracy

Time	Condition (A)		Condition (B)		Condition (C)	
	S %	CEP nm	S %	CEP nm	S %	CEP nm
daytime	1-3	0.12-0.36	5.2	0.62	6.7	0.8
nighttime	1-3	0.12-0.36	6.0	0.27	8.1	0.97
transition	1-3	0.12-0.36	6.5	0.78	10.8	1.3

S = Spacing between lanes

Transmitters and receivers

Apart from the propagation conditions, the accuracy of Omega position signals primarily depends on the frequency and phase stability of the individual ground transmitters. The necessary stability of these values is achieved by using in each ground transmitter 4 cesium-atom frequency standards and by according one transmitter the function of a master transmitter that is to phase-synchronize the other transmitters.

The carrier frequencies are generated in the frequency synthesizer in four independent channels that are continuously compared with each other. The carrier frequencies are then passed through the power stages and a power of about 130 kW is available across the transmitter output. The ground transmitters have the AN designation FRT-88, the associated control units have the AN designation FRN-30 (F = fixed ground, R = radio, N = navigational aids, T = transmitting).

To achieve a tolerable antenna efficiency, large-dimension transmitting antennas are required for the Omega ground stations. Where geographic conditions permit, valleys with suitable slopes are preferred for radiator element mounting. The ground transmitter in Japan, however, will receive a 450 m high radiator. Its upper end will carry 16 radial wires. Buried in the ground are 90 wires, each 200 m long. The voltage at the insulator base will have 170 kV.

Because of the very low bandwidth of about 20 Hz, all ground transmitters have to change over the tuning of their antennas for each of the 3 carrier frequencies by automatic means; finetuning means compensate for the effects of wind and weather. The effective radiated power (ERP) of the Omega system is only 10 kW in spite of the costly antenna system [6].

A number of receivers is offered for the analysis of the Omega signals. The prices fluctuate considerably depending on whether the user wants a simple, fully automatic operation or wants to invest his own efforts for operation and analysis. The circuit designers are also offering rather different solutions. Generally the Omega carriers pass through an amplifier and a filter stage and are then translated into the low-frequency range of 1000 Hz. In the course of further amplification, the bandwidth is narrowed down to about 50 Hz. Some designers eliminate the frequency translation and employ tuned RF receivers. In almost any case the Omega signals are processed by integrated circuits with control and tracking devices. Since the signals from the individual ground transmitters are radiated at different times in accordance with Table 2 and have different propagation paths depending on the user's position, the receiver should feature storage means for subsequent analysis of the phase difference. This is usually accomplished by digital circuits [6-10]. As useful position signals have to be expected from 4 ground transmitters, the corresponding circuits should be available in quadruplicate. Widely used are digital control circuits of Type II (velocity/zero-error control circuits).

Some known models of airborne units are: CMA-723 and CMA-719 by Canadian Marconi Co., AN/ARN-99 by Northrop Corp., ORN-101 by Litcom, M2 and MN by Sercel.

Mounting of antennas on ships usually presents no problem. In the case of aircraft, difficulties are frequently encountered because disturbing harmonics of the 400-Hz airborne power supply are picked up. A remedy consists in mounting two crossed iron- air coils in a suitable location to form the antenna [6].

3. LORAN A

LORAN is an acronym for the field of application (Long Range Navigation). The origins of this navigational aid reach back into the year 1941. Then a rapidly operational long range navigation system was wanted for the allied air forces and navies. Selected was a hyperbolic system in the 2-MHz region. Systematic investigations of propagation had shown that both large ranges and satisfactory accuracy for pulse-modulated position signals could be expected in this frequency region considering the state of the art. The decision was probably controlled by the findings that the reflecting ionosphere layer has an unexpectedly stable altitude of 100 $\pm$ 2.5 km in nighttime so that the sky-wave propagation, too, could be exploited for position fixing [12]. Soon trial operations with 100 kW transmitters were started on the Bermudas. The results were uncommonly encouraging and the system was therefore promoted. In spring 1943, the first LORAN chain resumed continuous operation. The meantime running production by the US industry soon yielded the necessary quantities of ground and airborne equipment units. By the end of World War II, 70 transmitters were in operation and about 50,000 airborne and 5,000 shipborne receivers had been delivered [4,12].

Meantime 81 transmitters radiate position signals in the LORAN A format. For the 47 transmitters served by US personnel, $ 250,000 per ground transmitter are required. These transmitters provide a coverage along all coasts of USA. The operating costs of the whole system will amount to $ 600 million during the next 30 years according to present estimates [10]. The basic configuration of LORAN A is a pair of transmitters with cooperating functions, the master transmitter M and the master-synchronized slave transmitter S. Under the pressure of war LORAN A was rapidly implemented without the delta or wye arrangement of Fig. 1 that was later found so useful. Rather, the individual pairs of transmitters were roughly installed along the most important coasts and islands.

The lines interconnecting the master transmitter M and the slave transmitters (S_x, S_y and S_z in Fig. 1) are called the bases of a hyperbolic system. Their length greatly depends on the frequencies chosen and the mission of the system. In LORAN A, initially called the Standard LORAN, the base lines are usually several hundred kilometers long, thus reliably providing sufficient field strength of the ground wave to synchronize the slave transmitter.

For special missions of the allied air fleets over Central Europe, a modification of the Standard LORAN was in use: the Skywave Synchronized Loran System (SS LORAN). Here the base lines are over 2000 km long. Synchronization between master and slave transmitters could be achieved only with the aid of the sky-wave propagation. This method was abandoned after the war [4].

The ground transmitters of LORAN A are dimensioned for a pulse power of 100 kW. However, more powerful output stages have also been developed. The pulse duration of LORAN A is 45 μs , measured between the half-amplitude points [4]; the rise and fall times are 10 μs each. The following operating channels were selected as carriers of the positions signals in LORAN A: 1,950 kHz in channel I; 1,850 kHz in channel II; 1,900 kHz in channel III; and 1,750 kHz in channel IV. Primarily the channels I and II were allocated to the transmitters. In each channel, a substantial number of ground transmitters operate with different pulse frequencies. In the airborne equipment, the individual pairs of transmitters and again the master and the slave transmitter of each pair have to be identified easily and uniquely. This problem was solved very elegantly. To separate the M from the S pulses, the latter follow with an adjustable time offset. This time offset is chosen so that the S pulses are released not before the M synchronizing pulses are released, thus eliminating the ambiguity of the delay measurement referred to the perpendicular to the center of the base line; on the other hand, the S pulses are radiated with another delay by about one half of the pulse repetition time so that a two-trace reading as shown in Fig. 3 is obtained. The M and S pulses thus have exactly identical pulse repetition frequencies and are strictly synchronized; however, they are radiated with a controlled time offset.

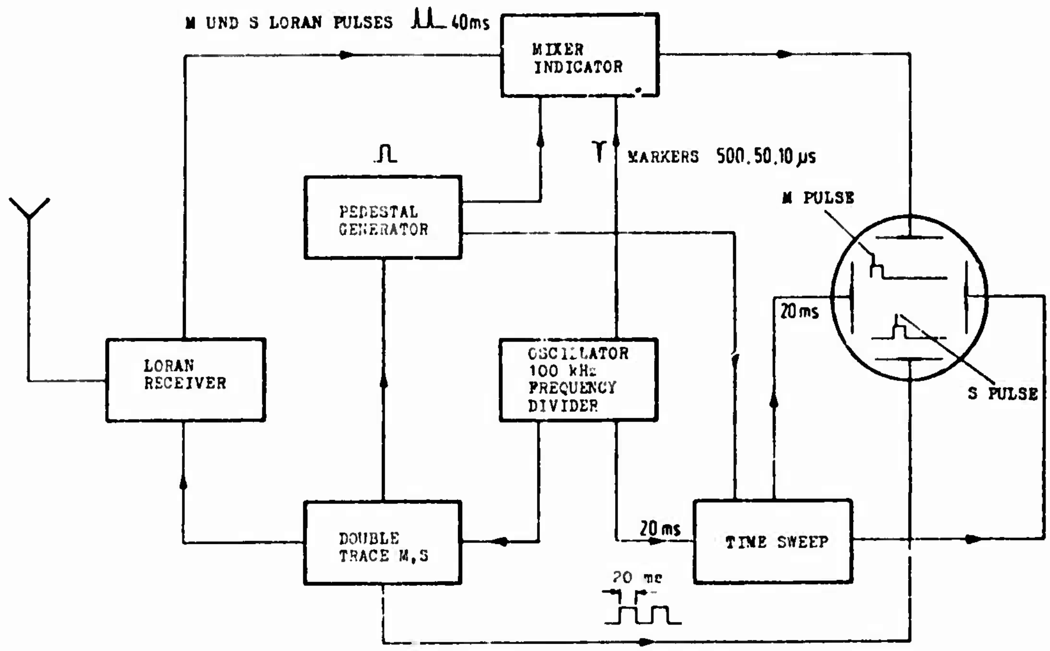


Fig. 3 Principles of the LORAN A analyzing circuits.

SS-0	100.0 ms	SL-0	80.0 ms	SH-0	60.0 ms
S-0	50.0 ms	L-0	40.0 ms	H-0	30.0 ms

Further subdivision and a resulting increase of the number of channels is obtaines by inserting idle intervals of 0.1 ms in the pulse sequence. These additional pulse sequences are identified by the numerals 1 through 7. From the definition it follows that the sequence of the individual pulses in time is exactly fixed, but the pulse-sequence frequencies are fractions. Table 5 shows these relationships. For the LORAN C transmitters operating in the North Atlantic, for instance, the code SL-7 means a super-position of pulses separated by 79.3 ms, which corresponds to a pulse repetition frequency of 12 39/64 Hz.

Table 5 Relationship between pulse separations, pulse frequencies, and the corresponding codes in LORAN charts and LORAN tables

1 Code	2 Separation in ms	3 Frequency in Hz	4 Code	5 Separation in ms	6 Frequency in Hz	7 Code	8 Separation in ms	9 Frequency in Hz
SL 0	80,0	12 1/2	L 0	40,0	25	H 0	30,0	33 3/9
SL 1	79,9	12 33/64	L 1	39,9	25 1/16	H 1	29,9	33 4/9
SL 2	79,8	12 34/64	L 2	39,8	25 2/16	H 2	29,8	33 5/9
SL 3	79,7	12 35/64	L 3	39,7	25 3/16	H 3	29,7	33 6/9
SL 4	79,6	12 36/64	L 4	39,6	25 4/16	H 4	29,6	33 7/9
SL 5	79,5	12 37/64	L 5	39,5	25 5/16	H 5	29,5	33 8/9
SL 6	79,4	12 38/64	L 6	39,4	25 6/16	H 6	29,4	34
SL 7	79,3	12 39/64	L 7	39,3	25 7/16	H 7	29,3	34 1/9

Knowledge of the structure of the individual Loran pulse frequencies facilitates understanding of the functional interworking between the LORAN receiver and the indicator. Here the delay difference has to be derived from the staggered reception of M and S pulses; this difference corresponds to the path difference. Fig. 3 shows the measuring principle used for this purpose. For the sake of simplicity, all events are represented for a pulse-repetition frequency of 25.0 Hz, corresponding to pulse separations of 40.0 ms (code LO).

The adjusting and measuring functions are just the same for the other pulse-repetition frequencies. Experience showed that the versatility of LORAN could best be used with the aid of a CRO display unit. A double trace is written on the CRO screen. When the receiving station is properly adjusted, the M pulse appearing across the receiver output is indicated as a small pedestal at the beginning of the upper trace (Fig. 3). The subsequently received S pulse is offset by the time $t_b + t_c$ where t_b corresponds to the propagation time along the base line while the delay t_c is chosen so that the S pulse is displayed in the second, lower trace of the CRO at each point of reception within the hyperbolic field. In the simplest case, $t_b + t_c$ is the half of the pulse separation time of the M transmitter, in our example 20 ms. The sweep frequency corresponding to the pulse frequency is adjusted on the CRO in compliance with the table data or navigation charts and manually adjusted until the pulses remain stationary. The sweep voltage is derived from a crystal oscillator having the frequency of 100 kHz; the sweep voltage has a stability that is fully satisfactory for the operation. From the sweep stage in Fig. 3 the sweep voltage is applied to the horizontal-deflection plates. In our example, it has the frequency of 50.0 Hz because while the two LORAN pulses are repeated with a separation of 40.0 ms, the CRO beam has to be deflected across the screen every 20.0 ms per trace (including retrace). Moreover, a square-wave voltage in synchronism with the row tracing should raise the beam to the M trace and lower it to the S trace (voltage from the stage "double row M, S" is applied to the vertical deflection plates).

The CRO screen has the function of a memory in this measuring method. Because of the phosphor afterglow, both pulses appear for the observer at the same time they are also received and written in succession. The length of the trace corresponding to the separation of the two pulses is a measure for the difference in propagation time, or the time difference (Fig. 4). The measuring procedure for this difference requires the observer to place an adjustable pedestal under the M and S pulse, respectively, so that the M and S pulses appear seated on the initiation of the pedestal (Fig. 4a).

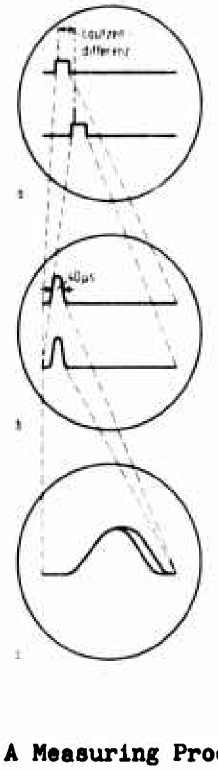


Fig. 4 LORAN A Measuring Procedure

These pedestals with the M and S pulses are now expanded as shown by the dashed lines so that the M and S pulses are expanded too (Fig. 4b). Now the square-wave voltage for the double trace is switched off and the M and S pulses are written in one trace only (Fig. 4c). The leading edges of both pulses are made to coincide. In older equipment units, the time difference corresponding to this shift had to be determined by a separate measuring procedure ("markers" in Fig. 3). In modern equipment, the delay time is measured with a calibrated, variable delay network.

Suitable planning of the carrier and pulse frequency distribution largely avoids the interference of pulses from neighbouring pairs of transmitters in the delay alignment. Interfering pulse sequences that might be displayed on the screen will run across the screen because of their different pulse frequencies and hardly disturb the analysis of the stationary pulses.

The latest equipment models offer numerous simplifications of operation. In particular, the individual circuits are adjusted only once and by automatic means thereafter. Finally, the receivers are designed so that they can receive not only LORAN A pulses, but also LORAN C pulses. The price of older units was about 1000 dollars, that for modern units is up to 10,000 dollars.

The range is very different depending on whether propagation over land or sea is involved, daytime or nighttime conditions apply, ground waves or sky waves are used. The ground wave over the sea has a range of about 700 to 800 nm in daytime, bit only 450 nm in nighttime. The sky wave over sea yields ranges up to 1500 nm. Over land and in daytime, the ground wave has useful field strengths between 200 and 500 nm, at night with the sky wave similar conditions as over sea [1, 13, 14].

In 95% of all cases, a value of 0.2 to 0.6% of the observer distance from center of the base line of the LORAN A transmitters is quoted as the accuracy in determining a line of position. In generalized form, an accuracy of 1 to 5 nm can be quoted [1,10].

4. LORAN C

The LORAN A system aided the allied air force and navy; however, several disadvantages could not be overlooked, especially uncertainties by ionosphere propagation in the 2-MHz region. Therefore, better solutions were wanted even during World War II and it was believed they could be found in the frequency region around 180 kHz. In 1944, orders for experimental systems were issued, but upon completion of WW II all work was discontinued [13].

In the subsequent years, the US air force still demanded a long-wave LORAN system. During extensive experiments in the Arctic, experience on long-wave propagation was gathered from 1946 to 1948 and efforts were made to separate the ground wave from the sky wave in the receivers. Yielding to the pressure of military specifications, especially for a range of over 1000 nm, finally the long-wave hyperbolic LORAN C system was developed. The first chains for military use were installed in the north-east Atlantic and in the Mediterranean area. Now 8 chains with 34 transmitters are in operation, covering about 4% of the surface of earth [10]. Fig. 2 shows the locations of these chains. It may be seen that a master transmitter is often surrounded by 3 or 4 slave transmitters as was also shown in Fig. 1. The distance between the slave to the master transmitters, i.e. the base line of a chain, is between 1000 and 2000 km long. The accuracy, about ten times better than with LORAN A, was achieved by coarse and fine position fixing, a combination of delay and phase difference measurements; the ground wave can be separated from the sky wave by application of pulse technology [1, 2, 10, 12].

LORAN C operates with a single fixed frequency of 100 kHz for all ground transmitters. The internationally settled permissible bandwidth in this range from 90 to 110 kHz is exploited by optimum shaping of the pulse envelope without interfering with the adjacent channels as in LORAN A; the individual LORAN C chains are identified by their pulse frequencies and additionally by pulse-phase coding. The same combinations of letters and figures as shown in Table 5 are employed as codes.

Owing to the fully automatic analysis, the operation is the simplest possible; the chain code, e.g. SL-3, applicable to the area in question, is taken from the charts and adjusted in the receiver; then the delay differences of 2 hyperbolic lines of position appear in the two windows and are looked up in the chart. Their intersection is the instantaneous position of the observer who may also check the quality of the information on a small CRO screen.

System description and signal format

Contrary to LORAN A, the position signals are no longer individual pulses in LORAN C (for discrimination between master and slave transmitters), but pulse sequences. Fig. 5 shows the structure and sequence of the various position signals. Each pulse sequence comprises a number of single pulses following a strict format.

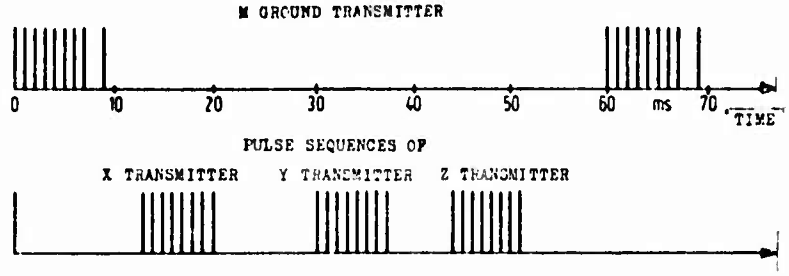


Fig. 5 LORAN C Signal Format

The pulse sequence of a LORAN C slave transmitter comprises 8 pulses at intervals of 1 ms. The sequence duration is therefore about 7 ms. The M sequence is characterized by a ninth pulse following this sequence after an interval of 2 ms (Fig. 5). Hence, the duration of an M sequence is about 9 ms. The phase of each individual RF pulse has a fixed relation to the phase of a reference frequency and thus permits phase coding. These measures are employed to identify the individual M and S transmitters by reliable and automatic means in the airborne equipment. The effect of long sky-wave sequences is eliminated and the accuracy of the ground-wave measuring is ensured under all circumstances. The pulses of a sequence may be phase coded as shown in the following example where + means in phase with the reference frequency and - out of phase:

M: ++--+-+-+ or +--++++++-

S: +++++--+ or +-+-++--

Radiation of 8 pulses in a sequence increases the mean ERP without increasing the peak power of the transmitters. Moreover, satisfactory delay measurements are feasible even in the presence of a low signal-noise ratio: even when the signals are 20 db down referred to the noise, LORAN C position signals are still useful.

This phase coding provides the additional means of transmitting simple data apart from the position signals. Thus the ninth pulse of the M sequence supplies information on the operational reliability of the individual LORAN C chains. If the ninth pulse has the constant interval of 2 ms to the preceding sequence, this hyperbolic chain can unreservedly be used for navigational purposes; if this interval of 2 ms fluctuates because of keying in the ground station, this indicates a disturbance causing the airborne equipment to generate an alarm signal.

The M sequences are radiated every 50 to 100 ms; the S sequences follow at offset, but synchronized times (Fig. 7). Modern receivers permit direct analysis of these position data in spite of the high repetition rate. Storage registers in the receiver store the M data until S data come in. In the ground transmitters, the S sequences are offset in time so that each sequence can properly decay before the next sequence is transmitted. Consideration is given to the sky-wave propagation. Generally the slave transmitter X is offset by 11 ms, the sequences of the Y and Z transmitters correspondingly more (Fig. 5). Some LORAN C chains have 4 slave transmitters (Fig. 2), which results in particularly satisfactory intersection angles of the hyperbolic lines of position. In such cases, the fourth S transmitter is identified by the letter W and its pulse sequence precedes that of the X transmitter [1, 2, 14].

The signal format above described allows a coarse position fixing with high information rate. The desirable fine position fixing is achieved by measuring the phase difference between the individual carrier oscillations of M and S transmitters whose phase relations from sequence to sequence and within each sequence have to be synchronized.

Position fixing requires separation of ground-wave and sky-wave signals in the airborne equipment. In LORAN A this problem was simply solved by observing the CRO screen of the indicator unit. For LORAN C this trivial method is no longer satisfactory. Here a special technique is employed: based on the experience that the sky wave arrives at least 30 μs or 3 full carrier cycles after the ground wave when the carrier frequency is 100.0 kHz, the phase measurement is obtained during this period so that the ionosphere-reflected signals have no effect. For this purpose, a current gate is provided in the receiver and will pass only the first three oscillations of the LORAN C pulse.

The envelope of the LORAN C pulse is tightly controlled. As may be seen from Fig. 6, the leading edge of this pulse increases to its amplitude maximum within 7 oscillations or 70 μs . The exponential decay takes about 20 oscillations so that the complete pulse has the duration of about 270 μs .

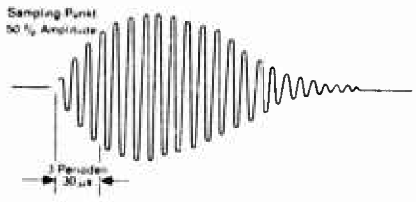


Fig. 6 LORAN C Pulse

The accurately defined leading edge of this pulse has a fixed sampling point after the first 2.75 oscillations of the carrier or 27.5 µs after the start of the pulse. Since the pulse envelope shaped by the transmitter has a turning point that coincides with the sampling point, the latter can be accurately determined in the receiver by differentiation of the leading edge [1, 2, 14].

The accuracy and reliability of the radiated signals have to meet stringent specifications; they are therefore subjected to a control centralized in the M or master ground transmitter. The slave transmitters X, Y and Z receive the signals from M and correct their own signals accordingly.

Ground and Airborne equipment

Two transmitter types are used in the ground stations of LORAN C. In the AN/FPN-44 model, the power stage has a four-tube push-pull parallel amplifier with water-cooled triodes. The pulse peak power is 400 kW, in the sampling point 100 kW. Oil-impregnated paper-mica capacitors are employed to achieve the high transmitter power output. The transmitter antenna is about 190 m high.

The AN/FPN-45 transmitter has a pulse peak power of 3000 kW, in the sampling point 750 kW. It works into a 410 m high antenna [2].

Of course a system offering the user such excellent position data cause for sophisticated receivers in order to exploit all possibilities. True LORAN C position fixing is also feasible with small and even portable receivers; for full evaluation of the signals, however, large-scale receivers are required that are best combined with navigation computers. The latest receivers have solid-state stages and repeated circuits are combined into modules [2, 15]. The only tubes used are in the CRO and in some switching stages. The CRO screen of the display unit is no longer used to continuously adjust and check pulses as in LORAN A; rather, the display unit serves as a means to check the functions of the airborne equipment and to accurately adjust the current gate for the phase-difference measurement. All large-scale receivers are equipped with tracking and tracing circuits so that even the navigator of a supersonic aircraft is continuously supplied with optimum position data in an automatic process requiring no readjustment.

The LORAN C system is somewhat related to the OMEGA system. Common to both is the hyperbolic principle, the use of highly stable frequency standards in the ground transmitters, and the phase measuring for fine position fixing. Therefore the problem has been investigated if it were possible to exploit the advantages of both methods by a single, integrated receiver. Although separate antennas, filters, and amplifiers would have to be used, many units in the analytical stages could be integrated. So far these efforts have only resulted in trial models [6]. Modern LORAN A receivers can also receive LORAN C signals, but make use only of the coarse position signals.

Range and Accuracy

The mean range of the ground wave is 2000 nm over sea and 1200 over land. The accuracy (mean error-circle radius) depends on the chain configuration and the intersection angle of the hyperbolas at the point of observation. It is also affected by the noise level. At a distance of about 300 nm from the M transmitter, the accuracy is 250 nm and at a distance of 800 nm about 1500 nm. Detailed data on range and accuracy have been compiled elsewhere [1].

The advantages of LORAN C are somewhat narrowed down by natural and man-made noise. Since the position signals are also available to military users, the effects of jamming stations have to be taken into consideration. Large-scale receivers contain various bandstop filters for any type of noise and narrow-band tracking servos in combination with the sampling error detectors.

In European space, for instance, the receiving conditions are rather unfavourable where numerous powerful transmitters operate near or in the frequency band of LORAN C (example: Decca). In effect these disturbing transmitters are equivalent to an additional noise of about 20 to 30 db, thus reducing the useful range of the LORAN C ground transmitters. The signal of the M ground transmitter Färöer, for instance (northern European chain, SL-3, 400 kW) provides a reliable range of 1100 nm over the

Atlantic; the same signal is difficult to receive at a distance of only 650 nm in the noise environment of South England [14].

5. References

1 van Etten, J.P.: Fundamentals of Low and Very-low-Frequency Hyperbolic Techniques. Electrical Communication Vol. 45 (1970), no. 3.

2 van Etten, J.P.: LORAN C - System and Product Development. Electrial Communication Vol. 45 (1970), no. 2.

3 Kramar, E.: Hyperbelnavigation - Geschichte und neue Wege, Interavia no. 2 (1969).

4 Stanner, W.: Leitfaden der Funkortung. Deutsche Radar-Verlagsges. mbH, Garmisch-Partenkirchen 1960.

5 Blackband, W.T.: Advanced Navigational Techniques. AGARD Conference no. 28. Technivision Services, Slough, England 1970.

6 Institute of Navigation: Proceedings of the first Omega Symposium, 9-11 November 1971. The Institute of Navigation, Washington D.C.

7 Swanson, E.R.: Omega. Navigation (USA) 18 (1971) no. 2, pp. 168-175.

8 Dickerson, N.C.: Automatic Omega Receiver: Computer-Augmented Approach to Low Cost Marine Navigation. Navigation (USA) 18 (1971) no. 2, pp. 147-154.

9 Heigl, J.: The quiet revolution in radio navigation. Signal 26 (1972) no. 6, pp. 21-24.

10 Pierce, J.A. and Woodward, R.H.: The Development of Long-Range Hyperbolic Navigation in the United States. Navigation (USA) 18 (1971) no. 1, pp. 51-62.

11 Sowjets Using Omega like Navaid System. Aviation Week no. 21, 1971, pp. 22.

12 Pierce, J.A.: Loran. Mc Graw Hill, New York, 1948.

13 Freiesleben, H.C.: Navigation. Matthiesen Verlag Lübeck 1957.

14 Blanchard, W.: Flugerprobung des Betriebsverhaltens von LORAN C. Internationale Navigationstagung 28-30 Oct. 1969, Hamburg.

15 Decca LORAN C/A for United States Air Force Jet Fleet. Decca Navigator News, no. 61, April 1972, pp. 18-19.

(B) MEDIUM DISTANCE AIDS

(VHF Omnidirectional Radio Beacons)

by

Dr Günter Höfgen, Department Head
Standard Elektrik Lorenz, AG
Stuttgart, Germany

TABLE OF CONTENTS

MEDIUM DISTANCE AIDS

(VHF Omnidirectional Radio Beacons)

by

Dr Günter Höfgen, Department Head
Standard Elektrik Lorenz, AG
Stuttgart, Germany

I. VOR

Summary

VOR (VHF Omnidirectional Radio Range) is a radio aid for aircraft guidance. It is an omnidirectional radio beacon providing the angle between aircraft and North, seen from the ground station. The VOR ground station radiates an azimuth-dependent signal that is analyzed as the bearing information by the aircraft receiver. The pilot guides the aircraft along the course selected with the aid of constant azimuth indication. The range is of the order of 100 to 150 nm.

1. Introduction

VOR (VHF Omnidirectional Radio Range) is a radio aid for aircraft guidance introduced on a world-wide scale[19]. It is an omnidirectional radio beacon providing the angle between aircraft and North, seen from the ground station, as the azimuth information. The present network is determined by a number of VOR ground stations located along the airways. The pilot guides the aircraft along the course selected with the aid of constant azimuth indication. This course runs either to or from a ground station. Whenever any change of course is necessary because of the airway configuration, this is generally performed when the aircraft flies over a VOR ground station.

The number of ground stations is determined by their ranges and the requirement that a ground station will be installed at every bend or branching point of an airway. The range is limited by the quasi-optical propagation conditions of VHF (metric waves) which confine the application of the VOR method to short-range and medium-range aviation. The range is of the order of 100 to 150 nm.

The combination of VOR with DME or TACAN, providing distance measurements, is an important prerequisite for area navigation, i.e. for the possibility of flying any course in the area.

2. Principles

On principle the VOR ground station radiates an azimuth-dependent signal that is analyzed as the bearing information by the aircraft receiver. The azimuth-dependent signal consists of a 30 Hz frequency phase-shifted to a reference signal in proportion to the azimuth angle. The azimuth-dependent signal is generated by rotation of a figure-of-eight radiation pattern at VHF (e.g. rotation of a dipole) at the speed of 30 revolutions per second (rps). This pattern is superposed on the omnidirectional radiated carrier in the frequency range 108-118 MHz, which results in an amplitude modulation (AM) of the carrier with 30 Hz. For proper analysis of the azimuth information in the airborne receiver, the ground station additionally transmits a reference frequency of 30 Hz. This reference signal is used for the frequency modulation (FM) of a sub-carrier of 9960 Hz having a frequency swing of ±480 Hz. The sub-carrier is also employed to modulate the said VHF carrier, again by AM. Through the use of FM and AM the azimuth-dependent signal and the reference signal are well decoupled, although both have 30 Hz; their phase shift is analyzed in the airborne receivers and provides the azimuth because the in-phase condition is adjusted in the North direction between the azimuth-dependent and the reference signals. Moreover, the carrier is amplitude-modulated with voice frequencies or speech (300-3000 Hz) and with the identity (1020 Hz). Fig. 1 shows the frequency spectrum.

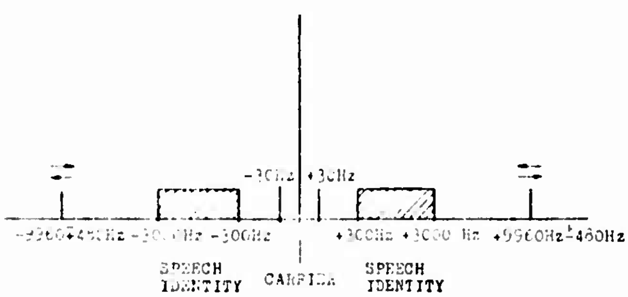


Fig. 1 VOR frequency spectrum

3. Operation Principle

The azimuth-dependent VOR signal is generated by rotating a figure-of-eight radiation pattern at a speed of 30 rps. The radiation of the rotating pattern is imposed on the simultaneously radiated carrier frequency; the latter is thus amplitude-modulated with 30 Hz and an azimuth-dependent phase angle in such a way that a change in azimuth corresponds to a proportional change of the electrical phase angle.

Practical structures for the generation of the rotating directional pattern so far provide a physical rotation of a dipole or a goniometer. In the former case, the rotating dipole is fed with the unmodulated carrier energy. In the goniometer case, fixed antennas radiate the power while the goniometer rotates with the speed of 30 rps. It is fed with the unmodulated carrier energy, and supplies amplitude-modulated 30 Hz signals with suppressed carrier to two separate outputs. The envelopes correspond to the sine and the cosine of 30 Hz (phase shift 90°).

Feasible is also an electronic solution without physical rotation based on the goniometer principle; in the following, only this solution will be considered.

The goniometer outputs are connected to two separate antennas, e.g. crossed dipoles, or to two pairs of loop antennas mounted at an angle of 90° in respect to one another where each sets up a figure-of-eight pattern (Fig. 2). Pattern 1 is associated with the first goniometer output, pattern 2 with the second.

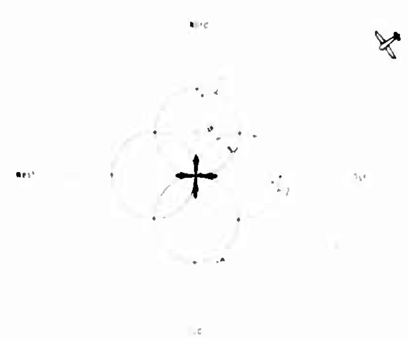


Fig. 2 Antenna patterns in the horizontal plane

The airborne antenna provides the receiving voltage

$$V_R = U_e \, (\cos\vartheta \, \sin\omega t \cos\Omega t + \sin\vartheta \, \cos\omega t \cos\Omega t)$$
$$= U_e \cdot \sin(\omega t + \vartheta t) \cdot \cos\Omega$$

where Ω = angular frequency of the carrier
ω = $2\pi \times 30$ Hz
U_e = peak value of the receiving voltage

This equation indicates that the phase angle of the 30-Hz-frequency equals the angle of azimuth ϑ.

Modulation in an electronic goniometer presents difficulties because electronic modulators meeting stringent specifications in respect of stability and linearity[5] have to be used. Here a special approach is being offered by STANDARD ELEKTRIK LORENZ AG (SEL) inasmuch as the modulation problem is separated from output power generation. Modulation is thus obtained at a low level with good stability and linearity. Fig. 3 shows the block diagram of this electronic goniometer.

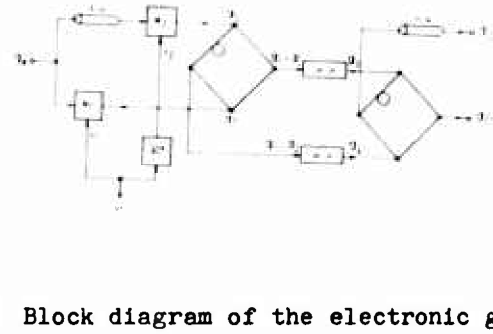


Fig. 3 Block diagram of the electronic goniometer

The input voltage $U_e = U_e \cos \Omega t$ is applied to the modulator M_1 directly and to the modulator M_2 through a quarter-wave cable providing a phase shift of 90°. The modulating voltages are:

$u_1 = U_m \cos \omega t$; $u_2 = U_m \sin \omega t$ — U_m is the peak value of the modulating voltage.

Thus the following modulator output voltages are obtained:

$U_1 = U_a \cos \omega t \cos \Omega t$; $U_2 = U_a \sin \omega t \sin \Omega t$ — U_a is the peak value of the modulator output voltage.

These voltages are transformed into unmodulated signals, namely, the upper and the lower sideband, in a bridge circuit. The bridge forms their sum and difference and thus generates the lower and upper sideband frequencies, respectively.

$U_1 + U_2 = U_a \cos (\Omega - \omega)t$; $U_1 - U_2 = U_a \cos (\Omega + \omega)t$

These two sideband voltages feed two power amplifiers from which the required goniometer output power is obtained.

The amplifier output voltages U_3 and U_4 are transformed into the goniometer output voltages with the aid of another bridge circuit:

$U_{a1} = U_A \cos \omega t \cos \Omega t$; $U_{a2} = U_A \sin \omega t \cos \Omega t$ — U_A is the peak value of the goniometer output voltage.

4. The VOR Ground Station

Here the VOR-S ground station by SEL will be described as an example. The transmitter is accommodated in a small shelter, the roof of which is both an electrical counterpoise and a supporting structure for the antenna array surrounded by a plastic cylinder for all-weather protection. The omnidirectional radiator with quadruple feeding and the crossed dipoles are all mounted on a plastic support plate. The four feeders of the omnidirectional radiator between the periphery and the centre point do both the balancing and transforming so that their parallel connection results in the desired impedance (50 Ω). A polarization cage topped by a cage-like extension under the plastic cylinder compensate for the vertical-polarization components of the crossed dipoles. There is another design of the antenna array where the radiator plate complete with polarization cage is topped by a second plate-cage combination that in turn is topped by the said extension for termination. This two-element antenna provides a field strength increased by 5 db under low elevation angles, which increases the effective range.

Fig. 4 is the block diagram of the VOR-S ground station. The transmitter comprises an exciter, a carrier transmitter, and an electronic goniometer. The modulator supplies the voltage U_{mod} to amplitude-modulate the carrier.

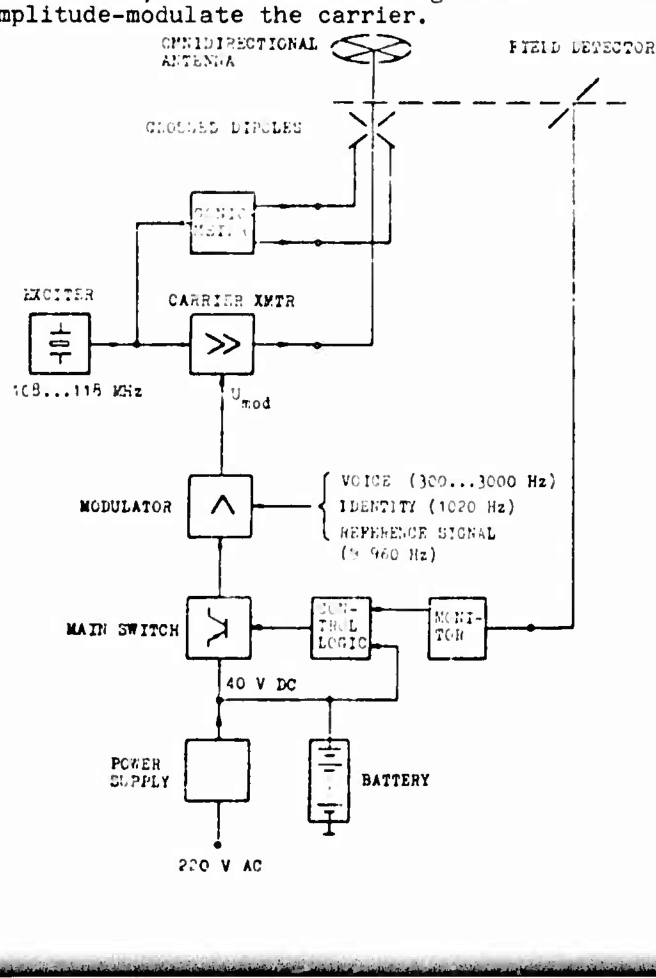


Fig. 4 VOR-S ground station; block diagram

Speech, identity, and the 9960-Hz subcarrier voltages are summed. The subcarrier contains the 30-Hz reference signal as FM. The main switch of the station is operated by a control logic that causes switching to the OFF state or to stand-by operation whenever an alarm is given by the monitor or the transmitter.

The monitor of the VOR-S station ensures that the radiated VOR signal complies with ICAO specifications[19]. Otherwise it gives alarm resulting in the said switching process by automatic means. A field detector located in the field of radiation detects the radiated signal, demodulates it, and feeds it to the monitor.

5. The Airborne Receiver

The airborne receiver picks up the signal of that ground station to which its channel selector (Fig. 5) was adjusted. After one or more frequency conversions and IF amplification with AGC, the VOR signal is demodulated.

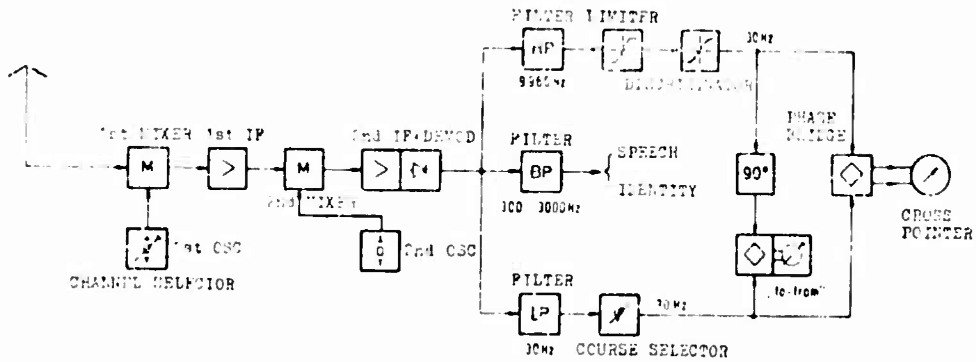


Fig. 5 VOR airborne receiver, basic layout

The composite signal is separated by means of filters into separate channels: the azimuth-dependent 30-Hz signal; the 9960-Hz subcarrier modulated with the 30-Hz reference signal; the station identity and speech. The 9960-Hz subcarrier passes through limiters to a frequency discriminator supplying the 30-Hz reference signal. Both 30-Hz signals feed a phase bridge, the one being directly applied while the other one first passes through the course selector.

By its electrical function, the course selector is a phase shifter permitting adjustment of any value between 0° and 360°. Here the selected course is adjusted, that is, an azimuth difference against North, which corresponds to an equivalent phase-angle difference between the two frequencies of 30 Hz.

The two phase-bridge outputs have a zero voltage difference in balanced state and are connected to the so-called cross-pointer instrument with a full-scale deflection of $\pm 10^\circ$. The phase bridge is balanced and the needle of the cross pointer is in its central position whenever the phase difference of the input voltages amounts to either 0° or 180°. (This description disregards a possibly necessary additional 90° phase shift of the input voltage for the purpose of bridge alignment.) Should the aircraft deviate from the selected course adjusted on the course selector, then the cross pointer would indicate a proportional deflection. When flying over a ground station, the phase difference of the 30 Hz will change by 180°; the phase bridge therefore remains in the aligned state.

A "to-from" indicator (Fig. 5) shows the pilot whether the azimuth is the bearing from the ground station to the aircraft or from the latter to the ground station. Contrary to a self-contained direction finder, VOR direction finding is independent of the aircraft heading (the longitudinal aircraft axis). For this reason, the to-from reading will correspond to the heading only whenever the latter coincides with the azimuth adjusted on the course selector. The to-from indicator operates on the same basis as the above described phase bridge with cross pointer, although it is much simpler because it gives only one qualitative information. Its main feature is the additional phase shift by 90° of the two feeding 30-Hz voltages. When the aircraft flies over a ground station, the sign of this 90° shift will change and the indicator will thus change from indicating "to" to indicating "from" the ground station.

6. System Accuracy

The errors affecting the VOR method are intrinsic errors of the ground station; the site error introduced by the surrounding terrain; the intrinsic error of the airborne station; and the pilot's errors. The intrinsic error of the ground station is relatively small, about 1°. Significant is the site error by obstacles reflecting the radiation. The receiver will then receive the obstacle azimuth in addition to the correct azimuth. Vectorial addition results in an error, the magnitude of which is dependent of the amount of radiation reflected, the RF difference between the phases of direct and reflected wave, and the difference between receiver and obstacle azimuth.

II. Doppler VOR

Summary

The designation Doppler VOR (DVOR) indicates its close relationship to the conventional VOR; the signals radiated by both, VOR and Doppler VOR ground stations, are compatible and can be processed by normal VOR receivers. Doppler VOR has an overriding advantage over VOR: a substantial reduction of the site error.

1. Introduction

The designation Doppler VOR indicates its close relationship to the conventional VOR; the signals radiated by both VOR and Doppler VOR ground stations are compatible and can be processed by normal VOR receivers.

Although a Doppler VOR ground station is more sophisticated and its cost therefore higher, it has an overriding advantage over VOR: a substantial reduction of the site error. Generally Doppler VOR is still useful in a terrain abounding in obstacles where VOR would be useless because of the high site error.

2. Principle

The functions of the two 30-Hz frequencies in VOR are reversed in Doppler VOR (DVOR): that amplitude modulating the VHF carrier is the reference signal and that dependent on azimuth is contained in the 9960-Hz subcarrier.

The most important feature of DVOR is the generation of this subcarrier, frequency-modulated with the 30-Hz frequency at a phase relationship depending on the azimuth. The frequency modulation is achieved through the doppler effect. The principle of DVOR can be explained with reference to Fig. 6.

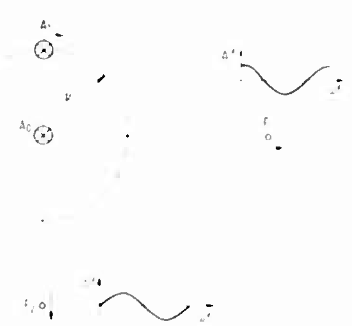


Fig. 6 Schematic presentation of DVOR principle

The central antenna A_0 radiates the omnidirectional carrier that is amplitude-modulated with the 30-Hz reference signal. Installed at a distance R from this carrier antenna is the sideband radiator A_1 that should be imagined as rotating around A_0 on a circle having the R radius.

The A_1 sideband frequency is offset against the carrier frequency by +9960 or -9960 Hz. If now the A_1 antenna is rotated at the speed of 30 revolutions per second (rps), the subcarrier is azimuth-dependent frequency modulated because of the doppler effect. The frequency swing is

$$\Delta F = F \cdot \frac{R \cdot \omega}{c}$$

where F = carrier frequency, approximate sideband frequency
c = velocity of light
$\omega = 2\pi \times 30$ Hz

In the frequency range between 108 MHz and 118 MHz, ICAO requires a frequency swing of ±480 Hz[19]. This calls for a circle diameter R = 6.5 to 7.1 m.

Fig. 6 also shows the variation of the sideband frequency as a function of time when received in the far field. This sideband frequency is frequency-modulated by the motion of the sideband radiator A_1. Assuming two positions E_1 and E_2 of airborne receivers and designating by Δf the deviation between the transmitted and the received sideband, we obtain

$$[\Delta f]_{E_1} = \Delta F \cdot \cos \omega t \qquad [\Delta f]_{E_2} = \Delta F \cdot \sin \omega t$$

This shows that the azimuth difference of 90^o between E_1 and E_2 in Fig. 6 is identical to the phase difference between the two frequencies of 30 Hz appearing as sideband FM at both receiver positions. Moreover it is generally true for DVOR just as for VOR that the electrical phase difference exactly corresponds to the azimuth difference of the receiver positions.

3. Operation

The above assumed physical rotation of the sideband antenna along the specified circle is actually impossible because the antenna would have to have four times the speed of sound. Therefore, this rotation is simulated by electronic means. For this purpose, a number of fixed antennas are mounted on a circle and fed in a sequence and at a rate that simulates rotation of the sideband radiation. In order to also simulate the continuity of the rotation, the radiation amplitudes of neighbouring antennas are made to overlap as shown in Fig. 7.

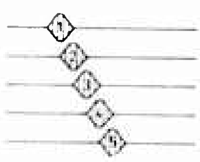


Fig. 7 Radiation amplitudes of No. 1, 2, 3 ..... antennas simulating the rotation

So far this description referred to the operation sketched in Fig. 6, namely, simulation of the rotation of only one sideband. In order to obtain from DVOR the VOR spectrum shown in Fig. 1 the other sideband has to be implemented. This is achieved by feeding the instantaneously opposite antenna on the circle with the second sideband, and by adopting the same feeding sequence and continuity provision for all antennas. In other words: the two sidebands thus set up are always in line with the carrier radiator in their centre and at an unvarying distance. The two sidebands consisting of frequencies differing from the carrier by +9960 and -9960 Hz, respectively, are combined in the far field at proper phase resulting in the carrier amplitude-modulated with the 9960-Hz subcarrier. A prerequisite is the provision of proper level and phase relationships between carrier and sidebands in the ground station.

The methods employing two sidebands are known as double-sideband (DSB) methods[16]. The DVOR spectrum employing DSB exactly corresponds to the VOR spectrum, but is characterized by higher implementation costs and difficulties.

A variant of the DSB method is the single-sideband (SSB) method. It is characterized by minimal cost, but has no VOR-compatible spectrum and substantial system shortcomings. The use of only one sideband entails the following potential errors:

- Rotation of the sideband results in an additional carrier modulation so that the reference signal (30 Hz AM) contains a phase error. The disturbing modulation is brought about by parasitic carrier radiation from the instantaneously radiating sideband antenna.

- Additional errors depending on the design of the airborne receiver type are introduced by the subcarrier having only one sideband and being additionally modulated in its amplitude by 30 Hz through the simulated 30-rps rotation. This results in cross modulation in the demodulator and generates together with other non-ideal receiver properties the errors.

Another variant is the alternating-sideband (ASB) method. This solution avoids high expenditure and realization problems by moderate sophistication and provides all the advantages of the DSB method. The ASB method[17] has been developed by Standard Elektrik Lorenz AG (SEL) and introduced world-wide with outstanding results. It employs both sidebands and results from DSB by eliminating every other sideband antenna in such a way that the resulting halved number of sideband antennas is an odd number. It can be shown[10] that the simulated rotation of the sideband radiation is equivalent to that in DSB although the transition from one antenna to the next does not involve the adjacent antenna radiating the same sideband, but the opposite antenna displaced one half division that radiates the other sideband. The ASB method permits satisfactory decoupling of the sideband antennas and is particularly suitable for electronic commutation.

4. The DVOR Ground Station

This section contains a description of the DVOR-S ground station by SEL. Employed is the ASB method. The construction is of the same kind as used for VOR-S. The transmitter equipment complete with the electronic antenna switching unit (commutator) is installed in a small shelter. The 39 sideband antennas as well as the carrier antenna in their centre point are all mounted on a counterpoise meshing arranged above the shelter. The counterpoise is 30 to 40 m in diameter and is generally mounted at a height of 3 to 10 m above ground. Because of the wide base of the antenna array, the monitoring dipole is mounted at a distance of about 200 m from this antenna arrangement.

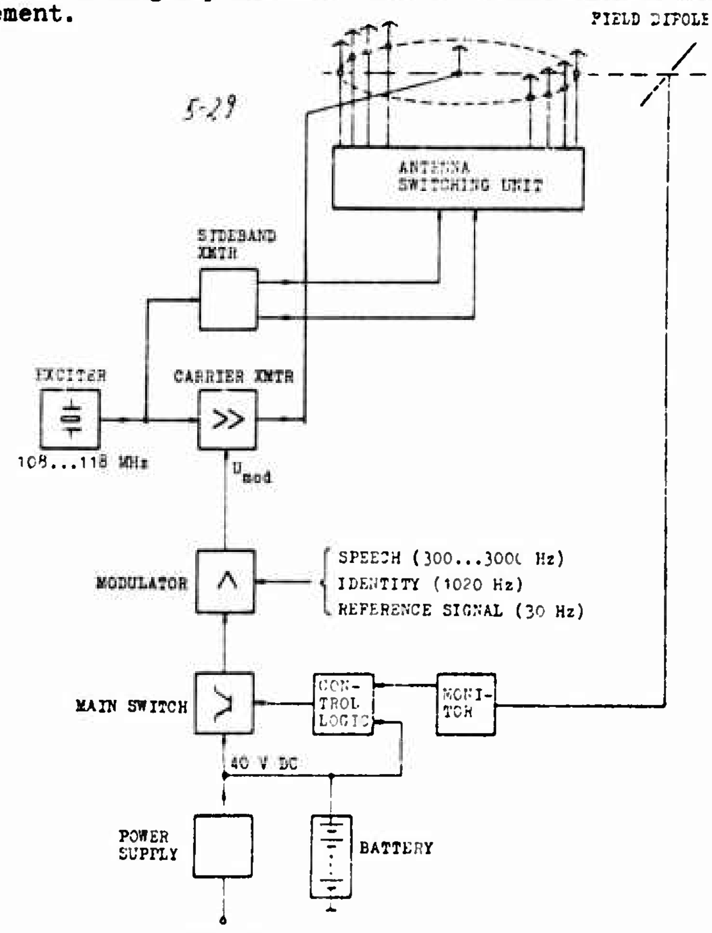


Fig. 8 Block diagram of the DVOR-S ground station

The ground station (Fig. 8) comprises the exciter, carrier transmitter, and sideband transmitter. The former two as well as the modulator, monitor, control logic, power supply, and main switch are of the same design as in VOR-S. In the DVOR-S sideband transmitter, the ±9960 Hz sidebands of the carrier are generated and amplitude-modulated to almost 100% in the power stages as timed by the antenna-switching frequency. The sideband generation is done by the same principle as in the VOR-S goniometer (by elimination of the second bridge circuit and by the use of 9960 Hz instead of 30 Hz modulating frequency). The envelopes are mutually shifted by one half of a cycle. The antenna switching unit takes care of the cyclic switching from antenna to antenna, each time in the voltage minimum. At the same time the about opposite antenna has its voltage maximum. Good decoupling is provided by the fact that only opposite and not adjacent antennas radiate at the same time; this decoupling is an important prerequisite for optimum suppression of parasitic radiation.

5. System Accuracy

The wide antenna base and the insensitive FM transmission of the azimuth-dependent signal constitute the substantial improvements of DVOR against VOR. A diagram published elsewhere[9] indicates the theoretical azimuth error due to reflection by an obstacle vs. angular difference between receiver azimuth and obstacle azimuth for both the VOR and the DVOR methods. It shows that

- the error maxima of DVOR are by about one order of magnitude less than that of VOR,
- the DVOR error is greatest when the angular difference between receiver azimuth and obstacle azimuth is least while the VOR error is greatest when the said angular difference is 90°.

The difference in the quality of course indication by VOR and DVOR is best illustrated by the records obtained from both systems over the same flying route. An example is shown in Fig. 9. In most cases the system error of the DVOR is less than 1° even if the site error is involved.

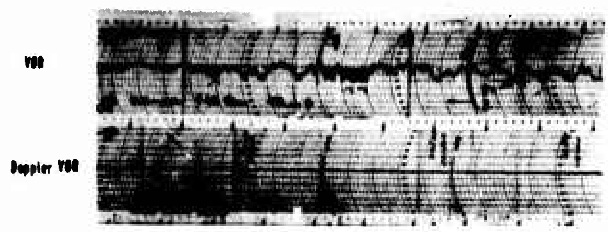


Fig. 9 Course indications for VOR and DVOR at the same route (Salzburg ground station)

References

1 Bärner, K.: Flugsicherungstechnik I, Hans Reich Verlag, München 1957, pp. 110-135.

2 Bärner, K.: Untersuchungen an Drehfunkfeuern. Luftfahrttechnik, April 1957, 82-88.

3 Crone, W.: Die Antennen des UKW-Drehfunkfeuers (VOR). SEL-Nachrichten, 1958, no. 2, 81-86

4 Bärner, K.: Kurzanzeigefehler bei Drehfunkfeuern. Z. f. Flugwissenschaft 8, 8-17, 1960.

5 Höfgen, G.: Ein elektronisches Goniometer. Frequenz 22, no. 2, 32-34, 1968.

6 Popp, H.: Festkörper-VOR, eine neue Generation von UKW-Drehfunkfeuern. Elektrisches Nachrichtenwesen 44, no. 4, 333-342, 1969.

7 Feyer, W.: Die Weiterentwicklung des Drehfunkfeuers (VOR) zum Doppler-Drehfunkfeuer (D VOR). Luftfahrttechnik-Raumfahrttechnik 14, no. 7/8, 186-193, 1968.

8 Berner, H.: Entwicklungsprobleme bei Navigationsantennen. SEL-Nachrichten 1968, no. 2, 76-80.

9 Anderson, S.R; Flint, R.B.: The CAA Doppler Omnirange. Proc. of the IRE 1959 May, 808-821.

10 Höfgen, G.: Das richtungsabhängige Signal des Funknavigationsverfahrens "Doppler VOR" und seine Auswertung in gebräuchlichen Empfängern. Diss. a. d. Techn. Universität Berlin, D 83 July 1963.

11 Kramar, E.: Doppler-Verfahren als Navigationshilfe. Nachrichtentechnische Zeitschrift 18, no. 1, 27-32, 1965.

12 Kramar, E.; Feyer, W.: Entscheidende Verbesserung der Mittelstrecken-navigation: Das Doppler-Drehfunkfeuer. Interavia, 1968, no. 2, 180-182.

13 Crone, W.; Popp, H.: Das Doppler-Drehfunkfeuer. Elektrisches Nachrichten-wesen 43, no. 2, 146-152, 1968.

14 Rüchardt, H.: Das Doppler-VOR-System. SEL-Nachrichten 1968, no. 2, 44-53.

15 Kramar, E.: The Doppler VOR - Its Foundation and Results in AGARD Conference Proceedings no. 28, 1970, pp. 129-142.

16 Kramar, E.; Steiner, F.: D. Patent 1123000, appl. 28.3.1959.

17 Earp, C.W.: D. Patent 1186522, appl. 18.2.1963 (Priorität England 19.2.1962).

18 Bärner, K.: VOR-System in L. Brandt, Funkortungssysteme für Luft- und Seefahrt, Verkehrs- und Wirtschaftsverlag Dr. Borgmann, Dortmund 1962, pp. 53-70.

19 ICAO, Aeronautical Telecommunications, Annex 10, Vol. I, Oct. 1965.

Preceding page blank

DISTANCE MEASURING METHODS

by

Dr Manfred Böhm, Head, Development Navigation
Standard Elektrik Lorenz, AG
Stuttgart, Germany

TABLE OF CONTENTS

DISTANCE MEASURING METHODS

by

Dr. Manfred Böhm, Head, Development Navigation, SEL, Stuttgart

Summary

Of various radio distance measuring methods principally possible only the ICAO-standardized DME (distance measuring equipment) is being employed in aviation. This equipment features round trip time measurement between interrogation pulses transmitted by airborne equipment, and reply pulses transmitted by a ground transponder. This transponder transmits a reply 50 /us after each reception of an interrogation. A search and track system in the airborne equipment discriminates own replies from replies to other aircraft. Every ground station is capable to serve 100 airborne equipments.

1. Introduction

The distance-measuring methods employing electromagnetic waves make use of the velocity of light c; the path ρ covered by a light or radio signal can be determined, if the transit time t is known, from the formula $\rho = c \times t$.

There are round-trip and one-way distance measuring systems. The round-trip systems operate with responding transmitters (or with passive reflectors as in the case of radar or radio altimeters). The responding transmitter will retransmit the received signal with an exactly defined delay (Fig. 1). In the case of one-way distance-measuring systems, identical and extremely accurate time standards are used in both the transmitter and the receiver locations (Fig. 3).

At present, the round-trip method is the only one used although the one-way method is more desirable for navigation because the number of users served is infinite, just as in the case of radio beacons for azimuth measurements. The reason for preference of the round-trip method is the high cost for sufficiently stable time standards that are indispensable for one-way operation. Lately, however, methods virtually constituting a mixture of one-way and round-trip methods have been developed for collision-avoiding systems (CAS). Here reasonably priced crystal-controlled oscillators are re-synchronized by round-trip measurements after relatively large intervals and may thus be used during such intervals as sufficiently accurate time standards for one-way measurements.

Furthermore, distance-measuring systems may employ either the pulse method or the continuous-wave or CW method; the latter is illustrated in Fig. 2. The CW methods where single-tone or multi-tone modulation is employed, are widely used for mapping (geodatic) while they are still in the planning stage for navigation systems involving satellites.

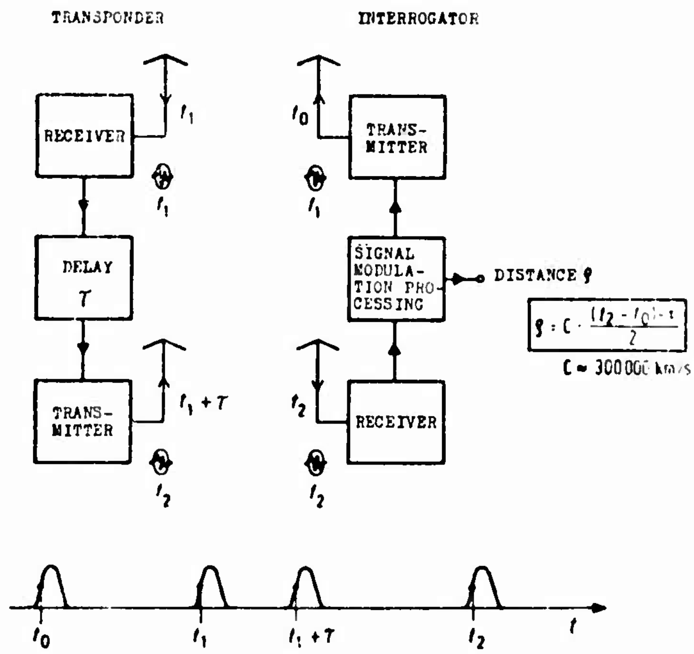


Fig. 1 Round-trip distance measurement by pulses

Therefore, the tone-modulation systems will not be described here.

To serve a substantial number of users with distance data, a special control is required involving either time or frequency division multiplex equipment. Therefore, a pulse method has widely been adopted for short and medium-distance navigation in aviation.

Its underlying principle that has also been recommended for future landing systems[13] will be described in the next section.

2. Distance Measuring Equipment (DME)

DME was introduced in 1959 and is now an internationally standardized distance-measuring equipment. It is usually operated in combination with the VOR azimuth-measuring system, but is also to be used in future instrument-landing systems (ILS). The balance of this section is a summary of international specifications[1] applying to DME:

Overview

The DME shall provide for continuous and accurate indication in the cockpit of the slant range distance to a ground station. The complete system shall comprise the airborne interrogator and the replying ground station or transponder.

SIGNALS MODULATED WITH ONE TONE FREQUENCY ONLY

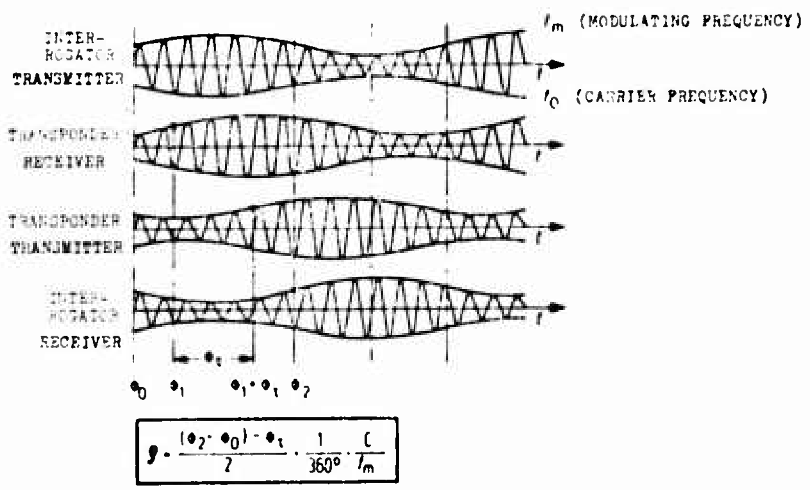


Fig. 2 Round-trip CW distance measuring system with tone modulation (block diagram similar to Fig. 1)

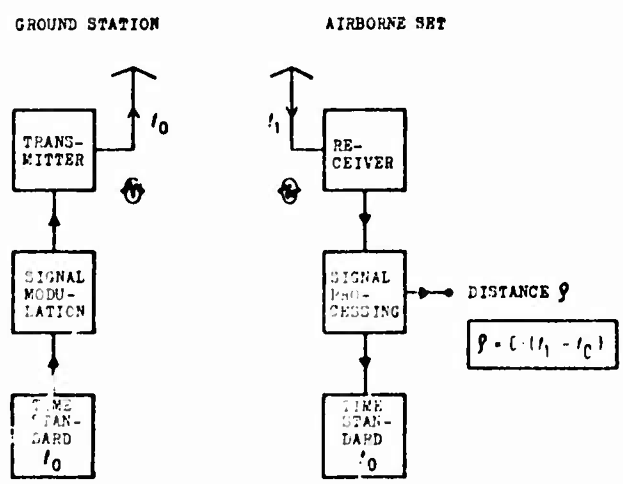


Fig. 3 One-way distance measuring with pulses

Accuracy

The distance error within the whole coverage area shall not exceed ± 0.5 nm or ± 3 % of the distance, whichever measured value is greater.

Properties

The system will be useful to measure the slant distance up to 20 nm beyond the radio horizon. It will be operated in the frequency band of 960 to 1215 MHz. The frequency spacing between interrogation and response will be [illegible] MHz. The allocation of frequencies to the various channels will be as in Table 1.

Table 1 Channel distribution for VOR/DME

Channel	VOR frequency[a] in MHz	DME interrogator frequency[b] in MHz	DME transponder frequency[b] in MHz
X		PSC 12 µs	PSC 12 µs
1	-	1025	962
.	.	.	.
16	-	.	.
17	108.00	.	.
.	.	.	.
59	112.20	.	.
60	-	.	.
.	.	.	.
63	.	.	1024
64	.	.	1151
.	.	.	.
69	-	.	.
70	112.30	.	.
.	.	.	.
126	117.90	1150	1213
Y		PSC 36 µs	PSC 30 µs
1	-	1025	1088
.	.	.	.
16	-	.	.
17	108.05	.	.
.	.	.	.
59	112.25	.	.
60	-	.	.
.	.	.	.
63	.	.	1150
64	.	.	1025
.	.	.	.
69	-	.	.
70	112.35	.	.
.	.	.	.
126	117.95	1150	1087

Notes:

a - channel spacing 100 kHz
b - channel spacing 1 MHz
PSC - pulse-spacing code time

The channels 1Y through 16Y, 60 X,Y through 69 X,Y, 70Y through 79Y, and 124Y through 126Y must not be used if their use will interfere with the secondary-radar systems operating with the same frequencies.

The statistically distributed pulse pairs transmitted by the interrogator and the transponder will have the shape and spacing sketched in Fig. 4. The spacing between the pulses of a pair will be chosen so as to provide unambiguous discrimination against other services.

•The mean repetition frequency of the interrogator pulse pairs will not exceed 30 per second, assuming that 95 % of the total time will be covered by tracking operation. If the search time is to be reduced, it shall be possible to transmit up to 150 pulse pairs per second.

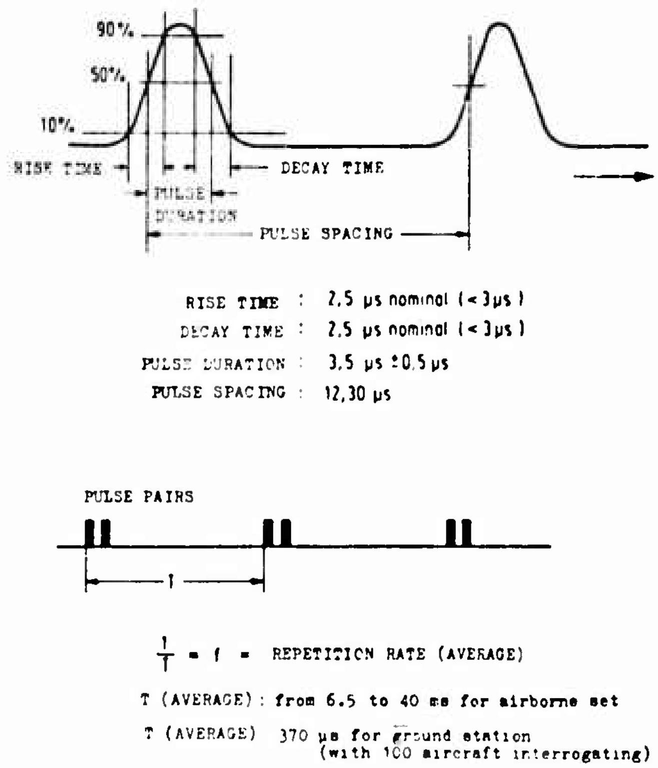


Fig. 4 DME pulse format

Capacity

Each responding ground station (transponder) shall be capable of serving up to 100 aircraft simultaneously.

3. Operation Principle

The summarized presentation of the preceding section will now be investigated in some detail.

Fig. 5 is a block diagram of the interrogating airborne equipment (the interrogator) and the responding ground station (transponder). The pulse peak power is between 50 and 2000 watts, depending on the type of set.

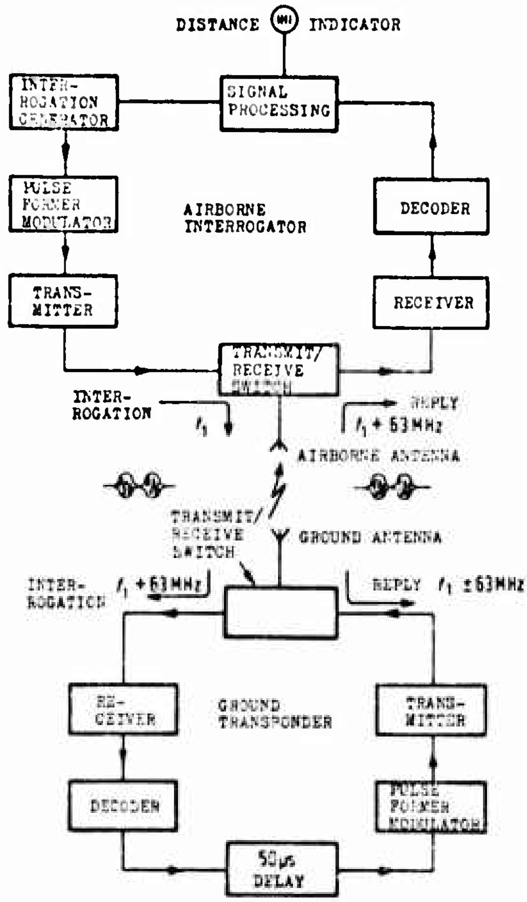


Fig. 5 DME block diagram

The interrogating pulses are received by the selected ground station and re-transmitted to the aircraft on a frequency spaced from the interrogating frequency by 63 MHz after a fixed delay of 50 μs, known to the interrogator. The transmitter power of the transponder may be between 1 kilowatt and 20 kilowatts.

In the airborne equipment, automatic circuits evaluate the delay between transmission of an interrogator pulse pair and reception of a transponder pulse pair, taking account of the fixed delay. The resulting time difference is converted into distance, and the distance is ind' ted (Fig. 5).

There are t' reason' using pulse pairs. In the first place, coding of the pulse spacing prov' a le means of identifying the channels X and Y, see Tables 1 and 8. In the sec pla the use of pulse pairs is a protection against the errone-ous evaluation of n -DME system single pulses as generated, for instance, by radar equipment in the L band.

In most equipment used the first pulse of a pair provides coding while the second serves for actual distance measuring. In new developments, however, it is the first pulse of an associated interrogator and transponder pair that is employed for distance measuring because this mode will substantially reduce reflection errors.

Each ground station can serve simultaneously 50-100 interrogating aircraft. Each airborne interrogator accepts only responses to its own interrogations. This is achieved by a statistically varied repetition rate of the interrogation pulses and a special, airborne search and tracking system.

As may be seen from Table 1, a total of 126 "X" channels is available, to be complemented by 126 "Y" channels. The channel spacing is 1 MHz, each channel offering a bandwidth of about 300 kHz.

Airborne Equipment

A somewhat refined block diagram of an airborne interrogating unit is shown in Fig. 6. The receive and transmit frequencies are always spaced by 63 MHz; therefore, a common frequency generator may be used. In older equipment, the frequencies were generated by means of 126 crystals; now the 126 frequencies are being generated by a synthesizer using one crystal in combination with a voltage-controlled oscillator (VCO) and a frequen y divider.

A common antenna is used for transmitting and receiving. A transmit/receive switch prevents interference upon the receiver during transmission.

The transponder pulses received in the aircraft are compared with the interrogations. However, all of the transponder pulses are received, about 3000 pulses per second. For this reason, the airborne DME has to perform two important functions, namely, filtering-out of the wanted responses from all other transponder pulses search and conversion of the round-trip time into a current indication of the distance (tracking).

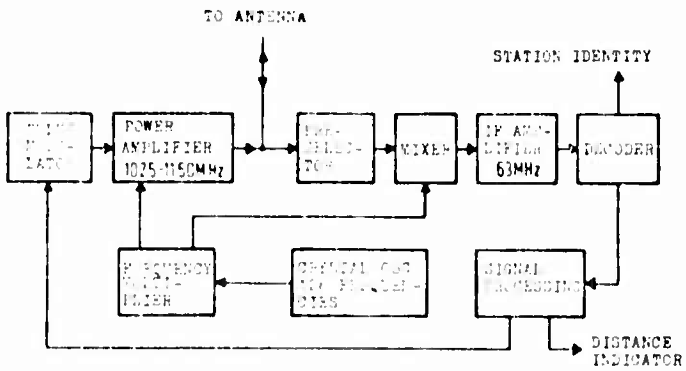


Fig. 6 Block diagram of an airborne DME

These functions can be implemented in a number of ways; all of them are based on the principle shown in Fig. 7.

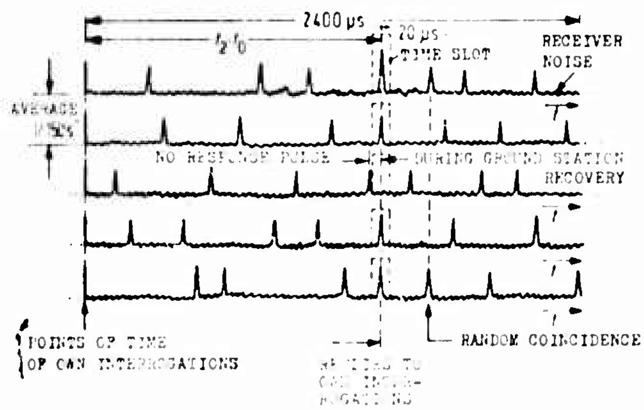


Fig. 7 The DME search process

This diagram displays the replies received in the airborne DME during the time of six sequential interrogations. The time slot is shifted towards longer round-trip periods during this time and is stopped if 80 % of the interrogations result in replies during the time-slot duration.

The characteristic search process is performed as follows: For each interrogation pulse, a time slot with the duration of about 20 μs is generated. Only a reply pulse received during these 20 μs will be processed. Otherwise the time slot is shifted for each new interrogation along a time scale corresponding to the maximum feasible distance. Considering a search interrogation rate of 150 per second and a duration of 20 μs of the time slot or gate, the analyzing circuit is ready to receive for only 0.3 % of the total time, or only 3 milliseconds for each second.

The shifting of the time slot stops as soon as a pulse incides several times in succession with a certain high rate. This occurs only in cases of replies to interrogations of the associated airborne DME, which is made possible by the statistical variation in time of the ground-station pulses.

If the ground station transmits 3000 statistically distributed pulse pairs per second, nine of these pulses will fall into the time slot in every second. Whenever the time slot has arrived at a point that corresponds to the correct distance, however, then almost 30 pulses will fall into the time-slot period. It is this ratio of 9 to 30, assuming a shifting rate of 18 km/s, that forms the criterion causing change-over from search to tracking operation.

The time spent for searching proper replies by shifting the time slot, should be kept as short as possible. Hence, a pulse-repetition rate of 150 per second is employed during the search. Depending on the equipment type used, the whole search time takes between 1 and 20 seconds.

A shift of the time slot at a rate of 18 km/s means that search over a range of 2400 μs, which corresponds to 360 km, would take 2o seconds. Advanced systems, however, reduce this search time to only 1 second.

When the proper position of the time slot has been found, then the interrogation rate is reduced to a value between 5 and 30 per second.

In the ensuing tracking mode, the time slot of 20 μs will change its position in synchronism with the shift of the replies to interrogations of the associated DME. The generation of the time slot is accelerated if the reply pulses fall into the first microseconds and is delayed if the pulses fall into the last microseconds of the time slot. The change of position of an aircraft during the time between two interrogation pulses is very small; hence, the interrogation rate can be reduced in this mode without any adverse effect. Generally 25 pulse pairs per second are used; however, even only 2 pulses per second have been found satisfactory in simulated airspeeds up to 5400 km/h.

Usually the time slot is associated with a tracking circuit that prevents changing-over to search just because response signals were lost for a short time. As a result of a signal failure, the time slot will thus either remain stationary for 10 sec (static memory) or continue to be shifted with the last tracking rate (dynamic memory).

The slot location, that is, the time spacing between interrogation pulse and slot, can be displayed by means of an analog voltage feeding a pointer instrument in very simple units. The error increases in proportion to the full deflection, but will not exceed the 3 % specified. In more expensive units, tracking servos with coarse/fine evaluation are employed where the inherent equipment error can be kept below 180 m regardless of the distance; in these units, however, the system accuracy deteriorates beyond the specified value because of various other parameters.

Ground Station

While the airborne DME should accomodate at least 126 X channels (to be supplemented by 126 Y channels at a later date), a ground station usually operates on only one channel. For this reason, the ground-station receiver can be made more sensitive, and its associated transmitter more powerful. The number of aircraft that can be served at the same time results from the assumption that, on the average, 95 % of all aircraft will be in the tracking mode with not more than 25 interrogations per second,and 5 % will be in the search mode with max. 150 interrogations per second. To serve 100 aircraft, the ground station (the transponder) has to transmit 3000 pulses per second. Although its pulse peak power is the same as in the airborne interrogator, the transponder has a higher power consumption for this reason. Most transponders operate with a constant mean duty cycle.

As long as no interrogation comes in, the transponder pulses are derived from receiver noise and constitute filling pulses. As long as less than 100 aircraft interrogate, the transponder radiates a mixture of response pulses and filling pulses. When 100 aircraft are to be served, then the transponder provides exclusively reply pulses. If more than 100 aircraft interrogate, then the sensitivity of the transponder receiver is reduced to a point where only the nearest 100 aircraft can be served.

The constant duty cycle mode has the following advantages:

(1) In an automatic process, the transponder is operated with the highest receiver sensitivity permissible at any time, the number of reply pulses transmitted being a readily evaluated, unambiguous controlling parameter.

(2) The transponder transmitter cannot be overloaded.

(3) The automatic gain control in the airborne DME can be simplified.

(4) When too many aircraft interrogate, service is withheld from the most remote aircraft where the aircraft operation is least affected.

To arrive at the simplest possible, reliable circuits, the transponder receiver is turned off during transmission. Moreover, the sensitivity of this receiver is reduced as a consequence of powerful interrogations so that the echoes of the latter cannot trigger replies. For this reason, some of the interrogations are always lost. According to estimates, this loss amounts to about 20 %. In other words: for every 25 interrogations, the airborne DME (interrogator) will receive only 20 responses on the average. This fact is taken account of in the dimensioning of interrogator circuits.

The delay between reception of an interrogation and transmission of the reply is fixed to 50 μs. This delay is permitted to vary by $\pm$ 0.5 μs, which corresponds to 0.04 nautical miles (nm). However, the resulting error is small compared with the ICAO-specified permissible system error of max. 0.5 nm or 3 % of the distance, whichever value is greater. When analyzed in the interrogator, the 50 μs are, of course, excluded from the time-to-distance conversion.

Under control of an external generator, each transponder transmits its identification code or identity which is usually the same as that of the associated VOR ground station. The identity is transmitted every 37 sec for the duration of about 3 sec. For this duration, the statistically distributed transponder pulses are substituted by equally spaced pulses with a repetition rate of 1350 per second. In the interrogator, these pulses excite a ringing circuit tuned to the frequency of 1350 Hz. The identity pulses are modulated by a three-letter Morse code in compliance with international specifications. During the transmission of the ground-station identity, the memory in the airborne DME, described above, becomes effective.

4. Development Trends

Further development of the DME system is characterized by the following trends: increase of measuring accuracy; increase of the number of users; and improvement of reliability and economy through exclusive use of solid state devices.

Increase of Measuring Accuracy

The distance-measuring accuracy has to be improved because DME is to be applied to additional tasks. In the first place, DME in combination with VOR and DVOR is to enable accurate area navigation to be implemented with the least possible aircraft separation. In the second, DME in combination with ILS should provide continuous and accurate distance indication so that the rather coarse marker indication of ILS can be substituted.

Improvement of the distance-measuring accuracy requires reduction of both the intrinsic equipment errors and the errors caused by multipath propagation. The simplest approach would be an increase of the channel bandwidth; the then resulting steeper pulse edges and less critical units as amplifiers and filters would reduce both error types. This approach is to be taken for the future microwave landing system (MLS) recommended by the Special Committee SC-117 of the Radio Technical Commission for Aeronautics (RTCA). This is feasible only because a new frequency band will be made available and because no compatibility with any already introduced signal format is specified. The new band comprises the frequencies 5003 to 5060 MHz.

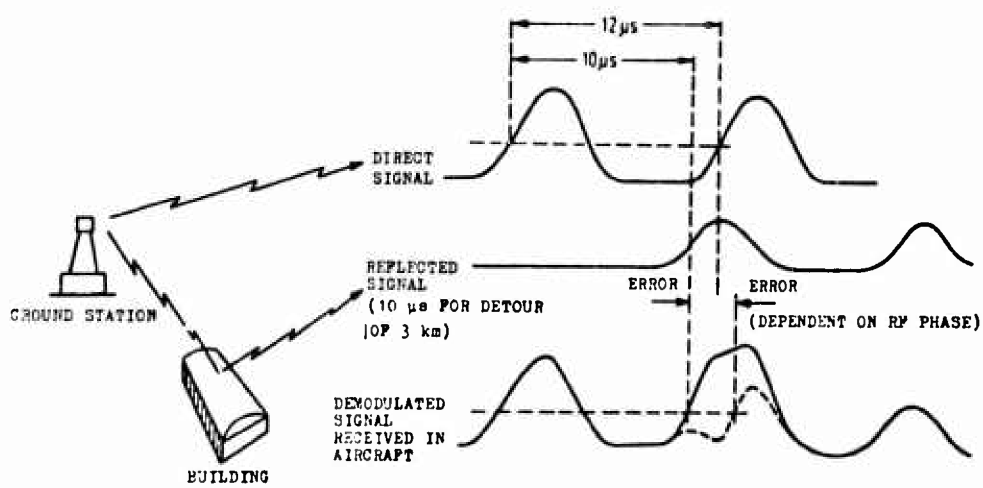


Fig. 8 Effect of echo signals upon the second pulse of a reply pulse pair.

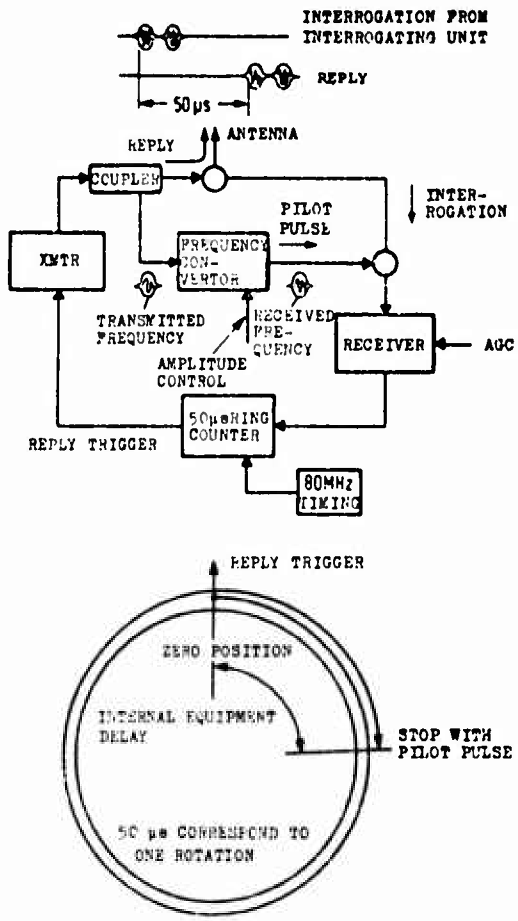


Fig. 9 Pilot-pulse mode employed in the transponder

An improvement of the ICAO-DME accuracy can only be achieved by refining the signal evaluation. An approach that has proved useful to reduce multipath propagation errors illustrated in Fig. 8 is to measure the leading edge of the first pulse of any pair because reflected pulses must necessarily be received later than the straight-path pulses.

The pilot-pulse mode (Fig. 9) is effective in reducing the propagation-time fluctuations within the interrogator and the transponder. Here the DME is not started by the interrogation video pulse, but by a pilot pulse that is decoupled from the interrogation pulse at RF level,converted in frequency, and fed into the receiving channel, thus eliminating intrinsic delays.

Increase of the Number of Users

The now 100 users per transponder can be expected to grow over 800 in future[10]. The most important measure to achieve this goal is the reduction of the number of pulses exchanged between all users and the transponder in unit time. If the present rate of 25 interrogations per second in the tracking mode can be reduced to 10 per second, for instance, then a factor of 2.5 is gained. Unfortunately the reduction of measuring pulses is somewhat in conflict with accuracy because the effect of statistically distributed errors can be kept low only by a sufficiently high number of individual measurements. In the long run another method seems to hold promise, a method that is being investigated for collision-avoidance systems (CAS). It is basically a one-way DME with the additional feature of synchronizing in intervals of several minutes by regular round-trip measurements the airborne time standard, a crystal oscillator with a stability of 10 to 0.1 parts per billion (ppb) to the transponder time standard. It should be remembered that the overall pulse requirement per user is very small even though each user transmits many interrogations and receives many responses; this is so because the interval between the individual synchronization processes is rather long (several minutes). This solution, although burdened with additional costs, allows for an increase of the number of simultaneous users.

At all times between the synchronization processes, the distance is measured by one-way DME. Measured is the time between transmission of transponder signals and their reception in the interrogator, both times referred to the transponder time standard to which the airborne DME is regularly synchronized.

Suppose the synchronization of the airborne time standard during the flight could be eliminated, i.e. a more stable time standard were available for economic airborne application. We would then have a true one-way DME that could not be saturated by the number of users - just as is the case with the rotating radio beacons. At present such solution is too expensive because it would involve the airborne use of atomic frequency standards. However, the dimensions and costs of such standards have been greatly reduced in the past ten years; hence, their introduction can be expected before the end of this century.

Improvement of Reliability and Economy

The technological development of ground and airborne DME is characterized by solid-state designing, micro-miniaturization, and digitizing. Tubes are still being used in transmitter amplifiers, but even here their substitution by solid-state stages is only a matter of time.

TACAN

by

Dr Manfred Böhm, Head, Development Navigation
Standard Elektrik Lorenz, AG
Stuttgart, Germany

TABLE OF CONTENTS

TACAN

by

Dr. M. Böhm, Head, Development Navigation, SEL, Stuttgart

Summary

TACAN (TACtical Air Navigation) is a radio position-fixing method for military short and medium range aviation. Since the 1950ies it has been introduced on a large scale in all NATO countries.

Each ground station currently provides azimuth or "theta" values to any number of aircraft and distance or "rho" values to max. 120 aircraft at the same time. Both types of values are either directly displayed (Fig. 1) or are inputs for a navigation computer.

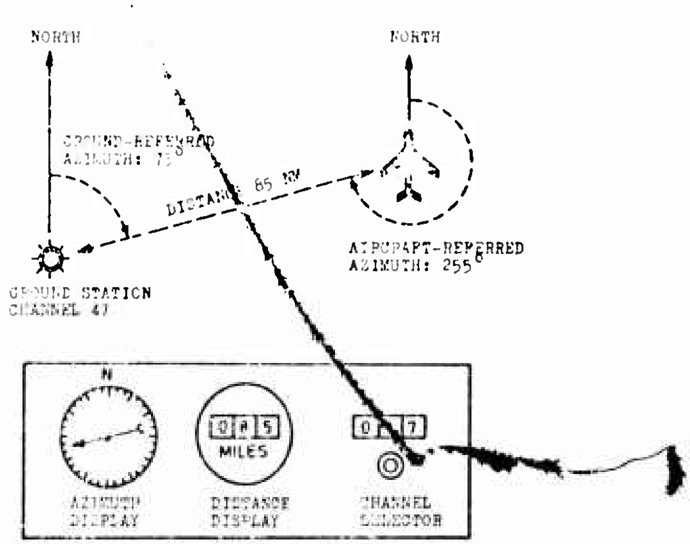


Fig. 1 Measuring and display of TACAN coordinates

1. Introduction

TACAN is accurately specified in MIL-STD 291B and in STANAG 5034. TACAN equipment operates in the frequency range from 962 to 1213 MHz just like DME. This range is subdivided into 252 X and Y channels spaced by 1 MHz. The range of TACAN is 300 km on the average. Five channels in the frequency region around 1030 MHz and five around the frequency of 1090 MHz are reserved for Identification Friend or Foe (IFF) or secondary radar (SSR), and are therefore not available for TACAN.

The association of TACAN channels to the various frequency subranges may be seen from Table 1. Just as in DME, pulse-spacing codes are employed, i.e. the spacing between two pulses constitutes a unique identification code.

Table 1 - Channel distribution in TACAN

(Total: 126 X and 126 Y channels)
PSC = pulse spacing code

	PULSES OF AIRBORNE INTERROGATOR		
	63 X channels PSC 12 /us	63 X channels PSC 12 /us	
	63 Y channels PSC 36 /us	63 Y channels PSC 36 /us	
962-1024 MHz	1025-1087 MHz	1088-1150 MHz	1151-1213 MHz
63 X channels PSC 12 /us	63 Y channels PSC 30 /us	63 Y channels PSC 30 /us	63 X channels PSC 12 /us
RESPONSES OF GROUND STATION TRANSPONDER			

2. Operation Principle

TACAN azimuth is measured by the airborne equipment by discriminating the phase shift between the azimuth-dependent rotating directional radiation pattern and the reference pulses independent of the azimuth, similarly as in the VOR system. To make these measurements possible, the ground station provides the following:

- a secondary radiator rotating around the fixed central antenna at a distance of 7.5 cm with the speed of 15 rps; this results in a rotating pattern used for coarse measurement (Fig. 2);
- nine antenna elements rotating around the fixed central radiator at a distance of 45 cm with the speed of 15 rps that are also excited by radiation coupling; they generate a nine-lobe pattern superimposed on the above pattern (Fig. 3);
- 15-Hz and 135-Hz reference pulses against which the said single-lobe and nine-lobe patterns are measured in order to indicate the azimuth (Fig. 4).

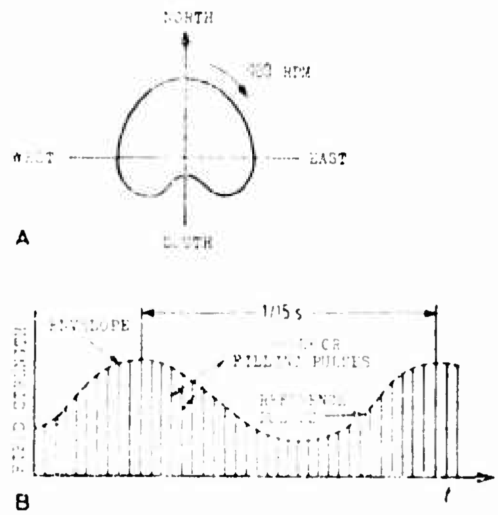


Fig. 2 TACAN azimuth single-lobe pattern

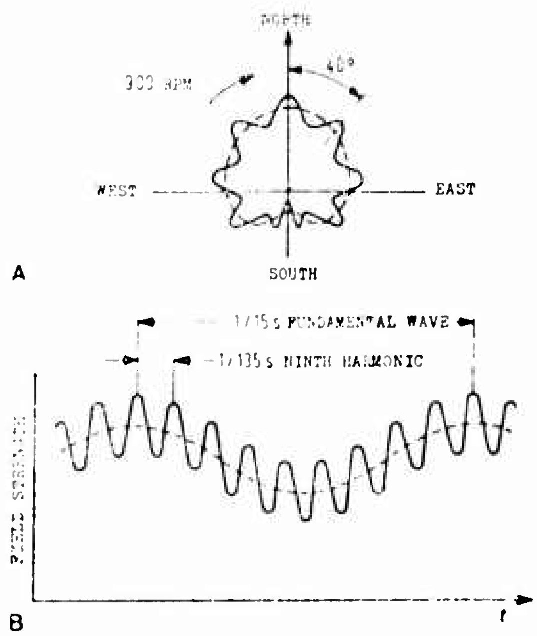


Fig. 3 TACAN azimuth nine-lobe pattern

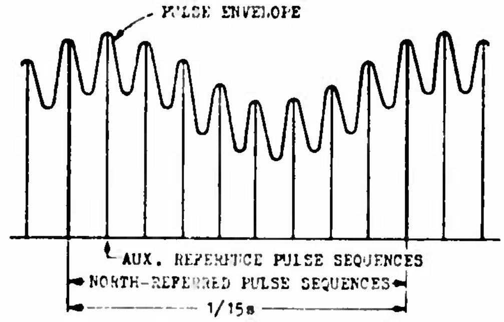


Fig. 4 TACAN reference pulses

The TACAN ground station thus basically consists of the DME described before and additional equipment that permits
(a) modulation of DME and filling pulses with an azimuth-dependent signal and
(b) additional radiation of reference pulses for azimuth measurements.
The TACAN airborne equipment is essentially a DME with attachments for azimuth measurement.

This provides a compatibility inasmuch as any airborne DME can process TACAN distance data while any TACAN airborne equipment can process DME distance data.

Compared with VOR/DME, TACAN features the following significant advantages:

- The TACAN antenna is smaller because of the higher operating frequencies (962-1213 MHz as compared with 108-118 MHz). This is important for mobile use on ships and land vehicles.
- The multi-lobe coarse-fine measuring principle improves accuracy and resolution.
- Azimuth and distance are measured with the same RF channel, which constitutes a more economic solution.

Compared with the doppler radio beacon (DVOR), a certain disadvantage may be seen in the application of the coarse/fine measuring principle for a certain class of users.

The TACAN Airborne Equipment

The airborne equipment is essentially a DME with associated azimuth-measuring unit. A block diagram of existing equipment is shown in Fig. 5 (azimuth part only; for DME see chapter: Distance Measuring Methods).

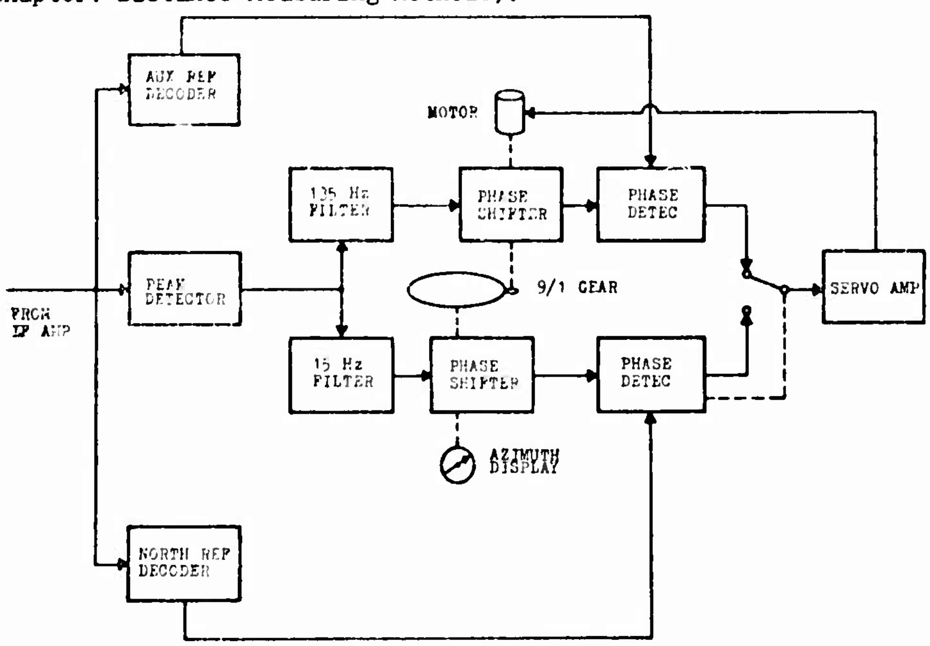


Fig. 5 Azimuth unit of TACAN airborne equipment

The signals received are decoded in this unit, i.e. only one pulse of a pair is processed. Then the 15-Hz and the 135-Hz components are filtered out from the amplitude-modulated signal. The North pulse triggers a corresponding 15-Hz reference frequency while the auxiliary reference pulses synchronize a 135-Hz reference frequency. The phases of the variable 15-Hz and 135-Hz waveforms are continuously measured against the reference waveforms with the aid of a tracking unit comprising two bridge circuits. A gear with the ratio of 9 : 1 is attached to the servo motor in order to interlink the 15-Hz and the 135-Hz systems. In the tracking mode, the servo motor is controlled by the 135 Hz as long as the latter signal is present and as long as the 15 Hz are tracked to an accuracy of 8 to 20 degrees. If the 135-Hz signal is missing or the coarse-measuring error exceeds 20 degrees, then the 15-Hz signal assumes control over the servo. In this way the 15-Hz signal provides unambiguity of the signal while the 135-Hz component enhances the measuring accuracy. Just as in DME, static or dynamic memories will bridge the gaps caused by short-duration fading of signals.

The TACAN Ground Station

The design of the TACAN ground station is sketched in Fig. 6. The antenna array comprises a central radiator designed to handle a bandwidth of about 250 MHz. These antennas have been built-up of up to 11 vertically stacked elements in order to achieve a high vertical antenna gain. Fig. 7 shows a typical vertical-radiation characteristic of a TACAN ground-station antenna.

The secondary antenna elements are embedded into the two plastic cylinders (Fig. 6) rotating around the central and fixed antenna. They are excited by radiation coupling and will amplitude-modulate the pulse sequence radiated by the fixed central antenna. The depth of modulation is about 20 per cent. Rotating synchronously with the plastic cylinders are the sender disks for the reference pulses and for the ground-station identity tone.

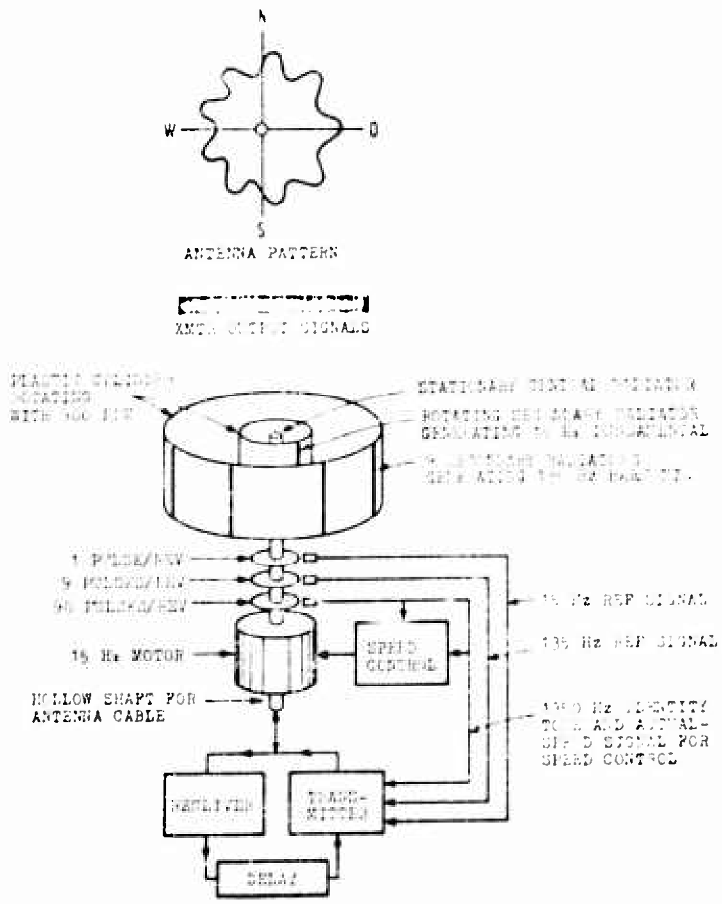


Fig. 6 TACAN ground station

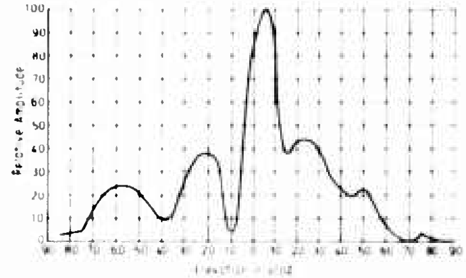


Fig. 7 Typical vertical-radiation characteristic of an antenna (TACAN ground station)

The rated speed of 15 revolutions per second (rps) of the motor (Fig. 6) for the antenna array is derived from a tuning fork and is controlled to an accuracy of max. 1%.

The North pulse sequence comprises 24 single pulses alternatively spaced by 12 and 18 µs. After decoding in the airborne set, 12 single pulses spaced by 30 µs appear across the decoder output. The reference pulse sequences are radiated 8 times for every revolution; the ninth sequence is substituted by the North pulse sequence. They consist of 12 single pulses spaced by 12 µs each. Reference pulses are transmitted with priority over filling and response pulses.

The station-identity tone of 1350 Hz is transmitted every 30 seconds in sync with the pulse sequences of the identity code, thus avoiding mutual interference. The identity code comprises 1350 pulse sequences per second; each sequence consists of 4 pulses with the following spacings: 12 µs, 100 µs, and again 12 µs.

VORTAC

Several countries have introduced the VORTAC system enabling TACAN-equipped aircraft to use the ICAO air lanes. The VORTAC ground station comprises a TACAN instead of a DME equipment (TACAN is compatible with airborne DME). The resulting VORTAC station is sketched in Fig. 8. Civil aircraft will obtain azimuth data from a VOR transmitter and the distance data from the TACAN transponder. Military aircraft obtains both azimuth and distance data from the same TACAN transponder. From the viewpoint of Air Traffic Control (ATC), both types of users can be treated in the same way.

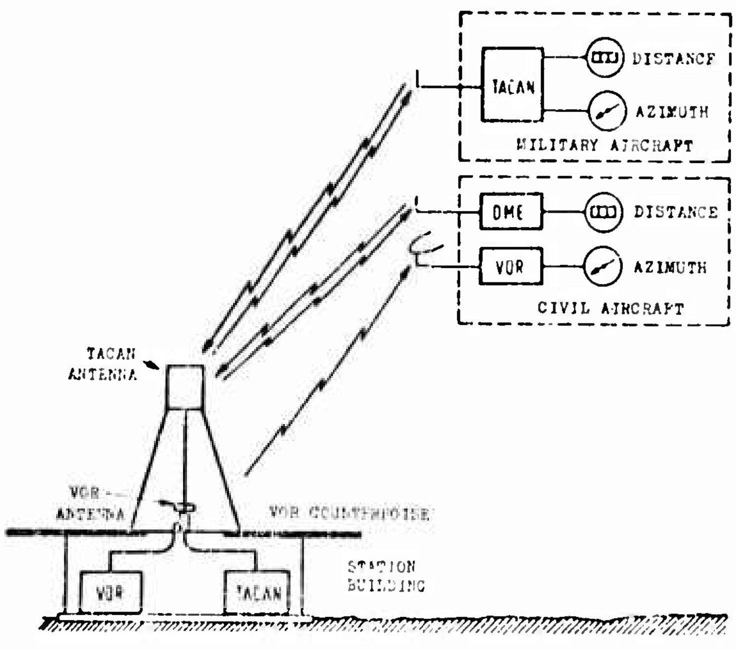


Fig. 8 VORTAC station

3. Development Trends

TACAN has been the first multi-lobe rho-theta system for coarse and fine measurements and has in some ways more than fulfilled expectations. The typical system error is 1 degree; however, TACAN will operate satisfactorily only as long as the 15-Hz system continues functioning. This system is sensitive to reflection interference because its measuring base is short. Particularly detrimental are reflecting obstacles located in the vicinity of the ground-station antenna. An effective remedy is the elevation of the vertical radiation diagram which, of course, will reduce the range for low-flying aircraft. Another remedy is the use of pulses because reflected pulses are received later than the original pulses and can be made ineffective. Other improvements are incorporated in advanced airborne equipment that makes use of only the leading edge of only the first pulse of a pair.

Further development work on TACAN reveals the following trends:

- improvement of ground station reliability
- increase of azimuth and distance measuring accuracy
- introduction of new and improved angle-measuring procedures
- multiple exploitation of TACAN channels for landing etc.
- development of highly mobile ground stations
- suppression of interference by frequency hopping.

The reliability of airborne and ground equipment is continuously being improved by the use of modern solid-state and digitizing technologies and by the elimination of physically movable parts.

The azimuth measuring accuracy is being increased by better antennas switched by electronic means, a wider base of the ground station, and careful airborne measuring of angular signals, predominantly using the digital technology. The distance-measuring accuracy is being improved along the same principles as employed for DME. New angle-measuring principles supported by wider antenna bases are being worked out to overcome the effect of multi-path propagation upon the measuring accuracy.

Multiple exploitation is suggested by the ample endowment of the TACAN frequency band with bandwidth and number of channels. Therefore, both data links within the position-fixing channels and the SETAC landing aid were developed on the basis of TACAN. The feasibility of incorporating additional services is being investigated.

The overall TACAN development trend will probably advance in the direction of multiple exploitation of the airborne TACAN equipment as a central unit of a so-called TACAN family comprising various highly mobile ground stations insensitive to interference and employed for various tasks of tactical radio navigation in air force and army.

4. References

1 ICAO, Aeronautical Telecommunications, Annex 10, Vol. I, Oct. 1965.

2 Kayton, M. and Fried, W. (ed.): "Avionics Navigation Systems", John Wiley & Sons, Inc., 1969, New York, (section on "Radio Navigation" by S.H. Dodington).

3 Poritsky, S. (ed.): Special Issue on VOR/DME. IEEE Transactions on Aerospace and Navigational Electronics, Vol. ANE-12, No. 1, March 1965.

4 Colin, R.I. and Dodington, S.H.: "Principles of Tacan", Electrical Communication, Vol. 33, No. 1, March 1956.

5 Sandretto, P.C.: Electronic Avigation Engineering, International Telephone and Telegraph Corp., 1958.

6 Kelly, R.J.: "A New Light-Weight DME", Proceedings of the East Coast Conference on Aerospace and Navigational Electronics, Baltimore, 1963.

7 Garfield, W.L.: "Tacan. A Navigational System for Aircraft", Proceedings of the IEE, England, Vol. 105, Part B. Supplement, No. 9, p. 298, March 1958.

8 Henney, K. (ed.): Radio Engineering Handbook (section on Avigation Electronics by A Casabona), McGraw-Hill Book Co., Inc., New York, 1959.

9 Metz, H.I.: "International Short Distance Navigation - after the ICAO February Meetings, Proceedings of the National Aeronautical Electronics Conference, Dayton, Ohio, 1959.

10 Report of Department of Transportation Air Traffic Control Advisory Committee (Alexander Report), Dec. 1969, Vol. 1 and 2, U.S. Government Printing Office, Washington, D.C.

11 Graciani, D.: "DME for the future", ITT-FACE, internal report, March 1971.

12 Vogel, H.: "Präzisions-Entfernungsmessung mit TACAN", SEL, internal report.

13 SC-117, Radio Technical Commission for Aeronautics: "A New Guidance System for Approach and Landing", Doc. No. DO-148, 18 December 1970, Vol. 1 and 2

14 DME-Bordgerät 3574B, LMT internal report, September 1971.

PART IV-B – SELF-CONTAINED NAVIGATION AIDS

INERTIAL NAVIGATION AND AIR TRAFFIC CONTROL

by
W.H. McKinlay
B.Sc., C.Eng., F.R.Ae.S., M.I.E.E., F.R.I.N.
Inertial Systems Department
Ferranti Ltd.,
Silverknowes
Edinburgh EH4 4AD - U.K.

PART 1

INERTIAL NAVIGATION AS A FACTOR IN ATC

PART 2

INERTIAL NAVIGATION SYSTEMS AND TECHNOLOGY

Preceding page blank

INERTIAL NAVIGATION

AND

AIR TRAFFIC CONTROL

by

W.H.McKinlay
B.Sc., C.Eng., F.R.Ae.S., M.I.E.E., F.R.I.N.

Inertial Systems Department
Ferranti Limited
Silverknowes
Edinburgh EH4 4AD
U.K.

PART 1 INERTIAL NAVIGATION AS A FACTOR IN ATC

SUMMARY

The history of the development of navigation within Air Traffic Control Systems is traced briefly and the requirements for effective navigation and flight path control are discussed. They are related to the adoption of inertial navigation and to the possible future extention of its use. Typical airborne system configurations are described together with the facilities provided by them. Reference is made to systems for both civil and military aircraft. The principles of Inertial Navigation are then described briefly with a review of the standards of accuracy and reliability being achieved.

1. INTRODUCTION

Any Air Traffic Control System depends for its operation on the precision with which aircraft can be confidently assigned to specific flight paths in space, and also in time. The integrity of the system, and hence the level of flight safety afforded by it, depends on the probability of aircraft achieving known standards of navigation accuracy, and/or their precise position being known to ATC.

These aims are obtainable by a combination of airborne navigation and ground-based surveillance, the former being of most interest here.

The alternative approaches to navigation are through the provision of ground-based radio aids and self-contained systems confined to the aircraft: in practice some combination has always been used.

In the two decades following the end of the Second World War the widest application of new technologies to navigation took place in the ground-based radio sector. In particular the air spaces over the most highly developed continents, the U.S.A. and Europe, were organized around extensive airway structures based on VOR, with the later addition of DME. Navigation over water and less highly populated land masses was achieved by the use of dead reckoning complimented by position fixes derived from whatever sources were available. Over the North Atlantic many operators used Loran A for this purpose.

The primary disadvantage of such classical navigation techniques was the work load imposed on the crew: a combination of data reduction, plotting and the exercise of judgement to weight the various information sources. There was a further consequence of importance in calculating acceptable separation standards: widely varying standards of accuracy and a real possibility of 'blunders', leading to a significant proportion of gross errors.

The operators of long range civil and military aircraft thus had a common use for an aid having the following properties:

Continuity and consistency of operation over all sectors, throughout the world.

Ease of handling by the crew: preferably pilot interpreted.

Highest possible accuracy.

Reliability.

The need for a world-wide capability and a desire to maintain the factors influencing accuracy and reliability under operator control pointed the way to a primarily self-contained solution.

Navigation by Doppler Radar was the first such solution to be commercially available but suffered from certain significant short-comings:

Accuracy dependent on the heading reference, in most cases derived magnetically.

Performance variations over water, in certain sea states.

A significant installation problem.

However Doppler was quite extensively used, particularly when the work load involved in interpreting the data was relieved by a relatively simple analogue computer, operating in along and across track co-ordinates.

In the late 1960s Inertial Navigation was introduced by a number of civil operators, following a period of development as a basic sensor in certain types of military aircraft. Apart from its basic superiority as a self-contained aid this wider application of IN was, and is, highly successful for a further important reason. Fortuitously the introduction of these systems coincided with the availability of a new generation of relatively reliable, compact, airborne digital computers.

The resulting inertial systems have significant advantages over the previous Doppler systems because:

Their advanced gyroscope technology has solved the problem of achieving an accurate heading reference.

Their high navigation accuracy has proved to be extremely consistent.

They have become more reliable because of advances in the technologies involved.

The inclusion of digital computing has provided:

A range of monitoring facilities and checks previously unobtainable.

Satisfactory pilot operation by using the computer to store way-points expressed as latitude and longitude.

Automatic flight guidance by versatile AFCS coupling.

Better and more comprehensive pilot displays.

By a combination of the above a higher degree of protection against 'blunders', or 'flight technical error'.

As will be seen later all of this has an immediate impact on the potential viability of an oceanic ATC system, where separation standards may have to be reduced. But in addition the full potential of Inertial Navigation has further implications for ATC in continental en-route air space, in Terminal Areas, and in approach and landing.

This is because the INS is not merely a navigation aid, but a sensor of the aircraft velocity vector. It produces this data in three dimensions, and also provides high accuracy data on roll and pitch attitude, and angular rates. It does so in a form free from noise, and immune to errors caused directly by aircraft manoeuvres. It is the potential foundation of a new generation of Guidance and Control Systems. The corresponding improvements in aircraft capability can in turn influence the evolution of future ATC systems.

2. REQUIREMENTS FOR NAVIGATION AND FLIGHT PATH CONTROL IN AN ATC SYSTEM

In the context of this paper it is of interest to distinguish a number of different flight regimes, and to see how Inertial Navigation applies to them.

2.1 Long Range En-Route Navigation

Here, for example in trans-oceanic sectors, IN is likely to be the sole means of navigation. The requirement is to achieve a given standard of accuracy, with a specified probability of its achievement.

A considerable amount of work has been done in assessing the accuracy and integrity required for N. Atlantic operations, and the majority of commercial INS equipped aircraft have been cleared largely with this route in mind.

Containing the probability of a conflict in this or any other route structure depends on the navigation accuracy achievable and the presence or otherwise of a known standard of surveillance. In general:

Present standards of separation are postulated on a mix of IN equipped aircraft and non IN equipped aircraft.

IN has established itself to the point at which the separations likely to be needed in the late 1970s could be achieved by its use.

But there is a point at which a further decrease in separations to below an order of 30 n. mls laterally would also require satellite surveillance.

The modelling techniques involved in such assessments have been described in detail elsewhere. (References 1 and 2).

With the recent downward revisions of the predictions of North Atlantic movements it appears that the capability now attributed to IN will become essential in the late 1970s if an operator is to enjoy as near to optimum routings as is possible. The separations employed will be of the order of 60 n. mls.

But it must be stated that the administrations do not require INS per se, but a demonstrable capability of the same order. It is possible, but not firmly established, that an alternative solution such as Omega will also prove acceptable.

The assessment of the acceptability of INS as providing a given capability over the North Atlantic has been a complex process involving a consideration of the following variables, among others:

INS errors propagate within a given system with time. The mode of propagation is relatively complex because of the number of possible error sources, and a system mechanization which interrelates them in numerous ways. The system is not positively error bounded.

A single INS is based on a single set of basic sensors and is thus essentially vulnerable to undetected failure. Multiple redundancy is called for, and the question of how a flight crew can diagnose the fact that a particular system is drifting away arises.

The original accuracy requirement for the North Atlantic, promulgated by the FAA in 1966, was the limitation of cross-track error to a maximum of ± 20 n. mls, along-track to ± 25 n. mls. This was to be determined and demonstrated on a 95% probability basis. (Advisory Circular 25-4, February 1966) (Reference 3)

2.2 En-Route Navigation in a Domestic Environment

Aircraft arriving in such an environment from a long range sector become ground referenced as opposed to self-contained when they are 'collected' in two ways:

By entering the coverage of the VOR/DME system,* which thus enables them to fly airways in the normal manner.

By coming under radar surveillance.

In the early days of INS development a relatively simple concept was accepted: such aircraft would then behave as if they were not INS equipped and be handled thus. It was conceded that if Area Navigation became adopted widely it would be based on the use of digital computers combining the data from a number of sensors, one of which might be the INS.

In fact it has now been observed that traffic densities and the resulting cost effectiveness of area navigation have not developed as rapidly as was once predicted, with a consequent slowing down of the adoption of the concept in practical equipment terms.

But meanwhile operators who have INS equipped aircraft have been monitored as to their compliance with the present VOR/DME based airways structure, and it is clear that INS is improving their performance. This is probably because most present day aircraft have somewhat unsatisfactory arrangements for direct VOR/AFCS coupling, the coupling often being achieved by visual instrument monitoring, the basic guidance mode being magnetic heading/AFCS.

When such aircraft continue to use INS/AFCS flight path guidance the total airborne system comes nearer to meeting a significant requirement which is particularly favourable to inertial navigation: an ability to couple closely to a desired flight path, to ignore wind changes or wind shear, and to manoeuvre precisely when track changes are required; e.g. over a VOR. Raw VOR continues as a basic monitor of track keeping. Beam noise is of course ignored. The role of the INS as a guidance tool based on the aircraft velocity vector relative to the earth's surface becomes apparent.

It has also become apparent that INS equipped aircraft already possess a form of Area Navigation in that the Arinc 561 computer can accept a series of way-points defined in terms of latitude and longitude. This permits the negotiation of special clearances which can if necessary depart from the normal airways structure. The requirements are that ATC can organize such clearances, and that INS capability as regards accuracy is sufficient for their execution. Such operations are now being conducted in the U.S.A. (Reference 4, Arinc 561)

It can thus be seen that the first generation of Inertial Systems to be used regularly en-route in the domestic environment are meeting specific requirements which have always existed:

Better dynamic track-keeping.

Better automation of certain manoeuvres.

The manner in which these requirements are met can only improve when Inertial Navigation is used in more flexible and comprehensive airborne system configurations, with provisions to mix IN and radio aids, or with more versatile pilot-equipment interfaces.

2.3 Terminal Area, Approach and Landing

Present TMA operations are based primarily on radar vectoring, followed by the acquisition of the ILS Localiser in the horizontal plane. Depending on the AFCS fit landings are possible over a range of weather minima.

The operational requirement in its simplest terms is to permit ATC to assign to arriving aircraft flight paths which will result in optimum longitudinal separation for landing, the departure problem being somewaht similar.

Significant trends include:

The need to discriminate in longitudinal separation to cope with wake turbulence, particularly from wide bodied jets.

The growth of automation on the ground to aid the controller in this task.

The desirability of improving the acquisition and subsequent holding of the runway centre line to facilitate parallel runway operations.

A need for better regulation of climb or descent profiles in the vertical plane.

The contribution of Inertial Navigation to these needs lies in its ability to generate extremely smooth velocity data in three dimensions, accurate in the short term. It is thus possible to smooth radio guidance and improve the general level of flight path control.

It has been pointed out that a clear aim of MLS is to provide much smoother approach guidance, with freedom from noise or beam bends resulting in ILS from multi-path effects, movements of other aircraft etc.

But it must be pointed out that the present ILS system is likely to co-exist with MLS for at least 15 years, or up to the 2000. Many aircraft cleared to Category 3 will have such clearance on the basis of the established integrity of ILS.

Considerable work has been done on both sides of the Atlantic to demonstrate the benefits of introducing inertial smoothing into automatic approach and landing systems based on ILS. (References 5 and 6). Therefore the potential contribution of IN to enhanced flight path control in three dimensions applies in the TMA down to the final approach phase.

The use of IN as an aid to movement on the ground is also a possibility.

3. AIRBORNE SYSTEM CONFIGURATIONS

3.1 History

The essential sub-assemblies in an Inertial Navigation System are:

(a) The Platforms, including the sensors.

(b) The Platform Electronics.

(c) The Computer.

(d) Special purpose Power Supplies Etc.

(e) Pilots' Controls and Displays.

In the earliest military systems (a) was usually mounted separately. (b) and (c) were contained in a rack mounted unit.

But in the latest military and commercial systems it is customary to pack (a), (b), (c) and (d) in a single box. Figure 1 shows such a unit, part of the Ferranti Digital Inertial Navigation System.

Systems for commercial aircraft conform to Arinc Standard 561, which standardises mounting arrangements, dimensions and interfaces with cooling air, power supplies and other avionic systems. (Reference 4)

The cockpit units consist of:

A Mode Selector Unit, used pre-flight to control system alignment etc.

A Control/Display Unit used for operational handling in the air.

3.2 Typical Configurations Used in Commercial Aircraft

A typical Arinc 561 System (Reference 7) has the signal interfaces shown in Figure 2. It will be seen that as well as functioning as an inertial navigator the system:

Provides the azimuth gyro reference for the compass system.

Acts as an altitude reference for flight instruments.

Provides autopilot steering.

The Control/Display Unit includes a key-board and alpha-numeric readouts, resulting in the following functions:-

Ability to set in 9 way-points, i.e. define a track.

Ability to select and read out:

Present Position (lat/long.)

Selected Way-point (lat/long.)

Cross-track Deviation/Track Angle Error

Distance or Time to Go.

Present Track Angle/Ground Speed

True Heading/Drift Angle.

Wind Velocity/Wind Angle.

Desired Track/System Status.

In practice a typical commercial installation consists of two or three systems to provide redundancy. With this aim there is typically (Reference 1) segregation of the systems in flight. The exception is the manual loading of flight plan data (way-point positions), when a single key-board one one CDU may be used to load two or three systems simultaneously.

3.3 Typical Military Aircraft Configuration

The tendency to develop special purpose integrated avionic systems for military aircraft results in a wider range of possibilities. For example the INS may be used as a navigation sensor feeding a Central Digital Computer in which the guidance computations are made.

Figure 3 shows the Control/Display Unit for the Ferranti System. It will be seen that the facilities of mode selection are included to avoid the need to install a separate cockpit unit for this purpose as in Arinc 561. But in this and the majority of military systems there is provision to insert way-points, and thus generate track guidance.

In many military systems other functions not relevant to ATC are to be found: e.g. concerned with weapon delivery or other special operational tasks.

It is also common to fit only a single INS in military aircraft. The redundancy for navigation is provided by TACAN and other aids. Radar and other sensors associated with the primary mission increase the total level of redundancy.

In some military aircraft present position data is displayed to the pilot by a Moving Map Display, such as the Ferranti unit shown in Figure 4.

However in general it may be said that commercial and military INS equipped aircraft, while having different overall system configurations, possess the same inherent navigation capability as seen by ATC.

4. OPERATIONAL CONSIDERATIONS IN THE USE OF INS

4.1 Principles

The principle of any INS (Reference 8) involves a cluster of inertial instruments, consisting of gyroscopes and accelerometers, stabilized against aircraft motion by a gimbal system. This is the platform.

The gyroscopes can detect any motion of the cluster relative to space about three measuring axes, roll, pitch and azimuth. They are the basic position sensors in the servo systems controlling the gimbals. By the inclusion of a fourth redundant roll gimbal the platform can be made topple-free through all aircraft manoeuvres.

The two vertical gyroscope channels are associated with two horizontal accelerometers which can be considered most simply as being aligned N-S and E-W. The possibility of inertial navigation results from the fact that an accelerometer maintained perfectly in the horizontal plane can furnish a signal capable of being successively integrated electronically to give velocity and displacement.

The problem of defining the vertical is tackled by constructing a feed-back loop around an accelerometer and a gyroscope so that the combination is 'Schuler tuned': i.e. it has a natural period of oscillation about the vertical of 84 minutes. Such a vertical is essentially pendulous, but does not respond to vehicle accelerations.

As the aircraft traverses the earth the cluster must be rotated to take account of its curvature. As the earth rotates the cluster too must be caused to rotate appropriately. The computer generates these correction terms from the basic geometry and furnishes commands to torque the gyros so that they precess in the desired manner.

The accuracy of inertial navigation therefore depends on the following factors, among others:

(a) Entering into the computer the precise geographical co-ordinates of the aircraft position on the ground at switch-on. (Essentially the accuracy is such that <u>ramp</u> position is required.)

(b) Levelling the platform precisely to the local vertical.

(c) Achieving a precise knowledge of the alignment of the instruments relative to true north: the initial azimuth alignment.

(d) Having gyroscopes of a very low intrinsic drift, and accelerometers of high accuracy.

(e) Controlling all the parameters involved in the subsequent measurements and computations extremely precisely.

There are two further rules. The alignment must be completed with the aircraft stationary at its ramp position. Also, since subsequent operation depends on processes of continuous integration, the system must be powered continuously, and run continuously.

4.2 Operating Considerations

The capability of an INS equipped aircraft within an ATC system is bounded by certain basic operational considerations which are common to most systems, although in detail they may be handled differently.

The basic principles of the operational aspects have been outlined and a brief technological description of INS is integrated in this paper but for in-depth information, the reader is referred to the bibliography listed under Section 7.

5. ACCURACY AND RELIABILITY

The accuracy of inertial navigation systems is generally specified as CEP (Circle of Equal Probability). This implies that the system will exhibit a circular error of the level specified on 50% of all flights.

There has been a steady increase in inertial navigation accuracy (Figure 5). In the early 1960s CEP's of the order of 2½ n. mls per hour were characteristic. But by the late 1960s proven equipments in civil and military service were realising accurac order of 1 n.m/hour CEP, and this has been the characteristic performance goal for high quality inertial systems since that date.

The fact that accuracies of this order can now be achieved in the rugged environment of a military strike aircraft as well as in the most benign commercial aircraft environment is indicative of the extent to which detailed problems have been solved. The accuracy depends heavily on the adequacy of the initial pre-flight alignment.

Complex considerations arise in any attempt to quote the reliability of inertial systems. In the early days there were predictions that the operational reliability would be low because of the unknown nature of the tails of the distribution of errors. In fact these predictions have not been realised and anomalous behaviour of this kind is less frequently encountered than obvious failure. This is largely because unsatisfactory operation of the instruments becomes apparent in the alignment process. Most systems include provisions for the processor to monitor the alignment and to display to the crew a confidence number, between 9 and 0 in the case of Ferranti systems. Failure to achieve the desired level of confidence in alignment is indicative that the system will not deliver its specified performance in the ensuing flight. The fact that an inertial system is effectively tested in this way before take-off has increased the effective reliability of these systems in the air. In general a distinction is made between:

Dispatch reliability: the extent to which INS failures prevent departures on schedule.

In flight reliability, quoted as an MTBF.

In practice MTBF's have increased from a few hundred hours to figures of the order of 1,000 hours. Dispatch reliability and the minimising of delays have been greatly assisted by the availability in the INS of comprehensive built-in test facilities including the use of the computer to diagnose fault conditions.

6. CONCLUSION

It has been shown that the wide application of inertial navigation has followed from the rapid development of the technology and the fact that such systems are truly self-contained. The fact that a modern INS includes a digital computer providing a high degree of flight automation is also significant.

In ATC terms the INS has made its greatest impact so far in long range sectors, particularly over the North Atlantic. Present levels of INS performance are sufficient to realise the separation standards which current predictions of traffic growth indicate as necessary in the 1970s.

In domestic air space INS aircraft have shown consistently good track keeping performance when flying on airways largely because the stability of the INS guidance data has eased the task of the automatic flight control systmm. The fact that an INS can be programmed to follow any desired route profile through a series of way-points gives it an intrinsic area navigation capability which is already being used.

In future generations of aircraft inertial systems will enable better flight path control in the TMA and there is scope for improving approach and landing performance by using the inertial reference in conjunction with radio aids.

7. REFERENCES

1. Tyler, Stepner and Sorensen: An ATC/Surveillance Modelling Approach for Specifying Lane Separation Standards; AGARD Conference June 1972.

2. P.G. Reich: Analysis of Long-Range Air Traffic Systems; Journal of the Institute of Navigation; January 1966, April 1966, July 1966.

3. Federal Aviation Agency Advisory Circular 25-4: Inertial Systems; 18th February 1966.

4. Arinc Characteristic 561: Inertial Navigation System.

5. N.H. Hughes, B.L.E.U., R.A.E., Bedford: The Use of Inertial Information to Improve Automatic ILS Approach Performance; AGARD Conference Proceedings No. 54, Hybrid Navigation Systems; September 1969.

6. R.J. Pawlak, DOT Transportation Systems Centre, Cambridge, Mass: Inertially Augmented Approach and Landing Systems.

7. J.H. Sheldrick, A.C. Electronic Division, General Motors Corporation: Inertial Navigation Systems for Subsonic and Supersonic Commercial Aircraft; Proceedings of SAE National Aeronautics Meeting, New York; April 24-27 1967.

8. C.T. Leonides: Inertial Navigation for Aircraft; Scientific American; Vol. 222, No. 3; March 1970.

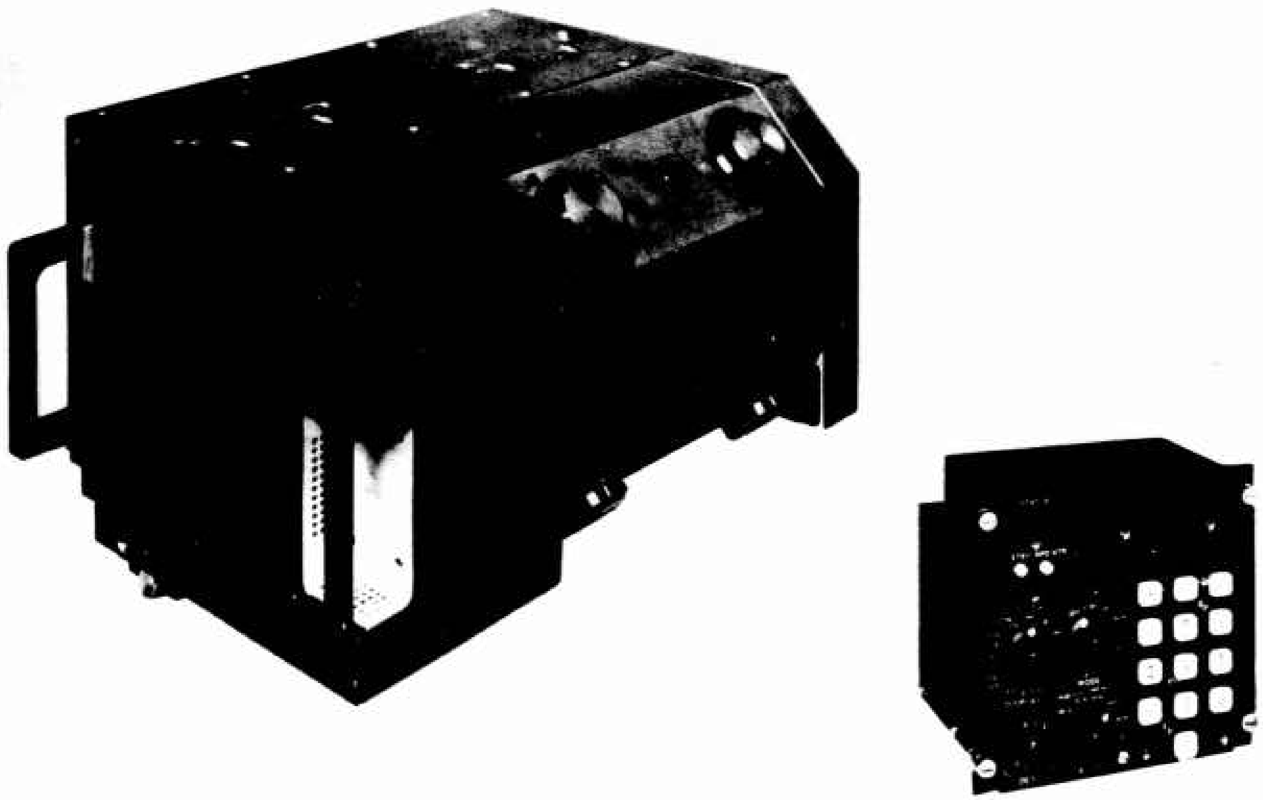
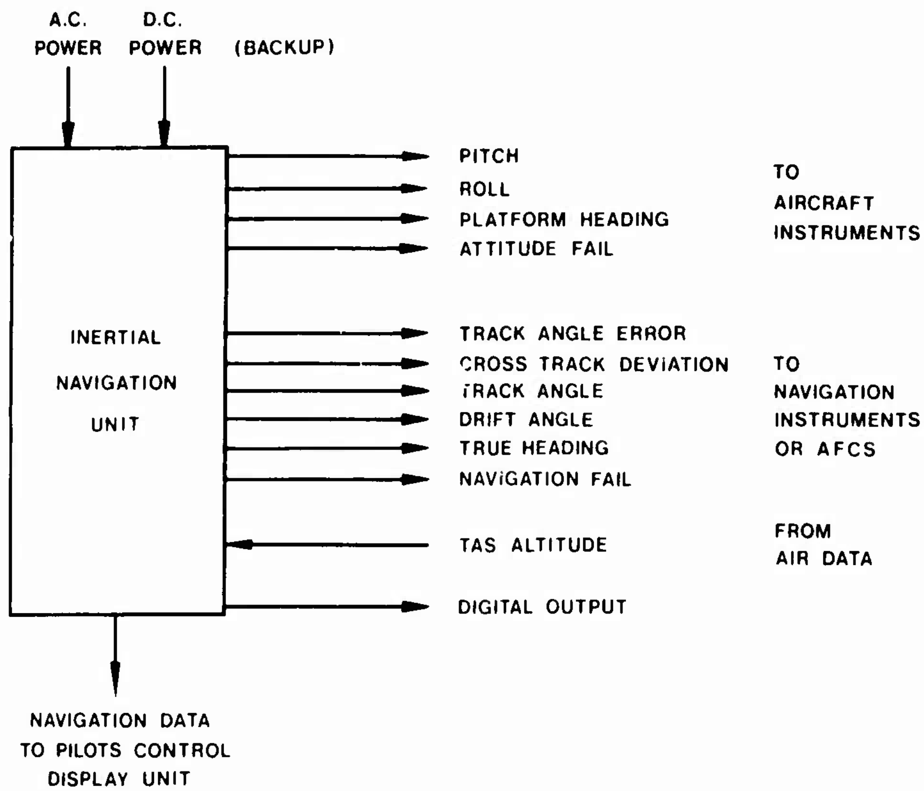


Figure 1
Digital Inertial Navigation System

FIGURE 2 PRINCIPLE ARINC 561

SYSTEM SIGNAL INTERFACES



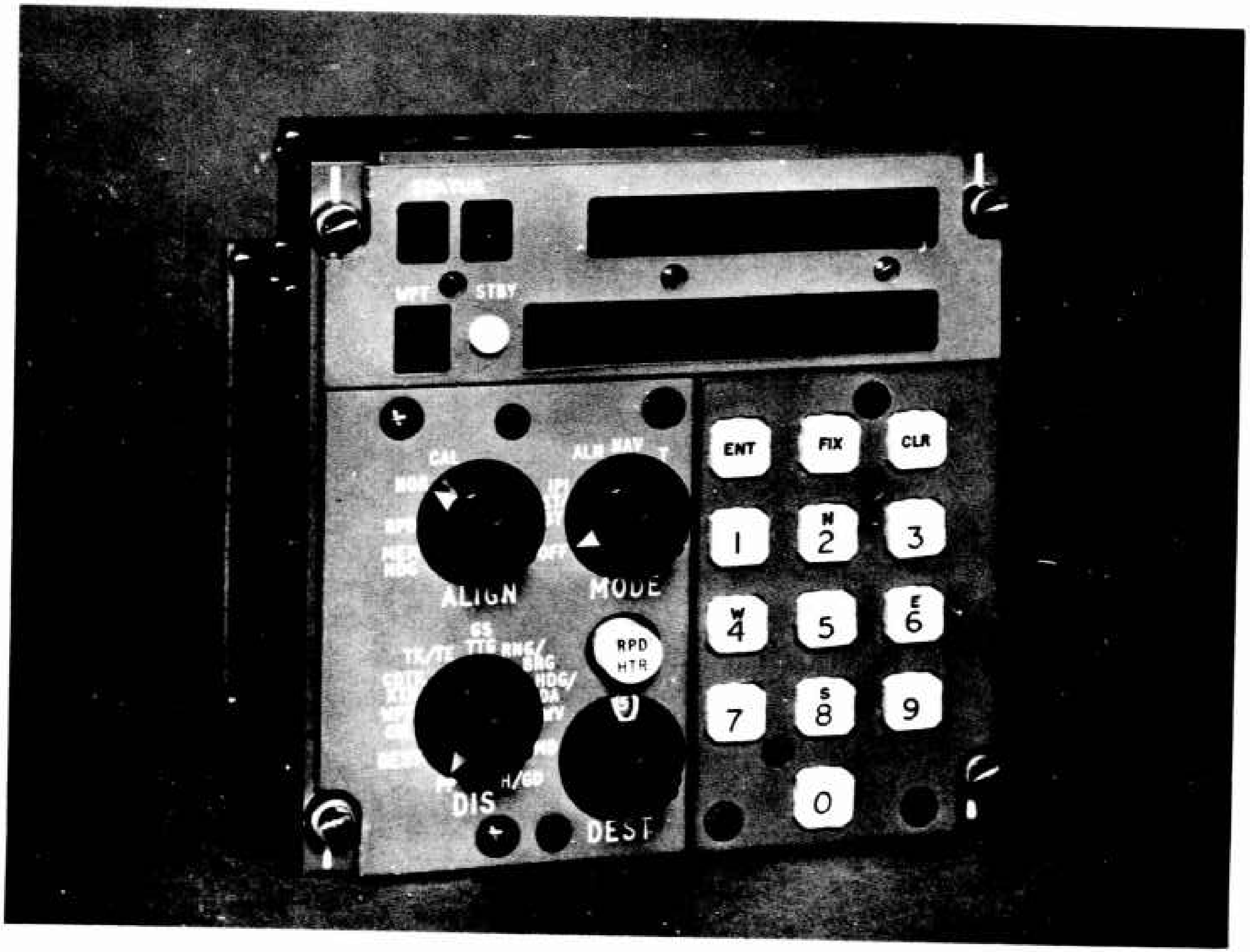


Figure 3

Typical Control and Display Unit

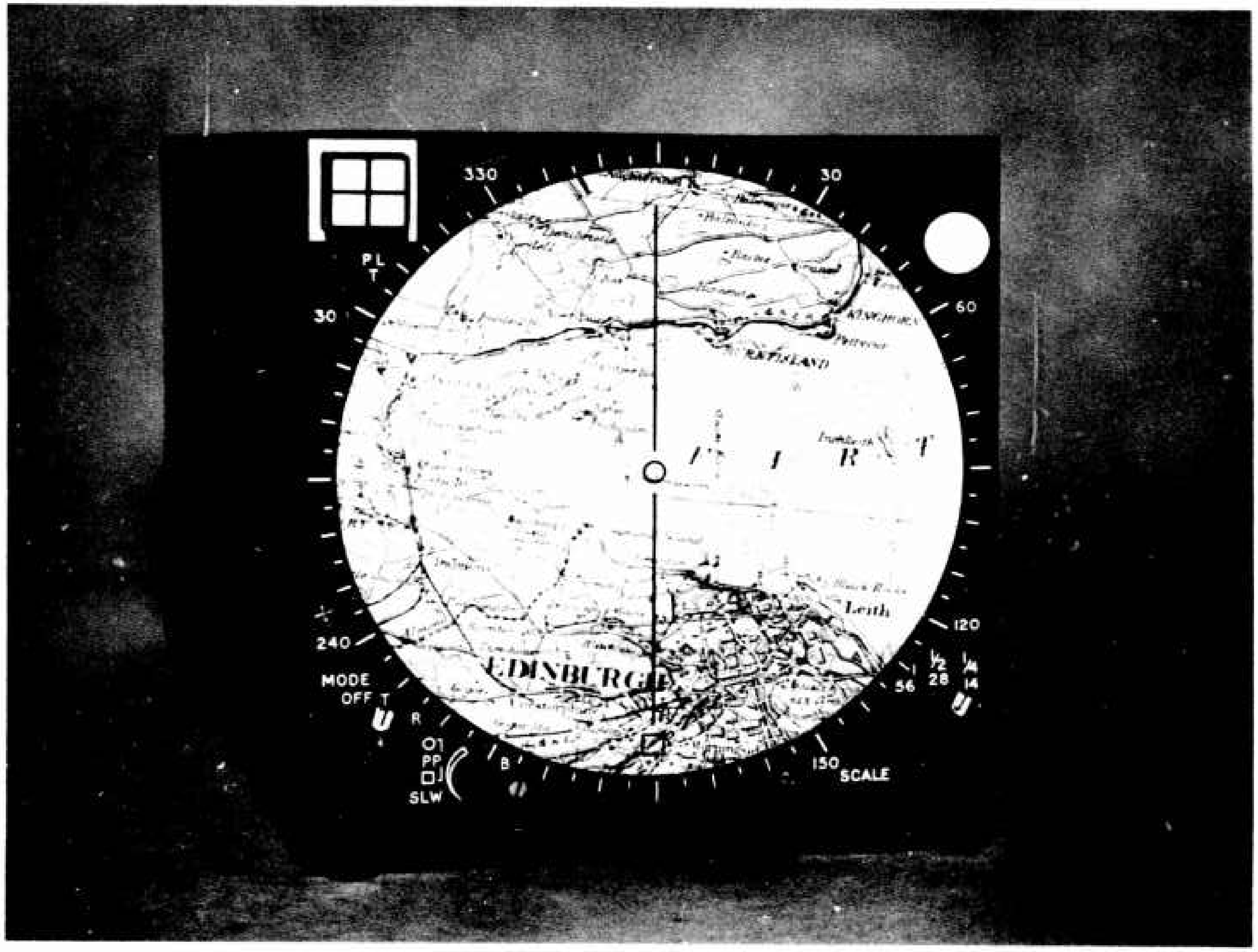
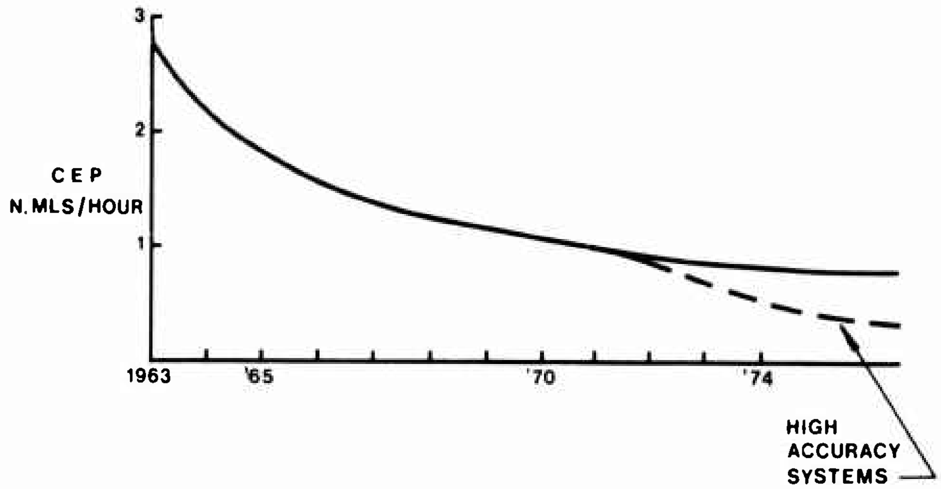


Figure 4

Moving Map Display

FIGURE 5 IMPROVEMENTS IN I N ACCURACY



NOTE:

THERE ARE OF COURSE, STILL CONSIDERABLE VARIATIONS BETWEEN DIFFERENT SYSTEMS IN DIFFERENT ENVIRONMENTS.

PART 2 INERTIAL NAVIGATION SYSTEMS AND TECHNOLOGY

SUMMARY

The principles of Inertial Navigation are reviewed in a simple manner, intended as background information for specialists in other fields. The mechanisation of a typical system is described and alternative methods are referred to. The problems of securing an initial azimuth alignment and compensating for errors within the instruments are referred to. There is then a brief description of the technologies involved in producing inertial instruments together with some indication of possible future developments. A section is devoted to the problems of specifying the accuracy of inertial systems and assessing it in practice. There are references to electronics and digital computation as applied to inertial navigation. In conclusion the significant improvements which have taken place in flexibility and ease of maintenance are referred to.

1. INTRODUCTION

The purpose of this second part is to provide a concise summary of the essential features of this technology for the benefit of workers in the more general fields of navigation and Air Traffic Control.

2. SOME FUNDAMENTAL PRINCIPLES

2.1 General

A full mathematical treatment of a basic INS is beyond the scope of this paper, and readers who are interested are referred to one of the many standard texts on the subject. However a qualitative visualisation of the principles involved is necessary to bring the technological problems into perspective.

The basic principle is that a suitable accelerometer can produce an output which can be integrated successively to give velocity and displacement along the axis of measurement. (Figure 1) Clearly an orthagonal triad of instruments in free space can produce three components from which a single acceleration, velocity or displacement can be derived.

Applying this principle to an aircraft moving on the earth's surface introduces the problem that an accelerometer tilted from the local vertical will also sense a component of 'g': see Figure 2. There is thus a need either to maintain the triad aligned to the vertical, or to know the misalignment and compensate for it.

In state of the art gimbaled systems this is achieved by mounting the sensors on a platform stabilized against aircraft rotational movements by means of gyroscopes. In the present example these are 'single degree of freedom' instruments. Thus the instrument 'cluster' consists of three gyroscopes and three accelerometers. (See Figure 3.)

Each gyroscope is essentially a directional 'memory' which provides a space reference about its sensitive axis. But an important feature is that by the application of a known torque the gyroscopic reference can be caused to precess at a known rate.

The long term vertical must be maintained by some mechanism which is independent of accelerations caused by vehicle motion, the task being to prevent tilt components from arising. This is accomplished (Figure 4) by the classical 'Schuler loop', in which the accelerometer/gyro combination is caused to have a natural frequency similar to that of a pendulum of earth's radius R. This is of the order of 84 minutes.

In the example being considered the platform is mechanized as follows:-

(a) The azimuth gyroscope is used to provide a directional memory, torqued to offset the effects of earth's rotation. It stabilizes the azimuth gimbal to true north. Hence, the two vertical channels operate in N-S and E-W axes.

(b) Components of the vertical gyro outputs are used to stabilize the cluster against roll and pitch disturbances.

(c) As a refinement there is a second redundant roll gimbal so that the complete gimbal set is fully aerobatic and cannot lock due to a Hookes Joint effect in the presence of a 90^o displacement, as for example in a loop.

(d) The accelerometer outputs are corrected for Coriolis acceleration etc.

(e) By means of a computer initialised to starting position the N-S and E-W velocities are integrated and converted to give continuous position data in latitude/longitude.

(f) The gyroscopes are all torqued to offset computed earth's rotation effects, and the angular orientation changes to compensate for movement over the earth's surface are similarly introduced.

In the above system the platform is 'north slaved': i.e. the horizontal accelerometers actually operate N-S and E-W at all times. However it is not always the practice to achieve this by torquing the azimuth gyro. In 'free' or 'wander azimuth' systems the horizontal channels are permitted to assume an arbitrary orientation in azimuth. Provided that this is known at all times the computer has simply to resolve the two measured components to derive N-S and E-W velocities, and hence change of latitude or longitude.

2.2 Strap-Down Mechanization

In the conventional gimbaled system the instrument 'cluster' is maintained physically oriented to the vertical by means of the gimbal system, and so isolated from aircraft motion. However an alternative is clearly to mount the sensors on the airframe itself in which case the rotational and acceleration effects due to aircraft motion will appear in their outputs.

It is possible to remove these effects by data processing: effectively the gimbal system is mechanized electronically. However this solution introduces further problems. For example most types of inertial gyroscope have a small physical arc of measurement: they are normally operated in a nulled condition through the action of the gimbals, needing only to be precessed to take account of earth's rotation, or vehicle motion around the earth.

When strapped down to the airframe the gyroscopes must accept high torquing signals so that they precess at the full airframe rotation rates. This involves accurate torquing about known axes at rates of say up to 300°/sec instead of say 30° per hour, which has considerable design repercussions.

2.3 Initial Azimuth Alignment

The accuracy of an inertial navigator depends on the orientation of the instrument cluster being precisely known in azimuth. An initial azimuth error has a number of effects. First order it introduces a navigation error cross-track. But it also causes the incorrect introduction of corrective torques to the gyroscopes, with a resultant complex propagation of system errors in flight.

In practice the precision of azimuth alignment sought is of the order of 4-6 minutes arc.

The process normally used is gyro-compassing.

It is simplest to visualise the system in the aircraft, stationary, at a known ramp position, with the N-S and E-W navigation channels aligned geographically, but with a small error. The platform is aligned to the vertical.

The effect of earth's rotation is to cause the platform to attempt to 'topple' about the local meridian. i.e. The E-W channel sees the full effect of earth's rate, while the N-S channel sees only a small component, due to the azimuth error.

In its simplest form the gyro-compassing process consists of torquing the azimuth gyro in a sense to reduce this signal to zero. It will be observed that:

(a) The vertical gyroscopes must have a drift significantly lower than the rate being sensed: essentially this implies full IN performance.

(b) Aircraft buffeting will introduce noise into the loop: this accelerometer noise must be filtered out.

(c) If the vertical gyroscopes are drifting the drift in the E-W channel can be calibrated. Any drift in the N-S channel is indistinguishable from an azimuth error.

Clearly it is not vital to physically rotate the platform. Many systems operate on the free azimuth principle in which case data processing is used to deduce the cluster alignment relative to north, and to filter out the effects of noise.

2.4 Compensation for Errors

2.4.1 Errors arise in gyro-compassing, and in subsequent navigation, from a variety of causes including:

(a) Unknown gyroscopic drifts.

(b) Incorrect torquing of gyros, for example due to the scale factor of the torquers themselves.

(c) Accelerometer errors.

(d) Errors in electronic interfaces.

2.4.2 In considering the errors it is important to appreciate the small magnitudes which are significant in Inertial Navigation. Table 1 gives some simple relationships to indicate the orders of magnitude involved.

Gyroscope drifts are significant in several respects:

(a) A random component is significant in that it cannot be compensated.

(b) A day to day component, appearing from one warm-up to another, can possibly be compensated.

(c) A known consistent bias can be catered for.

2.4.3 Two separate approaches have been adopted in state of the art systems.

(a) To manufacture instruments with minimum random drift but take some advantage of the ability to measure performance on the ground and compensate.

(b) Cluster rotation, in which the whole instrument cluster is turned in a defined manner. In alignment this enables drifts to be measured more comprehensively. The restrictions in 2.3 (e) do not apply. In flight drift effects tend to 'commutate out'.

It should be pointed out that systems of type (a) above can be submitted to a ground calibration if the platform is precessed in azimuth and measurements taken at more than one orientation.

But ground running of this kind is time-consuming and operationally inconvenient. In the limit it approaches scheduled maintenance.

The various manufacturers use combinations of techniques to contain the problems of gyroscope drift.

The technique employed by Ferranti Limited is to emphasise intrinsic instrument performance so that special calibration procedures are unnecessary in normal fast response military operations.

See Reference 1 regarding this, and the technique of rapid alignment.

3. INERTIAL GYROSCOPES AND ACCELEROMETERS

3.1 General Principles

The general principles of the gyroscope are well known. A rotating wheel has 'rigidity in space', i.e. in its perfect form its axis will point at a fixed point in space.

An applied torque will cause the wheel to precess about an axis 90° removed from the axis of application.

This is the basis by which a gyroscope can be torqued to any desired alignment, but also of course a potential error mechanism. Applied torques due to friction in the suspension, mass unbalance, etc all cause drifts.

The essence of the design of an inertial gyroscope is to minimise these drifts, and to make them several orders smaller than those encountered in conventional aircraft vertical or directional gyroscopes.

In practice some types of gyroscope are designed to produce measurements about one axis only, others about two. These are known as single and two degree of freedom gyroscopes.

3.2 Floated Gyroscopes

In conventional non-inertial gyroscopes the rotating wheel is supported by carrying its bearings on gimbals which permit the axis to take up any desired position irrespective of the motion of the supporting base. However any frictional torques generated by the gimbal system itself result in drifts. This led to the concept embodied in the first inertial gyroscopes, which was to use floatation as a method of frictionless support.

A single degree of freedom floated gyroscope is shown diagrammatically in Figure 5. The rotating wheel together with its motor are contained in a cylindrical float which is in turn contained in a housing. The latter is filled with a liquid of high density such that at the design operating temperature the float assembly is neutrally buoyant. The device behaves as a rate gyroscope in that when it is rotated about its measuring axis the float rotates in the housing, still supported by the fluid. In conventional rate gyroscopes such rotation is opposed by a spring restraint, spring deflection being a function of angular rate. But in this case the restraint is viscous, which leads to an integrating action such that the angular displacement of the float is proportional to the rotation about the input axis. The former can thus be detected by a pick-off as an indication of the latter.

Such single degree freedom rate integrating gyroscopes are both accurate and suited to operations in rugged environments. The instrument includes a torquer capable of producing the rotation desired to orientate the instrument. This in turn implies orientating the platform, since the gimbal servos of the latter operate in response to the pick-off outputs.

In an aircraft inertial system the platform carries three such sensors as well as the associated accelerometers.

The alignment of the float within the housing is maintained by a suspension, a simple pivot arrangement in the case of Ferranti instruments.

The fact that the instrument has a single defined operating temperature has led to the general acceptance that the platform and gyros must be at operational temperature before navigation can commence. This in turn has led to the provision of rapid heating arrangements to bring the whole assembly up to temperature as part of the initial pre-flight alignment process. This has been one of the dominant factors fixing state of the art alignment times at between 10 and 15 minutes. However recent work by Ferranti Limited (Reference 1) has resulted in much shorter alignment times of the order of 2 minutes. Gyro-compassing is carried with the gyros at the temperature found at switch-on and rapid heating to operational temperature is applied at reduced speed during alignment and accelerated to full speed for navigation. The reference describes the complex considerations involved in this improvement.

An extremely large number of two degree of freedom floated gyroscopes is in service in systems manufactured by Litton (Reference 2). A platform mechanised with two such instruments can of course possess four measuring axes, one of which is redundant.

The first generation inertial systems used in commercial aircraft (Delco and Litton) all made use of floated gyroscopes. Similarly they have been the basis of the early generation military systems, many of which are in service. They have been used extensively in space applications and marine systems. The possible user objection has been that such instruments have to be manufactured and filled with fluid in extremely high quality clean rooms employing skilled personnel. They are thus not a maintenance proposition in a normal avionics workshop. However in practice this has proved less of a disadvantage than was sometimes envisaged because of the very long life and high reliability of these instruments, which results largely from their rugged construction and the care introduced in their manufacture.

Both ball bearings and gas bearings are used to support the wheel in floated gyroscopes, depending on the application and a number of detailed considerations beyond the scope of this paper.

3.3 Tuned Rotor Gyroscopes

For many years it has seemed attractive to develop a dry gyroscope and thus escape the constraints of the floatation principle. In practice the dry instruments which have now entered production and service on a large scale are of the tuned rotor configuration. The operating principles are described in detail in Reference 3.

In a typical dynamically tuned instrument (Figure 6) the drive motor and its bearings form the base of the unit. The shaft carries a structure resembling a universal joint. It supports an annular ring free to pivot about one axis perpendicular to the axis of rotation. The former then carries a second set of pivots at 90° to the first and supporting the rotor. Such gyroscopes, embodying universal or Hooks joints have been in use for many years in gyroscopic gun sights.

The property of this arrangement is that when the joint is spun at high speed by the shaft the angular momentum of the rotor causes it to resist changes in its attitude. If the case is displaced through a small angle this in turn causes the gimbal to oscillate. In fact there is an effective torsional restraint between rotor and non-rotating reference. This torsional restraint, produced dynamically, has a negative co-efficient of spring rate, and is a function of gimbal inertias and rotation speed.

In a practical design the ball bearings and half axles which would constitute a normal universal joint are replaced by torsional elements with a positive co-efficient of spring rate. By suitable selection of this spring rate and the gimbal inertias the dynamically induced spring rate cancels that of the physical torsional springs at the selected speed of rotation.

In effect the rotor is provided with a frictionless gimbal system imposing extremely low torques on the rotor, which consequently is less liable to drift.

By a suitable pick-off arrangement such a device can be used as a two degree of freedom gyroscope having limited freedom, and suitable for employment in a platform in a manner similar to that of floated gyroscopes. It can also be torqued by the provision of a suitable torquer. Both electromagnetic and electrostatic torquers have been used.

The most difficult design problems are posed by the hinge arrangement itself.

Among the advantages claimed for dry instruments are a smaller number of functional parts and the fact that the drive motor is outside the gimbal system rather than within it.

There is a considerable amount of published literature regarding the analysis of the dynamics of these instruments, which is complex. The interest is to isolate possible sources of error and adjust the design to minimise them. In particular shaft vibrations can be induced by the imperfections inherent in the bearings. There is a particular mechanism which results in errors induced by vibrations occurring at frequencies equal to twice spin frequency. Damping effects and operation of the gyro at a spin speed different from the tuned speed can also cause errors. There can be temperature effects on torquers and pick-offs.

An interesting variation of the tuned rotor principle is found in the Ferranti oscillo-gyro. This instrument comprises a single axis suspension carrying a rotating bar, the position of which can be detected by a pick-off. While the element operates in only one axis relative to the shaft a two axis output can of course be obtained on a time sharing basis, and the instrument thus behaves like a two degree of freedom gyroscope.

It has also been pointed out in the literature that multiple gimbal arrangements are feasible and a possible method of containing some of the error sources.

3.4 Unconventional Gyroscopes

3.4.1 Electrostatic Gyros

Reference 4 gives more details of the evolution of this technology. Both gimbaled and strap-down applications are under development.

The instrument consists typically of a spherical envelope enclosing a rotor, the whole being evacuated. The envelope includes electrodes and the rotor is suspended in the chamber by the forces of electrical attraction. There is a feedback system which controls the suspension to keep the rotor centred.

Once suspended the rotor is brought up to operating speed by a rotating magnetic field in a manner similar to an induction motor. Since the suspension is frictionless no additional spin power is required once the rotor is up to speed. Arrangements are made to damp any rotor nutation.

Optical means are used to determine the position of the rotor within the housing. In a gimbaled system any misalignment is removed by the gimbal servos. An important point is that the ESG rotors are not torqued. Their spin axes remain essentially fixed in inertial space and the system is therefore a space stable system.

When an ESG is used in a strapdown application an optical or other technique must be employed to determine the angular rotor position relative to the housing.

The advantages claimed of the ESG are mechanical simplicity and low random drift, the error sources being predictable and repeatable. Errors can arise from mass unbalance or magnetic and electric fields.

3.4.2 Ring Laser Gyroscopes

This instrument is based on a ring laser having two light beams rotating in opposite directions. If the device is rotated about an axis normal to the plane of the beams a frequency difference between them can be detected.

The development of this device has involved tackling a number of difficult problems. At low rates of turn frequency locking between the two beams tends to occur, giving the effect of a threshold below which rates cannot be measured. The technology of building tubes with the necessary life and reliability is itself difficult. But systems based on the ring laser now exist and it is showing promise for strap-down applications. However in the context of aircraft inertial navigation systems and their impact on air traffic control the laser is not yet sufficient.

3.4.3 Other Possibilities

Free rotor gyroscopes, using spherical gas bearings, have been used in aircraft systems. Vibrating elements such as tuning forks can be used to sense angular velocities. Gases or liquids can be rotated, when they are subject to the usual dynamic forces which can conceivably be detected. It has been suggested that the inertial properties of atomic nuclei could be used as the basis of a sensor.

While work on unconventional sensors is certain to continue there are no immediate indications of their impending use in systems which might be used in aircraft.

The gimbaled system, using either floated or tuned rotor gyroscopes, is likely to form the basis of aircraft navigation for many years. While strapdown systems are not yet developed to aircraft standards of accuracy their use in aircraft references for flight control and navigation purposes is possible in the future, particularly as they can be arranged to achieve a greater variety of redundancy. The gyros used in these systems are likely to be either floated or tuned rotor, with the necessary torquing provisions.

3.5 Accelerometers.

A typical inertial quality accelerometer is shown in Figure 7. Such a device is capable of measuring acceleration along a single sensitive axis, 2 or 3 being incorporated in the instrument cluster depending on whether a vertical channel is desired.

The device consists basically of a pendulum. Any force acting along the measuring axis produces a relative motion between the pendulum arm and the instrument case. This motion is sensed electrically and the resultant signal is amplified, rectified and fed to a restoring coil, situated in the field of permanent magnets, such that a force is produced which opposes and restores the pendulum to its null position. Thus current passing through the restoring coil is proportional to the acceleration force applied and the accelerometer function as a force-feedback servo.

The feedback loop must be stabilised and this can be accomplished by viscous damping within the instrument. Alternatively there are instruments available using gas damping or in which damping is achieved electronically.

The principle characteristics required of such an instrument are:-

A large dynamic range, typically $\pm$ 20 g.

Short to medium term stability of bias: of the order of 10^{-6} g.

Scale factor stability.

Freedom from cross-coupling effects.

Accelerometer development has concentrated on realising this class of accuracy within the minimum weight and volume. The principle problems are concerned with the detailed design of the hinge, restoring system and pick-off. Both inductive and capacitiative pick-off systems have been used. In the example shown the pick-off is energised normally from an oscillator in the range 15-20 K. Hz.

4. PERFORMANCE CONSIDERATIONS

4.1 Specification of Accuracy

The parameter most commonly used to specify the accuracy of an inertial navigation system is CEP (Circle of Equal Probability). This is defined as the radius of a circle centred at the aircraft's actual position within which 50% of the position measured by the INS will lie.

It was pointed out by Amacker and Mason (Reference 5) that CEP was originally adopted as the statistical parameter applied to determine the equal probability of hit or miss with artillery shells. It was carried forward with the development of missiles and their guidance systems.

CEP is still the most commonly used terminology and it is implied in such statements as "A One Nautical Mile Per Hour System". But its use involves limitations both to the designer of an inertial navigation system and to the operator of an overall air traffic control system within which the problem is to assess the pathkeeping accuracies to be expected from different aircraft equipped with different sorts of systems. The simplest way to assess the CEP for an inertial navigation system is to check the aircraft's position carefully by external means at the end of each of a series of flights. The measured error is divided by the total flight time to give the rate of error propagation per hour. If a sample of flights is then examined the CEP in nautical miles per hour is the figure within which one half of the measured flights lie.

Among the shortcomings of this very simple minded approach are:-

It does not reveal many aspects of the system performance for the benefit of the designer or those requiring to assess the relative performances of different systems. This requires a much more comprehensive treatment allowing the results to be displayed in a manner which enables them to be related to the error sources.

For the pursoses of assessing navigation capability CEP is not an exacting enough requirement. A more probable requirement is to know the 95% error which can be anticipated bearing in mind not only the performance of the system but the possibility of flight technical error etc.

4.2 Assessment of Inertial Navigation Accuracy

There is a considerable literature describing methods of assessing the accuracy of inertial navigation systems. Analytical and modelling techniques can be used to predict accuracy or specific flight trials can be analysed.

An important point, particularly in long flights, is that the manner in which the errors propagate is dependent on the flight path. For example the latitude at which the flight takes place and the direction of flight both affect the error propagation mechanism. In addition, while the errors propagate with time, they do not do so in a linear manner. Cyclic errors are also present.

Reference 6 describes work carried out in the U.K. by the Royal Aircraft Establishment, and indicates the complexity of the considerations involved. It describes both practical flight test techniques and the production of a mathematical model of an inertial navigation system. The model concerned contained the following error sources:-

Gyro drifts, fixed, variable with time and as functions of flight path accelerations.

Fixed or variable accelerometer biases.

Gyro torquing scale factor errors.

Accelerometer scale factor errors.

Integration errors and biases.

The addition of errors resulting from the harmonisation of the inertial components would have further complicated the process, but in practice these are very small, of the order of 20 arc seconds 1 sigma.

A system of equations is derived containing all the commonly known periodic error patterns for inertial systems, Schuler (84 Minutes) Diurnal (24 hour) and Foucault (36 hour). These are not all significant on any given case but all exist.

The aim of any flight test programme to establish the performance of an INS is to discriminate between the results of various error sources. The authors use the mathematical model to produce error curves for position and velocity which would result from given error sources and the given flight path. These simulated error curves can then be compared with the observed errors, further processing of the data involving weighting the error sources and combining them until a fit is obtained.

At minimum the quick visual assessment of the results of an inertial trial normally demands that they be presented in terms of N-S and E-W position and velocity errors, which enables the observer to distinguish between the two navigation channels. An example is given in Figure 8.

Assessing the performance of an inertial navigation system by plotting the inertial position against information derived from external fixes is obviously the basis of any flight test technique, or technique involving the analysis of navigation data from more than one aid. Among the problems encountered in doing this are the following:-

(a) The errors in the external aids are critical. In practice the accuracy of inertial navigation is such that it is often difficult to secure cross checking information of the required accuracy. The most accurate aids, such as Decca and ground tracking radars, have limited coverage which is a disadvantage in long flights by fast aircraft. Reference 6 describes the combination of information from a number of aids to provide a reference.

(b) The frequency with which the position data can be acquired is also critical. Because the inertial errors propagate in an oscillatory manner the characteristics of the system under test may be concealed by too low a data rate, and depending on the position of an observation relative to the Schuler oscillation the use of small amounts of data can give misleading conclusions. The method used to derive radio position errors from the observations is also significant. Reference 6 quotes a procedure for inferring the shape of an apparent distribution from the ratio of the geometric mean to the route mean square of the sample. Other methods which were in use when the first commercial inertial systems were evaluated are described in Reference 5.

Finally it is of interest to contrast the treatment of the errors in an inertial navigation system with the treatment of navigational errors in general, specifically those resulting from classical navigation in which a human operator combines information from a number of aids using judgement. This subject was treated comprehensively for the first time in a paper published by E.W. Anderson in 1952 (Reference 7). A further paper by the same author in 1971 (Reference 8) gave the results of an investigation of the distributions that arise in practice when navigation errors are examined.

These papers throw an interesting light on a controversy which arose when it was first suggested that inertial navigation might be used as the sole means of navigation in long oceanic crossings. It was suggested that such a system would be acceptable if the error distribution was convex. But it was suggested that in a system having errors increasing with time the trend might be towards an exponential or concave distribution. The question was whether the tails or skirts of the pattern would be significant in practice using IN.

With hindsight it now seems possible to draw the following conclusions:-

Inertial performance in practice has proved extremely consistent and the tails of the distribution have not been an undue embarrassment. This probably reflects the consistency with which such systems can be manufactured and the results of the manufacturer's efforts to include pre-flight alignment procedures and built-in test arrangements.

While some blunders have occurred the availability within these systems of a digital computer and flexible man-machine interface have served to minimise them, and so this possible cause of tails has not become significant.

The general accuracy of the system has improved by a factor of 2 or 3 as the technology has improved.

As a result of all the above the earlier pressures to bound the errors of inertial navigation by means of a long range radio aid have not been successful, and operators have relied on similar rather than dissimilar redundancy.

When it was first suggested that inertial navigation systems in long range aircraft should be reinforced by long range radio aids most thinking was in terms of the mixing and cross-comparison being carried out by the crew. Subsequently various methods of treating the errors within a digital computer acting as a multi-sensor navigation system was suggested. The computing capability required for such multi-sensor systems has improved progressively over the years, particularly by the introduction of Kalman filtering, which makes possible an extremely sophisticated treatment of the error model including all aids used. New aids have become available, particularly OMEGA.

But the results achieved in practice using multiple inertial navigation systems have been much more spectacular than anticipated and the indications are that they can contain current or immediately proposed separation standards. It has not proved cost effective or desirable to introduce mixed systems on any large scale. The pure inertial navigator has scored on the grounds of intrinsic accuracy, ease of operation, fault diagnosis, initial cost and cost of ownership.

The only exception which has appeared to produce dividends in practice has been the automatic updating of inertial information by DME ranges for area navigation purposes in airways structures having the necessary deployment of DME. This at present applies only to the U.S.A. It has also been proposed that inertial and ILS information should be combined for approach purposes to contain beam bends in the latter.

5. Electronics and Computation

An inertial navigator requires the mechanisation of a series of equations for alignment and navigation (Reference 9), including the generation of correction terms. In practice these must take account of earth's radius to the extent of allowing for the form of the earth and aircraft altitude if maximum accuracy is desired. The computation rates required are not particularly exacting. For this reason it was found possible to mechanise early inertial navigators using analogue computers, the main problem being to secure the necessary standards of accuracy.

However the development of inertial navigation systems has coincided with rapid progress in microelectronics. This in turn has made it possible to produce small inexpensive digital processors.

Some of the earliest digital inertial navigation systems were based on the DDA (Digital Differential Analyser). Subsequent equipments employed general purpose processors including core store memories. But the development of digital components has been extremely rapid and the latest machines are universally based on solid state memories. In addition medium or large scale integration of electronic functions on a single silicon chip has reduced the number of packages in the typical processor. As a result the memory capacity and computing power available in most inertial navigation systems has reached a level at which spare capacity is available.

The principle tasks undertaken by the processor are:-

Handling the inputs from the accelerometers after these have been converted into digital form, and possibly integrated.

Generating commands to torque the inertial gyros through a desired angle.

Handling the basic equations to generate correction terms.

Co-ordinate transformations, for example in the case of the free azimuth system.

System self-test, reasonableness checks etc.

Housekeeping duties associated with the control and display unit. Great circle calculations for guidance between way-points.

The most critical design area is concerned with precision electronics forming the interface between the processor and the sensors in the platform. Other electronic functions include the provision of a number of special internal power supplies and platform environmental control.

The presence of an integrating function within an INS means that time is a significant parameter, and that interruptions of the integration process cannot be tolerated. It is therefore standard practice to input redundant power supplies. For example the Arinc 561 systems operate normally on 400 Hz AC supplies. A special external battery is provided and maintained in a state of constant charge. Should the primary supply be interrupted the system reverts automatically to the battery. Alternatively an INS can be powered normally from a high integrity DC supply. The largest power demand is concerned with rapid heating in the event of a start up from cold conditions. This can be drawn from an aircraft AC supply of lower integrity.

Within the system there are generated critical DC supplies for the gyro spin motors , since the instrumental accuracy depends on the precision with which the design rotation speed is achieved and maintained. A special supply driven from a precision oscillator, which may be associated with the computer, is therefore used.

6. Conclusion

The essence of inertial navigation technology is the mechanisation of some well defined tasks in a highly precise manner. There is no escape from this precision if a given standard of performance is to be achieved.

For this reason the development of the basic inertial instruments, particularly gyroscopes, has been an evolutionary process. The cycle from the first attempt to construct an instrument using given principles to the achievement of the desired accuracy on a production basis lasts many years, and it is necessary to produce relatively large numbers of instruments before a given accuracy standard can be established with confidence. The exacting accuracy requirements involved themselves create testing problems. The performance of an instrument has to be assessed on a statistical basis as has the performance of a final system.

The cost of inertial navigation systems has fallen progressively over the last decade in terms of real value. But attempts to produce a dramatic break through in cost have not succeeded. It has been rare to find a cost improvement between two successive generations much in excess of 20%. Production experience, improving the yield of high grade instruments and streamlining production and test techniques, has been as significant a factor as basic design.

The development of the associated electronic technologies has been forced by applications in other fields. The designers of inertial navigation systems have taken progressive advantage of new generations of electronic components producing larger numbers of functions per package with higher reliability. These improvements have been most rapid and significant in the area of digital processing.

For the operator the most significant improvements, apart from accuracy, have concerned operational flexibility and ease of maintenance. The former include improved ability to align pre-flight in a self-contained gyro-compassing mode. Alignment times have traditionally been of the order of 10 to 15 minutes, but systems with a rapid alignment mode are now possible. The latest Ferranti systems include such a facility.

Improvements in maintainability have been most marked in two areas: avoiding special setting up or calibration procedures as much as possible when main system components are changed and providing comprehensive self-test facilities. Modern systems include a number of levels of built in test, using both special purpose hardware and the facilities of the digital processor. With the aid of the latter it is possible to develop codes indicative of system performance or malfunctions in specific areas, and to display these to flight crews or maintenance personnel.

In the present context of manned military or commercial aircraft operating within a total ATC system the typical inertial navigation installation consists of one or more gimbaled platform systems using floated or non-floated gyroscopes. It is possible that strapped down systems or systems using novel sensors will ultimately enter this field, but there are no immediate indications of their achieving the standards of performance and cost required to bring this about.

It is sometimes suggested that a cost effective solution to navigation could be found in hybrid systems using software filtering to combine a low cost lower accuracy inertial sensor with a number of available radio aids. This has so far not happened, partly because of the high accuracy of pure inertial navigation and also because the hybrid systems so far proposed have not been able to equal the accuracy and versatility of an inertial navigator at a competitive price, given the need to operate over a wide spectrum of routes. Present day system accuracies have proved more than adequate for trans-oceanic sectors on the basis of pure IN, as well as providing a significant capability for operations off airways in continental air space. Only the most accurate ground based radio aids seem capable of producing an improvement justifying the costs of hybridisation. For example significant benefits are achievable with DME updating in areas in which the coverage is adequate.

It has also been suggested that an inertial component could improve flight guidance in the approach mode, particularly using relatively low quality ILS facilities. The extent to which this will be exploited in the future depends on the deployment of high quality ILS, and later MLS approach systems.

The author is indebted to the management of Ferranti Limited for permission to publish this paper.

7. REFERENCES

1. K.R. Brown and D. Anderson, Ferranti Limited: Gyro Characteristics for Rapid Gyro-Compassing; AGARD Conference Proceedings No. 116.

2. C.S. Bridge, Litton Industries: Developments in Inertial Navigation; Istituto Nazionale della Navigazione, Rome, Italy; February 1968.

3. R.J.G. Craig, Teledyne Systems Company: Theory of Operation of an Elastically Supported Tuned Gyroscope; IEEE Transactions on Aerospace and Electronic Systems; Vol. AES-8 No. 3; May 1972.

4. R.R. Warzynski and R.L. Ringo, U.S. Air Force Avionics Laboratory: The Evolution of E.S.G. Technology; AGARD Conference Proceedings No. 116; February 1973.

5. J.Z. Amacker and F.J. Mason, USAF: Specification of Inertial Navigation Accuracies; Navigation, Vol. 12 No. 3; Autumn 1965.

6. R.F. Stokes and S.G. Smith, R.A.E., Farnborough: Aircraft Inertial System Testing and Evaluation in the United Kingdom: AGARD Lecture Series No. 60; October 1973.

7. E.W. Anderson: The Treatment of Navigational Errors; Journal of the Institute of Navigation, Vol. V No. 2; April 1952.

8. E.W. Anderson and D.M. Ellis: Error Distributions in Navigation; Journal of the Institute of Navigation Vol. 24 No. 4; October 1971.

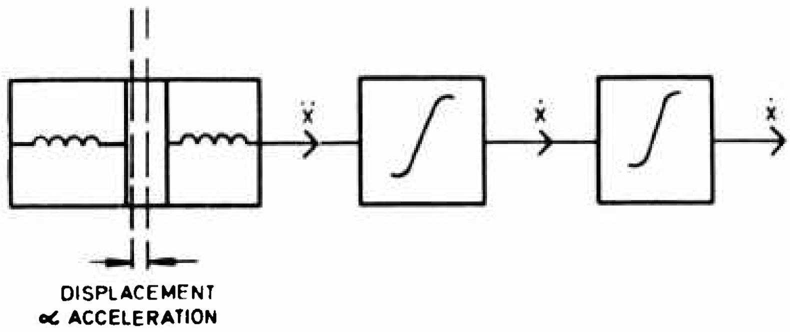


FIGURE 1
ACTION OF BASIC ACCELEROMETER

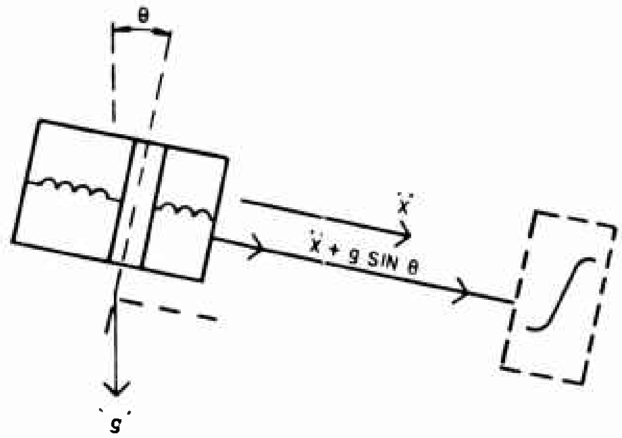


FIGURE 2
EFFECT OF TILT

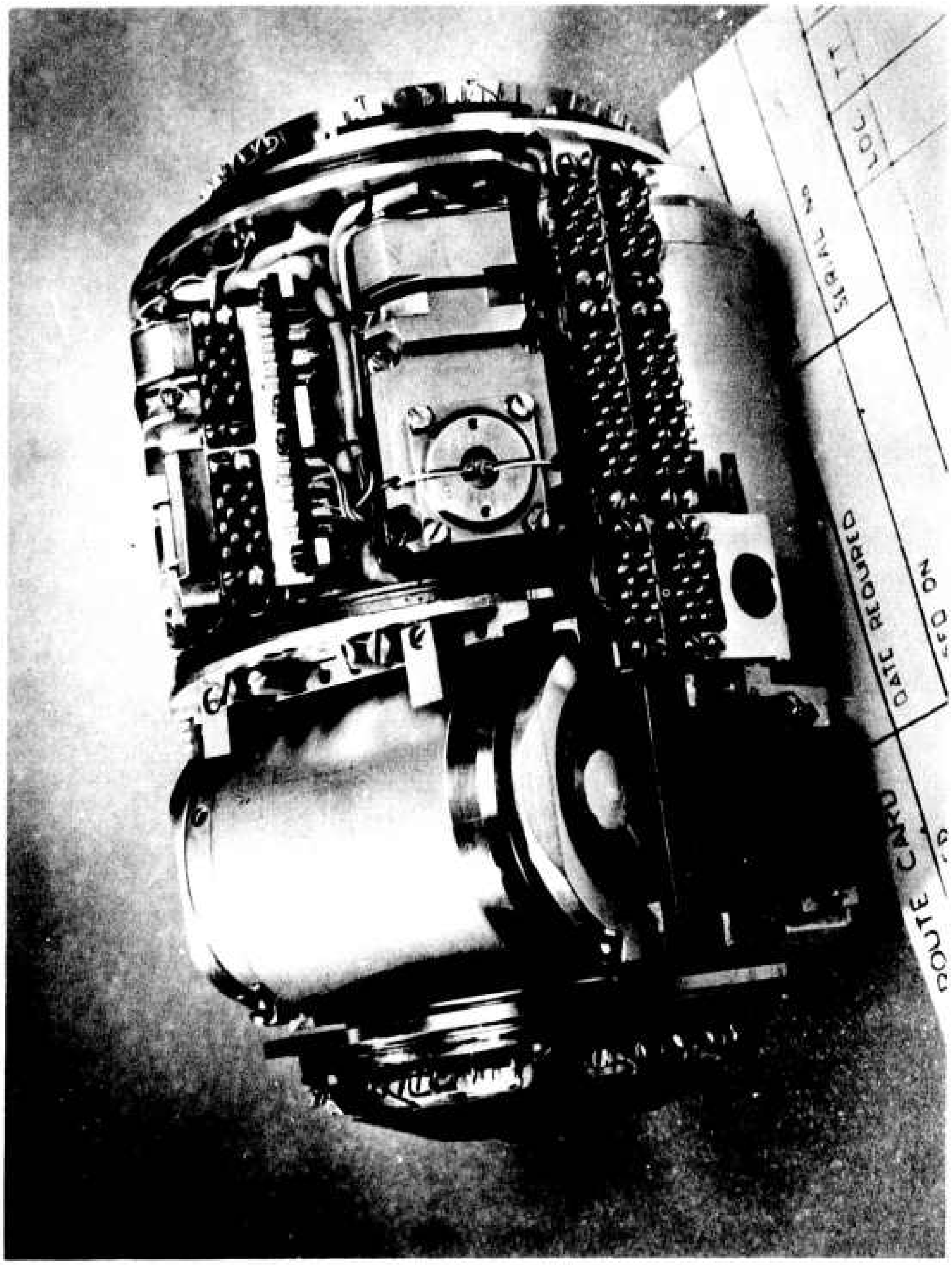


Figure 3
Instrument Cluster

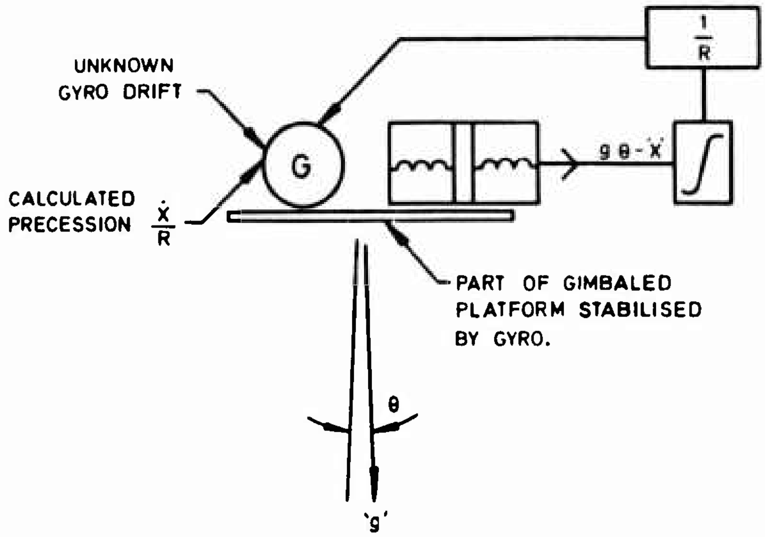
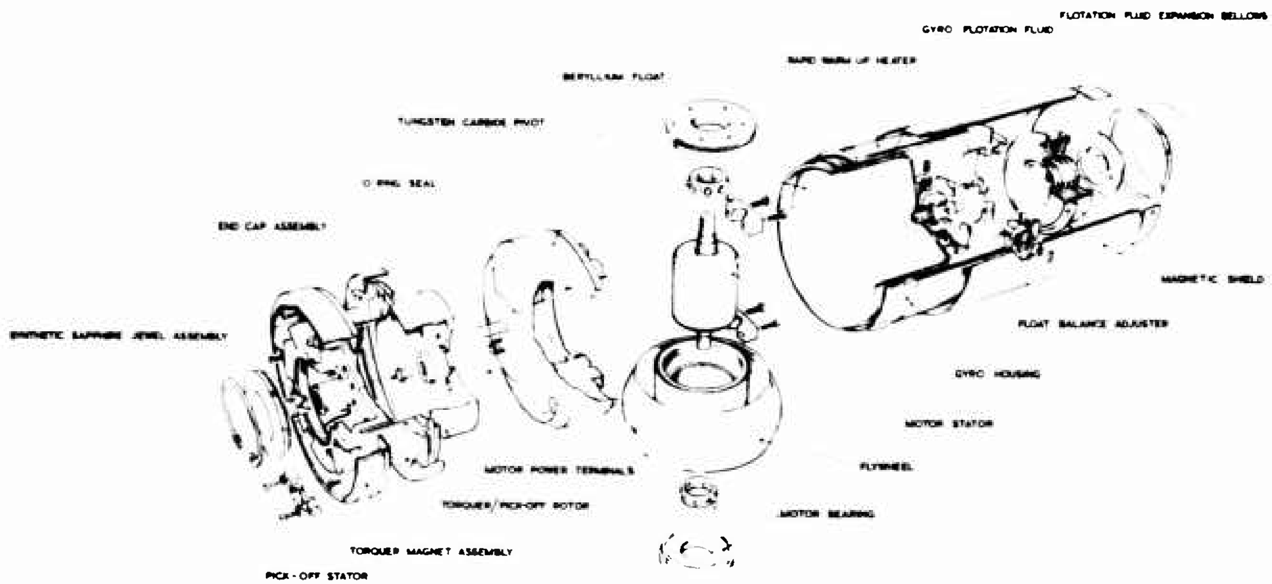


FIGURE 4
SCHULER ERECTION LOOP

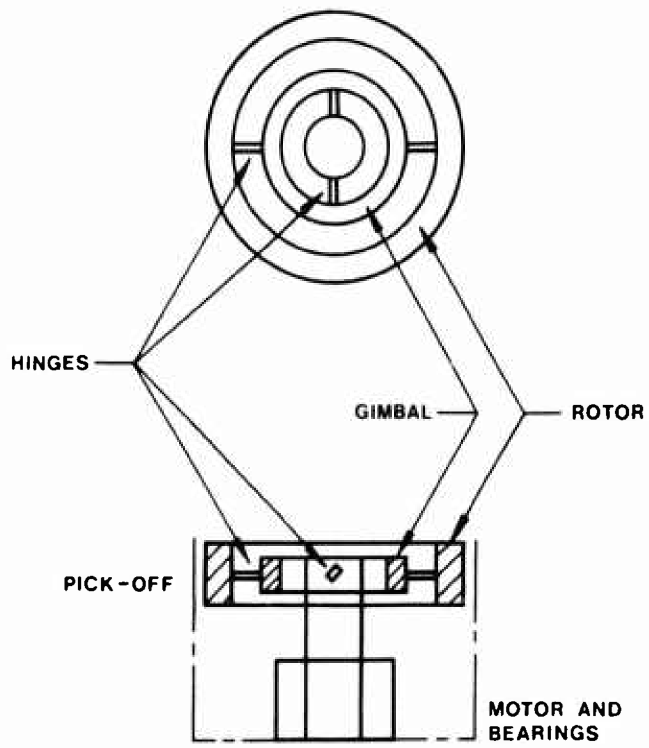


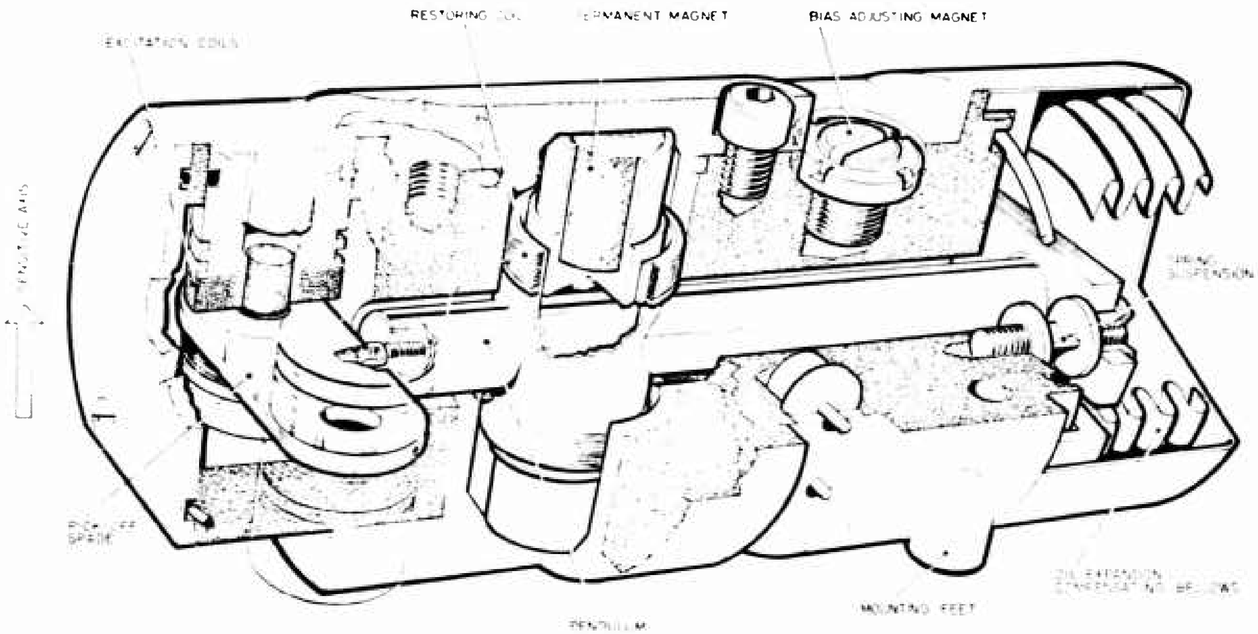
SINGLE AXIS RATE INTEGRATING GYROSCOPE

Figure 5

Single Degree of Freedom Floated Gyroscope

FIGURE 6 DYNAMICALLY TUNED GYROSCOPE

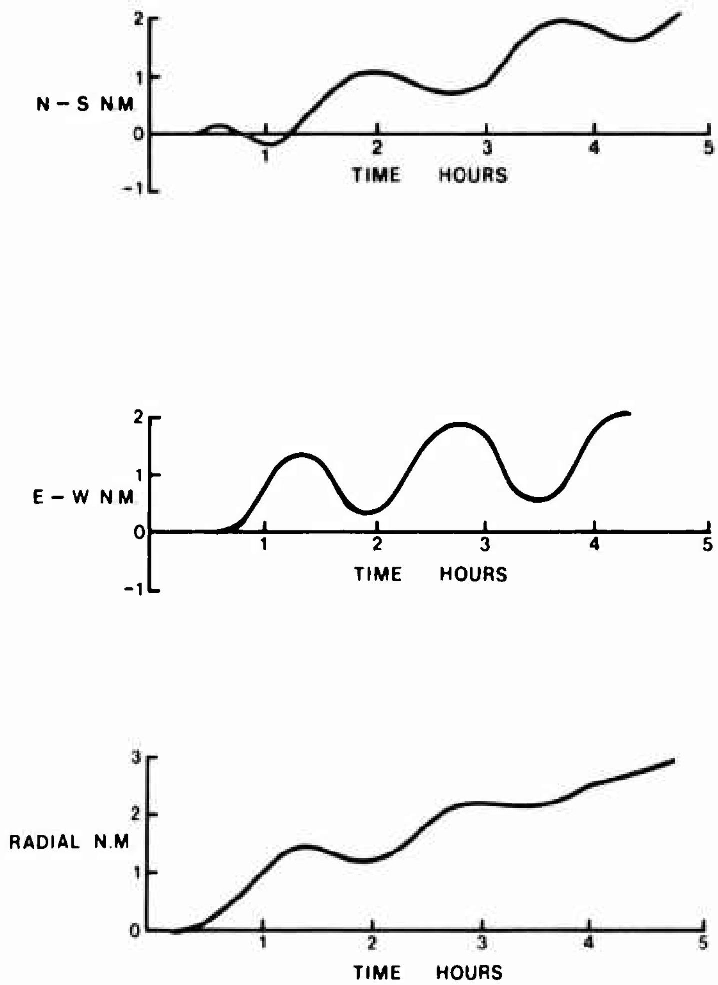




Single Axis Accelerometer

Figure 7

FIGURE 8 PLOT OF TYPICAL I N S ERROR WITH TIME



Pages 489 and 490 are Blank.

LANDING GUIDANCE SYSTEMS

by

Frank B.Brady, Aviation Consultant
500 New Hampshire Avenue, N.W., Washington D.C.

TABLE OF CONTENTS

LANDING GUIDANCE SYSTEMS

by

Frank B.Brady, Aviation Consultant
600 New Hampshire Avenue, N.W., Washington D.C.

SUMMARY

This publication reviews the evolution of aircraft landing guidance technology, from rudimentary nondirectional beacons and markers in the 1920's through the development and installation of the current standard instrument landing system (ILS) to the microwave landing systems proposed for future world standardization. The major milestones in landing guidance system development are depicted in Figure 0.1.

0.1 EARLY HISTORY

The need for approach and landing guidance during periods of restricted visibility was recognized soon after the end of World War I, and solutions were pursued by both civilian and military agencies. The U.S. Department of Commerce, Bureau of Standards experimented with a spark transmitter on the ground and a direction finder in the aircraft. In 1929, under Bureau of Standards responsibility and Guggenheim Foundation sponsorship, Lt. James Doolittle made a historic blind landing with the system diagrammed in Figure 0.1(a). A nondirectional beacon on the ground actuated a vibrating reed visual indicator in the aircraft to guide him over the beacon, at which point he turned outbound, descended to final approach altitude and returned to the beacon station. Upon receipt of a low-frequency vertical marker indication of being directly over the station, he began his final descent along a known compass course from the station. Another marker beacon (not shown) indicated the beginning of the runway. It is evident that Doolittle's approach foreshadowed the approach procedures for several subsequent systems.

In 1934, the U.S. Army developed the A-1, a radio-compass locator-marker system, shown in Figure 0.1(b). The pilot tuned his left-right radio compass to an outer station located along the extension of the runway centerline and maintained altitude until a marker beacon lighted a lamp in the cockpit, indicating he was directly over the station. He then tuned to the inner station, also along the extended runway centerline, and descended to a specified approach altitude. The inner marker beacon signalled his proximity to the landing field.

It was generally recognized, however, that some form of continuous glidepath guidance was needed. Pioneering work was done along these lines in Germany by Dr. Ernst Kramer, who developed a Lorenz-system constant-intensity glide path that the aircraft followed via aural (dot-dash) and visual (metered needle) instruments indicating fly-left, fly-right, or on-course. The glide slope was merely the underside of the beam envelope produced by the localizer. See figure 0.1 (c). Versions of the Lorenz system were tested by the U.S. Bureau of Standards in 1931 and the Bureau of Air Commerce in 1937, and were installed at airports in Germany, the U.K., and the U. S. But success was elusive, mostly because of multipath interference about which little was known.

The Lorenz constant-intensity glide path was abandoned in favor of equisignal straightline glide-slope and localizer courses, each formed by the intersection of two beams. The U.S. Bureau of Air Commerce, Civil Aviation Administration (CAA) supervised research and development on an equisignal system used in conjunction with an independent, redundant system of outer and inner locator-station marker beacons, shown in Figure 0.1(d). This was the forerunner of the ILS.

World War II intensified development of all-weather landing guidance technology on both sides, including automatic landings, militarized ILS, ground controlled approach (GCA) and use of microwaves. The U.S. Army version of ILS was the transportable SCS-51, illustrated in Figure 0.1(e). This became the standard for larger aircraft at major airbases.

The precision-approach radar GCA, the other standard, was the primary landing system in most military applications. The GCA, as shown in Figure 0.1(f), relied on radar fixes of the incoming aircraft to enable a ground controller to guide the pilot to a landing via a radio voice link. There were both fixed and mobile versions, and GCA could be employed almost anywhere and with any type of aircraft, at remote tactical airstrips and aboard aircraft carriers. It was especially valuable in all-weather landing for minimally equipped aircraft or less experienced pilots, as well as a backup for ILS.

The AN/SPN series of aircraft-carrier GCA landing systems, depicted in Figure 0.1(g), were developed in which optical systems were added to compensate for deck motion during the final approach, ground controllers gave way to computers and data-link systems capable of automatic landing guidance, and microwave scanning was introduced to increase range, capacity and accuracy.

0.2 INSTRUMENT LANDING SYSTEM

Soon after the end of World War II, the ILS was adopted by the newly formed International Commercial Aviation Organization (ICAO) as the standard approach and landing guidance system for international civil use. It is still the standard system and is installed at over 500 major airports throughout the world. The fully equipped ILS, shown in Figure 0.1(h), consists of a localizer ground transmitter, a glide-slope ground transmitter, marker beacons, plus an independent distance measuring equipment (DME - not shown in the figure), all of which work in conjunction with counterpart airborne equipment. The ILS uses VHF/UHF frequencies on 20 channels, one channel for each installation.

The localizer antenna array generates a straight and narrow azimuth course via intersection of two equisignal beams modulated to produce fly-left, fly-right, or on-course indications in the cockpit. The glide-slope antennas provide a glide path in essentially the same manner; however, the equisignal technique was replaced with a null-reference technique in 1953 in order to reduce glide-path angle variations due to snow buildup or tidal action in the ground plane, since the glide-slope signal depends on ground reflection for the formation of its beams.

For the lowest performance installations (Category I) the outer marker, located 4 to 7 miles from the runway, provides an independent check on the specified approach altitude versus position, and signals the pilot to start descent on the glide path. The middle marker, located less than a mile from the runway threshold, signals decision height (of about 200 feet in altitude); the pilot aborts the landing if he has not yet visually acquired the runway. An inner marker may be installed at higher performance facilities (CAT II or III) to signal a lower decision height of about 50 feet.

The glide-slope antenna is located adjacent to the runway opposite the glide-path intercept point (GPIP) in order to direct the aircraft to the touchdown zone.

Because the heavier and faster aircraft of today require a flared approach in the last seconds of flight to enable a smooth touchdown, a DME may be installed at CAT III facilities. DME, essentially an L-band transponder that replies to aircraft interrogation, primarily provides an independent and continual measure of range to the end of the runway, but also can be used in conjunction with a radio altimeter to lower the decision height to only a few feet for flare guidance.

In addition to ILS, airports over the years have been provided with a standard system of approach, runway edge, and centerline lights, plus painted markings, all in order to enhance airport visibility necessary for the final phase of landing. Attempts also have been made at fog dispersal.

0.3 NEED FOR A NEW SYSTEM

Despite over 30 years of successful use plus continuing research and development for improvements, the ILS has several inherent limitations which have precluded its universal use and restrict its future life.

One problem is siting. In order to install an ILS, the ground plane beyond the runway often must be extended, at great expense if possible at all. And the glide-slope and localizer signals are both adversely affected by reflecting objects such as hangars and intervening aircraft. Also, snow and tidal reflections can still appreciably affect the glide-path angle at some installations. Another problem is that the limited channels available in the VHF/UHF bands of the frequency spectrum restrict the number of interference-free installations in a geographic region. Also, ILS is limited to a single course, usually over an extension of the runway centerline, which prevents routing flights in curved or segmented approaches for aircraft separation, obstruction clearance, and noise abatement. In addition, military requirements for small or portable tactical landing systems obviate the large ILS antenna apertures necessary at VHF/UHF frequencies, which incidentally provide less signal resolution and hence less accurancy.

0.4 MICROWAVE LANDING SYSTEM

The foregoing limitations of ILS have given impetus to development of diverse landing systems, both civil and military, and mostly employing the microwave frequencies. In order to continue and reinforce the universal approach to a new landing system and to prevent adoption of a profusion of incompatible systems by different countries and organizations, an international body, the Radio Technical Commission for Aeronautics (RTCA) produced recommendations in the form of performance specifications in 1970 for a single microwave landing system (MLS) acceptable to all aviation interests. ICAO subsequently adopted essentially all of the recommendations and instituted a competitive development program. Australia, the Federal Republic of Germany (F.R.G.), France, the United Kingdom (U.K.) and the U.S. are submitting candidate designs to ICAO in 1975 for selection as the international standard, eventually to replace ILS plus the various military systems in existence.

The advantages of MLS are:

- Fewer adverse siting effects, including less ground-plane preparation
- Flexible approach paths
- Less performance degradation due to multipath effects
- Greater installation density: up to 200 channels at C-band versus a maximum of 40 for ILS.
- Possible simultaneous use of parallel runways due to greater resolution accuracy and greater area coverage
- Improved flare guidance
- Portability and military suitability due to smaller antenna size required for MLS
- Potential lower costs over the long term

All basic recommendations for MLS include an azimuth (AZ) transmitter-antenna and an elevation (EL) transmitter-antenna (see Figure 0.1(i)) as the minimum service for CAT I facilities. The AZ antenna array (equivalent in function to the ILS localizer) usually would be positioned beyond the stop end of the runway, and would cover a specified volume with signals that any equipped aircraft entering the volume could convert to the horizontal angle of its approach in relation to the runway centerline. The EL antenna (equivalent in function to the ILS glide slope) may be collocated with the AZ, but usually would be positioned adjacent to the runway opposite the GPIP, and similarly would cover approximately the same volume, only with vertical angle signal. Runway identification data, etc. would also be transmitted.

Another ICAO recommendation for MLS is that it should be modular in construction so that the different levels of service can be tailored to the various types of airports and aircraft; thus any aircraft can receive service from any airport commensurate with the least sophisticated equipment of either, whether airborne unit or ground station.

Additional services available at higher performance facilities include DME, greater coverage volumes, a flare antenna (often designated as EL-2), and back azimuth (BAZ) and back elevation (BEL) for missed approach or takeoff guidance.

There are many design considerations in developing the MLS. Among them are selection of:

- Carrier frequency compatible with available portions of the electromagnetic spectrum, propagation effects, and hardware requirements
- Signal format, including spatial, temporal, and also spectral features
- Angle data transmission, whether air-derived (as proposed by RTCA) or ground-derived (as proposed by France and the F.R.G.)
- Signal propagation method, whether mechanical or electronic
- Signal radiation technique, whether fixed fan beam, scanning beam, Doppler scan, or other
- Signal data coding, whether via frequency modulation in a frequency reference system (FRS) or via pulse modulation in a time reference system (TRS)
- Coverage region
- Guidance region
- Size and weight
- Primary power
- Radiated power, both peak and average
- Dwell time
- Duty cycle tradeoff
- Data rate tradeoff
- Data link capacity for auxiliary data
- Channeling tradeoff
- Reliability tradeoff
- Split-site versus collocation
- Compatibility with ILS and collocation with installed ILS
- Focusing
- Beam shapes and orientation
- Polarization, whether vertical or horizontal
- Coordinate system, whether conical, planar, or circular
- Flare data, whether DME with EL-2 or with radio altimeter
- Flare measurement, whether C-band or Ku-band El-2
- Range measurement, whether L-band or C-band DME.

Many of the above considerations mutually interact, and all interact with cost.

The Australian MLS candidate, called INTERSCAN, is a scanning beam TRS in which the measured time between a to-and-fro scan is proportional to the position angle of the aircraft. A to-fro scan is alternated with broad-beam transmission of auxiliary data. The various other azimuth and elevation functions, including BAZ, BEL and EL-2, are frequency-division multiplexed. The EL-2 antenna faces the runway from the side.

The U.K. MLS candidate employs a commutated linear array of antenna elements to generate Doppler signals in conical beams. The signal format is frequency-division multiplexed for the main signals (AZ,EL) but time-division multiplexed for the other signals. The array is scanned to-and-fro and up-and-down. An array thinning technique is employed to provide standby operation for purposes of integrity.

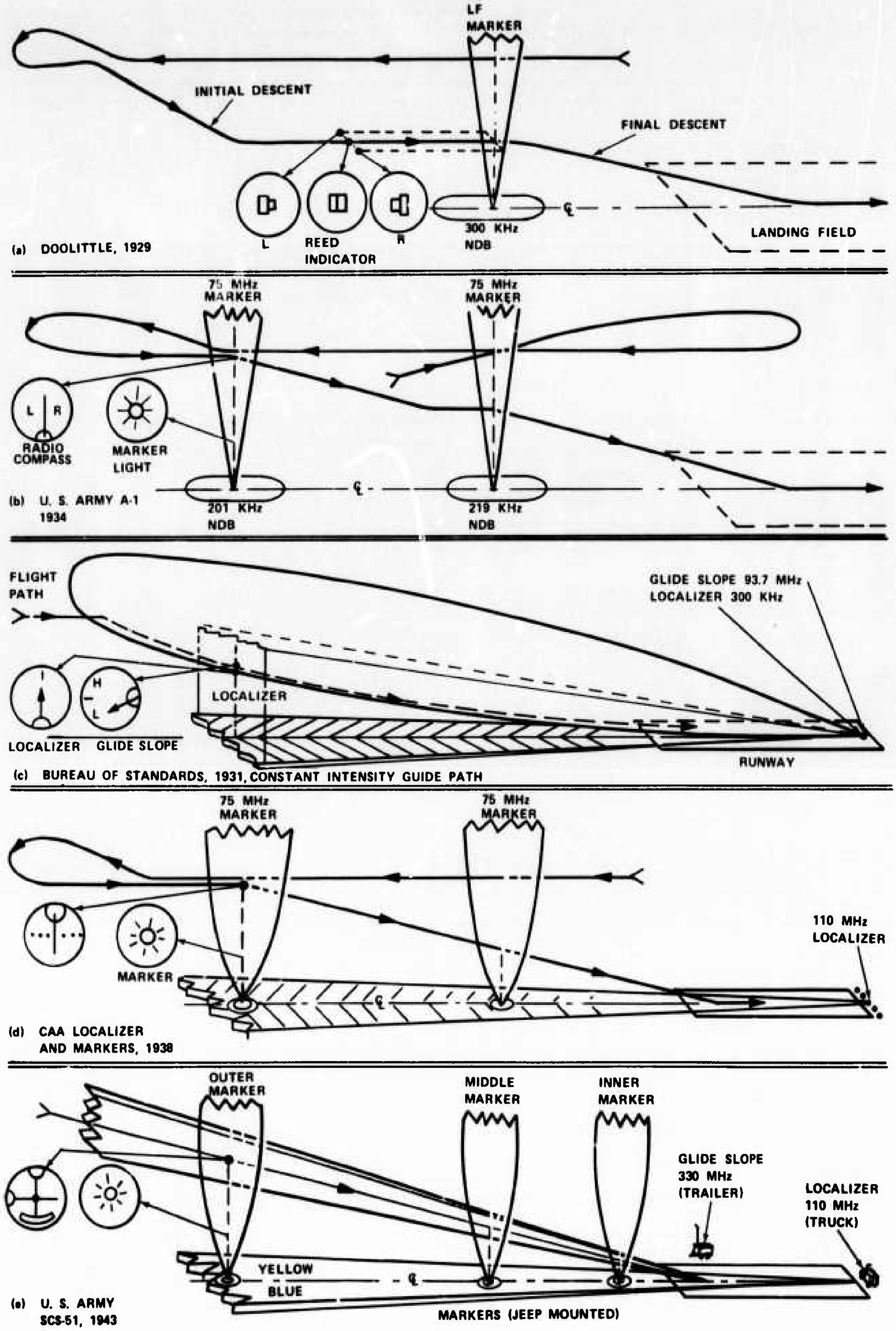


Figure 0.1 Evolution of Landing Guidance Systems (Early Types)

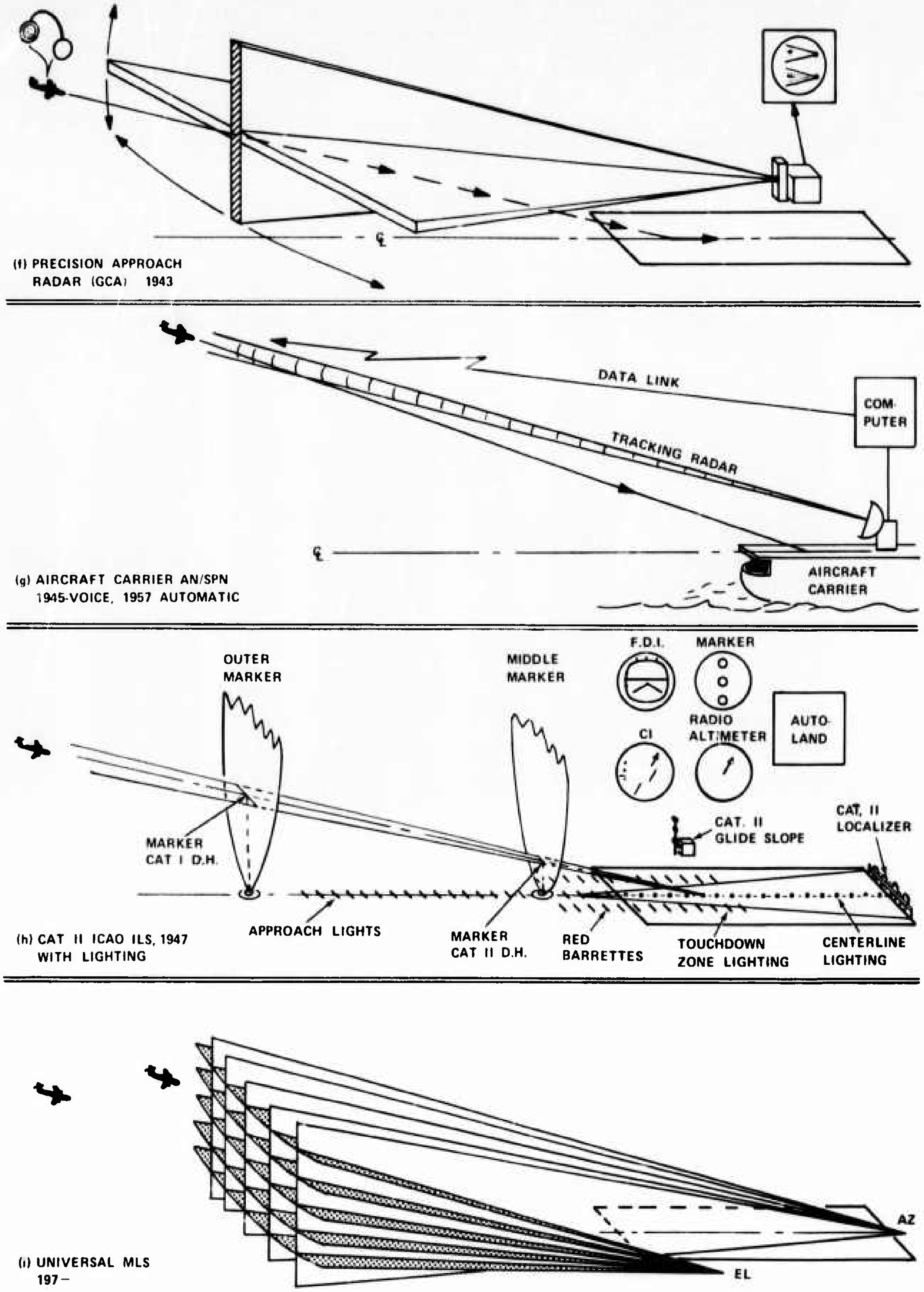


Figure 0.1 (Continued) Evolution of Landing Guidance Systems (Current Types)

The U.S. MLS program is the most elaborate, with two potential candidate systems in contention. One is a TRS scanning beam system and the other is a Doppler scan system. Both candidates operate at C-band except that flare guidance may operate at Ku-band. Each system is available in modular configurations, including versions for each performance category and coverage volume plus military tactical and aircraft carrier versions.

The MLS candidate from France is part of an integrated air traffic control (ATC) system. The MLS operates somewhat similar to a GCA combined with the current DME in that the aircraft interrogates the desired ground station, via time-ordered access, down linking aircraft identification. The ground station uses radar techniques to measure the aircraft position, and uplinks this data in continual replies. The time delay of retransmission supplies range data. The information is updated at a sufficient rate to provide virtually continuous steering data.

The F.R.G. MLS candidate, called the DLS, is similar to the French system, except that in DLS the DME is the main navigational aid, as part of an ATC-TACAN system, and the ground station is only a transponder, interrogated via random access. The position angles of the aircraft are determined by measurement of the interrogation waveform, and the angle data is retransmitted to the aircraft in a pulse position code. The AZ is a circular ground array. EL-1 and EL-2 are collocated, with EL-2 operating similarly to the Australian flare.

SECTION 1

EVOLUTION OF LANDING GUIDANCE TECHNOLOGY

1.1 EARLY HISTORY

The need for some form of landing guidance during periods of restricted visibility was recognized soon after the end of World War I and technical papers began to appear [1]. In 1919, the U.S. Department of Commerce, Bureau of Standards described experiments with a landing guidance system using a 300-kHz spark transmitter on the ground and a direction finder in the aircraft [2] [3]. To put this in perspective, these experiments were made 5 years after Lawrence Sperry made his dramatic Paris demonstration of a fully automatic stabilized Curtiss flying boat. Flying at low level, he stood in the cockpit with his hands above his head while his mechanic walked on the lower wing of the biplane [4].

A number of articles appeared in the early 1920's describing guidance system experiments on both sides of the Atlantic and from both civil and military sources. A paper was published in London in 1923 entitled "Automatic Landing of Aircraft" indicating the early military interest in the subject. Further such interest is evidenced in the files on fog research at the Instrument Landing Equipment Laboratory, Wright Field, dating back to 1923.

1.2 U.S. BUREAU OF STANDARDS AND GUGGENHEIM FOUNDATION EFFORTS

By 1928 U.S. commercial aviation had developed to a point where there was widespread recognition of the need for all-weather flight [5]. Far better instruments were available in the aircraft, and a system of aural radio ranges was beginning to supplement the lighted airways. This progress only served to emphasize the need for better landing guidance. With the U.S. Bureau of Standards as the center of responsibility, facilities were developed that led to Lt. James H. Doolittle's historic blind landing at Mitchel Field, New York, on September 24, 1929, under the sponsorship of the Guggenheim Foundation. Doolittle's feat was accomplished using a low-frequency range station on the ground that actuated a visual indicator of the vibrating reed type in the aircraft. A low-frequency marker beacon was also used to indicate the end of the runway.

The results of these tests, while successful, clearly indicated the need for some type of vertical beam guidance, and work was started on what was known as a constant-intensity glide path. The aircraft, equipped with a 100-MHz receiver, attempted to follow a path of constant field intensity formed by the lower envelope of a beam transmitted from a directional antenna array on the ground. Because of the attenuation in signal strength with distance, this signal produced a path that was steep as the aircraft started its approach and flattened out as the aircraft neared the transmitter. Success with the constant-intensity glide path proved elusive (little was known of multipath at that time), and eventually the concept was abandoned in favor of a straightline glide slope formed by the intersection of two beams. This so-called equisignal technique is used on the current world-standard Instrument Landing System (ILS).

1.3 PIONEERING WORK IN GERMANY

During the early 1930's important contributions to landing guidance were made by Dr. Ernst Kramer at the Lorenz Company in Germany [3]. Kramer's system comprised a vertically polarized combined localizer and glide-slope transmitter at 33.3 MHz that produced an equisignal azimuth beam much in the same manner as the current ILS. However, modulation was keyed to produce a series of Morse Code "e's" (dots) on one side of the course and "t's" (dashes) on the other side. In addition to the aural indications, a cockpit instrument was developed that gave the pilot a visual indication as to whether he was to the right, left, or on the course centerline. Kramer's system used marker beacons to signal the start of descent. The glide slope was a constant intensity type consisting merely of the contours of the underside of the radiation pattern that produced the localizer signal. Kramer's system went beyond the experimental stage and was installed at various airports in Germany. A Lorenz System was tested by the U.S. Bureau of Air Commerce at Indianapolis in 1937 [2]. A version called Standard Beam Approach (SBA) was installed at some military airfields in the United Kingdom prior to World War II.

1.4 EARLY U.S. ARMY TEST

The U.S. Army ran tests on its so-called A-1 System in 1934 and 1935 [5]. The A-1, a compass locator-marker system, was then declared the U.S. standard. The system included two low-frequency non-directional beacons installed in line with the runway approach centerline. Each beacon was equipped with a 75-MHz vertical beam marker transmitter. The pilot first tuned his left-right radio compass to the outer station and maintained his altitude until a marker beacon light indicated he was over the station. He then tuned to the inner station and started a controlled descent toward it, maintaining the rate of descent until reaching his minimum altitude. The inner marker beacon light signalled his proximity to the landing strip.

The high point in the history of the compass locator-marker system was reached in 1937 when Captains Holloman and Crane together with Mr. Ray Stout made a successful automatic landing in a Fokker ZC-14 monoplane at Wright Field [6]. Equipped with a Sperry A-3 autopilot, the aircraft used marker beacons to signal activation of various phases of flight. Limitations inherent in this technique prevented the system from going beyond the experimental stage.

Although the system was abandoned as the U.S. standard in 1937, outer and inner low-frequency homing stations together with VHF marker beacons later became an integral part of the standard ILS and served as an independent, redundant system. This backup was particularly important to pilots during the introductory phase of the ILS. It was, in effect, an early version of the concept of an "independent landing monitor."

1.5 BUREAU OF AIR COMMERCE

Responsibility for landing guidance system development in the U.S. was transferred in 1933 from the Bureau of Standards to the newly created Bureau of Air Commerce, predecessor to the Civil Aviation Authority (CAA) and today's Federal Aviation Administration (FAA) [5].

In an attempt to get away from the problems that had plagued the constant-intensity glide-slope system, a decision was made to use a much higher frequency. The Bureau of Air Commerce contracted with Massachusetts Institute of Technology (MIT) to develop a 750-MHz constant-intensity system [8]. This work was carried on until 1940, when field tests showed that ground reflections distorted the radiation pattern and prevented the much sought after smooth glide path.

In 1938 work was started by International Telephone and Telegraph Corp. (ITT) on a project sponsored by the CAA to develop a 110-MHz localizer, a 93.9-MHz constant-intensity glide-slope, and 75-MHz markers [2] [8]. The system was demonstrated in 1939 complete with monitor and remote control. The localizer of this system was basically the same as the current world standard. Disappointed with the results of the constant-intensity glide slope used in the system, the CAA contracted with ITT in 1940 to develop a straight-line equisignal glide-slope system at 330 MHz. This was the forerunner of the current ILS glide-slope system.

SECTION 2

WORLD WAR II ACTIVITY

2.1 INCREASE IN SYSTEMS DEVELOPMENT AND PRODUCTION ACTIVITY

The start of World War II caused a rapid increase in the tempo of landing guidance development. Responsibility for such development shifted from civil to military centers. Bad weather and attendant poor visibility caused major problems for military air transport, ferrying, tactical and strategic military operations; thus finding a solution to the all-weather landing problem was given high priority. The immediate result was a proliferation of ideas and programs with a number of intensive activities being carried out simultaneously. Some of the more important activities were:

- Sperry entered into a contract with the U.S. Army to develop a 3000-MHz (microwave) equisignal glide slope and localizer [2], shown in Figure 2.1.

Figure 2.1 Early 3000-MHz Sperry Microwave Glide-Slope System (about 1942)

- A team at Wright Field developed a programmed aircraft capable of automatic takeoff, navigation, and landing.

- During the late summer of 1941, Capt. Koster, a German pilot employing a system developed by Siemens, made a series of automatic takeoffs and landings in fog at Dispensee near Berlin [4]. Because of the success of this experiment, Koster believed that the problem of blind landing was solved.

- In November of 1941 Dr. Luis Alvarez of the Radiation Laboratory, MIT, working under the National Defense Research Committee, conducted initial tests to determine the feasibility of adapting radar to the problem of landing guidance and to determine if the information derived from radar could be transmitted by voice radio and effectively used by the pilot [9]. These tests marked the beginning of the successful Ground Controlled Approach (GCA) program, which is still in active and large-scale use by the military services.

• Another project at the Radiation Laboratory, MIT, resulted in the development of a 3-centimeter microwave system based on the equisignal principle, with simple crystal video receivers in the aircraft. The project, headed by Dr. Jack Buck of Canada, suffered from lack of reliable components but proved the feasibility of highly portable systems similar to those in tactical use today.

2.2 SHIFT IN EMPHASIS TOWARD MILITARY DEVELOPMENT

As a result of the entry into the war, emphasis was placed on finalizing development so that production could get underway.

2.2.1 AIRCRAFT RADIO LABORATORY, WRIGHT FIELD. During this period (1941 through 1943) Wright Field funded the development of a portable military system based primarily on the background work of the CAA and using the same localizer and glide-slope radio frequencies and audio modulations [2] [8]. ITT, the developer of the civil system, was the major contractor on the military effort. Although the number of frequency channels has increased and antennas have been improved, the basic ILS has not changed since that time. The original production system consisted of a six-channel localizer, a single-channel glide slope (soon expanded to three channels), and three 75-MHz marker beacons: an outer marker 5 miles from touchdown to indicate start of letdown, a middle marker 1 mile from touchdown to indicate the decision point (whether to continue descent or abort the approach), and an inner marker to mark the beginning of the useable landing area. The inner marker was dropped from the system; only recently has its use been revived in some installations.

2.2.2 U.S. SYSTEM STANDARDIZATION. By 1943 there were three strong contenders for a standard landing system in the U.S.: the Sperry Microwave System, the ITT VHF-UHF SCS-51 System, and the GCA developed at Radiation Laboratory/MIT.

The decision of the U.S. to standardize on the VHF-UHF ILS was made at a meeting of military and civil representatives at Pittsburgh Airport in 1943. As a result, a large-scale production effort for military implementation was started. GCA was favored for certain military applications, and the first production unit of GCA was delivered during that year to the Army Signal Corps by Gilfillan. The GCA operated as a tracking radar at X-band.

2.3 OPERATIONAL TESTING

2.3.1 GROUND CONTROLLED APPROACH. In June 1943, a crew including Dr. Alvarez carried prototype equipment aboard a British battleship to England and demonstrated the system for a 2-month period at a Royal Air Force (RAF) bomber base [9]. At the end of this period the RAF was convinced of the usefulness of the system.

2.3.2 SCS-51 ILS. Early production VHF localizers and markers were given a rigorous test during the operation of an Air Transport Command all-weather airway [5]. Airfields from New York to Newfoundland were equipped with the new SCS-51 system and Col. E. A. Cutrell, heading the effort, demonstrated the feasibility of landing military aircraft via ILS under severe conditions.

2.3.3 JOINT BRITISH-AMERICAN TRIALS. In January 1944 an important event to the future of landing guidance systems took place when a small team, headed by Lt. Col. F. L. Moseley of Wright Field, took the first complete prototype SCS-51 ILS to the United Kingdom (U.K.) for joint British-American trials [4] [10]. Not only were the trials successful and the system adopted for use by the U.S. 8th Air Force and units of the RAF, but Lt. Col. Moseley interested both British and American as well as other allied observers with demonstrations of automatic landing approaches using a displacement rate coupler he had developed at home and carried unofficially to the U.K.

G/C J. A. McDonald, Commanding Officer at RAF, Defford where the tests were conducted, wrote: "It should be recorded once and for all that development in auto-approach technique in this country (U.K.) involving use of new standard ILS owes its origin to this 'Black Box', first used at Defford in 1944."

During the conduct of the Defford test, F/L Barbour, upon the enthusiastic urging of the project officer, S/L F. C. Griffiths, modified a Rebecca-Eureka radar system to provide distance measuring equipment (DME) and also to permit automatic orbiting of the station at a pilot-selected radius. (The need for DME as part of landing guidance has long been recognized, but only recently has it been included as a standard part of tactical and civil systems.)

2.4 WARTIME IMPLEMENTATION

Although the high-priority production program did not make equipment available until late in the war, both the SCS-51 ILS and GCA were produced in quantity and were deployed in time to provide substantial relief in many parts of the world.

SCS-51 airborne equipments were carried aboard most tactical and transport aircraft with the exception of fighters. More than 30,000 airborne systems were built [2]. Ground stations were installed at a number of airbases and additional units were moved into France as the war entered its final phase. Systems were installed in critical locations throughout the world, including the China-India Hump operation, the Aleutians, and the Pacific theater.

GCA units were used extensively by Navy and Air Force units in the Pacific and were successfully used in bad weather operations in Alaska and particularly in the Aleutians, where five systems were installed. They were also active in support of the Pacific B-29 campaign [9].

Although both systems arrived too late to have a great impact on the course of the war, the worth of landing guidance systems was fully demonstrated, not only for military operations but also for the large-scale civil programs that were to follow.

SECTION 3

POST WORLD WAR II ACTIVITY

3.1 RETURN TO CIVIL PROGRAMS

Immediately following World War II the landing guidance activity was primarily concerned with applying the lessons learned from military experience and adapting military equipment to civil aviation.

3.1.1. ILS ADOPTED BY ICAO. With the background of military experience, the ILS was proposed as a world-wide standard system at the provisional International Civil Aviation Organization (ICAO) meeting held in Chicago in 1946. Agreement was reached and the ILS has been the international standard system throughout the noncommunist world and in parts of the communist bloc nations up to the present time.

3.1.2 U.S. CIVIL IMPLEMENTATION. In 1946 the U.S. CAA ordered 47 ILS and began installation at important airports throughout the country [2]. Elements of the system were assembled on site by CAA engineers. Airborne equipments used by the airlines during this period were surplus SCS-51 military equipment. The first authorized use of reduced weather minimums for civil operators occurred in 1947. In 1948 the CAA ordered eight GCA or Precision Approach Radar (PAR) equipments to supplement the ILS at important locations.

Instrument landing approaches in the U.S. prior to installation of ILS were accomplished by a low-frequency range technique in which an aircraft overheaded a range and then let down on one of four sector legs of the range on a timed descent using the barometic altimeter as a reference. This type of approach was limited to minimums of a 400-foot ceiling and 1-mile visibility. Introduction of the ILS cut these minimums in half at most locations. Even more important than the improved minimum was the increase in safety resulting from the use of precision approach guidance.

3.2 ESTABLISHMENT OF NEW BLIND-LANDING RESEARCH CENTERS

An important post-war development was the organization of specialized research establishments to solve the complex problem of all-weather landing.

3.2.1 BLIND LANDING EXPERIMENTAL UNIT. The Blind Landing Experimental Unit (BLEU) located at Bedford in the U.K. concentrated its efforts on automatic landing, and over a long period of development perfected a system called Autoland [4]. The system used basic ILS with a radio altimeter set to initiate a decreased rate of descent for flare as the aircraft neared the touchdown point. Initially, one Autoland feature was a system of magnetic leader cables buried in the runway. Magnetic loops were carried aboard the aircraft to sense the buried cables. This system gave highly precise, stable runway alignment guidance, but because of the additional equipment required and because ILS localizers had been improved to meet the full landing requirement, the magnetic system was dropped from the system. Most aircraft certified for automatic landing today use the other basic features of the Autoland system.

3.2.2 AIR FORCE ALL-WEATHER FLYING DIVISION. The U.S.A.F. All-Weather Flying Division, a part of the Wright Field complex, located at Wilmington, Ohio, also made major contributions to landing guidance technology. An all-weather airway was established between Wilmington and Andrews A.F.B., Washington, D.C., and a near perfect schedule was maintained regardless of weather [6]. The climax of the U.S.A.F. effort occurred in September of 1947 when a four-engine C-54 Skymaster transport aircraft was flown automatically from takeoff to touchdown across the North Atlantic to Brize Norton, England. Using information stored in punch cards plus a computer, the system was programmed for all phases of flight, including tuning in and homing on weather ships as a means of navigation.

3.2.3 LANDING AIDS EXPERIMENT STATION. The Landing Aids Experiment Station (LAES) located at Arcata, California, was a research station jointly sponsored by the U.S.A.F, Navy, and CAA for the express purpose of developing improved visual aids [11]. The site was selected because it has one of the highest incidence of fog in the U.S. Test flying was conducted mainly in weather conditions below a 100-foot ceiling and a quarter-mile visibility. The controlled nature of the testing and wide variety of systems tested provided the background for selection of the centerline and bar system of approach lights, which has become standard throughout the world. The LAES, in addition to its work with high-intensity approach and runway lights and visibility measuring systems, also tested fog dispersal systems and proved the economic feasibility of the Fog, Intensive Dispersal Of (FIDO) system which burned fuel oil under high pressure. However, when a system based on the Arcata techniques was later installed at Los Angeles International Airport, it was found to be ineffective because of the increased spacing between the rows of burners. Safety rules dictated that the burners be located well off the edge of the runway. As a result, the heat from the burners could not clear fog from the center of the runway as they had at Arcata. The project was subsequently abandoned.

SECTION 4

INSTRUMENT LANDING SYSTEM

The ILS is by far the most universally used landing guidance system currently in operation. Since standardized under the aegis of ICAO at the end of World War II, ILS is in use in virtually every country throughout the world. Because of ICAO's performance specifications, any ILS-equipped aircraft can expect satisfactory operation at any approved ILS airport installation. This is true even though both airborne and ground equipments are produced in many countries to widely different designs [12].

At present there are over 500 ILS installations in operation throughout the world and an approximately equal number to be added. This number will depend to some extent on the speed with which international agreement can be reached on a replacement for ILS.

4.1 FUNCTIONAL ANALYSIS OF ILS APPROACH AND LANDING

4.1.1 TRANSITION FROM ENROUTE. An ILS procedure begins with the transition from enroute flight to final approach. This may be accomplished by departing from the last VHF Omni Range (VOR) navigation station of the enroute flight on a radial that will intercept the localizer course approximately 7 to 10 miles from the runway [13].

4.1.2 LOCALIZER INTERCEPT. The aircraft intercepts the localizer course in level flight at an altitude (specified by the approach plate of the pilot's flight manual) and distance that place the aircraft below the glide slope. This allows the pilot to become stabilized on the localizer course before starting descent.

4.1.3 GLIDE-SLOPE INTERCEPT AT OUTER MARKER. The pilot continues level flight although the glide-slope indicator reads full-scale fly-up. As the aircraft intercepts the glide slope, the indicator starts to move toward center, and the pilot then makes the necessary power and trim adjustments to give a rate of descent consistent with the glide-slope angle. As he reaches the center of the glide slope, he receives the aural keying and visual flashing of the 75-MHz Outer Marker vertical beacon. The approach plate indicates the proper altitude at which the glide slope intercepts the Outer Marker for the specific facility being used. If he notes any significant deviation from the published value, before starting his descent he must determine whether the discrepancy is caused by an improper altimeter setting or a malfunction of some part of the system. With a normal interception, he is assured at this point that the key elements are working properly and he can safely begin his descent.

4.1.4 STABILIZED APPROACH. Descent from the Outer Marker involves keeping both localizer and glide-slope indicators centered by making small changes in heading and in rate of descent. Wind shear and turbulence during descent can cause deviations that must be corrected.

4.1.5 DECISION HEIGHT AT MIDDLE MARKER. If the approach is being made to Category I (CAT I) weather minimums (which can be as little as a 1,800-foot Runway Visual Range at a fully equipped airport), the pilot must have in view an element of the approach lights, runway lights, runway markings or significant ground references by the time he receives the aural-visual signals from the Middle Marker in order to continue his approach. If he reaches this decision height and does not have adequate visual reference, he must abort the approach and execute a missed-approach procedure. This usually involves a climb-out to a navigational fix where Air Traffic Control (ATC) can instruct him further.

4.1.6 FLARE. With the ground in sight the pilot continues his rate of descent until reaching a height of about 60 feet above runway elevation; he then slows his rate of descent so that he will further approach the runway on an exponential flight path.

4.1.7 DECRAB. Having checked rate of descent to an acceptable level, it is necessary for the pilot to remove any difference between the fore-and-aft alignment of the aircraft with the alignment of the runway by "decrabbing". The misalignment might be caused by last-minute course correction maneuvers or by a crosswind.

4.1.8 TOUCHDOWN. Touchdown is the initial contact of the aircraft wheels with the runway. For most aircraft it should be made at a rate of descent of about 2 feet per second, and with fore-and-aft alignment in close coincidence with the runway alignment. The touchdown should be near the runway centerline and at a suitable longitudinal distance beyond the point where the glide slope intersects the runway. This distance varies with the performance characteristics of the aircraft.

4.1.9 ROLLOUT. This portion of the landing begins with touchdown and ends with the deceleration of the aircraft to taxi speed or when the aircraft turns off the runway to enter a taxiway. Guidance must be provided during rollout when aircraft are operating in the poorest conditions of visibility (CAT III). Guidance must also be available under these conditions in order for the aircraft to clear the runway quickly and make it available for the next landing aircraft.

4.1.10 AUTOMATIC APPROACH AND LANDING. The procedures for automatic approach and landing on ILS are essentially the same as those for manual flight. Some automatic systems are limited concerning the angle of intercept with the localizer, and this must be taken into account in the procedure. Automatic systems may also have a limitation in the amount of acceptable windshear [14]. The major difference between automatic and manual landing occurs after reaching decision height. The automatic system continues on both localizer and glide slope until a predetermined height of about 60 feet is signalled by the radio altimeter. The system then starts a programmed flare that continues until just prior to touchdown, when, again on signal from the radio altimeter, the aircraft heading is automatically brought into alignment with the runway. At touchdown there is a transition from aerodynamic control to wheel steering, and the aircraft is maintained on centerline by this means.

4.2 ILS DESCRIPTION

The ILS as standardized by ICAO consists of a localizer for runway alignment guidance, a glide slope for elevation guidance, and marker beacons for providing key checkpoints along the approach [15]. Compass locator stations may be included in some systems, and DME are now being added to some ILS installations to provide continuous reading of distance for the airport during the approach. The geometry and general layout of the ILS are shown in Figure 4.1.

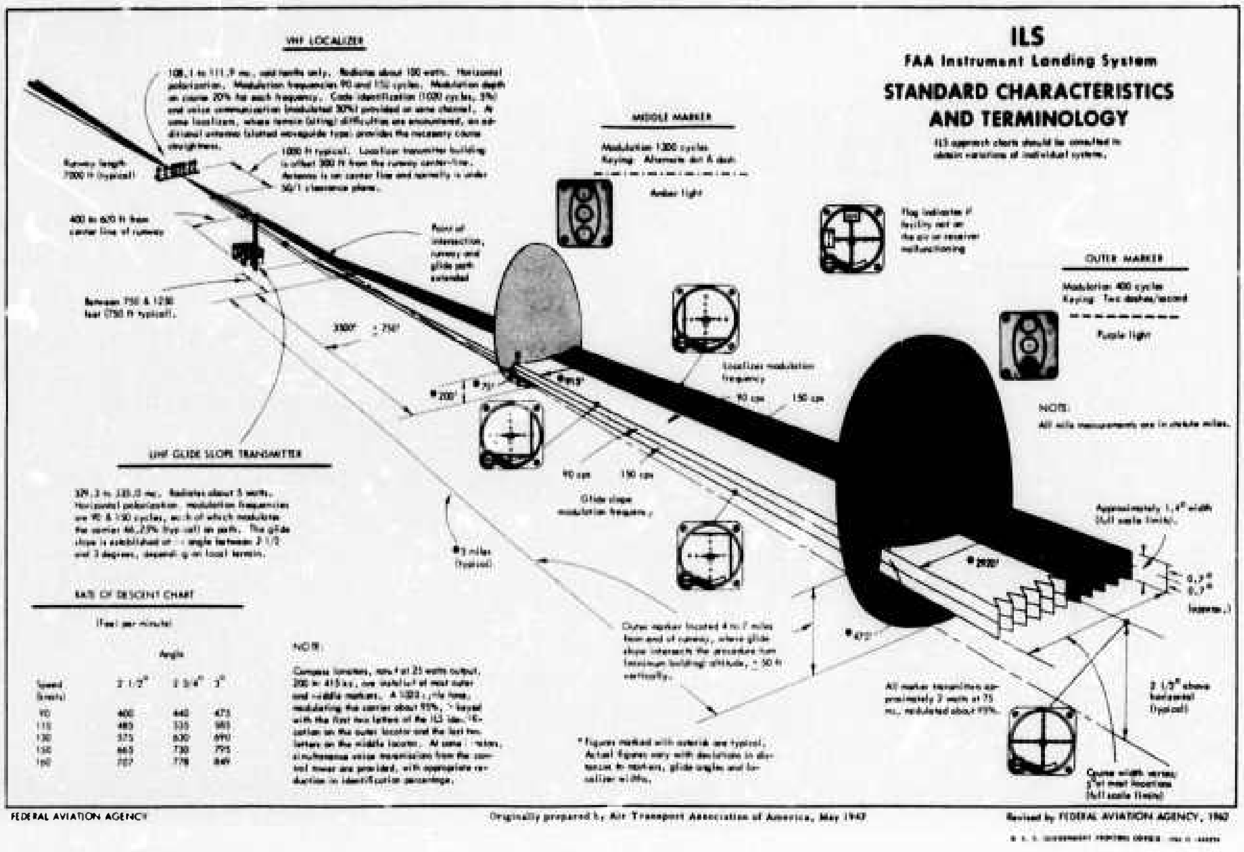


Figure 4.1 ILS Standard Characteristics and Terminology

Some ILS installations include both front and back courses and provide localizer signal indications throughout 360 degrees of azimuth. These are no longer required by ICAO at even the busiest airports, and most current systems limit coverage to the ICAO requirement of ±35 degrees and eliminate or greatly reduce the back course. The result of this reduction in coverage is improved multipath characteristics and conservation of frequency spectrum, which are critical in the case of the ILS.

4.2.1 LOCALIZER. The localizer, which provides lateral guidance, produces a course formed by the intersection of two field patterns [8] [16]. One pattern is modulated by an audio frequency that produces mainly 150 Hz to the right of course (as viewed from an aircraft inbound towards the runway) and the other pattern is modulated by an audio frequency of equal amplitude that produces mainly 90 Hz to the left of course. The "on course" is the vertical plane coinciding with the extended centerline of the runway.

Using, as an example, an eight-loop array system (Figure 4.2), the courses (front and back) are derived from four composite 90- and 150-Hz fields. In one of these fields the carrier sideband component is radiated by the carrier antenna pair, which produces both 90- and 150-Hz signals having constant RF phase in all directions. The other fields are the vector sum of three sideband pairs of antennas. The spacing between loops and the amplitude of the currents determine the radiation patterns. The patterns are designed to produce "solid" or "full clearance" left or right signals on each side of the course.

The two center loops are adjusted to be in phase and are fed with carrier energy modulated equally with 90- and 150-Hz tones. Only sideband energy is fed to each of the three loop pairs located on either side of the carrier loops. The three loop pairs on one side are in phase, but of opposite phase to the three loop pairs on the opposite side. The signals received in the aircraft will produce a "fly right" indication for the pilot when to the left of course in the predominately 90-Hz region. Similarly, a "fly left" indication will be produced for the pilot on the opposite side of the course in the predominantly 150-Hz region.

The present ILS localizer operates in the VHF band between 108.1 and 111.9 MHz with 20 channels spaced at 100 kHz. The system uses horizontal polarization to minimize multipath effects from reflecting structures and objects. Each station has an identifier, which is keyed 10 times per minute at a modulation frequency of 1020 Hz. The system is generally useable at a distance up to 25 miles at 1,000-foot approach altitudes and above. Line-of-sight attenuation due to earth curvature prevents reliable reception at distances greater than this below the 1,000-foot level.

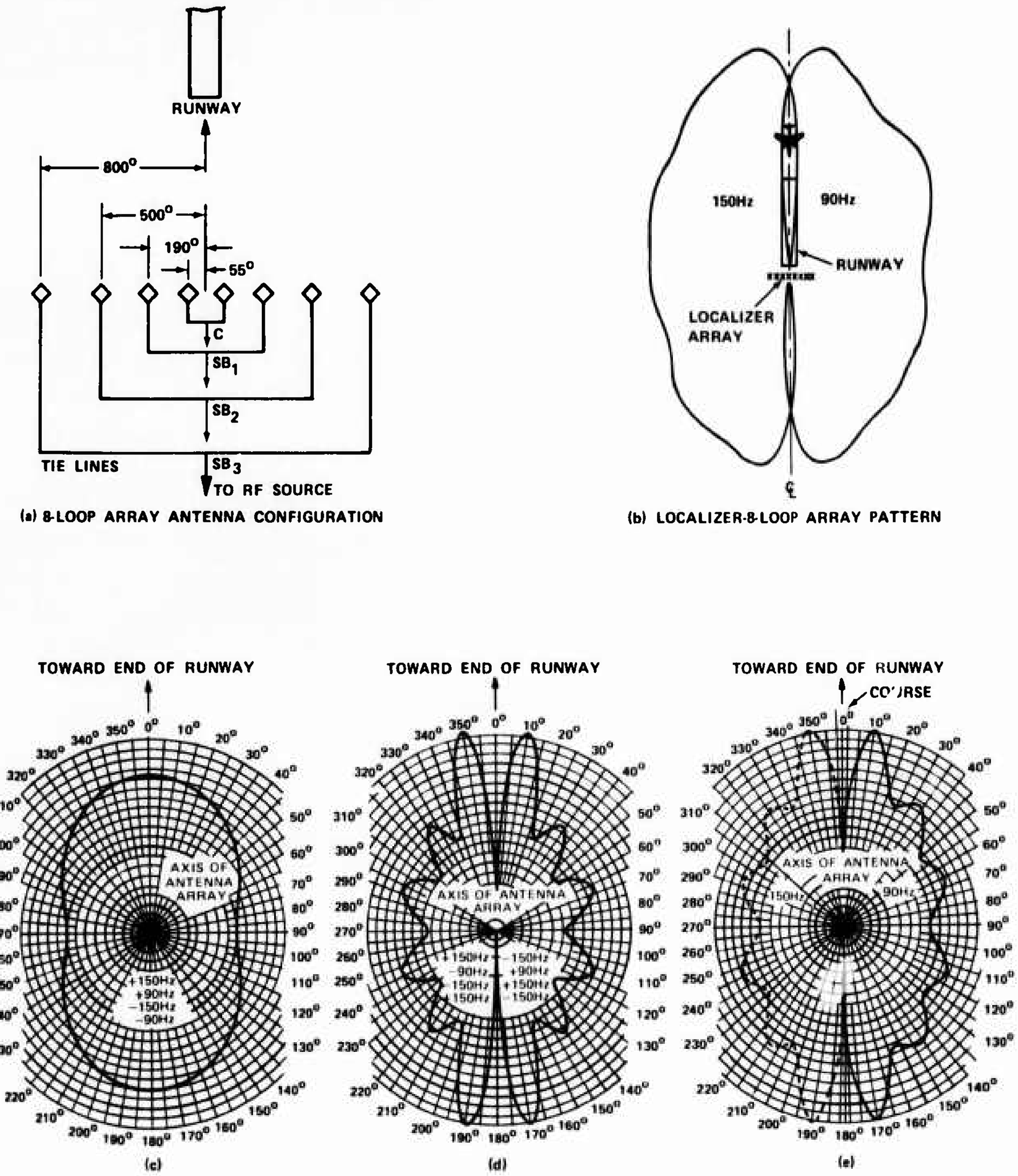


Figure 4.2 8-Loop Array Antenna Patterns

There have been many versions of ILS localizers, but all have been based on the same fundamental principles. Some of the improvements to the basic system are discussed in paragraph 4.4.1.

4.2.2 GLIDE SLOPE. The ILS glide-slope unit was first introduced as an equisignal system in which the course was produced by two overlapping radiation patterns, one modulated at an audio frequency of 90 Hz to provide a "fly-down" indication when it is the dominant signal and the other at 150 Hz to provide a "flyup" indication [8] [17].

4.2.2.1 Equisignal System. Two horizontally polarized antennas were mounted on a vertical mast in an image-type array to produce the desired radiation patterns. The lower antenna, mounted about 4 feet above ground, radiated the 90-Hz modulated signal, and the upper antenna, mounted about 19 feet above ground, radiated a multilobe pattern modulated at 150 Hz. The lowest lobe of the 150-Hz radiation overlapped the lower portion of the wider 90-Hz lobe. The intersection of the two lobes provided the "on course" of the glide slope. A similar overlap occurred at an upper intersection, producing a false course with opposite sensing. The false course was at a high angle so that it was not confused with the true course. The angle of the glide slope was adjusted by changing the height of the antennas.

Difficulty was experienced with the equisignal slope in the presence of heavy snow, and it was found necessary between 1948 and 1953 to convert all U. S. civil systems to a so-called null reference system [8].

4.2.2.2 Null Reference System. The null reference system, like the equisignal system, is an image-type array using two antennas. The lower antenna is located at one half the height above ground of the upper antenna. A reference signal at the carrier frequency (approximately 335 MHz), modulated with equal amounts of 90 Hz (fly-down) and 150 Hz (fly-up), is radiated from the lower antenna. The upper antenna radiates only 90-Hz and 150-Hz sideband energy. The first null above the ground provides the selection of the glide-slope angle, and the course width is contained within the region of the null. Phasing between the sideband signals radiated by the upper antenna and the reference signal of the lower antenna is such that, below the null, fly-up sidebands add and fly-down sidebands subtract. Above the null, the opposite phenomena occurs.

The greater height of the antennas of the null reference system (i.e., 28 feet for the upper antenna compared to 19 feet for the equisignal system) makes the null reference system much less susceptible to snow buildup or tidal reflection problems. Changes in ground plane of as much as 2 feet cause little shift in the path angle. However, the greater antenna height has an adverse effect in that it illuminates a greater image area and increased attention has to be paid to providing a large level ground area in order to obtain high-quality courses.

In later versions of the null reference system, the upper antenna was lowered 25 percent and the lower antenna set at one-third of the new height, creating a sideband reference system with a null at 3 degrees. Further improvement utilized "capture effect" principles that increased the integrity of the fly-up signals below the path. Three dipole antennas are used in the capture-effect glide slope. These systems are now in common use.

The glide slope, like the localizer, currently has 20 channels, but they are located in a band from 329.3 to 335 MHz, approximately three times the frequency of the localizer, and with 300-kHz spacing between channels. Localizer and glide-slope channels are paired so that they are selected by a single control in the cockpit.

4.2.3 MARKER BEACONS. Marker beacons have been an integral part of the ILS from the very beginning and, although the hardware has changed, the principle is essentially the same [8]. A CAT I ILS includes two marker beacons, the Outer and the Middle Markers. CAT II and III installations include a third unit, the Inner Marker.

The function of the Outer Marker, located about 4 to 7 miles from the runway threshold, is to signal the pilot to start descent. A 400-Hz tone is keyed at two dashes per second and causes a purple light to flash on the instrument panel.

The Middle Marker, located nominally at 3,500 feet from runway threshold, is to alert the pilot to the fact that CAT I decision height has been reached. A 1300-Hz tone, keyed with alternate dots and dashes, flashes an amber light on the instrument panel.

When a CAT II or CAT III ILS is used, the Inner Marker is to signal CAT II decision height. A 3000-Hz tone, keyed at six dots per second, flashes a white light to indicate decision height.

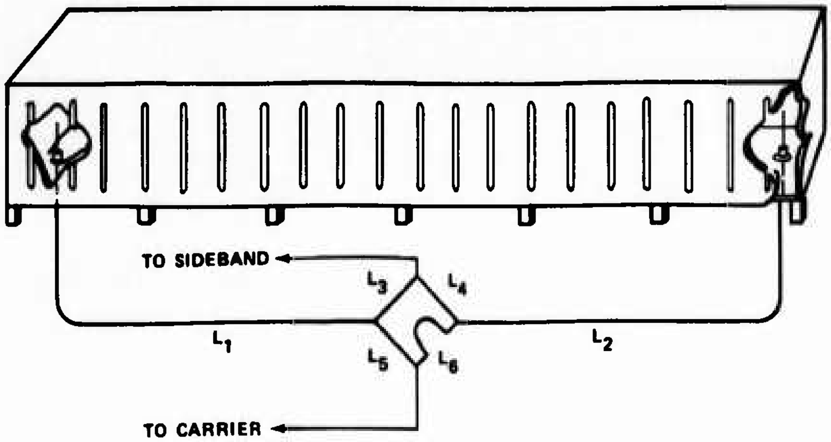
All marker beacon transmitters operate on the single fixed frequency of 75 MHz. Carrier power is generally limited to 1 to 2 watts. A conventional system uses two half-wave, cophased, collinear dipole antenna elements spaced above a 20-by-20 foot counterpoise, which in turn is elevated above ground to avoid snow problems. More recently, low-cost markers have been developed that are mounted on a single pole. The transmitter is contained in a weather-proof cabinet and an antenna consisting of two simple dipoles is used. Many of these systems are designed to be independent of commercial power, using a device such as a propane generator as a self-contained power source.

Markers, like other elements of the ILS, are carefully monitored, and a malfunction causes a warning signal to alert a ground controller, who can immediately notify the pilot.

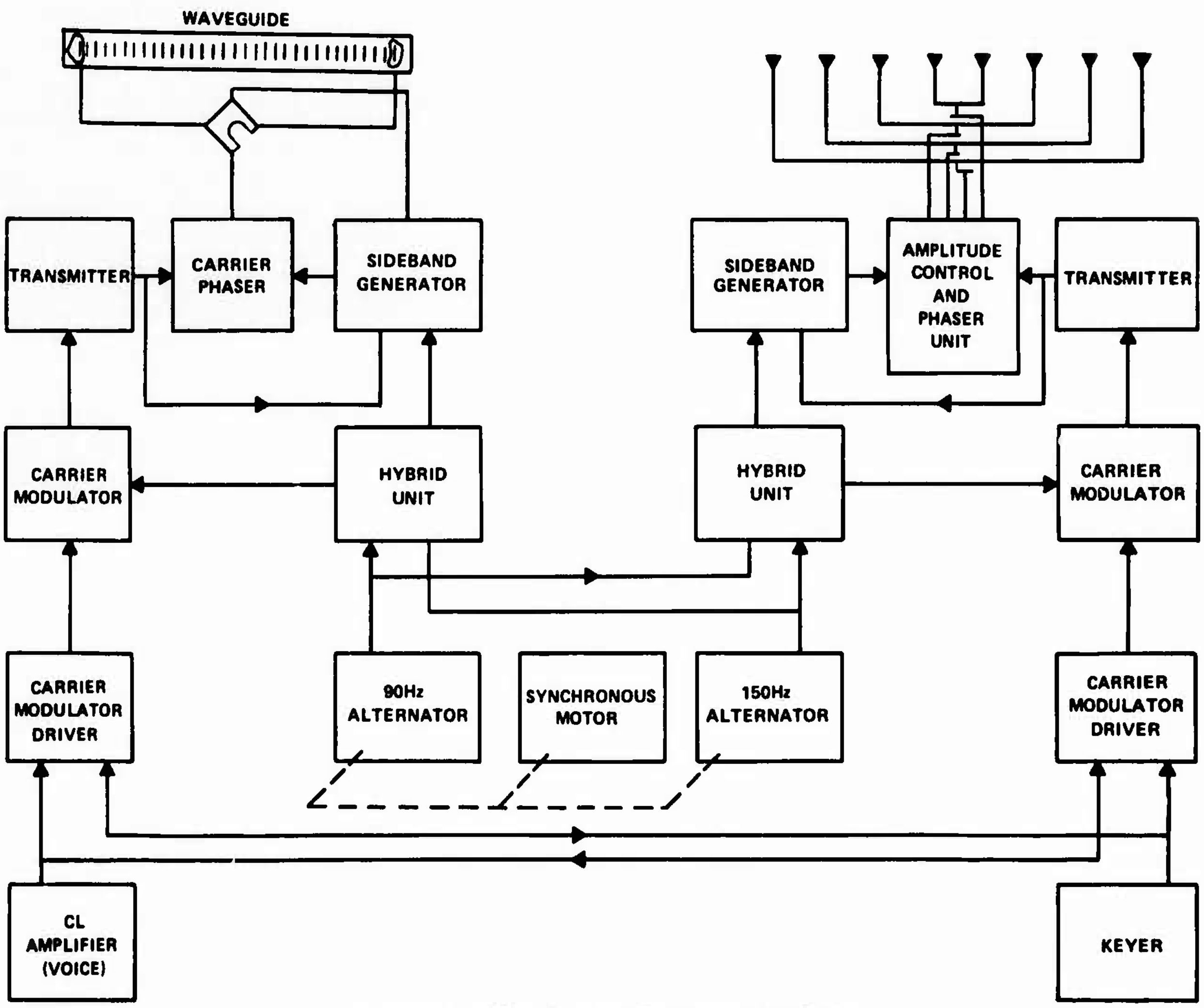
4.3 LIMITATIONS OF ILS

4.3.1 SITING. Implementation of ILS by both military and civil agencies has provided extensive experience which has shown certain weaknesses in the system. Both the glide slope and localizer signals have been affected by the presence of reflecting objects such as hangars, transmission lines, and nearby trees. The CAA developed a localizer antenna system based upon a large 90-foot aperture wave-guide array that provides narrow beams along the approach path (shown in Figure 4.3). This was supplemented by another array (shown in Figure 4.4) that provides the required signal coverage of at least ±35 degrees. The so-called wave-guide localizer improved the courses of many localizers, but because of the necessity of dual transmitters, it required complex antennas and additional monitoring, which were too expensive for smaller airports. In recent years an antenna system called the V Ring, shown in Figure 4.5, was developed that, when used in a 15-element array, produces a narrow beam characteristic along with the required angular coverage. These antennas are produced at a reasonable cost and require a single transmitter. Although these developments have reduced the multipath problem, they have by no means eliminated it; hence multipath continues to be a very serious problem for the system.

Another severe problem with the ILS is that the formation of beams by the conventional glide-slope antenna array depends upon a reasonably smooth, level ground plane in the area ahead of the antennas beyond the end of the runway. At many locations there is simply not enough level ground to provide the required reflecting surface. The result is either poor performance or inability to provide an acceptable glide-slope signal. To overcome this, many airports have resorted to extensive site alteration using landfill, or in some cases an artificial ground plane made up of screen material. Such site preparation frequently costs several hundred thousand dollars, and in the case of some extreme situations the estimated cost of suitable landfill runs to several million dollars.



(a)—WAVEGUIDE, BRIDGE, AND TIE-LINE ARRANGEMENT.



(b) BLOCK DIAGRAM OF ELECTRONICALLY MODULATED DIRECTIONAL LOCALIZER.

Figure 4.3 Wave-Guide Directional Localizer Block Diagram

Figure 4.4 Wave-Guide Localizer Showing Clearance Array and Monitor Probe

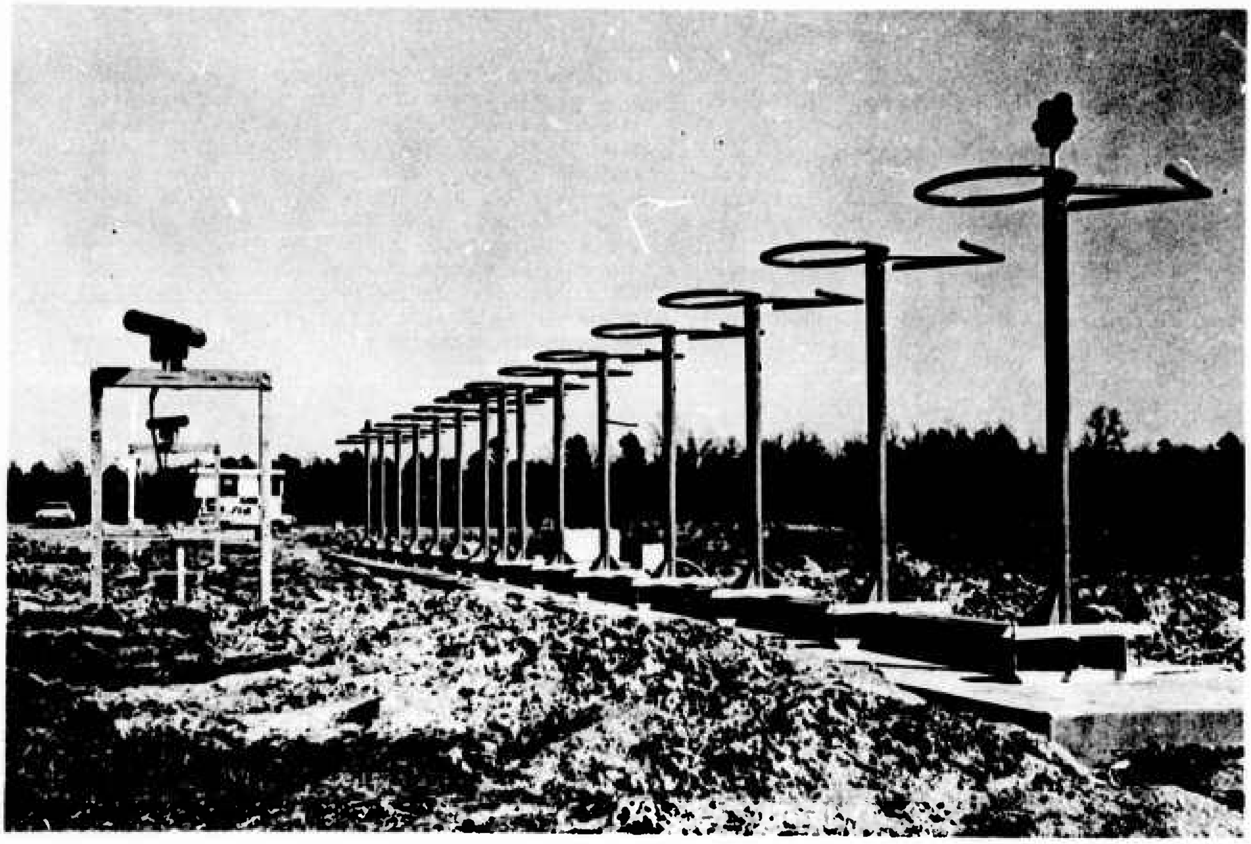


Figure 4.5 V-Ring Localizer Antenna Array

Most efforts to overcome this problem have involved special antenna design in which the signal is formed without aid of a reflecting ground plane. The difficulty is that at the UHF frequencies used, such large apertures are required that the antennas, including the present ILS glide-slope antenna mast of approximately 35 feet in height, constitute undesirable obstructions on the airport surface.

The glide-slope angle also was affected significantly by accumulation of snow. To correct this, the null-reference glide-slope antenna system was developed that has replaced all U.S. civil ILS glide slopes. This is a great improvement, though not a complete solution to the snow problem.

A similar problem was encountered where the glide slope is projected over a body of water. A rising and falling tide can also affect the stability of the glide-slope angle. Again, special antennas have been designed to cope with this situation.

4.3.2 FREQUENCY SPECTRUM AVAILABILITY. The limited availability of frequency spectrum inhibits future growth of standard ILS. Both localizer and glide-slope systems operate in portions of the frequency spectrum that, because of international allocations, have little chance of expansion to accommodate new facilities. The alternate solution, less separation between channels, already is being implemented; the current 100-kHz spacing between channels of the localizer is being cut to 50 kHz, and with a similar plan is going into effect for the glide slope. However, even doubling the paired localizer and glide-slope channels from 20 to 40 falls far short of meeting future requirements, particularly as airports grow and require multiple approach capabilities.

4.3.3 OVERFLIGHT INTERFERENCES. The ILS localizer signal is vulnerable to overflight by aircraft taking off directly in front of and over the transmitting antenna [18]. When this occurs, another aircraft using the facility for an instrument approach receives violent needle action for several seconds. A pilot making a manual approach can recognize and disregard these spurious signals, but it is very difficult to design an automatic system to take this erratic indication into account. Efforts have been made to minimize the problems of overflight, but the most successful solution to date is to control takeoffs so as not to interfere with aircraft on instrument approach. This, of course, has the effect of constraining the amount of traffic an airport can handle.

4.3.4 SINGLE PATH. The ILS is limited to a single narrow course, usually along the guide slope above an extension of the runway centerline. This prevents rerouting approaches to improve noise abatement over populated areas near the airport.

4.3.5 FLARE. The glide slope does not provide guidance to touchdown; for a complete landing, it must be supplemented with information obtained from a radio altimeter to initiate flare in order to touch down at an acceptable rate of descent.

4.3.6 MILITARY LIMITATIONS. Military usage of ILS is limited because the VHF and UHF frequencies employed require large antenna arrays unsuitable for portable tactical applications.

4.4 THE FUTURE OF ILS.

Most of the limitations of the ILS are basic to its relatively low radio frequencies. Very effective development work has been done to minimize these inherent problems, but ILS tends to be an expensive system to build and install, and its application is still restricted. Although it is generally acknowledged that ILS will be superseded by a microwave landing system (MLS), the tremendous investment in ILS design and installation plus trained personnel is almost certain to guarantee its continued existence for as long as two decades. Accordingly, a number of development programs are underway to extract the best possible performance out of the system during the remainder of its life [19]. New ILS will continue to be installed and older systems will be replaced by newer and better equipment. New techniques will be used to overcome deficiencies and limitations described in earlier paragraphs. These improvements can be categorized as: (1) antenna redesign to make the system less site-sensitive; (2) advanced mathematical modeling of installation factors for better control of reflecting objects and prediction of performance at a proposed site; (3) and revised monitoring, probably via the integral rather than the near-field type [20].

4.4.1 IMPROVED LOCALIZERS. Two localizer development programs in the U. S. show promise of significant reduction in site sensitivity; the Alford Array Antenna Element, which has already been put into limited service, and the Watts Prototype Slotted Cable Array.

4.4.1.1 The Alford Array Antenna Element. The Alford Array Antenna Element, shown in Figure 4.6, is an end-fire antenna in which a series of small radiating hoops are connected by an open, balanced transmission line [21]. One end of the transmission line is fed and the other terminated in its characteristic impedance. This array is used as a modular element; thus an array can be upgraded to higher performance such as CAT II or III by the addition of elements. The array permits integral monitoring, which eliminates overflight interferences to the monitor, a difficult problem for conventional systems.

The Alford Array can be used in apertures of 45 feet to 140 feet depending upon the performance required and the siting difficulty. The high directivity of the wide aperture makes possible operation at sites where reflecting objects are located as close as 5 to 8 degrees to runway centerline. The majority of sites can be served by arrays of much less aperture.

4.4.1.2 Watts Prototype Slotted Cable Localizer Array. A very promising new array currently under test has been developed by Watts Prototype of Alexandria, Virginia [22] [23]. Using a slotted cable array, the localizer presents a minimum obstruction with performance at least equal to the large wave-guide system. The slotted cable is used in a two-frequency capture effect system similar to the wave guide and clearance array system. The antenna itself (see Figure 4.7) consists of a 1-5/8-inch diameter transmission line cut to the length of the array. Sleeves are spaced equally along the transmission line, with one end of each sleeve fastened electrically to the outer conductor of the transmission line and the other end isolated from the transmission line by a dielectric sleeve. A probe attached to the free end of the sleeve passes through a hole in the outer conductor to provide electrical excitation to the slot.

Feed lines supplied from a hybrid unit apply excitation to each end of the antenna. Each successive sleeve is alternately reversed to take into account the 180-degree phase reversal for element spacing of one-half wavelength. The monitor pickup is a similar array running parallel to the antenna at a spacing of 26.8 inches and also functions as a reflector. The advantage of the monitor is the ability of checking radiation along the entire length of the antenna.

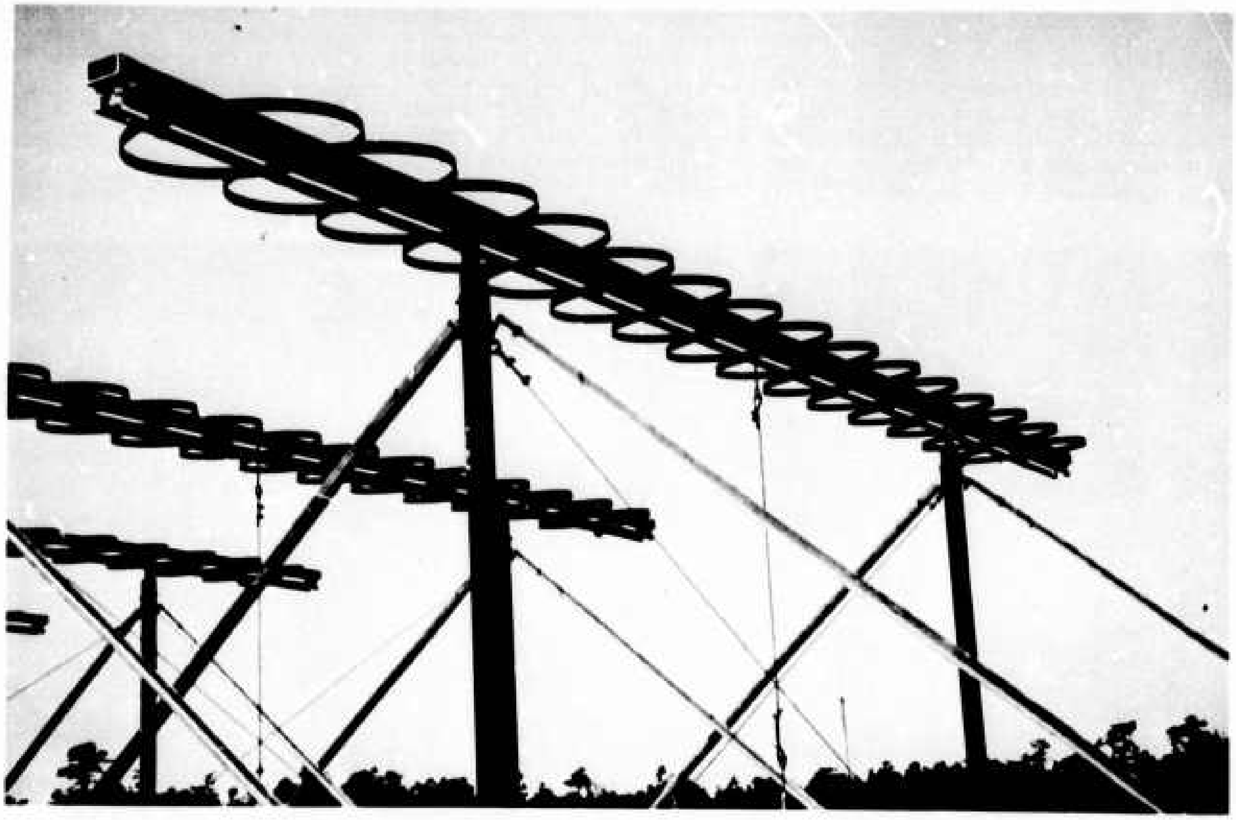


Figure 4.6 Alford Array Antenna Element Showing Element and Support Details

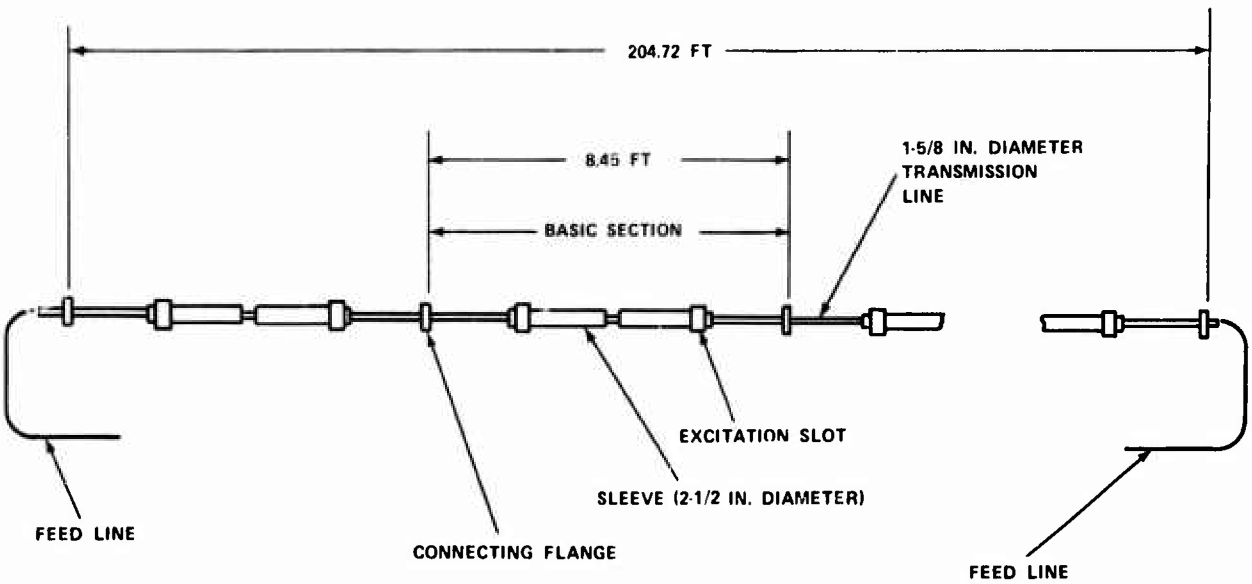


Figure 4.7 Watts Prototype Slotted Cable Radiator

4.4.2 IMPROVED GLIDE SLOPES. Most activity aimed at improving glide-slope performance has been in the development of antenna arrays capable of producing the glide-slope course with minimum ground illumination [24] [25].

4.4.2.1 NERA Glide Slope (Norway). The NERA glide-slope system is a 26-element wide-aperture array in which the high vertical directivity obtained from an antenna aperture of approximately 47 feet allows the signal structure to be shaped in the array instead of depending on image reflection of level terrain. The system has built-in pickup loops for monitoring [26].

4.4.2.2 Wave-Guide Glide-Slope Antenna. A wave-guide glide-slope antenna was originally developed by AIL Division of Cutler-Hammer to cope with the special conditions encountered at the LaGuardia Airport pier-type runways over water. The rising and falling tides caused an unacceptable shift in glide-slope angle when a conventional null-reference glide-slope transmitter was installed. The installation of a 60-foot wave guide, tilted slightly from vertical to provide the proper glide-slope angle, improved performance at the site to permit CAT I operation (see Figure 4.8). Later, Westinghouse Defense and Electronic System Center modified the antenna to meet CAT II requirements. A key part of the modification was length added to each end of the array for a total of 72 feet [27] [28].

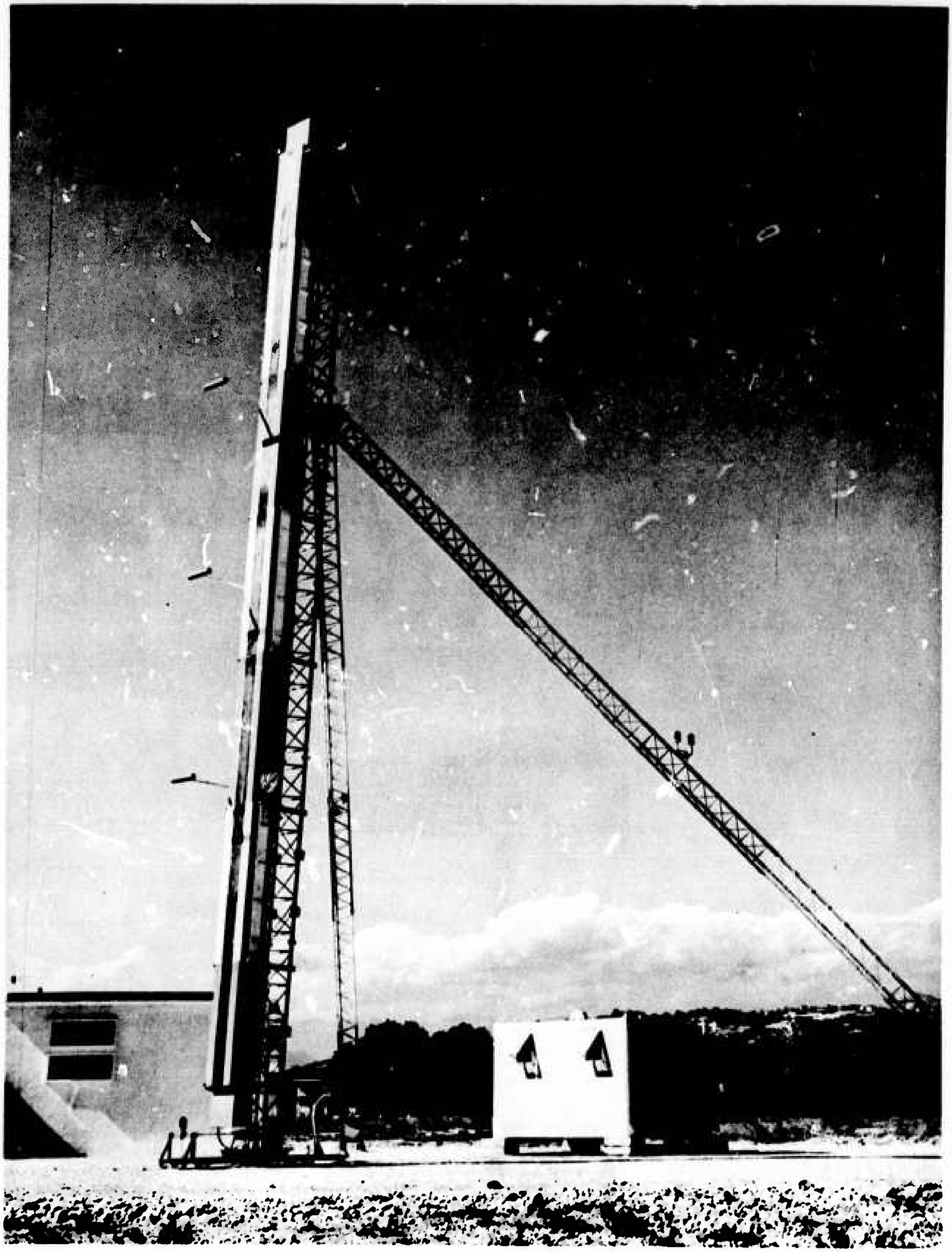


Figure 4.8 Wave-Guide Glide-Slope Antenna Does Not Depend on Terrain for Image Reflection

4.4.2.3 Watts Prototype Guide Slope. The two-antenna end-fire slotted-cable glide slope antenna uses a different approach than existing systems, and promises good performance for siting conditions under which other ILS systems are unworkable or impractical [19] [29]. The antenna array consists of two slotted cable antennas, each radiating circular phase fronts to control path angle over the specified azimuth sector. The two antennas (see Figure 4.9) have a common phase center located 200 feet from the runway centerline and typically 1,000 feet from the runway threshold. Azimuth coverage is nonsymmetrical in order to provide proper coverage over the approach end of the runway. A significant feature of the system is its low elevation profile (see Figure 4.10). The antennas are only 4 feet above ground, so that the proximity of the antennas to the runway edge does not create an obstruction problem.

4.5 VISUAL AIDS

4.5.1 EVOLUTION OF LIGHTING AND MARKING SYSTEMS. Early workers in the field of landing guidance made no provisions for lighting and marking systems, since the goal was blind landing, i.e., landing with reference only to instruments. Any compromise with a full solution was regarded with disdain. This idea persisted throughout the early development period. Not until landing systems development reached a greater degree of maturity was it realized that pilots would not land an aircraft filled with passengers based on the vagaries of a system of black boxes about which they knew little and sometimes trusted less.

The immediate answer was to restrict landings to conditions where there was a reasonable degree of visibility, measured by the extreme distances an observer could see objects or a limited candle-power light. Runway edge lighting was sometimes helpful at night, but early systems were too low powered for daytime operation and also were not beamed toward the approach. The pilot usually used his aircraft landing lights, but often this did more harm than good by "blooming" a light fog. Also there was no marking system to distinguish a runway from its surroundings.

The large-scale operations in England during World War II sometimes ended disastrously when aircraft returning to base were faced with impossible conditions of visibility. Extensive losses of aircraft and crews resulted and drastic action was taken in the form of FIDO (see paragraph 2.1). Huge quantities of raw gasoline were poured into trenches bordering runways and then ignited. The intense heat burned off enough fog for approaching aircraft to make a visual landing after a low approach on an electronic system. Although FIDO served its purpose, it was obviously an uneconomical system for routine operations. Although later attempts were made to improve operation, the system was abandoned. Other methods of fog removal such as seeding and ionization have been the subject of extensive experimentation, but with only spotty success. Such efforts are still going on, and the idea of eliminating the fog rather than trying to punch through it has not yet been written off completely.

The first steps toward improved lighting for poor visibility operation were to increase the power of runway edge lighting and to beam the lights toward the approaching aircraft for maximum penetration of fog and haze. It was realized that lights were required for day as well as night use and that daytime lighting against a bright fog background needed massive power to be effective.

With improved runway lighting came a realization that a visual gap existed between the end of the approach on the landing guidance system and the runway edge lighting. The answer was a system of approach lights, initially installed on an extension of the left edge of the runway, and later installed on the centerline back from the runway threshold almost to the middle marker.

With this greatly improved situation, operations were possible to much lower minimums, but pilots became acutely aware of coming to the end of a very effective system of approach lights and being faced with a "black hole" at the runway. The edge lights, though powerful, were too far from the centerline to be effective. Lighting engineers responded to the problem with "in runway" lighting in which powerful light sources are installed directly in the runway pavement almost flush with the surface. These provide centerline lighting for the entire length of the runway. In addition, lights are installed in a pattern to designate the touchdown zone in order to provide the pilot with visual clues required to make the critical flare and touchdown maneuvers.

In addition to the extensive improvements made in lighting, a parallel program has resulted in much more precise visibility measurement systems, consisting of ceilometers for measuring cloud height and transmissometers for measuring visual range. Still in the future is a system to measure slant range visibility so that the pilot can be given an exact report on what he can expect to see when he reaches decision height. Runway marking systems have gone through a period of development similar to that of lighting. The importance of good marking can be realized when one considers that under certain conditions of bright daytime fog, high-contrast markings (see paragraph 4.5.3) are more effective than even high-intensity lights.

4.5.2 PRESENT STANDARD LIGHTING SYSTEMS

4.5.2.1 Approach Lighting. Instrument approach lighting systems permit the transition from instrument to visual approach and bring the pilot within range of the runway lighting system. The configuration used in the U. S., shown in Figure 4.11, conforms to ICAO standards [30] [31]. The U. S. standard includes a series of short-duration high-intensity condenser-discharge lights collocated with steadily burning lights. These discharge lights are flashed sequentially, giving the appearance of a "ball of fire" racing along the approach. They run the full length of the approach-light lane twice in one second. The sequence flashing lights, even in conditions of poor visibility, are effective in giving the pilot advance notice that he is nearing or over the approach lights. The steadily burning lights are horizontal bars 14 feet in length which contain five sealed-beam high-intensity lamps aimed toward the approaching aircraft.

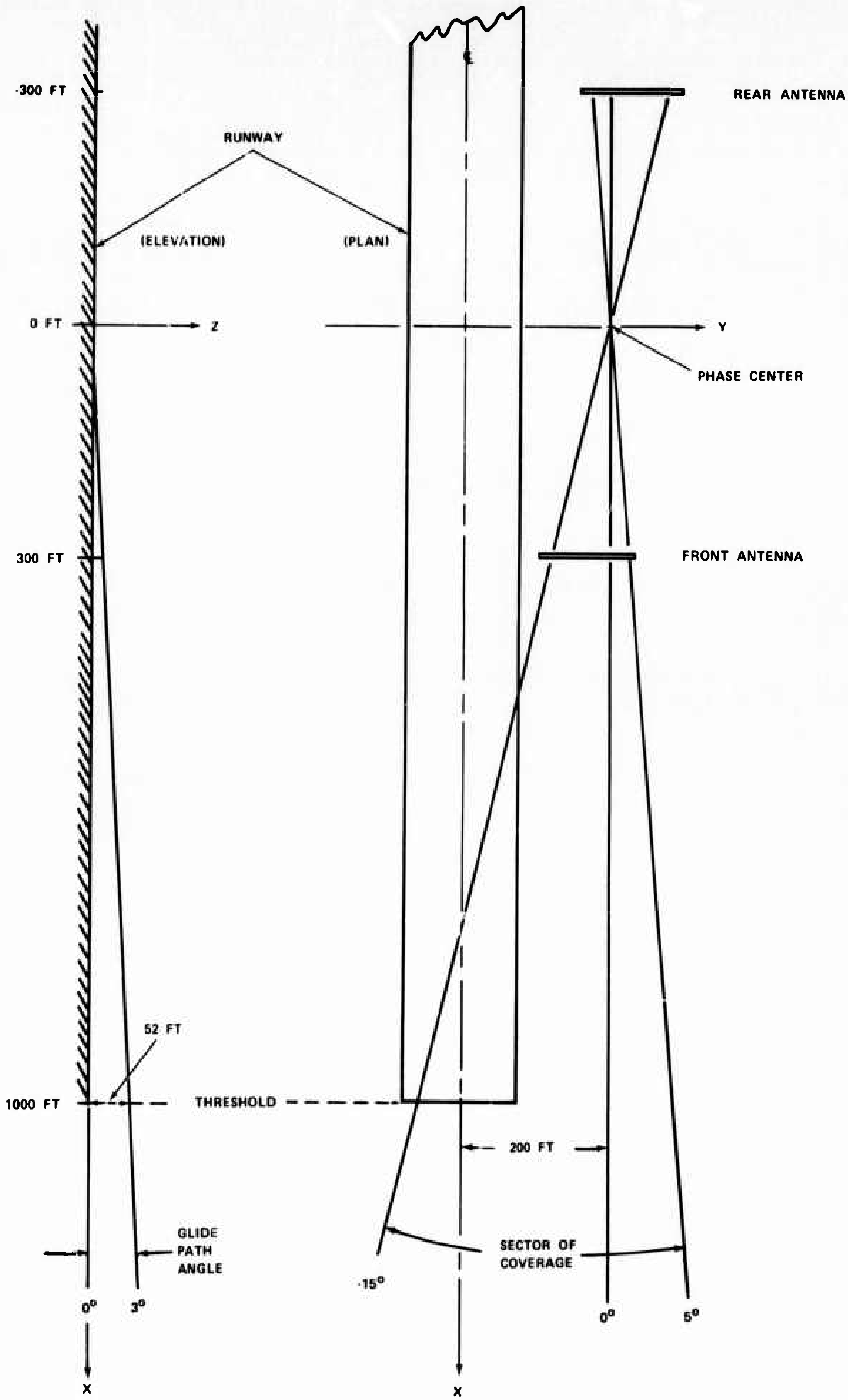


Figure 4.9 Two-Antenna End-Fire Glide-Slope Location With Respect to Runway

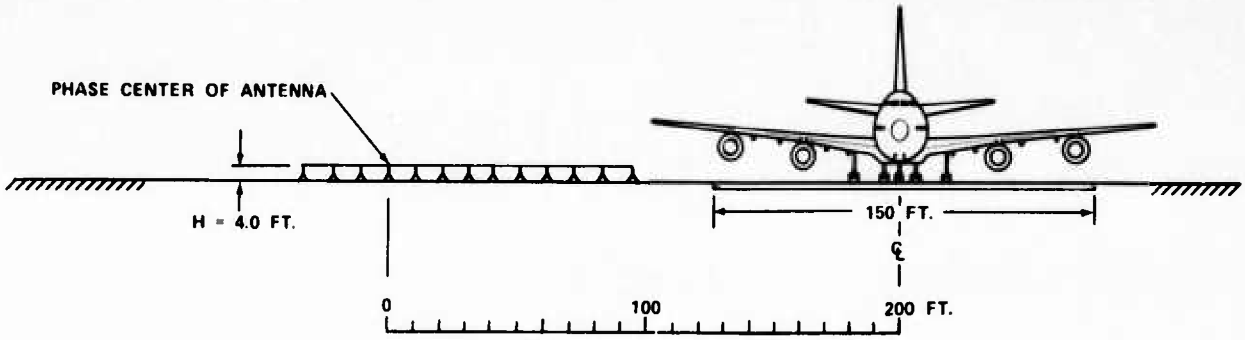


Figure 4.10 Boeing 747 Opposite Front Watts Prototype Glide-Slope Antenna

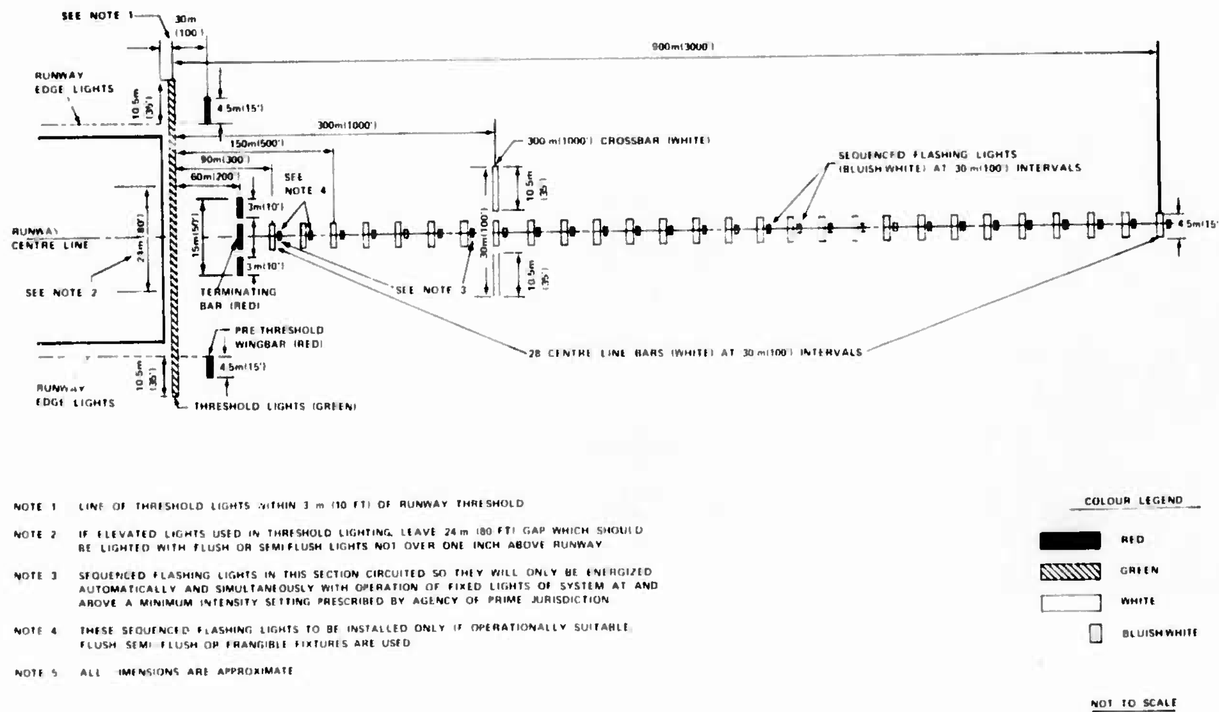


Figure 4.11 U.S. National Standard Approach Lighting Configuration (From ICAO Annex 14)

For CAT II operations, red barrettes have been added to supplement the centerline lights in the 1,000 feet preceding the threshold. These are parallel to the centerline as shown in Figure 4.12.

4.5.2.2 Runway Lighting. Runway lighting includes the edge lighting, discussed in paragraph 4.5.1 and used primarily to define the edges of the runway. Maximum longitudinal spacing between runway edge lights is 200 feet.

In-runway lighting has been added to many runways to make them eligible for CAT II operations. The complete, standard ICAO system for CAT II operation is shown in figure 4.12, including touchdown zone lighting, which generally extends 3,000 feet from the threshold, whereas centerline lighting continues the full length of the runway. The centerline lights are color coded, with white lights extending from threshold to within 3,000 feet of the far end, and then alternate red and white lights extend to within 1,000 feet of the runway end. The final 1,000 feet of runway are coded red.

4.5.2.3 Visual Approach Slope Indicator. The introduction of jet aircraft into commercial operation increased pilot difficulty in maintaining a proper glide slope during visual approaches and during the visual completion of instrument approaches. ICAO has standardized on the Visual Approach Slope Indicator (VASI), a system of lights installed on each side of the runway that provides color-coded indications of the position of the aircraft with respect to the proper glide-slope angle. The principle of the basic light unit used for VASI is shown schematically in Figure 4.13. Installation includes 2, 4 or 12 light units depending on the needs of the runway.

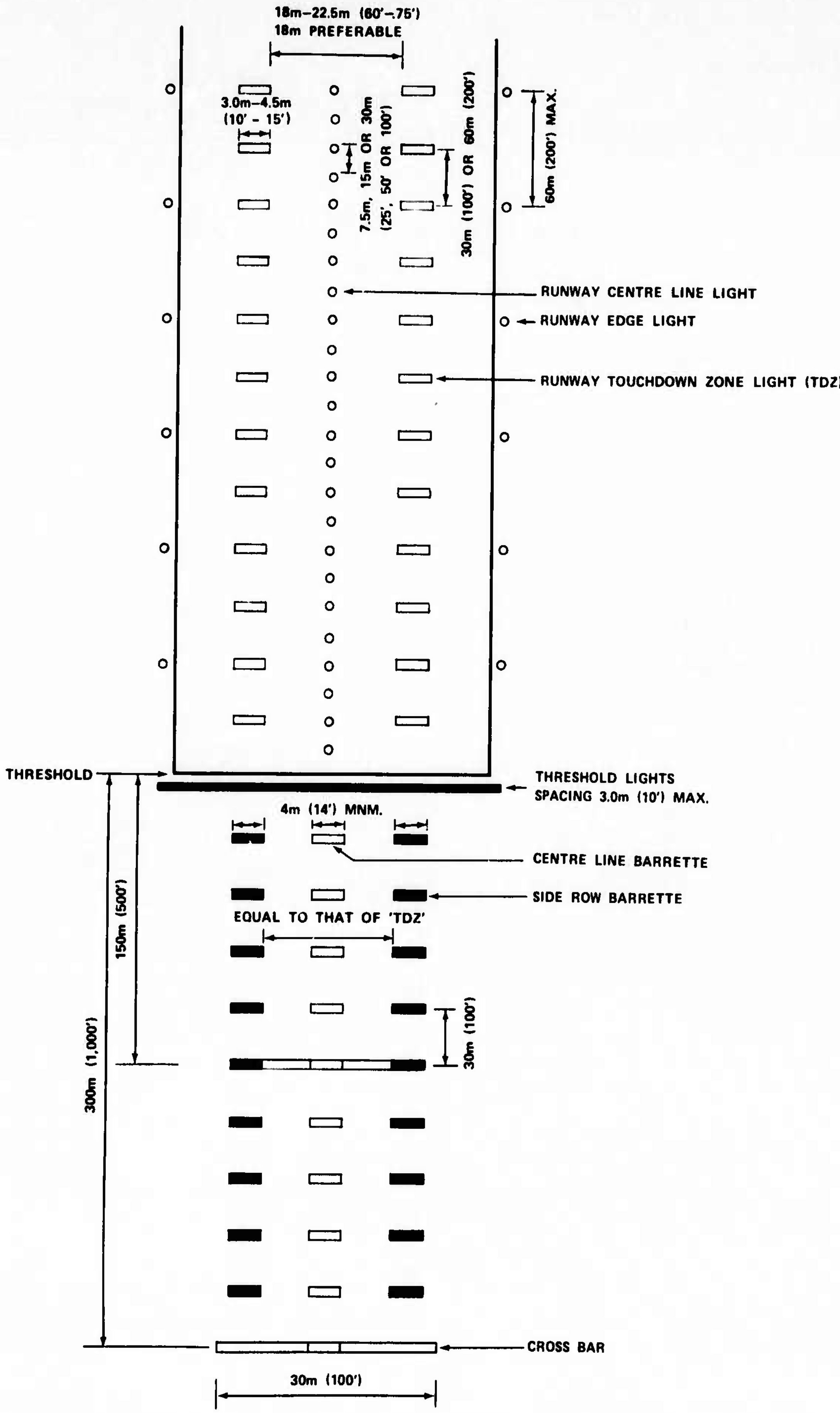


Figure 4.12 CAT II Configuration of Inner 1,000 Feet of Approach Lights and Runway Lights

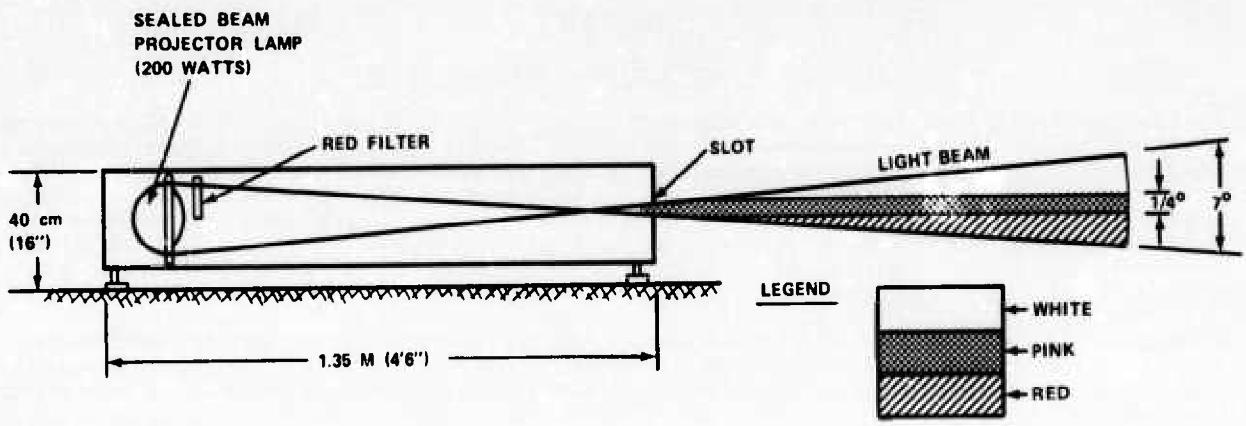


Figure 4.13 Visual Approach Slope Indicator (VASI)--Optical System

4.5.3 RUNWAY MARKING. Runway marking systems provide a number of important functions for landing guidance, including:

- Contrast so that the runway can be more clearly seen with poor visibility
- Identification of the magnetic heading of the runway
- Definition of centerline and edges of the runway
- Identification of the touchdown and rollout zones of the runway
- Taxiway turnoff guidance

The pattern (usually painted) of the standard ICAO runway markings and the dimensions used are shown in Figure 4.14.

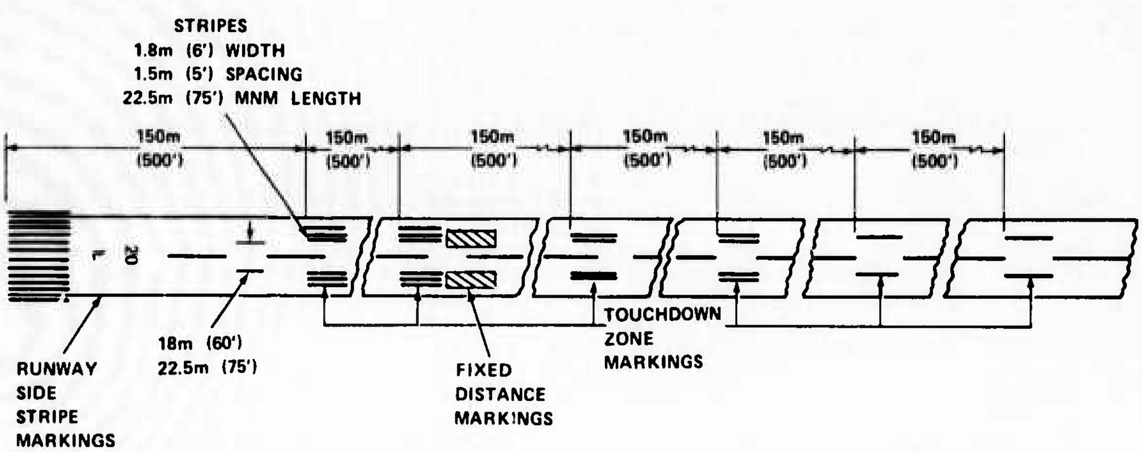


Figure 4.14 Standard Runway Marking Configuration (from ICAO Annex 14)

SECTION 5

MICROWAVE LANDING SYSTEMS CURRENTLY IN USE

The limitations of ILS, particularly with regard to siting, portability, and antenna size, have forced the development of a number of microwave landing systems (MLS) aimed at meeting specific, limited operational requirements [32]. A number of these systems have reached operational status and have been produced in sizeable quantities. Unfortunately, these systems are not compatible with each other, and an aircraft equipped with an airborne set is restricted to using companion ground equipment only. The proliferation of interim systems and the potential problems that certainly would develop as a result of noncompatible frequencies, modulation techniques, and signal formats stimulated the current worldwide effort to agree on a single standard MLS. This effort is described in Sections 7 and 8.

Although about 40 different microwave landing systems have been developed to meet various requirements, only a relative few have actually reached operational status and have been produced in significant quantities. These may be placed in three major classifications: aircraft carrier systems, military tactical systems, and civil systems.

5.1 AIRCRAFT CARRIER LANDING SYSTEM

There was early recognition that the ILS could not meet the stringent requirements of carrier deck landings. Sheer size of antennas required for VHF and UHF operation ruled out the possibility of siting on carrier decks. Even if space could be found, the problem of stabilizing such large arrays to compensate for deck motion is virtually insurmountable. Even assuming these problems could be solved, the limited area of a carrier deck still is not adequate to form the large ground plane needed to produce the lobe structure required for the ILS glide slope. For these reasons microwave landing systems for carrier deck use were developed early by the U.S. Navy and have been used for many years.

5.1.1 EARLY CARRIER SYSTEMS. The first operational U.S. systems, AN/SPN-8 and AN/SPN-12, were based on the GCA concept, but used stabilized active tracking radars to measure azimuth and elevation angles as well as distance. This ground-derived information was presented to a ground controller, who gave steering signals by voice link to the pilot as in a GCA approach. These systems depended on visual completion of the landing aided by a mirror optical system that gave the pilot a glide-path reference to the deck. The mirror system was later replaced by a more precise Fresnel lens system stabilized to compensate for deck motion [33].

5.1.2 ADVANCED SYSTEMS. The voice-link systems were superseded by the AN/SPN-10 and later the AN/SPN-42, which, in addition to having the voice capability, transmitted azimuth and elevation steering signals to the aircraft by UHF radio. The pilot's display of this information was in the form of the conventional crosspointer indicator, and the information could also be used to steer the aircraft automatically. Initially, the radar return signals were reinforced by passive corner reflectors, but because of short range and severe rain attenuation, transponders were added to the aircraft. Since a radar could track only one aircraft at a time, two radars were used to increase the acceptance rate [34].

5.1.3 AIR-DERIVED BACKUP SYSTEM. Responding to the need for increased range and capacity, the Navy-sponsored development of an air-derived microwave scanning beam system, the AN/SPN-41 [34]. In addition to providing much greater range and an unlimited capacity, this system was installed as an added facility rather than a replacement, and as such provided a totally redundant capability (see Figure 5.1). Agreement between the radar tracking system and the scanning beam system does not continue all the way to the deck; the scanning beam system is ignored when the aircraft is within 12 seconds of touchdown. Because a carrier deck is subject to the complex motions of roll, pitch, heading, yaw, and heave, carrier landing systems must compensate for each of these variables and transmit computer-corrected information to the landing aircraft. In addition, the computer stores aircraft performance parameters and provides optimum approach guidance for each aircraft type operating from the carrier.

5.2 INTERIM TACTICAL MLS PROGRAMS

The ILS has inherent limitations that inhibit or actually prevent its effective use as a tactical landing aid. The size and weight of the antenna arrays make it difficult if not impossible to design a configuration that has the needed portability and which can be quickly and easily installed. To obtain high-quality courses, extensive site preparation is often required. This limitation is simply unacceptable in a tactical environment.

Because of such problems military services of several countries have made substantial investments in various microwave systems in advance despite full knowledge that these systems will become obsolete when a standard MLS becomes a reality. Brief descriptions of these interim systems follow.

5.2.1 MADGE. The MADGE System was designed to meet the tactical requirements of the United Kingdom's Royal Air Force, and quantities are being produced to meet this need. Angle measurements, both azimuth and elevation, are ground derived using interferometric techniques. Distance is measured in the aircraft based on the round-trip travel of the signal between air and ground. The system operates at C-band.

Equipment in the aircraft includes a receiver-decoder. Digital address codes are used to identify both the aircraft and the ground system being interrogated by the aircraft. When an aircraft interrogates a ground station, the interferometer measures the angle during the period of the beacon pulse. This angle data is stored until called up by the data link.

The range-measuring system controls a range tracking gate in the aircraft receiver. The MADGE system accuracy is dependent on the spacing between two antennas. The ambiguities that result from wide spacing are resolved by reference to the phases of signals generated from pairs of decreased spacing signal sources.

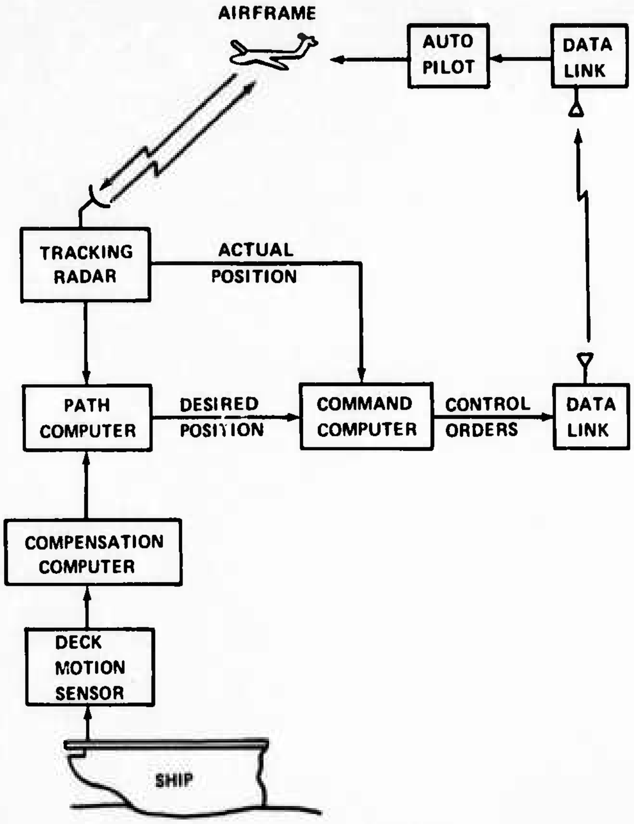


Figure 5.1 Automatic Carrier Landing System Block Diagram

5.2.2 TALAR. TALAR is a highly portable tactical system adopted by the U.S. Air Force for use by C-130 aircraft during the height of the conflict in Southeast Asia. Systems are still operated and maintained by the U.S. Tactical Air Command.

Operating at Ku-band, TALAR has both glide slope and localizer antenna arrays in a single case installed on a quick-setup mounting. The two antenna arrays are excited by a single magnetron (later a traveling wave tube) on a time-shared basis. TALAR guidance signals are formed by the intersection of two overlapping beams, each modulated at a different audio frequency and radiated from slotted wave-guide antennas.

The airborne system consists of a horn antenna feeding a tunnel diode detector and amplifier. A separate box houses the signal processing unit. The system operates on a single channel and uses conventional cockpit instruments and autopilot couplers already installed in the aircraft.

5.2.3 MARINE REMOTE AREA APPROACH AND LANDING SYSTEM. The U.S. Marine Corps has completed a development program for a tactical microwave landing system and is currently in the final stages of selecting hardware from two competing firms. The Marine Remote Approach and Landing System (MRAALS) is a Ku-band mechanical scanning beam system compatible with currently used U.S. Navy carrier airborne landing systems. The ground equipment, which includes a localizer, glide slope and DME, are all contained in a single small container that weighs less than 100 pounds. The system is rugged enough for tactical handling, has built-in supports, and can be set up and aligned in less than 10 minutes.

5.2.4 SYDAC. SYDAC is an ILS-compatible system operating at C-band and using equisignal beam forming techniques for both localizer and glide slope. Audio modulation frequencies correspond to the 90 and 150 Hz used in conventional ILS. A localizer signal is generated by a solid-state RF source and multiplier chain having an output frequency of 5003 MHz. A similar signal source produces a glide-slope frequency of 5225 MHz. The microwave converter in the aircraft has a high-stability oscillator operating at 4893 MHz, with the different frequencies corrected for input to localizer and glide-slope receivers.

5.3 CIVIL INTERIM MLS PROGRAMS

The pressure to overcome the deficiencies of ILS at civil airports with difficult siting problems has resulted in the development of a number of interim microwave landing systems as proposed solutions. Some of these have reached operational status in spite of strong opposition from organizations concerned about the chaotic situation that might develop if a number of noncompatible systems were allowed to be placed in service. It is argued by opponents that an interim system might gain enough support to forestall the implementation of a universal MLS program and yet lacks the features needed for the long-term future. Nevertheless, the interim systems briefly described in the following paragraphs have reached some degree of formal acceptance and are currently in limited operational use.

5.3.1 TULL AVIATION SYSTEM. The Tull Aviation Microwave ILS consists of a localizer and glide slope installed at approximately the same locations as conventional ILS. A step-scan technique is used for both localizer and glide slope, and audio modulation frequencies of 90 and 150 Hz are used for compatibility with the ILS airborne equipment. Three frequencies are transmitted to the approaching aircraft. When a frequency of 5005.2 MHz is received by the airborne microwave converter, an oscillator locks to a frequency of 4892.2 MHz. When additional frequencies of 5000.3 MHz and 5226.9 MHz are received, they are beat with the oscillator frequency to produce 108.1 MHz for the localizer receiver and 334.7 MHz for the glide-slope receiver. These signals are fed through slightly modified receivers to produce conventional instrument indications. Frequency selection is made by using the conventional ILS selector, giving a capacity of 20 channels. By changing the reference frequency, a total of 100 localizer channels and 95 glide-slope channels can be produced. The Tull System has recently been officially designated as the U.S. Interim Microwave Landing System, and although no government-sponsored installation program is planned, the system is eligible for Airport Development Aid Program funding [35].

5.3.2 TALAR IV C. The TALAR IV C system is a modified civil version of the tactical system described in paragraph 5.2.2. Improvements include 10 channels as compared to the single channel of the tactical system. A full monitoring system is included and a long-life traveling wave tube is used in lieu of the magnetron. Weather protection, including a rain shield and deicing for both transmitter and monitor, is part of the system. The system has been operating in a limited number of locations since 1971 and providing substantially lowered minimums at sites considered very difficult for ILS.

5.3.3 SETAC (German). The SETAC system is an adaptation of TACAN, the military Rho-Theta navigation system. Operating at L-band (1 GHz), it compresses the full 360 degrees of the navigation system to a 36-degree sector for precise localizer guidance. The basic system, SETAC-A, does not include elevation guidance but depends on improved accuracy DME to provide the basis for a computed glide slope [36].

SETAC-A produces its sector guidance by use of a rotating hyperbolic field generated electronically. Two signals having the same frequency, phase, and amplitude are emitted from sources spaced one wavelength apart, resulting in a four-lobed pattern. Change in phase of the two signals is caused by a drop in the frequency, which is repeated 15 times per second. The result is a rotating pattern that, viewed from a distance, appears to have an amplitude modulation of 15 Hz and a bearing-dependent phase, as in TACAN. By proper selection of the antenna spacing, the interval between zero points can be made to be 36 degrees. This is a tenfold increase in resolution over TACAN, where the spacing between zero points requires a full 360 degrees.

A Doppler glide slope, known as SETAC-E, is available that uses a separate transmitter at the conventional glide-slope location. The glide-slope signals are multiplexed with the SETAC-A signals, making use of the "dead time" of 60 to 70 ms following the transmission of each double pulse. An additional module is used in the aircraft to process the elevation signal. The advantage claimed is the use of existing airborne equipment already installed in military fleets requiring only modest alteration to provide full landing guidance capability.

SECTION 6

ADVANTAGE OF MICROWAVES FOR FUTURE WORLD STANDARD SYSTEMS

6.1 MICROWAVE SYSTEMS CAN MEET FUTURE REQUIREMENTS

In spite of the tremendous investment already made in both airborne and ground ILS equipment, and fully recognizing the extensive operational and technical know-how that has developed in its use, the limitations of the system, discussed in paragraph 4.3, are such that strong pressures are being brought to bear to replace ILS with a MLS. None of the interim systems are capable of meeting present and future requirements of all users for a worldwide standard system. Delay in achieving agreement on a universal standard system may lead to increasing investment in various nonstandard microwave landing systems which might force users to carry several sets of airborne equipment when crossing national boundaries or when operating into both civil and military airfields.

To avoid such chaos, effort was begun to develop a signal format capable of meeting the requirements of all users. The format would have to have enough flexibility to allow simple equipment, both airborne and ground, to be used for less stringent requirements and yet at the same time permit more complex but still fully compatible versions to be used to meet the most critical requirements of those willing to pay for the increased sophistication.

To meet these requirements, it was recognized that microwaves offered substantial advantages over the VHF-UHF frequencies used for ILS, particularly in the areas discussed in the following paragraphs.

6.1.1 RELATIVE FREEDOM FROM ADVERSE SITING EFFECTS. The use of microwaves provides relative freedom from adverse siting effects that have plagued the ILS from its earliest development and which continue to resist the best efforts of research and development. Although substantial ILS progress has been made, it is generally recognized that basic physical limitations are being approached, whereas the better control of beam shapes made possible by the smaller antenna apertures at microwave frequencies give designers a powerful tool for relieving this problem.

6.1.2 OPERATIONAL FLEXIBILITY. Microwave landing systems offer the possibility of greatly increased operational flexibility. Flight operations using ILS are limited to a single straight-line path normally projected along the extended centerline of the runway (or in some cases, slightly offset a few degrees to meet local siting restrictions) and along a single straight-line glide slope approaching the runway at a fixed angle. The use of microwaves gives the terminal area designer a new set of tools to increase runway and airport capacity, ameliorate noise pollution, and avoid traffic interference between adjacent airports.

6.1.2.1 Curved, Segmented, And Multiple Paths. The ability of microwave systems to provide aircraft with position data required for curved or segmented paths in both the horizontal and vertical planes will allow precision guidance around obstructions and noise-sensitive communities. Segmented glide-slope approaches will permit higher initial approaches, with an attendant reduction in noise level beneath the approaching aircraft. Parallel runways can be served by simultaneous landings of aircraft utilizing accurate lateral and vertical separation for curved approaches with a high degree of assurance. Cockpit selection of glide-slope angle is possible with MLS and may be useful in optimizing the approach angle for each specific aircraft, or in providing guidance for maximum-angle noise-abatement procedures.

6.1.2.2 Flare Guidance. Although the ILS was originally designed to provide landing guidance to touchdown, it was able to achieve this only because of the aircraft performance characteristics of that time. An aircraft with a 70-knot approach speed flying a 2-degree glide slope (the angle used in early experiments) produced a rate of descent acceptable for touchdown. This proved to be a temporary situation, however, and both aircraft approach speeds and glide-slope angles have increased significantly so that maintaining full rate of descent to touchdown today most likely would end in disaster. Rather than attempt to provide beam guidance to touchdown, landing system designers have resorted to other methods, usually involving radio altimeters, accelerometers and other aids to program a reduced rate of descent. The glide slope normally is not used below 50 feet. The use of radio altimeters for initiating flare leaves much to be desired because of lack of uniformity of terrain profiles just ahead of runway thresholds from airport to airport. The use of microwaves will allow the extension of precise ground-based landing guidance almost all the way to touchdown, uniformly at each airport. In most proposed systems, flare guidance requires a special transmitter used in conjunction with range information from a DME to compute the flare path. It is likely that flare guidance will be an optional feature.

6.1.3 AIRCRAFT CARRIER OPERATION. The advantages of microwave landing systems on aircraft carrier decks has been fully proven by years of operational experience. It is likely that present carrier microwave landing systems will be replaced by the universal MLS to provide carrier-based aircraft with the capability of operating compatibly with ground-based military systems and at joint civil-military airports without having to carry separate airborne equipment.

6.1.4 PORTABILITY AND TACTICAL SUITABILITY OF MICROWAVE SYSTEMS. As in the case of aircraft carriers, ground-based interim microwave landing systems have been successfully used in tactical operations and have shown their superiority over the ILS for this application. The lightweight, small-size and low-power requirements and the quick set-up time have given military tacticians new flexibility in planning forward area air operations.

6.1.5 FREQUENCY SPECTRUM AVAILABILITY. The problem of limited frequency spectrum availability was discussed in paragraph 4.3.2. The 20 channels presently in use at VHF and UHF are inadequate to meet the requirements for expanded use of the ILS. Even the planned increase to 40 channels will not meet the long-range requirements for landing system implementation. And this increase in number of channels is being accomplished only by cutting the spacing between channels in half. There is almost no chance of increasing VHF or UHF spectrum allocations for landing guidance systems because of the heavily vested interest in facilities adjacent to the ILS bands. This is a powerful argument in favor of using microwaves, where portions of the frequency spectrum, although by no means unlimited, are much more available than at the ILS frequencies. Current estimates are that 200 channels may be needed to meet long-term requirements, and microwave systems are being developed with that capability.

6.1.6 LESS SUSCEPTIBILITY TO AIRCRAFT INTERFERENCE. Microwave systems have proven to be less susceptible than conventional ILS to the effects of multipath caused by aircraft reflections. Such interference often occurs when one aircraft is following another during an ILS approach. The following aircraft may receive erratic indications as the first aircraft comes between it and the transmitter. More severe interference occurs when an aircraft taking off flies low over the localizer transmitter. These problems are greatly reduced at microwave frequencies.

6.1.7 POTENTIALLY LOWER COST OF MICROWAVE SYSTEMS. Although a complete CAT III MLS offering a full range of services consisting of curved approach guidance through wide azimuth angles, precision DME, guidance for segmented glide slope, flare guidance to touchdown, rollout, and missed approach guidance is likely to cost more than the most advanced ILS giving comparable if not equivalent service, the total implementation of a MLS program should cost less than continued implementation of ILS because:

- Site preparation costs will be virtually eliminated.

- MLS can be tailored to the required degree of sophistication.

- The installation costs of MLS are less because of smaller equipment size.

- Use of DME may eliminate real-estate costs of marker beacon sites outside of airport boundaries.

- Collocation of MLS azimuth and elevation transmitters or even time sharing a single transmitter, is feasible for the least critical MLS version.

SECTION 7

MICROWAVE LANDING SYSTEM DESIGN CONSIDERATIONS

A microwave landing systems (MLS) is defined, for the purposes of this section, as any low-approach and landing systems that operates in the microwave region of the frequency spectrum. Prerequisites to the MLS equipment design effort are the selection of a carrier frequency and a signal format with sufficient information content to accommodate operational requirements. Each of these two factors are discussed in this section, along with a description of the basic techniques for generating the necessary signals-in-space for landing guidance. Emphasis is directed towards air-derived data systems since this category has been the mainstay of commercial aviation for many years. This section draws heavily on a draft paper prepared for the FAA [37].

7.1 SELECTION OF CARRIER FREQUENCY

The selection of an operational carrier frequency is a fundamental decision that must be made before embarking on the design of a new MLS. Factors that must be considered in this selection process are discussed in paragraphs 7.1.1 through 7.1.3.

7.1.1 FREQUENCY AUTHORIZATION. Both national and international regulations dominate technical considerations in the selection of a carrier frequency. Frequency allocations other than those already established for this type of service would require tedious administrative efforts that could last for several years before approval of new frequencies was obtained, if at all. Current frequency allocations permit the implementation of microwave landing systems only in specified regions of the C-, K-, Ku-, and Ka-bands, depending upon equipment function, plus L-band for the DME component.

7.1.2 PRECIPITATION ATTENUATION. Rain greatly attenuates RF signals at higher frequencies, resulting in a considerable loss of signal above C-band. For example, in a very heavy rain of 2 inches per hour, the attenuation of a Ku-band signal over a 10-kilometer path is about 30-dB higher than the attenuation of a C-band signal over the same path. Equivalent performance at the two frequencies would require not only a Ku-band signal power increase over a C-band signal by a factor of 10^3 to overcome attenuation, but also an additional factor of 9 to compensate for the larger energy collection area required of a C-band antenna. Values of attenuation caused by various rainfall rates are difficult to determine with reasonable accuracy, since the attenuation varies considerably with raindrop particle sizes and distributions. Furthermore, the exact values of attenuation are difficult to measure because the rain rate over a measurement path normally varies substantially along the path.

An interesting presentation of the average theoretical and empirical precipitation effects is illustrated by Figure 7.1, where the range in clear weather, R_c, is related to the range in rain, R_r, for various rain rates at Ku-band. For example, a Ku-band MLS having a clear-weather range of 90 miles will be reduced to a range of 5 miles at a rain rate of 2 inches per hour.

7.1.3 HARDWARE IMPACT. The size and weight of the ground equipment of a landing guidance system are affected by the selection of carrier frequency. The antenna aperture size, which normally determines the largest equipment dimension, is inversely proportional to the carrier frequency for a specified signal beamwidth. For example, a C-band aperture is about three times larger than a Ku-band aperture for equivalent signal beamwidth. However, the weight of the equipment may increase with carrier frequency, since lightweight, solid-state RF power sources are more readily available at the lower frequencies, and these solid-state sources require less primary power and thus a lighter power supply.

7.2 SIGNAL FORMAT CONSIDERATIONS

Establishing a signal format that is capable of meeting the operational requirements of the MLS is a prerequisite to hardware design, as illustrated by Figure 7.2. The signals in space generated by the MLS represent the signal format, and contain spatial, temporal, and spectral ingredients.

7.2.1 SPATIAL INGREDIENTS. Spatial features that must be considered in the design of an MLS consist of: (a) coverage and guidance regions, with the coverage region being the spatial domain in which the MLS signal provides the necessary information for steering the aircraft into the guidance region, and the guidance region being the spatial domain in which the MLS signal provides the necessary steering information to maintain an on-course flight trajectory; (b) beam patterns, which define the shape and dimensions of the radiation envelopes emitted by the transmitters; (c) polarization, which defines the orientation of the electric vector of the radiation; and (d) siting constraints, which provide guidelines for locating the ground transmitters, including (1) the availability of real estate, (2) the location of objects that may shadow, reflect, or reradiate the signals, (3) near-field radiation regions in the vicinity of the ground transmitters, in which the signal is poorly defined, and (4) special siting constraints unique to the selected technique. An example of the latter is that the elevation scanning beam for the flare maneuver must be small enough to fit entirely between the receiver antenna on the aircraft and its ground image at touchdown, since this requirement controls the maximum distance that the transmitter can be located from the touchdown point.

7.2.2 TEMPORAL INGREDIENTS. The temporal ingredients are the time-dependent features of the signal format. These include: (a) duty cycle, which is that percent of the time during which a particular signal is available at the aircraft receiver for processing; (b) data rate, which is the number of times per second that a particular signal is available at the aircraft receiver for processing (it is necessary to distinguish data rate from data update rate, the latter being the number of times per second that an effectively processed signal is available for display or autopilot control); (c) timing cycle, which is the temporal description of one complete sequence of signal events; and (d) temporal features that are unique to a particular technique, such as the dwell time of a scanning fan-beam signal, which is the time interval during which the particular scanning beam illuminates the aircraft antenna.

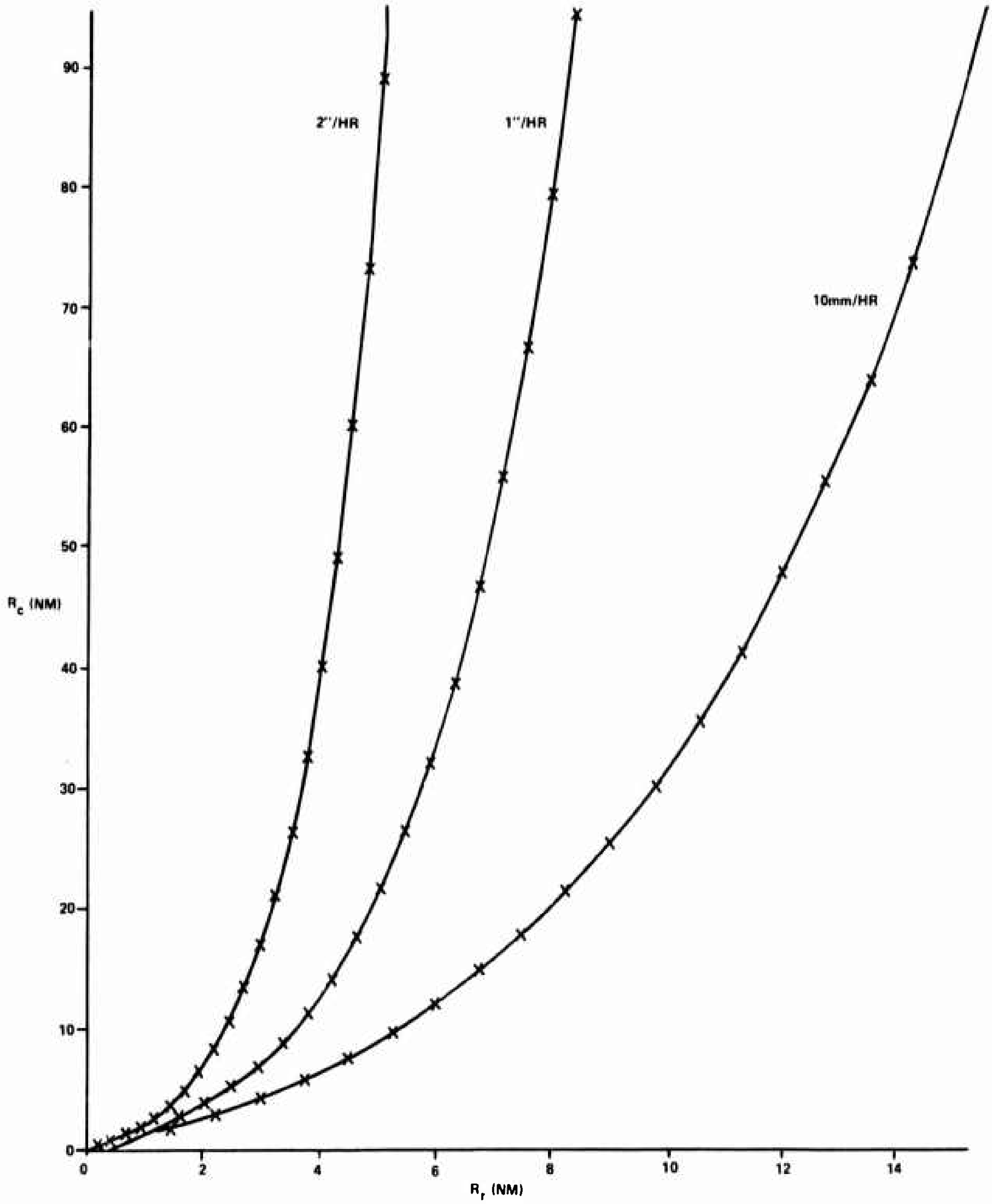


Figure 7.1 Precipitation Effects on MLS Performance at Ku-band

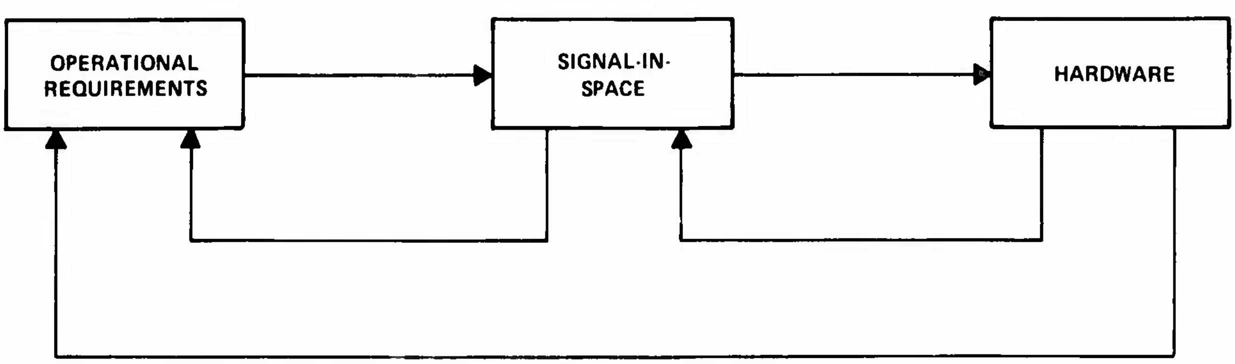


Figure 7.2 Signal Format in Context

7.2.3 SPECTRAL INGREDIENTS. The spectral ingredients are those features that define the characteristics of the transmitted signals. These include: (a) channelization, which refers to the specific regions within the electromagnetic spectrum that will be occupied by the MLS signals; (b) data code, which refers to the modulations that define the spatial orientation of the signals and identify the MLS facility and its conditions; and (c) spectral contaminants, such as noise, electronic countermeasure (ECM) interference, and other factors that degrade the quality of the MLS signals.

7.3 AIR-DERIVED ANGLE DATA

Many of the numerous techniques that can be employed for landing guidance systems at nonmicrowave frequencies also can be applied to microwave frequencies. The technical feasibility of three of these techniques for air-derived angle data systems have been demonstrated by experimental and operational hardware, i.e., fixed beam, scanning beam, and Doppler systems. The salient features associated with each of these systems including the design parameters that must be considered are discussed in this section.

7.3.1 CHARACTERISTICS OF FIXED FAN-BEAM MLS. Fixed fan beams for air-derived data microwave landing systems are similar in concept to the VHF/UHF ILS described in Section 4. This technique provides only one fixed localizer course and one fixed glide-slope angle for the low-approach phase of the landing, and normally does not provide sufficient information content for flare and touchdown. While having the limitations previously discussed, the fixed fan-beam system possesses several advantages compared to the more sophisticated systems, as discussed in the following paragraphs.

7.3.1.1 Compatibility With Conventional ILS. Fixed fan-beam systems have been developed that make use of ILS modulation frequencies and need only an antenna and converter to receive signals for processing through existing ILS airborne equipment already installed. This feature is economically attractive, but technically limited. Several such systems are described in paragraph 5.3. Operationally, the instrument approach procedures for fixed fan-beam equipment are similar to those now used with the standard ICAO ILS and thus require little pilot familiarization. However, the universal MLS will also have many compatible features.

7.3.1.2 Compatibility With Universal MLS. The avionics required to process the modulations of a fixed fan-beam system are capable of being designed also to process any future universal MLS signals, provided the universal MLS technique is known in advance. The feasibility of such an interim scheme thus depends on the final selection of MLS technique and also on whether the complications of processing are worth the effort.

7.3.1.3 Equipment Cost. The equipment cost of the fixed fan-beam systems for both ground station and airborne station is generally considered much less than for the other microwave landing systems described in this section. This is because the ground station requires neither motors to rotate antennas nor sophisticated phase arrays and angle code generators. Also, the airborne station is simpler, since it only requires circuitry to measure the relative amplitudes of the overlapping beams and does not require devices such as tracking gates.

7.3.1.4 Reliability. The simplicity and advanced development and operational experience of fixed fan-beam systems guarantee their present advantage in reliability. However, the question of reliability is one which must be answered in terms of state-of-the-art of the techniques selected. For example, a phased array with its large number of components may now be less reliable than a fixed fan beam or a mechanically scanned beam, but improved component designs may drastically change this in the future.

7.3.1.5 Duty Cycle. The fixed fan-beam systems have a duty cycle potential of up to 50 percent for the localizer (azimuth) information and 50 percent for the glide-slope (elevation) information. In contrast, the duty cycle for a scanning fan beam is well below 10 percent, since the aircraft is illuminated by the beam during only a small segment of the scan. The effective duty cycle for the Doppler systems is comparable to that for the fixed fan beam.

7.3.1.6 Data Rates. The fixed fan-beam data rate is limited only by the beam modulation frequencies used and by the amount of smoothing required, whereas the scanning fan-beam data and data renewal rates are related to the more limited number of times per second that the beam illuminates the aircraft receiver antenna.

7.3.1.7 Data Link Capacity. In addition to angle guidance and possible DME information, provision must be made for adequate data link capacity, including such items as runway identification and weather conditions on the landing runway. These requirements are all compatible with fixed fan-beam systems.

7.3.2 FIXED FAN-BEAM DESIGN PARAMETERS. There are several design parameters that must be considered in the design of even the simplest MLS, including fixed fan-beam systems. These must be examined in a quantitative manner and values assigned to each.

7.3.2.1 Carrier Frequency. Practical considerations, in particular the available allocated frequencies, constrain the selection of carrier frequency to either C-, Ku-, or Ka-bands. The criteria for the selection among these frequencies is discussed in paragraph 7.1.

7.3.2.2 Channeling. It is likely that the MLS will not be operating in a space diversity environment, particularly in tactical situations and at high-density airports with adjacent runways. Thus signals from more than one MLS ground facility can illuminate the aircraft, and it is desirable to separate the signals on a frequency diversity basis. The frequencies of the channels used should be spaced closely enough to avoid the waste of spectrum, but separated enough to avoid cross talk between channels.

7.3.2.3 Radiated Power. In determining the values for the radiated power design parameter, both peak power and average power must be considered, since these are related to the portion of the total time that useful guidance signals are transmitted. The radiated power requirement also is functionally related to the operational requirements of the system: higher power is required for greater coverage, higher data rates, or higher signal-to-noise ratios. The radiated power requirement has a significant impact on the hardware design. For instance, when radiated power greater than a few watts is needed, vacuum tube technology is currently required for implementation. Only when less power is required can the more efficient solid-state RF power sources be considered.

7.3.2.4 Coverage Region. The coverage region is that volume of space in which an MLS signal provides sufficient information content to direct an aircraft therein towards the guidance region. The most critical coverage region probably is that volume of space below the selected flight path, in which the processed MLS signal generates a steering signal that directs the aircraft to "fly up."

7.3.2.5 Guidance Region. The guidance region should be large enough to allow all of the normal maneuvering required by an aircraft during approach and landing. It should be small enough on the selected flight path such that the scale of indications will permit the most precise determination of deviation from course centerline.

7.3.2.6 Split-Site and Collocated Operation. Optimum siting of a landing system from an operational point of view normally requires a split-site configuration, with the azimuth transmitter located on the runway centerline extension beyond the stop end of the runway and the elevation transmitter located adjacent to the runway at the glide-slope intercept point (GPIP). However, substantial economies result from collocating azimuth and elevation antennas at a single site, usually opposite the GPIP. Although collocation limits the system from being used at the lowest minimum condition, it is desirable at low-activity airports and consideration should be given to include collocation capability in the design of a system.

7.3.2.7 Synchronization. The signals from the azimuth and elevation transmitters must be separated at the receiver, either by space, frequency, or time. Space diversity cannot be achieved because both vertical and lateral guidance signals must be received at all points along the flight path. Frequency diversity is achievable, but is considered somewhat wasteful of frequency spectrum and also requires multifrequency processing capabilities by the aircraft receiver. Time diversity, or synchronization, appears to be the most useful technique to separate the azimuth and elevation course signals. This synchronization is simple to achieve in the collocated facility. However, for the split-site configuration it is necessary to establish a communication link between the two sites for transmission of the sync signal. This communication link probably should be a landline, since at these line-of-sight frequencies a radio link could be interrupted by taxiing aircraft, land vehicles, or people.

7.3.2.8 Angle Modulations. The most convenient technique to identify the four beams radiated by the fixed fan-beam MLS, i.e., "fly-up", "fly-down", "fly-left", and "fly-right", is providing a unique modulation on each of these beams. These angle modulations can be: (a) audio tones, such as those used with the conventional ICAO ILS, (b) pulse spacing, as used by currently operational mechanical scanning beam systems, (c) FM tones, as used by scanning beam systems now under development, or (d) pulse repetition rates, such as those used by currently operational fixed-beam systems.

7.3.2.9 Auxiliary Modulations. Auxiliary modulations (AUX data) are those that convey information other than angle information to the aircraft. These include beam identity, station identity, and weather conditions at the landing site. Except for beam identity, these modulations are characterized by low data rate requirements, since it is normally necessary to transmit this information only once every several seconds. Accordingly, the modulation can be superimposed on the angle modulation in the form of a Morse code or any convenient coding arrangement.

7.3.2.10 Beam Shapes and Orientation. An early decision in the design of a MLS, normally based on operational considerations, is whether the fan beams should have a planar or conical configuration. The planar configuration normally is preferred for the azimuth course beam, particularly at high azimuth angles, but not always for the elevation beam. The conical elevation beam has an advantage in a tactical environment for low-performance aircraft or helicopters that may arrive at the landing facility from any direction, since the elevation angle is independent of azimuth angle. The planar elevation beam has a particular advantage for conventional high-performance aircraft landing under CAT II or CAT III conditions, since a conical beam becomes noticeably hyperbolic at a 50-foot minimum guidance altitude when transmitted from a site offset from the runway centerline. However, an airborne computer with range input can perform the necessary coordinate conversion. In addition to the planar vs. conical considerations, the fan beams must be shaped to provide proper information in the operationally required coverage and guidance region.

In order to avoid significant reflections of side-lobe energy into the main lobe, it is desired to keep the side lobes of the radiation pattern at the minimum feasible level. Normal practice in landing system antenna design is to keep side lobes at least 20 dB below the main beam.

Radiation reflected by the ground plane tends to contaminate both the azimuth course and the elevation guidance signals, and every effort should be made to reduce this multipath radiation. This normally is accomplished by shaping the beam such that there is a sharp power cut off at lower elevation angles, i.e., at those angles below the useable portion of the elevation signal.

7.3.2.11 Polarization. The polarization of the radiation must be considered in the design of any type of MLS. This is necessary to assure compatibility between the transmitter antenna and the receiver antenna, since horizontally polarized radiation will be severely attenuated by an airborne antenna designed to receive vertical polarization, and vice-versa. A design factor that should be considered in overall system performance is that when an aircraft banks, the antenna also banks, and the energy received by the aircraft antenna will be proportional to the cosine of the bank angle. Circular polarization could be generated for the transmitted radiation and accepted by both vertically and horizontally polarized aircraft antennas, but the signal would be at least 3 dB below that of matched polarization antennas. Circular polarization has an advantage when applied to radar landing systems, such as the GCA, since backscatter from precipitation is considerably reduced. This is because the direction of the circular polarization reverses when the radiation is reflected from a circular object, as approximated by a raindrop in cross section.

Another factor that should be considered in the selection of the polarization of the MLS radiation is the magnitude of reflections from dielectric surfaces such as the ground. Vertically polarized radiation is subject to more attenuation upon reflection from the ground than is horizontally polarized radiation, particularly in the vicinity of the Brewster angle. However, it does not necessarily follow that horizontally polarized radiation is attenuated more than vertically polarized radiation upon lateral reflection from objects such as aircraft and hangars, since the polarization attenuation advantage applies only to reflections from dielectric material and not normally to specular reflections from metal objects. The general view at present is that the MLS signal should be vertically polarized.

7.3.2.12 Data Rate. Data rate requirements are determined by the ability of aircraft to respond quickly to deviation signals, and this varies widely between classes of aircraft. The most critical requirement is generally considered to be the flare maneuver, in which rapid changes in pitch attitude is required to follow the precise curved path to touchdown. The proposed data rate for flare has been increased from 10 Hz to the neighborhood of 40 Hz in order to further alleviate multipath effects.

7.3.2.13 Focusing. Focusing is the technique for achieving a far-field radiation pattern throughout the region in which the signal is intended to convey useful information. Any equipment that monitors the transmitted signal also should be in the far-field radiation region. Transmitting antennas normally focus the radiation at infinity, and the depth of the far field for this focusing extends from a distance of about D^2/λ to infinity where D is the aperture dimension and λ is the wavelength of the radiation. In the event that it appears that focusing is necessary, the focal distance must be carefully selected. Focusing at very short distances to bring the far-field region closer to the transmitter brings the outer limits of the far field to a finite distance much too close to the transmitting antenna.

7.3.2.14 Primary Power. The selection of a primary power source for an MLS is determined by both the power requirements of the ground stations and the available facilities at the landing site. System design of tactical systems should allow for normal variation of voltage and frequency encountered in field generators.

7.3.2.15 Size and Weight. The dimensions of landing system equipment normally are determined by the required antenna aperture dimensions to obtain the necessary beam widths. The permissible size of the equipment is constrained by regulations that prohibit the installation of large structures in the immediate vicinity of runways. The weight of the ground-station equipment is not a significant factor except when it is necessary to transport the equipment frequently, such as may occur in a tactical environment.

7.3.3 CHARACTERISTICS OF SCANNING BEAM MLS. Scanning beam MLS are similar to fixed fan-beam MLS except for beam motion. This motion provides considerable operational advantages, but also adds a commensurate degree of complexity to the system. Accordingly, it must be determined whether the operational advantages justify the incremental complexity. These factors are discussed in the following paragraphs along with a description of the additional design parameters associated with scanning beams.

7.3.3.1 Operational Advantages. An operational advantage of a scanning beam MLS is that the guidance region can be selected at the airborne equipment almost anywhere within the coverage volume. An associated advantage is that the scanning beam information can be readily combined with DME information to provide guidance for curved and segmented approaches to the landing facility.

7.3.3.2 Incremental Complexity. The complexity of the scanning beam MLS is associated with technology for: (a) generating the beam motion, (b) modulating the beams as a function of pointing angle for angle signature systems, and (c) where applicable, synchronizing the azimuth and elevation course signals. The simplest mechanization for beam motion is rotation of the antennas. This technique normally produces planar fan beams. The beam motion also can be achieved by electronic scan techniques via frequency or phase. The frequency scan can be obtained by applying a varying carrier frequency to a slotted wave guide. Since the squint angle of the radiation emanating from the slots is a function of frequency, a conical beam pattern is generated from a linear slot array. This technique is somewhat wasteful of frequency spectrum, but the beam scanning motion is as smooth and continuous as it is with the mechanical scan system. Beam motion also can be achieved by changing the RF phase relationships among several point sources of radiation within the antenna. The beam is generated by the vector addition of the radiation from these point sources. Normally, this is accomplished with diode phase shifters, resulting in a quantized rather than a continuous motion, which is tolerable providing the magnitude of the steps are tiny, such as only a small fraction of the beamwidth.

Some form of modulation is necessary to identify the beam pointing angle when it illuminates the aircraft. A simple modulation technique is to illuminate the entire coverage volume from a separate antenna with a timing pulse prior to the initiation of the scan. The beam pointing angle then is related to the time difference between aircraft illumination by the reference pulse and the scanning beam. Another method is to modulate the scanning beam with a signal that varies as a function of beam pointing angle. This removes the requirement for providing a reference signal from a separate antenna, since the angle signature is impressed as modulation on the scanning fan beam. These and other necessary complexities of the scanning fan beam MLS are evident in the subsequent discussion of its design parameters.

7.3.4 SCANNING BEAM DESIGN PARAMETERS. All of the previously discussed design parameters associated with fixed fan-beam MLS also are relevant to the scanning beam MLS. These parameters as applied to the scanning beams are discussed along with those that apply uniquely to the scanning beam system.

7.3.4.1 Channelization. The larger guidance and coverage volumes associated with a scanning beam system result in a higher probability that signals from one scanning beam system will intrude on the spatial coverage volume for signals of a similar nearby facility. This imposes a more severe channelization requirement for scanning beam systems than fixed fan-beam systems, although the bandwidth for a time reference system (TRS) is considerably less than for the frequency reference system (FRS) and still somewhat less than the bandwidth required for the proposed Doppler format.

7.3.4.2 Radiated Power. Several factors must be considered in comparing the radiated power requirements for fixed fan beam vs. scanning beam. These factors all relate to the signal-to-noise ratios required by the data processor, and the time-per-bandwidth factors associated with the radiated signal. The guidance information for a fixed fan-beam system is not extracted from the maximum power points of the beams, but rather from the cross-over region, which may be several decibels down from the peak. However, information contained at the peak of the lobe is useable for extracting angle information with the scanning beam systems. The scanning beam systems normally have a larger requirement for radiated power in acquisition than for tracking, because the time-per-bandwidth product is large when the airborne station is operating in the acquisition mode. But when operating in the tracking mode, the aircraft receiver is capable of excluding all noise that is not within the dwell time of the received signal. Another parameter related to the required power in angle signature systems is the frequency of the modulation that determines beam pointing angle. This frequency should be set at the minimum useful value in order to reduce the bandwidth of the signal. The foregoing factors suggest that the scanning beam system will require greater power than a fixed fan-beam system.

7.3.4.3 Coverage Region. There are no theoretical limitations to the angular coverage of a scanning fan beam system using mechanical scan techniques. The coverage for electronic scan system is limited by the beam broadening characteristics of the high-angle radiation generated by the electronically scanned arrays.

7.3.4.4 Guidance Region. A well-designed scanning beam system provides a guidance region that is only about one beamwidth less than the coverage volume of the system. This region can extend downward to an angle that is approximately a half beamwidth above the ground plane.

7.3.4.5 Angle Modulations. In order to extract information from the scanning beam, it is necessary to know its pointing angle at the time the beam illuminates the aircraft. One method by which this is achieved is the Time Reference System (TRS), as represented by U.S. and Australian candidate systems described in Section 8. Another method is to impress a modulation on the beam that varies as a function of the pointing angle. The modulation can be of any convenient form, including pulse, AM, FM, or PM. Constraints imposed on angle modulation selection are that it should not waste the frequency spectrum, and that the modulation should be capable of being processed by the airborne station to meet the required beam pointing angle accuracy. Pulse modulation probably is more wasteful of the frequency spectrum. The pulses must be sufficiently short such that several pulses can be transmitted during the time that the scanning beam illuminates the aircraft. The bandwidth occupied by a pulse is related to the reciprocal of the width of the pulse. The frequency of the beam pointing angle modulation must be high enough such that several cycles occur while the beam is illuminating the aircraft, in order to assure effective processing by the airborne station. Also, the variation of modulating frequency with beam pointing angle must be high enough such that the airborne processor can measure this angle within the required system accuracy. These factors limit the data rate for systems in which angle signature is provided by modulation, yet high data rates are desirable, particularly for averaging multipath effects.

7.3.4.6 Auxiliary Modulations. In the fixed fan-beam systems the same modulations are used for both beam identity and guidance information. This same feature could be achieved with scanning beam systems, provided the modulating frequency spectrums were different for each scanning beam. However, this frequency diversity could result in inefficient, wasteful use of the frequency spectrum. Accordingly, it may be desirable to use the same segment of the spectrum for all the scanning beams at a particular facility, and use an independent modulation to identify each beam. This identity modulation could be superimposed on the beam pointing angle modulation, provided care is taken to avoid adverse effects of modulation cross products. The arrangement for identifying the particular beam when it illuminates the aircraft would be a sufficient identification technique, except for the AGC requirements of the airborne receiver imposed by the large, dynamic range of signal strengths from successive beams. For example, between decision height and touchdown the aircraft is much closer to the elevation transmitter than it is to the azimuth transmitter; thus it is necessary to reduce the receiver gain when the aircraft is illuminated by the elevation beam. This can be accomplished by transmitting a signal, prior to transmitting the elevation scanning beam, with modulation conveying information to the aircraft receiver that the next beam it perceives will be the elevation beam. This will enable the receiver to set the AGC to accommodate the signal strength of that beam. Other auxiliary modulations, such as station identity and weather conditions, can be provided in the manner similar to that provided by the fixed fan-beam systems.

7.3.4.7 Beam Shapes. The planar vs. conical considerations are substantially similar for both fixed-fan beams and scanning beams. Beyond that, there are no basic similarities in beam shapes. While a fixed fan-beam system normally provides two fixed beams for each function, a scanning beam system's cycle of a single scan beam can be provided by mechanical rotation of the transmitting antennas. Beam shapes independent of pointing angles are obtained by this technique. The shape of a beam generated by electronic scan techniques normally is a function of beam pointing angle.

7.3.4.8 Duty Cycle and Data Rate. The duty cycle and data rate available to scanning beam systems are considerably less than those available to fixed fan-beam systems. The fixed fan beams continuously illuminate the aircraft, while the scanning fan beams illuminate the aircraft only during a brief segment of each scan. The data rates of scanning fan beams generated by rotating antennas normally are limited by the mechanical forces associated with higher rotation rates. While there is no theoretical limit to the generation of data rate for electronic scanning beams, the rate nevertheless is limited by the dwell time requirements for the airborne processor.

7.3.4.9 Size and Weight. The size and weight considerations normally are similar for fixed fan-beam and scanning beam systems. There are situations, however, where the scanning beam system will require considerably larger apertures than the fixed fan-beam system. For example, if it is considered desirable to generate planar beams by electronic scan techniques, the necessary circular array will have a much greater aperture dimension than an equivalent linear array for generating conical beam shapes.

7.3.4.10 Dwell Time. The dwell time is that period of time during each scan in which the scanning fan beam illuminates the aircraft. It is necessary that the dwell time be of sufficient duration for the airborne processor to extract beam pointing angle information. Dwell times of between 1 and 3 milliseconds normally are considered necessary to extract this information when the angle is defined by the modulation on the beam. While it is desirable to maximize dwell time, this can only be accomplished by adversely affecting other important system parameters. It can be increased by enlarging the beamwidth, but this would result in more severe multipath problems. It can also be increased by decreasing the angular velocity in the scanning beam, but this would result in a lower data rate for the system. Thus, the system parameters of dwell time, scanning velocity, and beam width are interrelated and the effect of each must be examined in relation to the others. The U.S. and Australian TRS candidate systems can function with a small dwell time, thus permitting higher scan velocities and data rates, because the pointing angle is indicated by the measured time between successive "to-fro" scans, there being no modulation to define the pointing angle.

7.3.4.11 Scanning Motion. The preferred scanning motion of the beam is a continuous angular velocity throughout the scan region. This is most easily accomplished with a mechanical scan system. Electronic scan systems require moving the beam in steps rather than in a continuous motion. While additional switching circuitry is required to decrease the step size, it is necessary to decrease this parameter to its minimum feasible value, since large quantized motion has an adverse effect on system accuracy.

7.3.5 CHARACTERISTICS OF DOPPLER SCAN SYSTEMS. The Doppler MLS provides signals throughout the coverage volume that provide self-encoded angle information. No independent modulation is required to indicate the azimuth and elevation angles of the aircraft relative to the respective ground stations. This is accomplished at the ground station by transmitting a reference signal and simultaneously a signal that electronically simulates motion of the ground antenna by commutating, i.e., sequentially illuminating, the elements of a linear array. It can be determined by elementary Doppler principles that the carrier frequency will be shifted at the aircraft receiver by an amount equal to:

$$f_d = \frac{v}{\lambda} \sin \theta$$

where:

f_d is the Doppler shift

v is the commutation velocity

λ is the wavelength

θ is the angle of the aircraft measured from a line perpendicular to the array.

In order to avoid standing waves and facilitate data processing at the airborne station, the basic carrier frequency of the commutated signal is offset from that of the reference signal as illustrated by Figure 7.3. An interesting and useful phenomenon associated with the Doppler technique is that the direction of the commutated velocity reverses when the signal is reflected. This provides a tool for the rejection of multipath signal by the use of filters.

7.3.5.1 Incremental Complexity. The complexity of the Doppler MLS is associated with the mechanization of the commutated array. The essence of this problem is the design of high-speed switching hardware, which is more readily available at C-band than at Ku-band. Thus, it imposes a penalty at present upon extending this technique to Ku-band.

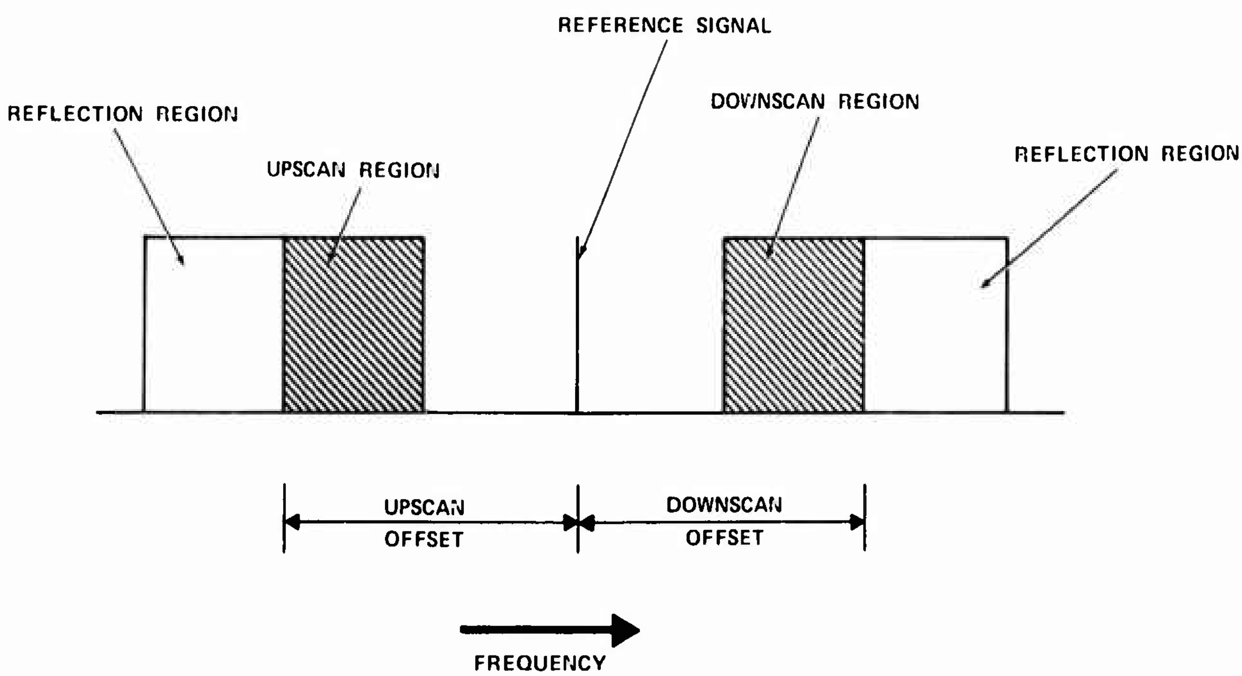


Figure 7.3 Doppler Channel Occupancy

7.3.6 DOPPLER SCAN DESIGN PARAMETERS. Design considerations of many parameters for fixed fan-beam and scanning beam systems apply to Doppler scan systems and have been covered in preceding paragraphs. The parameters of Doppler systems with design characteristics unique to that system are discussed in the following paragraphs.

7.3.6.1 Radiated Power. The peak radiated power for the Doppler MLS is considerably less than that for the radiated scanning beam MLS for equivalent performance. However, the average radiated powers for these two techniques are approximately the same. The accuracy and useful range are more dependent upon average power than on peak power, as will become evident in the subsequent discussion of other design parameters.

7.3.6.2 Beam Shapes. The radiation from each element of the commutative array fills the entire coverage area of that ground station. While the spatial beamwidth is large, it is more useful in the design of a Doppler system to consider its equivalent beamwidth (EBW) in the frequency domain, i.e.:

$$EBW = \lambda/L \text{ radians}$$

where:

λ = the wavelength of the carrier frequency

L = the antenna aperture dimension

The surfaces of position associated with constant Doppler shifts are conical rather than planar, but a computed planar surface can be established by transmitting a cone pair from orthogonal arrays.

The equivalent to side lobes in the amplitude domain for the scanning beam system is "spectrum side lobes" in the Doppler system. Commutation and switching frequencies introduce these pseudo-side lobes, and the system should be designed such that these contaminants do not interfere with the processing of the signal.

7.3.6.3 Duty Cycle and Data Rate. The duty cycle of a ground station such as azimuth or elevation in the Doppler system is that portion of the total time in which the signal from that station illuminates the aircraft. In a well-designed Doppler system, this period of time will be substantially the entire time that is allocated to the particular ground station.

7.3.6.4 Focusing. Focusing, in the usual optical understanding of the term, is not necessary in the Doppler system. The elements that radiate the signals have a wide beamwidth, and thus the far field starts very close to the array. There is a focusing-equivalent problem in the Doppler system in that when the aircraft is close to the commutating array the Doppler frequency received from one end of the array is noticeably different from that of the other end of the array.

7.3.6.5 Angle Modulations. It is not necessary to impress modulations on the Doppler radiation to indicate angles, since the radiation is self-encoded, a significant feature of the Doppler MLS.

7.4 AIR-DERIVED RANGE DATA

Range data normally is acquired independently of angle data in a MLS. It may be processed with the angle data to provide guidance signals for segmented or curved approaches. Range data also provides information for course softening when the aircraft is in the flare and touchdown zones, and particularly during rollout. Course softening is an operation that provides a guidance error signal proportional to the actual distance that the aircraft is off course, rather than the angle (measured from the ground station) that it is off course. There are two basic methods for acquiring range data: interrogation and synchronous clocks.

The interrogation method normally operates via an interrogation signal transmitted from the aircraft and received by a ground beacon transponder that in turn sends a response signal to the aircraft. The measured time between the original interrogation transmission from the aircraft and the reception by the aircraft of the beacon response transmission is related to the distance between aircraft and beacon and enables airborne computation of the distance.

The synchronous clocks method requires that both the airborne and ground stations have very accurate clocks that are synchronized with each other. Signals initiated by the ground station clock will arrive at the airborne station "out-of-sync" with the airborne clock, and the degree of this difference is a measure of the distance between the stations. An advantage of the synchronous system is that the range measurement equipment will not saturate in a high-density aircraft environment.

The interrogation and synchronous systems could be effectively combined, with occasional interrogations updating the airborne clock. The subsequent description concerns the interrogation DME. The design of an effective synchronous DME for operational use must await further developments in clocks to increase accuracies commensurate with propogation distances of a few feet and short-term stability (noise) characteristics, and lower long-term drift (bias), all at low cost, particularly for the airborne unit.

The conventional interrogation type of DME is illustrated by Figure 7.4. The ranging circuit in the airborne station initiates the transmission of an interrogation signal, which is received by the ground station beacon. This received signal is delayed in the ground beacon transponder and then retransmitted to the airborne station. The ranging circuit measures the time interval between the airborne transmission and the reception at the airborne station of the ground beacon transmission, and uses this time interval as a measure of the distance between the two stations for computation and display. One ground station normally is capable of serving at least a hundred airborne stations. This imposes a requirement that the airborne station identify the particular transmission by the ground station that is the reply to its particular interrogation. The airborne station also must reject noise pulses transmitted by the ground station or by other facilities within the operating frequency of the ground station. Several design parameters must be considered to implement this interrogation DME concept, as discussed in subsequent paragraphs.

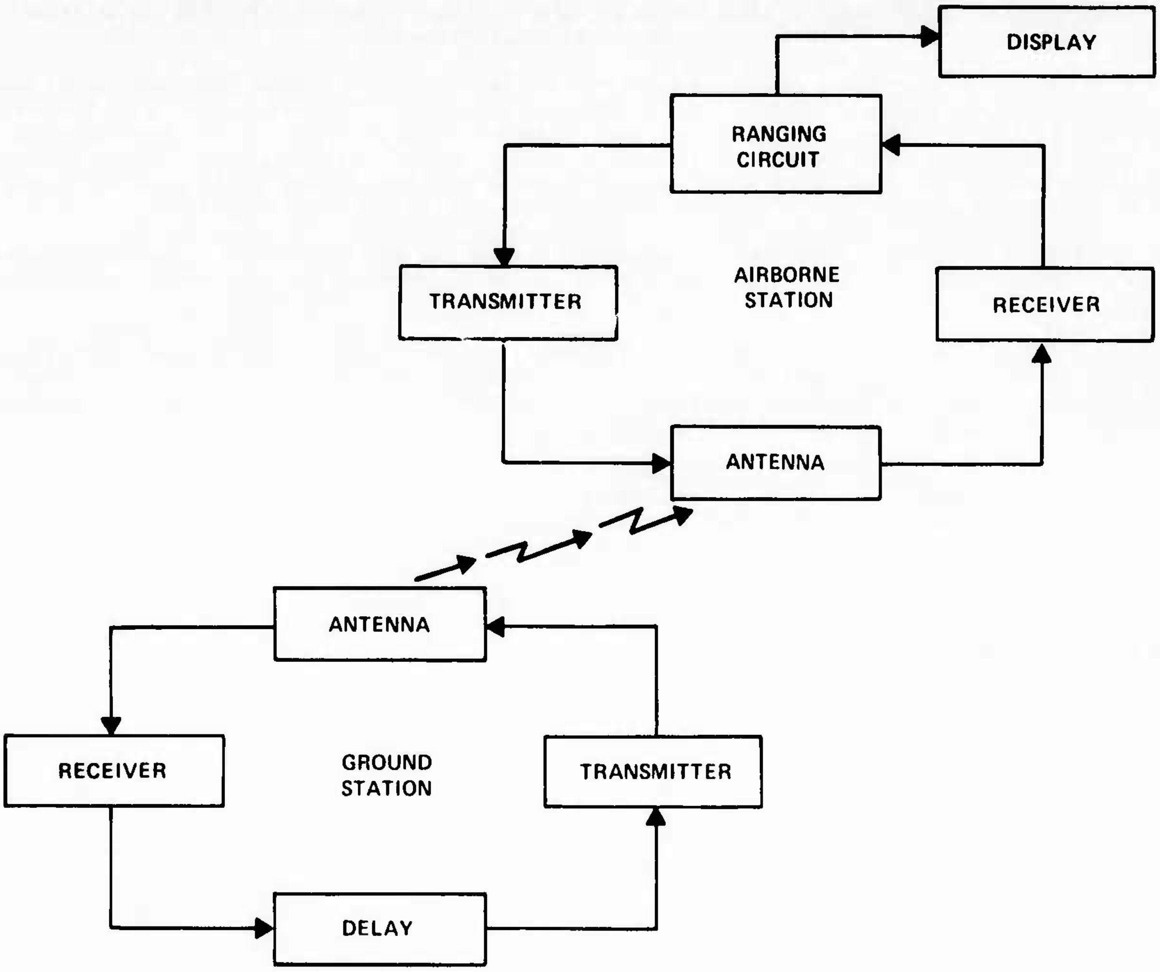


Figure 7.4 DME Block Diagram

7.4.1 CARRIER FREQUENCY. Most of the currently installed DME's operate in L-band at frequencies slightly above 1 GHz. There are technical reasons why the DME designed for use with MLS may operate at a higher frequency than L-band. Shorter pulse rise times associated with greater system accuracies are available, since there is more spectrum space to accommodate wider bandwidths at higher frequencies.

7.4.2 CHANNELIZATION. The channelization requirements are more severe for the range data than for the angle data subsystem of a MLS because of the broad spectrum occupied by the short DME pulses. A challenging but necessary effort is to design DME such that the airborne station effectively processes the desired pulses from the ground station being interrogated and rejects pulses transmitted by other nearby ground stations, yet with minimum channel separation.

7.4.3 RADIATED POWER. The peak carrier power supplied to the DME antenna normally is considerably larger than the power supplied to the azimuth and elevation antennas. There are several reasons for this greater power requirement. First, neither the airborne nor ground station antennas can have highly directional features because of the considerable volumetric coverage required for these signals. The duty cycles of the useful transmissions are also rather low, since the pulse durations are quite short as compared to the time between pulses. Furthermore, the receiver must have a rather wide bandwidth so as not to distort the leading edge of the pulses, and yet system noise is directly proportional to the bandwidth. The aggregate of these factors normally results in a peak-power requirement of several hundred watts for the DME, as compared to a peak power requirement of only a few watts for the angle data subsystem.

7.4.4 COVERAGE REGION. The DME ground station should accept interrogation signals from any aircraft located within the volumetric coverage of the associated angle guidance system, and the response pulse transmitted by this DME ground station should illuminate all aircraft within this coverage region.

7.4.5 GROUND STATION COORDINATES. The usual siting preference for the current DME ground station is collocation with the ILS localizer course transmitter. This arrangement permits aircraft to continue to receive range information through rollout, which is considered essential for CAT III system operation, and therefore collocation of DME with the MLS azimuth transmitter is desired.

7.4.6 SYNCHRONIZATION. Range data normally is provided in a different region of the frequency spectrum than is angle data. This frequency diversity obviates time synchronization of range data and angle data signals.

7.4.7 MODULATIONS. The basic modulations on the DME carrier are short pulses. These pulses normally are transmitted in pairs, with the interpair spacing constituting the measure of range. The intrapair pulse spacing can be used to convey other useful information such as station identity. The ground station transmits pulses at random time intervals when it is not being interrogated by airborne stations, so that signals are continuously available at the airborne station receiver in order that the AGC can be automatically reset at the proper level for reception of the response pulses.

7.4.8 BEAM SHAPE. As stated, the DME should provide a signal throughout the volumetric coverage of the MLS. While the coverage should be broad, there should be a sharp cutoff in radiated power near the ground to minimize ground plane illumination. A considerable illumination of the ground plane could result in a loss of DME signal for an aircraft at a low altitude near the fringe of the MLS coverage region, because in this region a phase reversal due to ground reflection plus the high magnitude of the reflected signal at these grazing angles could result in a substantial cancellation of the direct signal.

7.4.9 POLARIZATION. Vertical polarization is preferred for the DME signals in order to take advantage of the Brewster angle effects for low-angle reflections, although at grazing angles this advantage would not be significant.

7.4.10 DUTY CYCLE. The duty cycle of DME normally is extremely low since an airborne station interrogates a ground station at a rate of less than 100 Hz with pulses of less than 1 microsecond in duration. This low duty cycle requires the use of effective gating techniques to exclude system noise during those time domains when no useful information is being transferred.

7.4.11 DATA RATE. There are conflicting data-rate criteria. A high data rate improves system performance in both the acquisition and tracking modes. When in the acquisition mode, the larger the number of interrogation and response pulses, the more rapid the correlation within the moving acquisition gate. When operating in the tracking mode, the higher the pulse rate, the greater the accuracy (via application of smoothing techniques), since the improvement in accuracy is related to the square root of the number of pulses in the smoothing time interval. However, there is an adverse effect of high data rate, in that the ground station operation can become saturated with fewer interrogating aircraft, resulting in a reduced overall system capacity.

SECTION 8

INTERNATIONAL DESIGN PROPOSALS FOR MLS*

New precision approach and landing systems, referred to as MLS, have been under development on an international basis for the last 5 to 6 years. Five countries, Australia, France, the Federal Republic of Germany (F.R.G), the United Kingdom (U.K.) and the United States (U.S.), are submitting candidate designs to ICAO for selection as an international standard [40] [41]. This new system will eventually fully replace the currently used VHF/UHF ILS.

This section discusses some of the issues associated with design of a microwave landing system and then briefly summarizes each of the landing system programs of the various countries and their candidate systems. The emphasis of the description is on signal format, since it will be the end result of an international standardization. The major issues in making this selection can be grouped under headings of technical performance, integrity, and implementability. In this presentation the design features affecting these issues are stressed.

The background and general guidelines for the U.S. development are provided in the final report of the Radio Technical Committee for Aeronautics (RTCA) Special Committee 17 (SC-117) [41], which contained international participants. The operational requirements set down by ICAO mirror the SC-117 report very closely. MLS operational features enable more varied approach and all-weather landing capabilities than those presently available with ILS. For example, smaller performance degradation due to multipath effects, flexible approach paths, and flare guidance under CAT III conditions are realizable with the new system.

8.1 DESIGN ISSUES

MLS is a compatible design in that each equipped aircraft (general aviation, aircarrier, or military) can operate with any ground station facility. The selected design must consequently accommodate a variety of landing system configurations, such as small remote sites, aircraft carriers, STOL ports, and sophisticated terminals requiring guidance information through touchdown and roll out. Avionics demands vary from those associated with a simple fixed-path approach at CAT I conditions to those required for curved paths and automatic blind landings with high-performance aircraft.

Elements of the ground facility common to all the proposed designs and their location are shown in Figure 8.1 for a typical split-site configuration. Basic information is obtained from the azimuth antenna (AZ), located at the stop end of the runway, and the main elevation antenna (EL-1) which is offset opposite the glide-path intercept point (GPIP). Azimuth angle coverage is available over a sector of up to ±60 degrees, and elevation angle coverage is provided from 0 degrees up to 20 degrees. The maximum range is at least 20 nmi under heavy rainfall conditions. The maximum range information is available from the precision distance measuring equipment (DME) located adjacent to the azimuth antenna. This fundamental data base is augmented by a flare guidance antenna (EL-2) and a back-course azimuth antenna (BAZ) facility for missed approach or departure guidance where required.

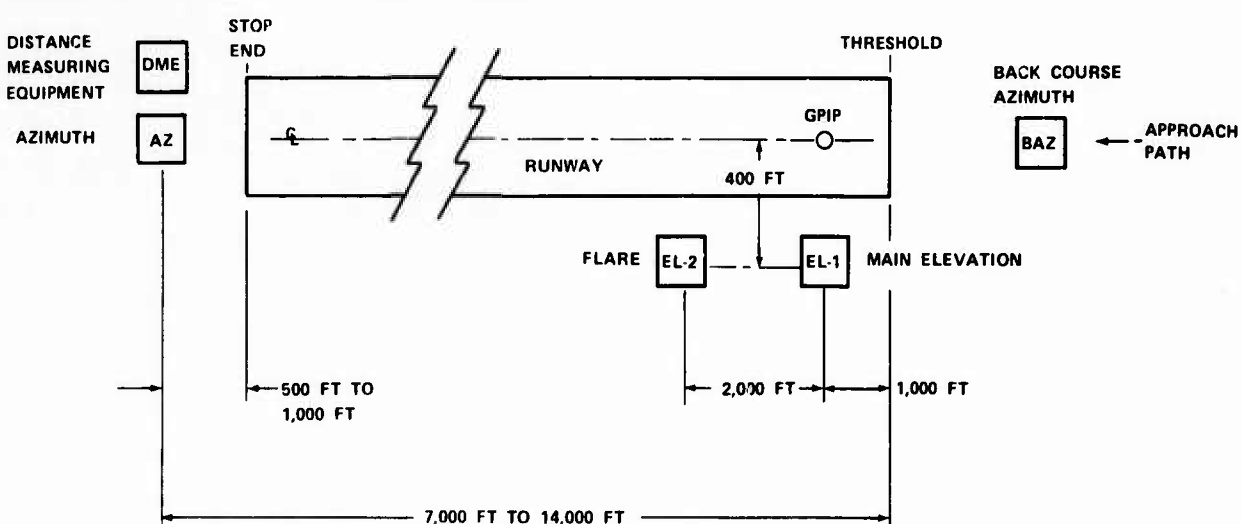


Figure 8.1 Typical Siting Arrangement for Various MLS Elements

Some of the basic issues associated with the provision of this information in the MLS are:

- Should the information be derived in the air, or should it be derived on the ground and transmitted up to the aircraft?

- By what technique should the angle information be obtained?

---

*This section is primarily based on two reports; one by S. Ahmed Meer and Stanley R. Jones of the Mitre Corp. under the direction of Project Leader Frederick C. Holland, prepared for an invited session on Air Transportation Systems, Man and Cybernetics, Dallas, Texas, October 2-4, 1974 [38]; and the other, basic documents used by the U.S. FAA MLS Technique Selection Steering Committee of the MLS Central Assessment Group in presenting their technique recommendation to the MLS Executive Committee [39].

- What equipment is requi.ed for the range measurement?

- What auxiliary data capability is desired?

Some aspects of these issues are reviewed in the following paragraphs.

8.1.1 AIR DERIVED VS GROUND DERIVED. Proponents of the ground-derived landing information approach have referred to the concept as "the integral approach" since it combines the landing or navigation function with the surveillance or ATC function. On the other hand, support for the air-derived concept is often based on the fact that this provides a desired separate and independent implementation of the navigation and surveillance functions.

8.1.1.1 Advantages of Ground-Derived Concept. No computation is required in the aircraft for position determination; a single computer at the ground site (with backup) can process data and then send smoothed position information to each aircraft. This should be particularly advantageous in such instances as deck motion compensation for aircraft carrier operations and operations involving a fleet of aircraft with a complex requirement such as curved approaches. Safety of a military ground station against electronic countermeasures (ECM) may be easier to assure in a ground-derived concept. Total avionics for both surveillance and navigation may also be minimized by integration of these functions. Avionics cost for a minimum landing capability still may be comparatively high, however, with little spread in cost between the simple and sophisticated user, since they must all carry basically the same complement of a transponder and data-link receiver.

8.1.1.2 Advantages of Air-Derived Concept. Costs for a minimum-configuration ground site are less in the air-derived concept since even the simplest site for a ground-derived system fundamentally requires a significant computation for position estimation and roll call as well as a data link capability.

Transmission of basic data from the ground, with decoding and processing in the air, avoids a potential capacity problem associated with the ground-derived concept, which must operate as a high frame-rate surveillance system that in addition provides a data link service. The air-derived system likewise removes a potential problem of transport delay associated with making position estimates on the ground and then relaying them to the aircraft. Finally, it may be argued that the air-derived concept is intrinsically more reliable because the ground-derived system encounters the following limitations relative to an air-derived system:

- The ground site is more complex than is the site for an air-derived system, which increases failure and maintenance problems.

- Each data update for the ground-derived approach requires success on a three-way link transaction (interrogation-reply-uplink data) rather than on the one-way link in air-derived operation. A failure on any one of these link relays means a consecutive loss of information on the next scan of the ground-derived system.

- The higher complexity of the ground-derived system may lead to a less graceful failure.

These factors were considered by RTCA in their recommendation of an air-derived system. An important consideration in their choice was the belief that safety and integrity of the present ATC philosophy is due to the independent navigation and surveillance systems. This choice is generally endorsed by pilots, who tend to prefer cockpit-derived navigation data independent of the ground surveillance system.

8.1.2 ANGLE MEASUREMENT. Angle measurements for the ground-derived proposals emphasize the potential advantages of either computer processing or signal correlation. One of the concepts is based on outputs from interferometers operating at C-band, while the other ground-based system obtains the basic measurement from a modified L-band TACAN/DME signal with special ground arrays sited at the azimuth and elevation locations.

The air-derived designs involve variations of the basic angle measurement techniques, scanning beam frequency reference system (FRS), and Doppler scan, indicated in Figure 8.2(a) and (b) respectively, plus scanning beam time reference system (TRS) as exemplified by the Australian candidate (see Figure 8.3). Both techniques radiate azimuth and elevation guidance signals in space. The Doppler signal is created by the apparent linear motion of a CW source. The Doppler shift on the received signal serves as the measure of the angle between the receiver and the virtual direction of motion of the source. The scanning beam signal in space, on the other hand, is created by scanning a narrow beam whose transmission is coded by the beam's pointing angle. The scanning beam (FRS) angle position is encoded by a separate frequency modulation (FM) on the beam, which is decoded when the center of the beam illuminates the receiver. The scanning beam (TRS) angular position is inferred from the time between beam passes.

Although the TRS concept is restricted to "to-fro" and therefore electronic scan implementations, the method of beam position encoding allows either mechanical or electronic scan realizations. As mentioned above, Doppler encoding is typically obtained by electronic commutation of a radiating source along an array of elements. It may, however, be implemented by a beam-port concept whereby overlapping beams are simultaneously coded in the proper frequency relationship. This method has also been extended to a Doppler configu-

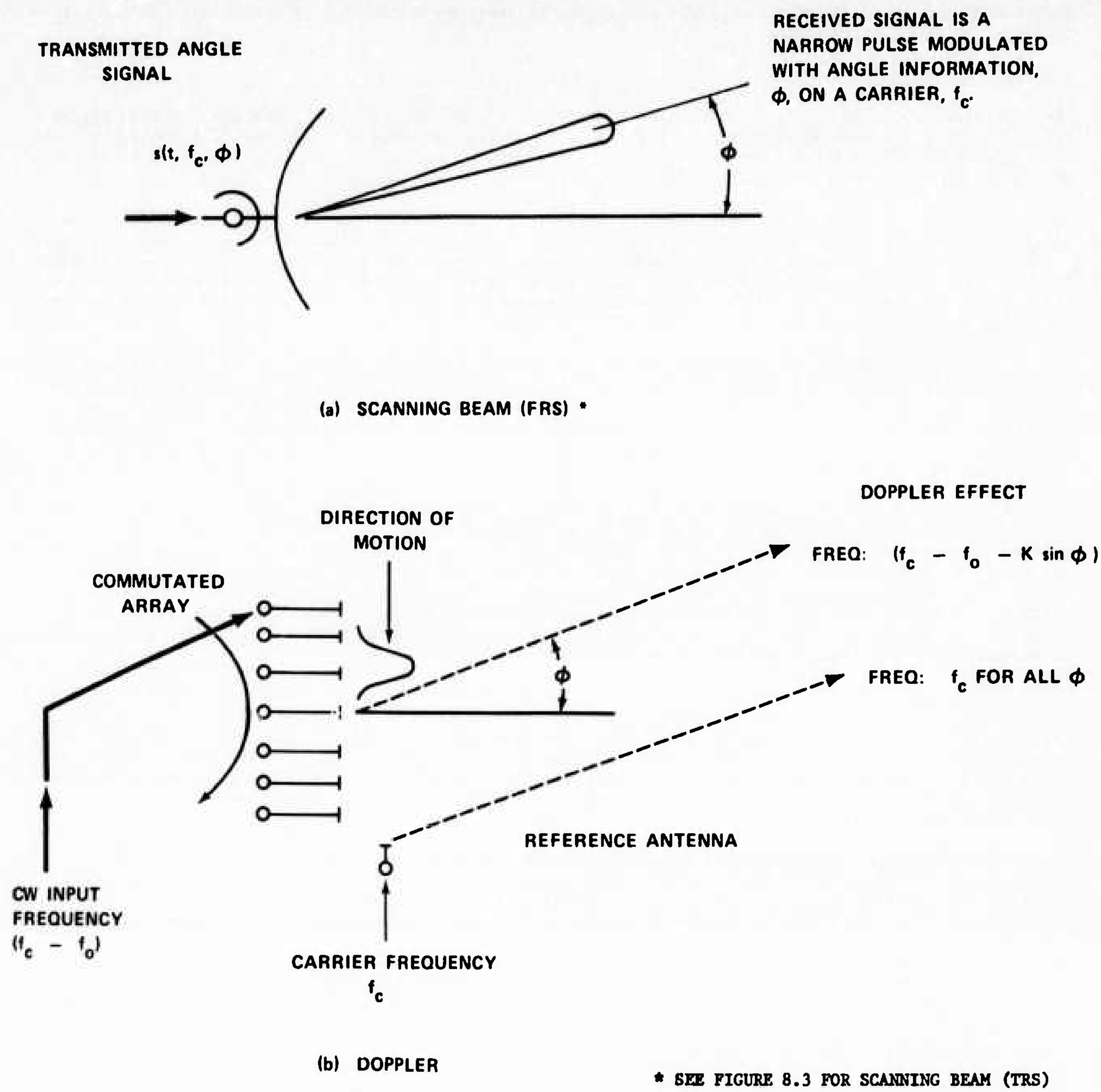


Figure 8.2 Scanning Beam (FRS) and Doppler Scan Techniques

ration developed from a sequential variation of the phase relationships between the progressive phase modes on a circular antenna.

8.1.3 COORDINATE SYSTEM. In addition to the technique of angle encoding, there is the question of what should be the angular coordinate system used in this process. Either Doppler or beam scanning of fixed linear apertures produces angle data in conical coordinates. Physical rotation of linear apertures, electronic scanning of circular arrays, and certain other quasi-optical systems provide the information in planar coordinates. Both scanning beam and Doppler may be implemented in either conical or planar coordinates.

8.1.4 RANGE MEASUREMENT. The RTCA report provides for a new C-band DME to be implemented in the MLS. The high accuracy available with this new design is particularly appropriate for the flare guidance descent rate calculation as well as for aircraft carrier operations. Many now believe, however, that adequate range data for the MLS may be obtained at less cost by modification of the currently deployed L-band DME. Resolution of this question, as well as the question of what type of range data (low accuracy DME or marker beacons) is required for low capability users, can be made somewhat independently of the other air-derived system issues.

8.1.5 AUXILIARY DATA CAPABILITIES. An integral data link is available in all MLS proposals. Basically, this link must provide reliable identification of various kinds of transmitted guidance information (function identification). Additional capacity is available for sending information called auxiliary (AUX) data, concerning the ground-site capabilities or data related to position computation in the air-derived systems. AUX data contains runway identification, site configuration, distances, and height difference between the MLS ground antennas, and weather information. Other AUX data transmission such as runway condition or wind shear is also possible, but this AUX link capability must not impair the basic performance of the system for aircraft landing. Some of the practical aspects of these design issues are illustrated in the following overview proposed designs.

8.2 AIR-DERIVED MLS

The Australian, U.K., and U.S. microwave landing systems are all air-derived. They also closely follow the frequency bands and channel assignments of SC-117. C-band in the range of 5,000 MHz to 5,200 MHz is used for the main functions of azimuth (front and back), elevation, and range.* Ku-band in the range of 15,400 MHz to 15,700 MHz is available for the flare-out function.

The total number of channels for each function in 200. The 200 "angle" channels for azimuth (front and back) and elevation have a bandwidth of 600 kHz each. These C-band channels also transmit AUX data, which is multiplexed with the angle data. Multiple tone codes or binary codes are used for AUX data. Each DME channel has a bandwidth of three MHz, three times the present L-band DME. To obtain 200 "effective" channels, each of 20 DME frequencies employs 10 independent waveforms or codes as opposed to the present two codes (X and Y) in the L-band DME.

The flare-guidance function is allocated 900 kHz per channel at Ku-band. However, all countries are also evaluating the use of C-band for flare. In comparison to Ku-band, an all C-band system has the obvious advantages of: (a) one airborne antenna and one RF front-end, (b) lower rain attenuation, (c) higher efficiency amplifiers, (d) lighter transmitters, and (e) spectrum conservation. The main disadvantages compared to Ku-band are the basically larger EL-2 ground antenna and the near-field effects on the antenna beam-width that lower accuracy.

The flare signal structure, whether C-band or Ku-band, is similar to the other angle functions, so that it can be integrated with the other angle functions in the avionics. The DME, on the other hand, is independent of the angle functions except that the 200-channel designs are generally arranged to permit the use of common synthesizers. The signal spectrum use in each 600-kHz C-band channel depends on the technique. The differences among each technique are therefore better understood by looking at the manner in which this bandwidth and the corresponding time are subdivided in order to multiplex the various angle functions and AUX data. Only the U.S. is supporting a significant DME effort. The following discussion is restricted to the angle measurement features of each design.

8.2.1 AUSTRALIAN INTERSCAN. The Australian MLS uses the TRS concept and is called INTERSCAN (Time Interval Scanning). The technique is based on the scan of a narrow beam across the sector of coverage at a precise rate. The scanning speed is uniform with the beam, starting from one extremity of the ±66° azimuth component coverage sector and moving to the other and then back again to the starting point, thus producing a "to-fro" scan as shown in Figure 8.3. In each scanning cycle, two pulses are received by an approaching aircraft; the time interval between the "to" and "fro" pulses is proportional to the angular position of the aircraft.

The "to-fro" scan is followed by data transmission over a separate broad-beam pattern for a duration roughly equal to the scanning time. Data is transmitted on eight tones. Three of the tones are used to identify the function and are called function identification (FCN ID) data. The other five tones are used for AUX data. The eight tones are phase modulated on the same carrier radiated on the scanning beam.

Frequency division multiplexing (FDM) is used to multiplex the various azimuth and elevation functions. Each of the 600-kHz C-band channels specified by SC-117 is divided into five subchannels of 50 kHz each with 70-kHz guard bands between the channels, as shown in Figure 8.4(a). The Australian design includes an additional back-elevation (BEL) signal and also employs a C-band EL-2 flare signal. The timing for the main azimuth (AZ) and back azimuth (BAZ) are similar, as shown in Figure 8.4(b). The complete cycle of the azimuth function is 50 ms, half for the angle scanning and the other half for the FCN ID and AUX data. The resulting azimuth angle data rate is 20 Hz. The angle data rate for elevation functions is double the azimuth rate, as shown in Figure 8.4(c), with a 25-ms period for a complete cycle. A multiple-feed reflecting antenna is used to uniformly scan at the relatively high data rates. The feeds are commutated electronically to scan the sector.

The proposed Australian EL-2 flare concept is novel in that the antenna will face the runway from the side, as shown in Figure 8.5(a). The azimuth axis of the EL-2 is therefore at right angles to EL-1. The EL-2 antenna has a wide horizontal coverage centered on the direction of the touchdown point. Planar beams are employed for both azimuth and elevation.

In the airborne receiver, the five functions are first amplified in a common RF and intermediate frequency (IF) section before being separated into five narrow-band (50-kHz) IF channels, each controlled by its own automatic gain control (AGC). The output of each channel is directly identifiable with the corresponding function. In addition, the FCN ID data tones verify the signal. The bandwidth of the scanning pulse and the data is 13 kHz, with an allowance of ±10 kHz for ground and air oscillator instabilities and a Doppler shift from aircraft speed of ±2 kHz. These total the allocated 50-kHz bandwidth. The specified airborne receiver noise figure is 9 dB.

Multipath rejection techniques used in INTERSCAN are similar to other scanning beam techniques. Pulse-width discrimination is used to detect the presence of multipath signals within the beam so that incorrect data can be discarded. The airborne AGC circuits are designed to select only the greatest amplitude pulses during acquisition; after acquisition only the pulses exceeding an amplitude threshold are processed. Time-gate trackers are also proposed. Evaluation of these techniques is still continuing. MLS configurations for general aviation aircraft for CAT I and lower category are also included in the Australian MLS proposal.

*The Australian design does not presently propose a special DME.

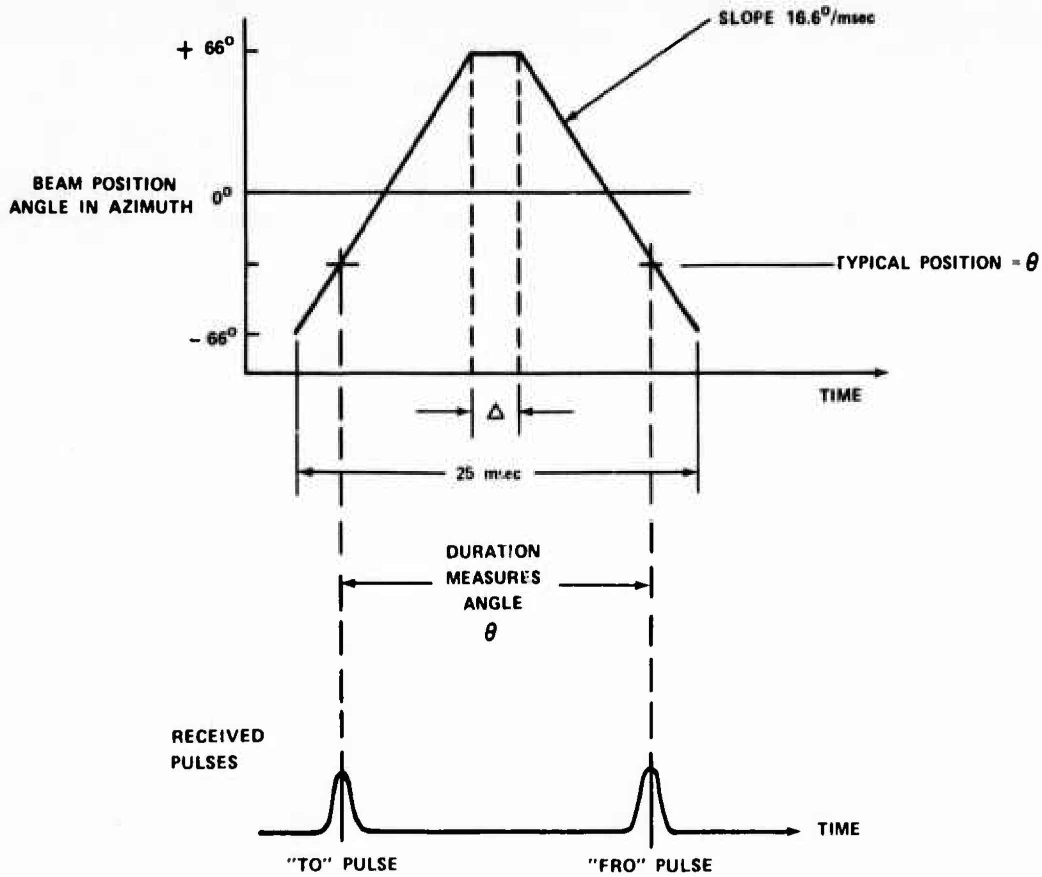
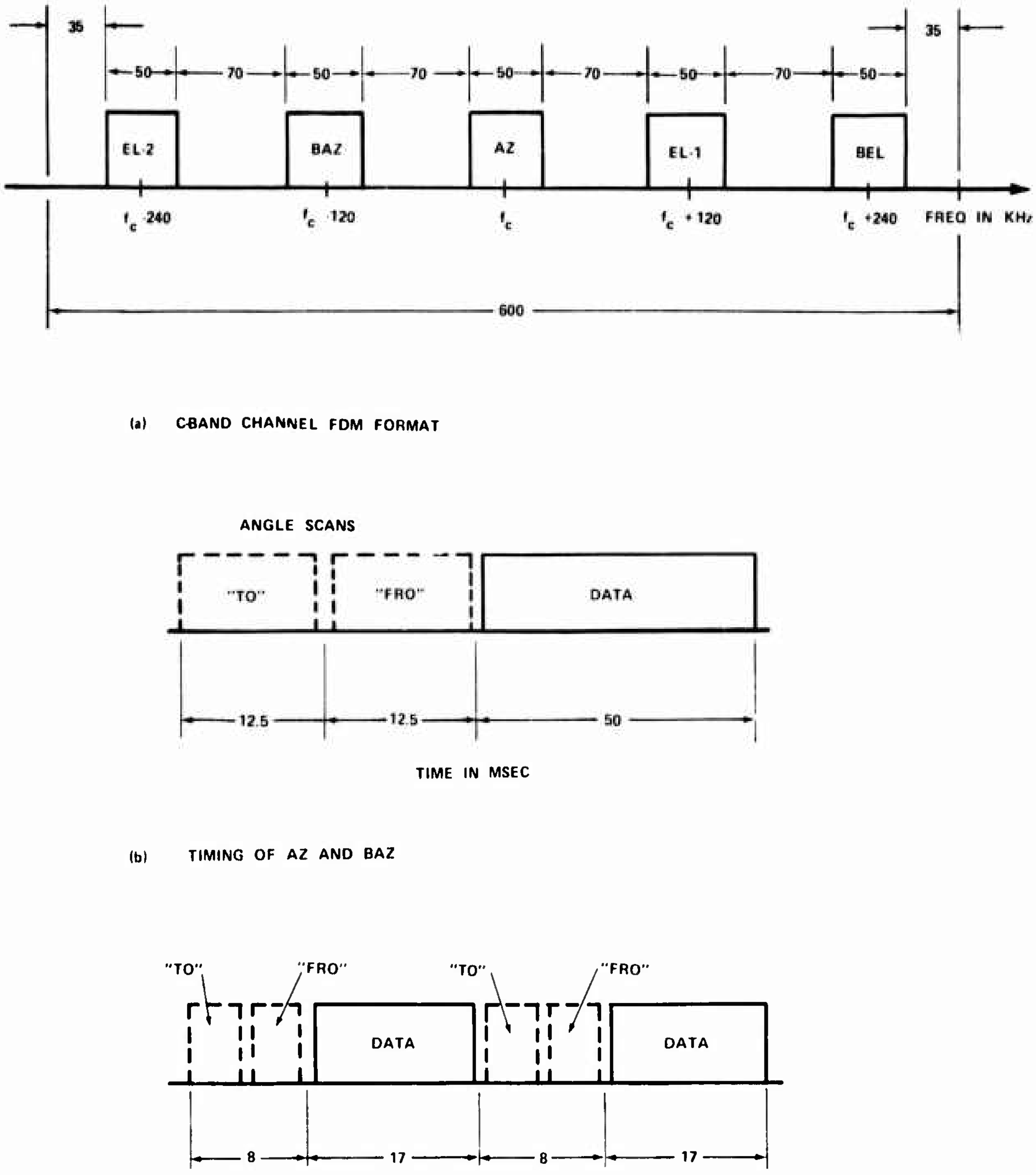


Figure 8.3 Australian AZ Scanning Signal

8.2.2 UNITED KINGDOM'S DOPPLER MLS. The U.K. was the earliest proponent of Doppler MLS. Their efforts have resulted in a fairly extensive theoretical and experimental basis for the use of a Doppler-effect signal. The proposed Doppler signal is generated by a commutated linear array, resulting in conical beams. Auxiliary arrays for conical-to-planar beam conversion are also included for CAT III configurations.

The signal format is basically frequency-division multiplexed (FDM). The 600-kHz C-band channel is divided into four subchannels as shown in Figure 8.6. The design emphasizes the importance of the AZ and EL-1 functions by dedicating subchannels to these functions. The AZ function is allocated 50 kHz, and EL-1 is allocated 15 kHz. The other 60-kHz subchannel is time-division multiplexed (TDM) with BAZ and signals from the auxiliary arrays. The fourth 15-kHz subchannel is for data. At present, EL-2 is transmitted separately on Ku-band as suggested by SC-117. The U.K. Doppler array is scanned alternately to and fro (and up and down). To get the same Doppler offset in the up-and-down scans, the frequencies to the Doppler array and the reference antenna alternate. The resulting signal at the aircraft is always single sideband; the sideband or offset frequency is 24.96 kHz for AZ and 14.96 kHz for EL-1. The Doppler coding sensitivity is 135 Hz per degree.

The array lengths are 120λ for AZ, 60λ for BAZ, and 90λ for EL-2, where λ is wavelength at the corresponding band. To satisfy the sampling theorem, the number of array elements must be equal to twice the number of wavelengths in the aperture. However, the U.K. design attempts to minimize the number of array elements by using either a course or fine configuration interferometer-like approach to thin out the number of array elements. This is done by making the reference antenna into a small array, colinear with the corresponding main array. By commutating the reference in the same direction as the main array, the change in relative spacing between reference and main array elements can be kept to the order of λ/2. Thus, instead of 240 elements in the AZ antenna array, the U.K. design has 64 main array elements and four reference array elements. Similarly, the EL-1 antenna has 32 elements in its main array and four elements in the reference array. The EL-2 main array has 96 elements, and 13 elements in the reference. Figure 8.7 illustrates the element-thinning technique for an azimuth array. A standby thinned array interlaced with the main thinned array is also being considered as shown in the figure; automatic switchover from the operational to the standby equipment is proposed for CAT III integrity requirements.



(a) C-BAND CHANNEL FDM FORMAT

(b) TIMING OF AZ AND BAZ

(c) TIMING OF EL-1, EL-2 AND BEL

Figure 8.4 Australian Signal Format

The elevation array elements are wave-guide horns that feed into a parallel plate region which provides an H-plane aperture to generate a ±30° sectoral pattern in azimuth. The basic element in the azimuth array is a monopole in a wave guide. This couples into an extended ground plane in order to achieve cutoff at low elevation angles. To improve the elevation cutoff thus obtained, consideration is now being given to wave-guide column arrays using slots in the narrow wall. The design objective for the azimuth array is to produce a 30° sectoral pattern in elevation with a lower edge cutoff rate of 10 dB per degree at the horizontal.

The FDM format used requires a frequency stability of 6 parts per million o[illegible] 1. This has been allocated as 2 p.p.m. to ground, 3.7 p.p.m. to the avionics, and the remaining 0.3 p.p. [illegible] aircraft Doppler shift.

Interference in the FDM format is being evaluated. The worst conditions are expected to be associated with coupling of the EL-1 signal into the AZ signal subchannel when the aircraft is at the GPIP. The next worst case is the interference to BAZ from AZ when an aircraft overflies the AZ array at an altitude as low as 100 feet. The designed protection ratios for these conditions are 73 dB and 63 dB respectively.

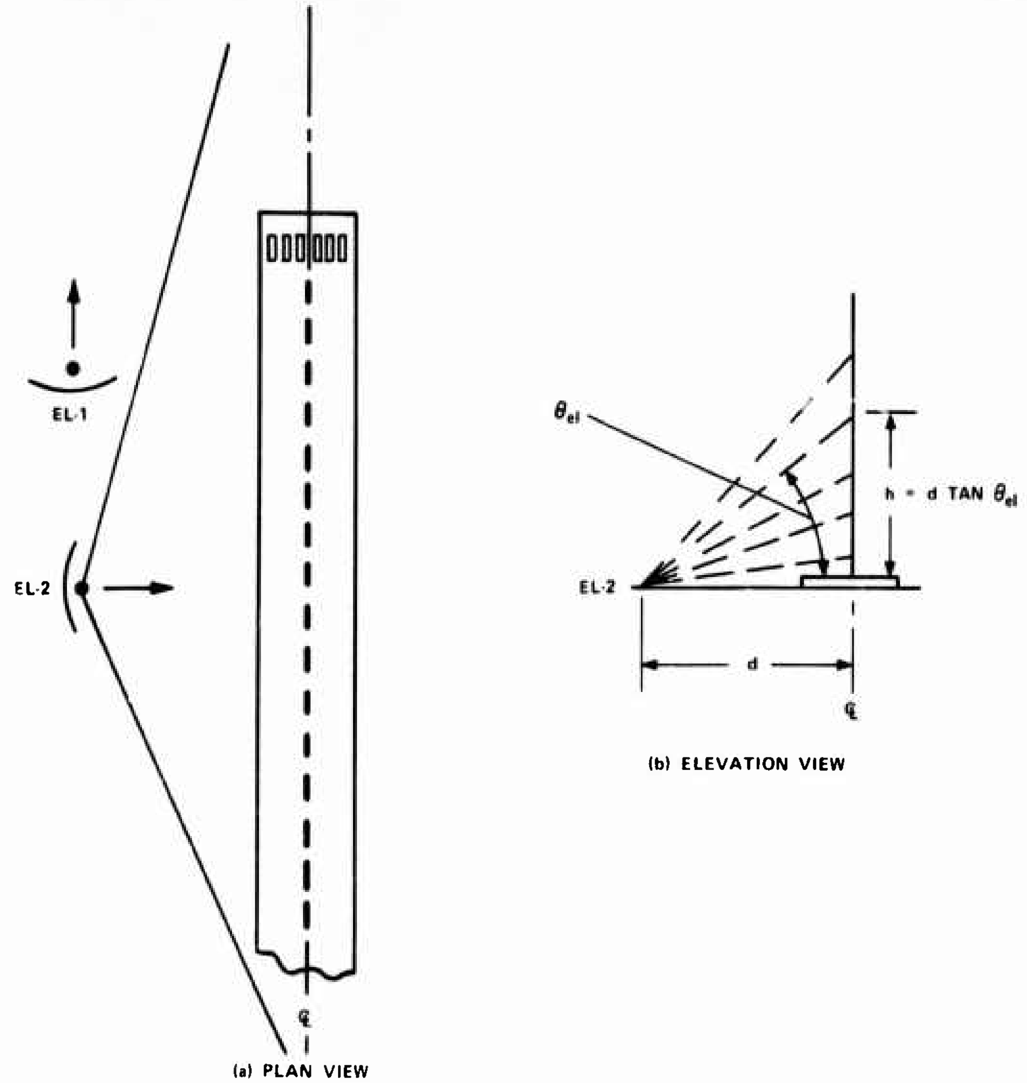


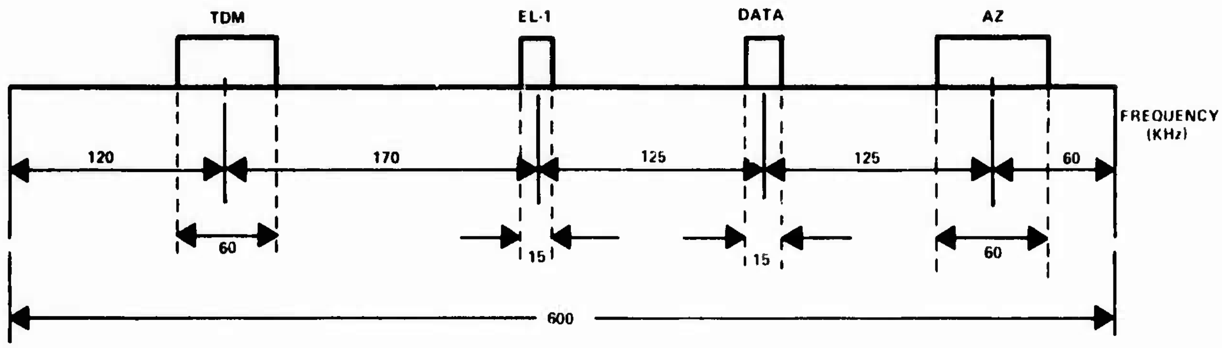
Figure 8.5 INTERSCAN Flare (EL-2)

In the airborne receiver, the FDM channels are separated in the IF. The doppler decoder contains a narrow-band tracking filter with automatic search, acquisition, and validation. The tracking filter is essentially an automatic frequency control (AFC) loop that centers the signal into a fixed narrow-band filter. Figure 8.8 shows such a basic tracking filter, is used with slight variation in both U.K. and U.S. designs.

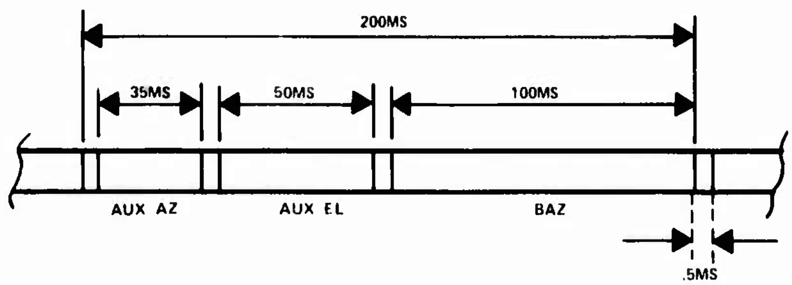
8.2.3 UNITED STATES MLS PROGRAM. The U.S. MLS Program is the most extensive of the international efforts. The U.S. has sponsored the development, design, fabrication, and test of MLS hardware representative of the two basic techniques, i.e., scanning beam (called conventional scanning beam or CSB) and Doppler. A comprehensive test and evaluation program clearly demonstrated that either technique can meet the stringent requirements associated with an all-weather landing system.

Results of this feasibility demonstration phase test and evaluation are being used to narrow down from among two CSB designs and two Doppler designs to a single U.S. MLS technique. The organization and philosophy of the selection process implemented in the U.S. MLS Development Program is described in a companion paper [42]. Detailed data concerning the four designs are presented in a recent progress report on the U.S. MLS Development Program [43]. The U.S. has followed SC-117 recommendations very closely; three of the designs plan to use Ku-band for EL-2, and three designs use TDM. One Doppler design, however, uses a C-band EL-2, and one CSB design uses separate 600-kHz C-band channels for EL-1 and AZ.

The U.S. candidate systems possess many common features. In particular, each fully meets the functional requirements for landing guidance systems as defined by ICAO and as detailed and expanded by the FAA during the evaluation process. The differences concern the operating principles of these technologies and the timing sequences by which the guidance data is provided to the aircraft. The demonstrated operational and functional capabilities of these systems and the timing sequence features of the signal formats are described in the following paragraphs.



(a) C-BAND FDM FORMAT



(b) TIMING OF TDM SUBCHANNEL

Figure 8.6 U.K. Signal Format

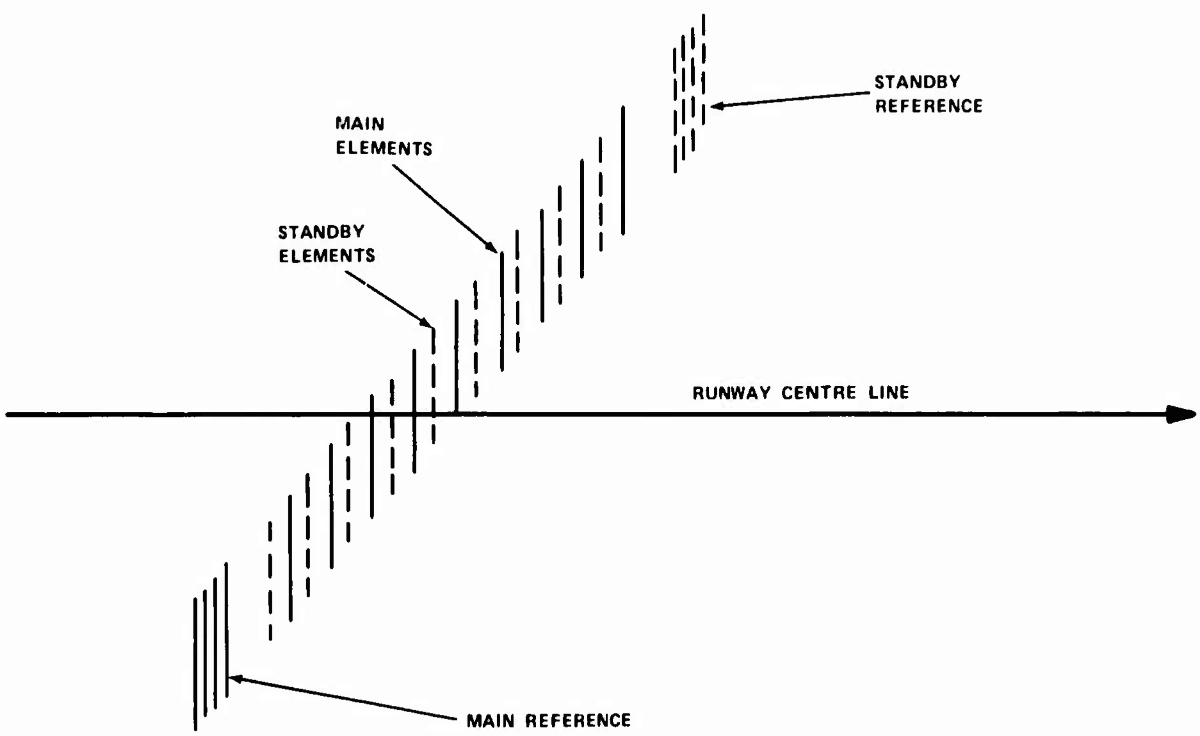


Figure 8.7 U.K. AZ Array Thinning and Redundancy

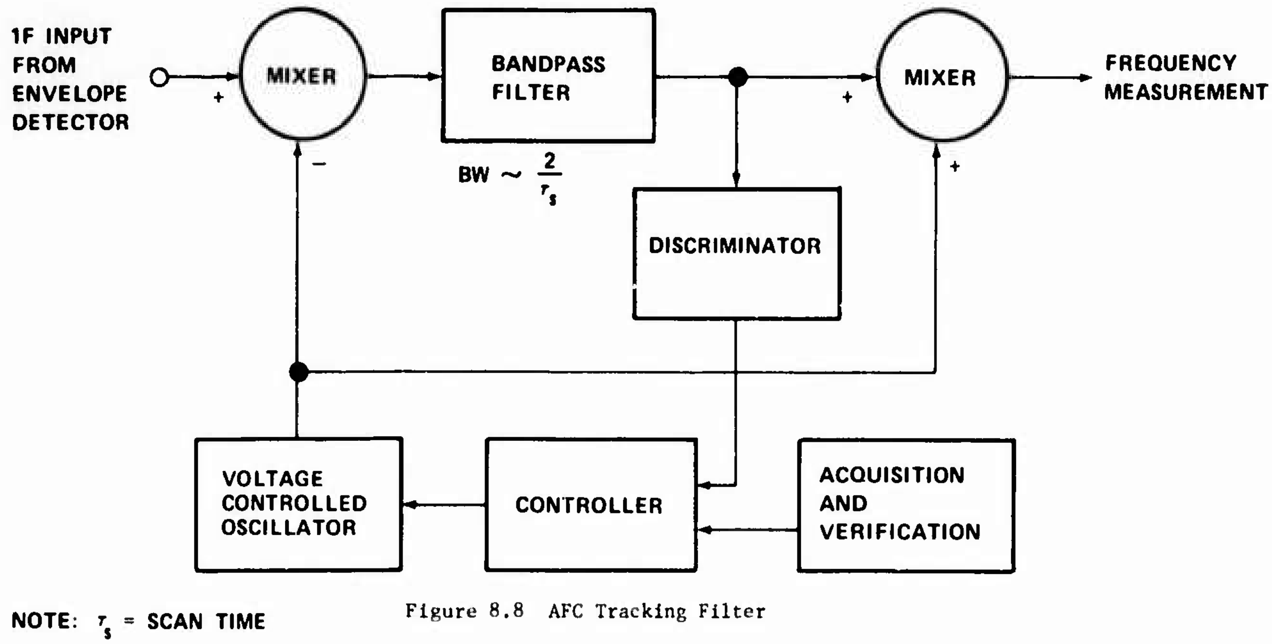


Figure 8.8 AFC Tracking Filter

8.2.3.1 Operational Capabilities. The MLS provides the following operational capabilities in different combinations for various configurations:

- Guidance so that closely-spaced parallel runways may be used to accommodate high-density air traffic.
- Guidance so that offset, segmented, or curved paths may be used as an aid to increased capacity in the terminal area and to aid in noise abatement, and obstruction clearance.
- Guidance so that offset or segmented paths may be used as an aid to wake vortex avoidance.
- Guidance so that a high-integrity CAT IIIb landing can be made available at a broad range of airports.
- Missed-approach and departure guidance to reduce decisions heights and airspace requirements.
- Positional data to aid in advanced metering and spacing.
- Low-cost versions of ground and airborne equipment to provide lower performance landing service at general aviation airports.
- Versions to give landing services for vertical and short takeoff and landing (V/STOL) operations.
- Versions for tactical military operations.
- Versions for operations on aircraft carriers and military shore bases.

These systems are designed so that the simplest airborne MLS equipment configuration can obtain commensurate service from the most complex MLS ground configuration; conversely, the most complex airborne configuration can obtain commensurate service from the simplest MLS ground configuration.

The MLS will use a compatible signal format for both civil and military applications.

8.2.3.2 Classes of Service. Various functional capabilities are required to provide the classes of service necessary to meet the current and foreseeable operational requirements. The functional capabilities are provided in modular form so that services appropriate to differing requirements can be formulated. Requirements specified for each service reflect the range of capability inherent in the MLS technique and signal format. Subsystem functional requirements do not restrict the performance of any particular system but indicate a range and combination of services available to the user. For example, while the azimuth coverage for the full capability azimuth subsystem is specified as ±6Q°, the user can limit that coverage to ±10° or whatever value the particular application might dictate. Conversely, it is possible to obtain wide-angle coverage, ±60°, for the lowest-level accuracy subsystem.

Lateral service available with MLS provides precision guidance throughout the coverage specified for each level of subsystem. The MLS is compatible with the lateral path shapes used within the coverage region (e.g., multipath straight and curved segments). However, the actual path shape for other than simple, straight-in approaches will be determined by the navigation equipment carried in the aircraft, such as a computer not part of the average airborne MLS complement.

In addition to precision azimuth information provided in the same measuring units at all installations, the MLS furnishes a means to adjust the effective course width to provide an acceptable deflection sensitivity at runway threshold regardless of runway length.

Azimuth and DME rollout service are to be provided on the surface of the runway for guidance of the aircraft during deceleration to the initiation of turnoff onto a taxiway.

An elevation subsystem will provide precision guidance for vertical service from 2° to 15°. As with lateral service, the MLS is compatible with the various vertical path shapes usable within this coverage region (e.g. multiple segments of different descent angles). The actual path shape for other than straight descents will be determined by equipment carried in the aircraft, such as a computer not part of the average airborne MLS complement. Vertical position flare guidance (EL-2) is provided from near the runway surface throughout the touchdown zone.

Back azimuth (BAZ) for missed approach or departure lateral guidance will be provided. Distance service (DME) is also provided as part of MLS. All MLS implementations provide runway identification signals in Morse code and signals to identify the MLS functions being transmitted by the ground equipment (e.g., azimuth, elevation plus AUX data such as MLS operational status, azimuth scale factor, or minimum glide slope.

8.2.3.3 Channelization. The MLS operates at C-band, except that flare guidance may operate at Ku-band. An all Ku-band MLS also may be used for some military configurations. The civil channels are "hard-paired" up to three frequencies (DME, angle, and flare guidance) for each MLS installation without overlap. The number of channels reserved allows noninterfering operation in the projected worst-case densities of the foreseeable future.

The MLS provides 200 channels for azimuth and elevation information in the frequency range of 5000 to 5125 MHz (C-band). Two hundred associated channels are provided for distance information in the frequency range of 5125 to 5250 MHz (C-band) and 200 associated channels for flare guidance in the frequency range of 15,400 to 15,700 MHz (Ku-band). Each channel defines all frequencies and DME time-coding associated with an individual ground system installation, which may contain C-band azimuth, elevation, and DME subsystems plus either a Ku-band or C-band flare subsystem. The channel plan and signal format allow a geographical deployment of the MLS adequate to meet reasonable projections of system installations without system-to-system interference.

8.2.3.4 Siting. Representative siting for the MLS ground stations is illustrated by Figure 8.9 for the expanded system. Stations such as flare and back azimuth may be deleted for lower category landing services.

8.2.3.5 Coverage. The MLS provides the near-runway coverage illustrated by Figures 8.10, 8.11, and 8.12 and the long-range coverage shown by Table 8.1. The accuracy abilities associated with this coverage volume are shown by Figure 8.13.

8.2.3.6 Conventional Scanning Beam TRS Signal Format. The basis of the CSB MLS candidate concept is the use of ground-based antennas to generate narrow fan-shaped beams. The optimum CSB technique established by the test and evaluation program is the time reference system (TRS), in which beams are scanned over the volume to be covered contiguously in a clockwise direction and then in a counterclockwise direction, ("to-fro") similar to the Australian system described in paragraph 8.2.1. An airborne receiver located in the coverage region senses the passage of a beam in both directions and decodes its angular position from the time displacement of the two beams. This position is used in deriving guidance signals for display to the pilot or for interface with an autopilot.

All angular functions are separated by time diversity and may use a common frequency channel. By transmitting azimuth, elevation and other information cyclically on the same frequency, an airborne receiver can use the same receiving and decoding circuits to process both, resulting in a substantial cost saving.

Autopilot studies have shown that a data update rate of as low as five scans per second provides an excellent margin of safety over the minimum usable scan rate. However, environmental constraints (multipath) dictate the higher update rates actually employed for optimized performance.

The CSB TRS scan format is shown in Figure 8.14. The vertical guidance data rate is more than twice as great as for lateral guidance, because there is less angular sector to sweep, and higher data rates are required for inbeam multipath averaging.

Specific features of the CSB/TRS system include:

- Coverage - In addition to the coverage demanded by the requirements, the proposed format provides functional time slots for future implementation of 360° azimuth coverage and 0° to 20° back elevation coverage.

- Accuracy - The TRS format will provide better "clean site" accuracy than the FRS formats, by elimination of quantization and thermal noise errors associated with the frequency coding on the beam.

- Multipath rejection - The TRS format will provide improved performance in multipath situations by use of substantially higher data rates, which will provide effective "multipath averaging" and which will allow the use of effective error limiting and outlier rejection techniques.

- Performance - The higher accuracy and data rate provided by the TRS format will result in significant improvements in coupled airframe performance, exhibited by smoother and more precise path following for curved approach and autoland operation.

- Interference resistance - The well confined spectrum of the proposed technique reduces cross-channel interference and interference to other services. The nature of the TRS scanning beam angle measurement process makes it highly resistant to interference from non-MLS sources, such as C-band radars.

- Site Immunity - The standard benefits from CSB MLS of spatial control and use of narrow scanned beams are provided. Techniques for implementing azimuth hopover (or centerline emphasis) with electronically scanned phased arrays are being investigated. The improved multipath rejection via high data rates also increases site immunity by reducing the multipath effects of large nearby structures.

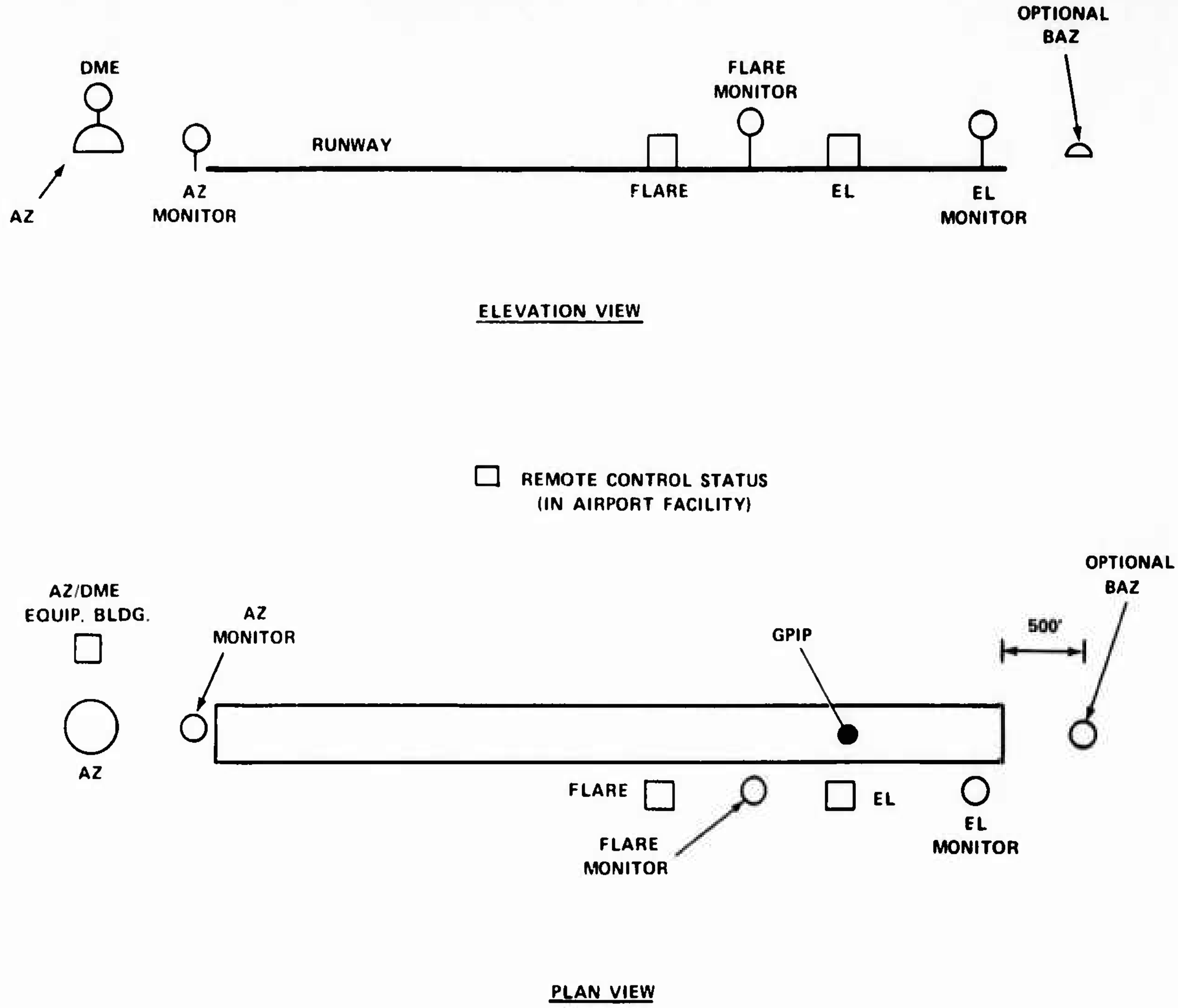


Figure 8.9 U.S. Nominal High Performance Siting

- Hardware realizability - Virtually all types of scanning beam antennas compatible with the proposed format have been demonstrated, including: linear phased arrays and beam-port antennas (Bendix), torus (AMSCAN) antennas and mechanical scan antennas (TI).

- Integrity - The integrity of TRS over FRS is improved by: (a) redundant FCN ID encoded on the omni ID message by disk phase-shift (DPSK) and via a unique keying time relationship established in the signal format between the Barker sync pulse in the omni-ID signal and the midpoint of the "to-fro" scan format for each function (this is a form of "identify on omni/verify-on-beam") (b) use of electronic linear arrays with a high degree of fail-operational capability, (c) use of high data rates for improved motion-multipath averaging to reduce multipath effects, (d) adjustable coverage limits in the ground MLS equipment without resort to a data link, (e) simpler transmitter implementation, which reduces the number of monitored functions and increases the system reliability, and (f) simpler field monitor implementation.

- Monitoring - The simpler TRS system concept proposed will ease the monitoring requirements.

- Channelization - The proposed TRS format fully satisfies the functional requirements for MLS channelization, and in fact offers the possibility for doubling the number of MLS channels or reducing the MLS spectrum occupancy by reducing the channel separation to 300 kHz. Also, the channel plan keeps the C-band or Ku-band option open for flare and military systems; it is arranged to allow modular receiver configuration and use of a common synthesizer, and it places DME at the top of the frequency band to protect the radio astronomy experimental band.

- Rate information - The TRS format provides much higher data rates than earlier recommendations and this plus the improved accuracy will result in substantial improvements in MLS position-derived rate information.

- Special System Considerations - The proposed format allows either C-band or Ku-band implementation for the flare and military systems. It also allows for mechanical scan implementation of the military tactical systems if desired, and the higher data rate (40 Hz for AZ and EL) will improve aircraft carrier performance.

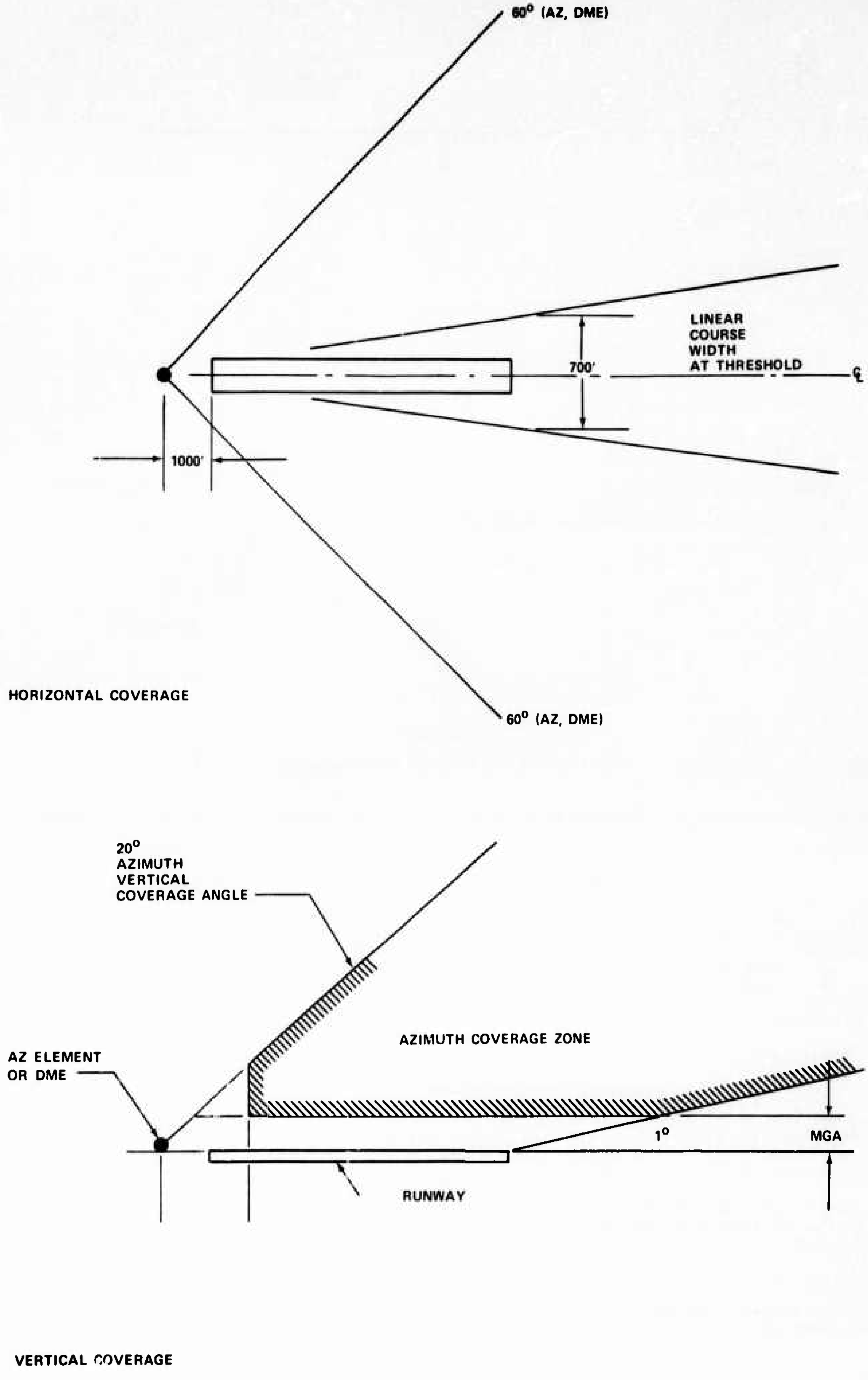


Figure 8.10 Azimuth and DME Subsystem Coverage Near Runway

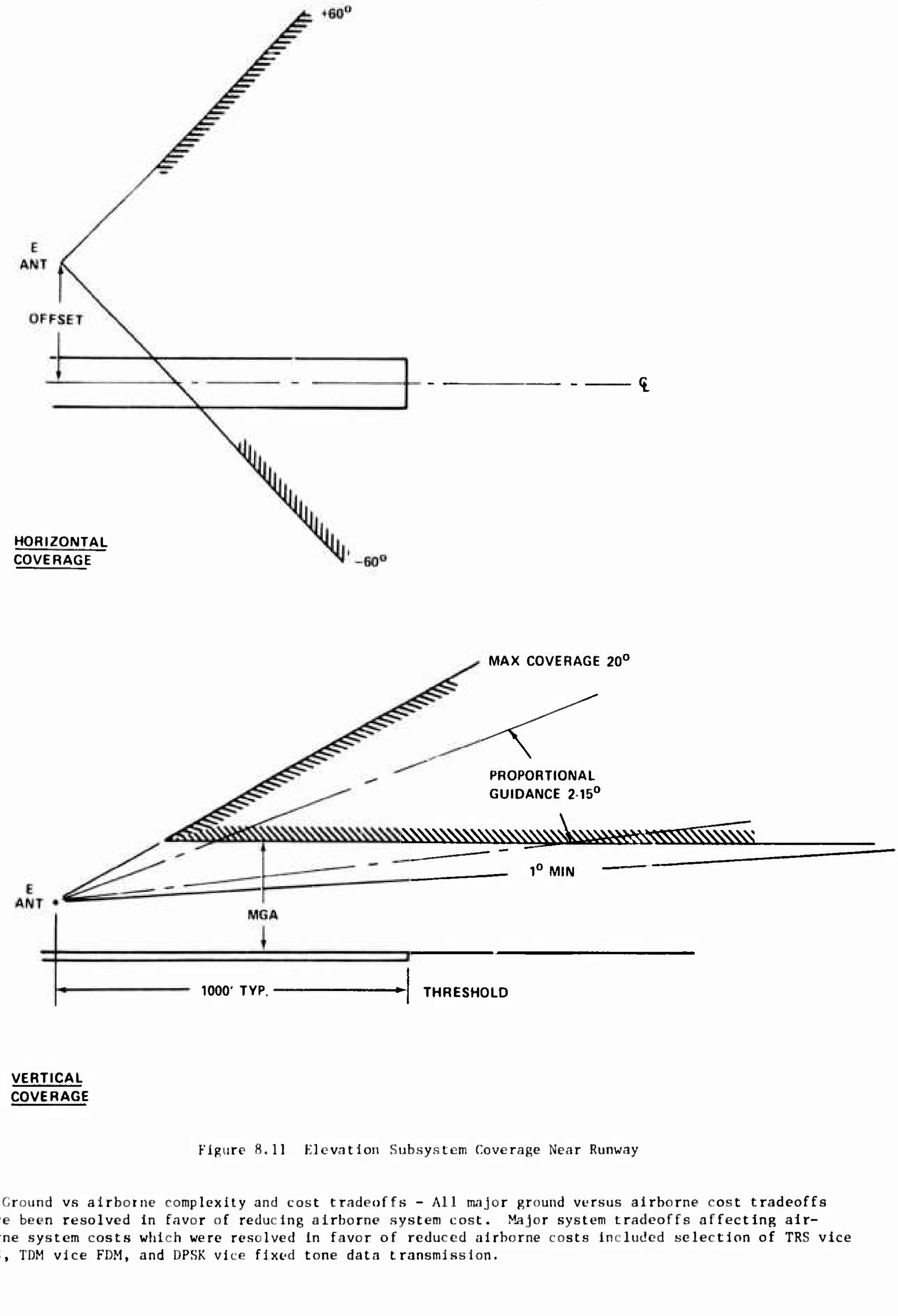
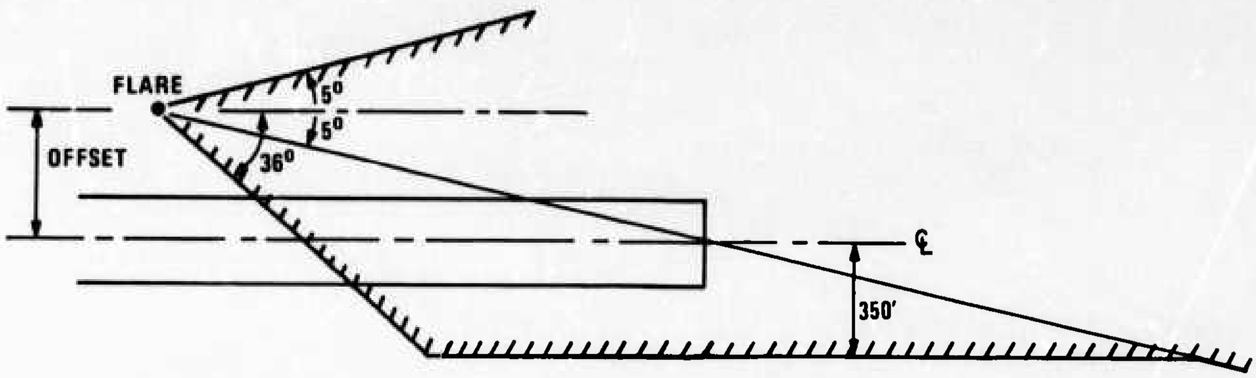
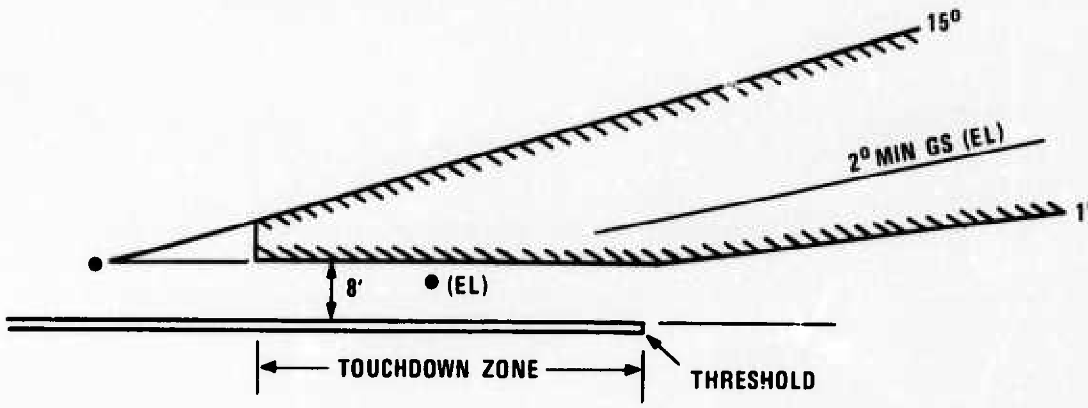


Figure 8.11 Elevation Subsystem Coverage Near Runway

- Ground vs airborne complexity and cost tradeoffs - All major ground versus airborne cost tradeoffs have been resolved in favor of reducing airborne system cost. Major system tradeoffs affecting airborne system costs which were resolved in favor of reduced airborne costs included selection of TRS vice FRS, TDM vice FDM, and DPSK vice fixed tone data transmission.



HORIZONTAL COVERAGE



VERTICAL COVERAGE

Figure 8.12 Flare Subsystem Coverage Near Touchdown

TABLE 8.1 LONG-RANGE COVERAGE ACCURACY ABILITIES OF MLS GROUND FUNCTIONS

Functional Element	Coverage Width	Altitude (1000 ft.)	Range (nmi)	Horizontal Coverage (deg)	Proportional Horizontal Guidance (deg)	Vertical Coverage (deg)	Range of Glideslopes (deg)
Azimuth	W	20	20	±60*	±60	1 - 20**	-
	M	20	20	±40	±40	1 - 20**	-
	N	20	20	±35	±10	1 - 15**	-
Back Course	M	5	5	±40	±40	1 - 20**	-
Elevation	W	20	20	±60	-	1 - 20	2 - 15
	M	20	20	±40	-	1 - 20	2 - 15
	N	20	20	±10	-	1 - 15**	2 - 8
Flare	-	2	5	±5	-	2 - 15***	-
DME	W	20	20	±60	-	1 - 20**	-
	M	20	20	±40	-	1 - 15**	-
	N	20	20	±10	-	1 - 15**	-
DME (BC)	M	5	5	±40	-	1 - 20	-

*Coverage clearance extends beyond the proportional coverage region.
**Azimuth and DME information is provided from 1° or the obstacle clearance plane, whichever is greater.
***Flare guidance is provided at low angles whenever elevation guidance is also provided at low angles.

W - Wide
M - Medium
N - Narrow

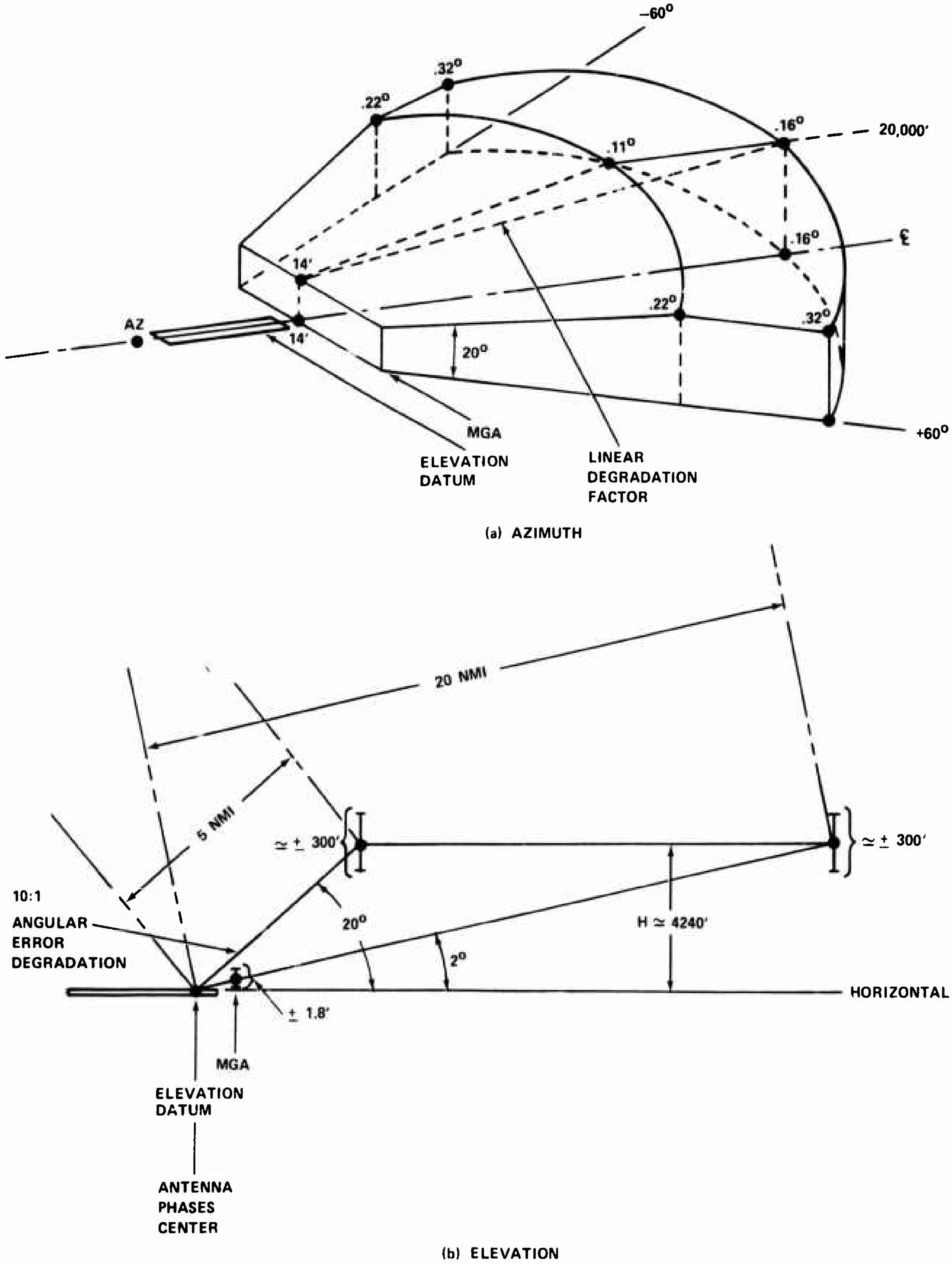


Figure 8.13 Allowable Degradation Characteristics for CAT-III Subsystems

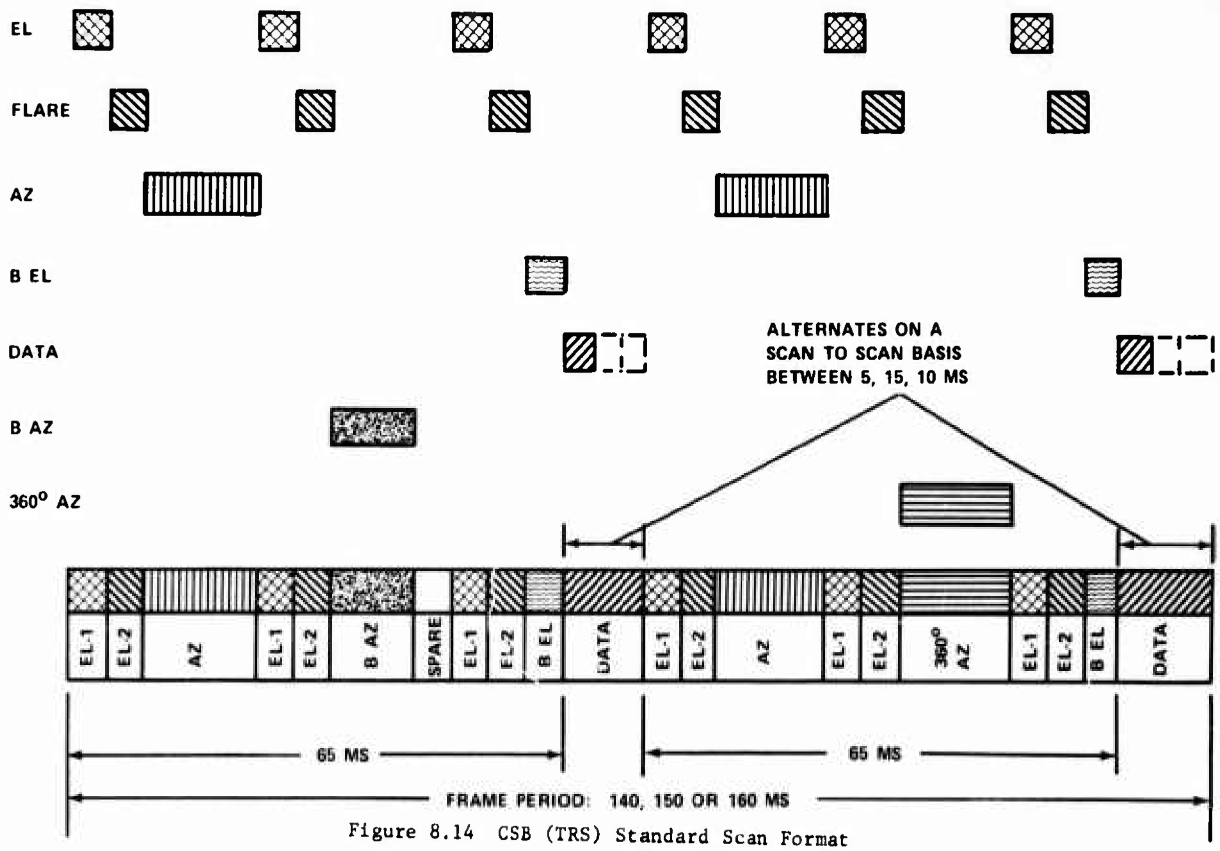


Figure 8.14 CSB (TRS) Standard Scan Format

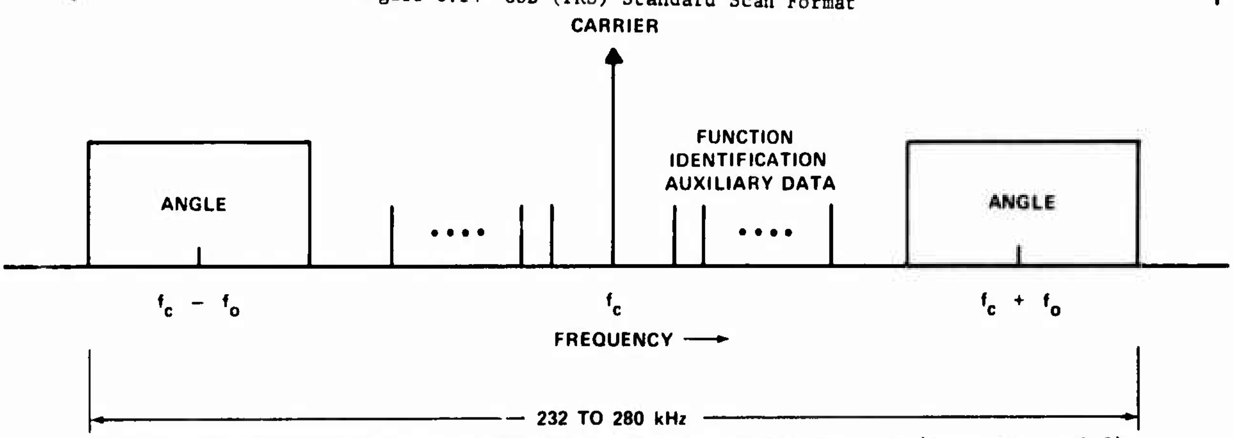


Figure 8.15 RF Spectrun of CSB and Doppler for Single Function (e.g., AZ or El-2)

8.2.3.7 Doppler Signal Format. In the Doppler MLS candidate, the frequency coding is inherent in the Doppler effect. The RF spectrum is a baseband signal superimposed as a double sideband on the carrier as shown in Figure 8.15. The RF spectrum results from the addition of the signals in space, that is, one transmitter feeds a reference antenna to radiate the Doppler reference carrier and the AUX data, while a separate transmitter feeds the Doppler array to radiate the baseband angle signal, shifted in frequency to f_c+f_o and f_c-f_o as shown in Figure 8.16. The two sidebands in the Doppler signal are radiated inphase with the carrier, resulting in an amplitude-modulated signal that can be demodulated with an envelope detector. This signal structure applies tto the various functions of azimuth, elevation, and flare. The largest fraction of the 200-ms period is allocated to azimuth, as shown in Figure 8.17. The flare function is transmitted twice, resulting in a 10-Hz data rate for this function. Guard bands are positioned between each function, and additionally, an identification signal is transmitted with each function so that the timing sequence need not be "hard wired" into the avionics. Although EL-2 is on Ku-band, it is time multiplexed with C-band angle functions in order to use common IF's and angle decoders in the aircraft.

One basic difference between the CSB and Doppler signals is their time duration. The passage of the CSB signal produces a relatively short pulse of approximately 1 ms, while the Doppler signal duration is 10 ms to 70 ms, depending on the angle function. However, the single-scan duration of the Doppler is of the same order as that for the CSB received signal, such that in a given function period the Doppler AZ designs provide up to 26 scans and about half this number for other functions. The Doppler signal format is essentially a pure form of time multiplex among all angle functions and auxiliary services within a 600-kHz channel bandwidth. Two basic frame lengths are used: 200 ms for CAT III systems and those systems featuring ECONAV, an economical 360° azimuth navigation unit suitable for multiple runways, and 100 ms for basic guidance and aircraft carrier landing applications.

Data update rates of 5 per second or more are available from all functional elements except EL-2, which has 10 per second, and the ECONAV, which has a minimum of 1 per second. The format provides for AZ and EL with a 10-Hz update rate, together with a high-capacity AUX data rate for application to aircraft carriers.

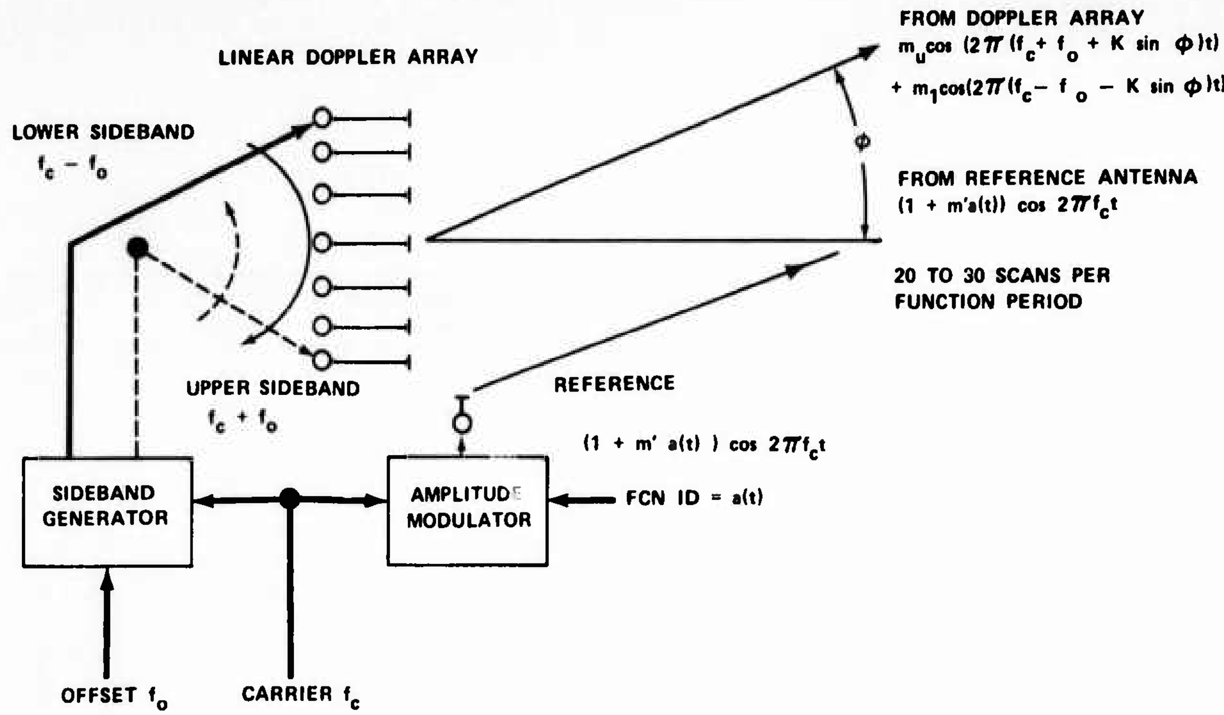


Figure 8.16 Doppler Signal Generation

The carrier frequency is placed in the center of the allocated 600 kHz with an angle sideband spaced 100 kHz from it. The sideband changes position above and below the carrier with each scan. The apparent Doppler shift as experienced by the receiver extends 25 kHz on either side of the sideband over the range of azimuth coverage. Function ID, AUX data, and out-of-coverage indication (OCI) are modulated on a 50-kHz sideband time multiplexed with angle.

The format is fully compatible with various airborne processing systems, including:

- Full time-slot (multiscan) counting
- Scan-at-a-time (uniscan) counting
- Use of fixed sector filters
- Angle tracking filters

The format is fully compatible with various Doppler ground antenna implementations, including:

- Basic commutated array
- Commutated reference array

The Doppler signal format provides for various combinations of the following functions by assigning a time slot to each function:

- AZ or EL-1 guidance elements operating at C-band, suitable for the full range of capabilities.
- DME elements operating at a separate portion of C-band with separate but correlated channels.
- An EL-2 element for flare guidance to touchdown on paired channels at Ku-band or C-band.
- BAZ and optional BEL guidance elements for missed approach and departure service, operating at C-band.
- An azimuth guidance element having 360° coverage for navigation guidance (ECONAV), operating at C-band.
- Optional Ku-band AZ and EL guidance elements for tactical applications.
- Growth capability for future optional functions.

These functions are available on a single-channel assignment with the following limitations:

- Full-capability configurations for hub airports can have any combination of AZ, EL-1, DME, EL-2, BAZ, and BEL.
- Configurations for nonhub airports can have any combination of functions AZ, EL, DME, and ECONAV.

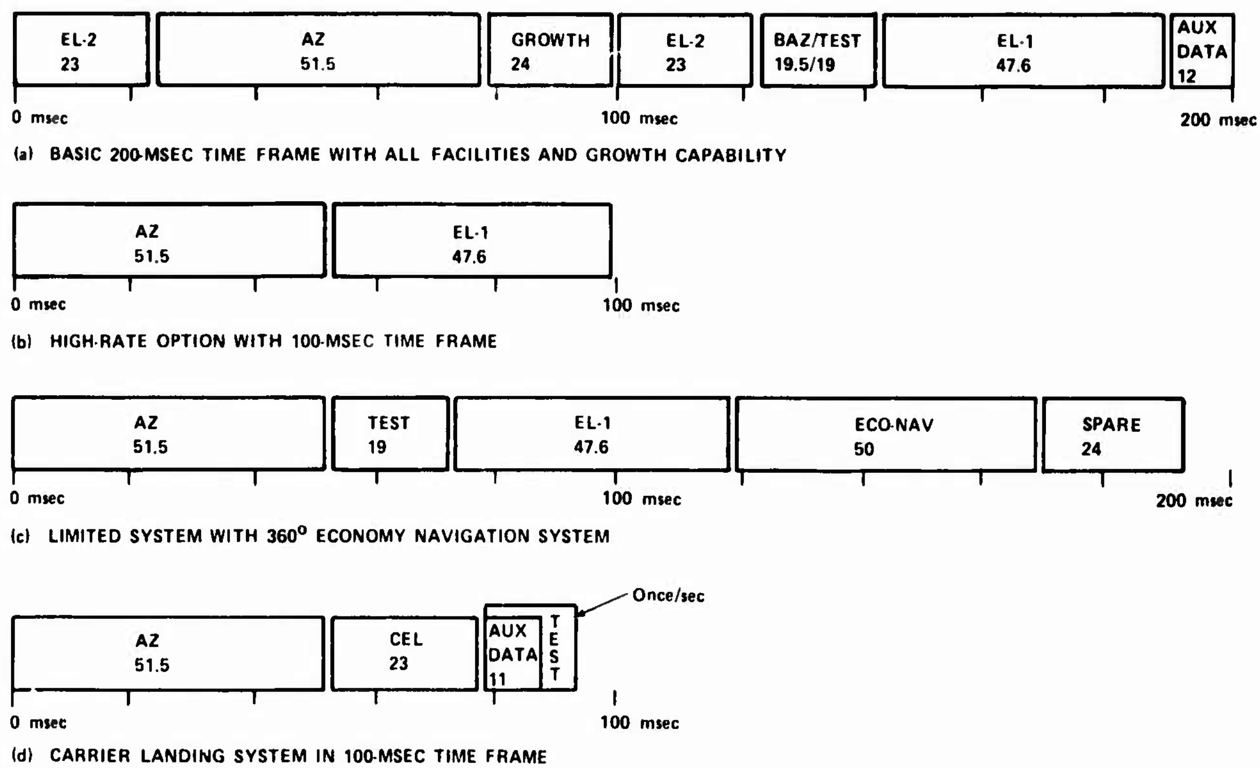


Figure 8.17 Doppler Scan Format

- Tactical configurations can have combinations of functions AZ, EL, and DME.
- The ECONAV is also available on a separate channel.

The channel plan is generally subdivided as follows:

- Basic angle services of 5000 to 5124 MHz
- DME air-to-ground service of 5188 to 5250 MHz
- Flare and tactical service available at 15,408 to 15,588 MHz
- Reference diversity
- Beamport antenna
- Modal antenna

Doppler arrays are usually of two generic types, commutated and beamport. The commutated antennas are scanned bidirectionally. The beamports generate an identical signal in space through the use of phase shifting techniques. The two types of Doppler arrays are fully interoperable.

8.3 GROUND-DERIVED MLS

The French and F.R.G. MLS designs are both ground-derived systems [44]. Figure 8.18 shows the common features of these systems. The position of the aircraft is measured from the ground and then transmitted back via an appropriate data link. Such a system is also capable of providing aircraft position directly to ATC for surveillance functions, as indicated in the figure.

Ground-derived systems can use various signal design and processing techniques. Systems differ mostly upon the user access procedure employed, either time-ordered access (the French approach) or random access (the F.R.G. approach). Time-ordered access systems invoke scheduling by the ground station, wherein each aircraft is individually addressed in turn. Range is used as the criteria for sequencing users via the data uplink. Random-access systems work on a cycle initiated by each individual aircraft (similar to present DME) and not synchronized to other aircraft using the same facility. Thus, aircraft interrogations occur randomly at the ground receiver. The ground station retransmits each of these with a known fixed time delay. Each aircraft receives all retransmissions, occurring randomly, out of which it selects its own reply, after an initial acquisition process.

In either approach, on receipt of the aircraft signal the ground system measures the azimuth and elevation angles. These measurements are then transmitted via the data uplink or on the interrogation reply. Onboard the aircraft, the uplinked angles are read, and the timing of uplink is used to measure the two-way range of the aircraft.

The F.R.G. ground-derived angular measurement technique uses interferometer processing methods. A reception technique that is the inverse of the Doppler signal transmission concept of air-derived systems may also be used. At this stage, France has not determined which processing technique it will use.

The range measurement in both approaches is based on two-way range, and it is an integral part of the angle system.

8.3.1 FRENCH MLS CONCEPT. The French MLS program is part of an integrated ATC surveillance, navigation, and landing system and is in the conceptual stage [45]. The design and experimental work has been mainlv on an ATC L-band data link. The concept is to transmit a narrow-band CW-PSK modulation on a C-band downlink that serves as the signal for angle measurement. For the uplink, L-band channels within the 1540-to 1650-MHz band are advocated. Two hundred channels with a total width of 250 MHz are being considered for the C-band link.

Figure 8.19 shows the uplink and downlink message formats, each 80 bits long with a duration of 4 ms. Each aircraft has a unique address, which enables it to identify the message directed to it. The uplink also contains the address of the next aircraft to be interrogated by the ground station. The position information derived by the ground may be transmitted in either polar coordinates or Cartesian coordinates. Frame check sequence (FCS) is included in the uplink for validation.

The ground station initiates commands to the aircraft through the uplink; the aircraft replies in the requested mode. Thus, the aircraft acts as a transponder for range measurement. Angular measurements are

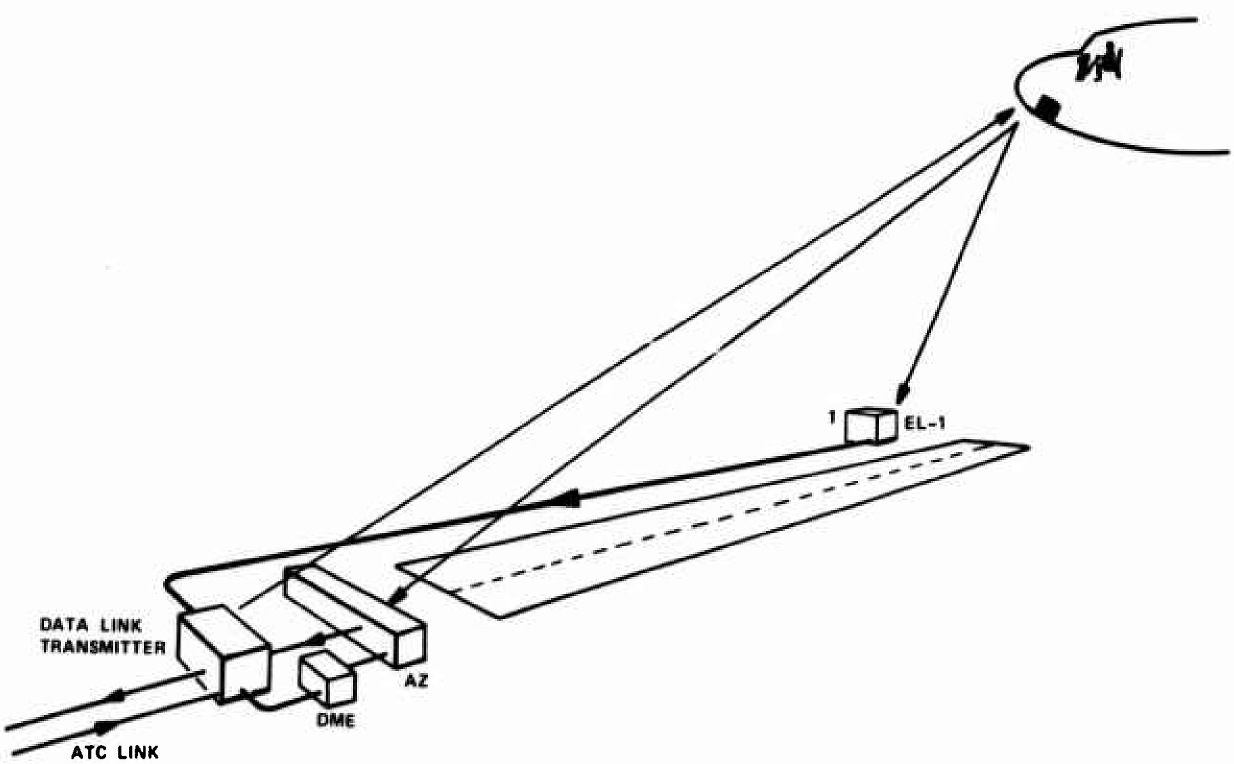


Figure 8.18 Ground-Derived MLS

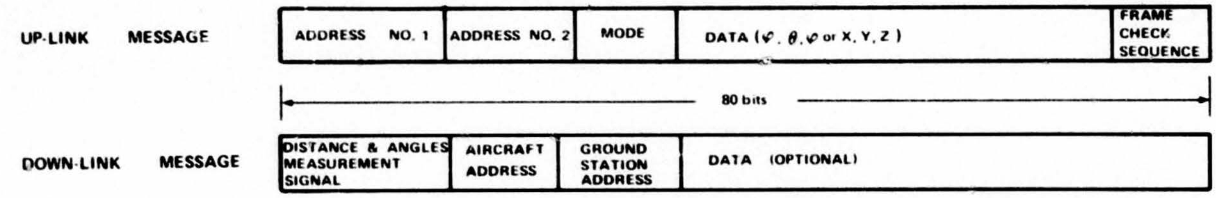


Figure 8.19 French MLS Message Format

derived from the downlink signals received by the azimuth and elevation sensor arrays. The computed angles are uplinked to the aircraft 20 ms after the initial request by the ground.

The first part of the downlink message of Figure 8.19 is used for angle measurement. This is followed by the aircraft and ground station addresses. The optional data part of the message may be used for requested ATC services or to respond to ATC ground-to-air messages. A bit rate of 19.2 kbps is used on both uplink and downlink. In the integrated surveillance and landing system, France proposes a second uplink for ATC messages. The ATC uplink would also have a data rate of 19.2 kbps. The two uplinks are TDM on the same frequency channel, resulting in a net data rate of 38.4 kbps.

Forty-eight aircraft can be simultaneously handled by the French design at a 5-Hz data rate. Moie aircraft can be accommodated at lower data rates. A variable mix of data rates may be used.

As discussed briefly, aircraft acquisition or access to the system is an important design issue. Of the two versions being considered, the integrated version uses the ATC data link, while the other simpler version uses a single transmitter aboard the aircraft as a transponder beacon for landing guidance as well as for other ATC and navigation purposes. In the integrated version, each aircraft requests service from ATC on the downlink. An address is uplinked to the aircraft via the ATC data link indicating the assigned data rate and the systematic delay scheduled between the ground response and the next aircraft transmission. The aircraft is thus engaged. In the simple version, the aircraft listens in on the ground transmissions to intercept a vacant time slot, and then uses the available address. In the next ground interrogation it replies on the received address. The time slot is then declared engaged by the ground station.

Preliminary ground equipment characteristics include AZ sensor apertures of 6 to 12 feet and an EL-1 height of 6 feet. The horizontal coverage for both is ±45°, the vertical coverage for both is 0 to 20°. The design range is 30 miles. No EL-2 is being considered; continued use of radio altimeter for flare is proposed instead. Range accuracy of the system is expected to be about ±50 feet (1 σ). This is expected to be achievable with a 200-KHz uplink and a 400-KHz downlink assigned to the previously discussed data format.

8.3.2 F.R.G. DLS. The F.R.G. ground-derived system is called DLS (DME based Landing System). The system originated from the military SETAC and ORTAC programs [45]. Although ground derived, the system philosophy is for the pilot to maintain responsibility, with the ground station serving only as an automatic signal processor and transponder.

DME is the main ICAO ranging aid used for enroute navigation [46]. It is also part of the military TACAN system, which, in addition to range, supplies bearing information. Since there are 20 DME channels assigned for ILS, the F.R.G. proposes using these together with 19 other channels available for national use in each country. These 39 channels will be doubled when the new 50-kHz spacing is introduced, resulting in 78 available channels versus the 200 channels proposed by SC-117.

The airborne part of the DME, called an interrogator, is a transmitter and receiver. The ground station operates as a transponder. The ranging waveform is a pair of 3.6 μs pulses; channel bandwidths are 1 MHz. The aircraft interrogation rate is random. The ground transponder is simply a repeater that returns the received waveform, albeit with a frequency offset of 63 MHz and a time delay of 50 μs.

The DLS system is essentially an add-on to the existing DME, both interrogator and transponder. On the ground, AZ and EL sensor arrays, shown in Figure 8.18, are installed in addition to the omnidirectional DME antenna. The interrogation waveform received by these arrays is processed to compute the angles of the interrogator. The computed AZ, EL-1, and EL-2 angles are retransmitted by the transponder in a pulse position code by adding three additional pulse pairs to every response.

The airborne receiver acquires the nominal DME relay with the standard DME search procedure [46]. Following that, the airborne unit acquires the angle pulses. The basic DLS waveforms are shown in Figure 8.20. The first pulse-pair is the normal DME reply. The second pulse-pair is transmitted after a delay, T_{AZ}, which is proportional to the received azimuth angle. The delay between the first and third pulse-pairs, T_{EL-1}, is proportional to EL-1 elevation angle. Finally, the EL-2 flare signal is sent on a fourth pulse-pair.

In standard DME, the RF phase of the two pulses is not used for modulation. In DLS, the phase of the second pulse is coded into four levels, ±90° and ±180°. This differential phase coding is being considered for general data transmission for both uplink and downlink and also for station address. The ground-station address capability allows cochannel operation.

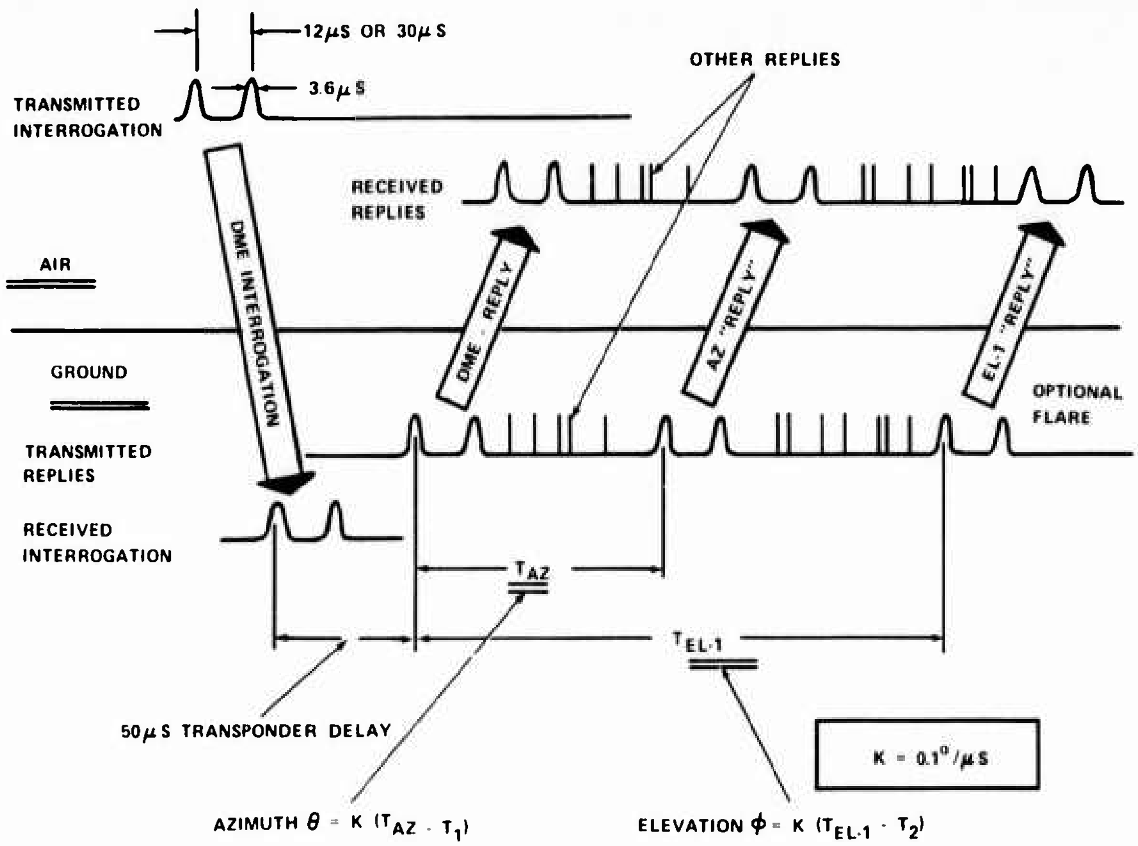


Figure 8.20 F.R.G. DLS Signals

For AZ, either a circular ground array with 360-degree coverage or two linear arrays at right angles is used. For EL-1 and EL-2, a collocated horizontal and vertical array is used. The horizontal array is similar to the AZ array. The signals from these two arrays are combined to give the EL-1 angle. For flare, the F.R.G. follows the Australian concept of Figure 8.5, deriving the flare angle from the same array as EL-1. The DLS antenna apertures are relatively smaller than the other designs. At L-band, both the horizontal and vertical apertures are 10 wavelengths. DLS designers claim the resulting accuracy against multipath will be better than ILS by a factor of ten. Further experimental work is being done to validate these predictions.

REFERENCES

1. Aircraft Radio Laboratory; "Instrument Landing System Bibliography;" Wright Field; Dayton, Ohio; 1919-1941.

2. Henry I. Metz; "A Survey of Instrument Approach Systems In the United States;" IRE Transactions on Aeronautical and Navigational Electronics; Vol. ANE-6; June 1959; No. 2; P. 78.

3. Robert M. Richardson; "Avionics Newsletter;" Litton Industries; No. 7; February 27, 1968; P. 119.

4. R. W. Howard; "Automatic Flight Controls in Fixed Wing Aircraft, The First 100 Years;" Aeronautical Journal (UK); November 1973.

5. E. A. Cutrell; "Operational Flight Testing of Early Instrument Landing Systems;" IRE Transactions on Aeronautical and Navigational Electronics; Vol. ANE-6; June 1959; No. 2; P. 67.

6. J. L. Anast; "All Weather Landing;" IRE Transactions on Aeronautical and Navigational Electronics; Vol. ANE-6; June 1959; No. 2; P. 75.

7. Frank G. Kear; "Instrument Landing and the National Bureau of Standards;" IRE Transactions on Aeronautical and Navigational Electronics; Vol. ANE-6; June 1959; No. 2; P. 61.

8. W. E. Jackson, et al; "Federal Airways Systems;" "Instrument Landing Systems;" Pp. 258-279.

9. Chester Porterfield; "Ground Controlled Approach -- Development and Early Operational Use;" IRE Transactions on Aeronautical and Navigational Electronics; Vol. ANE-6; June 1959; No. 2; P. 71.

10. Group Capt. J. A. McDonald, RAF; "Aircraft Automatic Landing;" Journal of the International Aviation Service of the British Petroleum Company, Ltd.; No. 17; Britannic House; Finsbury Circus, London, E.C. 2.

11. Air Force Navy Civil Landing Aids Experiment Station; "Landing Aids Experiment Station;" Final Report; Arcata, California; 1949.

12. Frank B. Brady; "All Weather Aircraft Landing," Scientific American; March 1964; Vol. 210; No. 3.

13. U. S. Federal Aviation Administration; "United States Standard for Terminal Instrument Approach Procedures (TERPS);" Second Edition; February 1970.

14. Radio Technical Commission on Aeronautics; Report of Special Committee 79 Paper "Standard Performance Criteria for Autopilot and Autopilot Couplers;" March 1963 31-63/DO-118.

15. International Civil Aviation Organization; "Aeronautical Telecommunications;" Annex 10.

16. U. S. Federal Aviation Administration; "Instrument Landing System (ILS);" September 1971, Localizer Manual FV302; Catalog No. 40206; FAA Aeronautical Center.

17. U. S. Federal Aviation Administration; "Instrument Landing System (ILS);" October 1970; Glide Slope Manual FV-301; Catalog No. 40206; FAA Aeronautical Center.

18. Frank B. Brady, et al; "Blanketing and Interference Between Aircraft on Army Air Forces Instrument Approach Systems;" July 1945; Operational Test Report No. 2; Radio and Radar Section; Wright Field.

19. J. J. Battistelli; "The Conventional ILS -- So What's New?"; Journal of the Institute of Navigation; Summer 1974; Vol. 21; No. 2; P. 170.

20. G. Chine, L. Jordan, D. Kahn and S. Morin; "Instrument Landing Performance Prediction;" January 1974; Report No. FAA RD-73-200.

21. (Anon.); 'Description of the Alford Localizer Arrays;" October 1971.

22. Clinton G. Hollins; "Design Plan For A Medium Aperture Localizer Antenna;" March 1974, Prepared under FAA Contract DOT-FA73WA-3291.

23. William E. Kennedy, Richard H. McFarland; "Environmental Study of the 200 Foot Aperture Slotted Cable Localizer Antenna Array;" April 1974; Report No. FAA-RD-74-94.

24. Richard H. McFarland; "Investigations To Provide Improved Glide Slope Operation During Periods of Ground Plane Snow Cover;" April 1974; Vol. 1; Report No. FAA-RD-74-69, 1.

25. J. J. Battistelli and R. H. McFarland; "Investigations To Provide Improved Glide Slope Operation During Periods of Ground Plane Snow Cover;" 1974; Vol. II; Report No. FAA-RD-74-69,11.

26. Ole P. Hakonsen; "Improved Antenna Systems For ILS Localizers;" IEEE Transactions on Aerospace and Electronic Systems; Vol. AES-8; No. 3; May 1972; P. 273.

27. R. A. Moore, G. Moussally, T. Parker and S. F. Payer; "Instrument Landing System Glide Slope Wave Guide Antenna Modification and Test;" June 1973; FAA Report No. FAA-RD-73-78.

28. R. A. Moore, G. Mousally, T. Parker and S. F. Payer; "Analysis of Instrument Landing System Glide Slope Broadside Antenna;" August 1972; Report No. FAA-RD-72-139.

29. C. B. Watts, Jr. and Clinton G. Hollins; "Two Antenna End-Fire Slotted-Cable Glide Slope;" October 1972; Interim Engineering Report No. 1.

30. ICAO International Standards and Recommended Practices; "Aerodromes;" Annex 14; Sixth Edition; September 1971.

31. U. S. Federal Aviation Administration; "Basic Flight Manual and ATC Procedures;" FAA Airman's Information Manual Part I; February 1973.

32. Hugh W. Cowin; "MLS and the Industry;" Flight International; 26 July 1973; P. 135.

33. A. Brodzinsky; "Electronic Landing Aids for Carrier Aircraft;" IRE Transactions on Aeronautical and Navigational Electronics Vol. ANE-6; June 1959; No. 2; P. 95.

34. Myron Kayton and Walter Fried, et al; "Avionics Navigation Systems;" John Wiley and Sons, Inc.; New York; 1969.

35. William C. Reed; "U. S. Selects an Interim Standard Microwave Landing System (ISMLS);" Draft Paper; unpublished.

36. Klaus Dieter Eckert and Hans Joachim Roper; "SETAC -- A New Approach and Landing System;" Interavia, February 1970; Pp. 189-191.

37. N. J. Cafarelli, "Landing Guidance Systems;" Draft Report; Vitro Laboratories; April 24, 1974.

38. S. Ahmed Mear and Stanley R. Jones; The Mitre Corporation; "International and U. S. Design Proposals for a Microwave Landing System;" September 1974; Report No. M74-87.

39. U. S. Federal Aviation Administration; MLS Program Office (ARD-700); "MLS Technique Selection;" December 20, 1974.

40. L. L. Sanders and V. J. Fritch, Jr.; "Instrument Landing Systems;" IEEE Transactions on Communications; Vol. COM-21; No. 5; May 1973; Pp. 435-453.

41. Radio Technical Commission for Aeronautics; Special Committee 117 "A New Guidance System for Approach and Landing;" December 1970; Document No. 14.

42. U. S. Federal Aviation Administration; "Progress Report on the U. S. MLS Program;" Systems Research and Development Service; MLS Office.

43. Gene Jensen; "Involving the Expert and Aviation Community in the Decision Making Structure of the U. S. MLS Program;" Proceedings of 1974 IEEE Conference on Systems, Man and Cybernetics; Dallas Texas; October 2-4.

44. Alexander B. Winick; "United States Landing System (MLS) Development Program;" Presented at Joint USSR/US Symposium on Aeronautical Technology; Moscow; July 23-27, 1973.

45. U. S. Federal Aviation Administration; "International Microwave Landing System (MLS) Symposium;" System Research and Development Service; April 1974; Report No. FAA-RD-74-56.

46. S. Ahmed Meer; The Mitre Corporation; "Study of the VORTAC System and its Growth Potential;" November 1973; MTR-6547 (FAA-EM-74-9).

Preceding page blank

PART IV-D – SURVEILLANCE

Preceding page blank

DIGITAL RADAR DATA PROCESSING FOR ENROUTE AIR TRAFFIC CONTROL

by

John A.Varela

The MITRE Corporation
c/o FFA, NAFEC
Atlantic City, New Jersey 08405
USA

TABLE OF CONTENTS

DIGITAL RADAR DATA PROCESSING FOR ENROUTE AIR TRAFFIC CONTROL

John A. Varela
The MITRE Corporation
c/o FAA, NAFEC
Atlantic City, New Jersey
08405

SUMMARY

The Federal Aviation Administration of the United States is bringing digital Radar Data Processing (RDP) into operation in its twenty Air Route Traffic Control Centers (ARTCCs) in the contiguous states. More than half of the Centers are operational with RDP now, and all are scheduled to be operational by fall of 1975. The system has been extensively tested, its performance measured, and standards established for system performance. This paper describes the functions of primary and secondary surveillance radar target detection, and the processing of target data in the central computer at the ARTCC. The latter includes filtering the data, conversion to common coordinates, correlation with automatic tracks, track smoothing and prediction in several modes, and measurement of data quality in real time. Data on measured system performance are given, and the paper concludes with a short description of current efforts to improve system performance.

INTRODUCTION

Figure 1 is a simplified block diagram of the RDP system. Video data from pre-existing types of long-range (and one short-range) primary and secondary surveillance radars are input to the Common Digitizer (CD),[1] which processes the input video and transmits detected targets in digital messages to one or more ARTCCs and, from some sites, to military users. The Weather and Fixed Map Unit detects the outlines of weather cells at two levels of intensity and gives them to the CD for transmission, and generates signals to the CD to cause blanking in manually selected areas.

The computer in the ARTCC, an IBM model 9020A or 9020D,[2]accepts digital data from many radars, up to a design limit of fourteen. The computer converts the data to a common coordinate system, monitors the quality of the input data, automatically tracks data of interest to Air Traffic Control, responds to controller entries, and sends the combined data to the display system.

Testing of this system began in 1969, with an early Model of the enroute ATC system. This Model, which used an interim display system, was installed at the National Aviation Facilities Experimental Center and at the Jacksonville, Florida, ARTCC. The RDP system was subjected to extensive testing [3, 4, 5] and, with refinements, became the national system that is now nearing completion. The test methodology for implementation of the national system [6, 7] was developed from the Jacksonville experience.

This paper describes the functions performed in the Common Digitizer and in the computer. The performance of digital target detection and the computer processing, as measured on operational equipments, is given following each functional description. It is assumed the reader is already familiar with Primary and Secondary Surveillance Radar operation and terminology.

TARGET DETECTION

There are two models of Common Digitizer: the AN/FYQ-47 and AN/FYQ-49. (There is also one Common Digitizer designated AN/FYQ-56, a modified AN/FYQ-49 that processes data from the one short-range radar in the enroute ATC system.) The AN/FYQ-49 contains all the basic aircraft target detection and reporting functions, and is used by the FAA at FAA-only radar sites. The AN/FYQ-47 contains additional functions for military use that will not be described here. The AN/FYQ-49 is organized into the following equipment groups:

Azimuth, Range, and Timing Group: provides basic range and azimuth data. As normally configured, it accepts azimuth change and reference pulses from the antenna pedestal, accepts radar pretriggers from the PSR, and generates the interrogation triggers for the SSR.

Radar Quantizer Group: contains two units that accept PSR analog video and convert it to digital pulses for further processing.

Beacon Reply Group: accepts SSR analog video, quantizes it, and detects and processes the reply codes.

Target Detection Group: performs statistical target detection on data supplied by the Radar Quantizer and Beacon Reply Groups.

Target Processing Group: controls the processing of data for detected targets.

Output Buffer Group: selects and formats output messages and buffers them in core memory for output.

Memory Control Group: contains the core memory.

Test Target Generator, Map Outline Generator, Performance Monitor Group, and Simplex Patching Group: self-descriptive.

The following text is a functional description of processing that occurs primarily in the Radar Quantizer, Beacon Reply, and Target Detection Groups.

Primary Surveillance Radar Target Detection

Figure 2 is a simplified block diagram of the PSR target detection process.

Two identical quantizers are used; one for Moving Target Indicator video, and one for log video. Each is a single-level quantizer, producing either ONE or ZERO for each 1/4 nm increment in range.* A Hit Width Discriminator prevents multiple range reporting of single targets by applying a preset value matched to the pulse width of the radar to suppress runs of ONEs out of the quantizers.

A slow feedback loop compensates for thermal noise drift by comparing the proportion of ONEs over a total sweep to a manually-entered percent noise setting, and adjusting the quantization threshold. A fast feedback loop was provided with the machine to adjust the quantization threshold for clutter cells, but this function was disabled as a result of testing, and the logic has been converted to use as a clutter sensor. [8]

The clutter sensor selects the MTI or log quantizer output for further processing. (There are also manual switches to set up areas in which MTI is unconditionally selected.) The clutter sensor counts ONEs out of the log video quantizer in a region 5 range cells deep centered on the cell of interest, and 3 sweeps wide. When the number of ONEs in the window exceeds 5, MTI quantized video is selected. When the number of ONEs drops back to 2, log quantized video is selected. This logic has been demonstrated [8] to improve target detection by reducing the use of MTI video to regions of actual clutter, as compared to the grosser control provided by the manual switches. In addition, by making the logic adaptive the modification reduces the need for operator intervention.

Whenever the clutter sensor is selecting MTI for the target detector, all outputs of the detector are tagged as being in Zone 3. The use to which this is put and the origin of the "Zone 3" terminology will be described later.

The CD uses a sliding window target detector. There are 1,000 range-ordered sliding windows, providing 250 or 500 nm range depending on the width of the range cells. The window length is variable from 10 to 17 sweeps, with the value adjusted to match the beam width of the individual radar. Each window is moved from core, shifted one bit, and loaded with the quantizer output in range synchronization with the radar sweep. A target START is declared when the number of ONEs in a window reaches the Leading Edge Threshold, T_L. At that time the signal leading edge position within the range cell is noted to permit target reporting to accuracy of half a range cell. The start azimuth is stored in units of 4,096ths of a circle, called Azimuth Change Pulses, or ACPs. When the number of ONEs drops to the Trailing Edge Threshold, T_T, target STOP is declared, and the stop azimuth is averaged with the start azimuth to give the reported azimuth.

The thresholds can be raised by the Automatic Clutter Eliminator, or ACE, to suppress false target reports in clutter regions. The ACE sums the ONEs in eight sliding windows: the cell of interest, the four preceding cells, and the three following. The sum is input to a function generator that provides a raised T_L and T_T to the detector. There are three switch-selectable function generator curves giving nominal probabilities of false alarm of 10^{-4}, 10^{-5}, or 10^{-6}.

The final step in target detection is run length discrimination, which rejects targets less than a minimum or greater than a maximum run length. (Run length is the difference between target start and stop azimuths.) The run length criteria are switch-selectable in two zones separated by a switch selected range. As built, the CD had three zones. In the modified CD, the minimum run length discrimination zones are altered to permit Zones 1 and 2 to encompass the full range of the radar, while Zone 3 becomes a floating area declared to exist whenever the clutter sensor is set. The data count within Zone 3 is accumulated for an entire 360° radar scan and used as the measure to adjust the minimum run length discrimination threshold for the succeeding scan. The data rate boundaries are 32 messages per scan to decrease the threshold, and 64 to increase it. The threshold shows a tendency toward values near 18 ACPs. This adaptive logic permits the use of a higher probability of false alarm setting in the ACE, thus improving system sensitivity in clutter and reducing the areas in which ACE raises T_L to values greater than the width of the sliding window (thereby preventing the detection of any targets), while restricting the total data count to values manageable by the rest of the system by discarding targets of short run length that are least likely to be valid.

PSR target reports are output by the CD in messages containing the azimuth in ACPs, the range in units of 1/8 nm, the run length in units of 2 ACPs, and the time the report spent in storage before transmission, as well as control and parity bits.

The measured performance of this system [9] is tabulated below. All measures are taken within the coverage envelope of the radar as commissioned in the pre-existing non-digital system, and in clutter-free areas. The reason for the latter restriction is that no acceptable method of quantifying clutter intensity and assuring comparability of results at different sites has been developed.

Blip/scan ratio (proportion of scans in which target is detected): at least 91%

Range accuracy: 1/8 nm, one standard deviation

Azimuth accuracy: 2 ACPs (0.176°) one standard deviation

---

* The CD can also be set up with 1/2 nm quantization, which has the effect of doubling the maximum range of the system while halving the detection resolution.

Secondary Surveillance Radar Detection

SSR target detection is somewhat simpler than PSR target detection, as there are no clutter sensing, clutter elimination, or run length discrimination. The single quantizer accepts only those pulses that meet fixed amplitude and pulse width criteria and sends them through a shift register. Gates are used to sample two points separated by the time between the framing pulses and, when they appear, generate a bracket pulse.

The bracket pulse initiates reply code sampling at the taps of the shift register. Modes 2, 3/A, and C codes, as well as SPI and X pulses are decoded and included in target report messages to the ARTCC. The bracket pulse also initiates the garble sensing logic, that checks closely-spaced, interleaved, and overlapped replies for garble conditions.

Mode 3/A bracket detections cause ONEs to be loaded into eleven-bit sliding windows that are distinct from those used for PSR target detection. The sliding windows are shifted on Mode 3/A interrogation cycles only. The mode interlace is normally Mode A, C, A, C, or Mode A, 2, A, C. Fixed T_L and T_T are used; typical values are 6 and 2.

SSR codes are subject to validation when the count in the sliding window reaches the Validation Threshold, T_V, which is typically set one higher than T_T. If two successive ungarbled replies of the same mode and code are received, the code is validated for that mode and is so identified in the output message. If validation is unsuccessful, the last ungarbled code received is included in the output message. If all codes are garbled, the code field for that mode will contain all zeros.

When, as is the usual case, a single aircraft is detected as both an SSR and a PSR target, the CD will report only the SSR target and will set the "reinforced" bit in the output message. This condition is declared to exist if SSR and PSR targets in the same range cell overlap in azimuth between target starts and stops.

The measured performance of the system within 4/3 Earth radius coverage is as follows: [9]

Blip/scan ratio: at least 95%

Range accuracy: 1/8 nm, one standard deviation

Azimuth accuracy: 3 ACPs (0.264°), one standard deviation

Mode 3/A code reliability (proportion of returns with correct code): at least 94%

Mode C code reliability: at least 93%

PROCESSING IN THE CENTRAL COMPUTER

The central computer complex at the ARTCC accepts data from a number of CDs ranging from a low of three at New York to a high of eleven at Salt Lake City, with a design limit of fourteen. To reduce the processing load on the computer, the first step in processing is to apply a polar coordinate mask to the entering data and discard all target reports from far outside the Center airspace boundaries or in areas of multiply overlapped coverage.

The chosen approach to multiple radar data processing is to combine the data from all the radar sites into a single cartesian coordinate plane before any further processing. The obvious alternative would be to process and track data on a single-site basis, and then to pool data only when needed, as for example for display. The advantages of the chosen method are:

No complicated logic is required to track an aircraft passing from one site's coverage to another's.

The quality of automatic tracking is improved by the availability of data from supplementary sites to fill in occasional data misses from the preferred site.

When a radar site goes down, data from other sites are immediately available to fill in the coverage.

Since only one coordinate system is used for all processing, including flight plan data, many operations are simplified. For example, track positions are transmitted computer-to-computer during interfacility handoff of control. The process would be greatly complicated if each facility had multiple coordinate systems.

The common coordinate system uses the stereographic projection from the surface of the earth to a plane tangent to the earth. Radar data are converted from polar coordinates with origin at the radar site to a stereographic plane tangent at the radar location, and transformed from the radar plane to the system plane that is tangent at a point central to all the radars. The coordinates are, lastly, translated from an origin at the point of tangency to an origin at the lower left, so that all data will fall in an area of the first quadrant no greater than 1024 x 1024 nm.

The stereographic projection was chosen over alternatives such as the Lambert Conformal and Gnomonic projections because its total error in conversion and transformation is less. Stereographic projection has the property that angles are preserved but distances are not. For projecting areas of the earth's surface of the size that are of interest in radar data processing, the worst case total projection errors are well under 1/2 nm. The projection equations, a discussion of the errors, and a large bibliography can be found in Reference 10.

The next section outlines the processes of correlating and tracking data of interest to Air Traffic Control, and the succeeding section discusses the process of real time quality control of radar data.

Correlation and Tracking

Automatic tracking of radar data is initiated through operator input or automatically upon receipt of a discrete code* matching that assigned to a flight plan in the computer storage. Once a track is started, there are two processes performed: correlation, which is the identification of which target reports belong to which tracks, and tracking, which is the process of smoothing the track position and velocity and predicting the future position of the track for the next correlation. The correlation process is done once each second as data are received but tracking operates once every six seconds. The interval between operations of the tracker is called a subcycle; two subcycles make a full cycle. As will be seen, certain operations are performed only in the second subcycle.

The system plane is divided into a rectangular array of 16 x 16 nm Radar Sort Boxes. An ordered set of registers in core memory contains information on each Radar Sort Box identifying one or two radar sites Preferred for coverage in that box, one site to be used for Supplementary coverage, and the sites to be used in the event one of the former sites should fail. A target report will be subject to further processing only if it falls in a Radar Sort Box for which its radar site of origin is Preferred or Supplementary.

A congruent array of Track Sort Boxes is offset one-half box width in each direction from the Radar Sort Box array. Thus, each Radar Sort Box is overlapped by four Track Sort Boxes. A target report is a candidate for correlation with tracks in the Track Sort Boxes that overlap its Radar Sort Box.

At the time of coordinate conversion from polar to stereographic coordinates, all target reports are corrected for slant range error. If a target report contains a validated Mode C code, it is given an exact slant range correction. If not, it is first given an approximate correction and then successive tries at exact correction as it is compared with tracks for which the altitude is known by other means. Similarly, each return, as it is compared against successive tracks, is displaced at the track velocity to a reference time at the middle of the current tracking subcycle.

Correlation is done in two modes: discrete code and standard. Procedures ensure that only one aircraft in a Center's airspace will be assigned a given discrete Mode 3/A code. Therefore, a match between a received and a stored assigned discrete code gives high assurance of proper identity and the correlation logic can be simpler.

The search areas for standard correlation are illustrated in Figure 3. The primary search area is a filter applied before exact slant range and time corrections to quickly reject target reports certain to be outside the large search area. Each target report is identified as in the Large Search Area, in the Small Search Area, or uncorrelated.

Discrete correlation omits the primary search area check and goes directly to compare the track-datum deviation against small and large search area radiuses of 1 and 6 nm, respectively.

Clearly, the search area determination will not resolve cases of ambiguity involving same-code non-discrete SSR reports, wrong-code discrete SSR reports, and PSR reports, where tracked aircraft are near one another or, worse, an untracked aircraft's path crosses that of a tracked aircraft. To resolve ambiguity and give a figure of merit for later use, a Correlation Preference Value is calculated for each track-target report pair. Ordered from best to worst, the Correlation Preference Values are:

	Radar Datum Class	Track Class	Condition
1.	Mode 3/A SSR Datum, Preferred Radar	Beacon	Received Code = Assigned Code
2.	Mode 3/A SSR Datum, Supplementary Radar	Beacon	Received Code = Assigned Code
3.	Mode 3/A SSR Datum, Preferred Radar	Beacon	Received Code = Established Code
4.	Mode 3/A SSR Datum, Supplementary Radar	Beacon	Received Code = Established Code
5.	Mode 3/A SSR Datum Preferred Radar	Beacon	Received Code ≠ Assigned or Established Code or is unvalidated
6.	Mode 3/A SSR Datum Supplementary Radar	Beacon	Received Code ≠ Assigned or Established Code or is unvalidated
7.	PSR Datum, Preferred Radar	Beacon or Primary	Run length ≥ 18 ACPs
8.	PSR Datum, Supplementary Radar	Beacon or Primary	Run length ≥ 18 ACPs
9.	PSR Datum, Preferred Radar	Beacon or Primary	Run length < 18 ACPs
10.	PSR Datum, Supplementary Radar	Beacon or Primary	Run length < 18 ACPs

* A Mode 3/A code that does not end in 00.

Track Class is determined by operator entry indicating whether or not the aircraft is transponder equipped. Note that SSR data cannot be correlated with Primary Class tracks. A Mode 3/A SSR code becomes Established for a track if in three consecutive cycles the same code appears in the correlated target report.

If an ambiguity cannot be resolved by application of Correlation Preference Values, the report closest to the track is selected. To prevent erroneous correlation on false PSR target reports from clutter, a clutter detection logic is part of the correlation process. Counts are maintained of the number of PSR returns within the large search area of the track and, if thresholds are exceeded, correlation with PSR data will be partially or wholly suppressed.

On completion of correlation, once a second, the correlated target reports and all uncorrelated Preferred target reports are output to the display system. Data on the correlated reports are saved for use in the next operation of the tracker. If in a subsequent operation of the correlation process within the same tracking subcycle, a better (lower Correlation Preference Value or closer) target report is correlated with a track, the new report replaces the older as the correlated datum, and the old report is reprocessed for possible correlation with other tracks if it was originally correlated by standard correlation.

The tracker operates each six seconds, following an operation of the correlation process. Technically, it is a modified bi-modal second order $\alpha-\beta$ tracker, the two modes being determined by small and large search area correlation. When correlation is in the small search area, the target is presumed to be going in a straight line, and relatively heavy smoothing* is applied. Large search area correlation indicates an accelerating target, and lighter smoothing is done. Small search area smoothing is done each subcycle, in the cartesian coordinates of the system plane. Large search area smoothing is done only in the second subcycle and only if the large search area target report has better Correlation Preference Value than any small search area target report correlated in the complete cycle. Large search area smoothing is done in cartesian coordinates rotated to the track heading.

Four sets of smoothing constants are used:

Discrete code

Matched code (non-discrete 3/A target return with Correlation Preference Value of 1, 2, 3, or 4)

Non-matched code (Mode 3/A target return with Correlation Preference Value of 5 or 6)

Primary

Each set contains constants for small search area position and velocity smoothing, and for large search area lateral and longitudinal smoothing of position and velocity. A major product of the testing effort was the determination of best values for the smoothing constants. The complete tracking equations are in Reference 11, and the values of all constants are in Reference 12.

In addition to the small and large search area modes of tracking, there are also flight-plan-aided and free modes. The former is considered the normal mode and will be selected by the system in the absence of an operator request to the contrary, provided the track can be "matched" to the flight plan route.

The flight plan route is stored as a series of points in the system plane. To enter the flight-plan-aided mode, the program must be able to match the track's position and velocity to one of the route points or to one of the line segments connecting the points. If this can be done, the system knows where the aircraft is along its intended route of flight, and will use flight plan route segment speed and heading instead of the smoothed velocity for non-discrete tracks. It will also monitor the progress of the track along the route to ensure that the matched condition continues, and to update the time information that is stored with the flight plan and make outputs to the controller accordingly.

The final action of the tracker is to predict the track from its smoothed position at the smoothed velocity to the reference time at the middle of the next subcycle, ready for correlation during the coming subcycle. The smoothed velocity is sent to the display system, which shows it as a vector originating at the displayed position. Because radar separation is based on target rather than track position, the track is displayed to the controller as being located at the position of the most recently correlated target report. Those tracks for which there was no successful correlation for two subcycles are displayed at their predicted positions.

Many aspects of the correlation and tracking processes have, necessarily, been omitted from this short exposition. Among these are initiation criteria and transients, exception processing, extrapolation when target reports are missing, and interfacility transfer, as well as the details of the logic and equations. The reader with an interest in these topics should go to References 10, 11, and 12.

The performance of the correlation and tracking processes has been measured with data from operational radar sites and is, therefore, total system performance measured from the signal in space to the air traffic controller. The standards of performance applied to this system [9] are listed below. For reasons given earlier, performance with PSR data was measured only in clutter-free areas.

* Position smoothing is the process of moving the track from its predicted position toward the position of the correlated target report. Heavy smoothing means the track predicted position (which has been derived by averaging many previous target reports) is judged more accurate than the current report, so the track is moved a small part of the distance to the target. In light smoothing, the current report is deemed more accurate, so the track is moved further. Smoothing is similarly applied to the velocity estimate.

Unassisted Free Track Loss Rate: the frequency with which the system, unaided by operator intervention, loses track identity, measured in losses per track hour. The reciprocal of these figures is the mean unassisted track life, in hours.

For discrete code tracks, the track loss rate is essentially zero. For all others, the rate is 3.0 losses per hour for aircraft making procedure turns, and 0.37 losses per hour for aircraft going straight.

Track-Datum Deviation: the 99th percentile distance between the predicted track position and the correlated target report. For turning tracks: with discrete codes, 99% of correlated reports are within 1.10 nm of the predicted position; with non-discrete codes, within 1.95 nm; and PSR, within 2.7 nm. The corresponding figures for straight line tracks are 0.95, 1.24, and 1.7 nm.

Track-Trail Swap Probability: the per cent of instances that the tracker will follow the wrong target trail when the large search area is traversed by an untracked aircraft. (The probability of swapping trails between tracked aircraft is significantly lower.) The swap probability for discrete code tracks is essentially zero. For non-discrete code tracks intersected by aircraft transponding the same non-discrete code, the probability of a swap is 13% in crossing cases, 10% in overtakes, and 5% in head-ons. For PSR tracks intersected by PSR trails, the probabilities change to 21.7%, 16.7%, and 3.2%.

Real-Time Quality Control of Radar Data

In addition to monitoring test and status messages and total data counts from the CDs, the computer statistically samples the data to measure the accuracy of SSR and PSR collimation at each radar site, and the accuracy of registration of data from different sites after transformation to the system plane.

To identify a collimation error the system must find the PSR target report that corresponds to an unreinforced SSR target report from the same radar site. This is done as part of the correlation process. Similarly, to measure registration accuracy the system must find target reports on the same aircraft reported by different radar sites. This also is done in correlation, using discrete code targets. To eliminate errors caused by slant range correction, both target reports must contain mode C, and be more than 20 nm from the radar site at an elevation angle less than 8^{o}.

The collimation errors in range and azimuth are calculated when twenty-five samples have been collected by the correlation process. In both cases, the error is presumed to exist in the PSR data. Range error is calculated as the sum of the errors in the sample divided by the sum of the sample size and the number of reinforced SSR reports received during the sampling interval. The latter must be included in the calculation since, of course, they had near zero range error. Because a substantial azimuth error, up to half the sum of the beamwidths, can exist in a reinforced target report, the azimuth sample contains only target pairs that are separated in both azimuth and range. The azimuth error is calculated as the sum of the errors divided by the sample size.

Registration error is calculated for each pair of radar sites with substantial overlapping coverage. The error is defined as the mean difference in range and azimuth between an assumed true target position in the system stereographic plane and the reported target position in the plane. The true target position need not be known to compute the registration error. The error is computed for a particular site by comparing discrete SSR target reports from this site with simultaneous discrete SSR target reports from other sites on the same targets. It is assumed in the mathematical analysis that the range and azimuth errors present in a radar site are independent of the range of the target.

Figure 4 illustrates two radar sites and the sampling of data for analysis. In each calculation four unknowns, the range error and azimuth error for each of two sites, are to be computed. At least ten samples must be obtained in each sample area, well separated from the line connecting the sites to guarantee accuracy. The equations used for registration are much too complex to be presented in this short survey, but the interested reader will find them in References 13 and 14.

Corrections for the calculated collimation and registration errors can be entered into the computer, which will apply them to each target report when converting from polar to stereographic coordinates. However, the preferred method of correction is to change and range and azimuth presets that exist in the CD.

The performance standard for collimation and registration [9] permits a maximum bias error of 1/4 nm in range and 2 ACPs in azimuth from each error source. This yields a worst case error from these sources of 3/4 nm and 6 ACPs between PSR targets detected by different radars.

CONCLUSION

The radar data processing system is being used today to control enroute air traffic in the United States. Although this system is successfully performing its mission, there are several areas in which its performance can be improved. Among these are PSR target detection in clutter, and SSR false target detection and target splits. Many studies have been done to improve tracking performance, especially in support of advanced system automation functions, such as Conflict Alert, that require accurate tracking data. A few of these efforts are highlighted below.

To improve PSR performance in clutter, a package of improvements to the CD has been evaluated by the Federal Aviation Administration [15]. These improvements include an improved quantizer with new circuits for thermal noise and clutter suppression, and removal of the cell of interest from the Automatic Clutter Eliminator window to prevent the presence of a target from raising its own detection threshold. Tests show the modifications significantly increase the CD's see-through-clutter capability, and field trials at an operational radar site will begin soon.

The Automatic Clutter Eliminator is designed on the assumption that weather clutter is statistically uncorrelated sweep-to-sweep. There is reason to believe that this assumption is incorrect, and development is underway of an adaptive clutter eliminator that will measure the clutter correlation coefficient in real time. [16]

Among the other developments planned or in progress is the use of a mini-computer associated with the CD to suppress stationary and random false reports by comparing PSR target reports scan to scan. [17] The mini-computer can also be applied to suppression of SSR false target reports and splits, by examining target quality information that is available in the CD. An investigation of a similar process with terminal SSR data showed promising results. [18]

The number of tracking studies and reports is too large to attempt description in this short space or to include in the bibliography. In general, the studies have addressed three topics: tracking PSR targets through clutter, achieving the best balance of performance between tracker responsiveness to aircraft maneuvers and resistance to swapping to wrong targets, and improving track position and velocity estimation accuracy. The author will be happy to respond to inquiries on tracker performance studies.

REFERENCES

1. Burroughs Corporation, Transmitting Set, Coordinate Data, Models AN/FYQ-47 AN/FYQ-49, 1969, T. O. 31S5-2FYQ47-2

2. "An Application-Oriented Multiprocessing System," IBM Systems Journal, 6, 2, 1967

3. R. N. Caprio, E. P. Carrigan, T. Dunkers, J. K. Fogg, R. H. Watson, The MITRE Corporation, C40 Final Test Report, 1970, MTR-4414

4. R. A. Amato, The MITRE Corporation, C50 Final Test Report, 1970, MTR-4415

5. T. Dunkers, R. W. Lautenschlager, Dr. B. G. Sokkappa, The MITRE Corporation, Jacksonville Primary Radar Inputs and Tracking Performance Tests - Final Report, 1972, MTR-4211

6. J. B. Hagopian, The MITRE Corporation, SAFI Test Results, 1969, MTR-4402

7. E. P. Carrigan, The MITRE Corporation, Surveillance System Test Concepts for SINE, 1971, MTR-4454

8. Federal Aviation Administration, Evaluation of Selected Common Digitizer Modifications, 1974, informal report

9. T. Dunkers, Dr. B. G. Sokkappa, The MITRE Corporation, Surveillance System Performance Standards, 1973, WP-8842

10. Federal Aviation Administration, Model A3d2 En Route Stage A Computer Program Functional Specifications, Multiple Radar Data Processing, 1973, NAS-MD-320

11. Federal Aviation Administration, Model A3d2 En Route Stage A Computer Program Functional Specifications, Automatic Tracking, 1973, NAS-MD-321

12. Federal Aviation Administration, Model A3d2 En Route Stage A Computer Program Functional Specifications, Introduction to Specification Series, 1973, NAS-MD-310

13. Federal Aviation Administration, Model A3d2 En Route Stage A Computer Program Functional Specifications, Real Time Quality Control of Radar Data, 1973, NAS-MD-322

14. Dr. B. G. Sokkappa, The MITRE Corporation, Registration Computation for RTQC, 1971, WP-7681

15. H. L. McFann, W. C. Swanseen, Federal Aviation Administration, Evaluation of Improved Quantizer Detector Modification to the Production Common Digitizer, 1973, FAA-RD-73-54

16. Dr. L. P. Sinha, The MITRE Corporation, Radar Target Detection Employing On-Line Estimation of the Video Correlation, 1974, MTR-4240

17. Burroughs Corporation, Final Report, Moving Target Extractor (MOTE), 1973, 64170

18. J. E. Freedman, J. C. C. Jagernauth, The MITRE Corporation, Filtering False Beacon Targets, 1974, MTR-6619

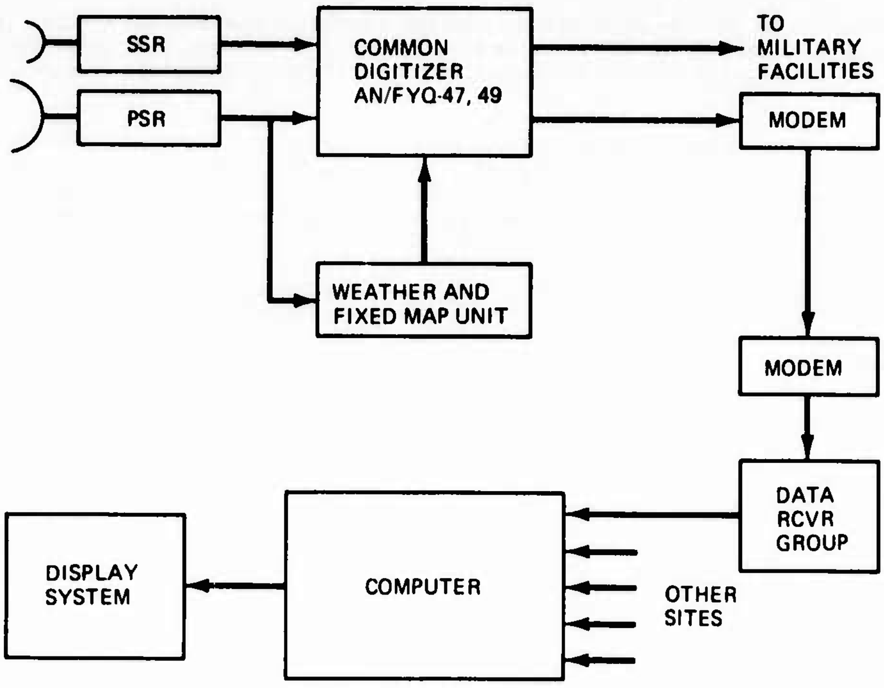


FIGURE 1
SYSTEM BLOCK DIAGRAM

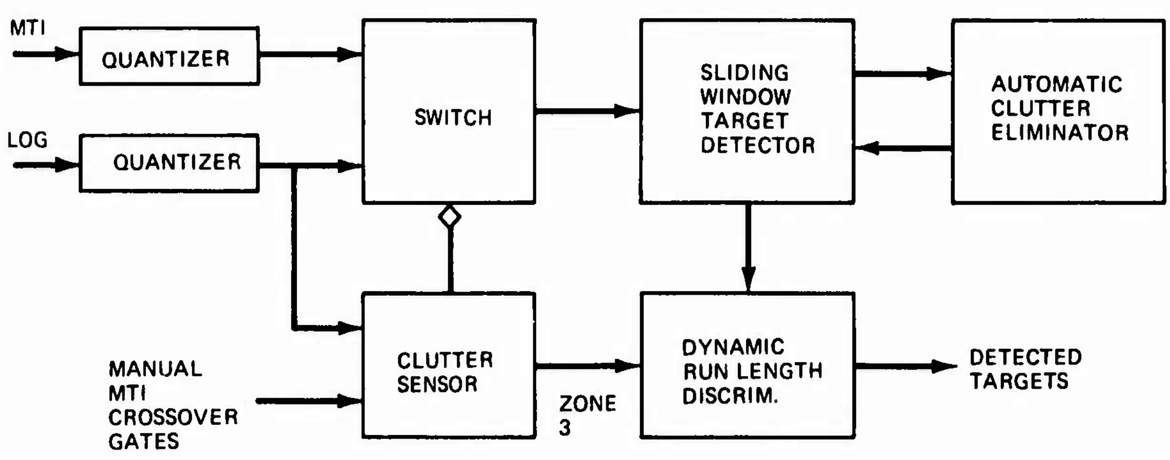


FIGURE 2
PRIMARY SURVEILLANCE RADAR DETECTION

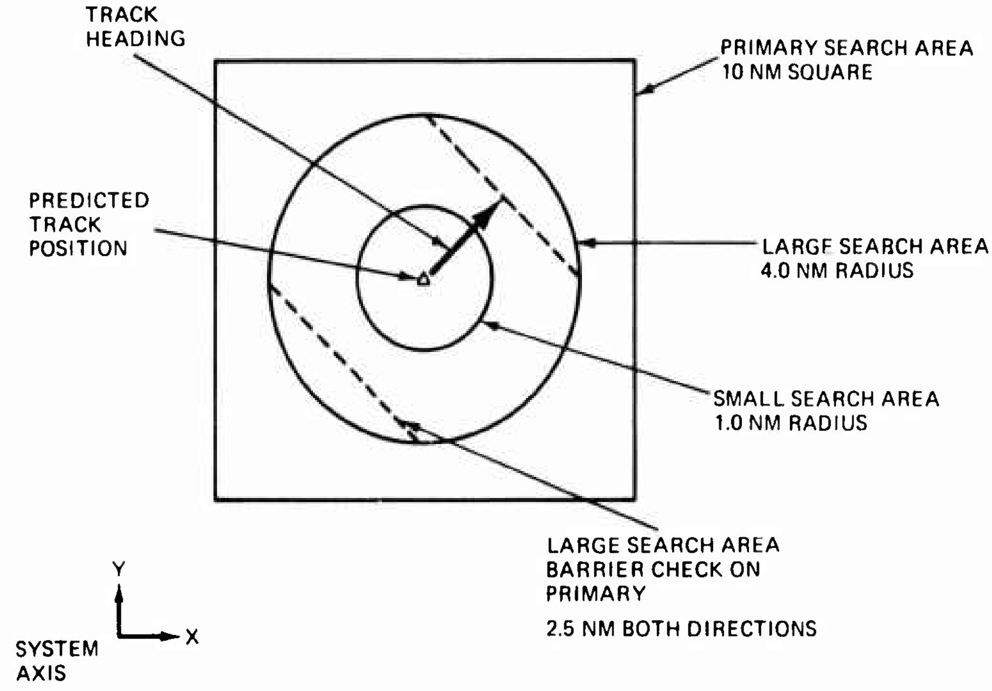


FIGURE 3
CORRELATION SEARCH AREAS

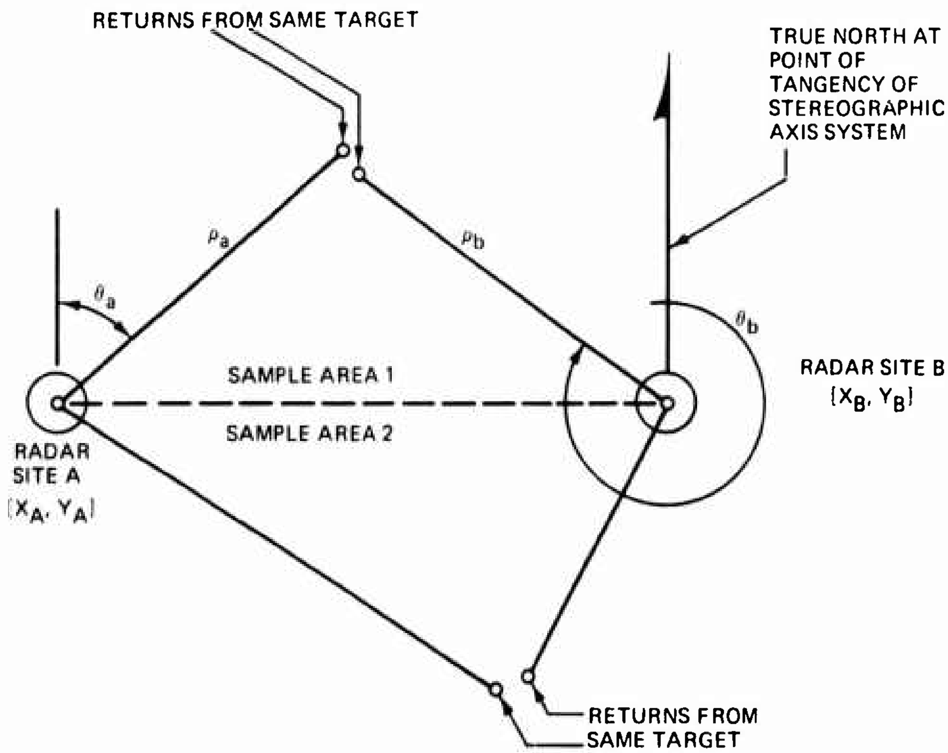


FIGURE 4
REGISTRATION SAMPLING

PART IV-E – VISUALIZATION

DISPLAY TECHNIQUES FOR AIR TRAFFIC CONTROL SYSTEMS

by
Richard Jenyns
European Organization for the Safety of Air Navigation
EUROCONTROL
72, rue de la Loi
1040 Brussels
Belgium

CONTENTS :

DISPLAY TECHNIQUES FOR AIR TRAFFIC CONTROL SYSTEMS

by
Richard Jenyns
European Organization for the Safety of Air Navigation
EUROCONTROL
72, rue de la Loi
1040 Brussels
Belgium

SUMMARY

With only rare exceptions,the cathode ray tube has consistently fulfilled the requirements for dynamic data displays in Air Traffic Control systems. The basic principles of the cathode ray tube are described together with some recent developments which help to keep it in the forefront of display technology. In recent years however more advanced techniques have emerged from the research and development laboratories which offer advantages not found with the cathode ray tube. Each of them does however have its own disadvantages. The most promising of these techniques, which may find applications in Air Traffic Control systems in the not too distant future, are discussed together with their characteristics and relative merits.

1. INTRODUCTION

The human aspects of the man-machine interface have been considered in (1). It is recognized that, as a preliminary to the specification of a data display system, the data required by the human operator and its most suitable format must be defined as a function of the individual control tasks. It is the responsibility of the display engineer to design the display system such that the information displayed to the controller will allow him to perform his tasks efficiently and with a low probability of error in an ambient environment which does not give rise to undue stress or fatigue or does not adversely affect the performance of the controller in any other way. In so doing,the display engineer must take advantage of all the techniques available to him,always bearing in mind their relative costs.

In an air traffic control system we can distinguish between two basic types of display :

- the dynamic data display and
- the tabular display.

The dynamic data display is usually a panoramic presentation of the air traffic situation as it evolves. The source of data for this presentation is either radar or flight plan information (intention) or a combination of both. In its simplest and most common form the raw radar data is presented on a circular screen or Plan Position Indication (PPI) in coordinates relative to the radar position. This is achieved by rotating a radial timebase about the point on the screen representing the radar position in synchronism with the rotation of the radar antenna. The timebase is triggered by the transmitter trigger and intensity modulated by the radar returns (video signals). Thus the distance of the plot from the centre of the screen represents the range of the target from the radar and the bearing of the plot relative to the top of the display represents its azimuth. Mosaic techniques can be used to combine,on a single screen, data from more than one radar.

In modern radar data processing systems the signals are usually digitized and processed prior to display (tracking, filtering, etc.). In this case target positions are converted into X, Y coordinates and represented by symbols on the screen. The synthetic data for display is held in a display buffer store and therefore addressing the display is random and the refresh rate can be chosen to suit brightness requirements dependent on the ambient light conditions and the remanence characteristics of the display medium. However the upper limit of display refresh rate will be a function of the total data to be displayed and the performance of the addressing and writing circuits of the display.

The dynamic data display can be considerably improved by the superimposition on the dynamic data of static data such as sector boundaries, air route networks, delineation of restricted zones, position of navigation aids, airports, etc. Further refinements can be made to such a display by the addition of alphanumeric labels related to targets and giving supplementary data such as altitude, call sign, etc.

The tabular display lists, in alpha-numeric form, information related to specific flights or particular control functions such as flight plan information, coordination messages, etc. or messages defining the present state of the system.

The most common medium for the display of dynamic data in civil ATC systems is the cathode ray tube (2). The cathode ray tube (CRT) offers advantages of versatility and simplicity. However in recent years a number of other techniques have passed through the research phase to emerge in advanced stages of development as potential replacements for the ubiquitous CRT. This paper is devoted to display devices. It will first describe the principles of the CRT and some of the recent major advances in CRT technology and will then review the new techniques which may, in the near future, find applications in ATC display systems.

2. CATHODE RAY TUBE DISPLAYS

2.1 Principles

The cathode ray tube (fig. 1) is basically an electron gun, or cathode, which generates a stream of electrons which are accelerated towards a phosphor-coated glass face plate. The impact of the high energy electron beam on the phosphor particles causes the latter to emit light by simple energy conversion. Following the cathode, a system of electron optics is designed to focus the beam to a fine spot on the screen and a deflection system will position the spot as required.

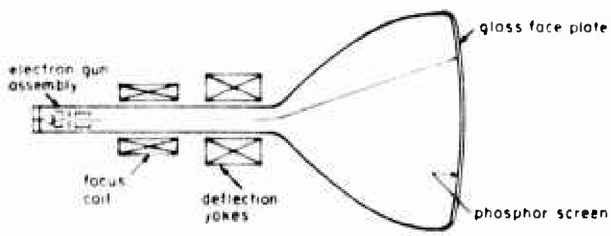


FIG.1 BASIC CATHODE RAY TUBE

2.2 Deflection

The deflection system can be either electromagnetic or electrostatic but the former is most common in dynamic data displays. Electromagnetic CRTs have coils, or yokes, placed around the neck of the tube and variation of the current in these coils will vary the magnetic field thus deflecting the electron beam. In early radar display designs a single magnetic deflection yoke was rotated around the CRT neck in synchronism with the radar antenna to produce the rotating time base. Separate fixed coils were required for vector and symbol generation. Present day displays generally have a fixed deflection system in the X and Y planes and, if a rotating scan is required, the deflection saw-tooth waveforms are amplitude modulated by the sine and cosine components of the antenna bearing. The angle of deflection of the beam, θ, passing through the magnetic field is given by :

$$\sin\theta = H.L\sqrt{\frac{1}{2V_b}\frac{e}{m}}$$

where H = magnetic flux density

L = length of magnetic field

V_b = electron beam potential

$\frac{e}{m}$ = electron charge/mass ratio.

Ideally the magnetic field should be uniform throughout the length of the deflection system. Fringe field effects at the ends of the deflection system will cause deflection defocussing i.e. the spot on the screen will become defocussed as it is deflected towards the perimeter of the CRT.

The so-called pin-cushion distortion is caused by the interaction of the X and Y deflection fields and is particularly prevalent in large diameter tubes. Correction can be applied by shaping the magnetic fields or by compensation in the coil current drive signals (3). The latter is preferable at the expense of increased complexity in the electronic circuits.

Electrostatic deflection is achieved by the electric field set up between a pair of electrodes on either side of the electron beam in the CRT neck. Two pairs are usually mounted orthognally and sequentially for X and Y deflection. The deflection angle of the beam, θ, passing through the electric field is given by :

$$\tan\theta = \frac{E.L}{2V_b}$$

where E = electric field strength between the electrodes

L = length of deflection field

V_b = electron beam potential.

Because deflection angle is inversely proportional to beam potential, deflection sensitivity can be improved by utilising Post Deflection Acceleration (P.D.A.). In this technique the beam is accelerated to an intermediate energy state prior to deflection and, following deflection, is accelerated to the normal high energy state. Although the technique can also be used with electromagnetic deflection CRTs, it is not so effective due to the fact that, in electromagnetic deflection systems, deflection angle is inversely proportional to the square root of the beam potential. Ideally the deflection system will position the spot anywhere on the screen extremely rapidly with minimum overswing. Typically, modern systems will deflect and settle across the diameter of 50 cm. CRTs in 20 μs .

2.3 Focusing

The focusing of the beam to a small spot on the screen can also be achieved electromagnetically or electrostatically, the former generally providing better resolution. However deflection defocussing is more severe with electromagnetically focused tubes.

Resolution is an important parameter defining the performance of a CRT and is dependent on the diameter and shape of the spot on the screen. It varies according to the position of the spot on the screen. The best resolution is obtained at the centre of the CRT. Deflection defocusing will adversely effect resolution if the spot is displaced from the centre. Techniques for the measurement of resolution are described in (4). The spot diameter typical of the CRTs of present ATC display systems is of the order of 0.3 - 0.5 mm for a 50 cm diameter display giving a resolution of 1000 - 1500 lines across the screen.

Laminar beam CRTs have been constructed to give a very high resolution (5). In this tube, beam crossover does not occur along the length of the CRT and the spot on the screen is an image of a hole in the cathode lens assembly and can therefore be focused to a very small size. Electron distribution across the beam is linear rather than gaussian and therefore the spot has sharp edges, giving an apparently high contrast. Geometric distortion and deflection defocusing are also insignificant.

2.4 Phosphors

The choice of phosphor for a CRT is dependent on the application (6). For raw radar displays the integrating effect of a long persistance phosphor enhances target detection, particularly in clutter. Long persistance phosphors retain the radar blip at a progressively diminishing brightness level during several scans and therefore provide an afterglow trail or history of the track position. This is useful for detection of manoeuvres, estimation of velocities, maintenance of identity, etc.

For computer driven displays of synthetic data, refreshed at a high rate, short persistance phosphors are required. Resolution requirements and choice of colour will also influence the choice of phosphor. Some phosphors have an initial high intensity, very short duration flash or fluorescent emission before the phosphorescent emission. This fluorescence, usually blue or ultra-violet in colour, may be used to activate a photo-sensitive device such as a light pen (7) for position designation. Phosphors are usually classified, according to their characteristics, into standard types (P1, P2, P3, etc.) defined by the Joint Electron Device Engineering Council (JEDEC) of the Electronic Industries Association (8).

2.5 Character Generation

A number of techniques are available for the generation of characters and special symbols for display on CRTs. The most common are the raster scan method and the stroke writing method.

In the former method the electron beam systematically scans the character area and is intensity modulated by the appropriate character signal which is coded and stored on normally a 7 x 5 or 9 x 7 matrix (fig. 2). With the stroke writing method the character is formed by deflecting the beam through a number of discrete interconnected segments (Fig. 3).

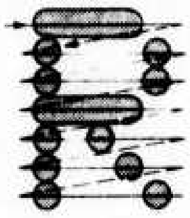


FIG 2 RASTER SCAN CHARACTER WRITING

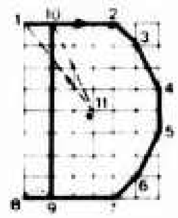


FIG 3 STROKE WRITTEN CHARACTER

For purely alpha-numeric displays the shaped beam character tube may be considered. This device has a mask in which are etched apertures in the shape of all the desired characters and symbols. For character selection the electron beam is first deflected through the appropriate aperture so that its cross section takes the shape of the character to be displayed. Further deflection circuits then position the beam on the screen. This technique gives very satisfactory character shape and a large selection of complex symbols.

For raw radar displays with a rotating scan, character, symbols and vectors are normally drawn during the dead time between scans. If the time thus available is not sufficient then radar scans must be "borrowed" for character writing. It is generally considered that up to one in seven scans can be used in this way without significant loss in radar data. In computer driven displays, with random access deflection systems, dynamic data and characters, etc. are presented randomly.

2.6 Display of Static Data

It is usually desirable to display on the dynamic data display, static data such as air route networks, sector boundaries etc.

For the raw display this is best achieved by use of video mapping techniques. The static data required for display is accurately etched on a photographic slide. The slide is scanned by a flying spot scanner in synchronism with the radar display timebase and the output is converted to an electrical signal by a photomultiplier. This signal is applied to the video circuits of the radar display. Although it is feasible to optically magnify the video map image before scanning when a number of range scale are required at the display, this has the undesirable effect of enlarging character and symbol sizes. It is preferable to have separate video map slides for each range scale required.

In the computer driven synthetic display system, the static data can be programmed and stored in the display repetition store as required together with the dynamic data. In this case the dynamic and static data are read out and displayed together. Another technique for the display of static data is the rear port tube (9) (10). The rear port CRT is similar to a conventional CRT but with a clear window in the cone through which a static image can be projected onto the rear surface of the screen. (fig. 4). The required static images are recorded photographically on slides and projected by conventional slide projectors. The major problem in the design of such a CRT is that, either the projection axis, or the electron beam axis, or both must be displaced from, and perhaps at an angle to, the CRT axis and therefore compensation must be included in the deflection system or the optical system or both to prevent distortion. Care must also be taken in the mechanical design to avoid registration errors. An advantage of the system is that the static data may be in a different colour to the dynamic data thus improving the legibility of the display.

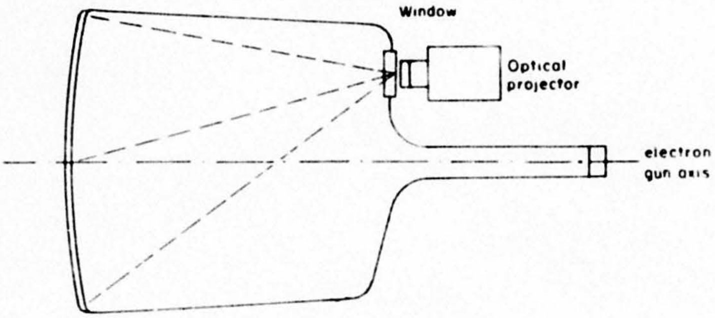


FIG 4 REAR PART TUBE

2.7 Bright displays

The brightness of raw radar displays is normally low and therefore, in order to maximise contrast and consequently target detectability, low ambient lighting levels are required. However this is incompatible with the necessity to read ancillary data such as flight progress strips, control panel indicators, etc. and the requirement for natural ambient working conditions. A number of techniques are available for the presentation of raw radar in high ambient light conditions.

The scan convertor or Radar Bright Display (RBD) is simple in principle (11). The raw radar data is written onto an electrical storage tube using the normal radar scan. The image thus produced is read out by scanning with a TV raster and displayed on a TV monitor at a high refresh rate, thus achieving a high effective luminance. This may be of the order of 10 times that of the conventional PPI display. The disadvantage of the scan converter is a reduction of resolution compared to the PPI.

The Direct View Storage Tube (D.V.S.T.) is capable of very high brightness levels. It is often necessary to erase the total picture (page erasure) and rewrite in order to modify any item of data on the screen. But versions with selective erasure do exist (12). Basically the image is written and stored as an electric charge on a storage mesh with a dielectric backing near the CRT screen. A flood gun, separate to the writing cathode assembly, evenly floods the storage mesh with electrons. When the information is stored on the mesh the electrons are allowed to pass through and are accelerated to screen potential to activate the phosphor. Picture resolution is limited by the storage mesh and the resolution of selective erasure is usually 2 or 3 line widths. Writing and erase speeds are also slower than for conventional non-storage tubes.

The Skiatron CRT which uses a scotophor or potassium chloride phosphor, can also be used in high ambient light environments. These phosphors, when bombarded by electrons, do not emit light but become absorbent to green light. Thus writing on the screen has the effect of producing a dark trace (magenta = white minus green) on a white background. This contrast is enhanced rather than impaired by increasing the ambient light level. The scotophor phosphors are long persistance phosphors and require heat for erasure. The erasure and settling process requires several seconds. Slow writing speed is also a disadvantage.

2.8 Colour

The potential of multi-colour displays has not yet been exploited in ATC display systems but there is no doubt that the added dimension of colour will facilitate the interpretation of the display by the controller. A number of techniques exist for the contruction of multi-colour cathode ray tubes.

The most common colour CRT is the shadow mask tube which is now the standard CRT for commercial colour television. The screen of this tube consists of phosphors of the 3 primary colours (red, green and blue) laid down in a systematic dot matrix as shown in figure 5. Near the screen and parallel to it is a mask with a pattern of holes arranged such that each hole is aligned with a triangular group of three colour phosphor dots.

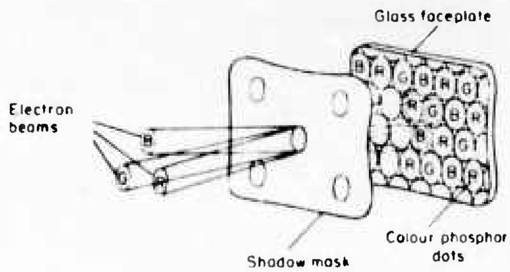


FIG 5 PRINCIPLES OF SHADOW MASK CRT

The CRT has three electron guns, one for each colour. The focusing and collimation of the electron beams is such that, on deflection to any part of the screen all three beams pass through the same hole in the mask but activate separately the three different phosphors. The size and proximity of the dots in a 3 colour group is such that, when activated, the appearance to the viewer is a single colour dot, the colour of which depends on the relative intensities of the 3 basic colours. The disadvantages of the technique include the limited resolution, which is governed by the minimum size of the colour dots and their spacing and the relative complexity of manufacture.

The single gun Lawrence tube adopts a simpler approach to the production of multi-colour displays but suffers from the same limitations of resolution as the shadow mask C.R.T. In this technique three phosphors of the three primary colours are laid down in a repetitive pattern in thin vertical stripes across the tube face. The colour stripes are divided by a wire grid. The electron beam from a single gun is directed towards the appropriate triplet of phosphor stripes and colour selection is made by the relative potential applied to the wires associated with each phosphor.

A colour CRT which is more applicable to the ATC display requirements for comparatively high resolution is the Penetron tube (13), (14), (15). In its most common form this CRT has two phosphor layers of different colours separated by a dielectric barrier. At a low EHT of some 6-8 kV the electron beam does not penetrate the inert layer and therefore only activates the first or inner phosphor which emits a specific colour, say red. If the EHT is raised to a maximum of 12-14 kV then the dielectric barrier is penetrated and the outer phosphor is also energized. Now, if the conversion efficiency of the second phosphor is greater, and the eye is more sensitive to its colour (e.g. green), then this colour will predominate. Intermediate EHT values will give intermediate colours as shown on the chromaticity diagram in figure 6 by the line joining the two principal colours (red and green in this example). Colour switching speed is limited by the need to change the final anode potential through some 6-8 kV. Current circuit technology can achieve this in less than 50 μs . However a change in beam energy will effect the focus, deflection sensitivity and brilliance and therefore compensation must be applied to the appropriate circuits as a function of colour (16).

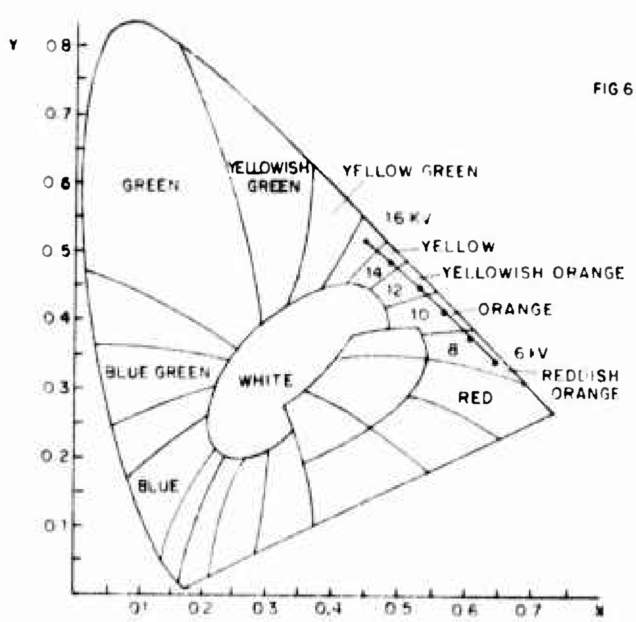


FIG 6 COLOUR CHARACTERISTICS OF TYPICAL 2 LAYER PHOSPHOR

The choice of phosphors for such tubes involves more than the choice of colour (17). The outer phosphor must be transparent and ,if sulphides are used ,they must therefore be milled to such a fine grain that many of the ideal properties such as persistance and conversion efficiency are effected. Rare earths and willemites are therefore preferable but current research is directed towards larger grain phosphors.

Post deflection acceleration applied to Penetron tubes would avoid the necessity for compensation when changing colour for deflection and focusing would be carried out at a constant beam energy. Another alternative is to use two guns, one for each principle colour operated at different cathode potentials below a common final anode potential. Colour switching signals would be applied to the respective grids. In order to compensate for differing deflection sensitivities of the two beams, the lower energy beam would have to be shielded for part of its passage through the deflection field. It has also been suggested that, in order to avoid the necessity of laying down two phosphor layers separated by an inert layer, a two phosphor mix could be applied, the particles of one phosphor being coated with an inert layer.

An alternative single gun polychromatic CRT is one in which the spectral emission of the phosphor is dependent on current density (18). This is usually achieved with a screen comprising a two phosphor mix. Phosphors normally have linear or slightly sublinear emission intensity/current density characteristics. Certain phosphors exhibit significant superlinear characteristics. The intensity/current density relationship of a near linear (A) and a superlinear phosphor (B) are shown in figure 7.

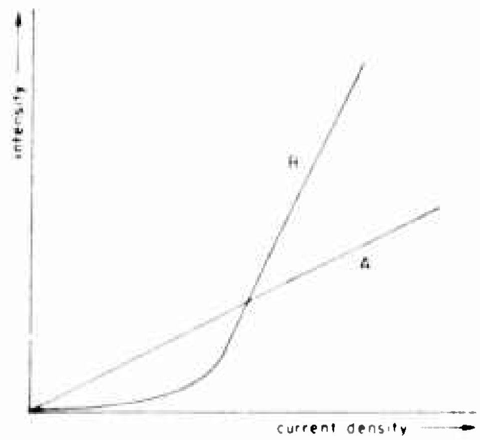


FIG 7 INTENSITY / CURRENT DENSITY CHARACTERISTICS OF CURRENT SENSITIVE COLOUR CRT

It can be seen that, if a screen combining these two phosphors is excited, than at low current densities the emission of phosphor A (say red) predominates. As current density is increased then the emission of B (green) becomes apparent until it predominates (19). A limitation of this technique at the moment is that the full colour range between the two phosphors cannot be achieved as is shown in figure 8 which represents the colour shift on a chromomaticity diagram as a function of current density. Another disadvantage is that brightness will tend to increase with an increase of current density. This can be compensated for if the sensitivity of the eye is less for the spectral emission of the superlinear phosphor than for the near linear phosphor.

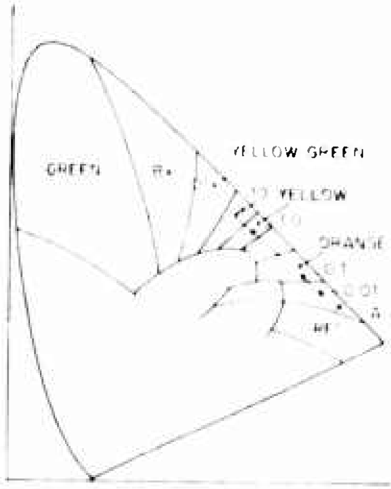


FIG 8 COLOUR SHIFT AS A FUNCTION OF CURRENT DENSITY ($\mu A/cm^2$) FOR TWO TYPICAL PHOSPHORS, A & B

3. LARGE SCREEN PROJECTION DISPLAYS

3.1 Requirements and Types

It is debatable as to whether there is, or ever will be, a requirement for large screen displays in civil Air Traffic Control. By large screen we refer to display areas measured in square meters, the information being shared by several operators. However techniques do exist, or are in an advanced state of development, and it is worth giving them some consideration. Briefly, large screen displays can be divided into two basic categories - the projection displays and the multi-element or panel displays. The latter class are considered later under advanced display techniques and therefore we will confine ourselves here to projection devices. Techniques involving photographic or etching processes such as those by which a CRT image is sequentially photographed, the film rapidly processed and projected, will not be discussed because of the inherent disadvantages of processing time and repetition rate limitations. Suitable projection techniques fall into three basic classes - CRT projection, light valves and laser displays.

3.2 CRT Projection Displays

The normal CRT image is not sufficiently bright for direct projection onto a large screen. However the Skiatron tube described above can be used with an external light source such as an xenon arc lamp. As has already been explained the potassium chloride screen of the Skiatron becomes absorbent to green light on being bombarded by electrons. The light from the external source, incident upon this CRT face, will be modulated by the image which may thus be projected through a suitable lens systems onto a large screen.

Transparent phosphors exist which become opaque when subjected to ultra violet light. Thus the screen of a CRT with such a phosphor will normally be opaque if flooded by ultra violet light. When painted by the electron beam, the screen becomes transparent to green light. Back projection by a high intensity light source and a suitable lens system will therefore allow the image to be reproduced on a large screen.

3.3 Light Valves

The basic principle of the light valve is the modulation of light by the deformation of an elastic reflecting or transparent film, the deformation being produced by a surface electric charge deposited by a scanning electron beam.

The first light valve projection technique was the Eidophor developed in 1939 at the Swiss Institute of Technology (20). The incentive behind the development was the direct projection of television images onto a large screen. The Eidophor uses a transparent dielectric oil film on a concave reflecting disc. The disc rotates through an oil bath and the oil film is smoothed and maintained at a constant thickness by a stationary bar. The disc is held in a vacuum and scanned by an electron beam similarly to the screen of a cathode ray tube. Intensity modulation of the beam causes a charge to build up on the oil film corresponding to the video input. The oil film is locally distorted by the attractive forces resulting from the electric charge. Where the film is undistorted, incident light will be directly reflected by the disc. Where this film has been distorted by the charge the light will be diffracted. The degree of diffraction is a function of the magnitude of distortion and hence the intensity of the charge. Thus, if a high intensity light source is directed onto the disc, a suitable optical system can be used to project this video image onto a screen. The Schlieren optical system is most commonly used for this.

The Schlieren optical system is shown in fig. 9. A high intensity light source such as a xenon arc is directed by a parabolic reflector and via a grating of parallel mirrors onto the disc. Any light falling on an undistorted area of the film will be reflected straight back via the mirrors to the light source. If however the surface of the film has been distorted the light will be diffracted, and reflected through the slits in the grating. A suitable lens system then focusses and projects the light onto a viewing screen. The persistence of the display is a function of the rate at which the deformation of the oil film decays and is therefore dependent on the conductivity and the viscosity of the oil.

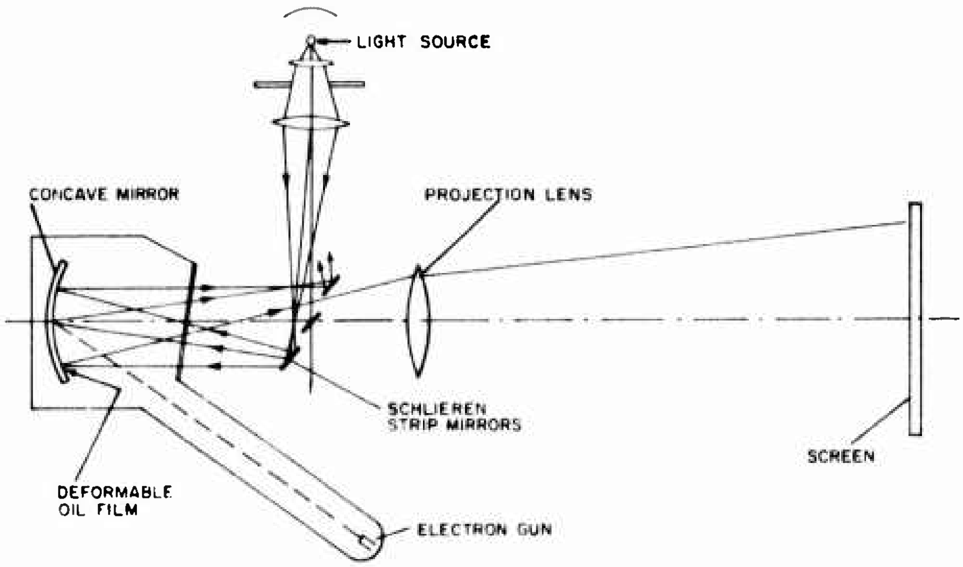


FIG 9 PRINCIPLE OF EIDOPHOR DISPLAY AND SCHLIEREN OPTICAL SYSTEM

Resolution is limited and is generally of the order of 1000 lines. The vacuum is continually maintained by pumping in order to reduce the concentration of impurities from vaporization of the oil. Nevertheless the cathode becomes contaminated leading to premature failure. Display of radar information is achieved by a scan converter to produce the raster scan necessary for the Eidophor. A colour display can be achieved by a rotating disc in front of the light source. The disc is divided into three equal sectors, each one being a filter to one of the primary colours. The system is then driven by three scan converters, one for each colour. The frame speed of the scan converters is 3 times that of the display, whereas the rotation rate of the disc is 1.5 times that of the scan converter frame repetition rate. Thus each half frame displays data in one of the three colours and after three frame repetition intervals both half frames have displayed all three colours.

The original Eidophor principle has been improved upon in later light valve developments. One improvement is the use of deformable plastic films to replace the oil and rotating disc. In one development the deformable surface has been separated from the cathode gun assembly thus eliminating cathode contamination as a cause of unreliability (21). This device is illustrated in fig. 10. The deformable film, a silicon rubber elastomer, is deposited on a mica membrane. The outer surface of the layer is a reflecting conducting film. The mica membrane divides the "dirty" vacuum with the elastomer from the "clean" vacuum with the electron gun assembly. Deformation of the elastomer is caused by charges built up on the dielectric membrane by the electron beam. In the version illustrated, flood guns are shown for erasing the stored charge or controlling the persistence.

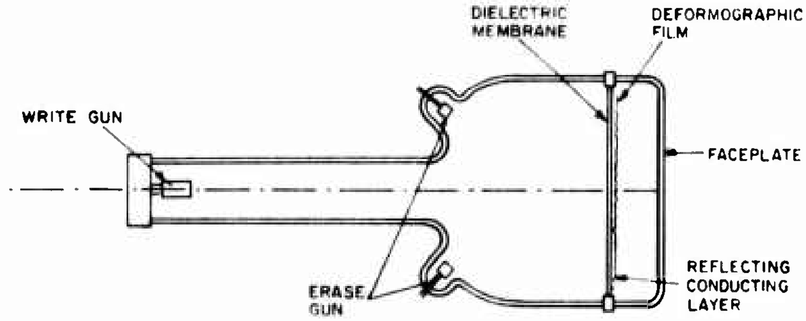


FIG.10 DEFORMOGRAPHIC STORAGE DISPLAY TUBE. (SOURCE REF 21)

A comprehensive summary of the principles of various light valve techniques with a review of the state of the art is given in (22).

2.4 Laser Displays

In its early days the laser was often dubbed as a solution looking for a problem and it was natural that applications should have been sought in the display field. Techniques have advanced such that several practical displays have been built using lasers (23). Other than holographic techniques for 3D displays, laser displays fall into one of two categories - those in which the laser beam is used to write information onto an optically sensitive material such as photographic film which is then used to modulate another light source as in a conventional projection system and those in which the laser light is projected directly onto the viewing screen.

Whichever the category, all laser displays have one characteristic in common : some means of deflecting the light beam is required. Much ingenuity has been applied to the scanning techniques (24) including the use of rotating mirrors, mirrors mounted on resonant fibres or piezoelectric crystals, etc. However direct deflection of the beam by electro-optic deflectors is the most promising for practical random access displays (25). A bistable deflection device can be produced by a birefringent crystal such as a calcite crystal preceded by an electro-optic polarization switch (26). Cascading n such devices will give 2n resolvable beam positions. The refractive index of a birefringent crystal depends on the plane of polarization of the light. The polarization switch therefore is arranged to rotate the plane of polarization of the light beam entering the crystal between two orthogonal positions. In one polarization state the beam will pass through the crystal undeflected, in the alternate state the beam will be deflected through an angle α . If the thickness of the birefringent crystals are in geometrical progression, linear displacement of the beam in discrete steps can be achieved (fig. 11). Various materials exist for the electro-optic switches which rotate the plane of polarization of light as a function of the applied electric field. The most common are nitro-benzine liquid and potassium dihydrogen phosphate (KDP) crystals. Both however have the disadvantage of requiring high voltage (kVs) to switch through 90°.

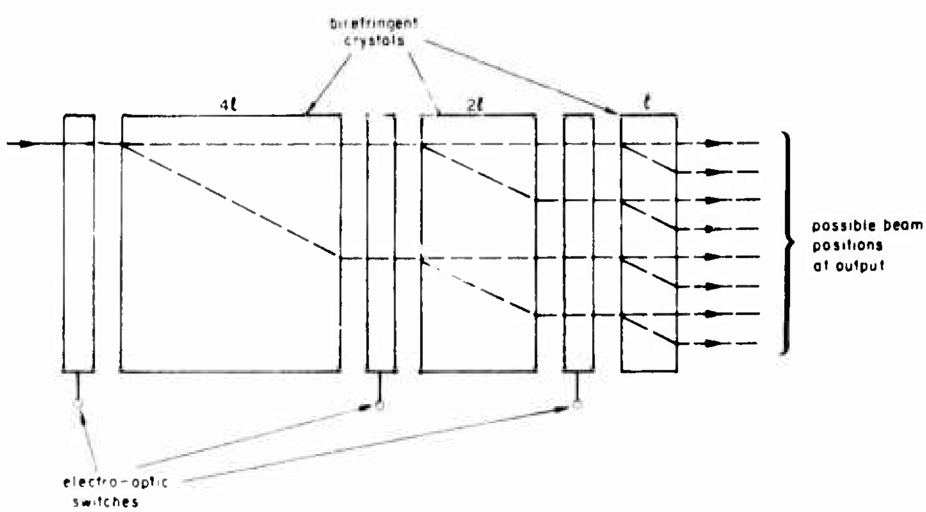


FIG 11 DIGITAL DEFLECTION OF A LASER BEAM

A display resolution of 1000 x 1000 lines is feasible by these techniques with deflection speeds of less than 1 μs between any two positions. Lasers can be used to write on photographic film or, better still, on photochromic materials. Photochromic materials are normally transparent but become opaque when subjected to light of a particular wavelength. Writing with a laser beam can be achieved at the same time as projection via an alternative light source (27). Disadvantages are the slow response of photochromic materials and the fact that erasure usually requires the application of heat.

4. ADVANCED DISPLAY TECHNIQUES

4.1 General

Except for limited use of electro-mechanical indicators, the cathode ray tube is the universal medium for the display of dynamic data in civil ATC systems. Development of the CRT has been such to keep ahead of any potential opposition. We have seen that CRTs with storage capability are now available as are colour displays with adequate resolution. Moreover bright display are available which can be viewed in high ambient light conditions. Nevertheless the CRT possesses disadvantages which impose constraints on the system designer. It is cumbersome in size and is not particularly rugged. It requires high voltages and power dissipation of the deflection circuits is also high. Being an analogue device there are problems of drift, jitter, distortion, etc.

However there are now a number of display techniques at a sufficiently advanced state of development, each having particular advantages, which may be considered as a potential replacement for the CRT in particular applications in future display systems. The more important and promising of these techniques will be discussed in the following paragraphs.

4.2 Electroluminescence

The phenomenom of electroluminescence (EL) is most simply defined as the direct conversion of electrical energy into light. It may,however, take one of two different forms. Junction electroluminescence normally occurs at a p-n junction in a single crystal. First observations are accedited to Thomas Round in 1907 but Lossev is recognised as the pioneer worker in this field whose experiments followed his recognition in 1923 of electroluminescence at a point contact on a crystal surface. Field effect electroluminescence occurs generally when high electric fields are applied to microcrystalline phosphors. This is sometimes referred to as the Destriau effect after the Frenchman who first observed the phenomenon in 1936. The most interesting characteristic of EL is that it may offer in the future the most complete gamut of visible colours of any active or passive display system. It has even been forecast that the synthesis of nearly all the 10^7 colours perceivable by normal trichomats may be possible (28).

Table 1 lists most of the materials in which electroluminescence has been observed (29). Although this list is not exhaustive,only those materials which are underlined are of practical value.

II-VI COMPOUNDS		III-V COMPOUNDS		OTHER MATERIALS	
ZnS,	ZnSe,	GaP,	GaAs,	SiC,	Ge,
CdS,	CdSe,	GaAsP,	GaAlAs,	Si,	NaCl,
ZnO,	CdTe,	GaInP,	GaN,	C(diamond),	AgCl,
BeO,	MgO,	GaSb,	InP,	ZnF_2,	CaF_2,
CaS,	SrS,	InSb,	InAs,	Al_2O_3,	Cu_2O ,
BaS,	PbS,	BN,	InAsP,	SnO_2,	TiO_2,
PbSe,	PbTe,	AIN,	AIP,	$BaTiO_3$,	$SrTiO_3$,
ZnTo.				$CaTiO_3$,	$KNbO_3$,
				$PbZnO_3$,	$CaWO_4$,
				$ZnSiO_4$,	ice,
				and other organic materials	

Table 1 : Materials in which electroluminescence has been observed. (source : ref. (29)).

In so far as junction electroluminescence is concerned, the most common materials used to date are GaP, GaAsP and GaAlAs (30). Figure 12 shows the spectral emission of these and other materials. GaP can emit in the red and the yellow-green area of the spectrum whereas GaAsP and GaAlAs are limited to the red-orange area. The power efficiency of GaP emitting in the green is not very high but, because this is near the peak sensivity of the eye (shown also in fig. 12), the luminous efficiency is as good as the red emission. Theoretically GaAs, which only emits in the infrared, can be used in conjunction with a visible emitting phosphor which is activated by infrared radiation to produce other colours. Recently a GaP light emitting diode (LED) has been announced, the colour of emission of which varies between red and green depending on the current drive.

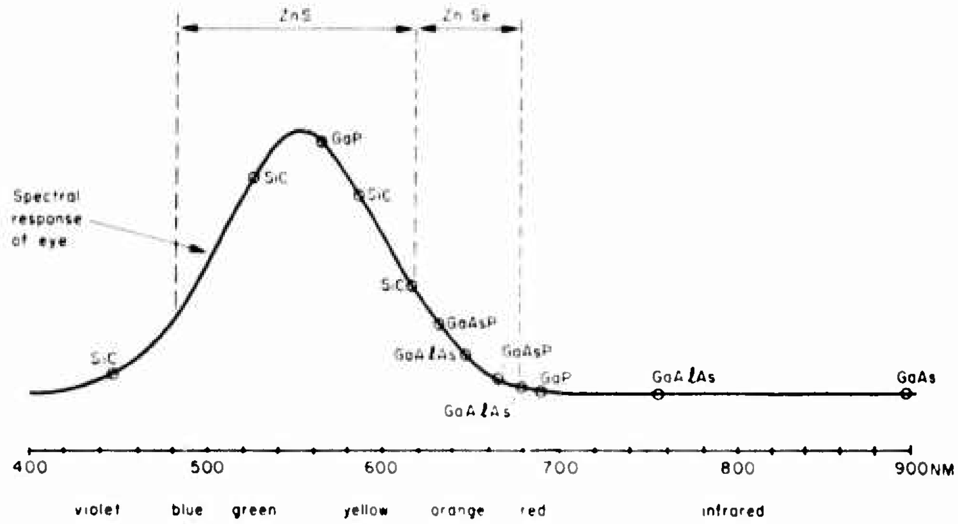


FIG 12 SPECTRAL EMISSION OF SOME EL MATERIALS SHOWN IN RELATION TO THE SPECTRAL RESPONSE OF THE EYE

Recently researchers both in Europe and the USA have reported blue emission from GaN and this material, which has a large band gap, should be able, with suitable dopants, to emit other colours. GaInP can also be made to emit anywhere between red and green but is, at the moment, very inefficient. Silicon carbide also possesses a large energy gap and, with suitable choice of material polytype and impurities, can be made to luminesce throughout most of the visible spectrum. The material is difficult to fabricate but recent progress in building electroluminescent devices has been reported (31).

The brightness of these techniques is usually high, typically of the order of 100-1000 Foot-Lamberts (ft-L) depending on the material, drive, etc. Higher luminances have been reported and, of course, can always be achieved by driving the device harder but with the consequence of reduced life.

An advantage of EL diodes is that the drive voltage is compatible with logic levels but their efficiencies are still such that current drivers are required to interface between the logic circuits and the light emitting elements. The reliability is of course, very high being similar to that of other semiconductor devices. The conversion efficiencies of these devices varies depending on the materials. Red GaP has one of the highest efficiencies, 6-7%, but for some materials the efficiency is 0.1% or even less (32). Research is now devoted to the materials technology and particularly to improving luminous efficiencies in order to reduce current drive requirements. Switching times are very fast, being of the order of nanoseconds and they exhibit a sharp threshold which, as we will see later, is an advantage in cross bar addressed matrix displays.

To form a display using these devices two approaches are possible, the multichip array and the monolithic array (33). The multichip array requires the bonding together of discrete elements to form a matrix addressable either as individual points or in x and y. The monolithic array is made up of p-n junctions diffused on a single substrate. With both techniques difficulties exist in the construction of a large matrix. To produce a matrix of only 100 x 100 points with the multichip method, it is necessary to lay down accurately and to wire bond 10,000 diodes. With the monolithic process a sufficiently large substrate could not be produced, and even if this were possible, a 100% device yield on such a large sample is beyond present day production capability.

Field effect EL devices are usually constructed by sandwiching the electroluminescent material suspended in resin, between two electrodes, one of which is transparent (fig. 13). Thus, in effect, the device is a capacitor. If a potential is now applied to the electrodes the resultant field will cause the device to emit light. These devices are normally ac energised because of the poor conduction paths in a phosphor particle structure. However, efficient dc devices are now also possible (34). Colour of emission will depend on the material used and the activator. The most common phosphors are of the zinc-sulphide (ZnS) family (35).

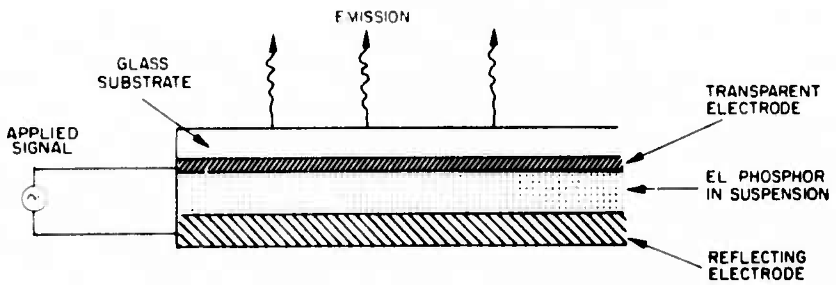


FIG 13 CONSTRUCTION OF FIELD EFFECT ELECTROLUMINESCENT DEVICE

It is also possible to produce a device, the colour of which is a function of applied frequency. For instance a phosphor doped with both copper and chlorine will emit green at low frequencies and blue at higher frequencies. Because the recovery time of the green centres is comparatively slow they become saturated as the frequency is increased and the blue emission thus becomes the dominant one.

A multi-coloured display has also been proposed based on a combination of flourescent stimulation and dielectric reflection (36).

One disadvantage of this technique is that brightness is limited, generally in the range of 5-50 ft.L. Researchers have reported values of 200 to 300 ft.L. but generally with very much reduced life which, even in normal circumstances, is not very high. Failure is not normally catastrophic but is evident by a gradual reduction of light output. Half life is typically of the order of 1000 hours. Drive voltages are of the order of 50-200 volts. Present research is devoted towards increasing brightness and reliability while reducing the required drive voltage.

The technique does, however, lend itself to large area displays of several hundred lines resolution. Fig. 14 illustrates how such a display would be fabricated. The phosphor is deposited between orthogonal sets of parallel electrodes such that a signal applied to one X and one Y electrode would activate the phosphor at the intersection. Using this technique a flat screen TV has been built and demonstrated (37). The screen, with a 33 cm diagonal has a limiting resolution of about 150 T.V. lines and a brightness of 10 ft.L.

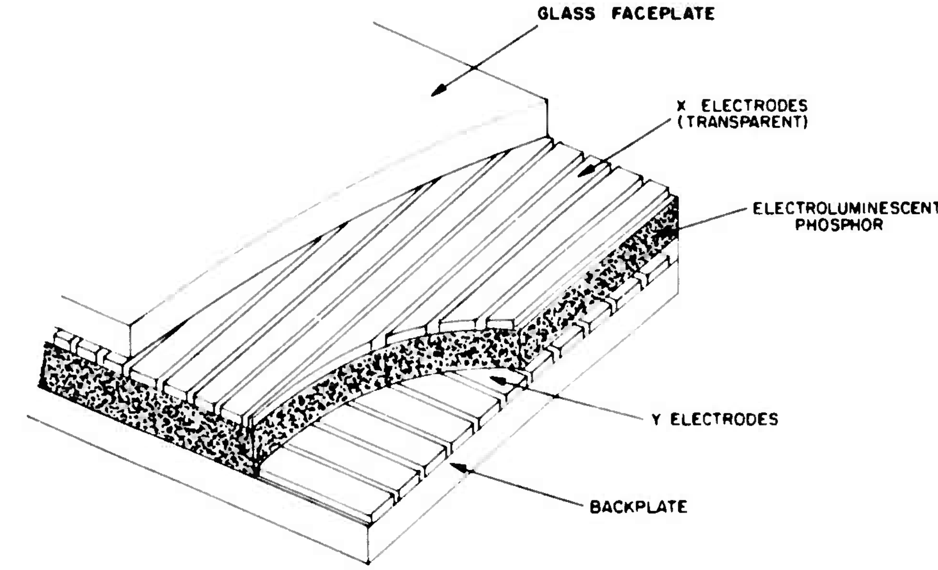


FIG 14 CONSTRUCTION OF EL DISPLAY PANEL

4.3 Liquid Crystals

Liquid crystals are organic compounds having the properties of liquids but with an ordered crystalline structure (38). Because they exist between the solid and the pure liquid state they usually exhibit these particular characteristics within a limited temperature range. Although many organic compounds possess the properties of liquid crystals only a few exist in this state at normal ambient temperatures. Liquid crystals can be classified into three different types according to their crystalline structure, smectic, cholesteric and nematic liquid crystals. These are illustrated in fig. 15. In each case the long, cigar shaped molecules are arranged in an orderly pattern under normal conditions.

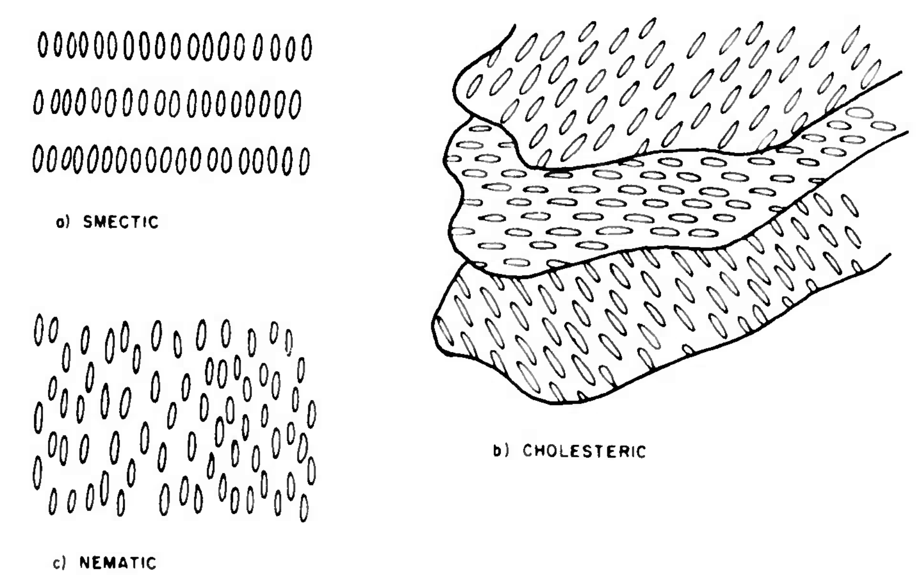


FIG 15 CRYSTALLINE STRUCTURE OF BASIC LIQUID CRYSTAL TYPES

The molecules in smectic materials lie in parallel layers with their axes perpendicular to the plane of each layer. Smectic materials do not normally respond to an electric field and therefore, except for one recent experiment in which a laser was used to write on a smectic panel, they are not considered practical for display applications.

In cholesteric materials the molecules are arranged in parallel layers,and in each layer, the axes of the molecules are mutually parallel and lie in the plane of the layer. Furthermore there is a rotation of the direction of the axes from one layer to the next. The angle of rotation is constant. An interesting characteristic of cholesteric materials, in so far as display applications are concerned, is that they will change colour under the influence of a dc field. Despite this, little effort has been devoted to their development as a display technique.

Nematic materials (39) demonstrate the least orderly crystalline structure. Their only characteristic is that, in a passive state, the molecules are aligned with their longitudinal axes mutually parallel.

Various electro-optic effects of nematic materials can be exploited to produce display devices (40). The most common of these are :

- dynamic scattering,
- electrically controlled briefringence and
- twisted nematics.

Dynamic scattering is a current effect. In nematic compounds the electric dipole moment does not lie along the molecular axis. The result of this is that application of an electric field causes ions to flow, resulting in a space charge building up,and the consequent shear forces cause turbulence of the molecular arrangement and changes in the refractive index. Thus, what is normally a transparent liquid, will scatter light when a field is applied. To exploit this effect a thin layer of nematic material, normally 5-30 microns thick, is sandwiched between two glass plates with transparent conductors deposited on their inner surfaces. The crystals can be either in a homeotropic or homogeneous state, i.e. the molecules are aligned either normal to, or parallel to, the electrodes. Desired alignment can be achieved by a number of different methods (41).

Figure 16 illustrates how dynamic scattering can be exploited. In the direct viewing mode (a) the light source is directed obliquely through the layer and the element will appear dark to the observer until a field is applied to the cell when the transmitted light will be diffused. Thus the cell behaves as a simple optical shutter. A reflective device (b) is similar to direct viewing except that ambient light is used together with a reflecting rear surface. The advantage of this technique is that the brighter the ambient light then the brighter will be the display. Figure 16 (c) illustrates how dynamic scattering can be used to construct a projection display.

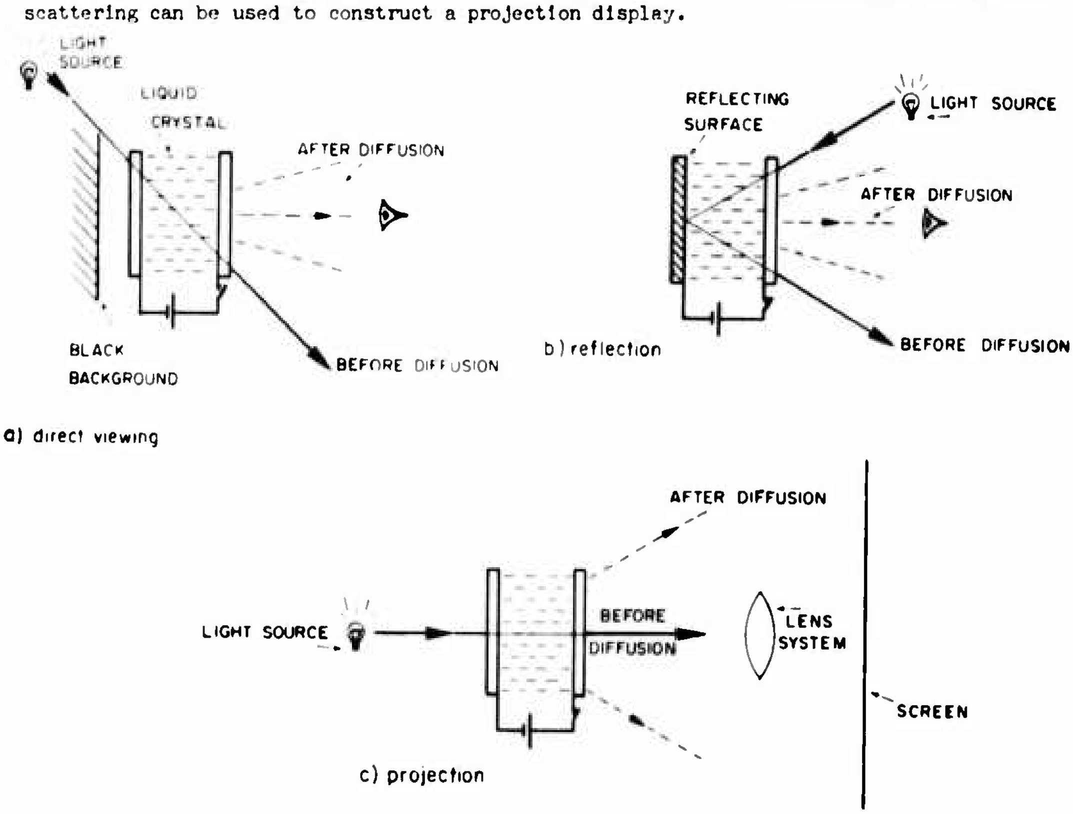


FIG 16: EXPLOITATION OF DYNAMIC SCATTERING EFFECT

An intrinsic disadvantage of dynamic scattering is the comparatively slow response. Because the cost of individual addressing of each display element would be prohibitive for a display of more than several characters, multiplexing must be used. This means that each element must be pulsed. But the voltage required to switch this device increases rapidly as the mark to space ratio of the pulse train is reduced (see figure 17).

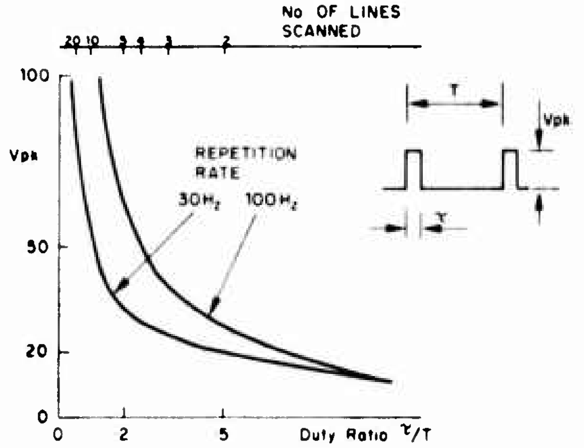


FIG 17 PEAK SWITCHING VOLTAGE AS A FUNCTION OF DUTY RATIO (Source: ref. 40)

For both X-Y addressed matrices and multiplexed alpha-numeric displays this is particularly disadvantageous because half addressed elements will turn on if the peak voltage exceeds by two or three times the threshold voltage. Dual frequency addressing can be used to overcome this disadvantage (42) (43).

Electrically Controlled Birefringence (ECB) is a field effect. If a nematic liquid is in a homeotropic state, i.e. the molecules are aligned perpendicularly to the electrodes, then the birefringence of the device can be controlled by applying an electric field. If light is passed through a polarizer, then through the device and through a second polarizer at 90° to the first, the device will appear dark if unactivated. Application of an electric field will cause light to be transmitted and the wavelength (or colour) of the transmitted light will depend on the voltage applied.

ECB, although offering a large multiplexing capability, suffers from a very narrow field of view and difficulties of device manufacture.

Twisted nematic devices have the molecules aligned parallel to the electrodes but the direction of the longitudinal axes of the molecules varies through the liquid layer from one electrode to the other. Plane polarized light passing through such a device will have its plane of polarization rotated. Application of a voltage will tend to straighten the helix arrangement of the molecules and change the angle of rotation of the plane of polarization. Thus, placed between two polarizers, this device has a high contrast ratio and a sharp threshold.

A comparaison of the three electro-optic effects and the present state of the art is given in table 2 :

	Dynamic scattering	ECB	Twisted nematics
Effect	Current	field	field
Drive (volts)	10-30	4	2-4
Current ($\mu A/cm^2$)	10	1	1
t (rise) (msecs)	10-20	10	10
t (decay) (msecs)	100-200	100	100

TABLE 2 : Characteristics of nematic liquid crystals.

Memory can be provided with liquid crystals by adding 1-2% of a cholesteric material to a nematic material. Writing is then possible at a low frequency whereas a high frequency will erase. Storage of several months has been achieved in this way but, at the moment, the amplitude of the erase signal is very high.

4.4 Gas Discharge

The glow discharge phenomenon which can be induced in certain gases under certain conditions has been exploited for display purposes. Both ac and dc devices have been developed. The ac device has inherent memory. However dc devices can also be built with memory. Both types are most suitable for large panel displays of a matrix of small closely spaced cells addressed by orthogonal sets of parallel electrodes. However both methods also lend themselves to bar matrix alpha-numeric indicators.

Fig. 18 shows the usual construction of a dc gas discharge panel.

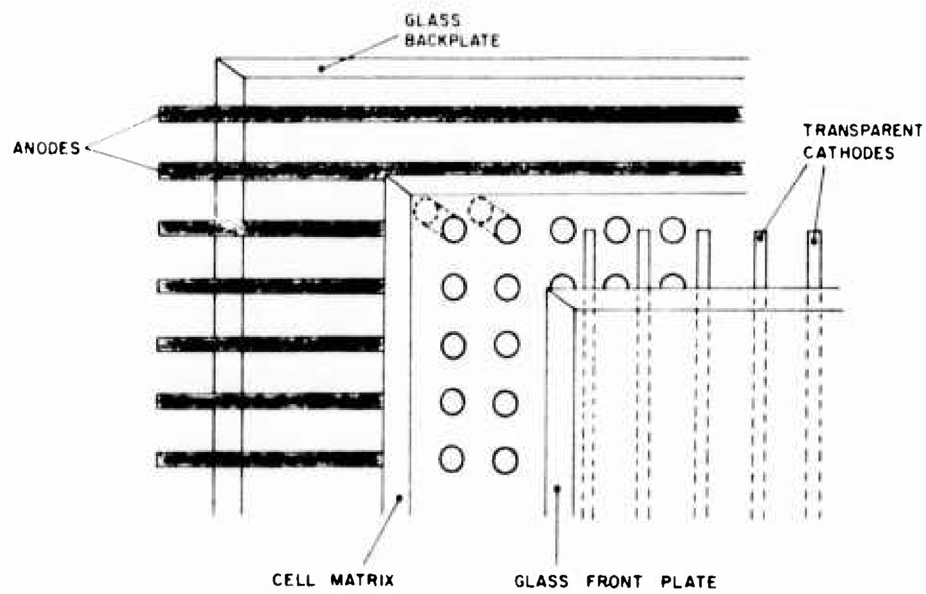


FIG 18 CONSTRUCTION OF A DC GAS DISCHARGE PANEL

The electrodes are typically thick film conductors fired onto the inner surfaces of the glass plates. They coincide with the rows and columns of cells which are etched in the photosensitive glass spacer. The cell contains a gas, the most common being neon. Appropriate signals applied to a pair of anode-cathode address lines will fire the cell at their intersection. The striking voltage depends on several factors such as gas used, pressure, spacing between electrodes, etc. Usually a bias voltage of 150-250 V is applied permanently between anodes and cathodes and half address pulses superimposed on both electrodes to ignite the appropriate cell or cells. The half address pulse applied to a single electrode plus the bias voltage must not fire any non-addressed cells. Therefore the spread of the striking voltages must be kept to a minimum. Mean brightness is typically 100 ft.L. Reliability is affected by sputtering but the effect of this can be reduced by recessing the cathodes and providing sputter traps between adjacent cells (44). Memory can be obtained by providing a resistor in series with each cell. This can be achieved by printing and firing the resistor onto the glass substrate using high resistivity thick film inks (45), or by evaporating thin-film resistors (46). Because in this case the addressed cells are permanently on, brightness values of 1000 ft.L. can be achieved.

Self scanning gas discharge alpha-numeric displays have been developed using a technique which allows the transfer of the glow discharge from one cathode electrode to an adjacent cathode (47). A panel using this technique has also been built with grey scale (48).

Fig. 19 shows the construction of an ac discharge panel.

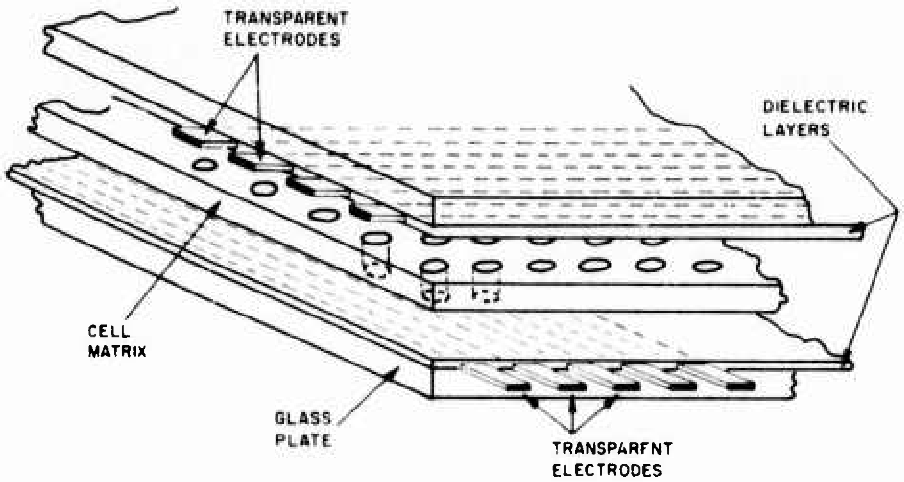


FIG 19 CONSTRUCTION OF AN AC GAS DISCHARGE PANEL

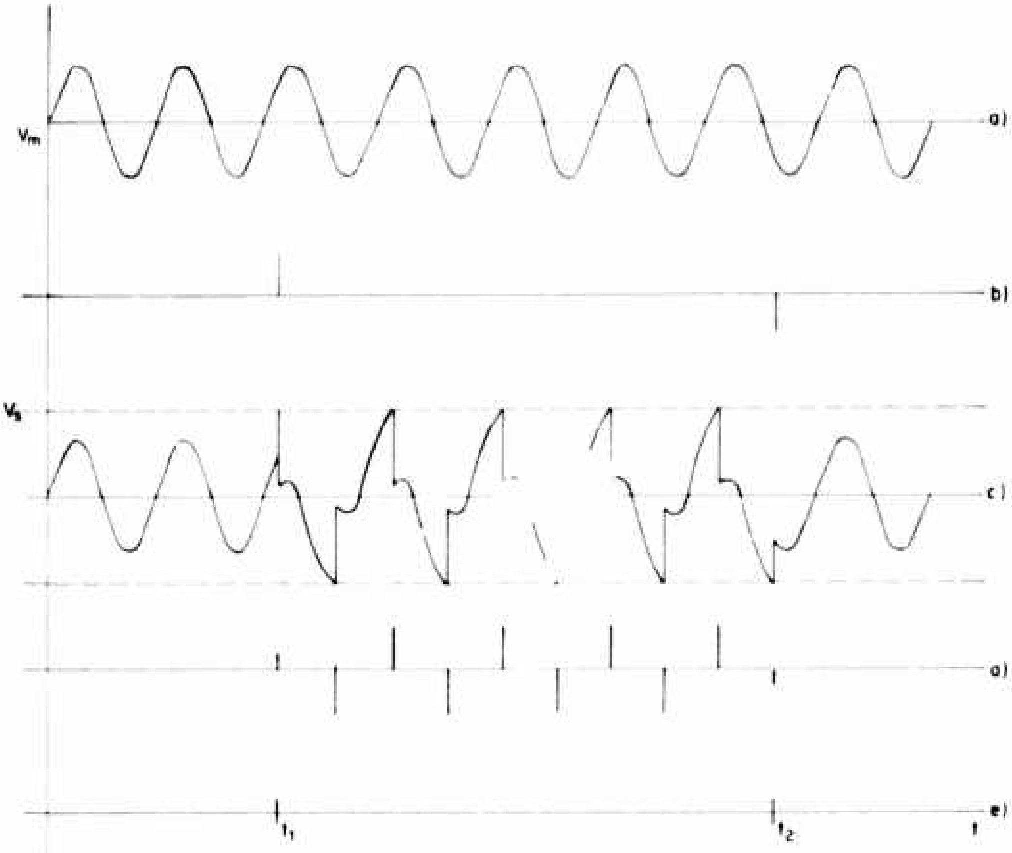
Comparing this with fig. 18, it will be seen that the essential difference is that the electrodes of the ac panel are not in contact with the gas. This is usually achieved by depositing a dielectric layer over the electrodes. The electrodes are thus capacitively coupled to the cells and the device must be ac driven.

The advantage of the ac panel over the dc device is that the former exhibits inherent storage (49). This is best explained by reference to fig. 20, which illustrates the voltage and current waveforms associated with the different phases of operation.

Consider the operation of a single cell. A maintaining voltage, Vm, is applied to the electrodes permanently. (fig. 20a). By capacitive coupling, a similar voltage exists across the cell but it is arranged that this cell voltage is below the required striking voltage, Vs, as shown in the first part of fig. 20c. At time t_1 a positive pulse (fig. 20b) is superimposed on the maintaining voltage thus taking the voltage across the cell above Vs. The cell fires and electrons and ions separate to create an additional wall charge voltage across the cell which opposes the applied voltage and extinguishes the discharge. The residual wall charge is such that it will now aid the voltage build-up during the next half cycle and therefore the maintaining voltage alone will be sufficient to take the cell voltage beyond Vs and cause another discharge. At each discharge a reversal of the residual wall charge will occur and the cell will fire every half cycle (fig. 20 c & d). This will continue until such time that the cell is erased. To achieve erasure a negative pulse (fig. 20b) is superimposed on the maintaining voltage at time t_2 such that it opposes the reversal of the wall charge, reducing it to zero. The cell then reverts to the off state.

A display panel has been built with multiple states offering the possibility of grey scale (50).

Typical operating characteristics include a maintaining voltage of 200-250V at 50-250 KHz. Half switching pulses are typically 30-60V. Brightness is of the order of 50 ft.L. but is dependent on frequency of the maintaining voltage.



a) Maintaining voltage (Vm) b) Write and Erase pulses c) Cell voltage d) Cell current

FIG 20 VOLTAGE AND CURRENT WAVEFORMS OF AC GAS DISCHARGE CELL

The most suitable gas for both dc and ac devices is neon which limits the colour of the display to the orange-red glow typical of this gas. However other colours are possible by using a gas such as xenon, which emits ultra-violet radiation, in conjunction with a phosphor deposited on the cell walls which is excited by UV light (51). Alternatively phosphors excited by low energy electrons can be used (52).

4.5 Addressing

The advanced techniques discussed above are generally suitable for alpha-numeric and dynamic data displays, although, because of particular characteristics, certain are more suited to one or the other of these applications.

In so far as alpha-numeric displays are concerned there are two possible methods of displaying a character : the segmented matrix and the dot matrix. These are illustrated in figure 21. The 7 bar segmented matrix (fig. 21a) is sufficient for the numerals 0-9 and a limited number of letters but the 16 bar version illustrated in figure 21(b) is required for a full alpha-numeric capability. With a dot matrix, the larger the matrix then the less will be the limitations on character shape or font. However, the larger the matrix the more complex will be character generation and addressing. A 35 point matrix (7x5), as shown in figure 21(c) is usually considered the best compromise.

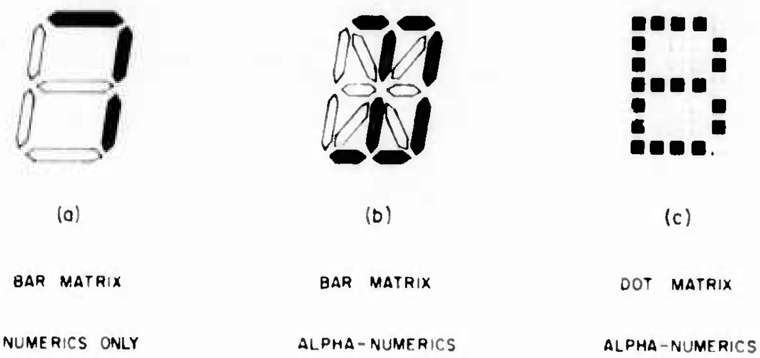


FIG 21 TYPICAL NUMERIC AND ALPHA-NUMERIC MATRICES

With the segmented matrix, character generation is simpler and therefore cheaper but a 7 x 5 dot matrix generally gives a more pleasing character font and the possibility of a larger number of special symbols.

Character generators are available, off the shelf, for both methods. In each case they accept a standard 6 or 7 bit data code such as an ISO or the ASCII code and convert this to a 7 or 16 bit parallel output for the segmented matrix, or a 35 bit parallel/serial output for the dot matrix. In this latter case the character generator outputs sequentially the 5 columns of the character format on 7 parallel lines. For such a character generator an additional 3 bit input is required to define the column to be output and to control its timing. Special character generators are also available which output the 35 dot character row sequentially on 5 parallel outputs. As for the configuration of an alpha-numeric display, we can distinguish three basic elements: the memory, the character generator and the display element. Either the character generator can precede the memory or it can come between the memory and the display. Furthermore the character generator may or may not be time shared. The configuration will depend on display capacity, characteristics of the display technique, etc. (53).

In a static array all characters are permanently on. This means that either each character has its own character generator interposed between the memory and the display or the memory follows the character generator, there being 1 bit of memory for each display element. The former solution is expensive in character generators, the latter is only suitable if the display technique has inherent or integrated memory.

If the display technique does not have inherent memory then time sharing the character generator requires that the display is continually refreshed. This must be done at a sufficient rate to avoid flicker. However the rate at which the display can be refreshed depends on the rise time of the display technique and the number of characters. Furthermore the effective brightness will depend upon duty cycle of the individual elements, the duty cycle being defined as the ratio of "on" time to "off" time.

For time shared arrays various options are open. A display can be refreshed character sequentially, i.e. one character at a time. The duty cycle therefore for an N character display is 1/N. For a 7 x 5 dot matrix display a 35 bit buffer memory must be interposed between the character generator and the display and an N position scanning circuit must be provided (see fig. 22).

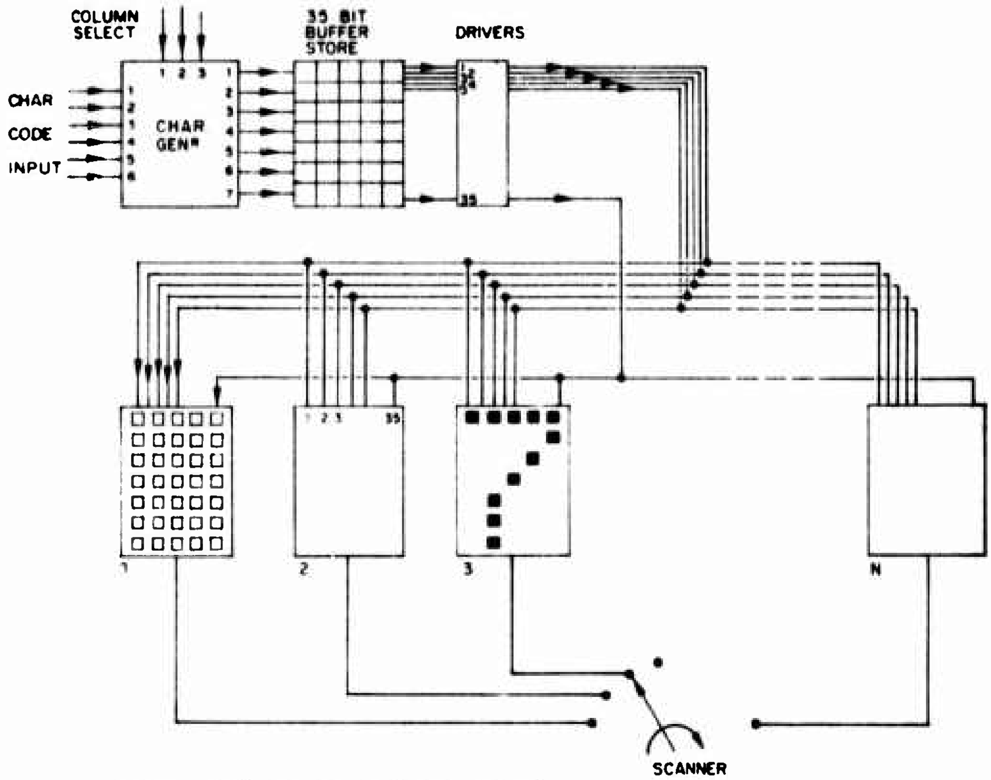


FIG 22 CHARACTER SEQUENTIAL ADDRESSING

Alternatively the display can be driven column sequentially as shown in figure 23. In this case no buffer memory is required and only 7 drivers instead of 35 but, on the other hand a 5 x N scanning circuit is required and the duty cycle falls to 1/5N.

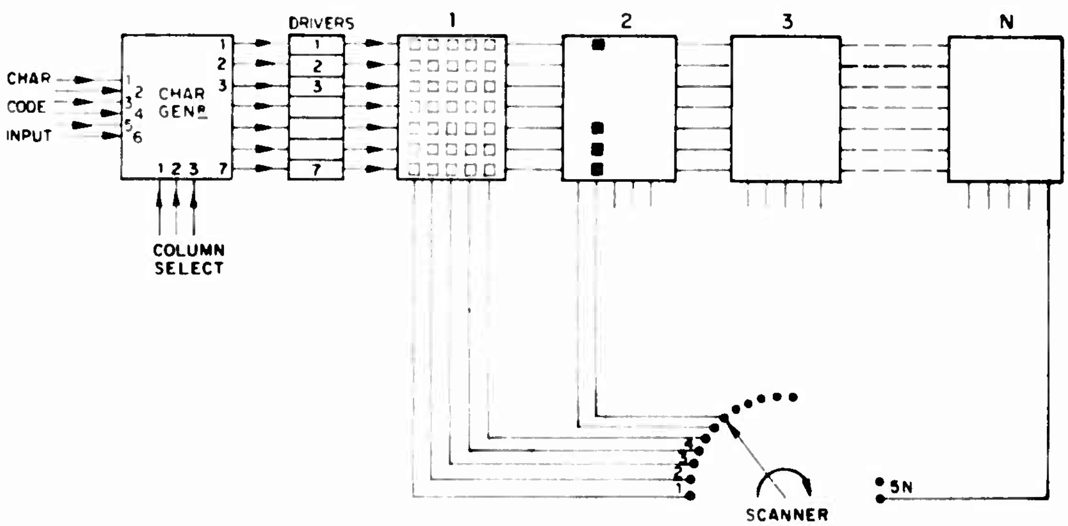


FIG. 23 COLUMN SEQUENTIAL ADDRESSING

Figure 24 illustrates a row sequential character drive. In this case the maximum duty cycle is approximately 14% but does not decrease much with an increase in the number of characters. A character generator which outputs the character information row by row is required.

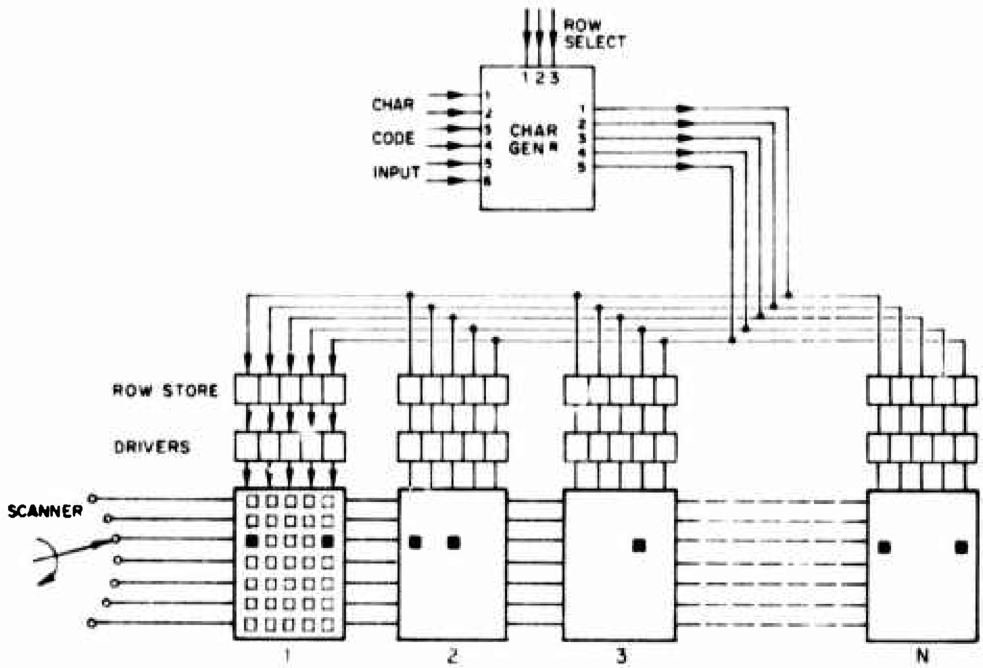


FIG 24 ROW SEQUENTIAL ADDRESSING

Referring to fig. 24 the sequence is as follows : the first character code is applied to the input of the character generator and the row select input is addressed to enable the output of the 5 bits of the character's first row. This information is steered into the row store of the first character. The second character is then presented to the character generator and its first row of information similarly fed to the row store of the second character. This is continued until the row stores all contain the first row information of all the characters. The scanner then selects the first row of the matrix and all the elements are driven in parallel. The row stores are then cleared and the process repeated for the second row. This continues until the last row has been displayed. The cycle is then repeated. The high duty cycle and simple scanning circuits are achieved at the expense of 5 x N bits of row storage and 5 x N drivers.

A simple calculation will show that, assuming a character generator response time of 1 μsec. and a display refresh rate of 250 Hz. then to display 100 characters the duty cycle will be approximately 12%. (This compares with 1% in the case of the character sequential method).

A dynamic data display refers to a display of synthetic data such as a computer processed radar display or a flight plan track display. We must therefore consider displays of several hundred lines resolution.

One way of exploiting advanced display techniques for a dynamic data display would be to provide a systematic scan similar to a television raster. This has been shown to be feasible using a piezo-voltaic crystal in conjunction with a field effect electroluminescent panel (54). But to benefit from such a technique would require the display store to be organized and scanned similarly to the display matrix. For a display of 1000 lines of 1000 points this would require 10^6 bits of memory for a monochrome display without grey scale.

Therefore the only practical means of addressing a dynamic data display is to provide random access. Proposals have been made to use scanning lasers to randomly address display panels but any such device would suffer from two of the disadvantages of the CRT which we would like to eliminate , that is the physical dimension and the requirement for switching voltages of several kV of present laser deflection techniques.

A practical alternative for random access addressing is coordinate or cross bar addressing in X and Y. Even so a display of 1000 x 1000 lines resolutions will require 2000 address lines. Considerable effort is now being directed towards the development of techniques for addressing matrix display panels and one of the potential solutions receiving some attention is the use of submatrices in the addressing circuits to reduce the number of lines.

There are other problems which must be taken into account when considering the display drive. As we have already seen the display must be refreshed at a sufficient rate to avoid flicker. Techniques with slow switching speeds impose constraints on the rate at which information can be written. Furthermore the duty cycle, and therefore the refresh rate, will affect the luminance.

One answer to some of these problems is a technique with inherent memory. This will not reduce the complexity of the scanning or selection circuits but it will eliminate the requirement for high refresh rates which imposes constraints on the amount of data displayed. Permanent data may be written once only and changing data will be erased and rewritten as required.

4.6 Advanced Techniques in ATC Display Systems

It is probably true to say that advanced techniques are now developed to such a point that they may be applied to certain ATC display requirements.

Early applications of the advanced techniques have been to alpha-numeric displays. The first devices to appear on the market were single character modules. These were soon extended to 3 or 4 characters per module and, no doubt with the fast expanding market for desk top calculators in mind, further developed to 9 and 16 characters.

There are many minor requirements in ATC systems for very limited alpha-numeric, or simply numeric, arrays such as digital clocks etc. which can therefore already be fulfilled by these techniques.

Large tabular data displays are becoming more common in ATC with the introduction of automatic data processing. Such displays have capacities of anything from 500 characters upwards. If we consider the state of the advanced techniques as they stand today, junction electroluminescence can be ruled out as a potential technique to fulfil such a requirement because of its unsuitability for large capacity displays. X-Y addressing would be essential to minimise drive and scan circuits. Furthermore to eliminate what would otherwise be an insurmountable problem of continuously refreshing the data, either a technique with inherent memory should be used or the display divided into segments, each one having its own character generator, drive and scanning circuits. The former solution at the moment could only be served by the gas discharge panel, but certain liquid crystals also have a potential to provide storage displays.

Dynamic data displays differ from alpha-numeric tabular displays in that they are essentially large area displays (>1000 lines resolution) but despite this large area the data density is generally low. They display a mixture of static or quasi-static data such as vector maps, etc. and continuously and sometimes rapidly changing data such as aircraft tracks. They will also contain characters and symbols and therefore character generation will be required. The problems of addressing such a display, as discussed above, are inhibiting the application of the these techniques to ATC requirements.

However a solid state radar display has recently been developed. Although it is simple in concept and limited in that it has one specific application, it represents a considerable step towards the use of these techniques in ATC systems. The display is a distance from threshold indicator (DFTI) intended for monitoring aircraft on the last few miles of approach before touchdown and is bright enough to be viewed in the high ambient lighting of the airfield control tower (fig. 25).

A matrix of GaAsP diodes forms the display medium. Spacing is 2.5 mm giving 1/4 NM resolution on the approach path. 640 diodes are used and brightness can be as high as 1000 ft.L.

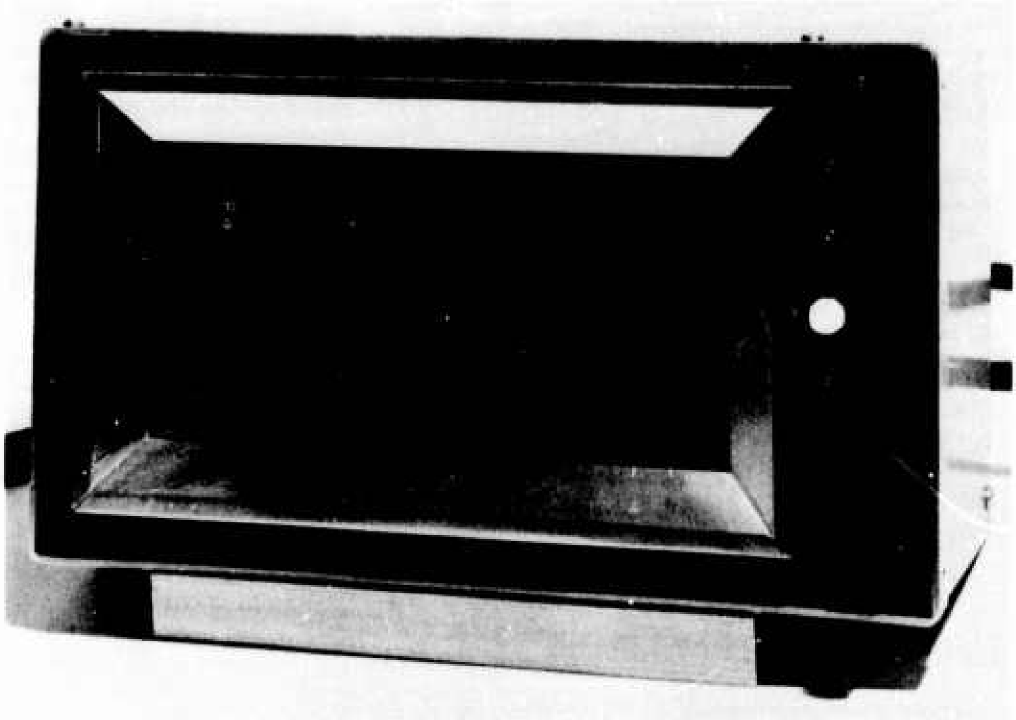


Fig. 25 : Marconi Distance From Threshold Indicator using GaAsP LED matrix (photo by courtesy of Marconi Research)

No one has yet produced a definite requirement in civil ATC for a large screen dynamic display. By large screen is meant a display area of 1 sq.m. or more. This would probably be a wall display. Its great advantage would be in coordination, for several viewers could all share exactly the same information. But to make it easily interpretable in such a role it would almost certainly have to be multi-coloured. Certain projection techniques exist for producing such a system but they suffer from various disadvantages (e.g. poor resolution, response time, etc.). Matrix techniques may prove to be capable of fulfilling such a requirement if it arises. If the problems which, at the moment, are holding back the development of a dynamic data display as discussed in the previous paragraphs can be overcome, then a large screen display will also almost certainly be feasible.

4.7 Economic Aspects of Advanced Techniques

For the choice of display technique in civil ATC systems cost will usually be the overriding factor. Therefore one of the first questions we must ask when considering the use of these techniques to fulfill our requirements is : "Can we do it cheaper than with a CRT?" But it is not sufficient to compare merely the costs of the various possible display elements. The cost of the interface and back-up equipment will often be a large proportion of the total system cost.

It can be assumed that, at the present moment all the techniques discussed above are cheaper than the CRT for an alpha-numeric display up to 100 characters capacity. Beyond this capacity gas discharge devices are particularly interesting, those using selfscanning techniques probably being the cheapest up to 2000 characters capacity and the ac discharge panel with inherent memory probably proving to be superior on a cost assessment up to and beyond 5000 characters. It has been suggested that by 1978 only liquid crystals will prove to be more expensive for a 500 character alpha-numeric display. For the complex display system, including both tabular and dynamic data displays of an automated ATC system it is believed that of the present potential advanced techniques ac gas discharge and dc selfscanning gas discharge techniques are the most feasible and that both of these may provide a display system for considerably less than the cost of a CRT system in 1980.

5. CONCLUSIONS

The cathode ray tube has been the universal medium for the display of dynamic data in ATC systems and, because of its versatility and other advantages, will continue to be so in the near future. However there is now emerging from the research and development laboratories a number of display techniques potentially more suited to the modern computer driven display systems. The incentive of commercial applications of these techniques is accelerating their development such that their performance is improving rapidly and their limitations and disadvantages compared to the CRT are being reduced or eliminated. The display designer of the future may be faced with a choice of techniques, the selection of which will depend on human factors, operational, technical and economic aspects of the particular application.

6. REFERENCES

(1) V.D. Hopkins :"The Psychologist's View", a previous paper in this AGARDOgraph.

(2) T. SOLLER : "Cathode Ray Tube Displays", MIT Radiation Laboratory Series, Vol. 22, New York, McGraw-Hill Book Company, 1948.

(3) A.E. POPODI : "Linearity Correction for Magnetically Deflected Cathode Ray Tubes", Electron Design News, January 1964.

(4) R.E. HALL : "Measurement Techniques for Cathode Ray Tubes" Electron. Eng., January 1967, pp. 86-90.

(5) "New Laminar-Flow Electron Gun promises brighter better CRTs", Electronic Design, 18, 1st September 1974, pp. 38, 40.

(6) S. LARACH, A.E. HARDY : "Cathode-Ray-Tube Phosphors : Principles and Applications", Proc. IEEE, Vol. 61, No. 7, July 1973, pp. 915-926.

(7) M.A. WOODS : "The Light Pen : A Graphic Input Device for Computer-Driven Displays", IEE Conference Publication No. 32, London, IEE, July 1967, pp. 230-237.

(8) Electronic Industries Association, "Optical Characteristics of Cathode Ray Tube Screens". January 1966, JEDEC Publication 16A.

(9) S. WOODCOCK, J.D. LEYLAND : "The Rear Port C.R.T." Electronic Equipment News, July 1969.

(10) J.W. ROBERTSON, F.J. SAVILL : "Rear Port Displays : Their Advantages, Limitations and Design Problems for Radar Applications", IEE Conference Publication No. 105, London, IEE, October 1973, pp. 207-212.

(11) G.T. NAGY : "Scan Converter Tubes and their Applications" Inform. Display, Vol. 2, March-April 1965, pp. 4-17 & May-June 1965, pp. 14-31.

(12) A.B.E. ELLIS : "A Direct View Storage Tube with Selective Erasure as a Data Terminal Display" IEE Conference Publication No. 80, London, IEE, September 1971, pp. 1-5.

(13) D.A. BOYLAND, D.J. NORMAN, D.S. ALLARD : "Bi-Colour Display Tubes" IEE Conference Publication No. 80, London, IEE, September 1971, pp. 39-47.

(14) G. MELCHIOR : "La Visualisation graphique en couleurs adaptées au contrôle de Trafic", L'Aéronautique et L'Astronautique, No. 35, 1972-73, pp. 13-18.

(15) A.F. MARTIN : "Penetration Colour Tubes are enhancing Information Displays", Electronics, 18th January 1973, pp. 155-160.

(16) H. WEISS : "Some Technical Aspects on the Use of Colour Displays for Air Traffic Control Systems" Report on the Controller-Computer Interface Seminar, Luxembourg, Eurocontrol Institute of Air Navigation Services, May 1973.

(17) R.E. CUSHERA : "Analysis of Phosphor Combinations for Multicolor, Multi-persistence Cathode Ray Tubes ", December 1969, Federal Aviation Administration, No. FAA-RD-69-57.

(18) T.E. SISNEROS, P.A. FAETH, J.A. DAVIS, E.H. HILBORN : Current-Sensitive, Single Gun Colour C.R.T.", Information Display, April 1970, pp. 33-37.

(19) F.J. AVELLA : Phosphor Screens for Current-Sensitive Single-Gun Polychromatic Cathode Ray Tubes" IEEE Transactions on Electron Devices, Vol. ED-18, No. 9, September 1971, pp. 719-724.

(20) F. FISCHER, H. THIEMANN : "Theoretische Betrachtungen über ein neues Verfahren der Fernseh-Grossprojektion", Schweiz Arch. Vol. 7, 1941 and Vol. 8, 1942.

(21) A. SANSOM, E.T. KOZOL : "Deformographic Storage Display Tube (DSDT) - A Light-Valve Projection Display Having Controlled Persistance", IEE Conference Publication No. 80, London, IEE, September 1971, pp. 325-329.

(22) J.B. FLANNERY : "Light Controlled Light Valves" IEEE Transactions on Electron Devices, Vol. ED-20, No. 11, November 1973, pp. 941-953.

(23) D.W. KENNEDY, C.R. GRAULING, A.J. DEVANEY, R.D. WRIGHT : "A Survey of Laser Display". Recent Advances in Display Media, Los Angeles, Technology Publishing Corporation, September 1967, pp. 93-109.

(24) V.J. FOWLER, J. SCHLAFER : "A Survey of Laser Beam Deflection Techniques", Proc. IEEE, Vol. 54, No. 10, October 1966, pp. 1437-1444.

(25) J.C. BASS : "A Review of Electro-Optic Beam Deflection Techniques" The Radio and Electronic Engineer, December 1967, pp. 345-352.

(26) T.C. LEE : "Light Beam Deflection with Electro-Optic Prisms" IEEE Journal of Quantum Electronics, Vol. QE-4, No. 7, July 1968, pp. 442-454.

(27) G.G. FULLER : "An Experimental Laser-Photochromic Display System", The Radio and Electronic Engineer, Vol. 39, No. 3, March 1970, pp. 123-129.

(28) E.E. LOEBNER : "The Future of Electroluminescent Solids in Display Applications" Proc. IEEE, Vol. 61, No. 7, July 1973, pp. 837-861.

(29) P.R. THORNTON : "The Physics of Electroluminescent Devices", London, E & FN Spon Ltd., 1967.

(30) D.G. THOMAS : "Light-Emitting Diodes", IEEE Trans. Electron Devices, Vol. ED-18, No. 9, September 1971, pp. 621-627.

(31) R.W. BRANDER : "Silicon-Carbide Electroluminescent Devices", Proc. IEE, Vol. 116, No. 3, March 1969, pp. 329-333.

(32) P.M. HEMENGER : "Materials and Technology for New Information Displays" AGARD Conference Proceedings , No. 96, October 1971, pp. 17.1-17.9.

(33) J. LEBAILLY, J.C. DUBOIS, J. FERTIN : "Dispositifs Electroluminescents Integrés" L'Onde Electrique, Vol. 50, June 1970, pp. 497-502.

(34) A. VECHT, N.J. WERRING, R. ELLIS, P.J.F. SMITH : "Direct Current Electroluminescence in Zinc Sulphide : State of the Art", Proc. IEEE, Vol. 61, No. 7, July 1973, pp. 902-907.

(35) E. SCHLAM : "Electroluminescent Phosphors", Proc. IEEE, Vol. 61, No. 7, July 1973, pp. 894-901.

(36) K.P. LALLY : "Multicolor Electroluminescent Displays", Recent Advances in Display Media, Los Angeles, Technology Publishing Corporation, September 1967, pp. 123-129.

(37) H. KARAWADA, N. OHSHIMA : "DC EL Materials and Techniques for Flat-Panel T.V. Display", Proc. IEEE, Vol. 61, No. 7, July 1973, pp. 907-915.

(38) I.A. SHANKS : "Liquid Crystal Materials and Device Developments", Electronic Engineering, August 1974, pp. 30-37.

(39) L.T. CREAGH : "Nematic Liquid Crystal Materials for Displays" Proc. IEEE, Vol. 61, No. 7, July 1973, pp. 814-822.

(40) A.R. KMETZ : "Liquid Crystal Display Prospects in Perspective" IEEE Trans. Electron Devices, Vol. ED-20, No. 11, November 1973, pp. 954-961.

(41) F.J. KAHN, G.N. TAYLOR, H. SCHONHORN : "Surface - Produced Alignment of Liquid Crystals", Proc. IEEE, Vol. 61, No. 7, July 1973, pp. 823-828.

(42) P.J. WILD, J. NEHRING : "Turn-on Time Reduction and Contrast Enhancement in Matrix Addressed Liquid Crystal Light Valves" Appl. Phys. Lett., Vol. 19, 1971, pp. 335-336.

(43) C.R. STEIN, R.A. KASHNOW : "A Two-Frequency Coincidence Addressing Scheme for Nematic Liquid Crystal Displays" Appl. Phys. Lett., Vol. 19, 1971, pp. 343-345.

(44) R.F. HALL, K.E. JOHNSON, G.T. SHARPLESS : "A Tabular Data Display using a Cross-Bar Addressed Glow Discharge Panel", IEE Conference Publication No. 80, London, IEE, September 1971, pp. 91-96.

(45) F. WALTERS : "DC Gas Discharge Matrix Displays" IEE Conference Publication No. 80, London, IEE, September 1971, pp. 7-11.

(46) J. SMITH : "DC Gas Discharge Display Panels with Internal Memory", IEEE Trans. Electron Devices, Vol. ED-20, No. 11, November 1973 pp. 1103-1108.

(47) W.J. HARMON : "Dot Matrix Display features Inherent Scanning Ability" Electronics, Vol. 43, 2nd March 1970, pp. 120-125.

(48) Y-S CHEN, H. FUKUI : "A Field Interlaced Real-Time Gas-Discharge Flat-Panel Display with Grey Scale", IEEE Trans. Electron Devices, Vol. ED-20, No. 11, Nov. 1973, pp. 1092-1098.

(49) R.L. JOHNSON, D.L. BITZER, H.G. SLOTTOW : "The Device Characteristics of the Plasma Display Element" IEEE Trans. Electron Devices, Vol. ED-18, No. 9, September 1971, pp. 642-649.

(50) W.D. PETTY, H.G. SLOTTOW : "Multiple States and Variable Intensity in the Plasma Display Panel", IEEE Trans. Electron Devices, Vol. ED-18, No. 9, September 1971, pp. 654-658.

(51) Z. VAN GELDER, M.M.M.P. MAATHEIS : "Principles and Techniques in Multicolour DC Gas Discharge Displays", Proc. IEEE, Vol. 61, No. 7, July 1973, pp. 1019-1024.
(52) D.C. KRUPKA, Y-S CHEN, H. FUKUI : "On the Use of Phosphor excited by Low Energy Electrons in a Gas-Discharge Flat-Panel Display", Proc. IEEE, Vol. 61, No. 7, July 1973, pp. 1025-1029.
(53) W. ROSENWEIG : "Organizations of Electroluminescent Diode Character Displays", IEEE Journal of Solid State Circuits, Vol.SC-5 , No. 5, October 1970, pp. 235-239.
(54) S. YANDS, Proc. I.R.E., 50, 2445, 1962.

Pages 605 and 606 are Blank.

USE OF COMPUTER IN AIR TRAFFIC CONTROL

by

B. Kirstetter, G. Maignan, V. Vachiery

European Organisation for the Safety of Air Navigation
EUROCONTROL
72, rue de la Loi
Bruxelles
1040
Belgique

Table of content

USE OF COMPUTER IN AIR TRAFFIC CONTROL

by

B. Kirstetter, G. Maignan, V. Vachiery

European Organisation for the Safety of Air Navigation
EUROCONTROL
72, rue de la Loi
Bruxelles
1040
Belgique

Summary

The present paper attempts to summarize the characteristics and applications of computers in Air Traffic Control. It is mainly based on the experience in the upper airspace.
After a short summary of the different applications, the general characteristics of hardware and software architecture are discussed. This includes a description of multiprocessor and multicomputer systems with their specific ATC oriented peripherals, real-time operating systems, programming techniques, data base and in particular reliability aspects and the associated problems of recovery management. This is followed by a chapter on the principles of data transmission in Air Traffic Control systems. The radar data processing deals mainly with mono- and multi-radar tracking aspects. Finally a summary of the different functions is given which can presently be provided by a flight data processing system.

1. Introduction.

It was at the end of the 50's that the first attempts were made to automate certain functions of Air Traffic Control by using Electronic Data Processing. The reason was a general feeling that, with the introduction of jet transport and the anticipated growth of traffic, serious consideration should be given to automation of ATC.

Almost in parallel several approaches have been started in different countries, all with the objective of increasing the capacity and efficiency of the system by improving coordination methods and thus releasing air traffic controllers from routine tasks.

A certain number of principles were laid down in the early days. For example the following two statements are extracted from a report of the first meeting of the Rules of the Air, Air Traffic Services and Search and Rescue division of the International Civil Aviation Organisation (RAC/SAR ICAO) of October/November 1958, throw some light on the initial concept of ATC automation.

"In view of the difficulty of ensuring that all aircraft will be fitted with special airborne equipment, which has been designed exclusively for air traffic control and which will always be serviceable in the air, automatic air traffic control equipment should be basically ground equipment and able to operate independently of special equipment in the aircraft."

"Unless it can be shown that the radar installations concerned have a very high degree of reliability in detecting targets under all atmospheric conditions, an automatic air traffic control system should not be based exclusively on the use of radar. This does not preclude radar as a primary source of information for the equipment, but any such use of radar should be supported by alternative means of obtaining the position of the aircraft".

Especially the second statement may be interpreted as a general suspicion about the integrity of radar information and the means to process it and indeed the first reliable radar data processing systems, which departed from the purely ground-based concept in that they relied heavily on the carriage of secondary radar airborne transponders, became only operational in the late 60's/early 70's.

Most of the systems which where implemented before that date therefore used flight plan information and pilot reports as the primary and sometimes only source of information.

After fifteen years of experience in ATC automation, computers are now generally accepted elements in modern systems but one has nevertheless to admit that computers have turned out not to be the magic remedy to the problems of ATC. Controllers may sometimes even feel that the additional workload imposed by the requirement to keep the system up-to-date is not always balanced by the services the system provides.

To arrive at the present situation one had however to overcome many setbacks, which caused in many cases considerable delays and cost overruns. There were the usual problems which anybody has who tries to apply data processing in his special application :

- the limits of a convenient utilization of "the computer" had not been fully appreciated and consequently the complexity of the software was frequently on a critical level.
- the management methods used for system definition and implementation was not always adequate for the size and complexity of the projects. In particular the software functions were sometimes only poorly defined.

Furthermore the technology required for a proper design and implementation was not always available and success of some projects were in particular dependant on progress in areas such as digital signal processing, graphic and alphanumeric displays, computer architecture and special programming techniques.

The present paper does not pretend to give someting like a philosophical approach to automation in ATC but is rather an attempt to summarize the application and characteristics of computers in ATC, whose implementation has been shown to be feasible. It is mainly based on the EUROCONTROL experience and data processing systems and makes frequent use of internal reports and documents issued on the occasion of courses and seminars.

2. Computer applications in Air Traffic Control (ATC)

Computers are used in ATC for the following purposes :

- radar plot extraction and transmission
- exchange of ATC information
- processing of flight plan and radar data in ATC centres

All three applications are mainly input/output driven and require fast response and high reliability. In the following a summary of the functions and the necessary peripherals is given for each application.

2.1. Radar plot extraction and transmission

The increasing size of the area of responsibility of ATC centres makes it necessary to implement several radar stations located at different places in order to achieve complete radar coverage. This entails a requirement to transmit radar information over long distances. This aim can only be achieved in an economical way if the large amount of information delivered by the radar itself can be compressed to such an extent that it can be transmitted through normal leased telephone lines.

With the present technology radar plot extraction and transmission control can best be accomplished by a combination of a special purpose hardwired target detection facility and a process control computer. The task distribution is frequently as follows : The target detection facility is responsible for analog/digital conversion, and some basic processing, e.g. distance and azimuth calculation. The computer performs primary (PR) and secondary (SSR) radar correlation, SSR code validation, discrimination of reflections, conversion into transmission format and control of transmission lines. The integral device is called "Radar Plot Extractor". In some extractors the computer tasks are mainly limited to formatting and transmission control task. The computers used for this purpose generally have a limited instruction set with an average execution time in the order of 2 μs. The system is duplicated for reliability reasons and performs parallel processing; switchover between the two computers in case of a failure is automatic.

The peripherals consist of :

- communications interfaces for medium speed synchroneous transmission (sometimes slow start/stop back channels) attached to a multiplexer channel,

- fast channel interfaces for reception of data from the target detection logic,
- magnetic tapes for initial program load and statistics recording,
- control facilities such as operator consoles, inter-processor communications and special monitoring devices.

2.2. Exchange of ATC information

Data which describe the dynamic environmental conditions (e.g. meteorological data) or a flight (e.g. flight plan) have to be transmitted to a multitude of ATC-sites. Traditionally this is done by semiautomatic means via the Aeronautical Fixed Telecommunications Network (AFTN), however even AFTN relay stations are more and more computerized to reduce the operating staff. Moreover the unprotected and slow AFTN telegraph channels (50 to 200 Baud) are being gradually replaced by protected medium speed lines (2400/4800 Baud) controlled by computers. This new network communicates with the data processing systems in ATC-centres and aerodromes.

The computer types are similar to those used in radar plot extractors, the reliability requirements are however higher than for Radar Plot Extractors as data are not repetitive. The processing system is duplicated, both parallel processing and restart techniques are used.

The peripherals generally used are :

- communications interfaces for medium speed synchroneous and slow speed start/stop transmission,
- magnetic tape and printers for logging and statistics gathering purposes,
- disks or drums for program swapping and to maintain the message queues,
- display or printer keyboards for local or remote input of data,
- control facilities as for the radar extractor,
- high speed interface (e.g. channel attachment) if the system is connected with a host computer.

2.3. Processing of flight plan and radar data in ATC centres

Within ATC-centres computers are used in increasing number to discharge the operational staff from routine work. In the course of development computers have first been used for the purpose of printing flight progress strips. The flight data were inserted manually; strip-printing was done instantaneously in a central position. Over the years systems became more sophisticated : they could receive data directly from AFTN, delay printing of strips until they were really required by the controllers and route them to the appropriate sectors. The most advanced systems actually in operation provide furthermore facilities to complete the strip information dynamically by the display of messages on electronic data displays (EDD) as a first step towards a system without strips.

The functions initially performed by computers in the field of radar data processing within ATC-centres were also rather limited, i.e. no attempt was made to improve by means of a tracking procedure the quality of the position information using a smoothing technique which correlates plots of the same target from several radars and antenna turns. The early systems where therefore only capable to receive plot messages sent by the extractor to filter them in such a way that each target is displayed only once and finally to convert the data into the format required by the radar displays. In the next generation of radar data processing systems tracking was introduced, sometimes limited on SSR data. In advanced systems the radar positions are additionally used to update the flight plan position in order to improve the precision of time values used by the system. This is especially important if time critical messages must automatically be transmitted to other centres. In a next stage of development the combined flight plan/radar information is used for conflict alerting. These systems are in a high degree EDD oriented and ask therefore for improved reliability.

No general computer characteristics can be given for this type of application because of varying requirements and system architectures. The computing speed, input/output capability and memory size reaches from the upper range of minicomputers for simple flight plan processing (FPP) - systems to large multiprocessor or multicomputer configurations for combined FPP/radar data processing (RDP)- systems. If a high input/output capability is required specialized computers frequently perform this task. Various restart or switch-over techniques ranging from slow manually controlled restart to fully automatic reconfigurations of computing,input/output and memory-elements of the multiprocessors are used. The typical peripherals are :

- disks or drums for memory extension and to safeguard data for restart purposes,
- magnetic tapes for legal and statistical recording,
- display keyboards to input flight-plan and system control data by operators,
- EDD's for the display of manually or automatically triggered flight information,
- EDD's with touch input device (TID) to enable controllers to communicate with the system,
- SDD's with the appropriate control facilities,
- printer for the display of flight progress strips and for logging of operational data,
- communications interfaces for the reception of radar data and for exchange of data with other centres,
- facilities for system control such as operator consoles, configuration switches and special devices for interprocessor communications and hardware monitoring.

2.4. Computer Systems

As already mentioned in the preceeding paragraph a large number of configurations are currently in use in ATC-applications. They range from simple uni-processors to very complex multi-processor configurations depending on the reliability requirements and the load to be handled and also include a multitude of input/output (i/o) choices.

2.4.1. Computer hardware

Computers consist of three basic components :

- the processor to decode and execute instructions and to control the i/o channels;
- the random access memory to store programs and data;
- the i/o channel(s) to perform the exchange of data between memory and i/o devices.

These components can be configured in several ways to form a suitable ATC system.

Uni-processor systems

In the simplest and most often used conventional uni-processor systems only one memory and processor-unit but sometimes several i/o channels exist. These elements must be completed by some control and monitoring hardware to make a workable computer.
The capacity of uni-processors used for ATC systems covers the spectrum from small minicomputers to very large computers :

- Minicomputers are used in application with much i/o, little processing and a simple data base. Their characteristics are typically as follows :
  . core or semi-conductor memory with a size from 4 to 64 K 16 bit words, some supervisor functions are sometimes in read-only memory.
  . instruction set is most suitable for logical and simple arithmetical operations with limited test on correctness of data and includes instructions for bit and character handling. The typical execution time is 2 µs.
  . frequently only one i/o channel (common data bus), but direct memory access for fast peripherals.
  . many priority levels for interrupts sometimes even with automatic switching of the processing environment, i.e. index registers and accumulators are automatically safeguarded when an interrupt occurs.
  . the reliability is very high (typical mean up-time 3000 to 10000 hours), therefore only limited error monitoring facilities exist, occasionally even no memory checking by parity bits.

- Large computers are ideal for applications with moderate i/o, extensive processing and a big complex data base. Their characteristics are typically as follows :

  . core or semi-conductor memory with a size up to 4 M bytes with byte or word structure, access control by hardware protection key. Recent systems have paging capability, a hierarchy of fast and slow memory and automatic correction of parity-errors.
  . powerful instruction set for logical, binary and decimal arithmetic processing with extensive tests on correctness of data, average execution time as low as 500 ns, sometimes automatic instruction retry on the occurrence of failures and special instructions for selective resetting of hardware.
  . several, independant multiplexer and selector channels, with capability for chaining of commands and retry of transfers on detection of unexpected conditions.
  . moderate reliability, depending on computer size mean up-time between 250 and 500 hours. Advanced and sophisticated hardware surveillance features and the modular hardware structure allow however to continue the system operation in a degraded mode or at least a controlled system shut-down.

Modular multi-processor systems

Single uni-processors do not satisfy the reliability requirements of ATC systems and have the further disadvantage that a completely new system must be installed when the load exceeds their capacity. To solve both the reliability and the load problem, modular multi-processor systems have especially been developed for ATC systems.

Modular multi-processors combine several independant memory modules, processors and i/o processors into one big computer. The principles of such a configuration are shown in figure 1 :

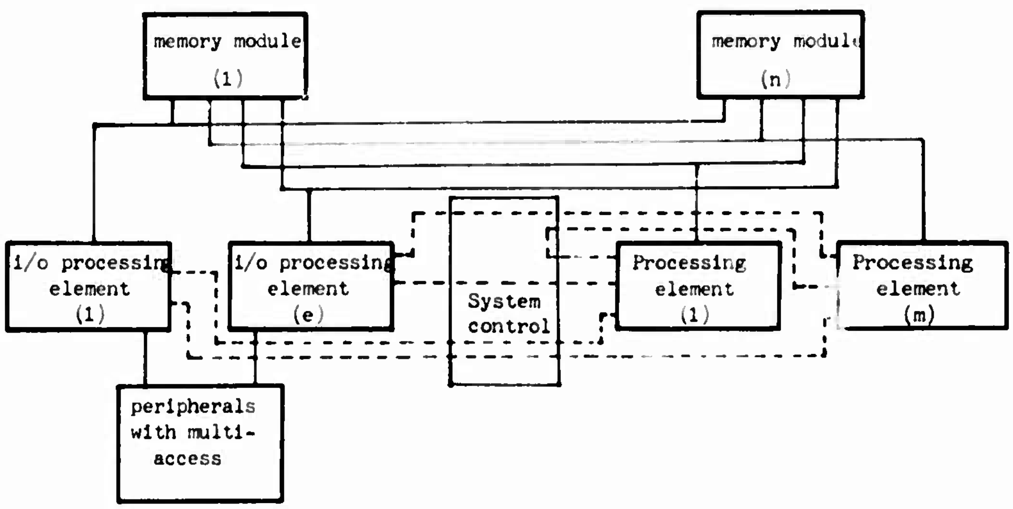


Fig. 1 : Multiprocessor system

A number of processing elements (PE) without i/o capability have access to a number of independant memory modules (MM). Specialised input/output processing elements (IØPE) with access to the same memory modules are responsible for the i/o operations. An IØPE again consists of a processor and i/o channels as described above.

PE's are the computational, logical and control elements and dispose therefore of special instructions for the control of the system, e.g. to define the status of MM's and IØPE's and the allocation of preferential storage to IØPE's. PE's inform IØ PE's of an i/o request and vice-versa IØPE's inform PE's of termination of i/o operations through special instructions. The real data are however passed via shared storage.

The configuration control and monitoring facilities allow, by manual intervention or automatically, to establish several modes of operation and control of the PE's, which monitor each other, MM's and IØPE's :

1. All elements cooperate in a single system in order to improve processing and i/o capacity.
2. Most elements operate as mentioned under 1, but some are in a stand-by mode and only become operational if an active element fails.
3. Elements are grouped to form several independant systems, e.g. the majority forms the operational system and a minisystem consisting of only one element of each type is under maintenance.

The best known special purpose ATC multi-processor is the IBM 9020 system. It is built of components from IBM 360/50 and 360/65 computers. Each system may consist of maximum of 4 PE's, 3 IØPE's and 12 MM's of 32 K words each. IØPE's are system 360 CPU's with 2 selector and one multiplexer channel each. The mean-up time is calculated to be of the order of 30.000 hours.

Multicomputer systems

Though standard multi-processor systems (e.g. UNIVAC 1100 series) fulfilling most of the above requirements are becoming more common, many ATC systems being actually operational or under development belong to the family of multi-computer systems composed of universal computers. (fig. 2)

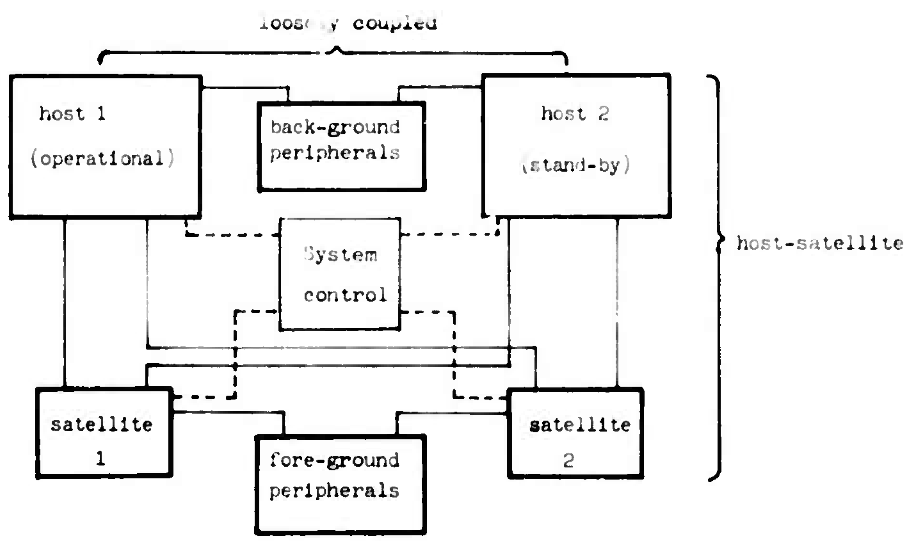


Fig. 2 : Multi-computer system.

Basically, two types of multi-computer systems can be distinguished :

- the hierarchical host satellite system with the main purpose to discharge the powerful host from i/o processing and
- the system with loosely coupled computers of the same size to achieve fail safe operation.

Host -satellite system

Host and satellite are linked via channel-to-channel links to exchange operational data and commands. They are generally in a master-slave relation, i.e. the satellite is under complete control of the host, which acts as the primary computing element of the system and which maintains the data base. It therefore disposes of the back-ground peripherals such as disk and magnetic tapes. The satellite or front-end processor performs i/o operations with transmission lines and the fore-ground peripherals such as displays and input devices used by the operational staff. It may be loaded with programs from a cassette recorder or from the host. Several satellites are frequently linked with the host, a typical example of such a system being minicomputer-equipped SDD's connected to a central computer.

The reliability of such a system is normally identical with that of the host. In some applications, however it is possible to use the front-end processors to provide minimum services during host outage e.g. it is possible to run the mini-computers in SDD's in an autonomous mode.

Loosely coupled system

Two or sometimes three identical of a capacity that anyone could handle the complete load are coupled via processor-to-processor links which are used for mutual status testing and abnormal condition alerting. Only one of the computers performs the actual work, the others are in a stand-by position. The exchange of normal data is via shared discs or occasionally via a high-speed channel-to-channel link. Foreground and background peripherals are via multiple access devices hardwarewise accessible from all computers. The exchange of data with a particular peripheral is normally restricted to the computer which has by a reserve command temporarily monopolised the access path. In many systems the reservation is unconditionally released by a time-out, in other systems this must be done by a command. In the latter case the computers must be capable to reset each other to force the release, when the computer which has reserved the peripherals fails.

The hardware reliability of a system which consists of two computers of the process control type is sufficient. A third computer is however advisable for systems which consist only of large universal computers due to the relatively low reliability of these computers.

Some existing ATC systems combine the host-satellite principle with the loosely coupled system, e.g. the MADAP/KARLDAP systems described later in this volume : a loosely coupled main computer system as described above controls a front-end processing system with a sufficient number of redundant members. Each main computer has access to each front-end processor and the background peripherals. All front-end processors are identical and have access to all foreground peripherals including the lines to receive radar data. Radar emergency processing is therefore possible during an outage of the complete main computer system.

2.4.2. Peripherals

Most of the peripherals of an automated ATC system operate in the technical background and only a very limited group is concerned with man-machine communications and is directly used by the operational staff, i.e. they work in the operational foreground. Man-machine peripherals consist of :

- synthetic data displays (SDD) for the presentation of radar information;
- the family of the electronic data displays (EDD) mainly used for the input and output of flight data. It includes the display keyboards and the EDD touch-input devices (TID);
- the different types of printers used for printing of flight progress strips and for logging purposes.

It is a common particularity of most of the above devices that they are nearly exclusively used in ATC applications, they are therefore very expensive and only offered by a very few firms. Their duplication is sometimes impractible for operational or financial reasons and full availability must therefore be achieved by a high inherent reliability and by a design and manufacturing which allow fast repair.

The auxiliary peripherals, which work in the technical background however are in most cases standard products, which are offered with identical or at least similar characteristics by a number of firms. The following devices belong to the auxiliary peripherals :

- discs or drums used for memory extension and to safeguard data for restart purposes;
- magnetic tapes are mainly used for logging purposes, but in some applications also to input off-line generated data and for restart purposes;
- line interfaces for the reception and transmission of radar, flight and meteorological data.

In practice all devices must normally be connected with several computers and must therefore be equipped with multiple access switches. They should be designed in such a way that a failing computer can never hold the connection for an indefinite period.

Man-machine peripherals

Synthetic data display (SDD)

In automated ATC systems the SDD also known as Plan View Display (PVD), is the most important source of information for the controllers. It displays digitized radar information generated by a computer system and in some models alternatively by a scan-converter. To perform its functions, the SDD must be able to display the following classes of data :

- alpha-numeric text.
- the dynamic track of aircraft as delivered by the radar data processing system and consisting typically of the position, the speed vector and an explanatory text label with callsign, SSR-code, flight level.
- the environment description such as routes, aerodromes, sector boundaries, markers.

For the active communication between the controllers and the system some input and designation facilities are provided they physically exist of :

- a light-pen for the selection of a target or an unlimited number of functions under program control by designating the corresponding target or function label.
- function keys for the execution of simple functions, e.g. for scale setting and
- touch input devices (TID) for the composition of more complex commands, e.g. commands including the callsign.
- a rolling ball for the target designation, it is normally used together with function keys and TID's or similar facilities.

In technical terms this means, the SDD must be able to display special and alphanumeric symbols and vectors of a good quality on a screen of about 50 cm diameter, e.g. the picture repetition frequency must for a bright display be in the order of 50 to 60 cycles/sec. The load to be handled is very high, typical maximum values are 2000 characters or special symbols, 500 vectors and 800 positionings with an average length of 1/5 screen diameter. Hardwarewise the SDD consists of a cathode ray tube with its associated control electronics for the generation of the deflection signals required for character and vector display, the power supplies, a processor and the picture repetition store (PRS). The latter form frequently a separate unit, which drives two or more screens. For the PRS both core and solid state memories are used. The processor is sometimes realised by means of firmware, but more frequently PRS and processor simply consist of a minicomputer, which performs the following tasks :

- control the high speed interface with one or two host computers (operational and stand-by), which deliver radar data and control information, e.g. for the light-pen operation.
- management of the data in the PRS, i.e. storing of radar data and if applicable control information.
- scaling, i.e. conversion of system coordinates into the coordinates requested by the particular SDD.
- generation of artificial afterglow by conserving the target position of up to five preceeding antenna turns.
- processing and if necessary forewarding to the host-computer operator inputs performed by means of rolling-ball, function keys or light-pen.
- and optionally direct reception of radar data from the radar extractor, i.e. by circumventing the host if it is down and processing of these data in order to maintain a limited display of radar information. The minicomputers in systems with this feature are rather independant from the host e.g. they have their own facilities such as cassette recorders for loading of programs and data.

Electronic data displays (EDD)

EDD's may be driven by a computer or an operator. They normally allow to input and display only alphanumeric data, a typical capacity is 25 lines with 80 characters each. The symbols displayed on the screen are either composed of a set of dots within a fixed matrix or for better readibility of a sequence of short linear or bent vectors.

Some recent EDD's have their own picture repetition store and input buffer and have therefore a high degree of independance from the controlling computer, this feature is vital, when due to the large distance EDD's must be connected via telephone lines. For local applications the memory of the controlling computer may be used as PRS. A variety of input devices exists :

- standard keyboards similar to those used for typewriter. They are general purpose input devices and are especially used to input variable length text strings of complex content. This combination EDD/Keyboard also known as display keyboard is generally not a special ATC equipment.
- touch-input devices (TID) are widely used as special input devices for the controllers to insert short, fixed format messages in a conversational mode, e.g. a clearance input. Physically they consist of an EDD and a set of overlaid keys or touches. The function of each key is dynamic under program control and is displayed on the EDD in form of a label under each touch.
- in a similar way light-pens can be used to converse with a computer system by selecting under program control a symbol or a function to be performed. Much effort has been spent in the past to discuss the merits of light-pen versus TID. For some time the TID had the advantage of higher reliability, but with the present technology a preference for the one or other solution depends more on the habits of individuals than on rational reasons.

Printers

Special purpose printers are generally only used to print flight progress strips. Their main characteristics are : the capability to print in two different colours or similar, to adjust on the beginning of the strip form not regarding the position of the form before the start of the print operation. Idealy the printer should also be able to cut and eject the strips. The printing speed is in the order of some hundred symbols/sec. These printers are generally not supplied by the computer manufacturer, to ease their adaptation to computer types, they are therefore not directly attached to an i/o channel, but use a medium speed half duplex serial transmission procedure and buffer the strip content before printing.

Back-ground peripherals

They are in general the standard peripherals of the selected computer type and require in general no special consideration. In small systems preference is given to fixed head discs or drums because of their better reliability, moving head discs are mandatory in large systems because of the required space.

The tape systems used for host and satellite computers should be compatible, to ease program production and analysis of data recorded by the satellite.

For the exchange of flight information, the common start/stop and synchroneous transmission procedures are applied. This allows the use of standard hardware interfaces. In multi-computer systems, multipliers provide the multiple access facilities. The format of the radar data transmitted by the extractors is frequently not compatible with the usual hardware. In many cases the development of special purpose interfaces can therefore not be avoided.

2.5. Software principles in ATC systems.

A computer in any application but in particular in real-time applications is only as effective as the software which runs it. The software required to run, develop and maintain real-time systems consists of :

- software specially produced to perform application oriented functions, e.g. reception and validation of flight plan data or processing of radar data. The operation of this type of programs will be discussed in detail elsewhere in this paper.
- system software which performs central functions such as resource management, scheduling of and communications between application programs and system control. This type of program is known under the names real-time operating system, supervisor or monitor.
- the data base which contains all static and dynamic information required by the system and application software.
- a special class somewhere between application and system programs is concerned with system start and stop and the management of back-up facilities. It is frequently known under the name recovery management.
- the tools and techniques required for program production and system maintenance.

2.5.1. Real-time supervisors

Early real-time systems were rather unstructured, i.e. no clear distinction between programs to perform the special application functions and programs responsible for central service functions existed. Such a structure was acceptable in small, simple systems programmed by one or two people only, but ended in complete chaos when systems and consequently programming teams became larger. In order to overcome this problem software systems had to be split into small program units with standard interfaces. Each program module in such a system with a modular structure performs a well defined function and is an independant unit commonly known under the name "task", under the control of a supervisor program. The term supervisor spans a spectrum from simple executives as used in a minicomputer which drives an SDD to a complete real-time operating system with the full range of support programs as for the 9020 system. The basic functions are however always the same :

- assynchroneous or cyclic scheduling of task execution according to task priority and availability of resources in a multi-programming environment.
- input/output (i/o) processing which includes interrupt handling as well as scheduling and control of the physical i/o operation.
- management of internal resources, i.e. allocation of core storage and processing elements in multiprocessor systems.

A multitude of extended central service and control functions are built around this nucleus in more sophisticated supervisors. They consist of :

- access methods for all peripherals including computer-computer links. They are the interface between basic i/o processing and i/o requests of application programs and provide a variable degree of convenience for the user, e.g. queuing of i/o requests and file management.
- facilities for recording of transient and permanent hardware errors for statistical purposes and to ease maintenance.
- system services such as timer management, gating facilities to control the access sequence to either programs or data, generalized routines for interception of abnormal conditions.
- debugging tools such as on-line modification of storage, tracing of program execution, general or selective dumps (snapshots).

Because of the stringent response time requirements and the particular way to use redundancy in an ATC-environment, special purpose supervisors are still frequently used in such systems. In order to make efficient use of the total capacity of the hardware, especially of large systems, the real-time supervisor should be able to run the normal operating system as a background job. This is of special importance, when a stand-by computer is idling most of the time.

2.5.2. Data base

One of the key problems in large ATC-systems is the organisation and design of the data base. Three conflicting requirements have to be satisfied :

- fast access to frequently used data e.g. radar stores by a many application programs
- safeguarding of dynamic data describing flights, the ATC - and hardware - environment and the system status for restart purposes.
- limitation of temporary locking of data bank access in order to avoid serialization of program execution.

In a good compromize the data base is therefore generally split into core resident parts, which contain the most frequently accessed data and disk resident parts which consist of the static environment description, rarely used data and the restart data. The structure of the dynamic data bank used by the application programs may be very simple i.e. in a single shot strip printing system it may mainly consist of a core resident flight oriented directory with references to the respective flight plan records on disc containing both raw and processed data.

In very complex systems with interdependent flight and radar data processing, e.g. in systems with conflict detection capability the data bank consists of many interweaved lists in order to minimize the access time if a multitude of search criteria must be satisfied. Search criteria may be in the simpliest case the callsign to access a specific flight, but it may also be the totality of flights which touch within a given time period a route-point or segment, i.e. all data elements of flights passing through a route point must be linked via pointers or indices with the data element that describes this point. This requirement leads to a large integrated data base used simultaneously by several tasks in a multiprogramming or multiprocessing environment. A locking mechanism is therefore required to control the access to such areas. This is achieved by the assignment of keys to those areas, they are inspected by each program that wants to use data in this area. If the key is already set the program is interrupted until the key is released, otherwise the key is set to inhibit access by other tasks.

The locking mechanism must be used with caution to avoid two critical situations :

- deadlocking, e.g. if two programs try to access the same two data areas with different lock keys the condition may exist that each program has enqueued itself to one of the areas and wants to access the other.
  Such a situation may occur if many areas with different keys exist.
- serialization of program execution e.g. most programs use the same area i.e. a new program can only get control on the locked area when the preceeding program has released the key regardless of its priority.
  This situation mainly occurs in large interlaced data bases.

2.5.3. Recovery management.

One of the areas where ATC data processing systems differ considerable from most other real-time systems are the recovery procedures to be supplied when the system fails. The term recovery management includes functions such as the reconfiguration of peripherals on the occurrence of device or control unit failures and systems recovery, i.e. system start, restart, stop and switching between computers. Successful recovery is bound to the existence of properly configured redundancy in the system controlled by the appropriate monitoring hard- and software and a continuous process of safeguarding of operational and system data during normal system operation for restoration purposes.

As a rough criterion derived from experience it can be assumed, that controllers performing on-route control in the upper airspace can tolerate without a requirement for special actions system outages in the order of :

30-60 sec. for the display of radar information
1- 5 min. for the display of flight plan information.

It goes without saying that a more rigorous criterion must be applied for approach control.

The lower figures apply for advanced systems with a high degree of automation which make extensive use of flight data display on EDD's. The higher figures apply for more conventional, flight progress strip based systems with separate flight plan- and radar data processing. In the following it is first attempted to discuss the general principles which have to be considered for device reconfiguration and system recovery, then the different action to be followed in a multicomputer system are described.

Reconfiguration of peripherals.

Full system operation must be maintained during outage of peripherals. The programs concerned with device reconfiguration have therefore the responsibility to control the use of :

- stand-by units of such peripherals, which are vital for the technical system operation but which are not used by controllers for ATC-purposes, e.g. magnetic tape units used for recording purposes or interfaces with communication lines;
- man-machine interfaces such as SDD's, EDD's and input devices for which normally no idling stand-by device exists. Another device of the same type in physical proximity has then to perform the functions of the failing device in addition to its own tasks;
- the intentional, temporary close down of single devices, complete ATC-working positions or sectors during periods of low traffic or to carry out preventive maintenance activities.

The conditions, which must always be satisfied, are to avoid complete loss of system or operational data and temporary unavailability to information used by controllers.

Ideally application programs should be relieved from all reconfiguration problems, i.e. the access methods should be the central place for all actions concerned with device reconfiguration. To achieve this, application programs must use fonctional, instead of physical device addresses and must be relieved from all queuing problems. The access method has to convert functional into physical addressees and vice versa and has in case of a failure, to attach the message queue of the failing device to the replacement device. In order to make sure that stand-by devices can virtually take over at any moment, their right functioning has to be monitored continuously by the regular exchange of test data or by a periodic switching between operational and stand-by devices. Unfortunately it is sometimes difficult if not impractical to accomplish such an ideal situation in a complex conversation oriented multi-processor system. The problem can be shown best by discussing the special problems for some peripherals in more detail.

Disc.

Discs have an exceptional position due to their capability to safeguard data for restart purposes after a system break down. That is why the data base should be duplicated, preferably on drives attached to different control units.

In the simplest way to organize data the content of both discs are identical, all write operations are therefore doubled, whereas read operations are only single. If one of the discs fails the system will continue its operation without any degradation. In order to protect the system against complete outage following a break-down of the remaining disc its access has to be inhibited during the copy procedure. The simplest way to do this is to enqueue all disc accesses until completion of the copy.

Displays

The information shown on a failing display must be transferred to an alternate device which has then to perform its own and the tasks of the failing device. This has to be done for all devices attached to one control unit if this control unit fails. Safeguarding can be carried out basicly in tow different ways :

- copies of all information shown on the display are maintained in a chain of records of a disc data set associated with this device. With such an organisation the conditions described for a centralization of the reconfiguration actions in the access method can easily be satisfied. A typical example of such an organisation is the display keyboard with a queue of messages attached to it, which can be dequeued manually by operator request;

- no direct copies of the displayed data but only the absolute minimum required for reconstruction of the data is kept on disc. It is obvious that in such a system the application programs, which had originally created the data will be involved in the restoration of the display. Such an organisation is well adapted to a system with data of a very dynamic structure, which are partly handled by the host and partly by a satellite.
- radar displays require only little action due to the periodic renewal of radar information.

Satellite computers.

Reconfiguration of satellite computers which perform front- and processing tasks for a large host computer is of variable complexity depending on the functions they perform. The relation between host and satellite computer may be master-slave or the satellite may operate rather independantly from the host. It goes without saying that the reconfiguration procedures in the two cases will be completely different.

- If a master-slave relation exists, the data transfer and therefore the complete reconfiguration process is normally under the control of the host. This includes requesting a dump for further failure analysis if an error occurs and loading during start-up. A slave in a stand-by position shall always be ready for immediate take over, i.e. it is loaded and runs a monitor program which makes a periodic exchange of test messages with the host. Once it becomes operational it is loaded with the tables and programs it requires to execute its task. Initiation and reconfiguration attached to the satellite should be under the control of the host to ease the restart of a stand-by satellite and to minimize interference problems. Application programs of the host are frenquently affected by the reconfiguration if no copy of the data displayed on the display is maintained on disc. The procedure described under displays applies.
- Independant front-and processing systems handle generally flight plan data, i.e. non repetitive data. But as in a master-slave relation, the exchange of data is under the control of the host. To prevent the system from loss or duplication of data or a critical delay of transmission during reconfiguration but also during a restart of host or front-and processor, the proper transfer of data between the two has to be under the surveillance of the usual communications book-keeping system, i.e. of a message numbering system associated with a positive or negative acknowledge of each message. Special synchronization messages have to assure resynchronization if the book-keeping system of one of the two has been corrupted.

System start techniques.

A variety of techniques are used to resume operation within a minimum of time after the break-down of a main computer in a system with redundancy on this level. Prerequisite of a restart with a minimum implication on the work of controllers is that all vital data used before break-down can be reconstructed during the restart. In systems with periodically renewed data this occurs automatically, otherwise all relevant data must either permanently be kept on disc or a stand-by computer must maintain its own data base.

Restart with data safeguarded on disc is the most frequent restart technique in real-time systems. In practice two types are used :

- The content of the dynamic data bank is periodically dumped onto disc in the check-point restart technique. During that copy period all data bank modifications must be inhibited, i.e. processing is basically limited on reception and buffering of data; all data-bank modifications, which occur between two dumps must be recorded in an appropriate sequential data set. This technique is frequently used for data banks with extensive linkage between different stores. On a restart on either the same or another computer the last dump is first updated from the sequential data set, than processing resumed from the last situation before the system failed.
- In a second technique, which has been favoured in the paragraph on the data base, the data base is already disc oriented and all information is automatically recorded on disc, whenever something is modified. On a restart only the core resident parts such as directories and event lists must be reconstructed from the data on disc. This technique is frequently used in big systems. In a multi-computer system as shown in figure 2 a typical sequence of restart actions is as follows :

Initialisation of the real-time operating system in the stand-by host-computer either by manual intervention or automatically after detection of a malfunction in the operational computer by the stand-by computer. This is sometimes followed by reloading of the front- and processors. Activation of the special restart programs, which perform functions only required during a restart phase such as the control of actions necessary for the restoration of normal operation. Testing of the proper operation of all real-time peripherals by the appropriate test programs and reconstitution of the data base selection e.g. loading of system tables. Because it always requires some antenna turns to rebuild the radar display, the normal radar data processing is started first however with inhibited outputs to displays. Re-processing of the individual, safeguarded flight plan records is resumed next. The coherence of data is checked record by record and finally the core resident directory and the associated event entries are recreated. Communications with the operational staff is only resumed once the data base has been completely reestablished.

This concerns the redisplay of radar information on SDDs , of flight plan related operator and controller information on the different types of EDDs and the continuation of strip printing. Finally reception of data from AFTN and other automated ATC centres is recommenced and the manual input devices are enabled for the input of new data.

The most advanced technique with recovery times depending on the system size and complexity in the order of 1 to 10 secs. is the switch-over between computers operating in parallel. The computers of such a system receive all data simultaneously, perform the same processing and maintain their own data base, but only one computer is authorized to make outputs. All members of such a system run special surveillance programs and exchange the results via the processor to processor link;computers delivering results which are obviously wrong are stopped. A continuous comparison of the results of normal processing is generally too expensive due to the time spent waiting for results. Parallel-processing is in practice only applied in systems with a straight forward processing and data base structre, i.e. in applications which include no or so very little interdependant events that a synchronisation of these events by comparison of results is feasible. Typical examples for parallel processing systems are : message switching systems in particular AFTN relays and operational telephone exchange systems.

2.5.4. Programming Techniques and Tools.

Because of the rapidly increasing programming costs and the higher reliability standards for software, advanced design- and programming- techniques, which allow to improve both programmer efficiency and program reliability are of growing interest for the production and maintenance of real-time systems. This can best be achieved by a straightforward program architecture, which allows easy modification and modular testing and by the use of high-level languages supported by efficient compilers.

Program design

The simplicity of program structures is the primary prerequisite for the production of reliable, easily readable and modifiable software systems. In extension of the well established principles already discussed in the paragraph on real-time supervisors, a program performing a task is further subdivided in separate pieces of independant code, which can be easily identified and hence implemented and modified.

The ideal program module is therefore a kind of self contained "black box" described by the functions it performs and by the data passed to and returned from it. This allows to understand what a module does without knowing how it does it. By this means the variety of connections between modules and hence the paths along which errors can propagate into other elements of the system are minimized. Obviously the applicability of this design principle does not depend on the programming language used.

If either a high level language or a suitable macro language is used, the same principle can be applied for coding. The price one has to pay for such a well structured program is an increased overhead in duty cycle and memory occupation.

Program languages

Most of the early real-time systems have been programmed in assembler language because of the good run-time performance, that is to say program size and execution time. But in the last few years a general trend toward high level languages can be seen in order to reduce program production and maintenance costs, to ease adaptation to other computers and to improve software reliability. Before comparing the merits of programming languages in some more detail their main functions shall be summarized :

Assemlber language

Assembler which may be considered for actual and future program development of large ATC data processing systems shall in particular provide the following functions :

- conditional assembly of basic statements
- linkage of separately assembled program segments
- macro instruction facility with the capability of nested calls

Programs written in assembler language are machine dependant but without doubt the most efficient in terms of run-time performance, that is to say size and execution time of object programs. Assembler programming is considerably eased if a set of comprehensive macro's is created for a specific application.

High level language

High level languages have been invented in order to enable non data processing experts to solve their problems with the help of computers and to run the same program without recoding on computers of different manufacturers. They can be classified into two basic categories :

- languages, which are independant from the run-time supervisor and which have consequently neither a core management for automatic pointer- based variables nor facilities for real-time processing such as task control or interrupt-handling. These languages generally contain statements for data declarations, assignment, branch and loop control and subroutine calls. They allow the insertion of program sections written in assembler language and the control of register allocation. The user must normally provide himself with the subroutines required for system communications and for input/output. Some languages however support some simple mechanisms for task and event control and for the easy implementation of access methods. Representatives of this language type used in the ATC field are ASTRE and MINICORAL.

- languages which provide a full set of real-time facilities and which consequently need sophisticated run-time routines to link with the appropriate supervisor. Representatives of this language type are PL/1 and JOVIAL. The high complexity of multipurpose languages such as PL/1 is the frequent reason of inefficient object code. Hence sometimes only a subset of the language excluding all inefficient features is used in real-time applications. This allows to make use of the extended error detection and diagnostic facilities of the respective compilers. The linkage of the run-time routines, e.g. to communicate with the supervisor, is frequently performed by special preprocessors.
- The display control languages used in the formatting process for input from and output to alpha-numeric and graphic displays constitute a special class of programming languages. They generally consist of a non-real-time program for the generation of mapping control tables and a set of real-time routines which perform by means of the control tables the actual mapping from internal computer representation into the desired format and the insertion of the control characters required by the hardware. It is the purpose of such a display language to ease modification of message content and format.

Comparison on merits

Software engineering is more and more influenced by a permanent decrease of hardware costs, a rapid increase of programming cost and by the trend to larger systems with higher reliability. Consequently, software production methods must be used, which allow to improve the efficiency of program production and maintenance even at the expense of requiring a more powerful hardware.

A solution is programming in a high-level language :

- programming in a high-level language is easier to learn and faster, because a "statement" does more than an assembler instruction. The advantages are however marginal for the initial writing of a program. A considerable saving of the total programming effort is, however, achieved if one also considers the modification and update during the normal life of a real-time system, that is to say over a period of 5 to 10 years. This is because programs written in a high-level language are easier to understand and the documentation automatically is always up to date.

- programs in a high-level language can within certain limits be adapted to other computers or supervisors by rewriting the subroutines for system communications. But this portability is considerably reduced if extensive use of machine oriented features such as control of register allocation is made.

- the biggest advantage of a high-level language is the capability of good compilers to detect most of the syntactical errors. This characteristic reduces the time spent during system integration and reduces additionally the probability of hidden errors, which are some of the reasons for low software reliability and high maintenance costs.

- the price one has to pay is the expansion factor for core occupation and duty cycle. For a good, modern compiler this expansion factor is however smaller than 1.5.

- furthermore, documentation is simplified because flow-charts generally become unnecessary.

Support programs

A set of special, non real-time programs is required during the development phase and to operate the ATC data processing system. These programs are used for the following purpose :

- system simulation to assist the designer in the selection of the appropriate hardware and in the elaboration of the software architecture and to ease system tuning especially if it is getting near the capacity limits. The system to be simulated is defined by its hardware characteristics i.e. processing and I/O capacity, storage size and by the software characteristics i.e. tasks, data base and by a statistical description of the input data which drive the system. The results, displayed as tables or histograms, are average and peak values for the utilization of the different resources and for the response time of the system. The models are generally programmed using special simulation languages.

- generation of the data base. These programs generally have two tasks :

  . preformatting of the data sets which are part of the dynamic data bank
  . bulding of the static data bank , which contains the environment description in computer format.

  Input for the program are statements to define the stores, their organisation and content. During program execution input data are checked on their validity and if necessary converted into internal format, inter-dependences between different data resolved, the different stores and records are chained and finally the data are transferred onto the real-time data sets. The preparation programs make extensive use of the utilities of the operating system, ideal is the input of data in a conversational mode through terminals.

- analysis of data recorded for legal and statistical purposes. These programs shall allow to reconstruct dynamic traffic conditions (play-back) for incident investigation and staff training, to print all messages which have been printed during a given period, to produce technical and operational statistics for the total system, individual sectors or working positions, to trace a flight on its way through the system and last but not least to evaluate the hardware error records.

- generation of test data. For system integration and testing, test data are required which allow the easy establishing of the various test conditions. A typical example is the generation of radar plots derived from the flight plans used by the flight data processing programs for the execution of maximum load tests.

3. ATC Communication

3.1. General

It is a truisme to state that an air traffic control data processing system must not be an "island of automation". This is the obvious consequence of the many data to be exchanged between ATC-centres when aircraft are transferred from one centre to another and so on. In practice data are presently exchanged between centres either in the broadcasting mode like flight plan, MET data, or AIS data or in the progressive step-by-step or conversational mode like updated flight plans, boundary estimates or radar handovers. In fact to be efficient, automated ATC-centres should be designed from the beginning as a computer network, the computers being strongly interconnected and exchanging data of the above nature. Up to now this aim could only be achieved in large countries like the United States of America, but unfortunately in many other parts of the world the problem is complicated by the international nature of the exchange and hence the difficulties to obtain a standardisation. One therefore has very often started by building "islands of automation" which are not very cost effective. Only now is one trying on a empirical and bilateral basis to interconnect the centres. So, to really understand the present situation, it is necessary to have an historical view of what happened during the past ten or fifteen years.

The manual ATC-centres are interconnected since a long time by the specialised AFTN telegraph network standardised by ICAO.
This network was mainly used in the past to transmit data of the broadcasting class that is to say MET data, AIS and flight plans. Data exchange was supplemented by the verbal exchange between controllers of the interactive and progressive data related to the handover from one centre to another.

At the beginning of the introduction of automation in ATC, the AFTN messages were of rather unstandardised format and the problem of using the AFTN network in an automated environment was raised.

ICAO implemented a group called "The Air Traffic Control Automation Panel" (or ATCAP) which defined and standardised a set of messages : FPL, CPL, CHG, EST and so on. The format and content of these messages was such that they could be understood by a human operator or processed directly by a computer. So the interconnection of automated ATC-centres with the outside started with this set of ATCAP messages transmitted via AFTN. Only the national nature of the problem in the United States gave the possibility of a strict standardisation of the computer-to-computer messages and procedures.

As a consequence of the requirement to allow manual and automatic treatment the ATCAP messages are difficult to process by computers because of their relatively poor syntax combined with a large amount of human errors. Only semi-automatic processing is presently implemented and a good amount of manual assistance is frequently necessary to make the computer understand the messages in particular the route field of the FPL.

For the EUROCONTROL area, the DATEX working group of EUROCONTROL and its Member States with participation of ICAO tries to solve this problem. This group standardised the ATC-message for automatic triggering of fligt plan activation in neighbouring ATC units via computer-to-computer links.

The problem of a remplacement of AFTN without any data protection by an advanced ICAO Data Interchange System (CIDIN) designed for the interconnection of automated ATC-centres has been treated since 1969 in the Automated Data Exchange System Panel (ADISP) of ICAO. The approval of a standard can be expected for 1976/77. The implementation of such a system on an international basis will take some time and provisions have therefore been made allow the integration of AFTN into CIDIN.

In the EUROCONTROL area at least 3 different data transmission systems are presently in operation :

- CAUTRA (Contrôle Automatique du Trafic) in France, which consists of several ATC-centres. It is a star network which uses a special procedure and non-standard messages in internal computer format.
- DUV (Daten Ubertragungs - und Verteilungs system) in Germany, which is mainly used to transmit flight data to a central flight plan processing system and to retransmit the processed data (strips) to the local ATC centres. It is a star network and uses a control procedure for alpha-numeric text, capable of supporting the integration of AFTN.

- point-to-point transmission of flight plan and radar data from MADAP to a Belgian MATRAC (Military Air Traffic Control Centre). The system follows the ATCAP recommendations and uses a control procedure for suitable alpha-numeric text.

3.2. Data flow and control (Fig. 3)

The main problem in communications between computers is the problem of data integrity, that is to say one has to make sure that data sent from a user process A e.g. a sending flight plan processing system to a user process B e.g. another flight plan processing system are correctly received irrespective of the way data are exchanged between A and B. In order to achieve this a set of operational rules governing the interaction between the two processes must be defined, these rules are generally known as a protocol. Hence user processes normally do not communicate directly with each other but by means of a specialised communications system, which can be a complex computer network with many nodes or a simple point-to-point connection. Similar rules must therefore be established for the data exchange between nodes and user processes and nodes. These rules are generally known as procedures. Lower levels e.g. the procedures shall be transparent to the higer level, e.g. the protocol.

Sophisticated protocols which allow opening and closing of connections, initiation and interruption of user processes and a complex control of the data flow within the network including recovery from failures are actually only used to control computer networks with heterogenous members, the best known being the ARPA (Advanced Research Project Agency) network in the United States.

ATC systems are still used as single purpose systems and one is therefore far from a standardisation on the protocol level. However, well established procedures for the exchange of one operational message as a single transmission (message switching) or in a sequence of small packets each one being transmitted separately (packet switching) between two computers (point-to-point) or via a computer network exist. These procedures include however some of the features and rules of a protocol, e.g. for link monitoring, recovery of lost or mutilated data and to establish or close connections.

A typical procedure may be described by the following elements :

- a set of control messages to be used to acknowledge messages, to establish or terminate the connection or to signal error conditions and to recover from them.

- a set of rules and agreements which define the behaviour of the transmitting and receiving stations under the different normal and abnormal conditions.

- the message content, that is to say :

  . a transmission code e.g. CCITT2 for AFTN
  . delimiters to indicate beginning and end and control information to define the type of transmission e.g. code used
  . a sequence number of the message and/or packet used as protection against loss of messages or parts of it
  . address of single or multiple destination(s) of the data
  . address of the origin of data, sometimes supported by the time of handing-in
  . the type of message and its transmission priority
  . the length of the data (text) field
  . the proper text to be transmitted
  . since transmission errors are likely to occur, the data must be protected against corruption. This is done by protection information, e.g. vertical and longitudinal parity, derived from the useful information, which is sent to the receiver together with the data. It is checked by the receiving station.

Message switching

Message switching systems use a store-and forward technique, that is to say a message originating in A with destination D travelling via points B and C will be received and temporarily stored in B and C until a circuit becomes available. The present computer-based message switching systems evolved from the automated telegraph switching centres and AFTN still consist of both types of centres. Consequently AFTN-messages are unprotected and the delimiters consist of a combination of rarely used symbols (ZCZC and NNNN). A limited link surveillance is performed by the regular exchange of a channel check message. Other control messages are used to request a retransmission of a particular message already received, to tell the transmitter that a message was lost, to request synchronisation on last message received or to inform the receiver on the last message sent to it. Computerised AFTN switching centres perform all these recovery functions automatically and must therefore store already transmitted messages for at least one hour.

Packet switching

Packet switching also uses the store-and-forward technique and has been designed for computer networks with data protection. The switching centres (nodes) of the planned CIDIN are interconnected through medium or high speed links, each of which will serve a given geographical area. The data processing systems in ATC centres will communicate with each other via this network.

Hence the CIDIN link procedure requires provisions to control the physical transmission between two nodes and to control the communication between the subscribers connected to it. Each packet has therefore a double header. The first contains the information for the physical link addressing one or several nodes, the second for the logical link, the subscriber.

3.3. Technical characteristics

Hardware function

The basic characteristics of the hardware used for data communication systems in ATC have already been described in paragraph 2.1.2.. It seems however useful to compare a system with a maximum of functions performed by specialised hardware with a more software oriented solution. The highest burden for the processor is the handling of line input/output. Three main solutions are actually in use. They require program intervention

- for each bit, this solution gives a maximum of flexibility at the expense of data throughput and implies a very short response time of the interrupt system. This principle is frequently used in small systems;
- for each character, this is a typical solution for specialised communications processors;
- for each block, this is the preferred solution, if a general purpose computer with a multiplexer channel is used. It requires a special hardware to detect the control fields and to report transmission errors.

Large, modern flight and radar data processing systems shift the line and procedure handling more and more into dedicated communications oriented front-end processors to discharge the central processing system. As a side effect this allows with a suitable system architecture to improve the system availability mainly because limited services can even be performed after emergency shut down of the central processing system by the front-end processing system.

Software functions

The essential functions performed by the software of a generalised data communication system using a protected link procedure can be summarised as follows :

- handling of the link control procedure as described in the paragraph 3.2. on data flow and control, this may include: detection, insertion and deletion of procedure oriented control information, basic error handling such as retransmission and checking of basic data in the message header;
- buffer and queue management. Several queues exist for each link :
  . the output queue with all messages to be output
  . the safeguarding queue with all messages already output but not acknowledged by the receiver
  . the queue with the acknowledge messages
  . a free record queue with all unused records.

All queues are maintained on disc for economical reasons and for back-up purposes. Incoming messages are after validation transferred from the input buffer to the output queue of the destination queue, simultaneously an acknowledged message is generated and inserted into the acknowledge queue. If the incoming message is itself an acknowledgement message, the message referred to is deleted from the safeguarding queue. Data from the output queue are transferred after transmission into the safeguard queue and kept for retransmission until they are acknowledged.

- routing and priority control is responsible for the selection of the appropriate output queue according to the address information found in the message header and for the assignment of the position within the queue.
- code and format conversion allows communication between different link procedures e.g. between the host and the satellite which performs the communication task. This may include the breaking of messages into packets at the origin and reassembly at the destination.
- management of the redundancy in the system, that is to say the switch-over from the operational to the stand-by computer, the reintegration of a computer to become stand-by in a duplicated system or a complete restart. The stand-by computer is generally idling and must rebuild its actual traffic situation from the different disc queues. Only AFTN switching centres sometimes perform parallel processing to achieve a character synchronisation thus avoiding the heavy retransmission of complete messages.

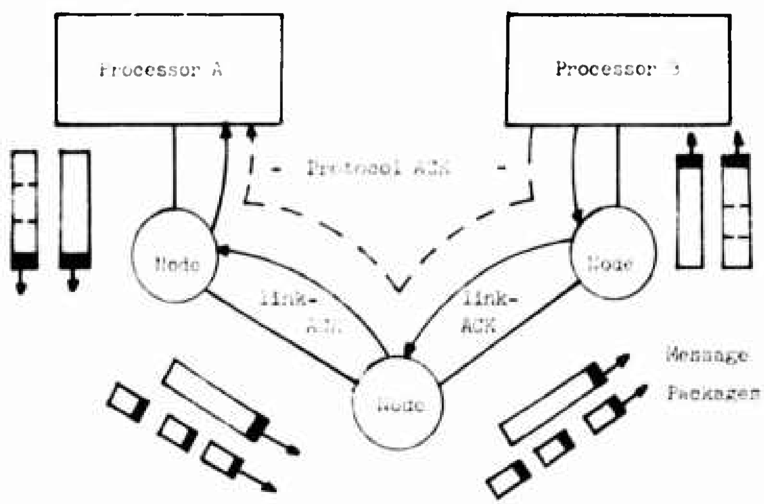


Fig. 3 : Control and flow of messages and packets in a network.

4. Radar data processing

Radar data processing comprises a number of functions necessary to transform analogue radar signals into a useful representation of the airspace situation on a Synthetic Data Display (SDD) and into the information required for a real-time update of flight plan data. The main functions in this process are : extraction and transmission of plots, radar tracking and display.
The related data processing functions are normally performed by distinct subsystems known as : plot extractor or radar digitizer, tracking system and display system.

4.1. Plot handling

It is the purpose of a plot extractor to convert video signals provided by the radar receiver into individual plot messages, each one containing information on the position and optionally the code of a particular target and to transmit the messages via telephone lines to ATC centres for further processing. The sequence of functions concerned with plot extraction can be summarized as follows :

- quantization (digitalization) of the video signals with elimination of information from fixed targets (MTI) and of noise from various sources, e.g. clutter by means of a correlation process;
- determination of the target position (range and azimuth) by correlation of successive hits of the same target during the same antenna scan by means of a sliding window detector;
- decoding of the SSR-code information and its validation by correlation of codes received from the same targets in successive antenna scans (in case of secondary radar only);
- association of primary and secondary radar information of the same target and combination into a single plot message;
- reduction of false plot rate in areas of hight clutter and further suppression of fixed targets by correlation of plots from successive antenna scans.

Generally the transmission of plots from the radar site to the ATC-centre is via leased telephone lines. The plot extractor is therefore also responsible for the buffering of plot messages during the scan of areas with a high target density, for the displatching to the different centres and for the management of the data transmission procedure.

Within the extractor plot messages undergo variable delays because of the time spent in the buffer during peak load periods. Each plot message is therefore time stamped in order to inform the receiving tracking system on the precise time of target detection. The time is frequently relative to north crossing of the antenna.

A high priority message indicating this event is transmitted without delay. The most relevant data in a plot message are :

- the target position in distance (ϱ) and azimuth (θ) relative to the radar site;
- the SSR-code including possibly the flight level if it is a secondary radar plot;
- a time delay due to queuing in the extractor;
- information on the plot quality.

4.2. Radar tracking

The plots obtained at successive antenna scans from a particular aircraft represent an uncorrelated set of position data. The purpose of the tracking process is to correlate the plots of successive antenna turns belonging to the same aircraft in order to obtain a clear identification, the more precise position and the speed vector of the aircraft in question. Monoradar tracking is the process of correlating plots from one radar, multiradar tracking is the same operation extended to several radars.

Monoradar tracking

The process of monoradar tracking can be outlined as follows :

In a steady state of the tracking system a track, described by its identification (normally the SSR-code), position and speed exists for each aircraft. The track is the result of a tracking process performed successfully in the past. In principle each plot received from the extractor has to be inspected, if it matches with this track.

In the simplest case this can easily be done by comparing the individual SSR-codes of plots and tracks. Tracking is then limited to the chaining of plots with identical codes. Otherwise in particular for primary plots this comparison must be based on the position information. This processing is more complex and represents furthermore a high load for the computer :

The first problem is therefore the reduction of the number of comparison steps. This can e.g. be achieved by dividing the airspace covered by the radar into azimuth sectors for plots and tracks. Plot-and track- sectors are of identical size, but dephased by half a sector . It is then sufficient to compare the plots of one sector only with the tracks in the two adjacent tracksectors, which cover the plot sector.

Comparison of tracks and plots must be done for a common reference time. For this purpose, the track position must be extrapolated for the time at which the plot from the next antenna scan is expected. All new plots which fall into a distinct area in the proximity of the calculated position are tentatively assigned to the track. The size of the area generally known as forecast window is a function of the track history and depends on parameters of the flight such as speed and turn attitude and on the quality of the data provided by the radar.

At the end of this process more than one plot may be assigned to a particular track or a group of tracks. Various techniques and algorithms have been developed to decide during the same or only next scan period on the most probable association of plot-track pairs to be used for further processing.

The selected plot is used to update track position, speed vector and the dimensions of the forecast window mentioned above. This calculation has the additional aim of smoothing the statistical variations of the plot positions provided by the radar. The relative weight of the track data versus the plot data in this process depends on dynamic factors as the trajectory of the target and the track history reflected in a quality factor and static parameters selected during system design and tuning. In this process a trade-off has to be made between sensitivity and smoothing. If the sensitivity is too high a low plot quality will be directly reflected in the track, that is to say the trajectory of an aircraft flying a straight line looks like a zigzag. If it is too low and smoothing consequently too high, plots of a fast turning aircraft can not be associated to its tracks which will therefore disappear.

Before the tracking process for a particular target has reached the steady state assumed in the above discussion, it must be through an initiation process to combine successive plots belonging to the same aircraft without knowing the predicted position of an already existing track. For aircraft with individual SSR-codes this process can easily be performed by comparing the plots of two successive antenna scans,and is normally automatic. For primary plots positional data of successive antenna scans must again be compared. Because of the generally lower quality of primary radar information, and the associated great number of false plots, this results in the frequent initiation of spurious tracks. In order to avoid a waste of computing power and because of the increasing use of secondary radar especially in the upper airspace, initiation of primary radar is in many systems manual, that is to say the controller has to designate by means of a rolling ball two plots belonging to the same object.

Multiradar tracking

Usually the information from several radar stations is required to cover the area of responsibility of a centre. A mechanism to select and combine at any instant the information from different radars referring to a particular aircraft is therefore necessary. This function, known as multiradar tracking, can be realized with different methods the best known being the

- NAS - system of the Federal Aviation Administration (FAA)
- MADAP-system of Eurocontrol and
- CAUTRA-system of the French Administration.

The NAS algorithm selects every ten seconds the preferred plot for a particular aircraft, from the various plots available from the different radars. The preferred plot is selected with the help old so-called "radar sort boxes". The common radar track is directly built from the preferred plot. The process is not so simple because in order to decide if a plot matches with a common track, one has to extrapolate the plot position which is provided at any instant by the radar head to the next ten second period of the common track.

Because of the lacking of monoradar tracking no speed vector is available for the chosen plot and one has therefore to use for the extrapolation of the plot position the speed vector of the track to which it is tentatively associated.
In the NAS, the plot extraction is made by hardware and the processing is made by a specialised IBM 9020 computer.

In the MADAP system, the process is different. The plots coming from the various radars are processed independently and in synchronism with each radar by a monoradar tracking process as indicated above in order to build "local-tracks". In areas where multiradar coverage is selected, local tracks are combined by a weighted averaging mechanism to build a common track. In order to update the common track, local tracks updated in synchronism with the various radars have to be extrapolated for the next period of the common track. This extrapolation is made by means of the local-track speed vector.
In MADAP, the plot extraction for the various radars is made partly by hardware and partly by software; the processing including the multiradar tracking is made on IBM 370/155 computers.

The radar tracking mechanism of CAUTRA is similar to that of MADAP with one important difference : the common track is a virtual entity.
In fact, at any instant in time the "common track" is nothing else than the preferred local track dynamically selected. Each local track is updated in synchronism with the particular radar it is coming from. There is a change of "rythm" and of "phase" when one jumps from one local track to another, however this has only little influence on the quality.
On the other hand, no further extrapolation is required and consequently one source of error is eliminated.

In the CAUTRA system the plot extraction is also made by hardware for the majority of the radar sites. The central processing is made on CII 10070 computers.

It is difficult to say which one of these processes is the best because if quantitative results are available for the NAS, only partial results are available for MADAP and practically no results are available for the CAUTRA.

When implementing a multiradar tracking system special consideration must be given to the problem of absolute precision. This is to avoid that different positions for the same object will be indicated by different radars. In order to overcome this problem a precise and sophisticated projection and coordinate conversion system, which includes slant range correction has to be used to compute the X, Y position from the ρ , θ positions in plot messages. Furthermore the precise position of each radar site must be known and the alignment of the radar must permanently be monitored. To solve the alignment problem some kind of real-time quality control is necessary, e.g. to consider the misalignment in the coordinate conversion process.

Display of radar information

Periodically, e.g. with the scan frequency of the radars, the track and associated data are transmitted to the display system. A description of the information shown on the SDD and on the further processing performed by the SDD can be found in 2.4.2.

Plan-track correlation

In order to allow identification of an aircraft represented as track on an SDD by its callsign rather than by its 4-digit SSR-code or to enable the update of flight plan data by means of radar data, callsigns and SSR-codes have to be correlated. For this purpose it is necessary to provide a link between the radar and flight plan data and to inspect on each creation of a new track with discrete code if a compatible flight plan exist. The correlation process is in particular very efficient if the management and allocation of SSR-codes is performed automatically by the flight data processing system.

4.3. Technical characteristics

A variety of technical applications for radar data processing in ATC have been developed and implemented. They range from completely hardwired plot extractors to very complex multiradar tracking systems.

Even in the field of plot extractors no clear trend exists and in fact both specialized, hardwired systems and solutions mainly based on fast process control computers as indicated in 2.1. have been implemented.

For radar data processing systems in ATC-centres three basic solutions can be distinguished :

- dedicated systems, either completely hardwired or with specialized processors with programs in read-only memory. Because of their simplicity they have a high reliability but are generally only designed to perform a straight forward monoradar tracking;
- dedicated systems realized with process control computers. Both stand-alone solutions and systems connected with a flight data processing which performs SSR-code management exist. This latter solution also allows a retransmission of correlated tracks for the purpose of flight plan updates;
- integrated radar data/flight data processing systems which run on large computers. This is at present the most common solution in big system because of the relative simplicity to exchange data between the programs responsible for radar data and flight data processing.

Radar information represents a high data rate - a typical load figure is 250 active tracks - and is very repetitive - a typical repetition cycle is 10 secs. The consequences for the software are therefore :

- optimization of programs and data base structure in terms of execution time
- no sophisticated restart procedures and no requirement for safeguarding of many dynamic restart data are necessary.

5. Flight Data Processing.

5.1. Introduction

Flight data processing systems have initially been developed to discharge the operational staff from the time consuming manual production of flight progress strips on the basis of flight plan information received via AFTN, telephone or R/T. The flight plan information had to be fed into the computer in a form suitable to produce strips without extensive computer analysis of the route data. The system output was normally immediate printing of strips at a single position.

The extent of services and hence the degree of sophistication of later systems increased, i.e. more powerful syntax and semantics analysis programs have been developed in order to simplify data entry and strips are now generally only printed at the time when and near the position where they are actually required.

The most advanced systems presently in operation supplement the information presented to the controller in the form of strips by the display of dynamic data such as last minute time modifications on EDD's. They also support the direct processing of flight data messages automatically received, the computer-assisted conversation with the system and the update of flight plans through radar information.

The basic information on a intended flight or position of a flight is contained in the flight plan message in its different forms of presentation. It therefore constitutes the only basis to plan ATC operation.

The rules governing the classification, routing and the content of flight data have been established by ICAO and are laid down in Doc. 4444. Flight data processing performed by computers is in particular concerned with the :

- acquisition of flight plan information
- processing of these data and
- display of data to controllers and transmission to neighbouring units.

5.2. Flight data acquisition

The conventional means to input flight plan data is manual via a typewriter. In more recent applications they are replaced by display keyboards which allow data entry in a conversational mode and hence an easy correction of erroneous data or the update of flight plans. In principle there should be no obstacle of treating flight data exchanged according to the methods described in the chapter on ATC communications completely automatically by the computer.
In practice there are however the problems of the low technical quality of data transmitted via AFTN and of the ambiguities in the logical interpretation of data. Both situations cannot normally be resolved by the computer and can only be solved by the manual intervention of an operator. One arrives therefore at a system, which operates automatically when this can be achieved with a reasonable effort but which has also the facilities to have difficult cases solved by human intervention. Such a procedure is very flexible and allows the smooth introduction of an increasing number of automatic functions.

The data used by a flight plan processing system for the description of a flight are : aircraft identification and type, SSR-capabilities, flight rules and status, aerodrome of departure and destination with the respective time estimates, the flight path described by air-routes, point references and level, the allocated SSR-codes and the time estimates at boundaries between responsibility areas.

The data are transmitted and generally also entered into the computer in form of standardized messages, the most common ones being the :

- Filed Flight Plan message (FPL) containing information to be provided to ATS units relative to an intended flight or position of a flight. It is used in the simultaneous mode, i.e. FPLs are sent simultaneously to all ATS units concerned, that is to say all stations receive exactly the same information about the flight. Because of this general nature and its early dispatching the FPL contains no boundary and SSR data.

- Activation message (ACT) which is proposed for Europe to complete the data in the FPL message by boundary estimate and SSR data and shall be exchanged between neighbouring automated ATS units.
- Current Flight Plan message (CPL). It combines the information of the FPL and ACT-messages and is transmitted in the step-by-step mode between computers, that is to say each ATS unit receives only the data needed by itself and by the subsequent units. The applicability of the CPL message depends heavily on the quality of the transmitted data i.e. flight level, boundary estimate and SSR-data. This condition is generally only satisfied, when the flight plan information is permanently updated by radar information.
- Repetitive Flight Plan message (RPL) which is used for the description of flights, which are activated on a periodical basis, but at least once a week. RPL s are usually kept on magnetic tape, sometimes, also on disc. Their message content and treatment is similar to that of an FPL received via AFTN. The periods and dates of validity are part of the message.
- Cancellation message (CNL) which is used to cancel flight plans specified by FPL or CPL.
- Messages which are used to modify data transmitted earlier as FPL or CPL.

Apart from these standardized formats a number of special messages are used in most systems e.g. to modify flight plans in a conversational mode after activation by means of touch input devices (TIDs) or light pens.

Further information received by the flight data processing system are the meteorological data expecially those required for time calculations such as the wind vectors and the temperatures.

5.3. Processing of flight plans

The actual processing of the flight data is concerned with the

- interpretation and validation of the flight data in terms of the particular ATC-environment
- creation and managment of events which describe the flight from the ATC point of view

Interpretation and validation

The first step during the interpretation and validation process is the build-up of a data structure which contains all information required at the appropriate time by controllers or neighbouring systems. Whatever organisation for a data structure may be found to be most suitable for a particular system, a logical record describing a flight should contain the following basic data elements:

- the initial, raw flight plan data formatted into its constituent fields which are the basis for all further processing. These data are generally supplemented by a number of warning data to indicate a possible operational problem, e.g. lack of present SSR-code at activation time. Such a situation has to be corrected or confirmed by operator intervention.
- the processed data which consist of
  - . the global data such as aircraft type or SSR capabilities. The extraction of these data is a straightforward process without special problems
  - . the extracted route information consituting a set of points which describe all relevant geographical locations of a flight with their attributes. The proper processing of these data mainly found in field 15 is one of the crucial problems to be solved.
- the status data which describe the status of a flight from the data processing point of view; this information is mainly required for restart purposes and is a dummy element before activation of the flight.

The programs responsible for syntax analysis and extraction of flight plan data are rather complicated and subject to frequent modifications. In order to give them therefore a maximum of generality and to divorce them from the particularities of the message syntax and the geographical environment, a data structure called static data bank is used for the interpretation of the flight plans and other data received by the system.

It contains tables with a description of the syntactic rules and the environment, e.g. the lateral and vertical airspace boundaries, the air routes, the reporting, coordination and boundary points, the airdromes and their access routes etc.
For each of the above items the relevant characteristics are given, e.g. for a route segment, its length and the authorized flightlevels in each direction. Especially in a system which uses FPLs which frequently do not specify boundary points, one may have problems to identify these points. One has to extract from the total route description that part which concerns the area in question. To achieve this also a description of the immediate surroundings of the area is required in order to recognize the boundary points. At the end of the validation process such a self contained record is stored on disc. Obviously a reference to each record must be kept in core. In a simple strip-printing system this may be just a directory. In a system with flight plan derived conflict prediction one may be faced with the design of a complex data bank structure which allows logical relationships to be made between the flight data records, e.g. to tie together all aircraft passing a particular point.

Event management

Management of events is the process which controls the execution of functions performed automatically by the system once a flight has received a time attribute. Each event therefore refers to a processing task as well as to data.

The effect caused by an event may be : display/erasure of data, enabling/disabling of communications and enabling/inhibition of inputs. The nature of an event may be linked to a time, e.g. reuse of SSR codes after plan cancellation or to a statistically defined of dynamically derived point on a flight path, e.g. crossing of a boundary.

Plan activation :

It is the process of building up the dynamic data tables from the "master record" which has been described in the preceeding chapter. It is triggered by the expiration of a time e.g. estimated time of arrival at a boundary or directly by an input e.g. to indicate the occurance of take-off. The activation function is combined with the calculation of significant flight parameters and subsequent events. The processing following immediately the activation includes in particular :

- allocation of the SSR-code to flights as required for plan/track correlation
- calculation of the estimated times over the significant points of the flight path
- determination of subsequent events to be triggered automatically during the further flight; e.g. printing of the remainder of strips or communication with other systems

Display and communication

Most display, erasure and communications events are linked to crossing of boundaries between systems or sectors. The result of such a transition is normally printing of strips or display of data on EDD/TIDs of the receiving and erasure on the sending side. Communications between systems is performed by means of CPL-, ACT- or similar messages.

Plan cancellation

This is the deletion of all data which refer to the flight to be cancelled. It may be be triggered at any time by a message input via keyboard or data link or it is performed automatically after the flight has left the zone of responsibility.

Control management

In sophisticated systems with a high degree of automation the assignment of tasks to controllers and the transfer of flights between controllers and control states is also performed under computer assistance.

Assignment of a complete flight or parts of it to a controller takes place either automatically or a purely geographical basis or, with a higher priority, manually by input order. The actual transfer of responsibility on crossing the boundary between two zones of responsibility may be proposed manually or by the system, however its acceptance is always manual.

On special itineraries (e.g. OAT flights) or in a holding state some processing functions e.g. divergence detection or even strip printing may be temporarily suspended. The occurrence of such an event may be conditioned by a manual operation or may be derived from the route description.

Generally display, erasure or transmission of data and enabling or disabling of some data entry functions goes in parallel with the above transfer of responsibility. Typical examples are the display of controller task messages or the inhibition of all facilities to modify flight plan data by controllers not being in charge of the flight.

Divergence checking

Aircraft positions may be recalculated in intervals of some seconds on the basis of the flight plan with the purpose of either displaying the "flight plan track" in case of absence of a radar track or for comparing the radar and flight plan position of the aircraft. A detected divergence may be used to update the flight plan automatically or to alert the controller, who has to command the update via input. Prerequisite for the implementation of this feature is the access to the correlated tracks of the associated radar data processing system.

Conflict detection and flow control

One of the main intentions of automation in ATC is the implementation of automatic control functions. Although several trials have been undertaken to use flight data in order to detect the possibility of conflicts along the predicted flight trajectory, no completely satisfactory solution has yet been found. This is mainly because of the lacking precision of the available data and because of the uncertainties during the climb-and descent phase. More promising are recent attempts to solve the problem of long range traffic planning based on flight level occupancy. The present operational "flow control" systems are completely self-contained and independant from the conventional ATC data processing systems.

5.4. Data output

The most frequently used output of flight data processing systems are still the flight progress strips, which have the major advantage that there is always a "hard copy" which will not disappear even in the catastrophic breakdown situation

The layout and content of data displayed on strips and EDD's vary from system to system, however the data mentioned in para. 5.2. for the description of a flight are found again. They are supplemented by the reporting point to which the message refers and by the overflight time. Some messages contain information for only one, others for several reporting points. EDDs are not the only means of displaying dynamic flight plan data. Occasionally limited information representing urgent tasks is also shown in tabular form on SDDs. Format and content of messages are subject to frequent modification. In order to avoid the resulting reprogramming effort the use of special display languages or table driven extraction and formatting programs is becoming common practice.

For the transmission of flight plan information to neighbouring centres CPL- and ACT - messages also described in para 5.2. are used. Extended communications facilities in the NAS-system allow the transfer of control on flights with computer assistance in an identical way between controllers of the same or different centres.

------ooOoo------

References

1. F. Dumanian, D.Clapp "Large Scale Systems - A Study of Computers Organisation for Air Traffic Control Applications" Report N° DOT-TSC-FAA-71-15

2. EUROCONTROL, several authors "Report on the Study Weeks for the Austrian Administration" EUROCONTROL Experimental Centre 22.4.1974 - 10.5.1974.

GENERAL ASPECTS OF DATA FLOW

by

Horst A. Neumann, Dipl. Phys., Dr. rer. nat.

Messerschmitt-Bölkow-Blohm GmbH

EDP-related system studies and system design

8012 Ottobrunn, F. R. Germany

Index:

GENERAL ASPECTS OF DATA FLOW

by

Horst A. Neumann, Dipl. Phys., Dr. rer. nat.

Messerschmitt-Bölkow-Blohm GmbH

EDP-related system studies and system design

8012 Ottobrunn, F.R. Germany

1. INTRODUCTION

Data processing, no matter if performed manually or automatically to deduce results according to logical instructions, proceeds consistently with a logical scheme of data flow. From a data source data are periodically or sporadically called by data acquisition devices. The acquired data are properly pre-processed and stored with respect to intrinsic data processing, which generally needs access to all assembled and filed sets of data. Any result of intrinsic data processing is either added to the data files or led off to a machine, usually by generation and transmission of control signals, or to a person, usually by generation of sound or speech or by display of characters and graphs.

Regarding very complex systems such as air traffic control, data acquisition and data processing is performed in cooperation of man and machine. Indispensable dialog usually is accomplished by output of machine's results via teletypes, lineprinters, plotters or luminous data displays and by input of man's acquired data, results and decisions via functional keys, keyboards and touch displays.

A block diagram of data flow in air traffic control is presented by fig. 1. Manual data input is combined with data acquisition, data output is differentiated with respect to the receiver, i.e. pilot (see: data output to aircraft) and controller (see: data display) respectively. Therefore main consideration is with automatic data flow. Nevertheless the diagram, as seen from a general point of view, represents implicitly all possible concepts of air traffic control executed on-ground, including even the concept of manual control, which leaves time-consuming tasks of data acquisition and preprocessing as well as all crucial tasks of intrinsic data processing to controllers.

In the following an introductory description of main features and of present status of each block or function is presented.

2. DATA ACQUISITION

2.1. Acquisition of radar data

Analog signals for slant range and azimuth derived from primary and secondary radar antennas may be fed directly to analog radar indicators. To avoid the disadvantageous wide-band transmission and to make radar data accessible to digital computers, the analog data are generally converted to digital data by digital plot extractors and, along with binary coded transponder replies (identity, altitude - see /1/), are transmitted to control centers via telephone lines.

Data renewal times of rotating antennas are within the range from about 10 seconds down to about 2 seconds. Data processing techniques relying on smaller renewal times and more reliable data, for example automatic conflict detection and resolution, call for phased array antennas or at least for discrete addressed interrogation of transponders. However these techniques are not yet available for civil purposes.

2.2. Acquisition of aircraft replies and a/c-reports

Position, altitude, identity of aircraft and supplementary data are frequently interrogated by controllers and replied by pilots via radio communication channels, or these data are reported spontaneously in case a stipulated flight level or reporting point is passed by the aircraft. Consequently acquisition of a/c-replies and -reports is more flexible compared to radar data acquisition, but imposes quite a lot of workload on controllers.

Usually the acquired data are recorded on paper strips (called flight progress or control strips), each of which is related with one announced aircraft. To make these data accessible to digital computers, either manual input or automatic radio transmission is required. Corresponding techniques, f.e. touch input display supported by sophisticated software, discrete addressed interrogation, data exchange by digital data link and automatic voice recognition, are still in the state of development and trial going along with eager disputes.

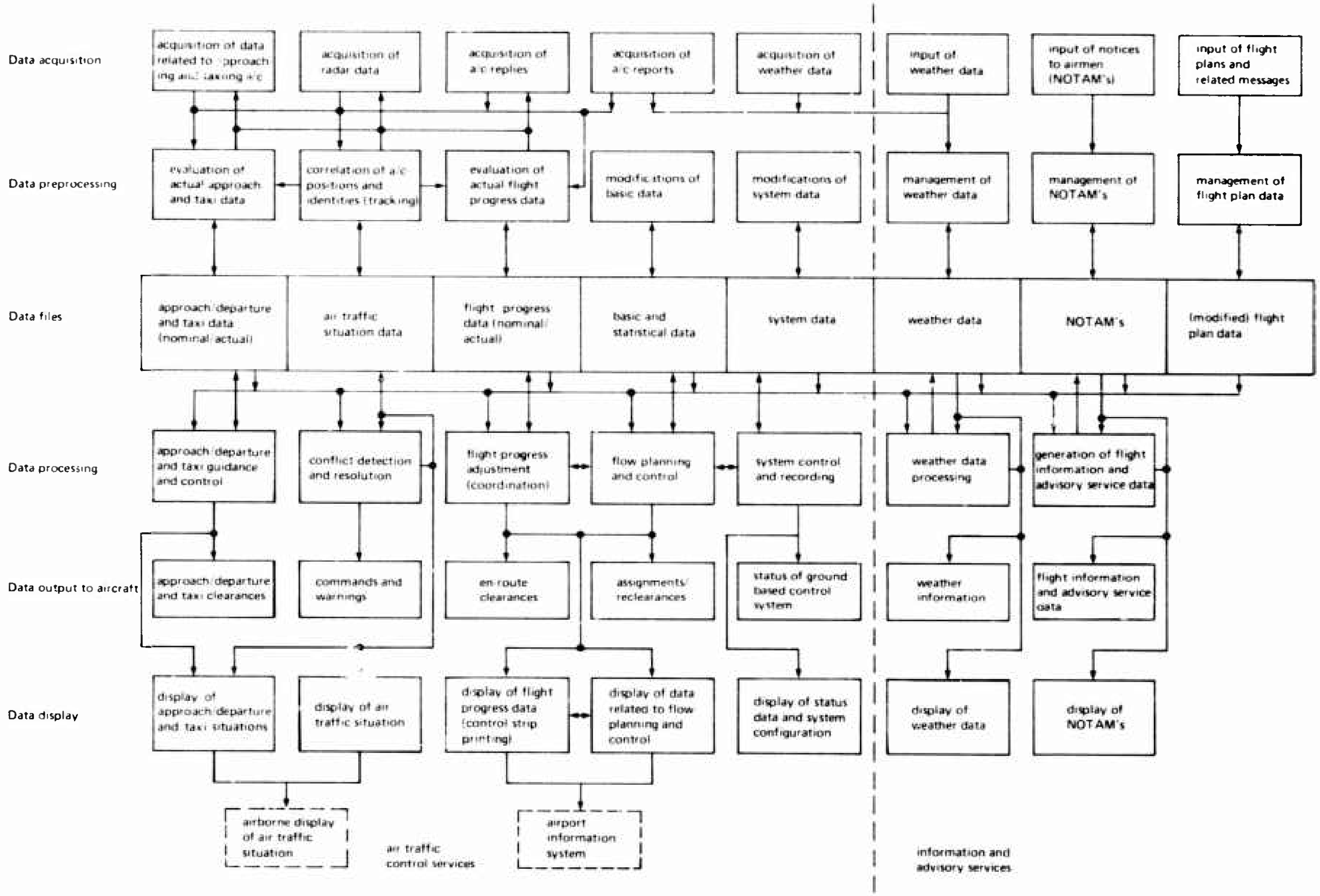


Fig. 1 Diagram of data flow

2. 3. Acquisition of data related to approaching and taxiing aircraft

At present acquisition of data related to approaching and taxiing aircraft is generally confined to voice communication. Approved automatic data acquisition is still pending. Merely plot extraction applied to airport surveillance radars and taxi guidance by use of optical or electromagnetic sensors are feasible.

Future microwave landing systems (MLS) offer a potential solution, i.e. derivation of digital approach data aboard and ground-based data acquisition via an advanced air/ground data link.

2. 4. Input of flight plans and related messages

Flight plans, which contain data of flight intention and of aircraft equipment, are usually delivered to aeronautical information service offices (see/1/). There contents of flight plans are checked and converted to teletype messages, which are transmitted to corresponding planning and control centers for data processing. On special occasions (f.e. if an aircraft is switched from visual to instrumental flight rules) flight plans are delivered on flight via radio communication.

Short-term messages related to flight plans, declaring cancellations, alterations, delays for example, are handled by controllers' assistants, who check and input revised data via teletypes or keyboards attached to tabular displays.

Other data related to information and advisory services, f.e. notices to airmen (see /1/), are also delivered to aeronautical information service offices and after approval are converted to digital messages for transmission and distribution.

2. 5. Input and acquisition of weather data

Weather data, particularly altitude-dependent temperature and wind profiles for defined geographical areas, are provided by meteorological stations 6 to 12 hours in advance. These data are usually transmitted via teletypes to flight weather service offices for further treatment.

Direct acquisition of weather data at present is confined to the ability of special radars resp. plot extractors to derive analog weather pictures resp. digital weather messages related to bad weather fronts. Future systems exploiting advanced digital data link may include controlled interrogation of temperatures and wind data along ATC-routes recorded by sensors of flying aircraft and of temperature, wind and visibility along runways and taxiways recorded by sensors on ground.

3. DATA PROCESSING

3. 1. Data preprocessing

Data preprocessing supports intrinsic data processing by verification and proper preparation of acquired data. Therefore primary and secondary functions may be distinguished.

Primary data preprocessing comprises the reception of data blocks, the check on imperfections of acquisition (f.e. errors of input, failures or interference during transmission etc.) and eventually the release of a repeated acquisition.

Whereas primary data preprocessing depends on the mode of acquisition, secondary data preprocessing depends on the type of data. In case departing, cruising, approaching and taxiing aircraft are causing radar data, a/c-replies and a/c-reports to be received, the identification and after that the continuous correlation of identity, position and altitude for each aircraft (i.e. the process called "tracking") is one important task, the evaluation of actual flight progress data (time-over, altitude, velocity, heading with regard to defined control points) is another one.

In conventional ATC systems tracking and data evaluation is performed manually by controllers. A more up-to-date conception aims at the utilization of digital computers to relieve controllers, but a lot of problems with data acquisition (arising from technical deficiencies of secondary radar and short-comings of digital plot extractors) and with performance of computers (related to lack of processing power and of reliability) are not yet solved statisfactorily. Practically all trials known to the author expanded routine tasks of controllers by adding tasks to be attributed to man/machine interface, such as release of operations, data input, switching and monitoring.

Less critical examples of secondary preprocessing of data include modifications of static data and management of dynamic data files.

3. 2. Approach/departure and taxi guidance and control

Even very busy international airports rely on automatic data processing for approach/departure and taxi guidance and control only marginally. There are four main reasons: first, real-time data for direct use in digital computers are not yet available (see above); second, logical models to yield the optimum sequence of approaching aircraft by appropriate selection and utilization of standard approach routes and procedures are extremely complex; third, corresponding software is of little value, as long as the flight progress data of all approaching aircraft are not continuously updated according to automatically evaluated actual data; last not least fourth, computer/controller/pilot/aircraft interface problems making up reaction times which harmfully disperse the computed overflight times assigned to the outer and middle marker are not yet solved.

Therefore approach/departure guidance and control is essentially performed by means of radar vectoring and depends mainly on decisions of experienced radar controllers. Compared to the theoretical optimum scattering of overflight times deviate up to ± 20 seconds. Therefore EDP-supported approach/departure guidance and control bears latent potential to improve runway capacity of busy airports, which are bottlenecks to air traffic today.

Automatic taxi guidance and control is understood to be an important supplement to any future landing system devised to overcome bad-visibility conditions.

3.3. Conflict detection and resolution

Main objective of tactical control is immediate conflict protection, i.e. conflict detection and conflict resolution. Careful monitoring of the air traffic situation, perfect judgement of the development based on experiences, quick reaction, direct communication and maximum responsibility are features. Therefore well-trained and experienced specialists (radar controllers) do the job. Automatic data processing, even with coming techniques, is only good enough for support.

At present electronic data processing is in the state of development and trial. Problems to be solved are related to those mentioned with tracking and approach guidance and control: acquisition of real-time data is imperfect with respect to the degree of entirety and accuracy; preprocessing of real-time data (tracking) is incomplete; logical procedures are very complex (see /2/); demands on processing power of computers are extremely high (see /3/); efficient computer/controller interface is lacking.

Demands on computer power ask for multiple data stream processing. Therefore associative array processors offer their capability. Results of simulations and real-life trials with the application of an associative processor prove, that at least conflict detection with regard to other aircraft and to obstacles could be automated (see /4/). This would provide the basic requirement for automatic transmission of conflict warnings (besides air traffic informations) to aircraft flying primarily according to visual flight rules (VFR) via speech generation and voice channels (i.e. AVAS = automatic VFR advisory service - see /5/) or via digital data link (i.e. IPC = intermittent positive control).

3.4. Flight progress adjustment (coordination)

Immediate conflict protection is very well supported by mediate procedures, i.e. procedures of conflict protection applied to air traffic in advance. These comprise: flight progress prediction; transformation of predicted flight progress data to nominal flight progress data; flight progress monitoring by continuous comparison of actual with nominal data; finally, according to actual situation, derivation of assignments to be transmitted to aircraft for better actual-to-nominal adjustment or, if not possible, updating of nominal flight progress data, i.e. nominal-to-actual adjustment.

However regimentation of air traffic with reference to a fixed air space structure (fixed routes, defined flight levels, restricted areas), consequently reduction of freedom to move is an attached precondition, because for practical reasons flight progress data have to be referred to defined (along route fixed or flexible) control points. Furthermore flight progress prediction asks for provision of flight plans describing flight intention.

Predicted flight progress data of different aircraft may be in conflict with each other. In such cases it is tried to solve the conflicts by modifications of flight plans. This coordination of predicted flight progress data yields the nominal flight progress data. From these data mandatory clearances with respect to aircraft movements are extracted.

There is one important point, which is particularly striking in case automatic processing of flight progress data or system overload is discussed. Some basic data describing the threshold values to be imposed upon air traffic flow are needed. These values are dependent on system capacities, such as are runway capacity, traffic flow capacity (related to control points) and control capacity, to mention the evident ones. However the most valuable definition and the evaluation of capacities is a problem, which is not yet solved satisfactorily.

This fact is a severe draw-back to application of automatic data processing to flight progress adjustment. Another problem is the lack of a comprehensive logical model for flight progress prediction and for solution of mediate conflicts, though the proposed and tested trajectory prediction model EROCOA /2/ developped for conflict detection is assumed to be a productive approach.

At present electronic data processing is confined to computation of predicted flight progress data on the basis of fairly crude models. The results are usually printed on control strips via teletypes about 10 to 30 minutes ahead of time to allow manual coordination by assistant controllers or special planning controllers. Control strips are largely used to place clearances agreed between planning controllers (nominal data) and a/c-replies/reports (actual data) and related notes on record. Issue of updated control strips replacing preliminary strips is released after input of actual data or of manually estimated data deviating distinctly from printed data.

Control strip printing was the first milestone passed towards a successful application of electronic data processing to functions of air traffic control. In some cases the computed data are not printed but presented on electronic data displays. However disadvantages must be accepted, since written record of data is no longer possible and failures of the EDP system may occur. One important advantage stands for compensation: Data input can be relieved essentially by program-controlled display of selected data for touch input; therefore input and automatic exchange of data related to coordination (actual data, estimates, requests, clearances, conditions of hand-over) between planning controllers is feasible, thereby avoiding time-consuming switching of telephone lines. Yet a critical reduction of data input with regard to real profit of an operational system is believed not to emerge unless automatic updating of flight progress data by means of radar data is established and the process of prediction is improved.

3.5. Flow planning and control

Problems with flight progress adjustment arise, when air traffic control positions are loaded heavily. Then coordination imposes considerable delays on announced flights or yields restrictions such as limited acceptance rates regarding inbound aircraft. The reason is, that intervention as consequence of short-term coordination of predicted flight progress data is coming too late. Only flow planning, i.e. long-term coordination of traffic flow being an integral part of planning resp. scheduling of air traffic can help to avoid air traffic jam in order to reduce severe restrictions. Unfortunately only raw flight plans of periodical and, at the most, of prospective flights are available in due time. Therefore the portion of immediately announced air traffic (composed of private, business and military aircraft flying according to instrumental flight rules) must be considered by statistical methods.

A valuable or even necessary supplement to flow planning is flow control, which is defined as medium-term control of all the actually announced flight plans and of related messages to detect impending overload of control positions and to balance load by either approval or modification of flight plans or, if possible, by well-timed extension of control capacity.

To sum up: flow planning is applied to raw flight plans (and statistical data) and is mainly aiming at air traffic jam protection comprising jam prediction and resolution; flow control is applied to announced flight plans (and related messages) and is mainly aiming at overload protection; finally flight progress adjustment is applied to approved flight plans and is mainly aiming at mediate conflict protection to support tactical control.

At present national flow planning offices are generally being instituted. Electronic data processing is usually confined to prediction of overlapping air traffic, whereas possible solutions are derived manually.

Flow control mostly is executed manually by planning controllers taking into account data on control strips and related messages. Often a simple procedure is applied: As soon as an impending overload of the adjoined radar controller is individually recognized by the planning controller the acceptance of further aircraft is restricted.

Problems with automatic data processing are similar to those mentioned with flight progress adjustment: lack of defined and evaluated capacities and deficiencies of logical models for overload prediction and solution.

3.6. System control and recording

Capacities depend on the status of the control system, for example on the extent of degradation of data acquisition and processing facilities, if the control capacity is regarded. Therefore the main object of system control is monitoring of the total system in order to discover troubles of operation or failures of components and, in that case, to determine implication with efficiency and capacity. Eventually reconfiguration by switching over to redundant components or, if not possible, an appropriate system degradation is attempted by system control functions

The aspect of recording is emphasized by the fact, that statistical interpretation of recorded data leads to analysis of air traffic flow and of system operation. Corresponding results may support forecasts of traffic flow for flow planning and control and of system capacities for flow control and flight progress adjustment.

Progress of data acquisition and data processing implies semi-automated system control. However at present nearly all system control is done manually. Recording usually is restricted to preservation of used control strips for statistical or even jurisdictional evaluation.

3.7. Weather data processing and generation of flight information and advisory service data

Processing of weather data occurs with flight progress prediction and with indication of bad-weather areas on radar displays. Electronic data processing is a basic requirement.

Generation of flight information and advisory service data beyond conflict warnings and advised flight progress data essentially means information retrieval and sorting, therefore is closely related to management of data files. Classical stack processing is suitable, therefore application of EDP is widespread.

4. DATA OUTPUT AND DATA DISPLAY

4.1. Data output to aircraft

Results of intrinsic data processing which serve for guidance of aircraft (i.e. clearances assignments, commands, warnings, flight information and advices) are transmitted from controllers to pilots via VHF-voice communication. Automatic data transmission via up-data-link to reduce controller's load is not promising unless extensive application of automatic data processing is achieved. Apart from this fact there are two procedural draw-backs: first, transmitted data need conversion to optical or acustical code to be quickly and safely recepted by pilots, which is very costly and - more important - very problematic with regard to anthropotechnical aspects; second, voice contact between pilots and controllers implies an important psychological power, which is too eminent to be relinguished. That is not to say, that techniques like AVAS or IPC (see: conflict protection) are without great promise.

4.2. Display of approach/departure and taxi situation

Nominal and actual approach/departure data gained by automatic data processing are presumably most clearly indicated by graphical display of approach/departure profiles with attached labels containing call sign and altitude. If two aircraft are found on crossing ascent and descent, the vertical projection of profiles seems to be more adequate for control and guidance. In such special cases a free area of the screen could be used for quick-look indication of projected profiles following touch-input.

Display of approach/departure profiles as indicated is in the state of research and development. Crude versions are occasionally tested. At present common display of air traffic situation (see below) is used for approach/departure guidance and control, whereas generally direct observation of runways and taxiways is utilized to control and guide landing and taxiing aircraft.

4.3. Display of air traffic situation

The key to tactical control is agreed to be a device indicating the actual air traffic situation. Display of tracked radar data represents the most advanced technique. Usually position data are displayed graphically in form of tracks (i.e. vectors or series of plots revealing velocity and heading), which are accompanied by labels containing transponder replied data (i.e. identity or correlated call sign and altitude) and on demand data related to the flight plan. Preconditions comprise digital plot extraction of position data, correlation of primary and secondary radar data, tracking of correlated data, correlation of plan data and radar data (also to support tracking) and last not least reliable hardware.

At present analog radar indicators with rotating radial and separate SSR-panel are still wide-spread. In case digital displays are already operationally used, extracted radar data usually are plotted directly, that is without extensive preprocessing. Although false plots are only suppressed crudely and correlation of labels often must be supported by manual input, improvements with respect to steady, bright and clear data display are acknowledged by controllers. To avoid problems arising with failures of computers processing data to be displayed, frequently the ability of equipment to switch over to direct display of analog data, also including the crossfading of labels, is provided.

4.4. Display of flight progress data

Procedural control and coordination of air traffic is based on the display of flight progress data. In the beginning of air traffic control assistants converted flight plan data to flight progress data manually and noted the results down on paper strips. At present control strip printing is generally introduced. But even today the control strips must be handed over to controllers, who use strip-tables for flexible arrangement of strips according to the expected air traffic situation. Also manual updating of control strips is still exercised besides occasional reprinting of strips (see: flight progress adjustment).

Though electronic tabular displays offer useful features for display of flight progress data, application is often refused because of lack of flexibility and reliability. As already mentioned tabular or touch-input displays will presumably not succeed generally before automatic updating of flight progress data is solved satisfactorily.

4.5. Display of data related to flow planning and control

Flow planning, i.e. long-term coordination of raw flight plans, is principally an off-line function, therefore is closely related to data stack processing and time-sharing. Tabular displays with attached dialog key board to influence coordination by manual data input and lineprinters for output of final results are adequately used.

Flow control causes a lot of data to be exchanged, just as flight progress adjustment does. Therefore a device with capabilities of relieved manual data input is required. At present touch input displays and related software are in the state of extensive trial.

4.6. Display of status data and system configuration

Display of status data and of system configuration serves system monitoring. Modern computers are equipped with electronic displays to supply actual status data on request and indicate critical status as long as logically possible. With extensive application of electronic data processing to support air traffic control, status of displays, data lines, data acquisition devices and related equipment could be supervised by a central diagnostic computer which is supplied with appropriate control signals.

4.7. Display of weather data and notices to airmen

Weather data supplied by weather information service usually are displayed to controllers on telescreens via scanning of printed or written notices.

Notices to airmen are normally printed by teletypes or lineprinters spontaneously on aspects of actual interest or on special request.

5. REFERENCES

/1/ H.G. Friedrich
Principal ATC Components
paper of AGARDograph on "Air Traffic Control" in hand

/2/ A. Benoit et.al.
Aircraft Trajectories - an approach to the calculation of air traffic trajectories for possible application in air traffic control
Eurocontrol Review, 2-II, 1970

/3/ H.A. Neumann et.al. of MBB
Systemstudie "Flugsicherung der 80er Jahre"
Band 3, 2. Auflage April 1974

/4/ E.E. Eddey
Ground Based Aircraft Collision Avoidance Using an Associative Processor
NAECON 70, pp. 302-306

/5/ G.E. Miller
ATC application of computer generated speech
ICAO-Bulletin, Oct. 1972, pp. 20-24

SOME TRENDS IN HARDWARE CONCEPTS FOR ATC COMPUTER

by

Wolfgang Nittner, Dipl. Ing.

Data processing systems

Messerschmitt-Bölkow-Blohm GmbH

8012 Ottobrunn, F.R. Germany

INDEX

SOME TRENDS IN HARDWARE CONCEPTS FOR ATC COMPUTER

by

Wolfgang Nittner, Dipl. Ing.

Data processing systems

Messerschmitt-Bölkow-Blohm GmbH

8012 Ottobrunn, F.R. Germany

1. INTRODUCTION

Growing air traffic necessitates the expansion of the control capacity of air traffic control. Manual operations must be replaced step by step by automated system modules. Therefore application of automatic data processing to ATC requires systems with the ability of organic growth and a certain inherency of future-orientation. Studying present-day hardware concepts three trends are recognized:

- acceleration of execution speed of a single processor by microparallel techniques
- increase of reliability and performance by multiprocessor-structures
- realization of special array-, associative- and I/O-processing subsystems.

Observing software-research it can be expected:

- facilitation of programming by utilization of higher level languages and by expansion of the hardware-software-integration facilities
- improvement of compilers, control programs and operating systems with respect to mapping strategies in virtual memory systems and to automatic recognition of parallel structures of procedures.

Besides future-orientation there are four main requirements for hardware-structures resulting from applications such as radar data tracking, conflict detection and resolution, long-term flight plan coordination and flow control, flight progress adjustment, recording and statistics application programs:

- reliability and safety
- storage capacity and processing power
- time behaviour
- hardware-software-integration and software related features.

From these aspects some technologies and important features of computer structures are outlined in the following sections.

2. RELIABILITY AND SAFETY

There is no question about reliability and safety of computer systems to be of great importance to ATC because of a certain dependence of human beings on these systems, which is especially true in the future, when it will be impossible to fall back to a manually operated system. Therefore a redundant system with graceful degradation capabilities must be established.

Generally reliability is achieved by the concepts of fail safe and fail soft. The concept of fail safe duplicates elements or modules, so that in case of a failure the system can continue operation at the former level. Fail soft is a further step in reliable system design and uses fall back technique to avoid a total system black out. A sufficient number of fall back levels must be established to meet all the important patterns of failure.

Computer hardware is an aggregate of different levels. With respect to reliability and safety circuit-, system- and structure-levels should be distinguished.

Many investigations have been made on the different computer levels to solve reliability problems. On circuit-level there are parity checks of memory-, processor- and I/O-modules and data- and address-busses, automatic error correction, residue checking of arithmetic and address-calculations and I/O-activities, time-out of processor- and I/O-executions, and checking of memory boundaries when references are made.

These checks can cause error conditions. If a failure has been detected which cannot be repaired automatically on circuit level, fail-safe-activities must be provided on system level. In this case the failed module must be replaced by a redundant one, so that no loss in system performance occurs. If a redundant module is no longer available, fail soft mechanisms must be provided which permit emergency operations by reduced system performance.

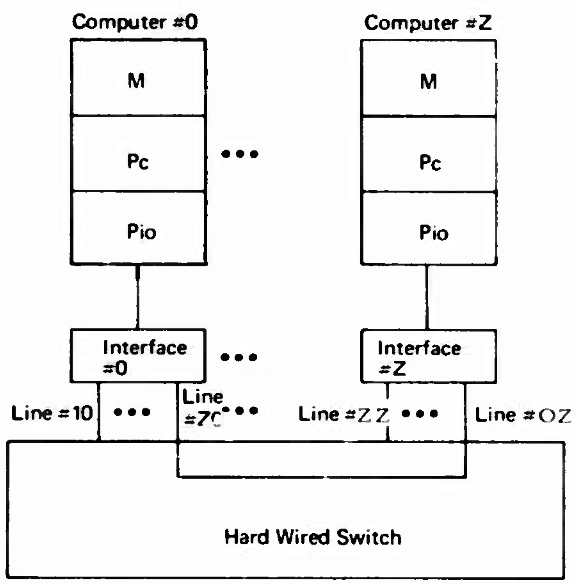


Fig. 1: Compound Single Processor Structure

The shaded areas mark the failed modules of a pattern of minimum system failure.

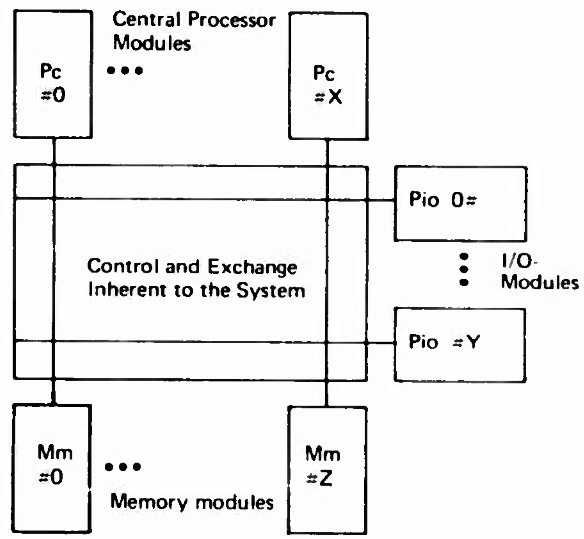


Fig. 2: Multiprocessor Structure

The shaded areas mark the failed modules of a pattern of minimum system failure.

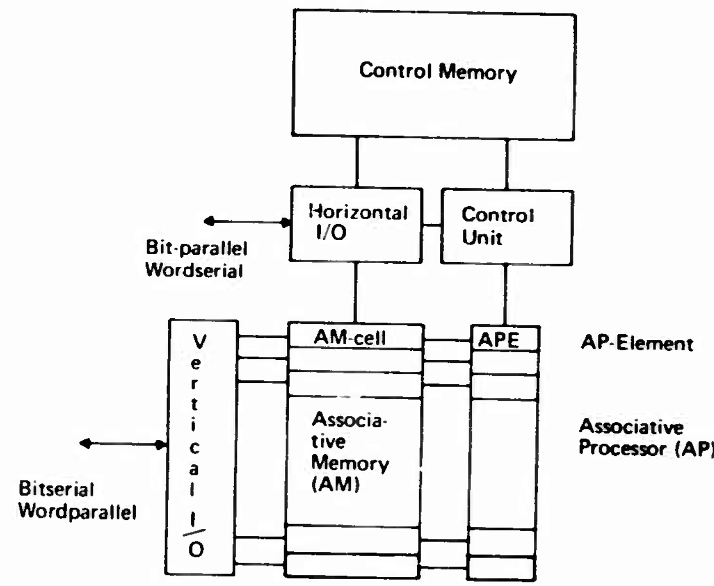


Fig. 3: Associative Processor Structure

The shaded areas mark the failed modules of a pattern of minimum system failure.

A further aspect is fault isolation and maintenance. Computer hardware operating in ATC-applications should have capabilities to isolate faults to a single card, and to integrate one or more maintenance diagnostic units to facilitate fault isolation and module test in parallel to the computer's operational work. These facilities minimize the time intervals of the degraded mode of operation and maximize system availability.

The question how redundancy and reconfiguration capabilities can be achieved yields to the organization of the computer structures to be installed. Associative processor, multiprocessor and compound single processor structures have distinguished features with respect to reliability and safety.

In case a multicomputer system (Fig. 1) consisting of more than two computers is regarded the installation of the inter-computer-connections is a difficult problem. Besides there are problems arising from the interaction of different operating systems, and from centrally supervising the operational computers. These disadvantages are avoided in a multiprocessor structure (Fig. 2) where the inter-module-connections and interactions, and central supervisory functions are already included within the basic system design. Conne[illegible]s between moduls are arranged within a matrix, which builds up highly redundant switches. Reconfiguration is supported by hardware, because each module is able to gain access to any other one.

Moreover, once a failure occurs in the multicomputer system one complete computer falls down; in the multiprocessor system the overall performance is only reduced by one module.

In case an associative processor structure is regarded (Fig. 3), the failure of an associative memory element does not result in a real reduction of performance because a single instruction stream controls multiple data streams at the same time. Of course the data not processed by one failing element are lost but all other data streams are not affected.

Failures of sequential units of the associative structure produce problems comparable to those of a single processor structure.

3. STORAGE CAPACITY AND PERFORMANCE

In an ATC-computer-complex on-line programs will not be the only functions sharing the resources of the pool in a given time interval. Care must be taken to allow the processing of off-line-jobs such as statistical evaluation of recorded data, which have to be inserted into the system at sporadic intervals. Furthermore, expansion of software packages will be an organically growing process, therefore the system lay-out must include an extended on-line time-sharing test system. These problems only can be solved by effective memory protection capabilities, mapping techniques which allow easy allocation and relocation of programs, and a hardware-supported time slicing algorithm, all of which make up a sophisticated multiprogramming and time-sharing system.

Thrashing is a phenomenon which can occur in such virtual memory systems when exchange of segments or pages extends an upper limit. This limit is defined by main memory capacity, page memory access time and transfer rate and internal performance. In a given system configuration thrashing can be avoided strategically by optimizing and estimating the memory requirements of a program or tactically by suspending those jobs which cause thrashing until more space is available. The first method can be realized by the software specialist supported by the compiler, the second method by the control program supported by hardware.

Because thrashing highly reduces the total system performance the balancing of the hardware configuration is an important factor. Another extreme which should be avoided in a virtual storage environment is the mere execution of CPU-bound jobs. In this case system balance is disturbed in the other direction reducing system performance, too. The criteria whether a job is CPU-bound or not in a virtual storage environment, depends on the speed of the processor and the paging devices. When CPU-bounding occurs, the paging mechanism is passive. Fig. 4 shows the qualitative dependence of system performance on virtual to real storage ratio in use. It can be seen, that in a well balanced environment system performance is increased.

The necessity to continuously expand automation of ATC-routines was already mentioned. In this respect it is interesting to consider the increase of throughput when more parallelism on system level is to be installed. If in a given system configuration the number of processors reaches a certain limit, the processing power remains constant or even falls back to lower values, because interferences with other resources grow rapidly. The most critical bottlenecks are the main memory modules, the number of which must have a balanced ratio to the number of the processor modules. Fig. 5 shows that even systems with low parallelism in their module structure do not yield the same factor of throughput increase expected by multiplying the processor modules.

4. TIME BEHAVIOUR

Real time data processing in ATC is connected with functions like radar data tracking, display of air traffic situation and conflict detection and solution. In these cases the computer must be able to process

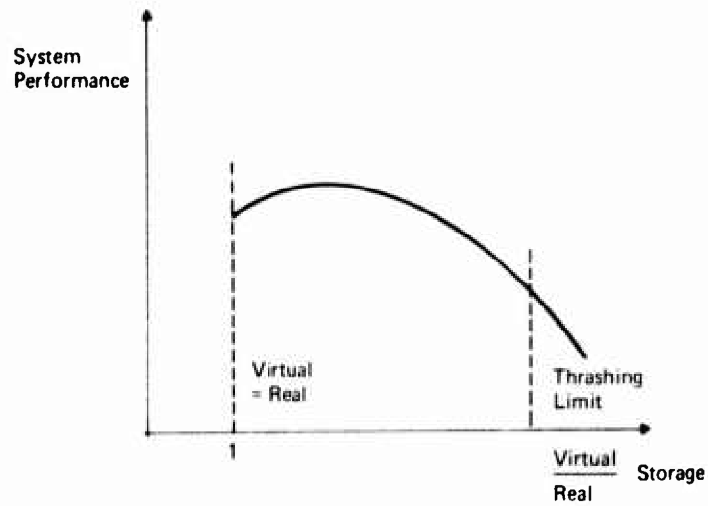


Fig. 4: System Performance in a virtual Storage Environment

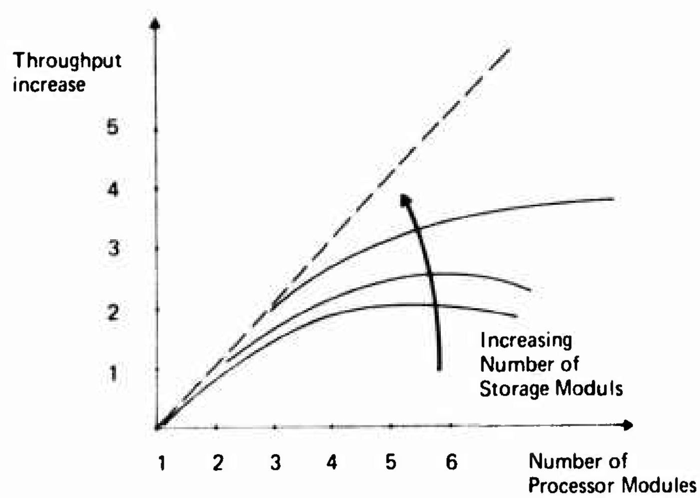


Fig. 5: Throughput Behaviour in Parallel Structures

an incoming data flow according to a certain priority scheme in such a way that the output of desired results or actions is not delayed beyond critical limits. Usually the radar data queues are scanned in sense-line triggered by clock pulses. Therefore the time behaviour of the real-time system is critical with respect to a high execution speed rather than to quick interrupt handling, because no critical process reaction response times are encountered.

Moreover, the tasks mentioned above have an inherent parallelism in their logic structure of data processing. Search instructions applied to a data base consume a great part of CPU-time available in conventional processor systems. Great advantages for the system's time behaviour may be obtained, if an associative processor (Fig. 3) is used, because the execution times of associative instructions are independent of the number of data (to which instructions are applied at the same time). For example, a search instruction is executed in about 1,0 microseconds by an associative processor, using a 10-bit data descriptor. This time is independent on the length of the list. In a conventional large scale computer an average time of about 30 μsec is required for the match, assuming a tree-structured list with a length of 1024 data words, a program loop of 6 instructions, and an internal performance of 2×10^6 instr./sec. In ATC-applications statically ordered lists regularly cannot be achieved. Therefore sequential search has to be applied, which would yield an average time of 1,5 millisecond on given conditions.

The high speed of 2×10^6 instr./sec of one sequential CPU is only available with pipeline machines and is accomplished by microparallel techniques, such as look ahead, look aside, interleaving and parallel arithmetic circuits.

The first method uses an IC-buffer which is filled with information before an access is required by the execution unit. Generally this is done in a sequential way (first-in-first-out organization), and is therefore suitable for instruction fetching.

The look aside buffer is used for temporary storage of data which are often used. Each time an access is required an associative address comparison is made on the complete buffer with practically no loss of time. If the requested operand is found in the associative buffer, the fetch is done at IC-speed, if not, a main memory access occurs and transfers are executed to the requesting CPU-register and to a free space in the associative buffer for a later likely usage. Because operands are generally accessed at random, no FIFO-organization is suitable.

Interleaving is a method which allows the CPU to access different memory banks at the same time via independent memory module controls. The last method uses different arithmetic units which can execute arithmetic, logic, boolean or shift operations etc. synchronously. All these techniques are illustrated in figures 6 and 7.

5. SOFTWARE-RELATED FEATURES

The last of the requirements stated concerns with that computer level, at which the hardware-implemented functions of a machine end and the software-implemented functions begin. During the design phase of a hardware-software-system, this level influences both the machine architecture and the language structure. The efficiency of compilation is improved with the increasing depth of the hardware-software-integration of a machine. An example of a minimal integrated machine is the conventional von-Neumann-computer where the execution of instructions is the only supported function. In highly integrated systems, problem-independent functions like storage management and protection, generation of parallel tasks and reentrant procedures and block structured languages are supported by hardware.

As an example of advanced hardware-software-integration, the generation of the object code of the expression A = (B+C)*D is outlined. This expression consists of a hierarchical operator structure which is to be transformed by the compiler, using an algorithm which yields the polish string ABC+D*=.

In the case of a general register or accumulator machine, this string must be processed by a further compiler routine. This is done by scanning the string, recognizing the temporary results BC+, establishing a scratch pad register and inserting operators and addresses.

In a highly integrated machine (with respect to arithmetic expressions) it should be possible to execute the polish string directly. This is easily accomplished in case a hardware stack organization is used. Then the final compiler routine needs only to scan the polish string and to insert the operators. During execution of the code, all instructions make references to the top of the hardware stack. Each time a reference is made the operands are pushed down or popped up automatically by hardware. Fig. 8 shows the arithmetic expression, the polish string, the code string generated by the compiler, and the expansion and contraction of the hardware stack, when the statement is executed.

Considerations based on a similar expression which is executed by a hypothetical accumulator, general register or stack machine yield a relationship in storage space of 100:94:90 and CPU-time of 100:92:86 respectively.

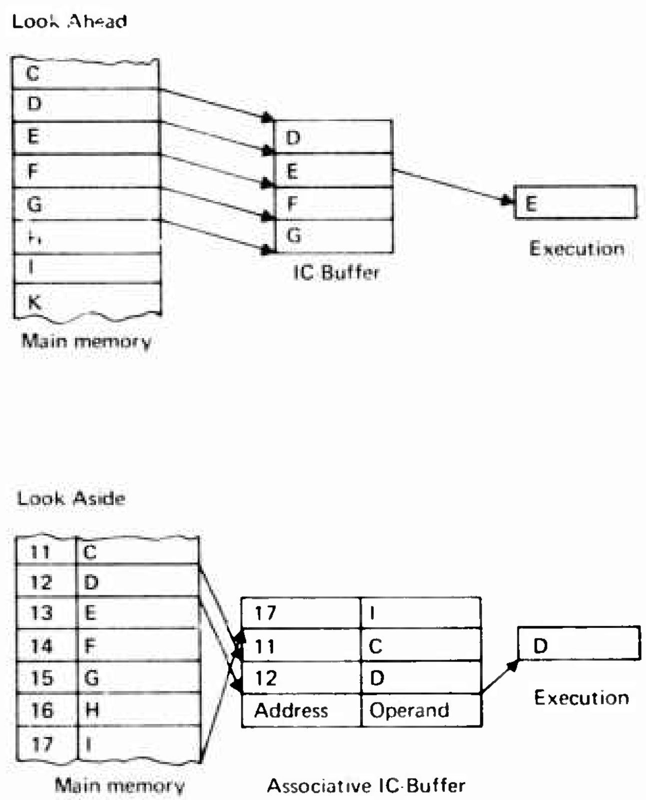


Fig. 6: Microparallel Techniques

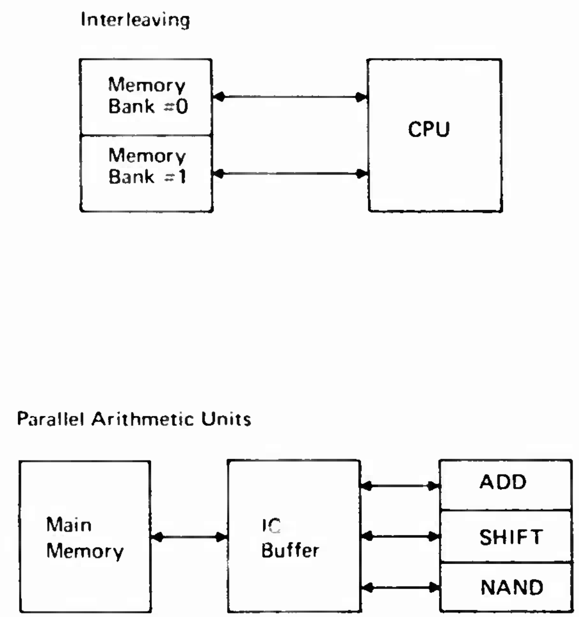


Fig. 7: Microprallel Techniques (Cont.)

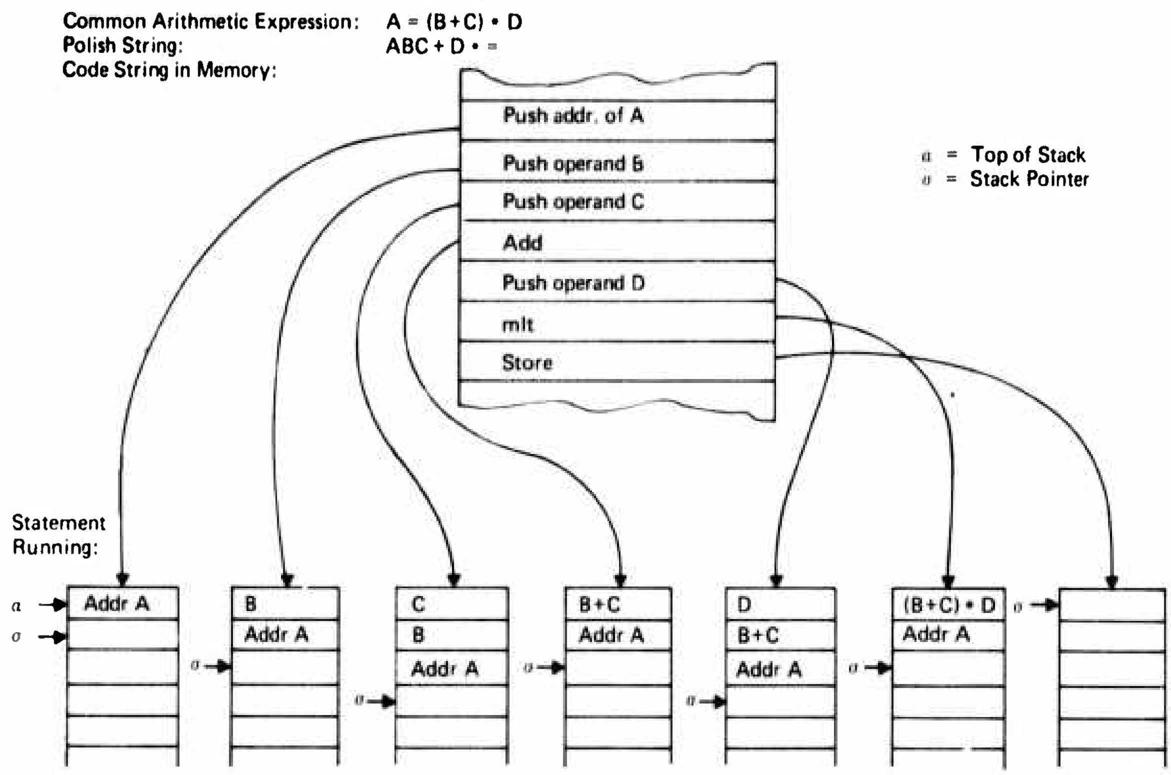


Fig. 8: Stack Organization and Polish Notation

Extending the hardware stack concept, a block structured language, the hierarchical structure of which is comparable to the arithmetic hierarchical structure mentioned above, can be transformed by the compiler in a similar way. For each block or procedure entry an activation record is prepared by the compiler and filled with fixed information at compile time. Storage for the variable information at run time is stack-organized and also part of the record. Each record is statically chained to the next neighbour of the hierarchical block structure until the main procedure is reached (tree-structured). This technique enables the maintenance of the valid environment at any position. Moreover, reentrancy is obtained automatically, because the information is stacked in the record. Therefore no comprehensive compiler standard routines are needed, to handle reentrant procedures. At run time, the instruction counter points to the instruction in execution and the environment vector points to the activated records of the environment in use. This vector is stack-organized and updated by hardware whenever a block or procedure entry or exit occurs. This sophisticated organization is far away from the von-Neumann-concept and could only be outlined here.

In ATC-applications a complex, safe and reliable software is required, also with the ability of organic growth. These requirements are supported by higher level languages, which enable a more structured programming. In a system with a suitable hardware-software-integration a compiled program yields an optimized object code because the hardware lay-out is suitable for corresponding statements in a similar manner as assembler instructions are for a more conventional hardware.

6. CONCLUSIONS

In this paper some important requirements for ATC-computer hardware were pointed out, and it was outlined what technologies and features are related with these requirements, which are considered to be reliability, storage capacity, time behaviour, and hardware-software-integration.

Reliability is achieved by the concepts of fail safe and fail soft, i.e. either by duplicating of modules or by using fall back and reconfiguration techniques, respectively.

In a virtual memory system the balancing of main and page storage yields an optimum of system performance. If during the organic growth of an ATC-system the throughput reaches a certain limit, more parallelism has to be installed. The most critical bottlenecks in this case will be the main memory modules, the number of which must have a balanced ratio to the number of the processor modules.

Real time data processing in ATC is concerned with functions which have an inherent parallelism with respect to the logic structure of data flow. Therefore great advantages regarding processing power are obtained in case an associative processor is incorporated, because the execution times of associative instructions are independent of the number of data to which these instructions are applied.

In ATC-applications a complex, expansible and reliable software is required. This is supported by the usage of higher level languages, which enable a more structured programming. A hardware structure with suitable software-related features enables the compiler to generate an optimized object code from a program written in a sophisticated higher level language just as it were written in assembler instructions for a lower hardware-software-integrated structure.

7. REFERENCES

/1/ H. Lorin
Parallelism in Hardware and Software
Prentice Hall 1972

/2/ E.I. Organick
Computer System Organization
Academic Press 1973

/3/ Bell & Newell
Computer Structures
McGraw-Hill 1971

/4/ P.J. Denning
The Working Set Model for Program Behaviour
Comm. ACM 11 (5), 323-333 (1968)

/5/ J.A. Rudolph
A Production Implementation of an Associative Array Processor
Proceedings of the FJCC, 229-241 (1972)

Pages 657 + 658 are Blank.

THE SATELLITE AS AN AID TO AIR TRAFFIC CONTROL

by

Keith D.McDonald
Chief, Aeronautical Satellite Division
Federal Aviation Administration
800 Independance Avenue, S.W.
Washington, D.C.
USA

TABLE OF CONTENTS

THE SATELLITE AS AN AID TO AIR TRAFFIC CONTROL

by

Keith D.McDonald

Chief, Aeronautical Satellite Division
Federal Aviation Administration
800 Independance Avenue, S.W.
Washinton, D.C.
USA

SUMMARY

With the evolution of satellite system technology, there has been increasing interest, attention and commitment directed toward exploiting the capabilities and potential of satellite-based systems for navigation and air traffic control (ATC). The utility of satellite systems has been extended from communications, surface navigation and geodetic applications into air and space, and to other functional areas such as position surveillance for air traffic control, precise time and time transfer, international maritime and aeronautical position location and reporting services, and collision avoidance. This chapter provides a summary of the basic principles upon which the satellite systems operate, an indication of their advantages and potential, a brief review of the historical developments, and a description of the system concepts and characteristics of selected satellite-based ATC systems which appear to be representative of the principal current candidates for filling this role in the future. Communication, surveillance, navigation and collision avoidance are discussed including certain operational aspects of employing satellite systems in these applications. Special attention has been given to those techniques which appear most promising for the over-ocean and continental ATC functions. However, substantive discussion is included for those satellite system developments which have been oriented toward position location and traffic control because of the significance of these efforts to the current state of satellite system technology. System concepts described include Transit, the Navy Navigation Satellite System; the expanded Transit, the Transit Improvement Program and the Two In-View concepts; the Defense Navigation Satellite Timing and Ranging (NAVSTAR) Global Positioning System; the NASA Position Location and Communications Equipment (PLACE) experiment; the maritime satellite program of the Department of Commerce's Maritime Administration; the Location Identification by Transmission (LIT) and the Satellite ATC and Navigation (SATAN) system concepts. Particular emphasis is given to the Aeronautical Satellite (AEROSAT) Program because of its international character and its application of satellite technology to immediate oceanic ATC needs. Satellite-based techniques for continental ATC discussed include the DOT's Advanced Air Traffic Management Systems (AATMS) concepts, and the FAA's recently developed ASTRO-DABS concept. Several of the more significant benefits and limitations of these techniques are covered and an indication is provided of future trends for the use of satellites in air traffic control.

PREFACE

During the past decade or more, there have been a significant number of investigations into the feasibility, practicality, and performance improvement which can be achieved by the use of satellites for position determination, time transfer, and other ATC-related functions. The application of this technology to the fields of air traffic control, surveillance, navigation, precise time transfer, system synchronization, and collision avoidance has been recognized and pursued through the development of a substantial variety of system concepts. Because of the specialized character and multiplicity of the system developments, a description is provided of the purpose, features, operating concept, advantages and limitations of many of these systems.

One of the better ways in which an understanding can be obtained of the utility and potential of the satellite as an aid to air traffic control is by, first, gaining some familiarity with the physical factors and measurement techniques associated with satellite-based systems, and second, reviewing the system concepts, implementation characteristics and performance of those systems which have either had a significant impact on the development of this field or which form the principal current basis for the evaluation, implementation or development of future satellite systems for ATC. This is the general structure of the Chapter. Air traffic control is treated in its broadest context, however, the principal areas of ATC assistance which the satellite provides are taken as communication and position determination. Data link, time synchronization and similar transmission capabilities are grouped within communication while position surveillance by a central ATC facility and vehicle navigation, or on-board position location, are grouped within position determination. The significant background of activity in satellite communications is not addressed but in general terms because of the advanced operational status of this field and the accessibility of this information in the literature. Navigation and position surveillance are treated in greater detail due to the increasing importance of improved position data acquisition to upgrading the performance of the ATC process and the capacity of the system. However, the ATC system clearly must incorporate a combination of communications, position data and a variety of other essential features such as automation, display, procedures and human competence to operate in an effective and efficient manner.

To provide a sound basis for the description and evaluation of the selected system concepts, a brief discussion of some of the principal factors influencing the performance of satellite systems is presented with particular attention to the various measurement error contributors, their frequency dependence, carrier frequency selection, selected signal structure alternatives, bandwidth requirements, satellite constellation configurations, ground station needs, and user equipment performance and cost.

The satellite-based system configurations selected for discussion are indicative of the system techniques under consideration by defense, civil, and international agencies. The extent of the activities may be seen by the following: The Defense Department is improving the capabilities of the operational Transit Navigation Satellite System and planning the development and experimental deployment of the precursor to a highly precise Defense Navigation Satellite Timing and Ranging System (NAVSTAR), the Department of Transportation is studying advanced air traffic management systems based largely on satellites, the Federal Aviation Administration is also investigating the ATC applications of satellite systems, the President's Science Advisory Committee ATC Panel completed investigations along these same lines, the Maritime Administration is conducting communication and ranging experiments on NASA's ATS-6 Satellite, and the DOT with Canada and the European Space Research Organization is working toward the deployment of an aeronautical satellite system which would include communications and position surveillance capabilities. Other satellite-based system concepts have been developed for consideration as candidates for the next generation ATC system, including systems designated LIT, SATAN, AATMS, and ASTRO-DABS, which will be discussed. The capabilities and potential benefits obtainable from these systems are briefly reviewed and evaluated in the context of projected future requirements and the existing and planned means for providing ATC-related services.

1. SATELLITE-BASED SYSTEM FEATURES

With the evolution of space technology, it has become apparent that satellite platforms provide a desirable means for accomplishing a variety of functions. Placing emitters in space results in significantly improved visibility of the users served and greatly expands the system coverage. Also, since the users are within line of sight, additional flexibility is achieved in the selection of the operating frequencies which may be used. For example, frequency selection need not be based upon the ability of the signal to propagate long distances around the earth, as in the case of LORAN and OMEGA, but may be chosen so as to minimize the measurement error or the propagation losses.

For the functions associated with air traffic control, satellite systems have a number of highly desirable features and some undesirable aspects. For communications or data link applications, in the large area coverage and the ability of the signal to penetrate into low elevation regions are beneficial. However, the distances which typically occur between a satellite and user are such, i.e. from several hundred to about 20,000 n.mi., that a significant range loss is involved. This normally is partially compensated for by a combination of antenna directivity on either the user vehicle, the satellite, or both; the use of high power transmitters and low noise receivers; and the incorporation of signal structures which improve the signal detectability and reliability. These techniques alleviate the range loss deficiency to an extent but system constraints, such as low cost, low gain user antennas, normally preclude incorporating sufficient complexity or sophistication to off-set this effect. A disadvantage is that satellite system receivers are usually more susceptible to interference and jamming than alternative local region systems.

Satellite-based systems provide an excellent means for structuring large coverage centrally managed ATC, surveillance, navigation and communications systems. This has traditionally been a significant feature of satellite communications systems, especially since the large payload operational feasibility of geo-stationary satellites, and the very large capacity (bandwidth) repeater systems have become available.

The space environment is relatively benign to orbiting spacecraft. If the satellite survives the accelerations and vibrations of its launch and placement into orbit, its chances of providing uniform, reliable service are excellent. There are, of course, significant risks associated with the launch of a satellite, since both satellites and launch vehicles are high cost items and booster reliabilities, although increasingly good, are not without failure.

In the position location role, satellites can be configured into orbital arrangements which provide excellent geometries for accomplishing the position determination function. This may be done with either synchronous satellite systems, in which case a regional coverage capability is provided; or with an orbiting constellation of satellites where worldwide coverage may be obtained with regional coverage provided on an intermittent basis.

2. HISTORICAL DEVELOPMENTS

The basic concept of the use of artificial satellites for communication and position location purposes has existed for over two decades. The initial demonstration of satellite capabilities occurred in 1958 when a communication satellite was launched. In 1960, the U.S. Navy's Transit Satellite System for ship navigation became available. This system, using polar orbiting satellites and user measurements of the doppler frequency shift indicated acceptable accuracies for many navigation applications. However, the long periods required to obtain a position fix and the intermittent coverage made this system unsuitable for civil aeronautical and other users.

The first satellite to achieve synchronous orbit, SYNCOM 2, was used in 1964 and 1965 to transmit low-rate digital data to aircraft and to receive transmissions. NASA's Applications Technology Satellites (ATS) 1 and 3, placed into geostationary orbits in 1966 and 1967, respectively, have been used in a continuing series of cooperative experiments between the FAA, NASA, and commercial airlines to demonstrate high-data-rate and two-way voice transmission between satellite and aircraft.

Successful ranging and position-fixing experiments were performed with these satellites on aircraft and ships during 1968 and 1969 using two-way transmissions and surface-based computation. More recently, the ATS-5 and ATS-6 satellites have been employed to obtain basic propagation and ranging data by NASA and the U.S. Air Force. Additionally, the U.S. Navy has employed its TIMATION I and II satellites to obtain similar ranging data and has performed precise time transfer experiments over intercontinental distances.

3. SATELLITE MEASUREMENT TECHNIQUES AND PERFORMANCE

It appears appropriate to briefly discuss the general elements of radio position determination systems, and identify the factors influencing system performance, at least those affecting accuracy and coverage. The user typically obtains information from several stations. This information normally takes the form of range or range difference measurements, at least for most LORAN, DECCA, DME, OMEGA and satellite system implementations. For hyperbolic systems, measurements are made of the range differences between pairs of stations; similarly, for ranging systems, direct measurements of propagation time which is translated to range is determined.

Ground-based systems typically operate in the ten kHz to several hundred MHz frequency range. The propagation characteristics of these systems are such that as the frequency increases, the ground level coverage of the system decreases markedly. However, the influences of the propagation medium are generally reduced with increasing frequencies which cause the propagational related errors to become smaller, improving the system precision.

The coverage obtainable from an earth orbiting satellite varies considerably with its orbital altitude. A satellite at about 600 NM altitude has an orbital period of about 1.5 hours and covers about one or two percent of the earth's surface at any instant in time. The coverage also depends upon the minimum elevation angle at which the satellite can be effectively viewed from the earth; typical values are in the order of five to ten degrees. At an altitude of 1000 NM, about seven to nine percent of the earth's surface is visible and at 5000 NM altitude (an orbital period of about 5.5 hours) the strong dependence between coverage and altitude decreases significantly. Between 30-35% of the earth's surface is visible at this altitude; this very gradually increases to between 40-45% at the geo-synchronous altitude of about 19,200 nm. It is apparent that large regions of coverage are feasible with small numbers of satellites. Additionally, the coverage is obtained on a "line-of-sight" basis which obviates the need for selecting frequencies which diffract around the earth's surface in order to obtain large area coverage. Frequency selection can be made with increased flexibility. Although the effects of the troposphere and ionosphere remain, they enter as error components in a somewhat different manner.

Error Contributors

To provide some perspective of the performance capabilities, limitations, and trade-offs associated with satellite-based position determination systems, we may briefly discuss the sources of measurement error influencing system performance. The principal contributors to system error are as follows:

- Thermal noise
- Man-made and natural interference
- Cosmic, or "galactic" noise
- Jamming
- System self-jamming
- Satellite orbital position error
- Geodetic location uncertainties
- Multipath effects
- Atmospheric absorption
- Tropospheric refraction
- Ionospheric refraction

Thermal noise refers primarily to the noise generated within the receiver; cosmic noise is noise of extra-terrestrial origin; geodetic uncertainties relate to the displacement of the satellite-based coordinate grid system to the geodetic system; absorption is normally of significance only in those frequency regions where strong atomic or molecular resonances occur for the constituents of the atmosphere; and the refraction effects are of interest in that they cause signal propagation delays. Figure 1 illustrates the one-way ranging errors resulting from the group delay characteristics of the ionosphere. The figure provides a nominal indication of the strong dependence of the ranging error with frequency and the variation in error for two conditions of antenna elevation (9° and 54°). The values of the ranging error for selected frequencies (0.4, 1.6 and 4.2 GHz) are indicated. At 400 MHz the errors are reasonably large and a dual frequency correction approach for determination of the group delay (similar to that employed by TRANSIT) is normally required to achieve high accuracy. In the L-Band region (1.6 GHz) the error is moderate and most of it can be estimated by applying operationally similar conditions to a "model" of the ionosphere which incorporates previously observed data. At 4 GHz and beyond, the influence of the ionosphere is negligible and neither dual frequency correction nor modeling is normally required. It should be noted that the ranging measurement errors caused by ionospheric effects for the microwave region are only a few feet.

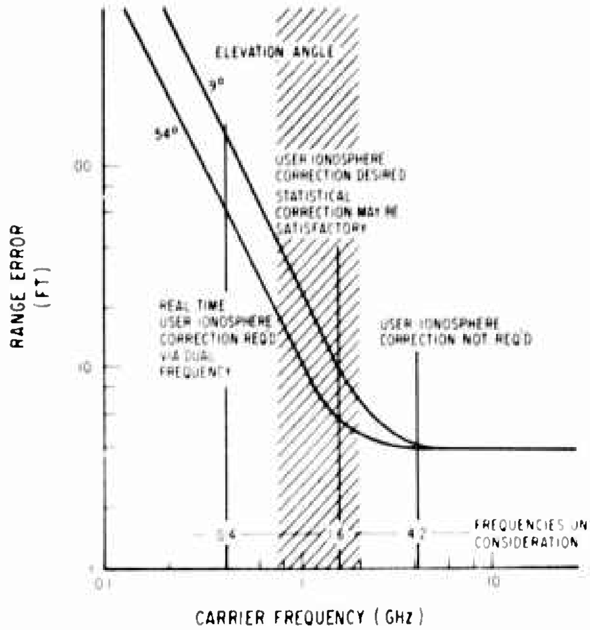


Figure 1 - Ionospheric Propagation Ranging Error with Frequency

<u>Satellite-Based Techniques for Determining Position</u>

A variety of satellite-based system concepts have been considered as candidates for performing the position location function. The system concepts are based upon the measurement parameters available to either the user as a central facility for performing the position determination. The space-based measurements employed in the determination of the position of an observer located on or near the surface of the earth may be classed in the following general categories:

(a) Doppler and doppler rate
(b) Angle and angle rate
(c) Range and range rate
(d) Measurement sums or differences
(e) Measurement combinations

Figure 2 illustrates a number of representative position determination techniques employing satellites. These system concepts indicate the number of satellites required, the measurements performed, the surfaces of position which are generated by the measurements and comments concerning the system performance and implementation requirements. The salient characteristics of the systems, based upon these measurement techniques, may be summarized as follows:

(a) <u>Doppler Measurement</u> - from a system of low altitude satellites is an approach of proven capability. Referring to A, this technique, successfully used in the Navy Navigation Satellite System, requires accurate knowledge of the satellite ephemeris and the user's velocity. This information, combined with measurements of the satellite-observer relative motion, as determined by doppler measurements of the satellite transit through its region of closest approach, provides an accurate determination of observer location. Limitations of this approach relate to the use of a relatively complex data processing procedure, and more importantly, the strong sensitivity between position error and the uncertainty in user velocity. Additional system characteristics of concern include the moderate time intervals (5-20 minutes) normally required for a doppler position determination and the large number of low altitude satellites required to assure a short waiting period between satellite transits.

(b) <u>Angle Measurement</u> - by a space-borne system, shown in B, normally includes a pair of crossed-baseline space-borne interferometers extending from a single satellite. Each interferometer baseline is of a dimension equivalent to a large number of wavelengths (e.g., several hundred). The length of the baseline establishes the number and angular dimensions of the grating-lobe structure subtended by the baseline aperture. It is the resolution provided by the grating-lobe structure of an interferometer which makes possible accurate angular measurement of the user's position in space. The angle measurement technique provides a desirable feature in that full coverage for a specified satellite visibility region may be obtained from a single satellite.

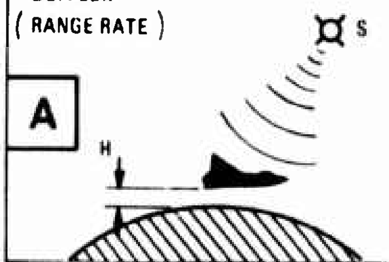
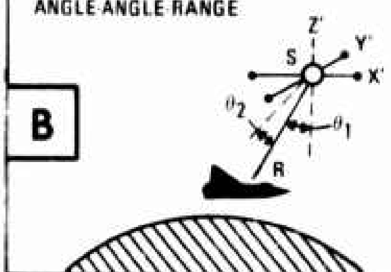
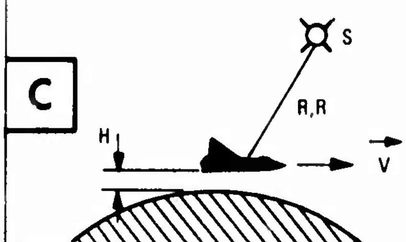
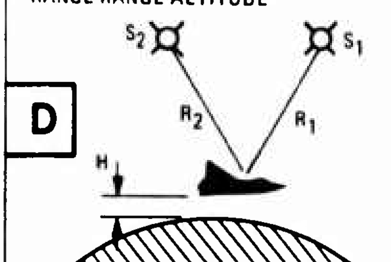
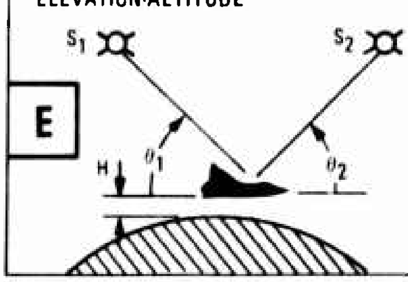
SYSTEM TITLE AND DIAGRAM	SATELLITES REQ'D / DATA PROVIDED	MEASUREMENTS PERFORMED (BY USER)	SURFACES OF POSITION	NOTES
DOPPLER (RANGE RATE) 	1 / X,Y	• DOPPLER FREQUENCY • TIME • ALTITUDE	2 HYPERBOLOIDS (1 PER PAIR OF MEASUREMENTS) 1 SPHERE	• PASSIVE SYSTEM • OPERATIONAL SYSTEM TRANSIT • EXCELLENT ACCURACY • PRINCIPALLY FOR LOW VELOCITY VEHICLES • NON-INSTANTANEOUS
ANGLE-ANGLE-RANGE 	1 / X,Y,Z	• 2 ANGLES-SATELLITE TO OBSERVER • RANGE-SATELLITE TO OBSERVER (R)	2 CONES 1 SPHERE	• ACTIVE SYSTEM • MEAS. OF H CAN BE SUBSTITUTED FOR R MEAS.
RANGE-RANGE RATE-ALTITUDE 	1 / X,Y (Z)	• RANGE TO S: R • RANGE RATE: R • ALTITUDE: H	2 SPHERES 1 CONE (1 HYPERBOLOID PER PAIR OF R MEASUREMENTS)	• ACTIVE OR PASSIVE • PASSIVE USER REQUIRES STABLE CLOCK • LOW ALTITTUDE SYSTEM • VELOCITY VECTOR OF OBSERVER MUST BE KNOWN • H CAN BE DET'D WITH MULTIPLE MEAS.
RANGE-RANGE-ALTITUDE 	2 / X,Y	• RANGE-SAT 1 TO USER • RANGE-SAT 2 TO USER • ALTITUDE	3 SPHERES	• ACTIVE OR PASSIVE • PASSIVE USER REQUIRES STABLE CLOCK TO DETERMINE R
ELEVATION-ALTITUDE 	2 / X,Y	• ELEVATIONS ANGLES: OBSERVER TO S_1 (θ_1) OBSERVER TO S_2 (θ_2) • ALTITUDE	1 SPHERE (Z) 2 CONES	• PASSIVE SYSTEM • RANGE MEAS. CAN REPLACE H MEAS.

Figure 2 - Representative Satellite-Based Position Determination Techniques

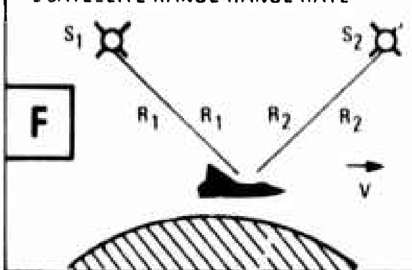
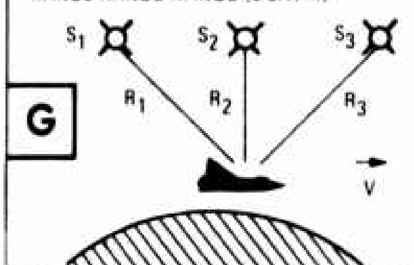
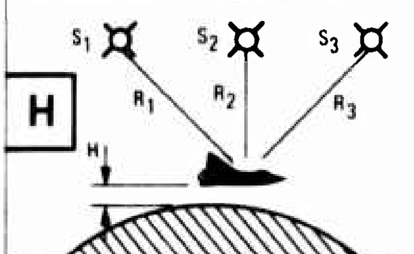
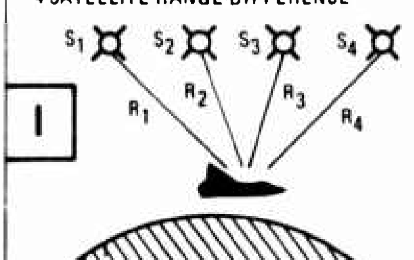
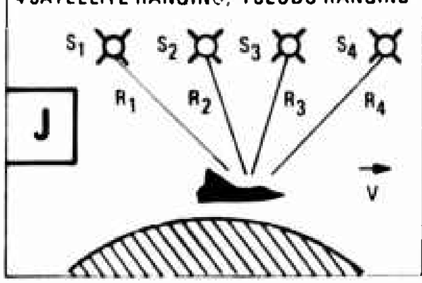
SYSTEM TITLE AND DIAGRAM	SATELLITES REQ'D / DATA PROVIDED	MEASUREMENTS PERFORMED (BY USER)	SURFACES OF POSITION	NOTES
2 SATELLITE RANGE RANGE RATE 	2 / SINGLE MEAS. X,Y,Z MULTIPLE MEAS'S: X,Y,Z $\vec{V}$ t	• RANGES TO SATELLITES R_1, R_2 • RANGE RATES TO SATELLITES $\dot{R}_1, \dot{R}_2$	2 SPHERES 2 CONES (2 HYPERBOLOIDS PER PAIR OF $\dot{R}$ MEASUREMENTS)	• ACTIVE OR PASSIVE • PASSIVE REQUIRES STABLE CLOCK AND KNOWLEDGE OF USER VELOCITY • LOW ALTITUDE SYSTEM • PASSIVE MULT. MEAS'S • NO CLOCK REQUIRED • LOW ALTITUDE SYSTEM • KNOWLEDGE OF USER MOTION REQUIRED
RANGE RANGE RANGE (3 SAT R) 	3 / X,Y,Z $(\vec{V})$	• RANGE, OBS. TO S_1 (R_1) • RANGE, OBS. TO S_2 (R_2) • RANGE, OBS. TO S_3 (R_3) (ALSO, RANGE RATES $\dot{R}_1, \dot{R}_2, \dot{R}_3$)	3 SPHERES	• ACTIVE OR PASSIVE SYSTEM • PASSIVE USER REQUIRES STABLE CLOCK • MEAS. OF $\dot{R}$'S ALLOWS DETERM. OF USER VELOCITY
RANGE DIFFERENCE AND ALTITUDE 	3 / X, Y	• RANGE DIFFERENCES: $\Delta R12 = R_1 - R_2$ $\Delta R23 = R_2 - R_3$ • ALTITUDE	2 HYPERBOLOIDS 1 SPHERE (Z)	• PASSIVE SYSTEM • HYPERBOLOID TECHNIQUE • PASSIVE USER DOES NOT REQUIRE STABLE CLOCK
4 SATELLITE RANGE DIFFERENCE 	4 / X,Y,Z	• RANGE DIFFERENCES: $\Delta R12 = R_1 - R_2$ $\Delta R23 = R_2 - R_3$ $\Delta R34 = R_3 - R_4$	3 HYPERBOLOIDS	• PASSIVE SYSTEM • HYPERBOLIC TECHNIQUE • USER DOES NOT REQUIRE STABLE CLOCKS
4 SATELLITE RANGING, PSEUDO RANGING 	4 / X,Y,Z, t $(\vec{V})$	• RANGES: R_1, R_2, R_3, R_4 • (ALSO RANGE RATES: $\dot{R}_1, \dot{R}_2, \dot{R}_3, \dot{R}_4$)	4 SPHERES	• PASSIVE SYSTEM • 4TH SATELLITE PROVIDES PASSIVE USER WITH ACCURATE TIME RERERENCE

Figure 2 (cont'd) - Representative Satellite-Based Position Determination Techniques

Accurate knowledge of the satellite baseline orientation in space is necessary for this technique. A variation of this approach is shown at C.

(c) Range Measurement - from multiple satellites as shown in D and E provides a means for instantaneous position location by determination of the intersection of multiple spheres of position. Minimally, determination of the ranges of two satellites, and the aircraft altitude are required for a unique instantaneous position determination. For a 3D determination, completely independent of observer-derived inputs, three satellites are required. Normally, the satellite systems associated with the ranging technique are not appreciably different from conventional communications satellites, i.e., they may consist of simple transponders. A sufficient number of satellites in a favorable geometric arrangement must normally be provided by the orbital constellation for the desired system accuracy to be obtained.

(d) Range and Range Rate Systems - for determining the position of a satellite with respect to an observation point on earth have been employed for some time. This technique is illustrated by F and G. The NASA Goddard range and range rate (RARR) satellite tracking system has provided results of excellent precision. The addition of the range rate, or doppler, information to the ranging data provides a supplementary approach for obtaining a position fix. Each pair of doppler readings generates a hyperbolical surface of position. Two sets of doppler determined surfaces of position provide a line of position which may be combined with the ranging data to establish a position fix. Incorporation into the system of range rate information in addition to multiple satellite ranging increases the complexity of the computer processing involved but provides a means for improving the accuracy of position determination.

(e) Range Difference Measurement - between pairs of satellites as shown in H, provides a method for obtaining hyperbolic position determinations from space-borne stations. Ground-based hyperbolic techniques such as LORAN, DECCA and OMEGA, have been used successfully for a number of years, particularly for long range over-ocean navigation. The application of this technique requires, minimally, one additional satellite to that of a ranging configuration since differences measurements between pairs of satellites are involved.

(f) Pseudo-Ranging and Time Measurement - can be accomplished as shown in I and J, which illustrates two 4-satellite ranging concepts. This configuration allows the user to operate passively (no transmission) and without a lights stable clock. It is possible to obtain a 7-dimensional fix, in that 3D position, 3D velocity and time are determined.

(g) Measurement Combinations - of many types are possible. Several examples of these have been shown in the system technique figures.

4. FIELDS OF APPLICATION

The applications of satellite system technology to air traffic control and related fields have been investigated and in some cases implemented for a number of years. Communications repeaters have been tested with aeronautical and ground based systems for both civil and military uses. A satellite system has been used operationally for over a decade for navigation, and a variety of time transfer and synchronization operations have been routinely accomplished from space. Additionally, a number of applications have been investigated, some of which do not directly relate to ATC, which indicate the versatility of appropriately configured satellite-based systems. These applications include area navigation, vehicle or unit location, collision avoidance and proximity warning, search and rescue, survey and geodesy, instrument landing system guidance, vehicle or satellite tracking, calibration for photo mapping, velocity determination and a variety of defense uses.

The application of satellite-based communication and position determination systems to air traffic control has undergone extensive investigation by several agencies of the government. Representative examples of these activities are: The Department of Transportation ATC Advisory Committee Report investigated and made recommendations concerning the utility of a space-based position data acquisition system which would operate within the ATC framework. The Ad Hoc ATC Panel of the President's Science Advisory Committee (PSAC) analyzed the feasibility of a satellite-based system for Continental United States (CONUS) air traffic control as an alternative to the planned upgrading of the current (third generation) Air Traffic Control Radar Beacon System (ATCRBS). The Department of Transportation (DOT) has funded substantive investigations of the feasibility, benefits and economic viability of satellite-based air traffic management systems through their Transportation Systems Center in Cambridge, Mass. Additionally, the Federal Aviation Administration (FAA) has investigated satellite-based systems as possible candidates to supplement or replace selected elements of the ATC system.

The FAA and the DOT are very much involved in the Aeronautical Satellite Program - a cooperative venture between the United States, Canada and Europe which may involve other partners. AEROSAT is configured to provide highly reliable oceanic ATC communications, data link and position surveillance services to over-ocean traffic which could result in providing more direct routes and reducing the cost of the Flight Information Region (FIR) structure.

There also has been considerably increased concern in the collision avoidance and proximity warning area, which in many system implementations involves the use of a space-based technique for time synchronization or position determination, or both. The use of space-borne systems for all categories of instrument landing approaches has been considered and is under investigation by the Defense Department and others.

The applications to vehicle location, geodesy and survey are readily apparent. It should be noted that one of the most accurate and frequently employed methods for obtaining geodetic information on a world-wide basis for some years now has been through the use of observations performed on orbiting satellites, e.g. the GEOS series of satellites. The capability provided by certain systems for the transferral of precise time from one location to another has been both scientifically and operationally useful. There are, as the list indicates, a number of Defense uses for space-based position location data, including the delivery of material, weapons and men, reconnaissance and various other missions contained within the items discussed.

5. SYSTEM CONCEPTS AND CHARACTERISTICS

Table I lists a number of satellite-based systems and system concepts which have been selected for description, comparison and evaluation. The sponsoring agencies are indicated in the table as are the principal functions of the systems, i.e., navigation (N), position surveillance (S), or communication (C) which includes data link capabilities. The priority associated with each of these functions in the context of each system's current configuration is indicated by the ordering in the right column of Table I. The other columns attempt to establish the development or operational status of the systems and are self-explanatory. The interpretation of system development status in Table I should be considered a judgemental interpretation based primarily on the information available. However, the table is believed to represent a reasonably valid indication of the current status for each system concept, even though the information in the table will require updating with time.

About half of the systems considered in the table are system concepts which have no clear experimental or operational plans. Similarly, it should be noted that only one satellite-based system is operationally deployed: the Navy Transit, or NNSS system. There is a notable lack of other systems which are planned, or approved for operational deployment. About half of the system concepts have proceeded to the point where plans have been configured for experimentally testing the concepts. In most of these cases, a limited experimental program has been funded which typically attempts to resolve certain technical issues on a limited deployment basis. Two of the system concepts which have progressed to the system test and demonstration stage are essentially long-standing single satellite system concepts, which, therefore, involve relatively small risk and modest funding levels for completion through this phase. We now may discuss the principal characteristics, including the similarities and differences, of the selected system concepts.

TABLE I
SPACE-BASED POSITION DETERMINATION SYSTEMS

System name	Sponsoring agency	Concept(s) only	Experimental system configured	Limited experimental program funded	Test and demonstration system deployed or planned (funded)	Operational system planned and approved	Current operational system status	N-Navigation S-Surveillance C-Communication
TRANSIT/ NNSS	DOD/USN						●	N
EXPANDED TRANSIT	DOD/USN		●	●				N
TIP	DOD/USN		●	●	●			N
TIV	DOD/USN	●						N
NAVSTAR	DOD/4 Svc		●	●	●			N, (S)
SYSTEM 621B	DOD/USAF		●	●				N, (S)
TIMATION	DOD/USN		●	●				N, (S)
PLACE	NASA		●	●	●			S, C, N
MARSAT	DOC/ MARAD		●	●				S, N
AEROSAT	DOT/FAA		●	●	●			C, S, N
LIT	CONTRACTOR	●						S, N
SATAN	CONTRACTOR	●						S, N
AATMS-I	DOT	●						S, C, N
AATMS-II	DOT	●						S, C, N
ASTRO-DABS	DOT/FAA	●						S, C, N

(a) Transit/NNSS

The Navy Navigation Satellite System (TRANSIT or NNSS) has provided position fixing information on a world-wide basis to surface and sub-surface users since 1964. TRANSIT is an all-weather, passive user navigation system that provides accurate navigation capabilities, normally latitude and longitude, to better than 0.1 NM. Timing signals transmitted from the satellites of the system are synchronized with Universal Time Coordinated (UTC) to within 200 microseconds. The TRANSIT system is composed of three subsystems: a constellation of satellites, a network of tracking stations that continuously monitor and update the satellite's navigation data, and the user's receivers and computers.

Operational use of the TRANSIT system began in July 1964 and the system has been in continuous operation since that time. The shipboard receiver was released for commercial manufacture and civil use in June 1967. There are now more than 150 commercial and 250 military receiver systems in operation using the navigation data from TRANSIT. The satellite constellation consists of a minimum of four (currently five) satellites in circular polar orbits at an altitude of approximately 600 NMs. The satellites transmit frequencies of approximately 150 and 400 MHz and the user determines his position by measuring and examining the doppler shift of these signals. There are four ground stations which track and update each satellite's ephemeris and time synchronization data. The control center at Point Mugu, California provides the system tracking and data injection facilities, the central computer center and an operations and communications center.

(b) Transit System Upgrading

For several years the Navy has been concerned with upgrading the capabilities of the TRANSIT system. A listing of techniques which have been considered for upgrading TRANSIT is shown in Table II. Improvements are planned in several areas but the long term details of the Navy program in this field are not completely established at this time. The Expanded TRANSIT system concept relates primarily to the expansion of the current TRANSIT satellite constellation for the purpose of reducing the average waiting period between satellite passes, or transits. The current system deployment has an average waiting period of about 100 minutes, but this period is latitude dependent being longer near the equator and shorter toward the poles. The Expanded TRANSIT concept also addresses the reduction in the time interval required to obtain a position fix, through an improved computer program in the receiver and certain other refinements, and the use of the system in a dynamic (e.g., aircraft) environment. To this end means for incorporating a satellite-based ranging signal have been developed which provide the features indicated. The ranging signal employs a pseudo-random noise (PRN) phase code modulation, such as BINOR or similar spread spectrum signal format for the transmission of precise time and for direct precision range determination. The PRN code allows satellites to be tracked individually, even though their doppler frequencies may cross, and provides maximum multipath suppression - a condition especially important to aircraft users. The PRN modulation is designed so as to be transparent to current TRANSIT navigation user equipment; if the current or an expanded satellite constellation continues to transmit the old signal any existing equipment can continue to operate with no change in performance.

TABLE II

TRANSIT UPGRADING TECHNIQUES

- Satellite Constellation Expansion
  - Reduces period between fixes
- Ephemeris Accuracy Enhancement-DISCOS
  - Improves fix accuracy
  - Allows longer period between updates
- Inclusion of Ranging Signal - PRN/BINOR
  - Allows ranging
  - Shortens fix interval
  - Improves accuracy
  - Improves operation in dynamic environment
  - Enhances AJ, acquisition performance
- Improved Fix Program in Receiver
  - Provides greater flexibility
  - Allows use of simpler receiver
  - Use of shorter doppler counts
  - Decrease of required transmission duration

Other improvements in TRANSIT have been considered in the TRANSIT Improvement Program (TIP) which is basically concerned with increasing the survivability of the system. These efforts have included hardening the satellites to radiation effects, enhancement of the accuracy of the satellite constellation ephemerides, and shortening the duration of the satellite observation interval necessary to obtain a position fix.

A technique for improving the accuracy and reliability of satellite orbital predictions is under development which has the important operational advantage of greatly extending the period of acceptable performance of the system in the event of loss of one or more ground stations. This system, called DISCOS, for disturbance compensation system, uses the detection of the relative motion of a freely suspended ball located within a container equipped with sensors which indicate the ball's position. The container is an integrable part of the satellite and therefore influenced by the drag forces acting on the spacecraft. The ball is in "free fall" and by sensing its motion it is possible to compensate for the drag (and other) disturbances by applying small thrusts to the satellite. The overall affect is to correct the orbit to one essentially free from drag and thereby significantly enhancing the stability and predictability of the satellite orbit. The first experiments of these new techniques are now in progress using the TRIAD-I (Transit Improved and DISCOS) experimental satellite launched in September 1972. Results from these and further experiments will be used to determine the configuration/cost schedule options for providing the most effective, useful TRANSIT system for the future.

It is interesting to analyze and review the need for the continuation of the TRANSIT system in the context of the various other operational and developmental system alternatives which appear promising or planned. Indications are that the operational TRANSIT system capabilities must be improved if the viability and utility of the system to support Navy missions is to be maintained to the mid-1980 time frame. The principal operative factor influencing the situation is the realistic time-table for the operational deployment of a follow-on precision world-wide navigation system, such as the Defense Navigation Satellite System (DNSS), the most promising current candidate to replace TRANSIT. Current planning for the DNSS indicates an experimental demonstration deployment in the latter half of the 1970's and possible operational deployment in the early 1980's. Also, a reasonably long (several to ten years) phase-over period is normally required for the introduction of any large new system such as the DNSS, and this is not usually initiated until the new system is fully deployed. It is likely that the early or mid-1980's is not an unreasonable estimate of the minimum required "lifetime" for the TRANSIT system.

(c) Two-In-View TRANSIT (TIV)

The TIV TRANSIT system concept was conceived as a moderate cost, world-wide, continuous, all-weather, high accuracy satellite-based system which would capitalize on the base of TRANSIT technology and serve both suitably equipped dynamic users and those users normally served by the current TRANSIT system. The TIV system is capable of providing significantly improved navigation information to dynamic users (typically aircraft) by utilizing in a "hybrid" manner the data available from the platform's dead-reckoning system, such as an inertial platform or air data system.

The system concept consists of a 30 satellite orbital constellation with five equally spaced polar rings and one equatorial ring operating at an orbital altitude of 1465 NMs. This arrangement provides at least two satellites always in view throughout the world and allows for all tracking support to be accomplished from installations within the United States.

Updated versions of the current TRANSIT (Oscar) satellites could be employed which transmit two continuous navigation signals at approximately 150 MHz and 400 MHz. Both signals are modulated with a pseudo-random noise (PRN) code, which is required to overcome cross-satellite interference problems. This modulation also yields moderate jamming immunity by providing processing gains of 30 db or greater. Both range/doppler navigation signals are tracked and used to correct ionospheric refraction effects, and the result is equivalent to the use of a single frequency transmission at 2500 MHz.

The simultaneous tracking of two continuously available satellites and the use of an improved satellite ephemeris, coupled with the exploitation of both doppler and range measurements, yields greatly improved navigation performance on a continuous and world-wide basis. Typically, the initial fix would be obtained after five minutes of satellite data, with subsequent updates provided every four seconds as long as satellites are continually acquired and tracked. In addition to the position and velocity information, the fix computation recovers time and frequency calibrations for the navigator's oscillator or clock. Since the initial studies, the TIV concept has not been supported by the Navy because of technical and operational considerations and is currently not considered as a candidate for system implementation.

(d) Defense Navigation Satellite Time and Ranging (NAVSTAR) System

The NAVSTAR global positioning system, which is conceptually planned to satisfy selected DOD navigation needs, and provide navigation data to an unlimited number of passive users. The NAVSTAR techniques typically involve simultaneous measurements by a user of range information to a number of satellites.

System Description

A NAVSTAR implementation consists of three basic parts, or segments: space, ground, and user equipment, as shown in Figure 3. The space segment consists of satellites which serve as platforms for the navigation signal emmiters. The orbital paths of the satellites are designed to provide multiple satellite visibility to the users and favorable satellite system geometries for position determination. Full-time, worldwide, instantaneous availability of the system's capabilities can be provided by a variety of satellite constellation arrangements; these typically involve the deployment of a 20-27 satellite space segment. The ground segment consists of a

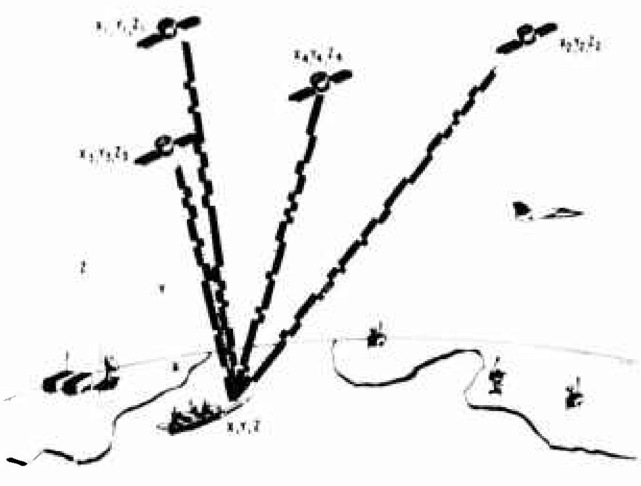


Figure 3 - Defense NAVSTAR System Concept

central control, monitoring, and processing facility, or "master" station, and a network of remote stations are used to determine orbit information and may serve as the space/ground communications interface for control of the satellites. In some concepts, special local calibration stations may be desired to refine and decrease certain error components within a local region where maximum position accuracy is of primary importance (e.g., within a tactical area of operation). The user equipment segment consists of receiving, processing and display equipment which accomplish the range measurements and from this data computes and displays position, velocity and time data as appropriate. User equipment can be configured to satisfy the varying needs of several general categories of users, ranging from high performance aircraft to foot soldiers.

Capabilities

NAVSTAR is characterized by long baselines in space and line-of-sight operation between users and satellites which inherently offers performance advantages in terms of accuracy, coverage, data availability and timeliness. Some of the more important capabilities of this satellite-based system as it applies to Defense (and other) needs are as follows:

- Position fixing to tens of feet in three dimensions
- Accurate determination of velocity vector and time
- Continuous availability of navigation data in real time
- Basis for worldwide common grid coordinate system
- Worldwide or regional coverage feasible
- Provides general navigation capability - minimizes proliferation of special purpose, or regional, systems

The passive user of the system normally requires at least four satellites in view as shown in Figure 5 to obtain both continuous three dimensional (3D) position and system timing information. This technique obviates the requirement for the passive user to maintain a highly stable "clock" for ranging (to three satellites) but does so at the expense of an additional satellite (i.e., four) always in view. The timing data provides the necessary frequent user updating capabilities which allows the ranging measurements to be made with a relatively low quality user clock. This technique is frequently termed "pseudo-ranging" because direct ranging measurements providing a 3D determination of position to the passive user are not obtained independently but are achieved only in combination with the timing data determined through the observation of a four satellite constellation.

General Navigation Characteristics

Navigation system limitations or deficiencies frequently limit the effectiveness or utility of military and other operations. There have been many investigations directed toward establishing the navigation system characteristics which are important to the accomplishment of Defense-related mission responsibilities. Knowledge of these general characteristics combined with detailed determinations of the features and limitations of specific system candidates provides an important base of data for the guidance and direction of the development of new capabilities and the upgrading of current resources. A qualitative indication of these navigation system characteristics, identified as essential or desirable for Defense navigation systems, are consistent with the capabilities of navigation satellite systems, as indicated earlier.

TABLE III. DEFENSE NAVIGATION SYSTEM CHARACTERISTICS

ESSENTIAL	DESIRABLE
• Worldwide coverage	• All ground stations on U.S. territory
• High Accuracy	• Minimum propagation limitations
• Common grid capability	• Acceptable operation underwater
• Continuous availability in real time	• Provide worldwide time reference
• Passive user/non-saturable system	• Compatibility and integrability with other military/civil systems
• Acceptable survivability, security, anti-jam	• Evolutionary growth from R&D to operational capability
• Operation with dynamic users	
• Satisfactory portability/size/weight	
• Minimal frequency allocation problems	
• Freedom from ambiguities	

There has been recognition by a significant part of the navigation community that satellite-based radio navigation systems are technologically feasible and that they can satisfy nearly all identified military navigation requirements. Feasibility of these systems has been established principally through the extension of the results of individual experiments, analytical investigations, and the use of applicable data and developments in the navigation satellite and related fields, especially satellite-communications.

Prior Investigations

We may summarize some of the recent efforts by the military departments which have assisted in providing the building blocks on which to base a NAVSTAR concept. These activities include a substantial background of studies and experimental work into system techniques, applications, technology and cost. Representative of recent experimental efforts are the launch and testing by the Navy of two experimental satellites, TIMATION I (launched in May 1967) and TIMATION II (in September 1969), which were placed in 500 NM 70° inclination orbits. The satellites used 150 and 400 MHz signals with side-tone modulation to determine propagation delays and provide the ranging measurements. These satellites demonstrated the feasibility of navigation by passive ranging from satellites of known ephemerides, and were used as test beds for the improvement of satellite-to-ground transmission links, spaceborne stable oscillators, satellite attitude stabilization devices, ranging signal measurement accuracy, and preliminary determination of the total position fixing error budget.

Medium altitude and high altitude (typically earth synchronous) systems concepts have been investigated, principally by the U.S. Navy and Air Force under the program designations TIMATION and System 621 B, respectively. These system concepts now will be briefly described.

(e) TIMATION System Concept

The TIMATION navigation satellite system concept was developed by the U.S. Navy in their efforts to determine the technical feasibility of a satellite-based system to provide worldwide, continuous, accurate 3D position and velocity to passive users. Position determination accuracies of tens of feet are estimated for the system, based upon analytical and experimental investigations. The TIMATION concept consists of a constellation of 27 satellites and a network of four ground stations. The ground stations may all be located on U.S. territory and can provide the tracking, time management and control of the system. The satellite constellation is comprised of nine satellites in each of three orbital planes at about 7500 NMs in the medium altitude region corresponding to an orbit period of about eight hours. The orbits would be circular, highly inclined (possibly polar), and the three planes would be equally spaced (in their ascending nodes). Ranging signals from the satellites would be transmitted on two coherently related carrier frequencies of approximately 330 MHz and 1.6 GHz. These signals would incorporate sufficient data to generate identity codes, synchronize pulse patterns, and provide navigation information as required. This arrangement provides multi-satellite (four or more) visibility with a near minimum number of satellites which may be deployed so as to involve tracking of the satellites from stations on U.S. territory.

The TIMATION system concept should not be confused with the TIMATION I and II experimental satellites described earlier, or with the TIMATION III (redesignated Navigation Technology Satellite I) which was launched on July 14, 1974. TIMATION III went into a circular 7500 NM, highly inclined orbit. Subsequent TIMATION satellites will form part of the NAVSTAR demonstration program and are planned for launch into 12 hour orbits. The TIMATION III satellite employs ranging signals at approximately 330 MHz and 1.6 GHz and incorporates a PRN transmitter for use with NAVSTAR experiments. These experiments are planned to investigate the influence of the error sources cited earlier on advanced navigation satellite system performance.

(f) System 621B

System 621B designates a satellite system which was proposed and developed by the U.S. Air Force to provide a continuous, highly accurate, 3D navigation capability to any number of users on a global basis. The system position accuracy was expected to be within tens of feet and velocity accuracy to a fraction of a foot per second under most conditions. The system consists of three major elements: A complex of high-altitude satellites transmitting navigation signals, ground stations to track and command the satellites, and user receiving equipment. Global coverage may be obtained by four constellations of satellites in geo-synchronous orbits. A typical constellation consisted of a center satellite in circular, near equatorial, synchronous orbit and four outer satellites in synchronous, inclined, eccentric orbits. This is the "rotating-x" constellation configuration which provides reasonably good system geometry and stability. Regional coverage is obtained by placing one constellation over the region desired.

The satellite transmissions were planned to provide a coded anti-jam navigation signal which includes satellite ephemeris and timing data, a clear navigation signal to aid initial acquisition and for use by less sophisticated users, and "housekeeping" telemetry. The satellite periodically receives command and control signals which update the satellite ephemeris data, adjust the frequency of the onboard oscillator, or "clock," and modify the signal coding or other signal characteristics consistent with security requirements. Each satellite constellation is typically serviced by one master ground station and two signal monitor stations. The use of two-way ranging equipment located at the master ground station allows the position and velocity of each satellite to be determined by a satellite tracking process. This position and velocity information is processed by the master station computer to obtain the satellite ephemerides which are then transmitted to and stored in each satellite. The master ground stations also monitor and periodically adjust the synchronization of the oscillators onboard each satellite thereby updating and improving the timing quality of the system.

Signal propagation tests were conducted in 1969 by the Air Force's Space and Missile System Organization (SAMSO) using the L-Band transmissions of NASA's Advanced Technology Satellite No. 5 (ATS-5). Flight testing of transmitters and receivers commenced in January 1972 at White Sands Missile Instrument Range, Holloman Air Force Base. A four channel airborne receiver was tested using a ground-based constellation of transmitters with a spread spectrum PRN signal structure providing the ranging and timing data.

Launches of the Navy's TRIAD satellite, at 400 NM altitude and TIMATION III at 7500 NM provide for further testing of station keeping features, ephemeris, signal structure, ranging accuracy, time transfer, and geodetic measurements. A joint Army/Navy/Air Force effort will be conducted with the introduction of an Army PRN modulator experiment on the TIMATION III satellite. Additional testing using airborne single channel receivers integrated with inertial navigator units has been investigated using the ground-based transmitter array from the White Sands tests. This system may be supplemented by the signal from the PRN signal modulator onboard the TIMATION III satellite.

User equipment designs have been configured to satisfy the varying needs of several categories of users, including high-performance aircraft, helicopters, naval vessels, artillery, and backpacks for infantrymen. The user equipment is passive in that it requires no transmission by the user to either the satellites or ground stations, allowing simultaneous position fix capabilities to an unlimited number of system users. The coded navigation signals from the satellites provide the users with range, range rate (velocity), and ephemeris data. A typical user receiver provides one or more channels to receive the satellite signals, a correlation processing system which correlates the received PRN code with an internally generated sequence to obtain the range measurements, and a computer to perform the position computations and related tasks.

(g) Maritime Satellite Navigation/Communication Concept (MARSAT)

The Maritime Administration of the U. S. Department of Commerce has been investigating for some time the feasibility and benefits to the U. S. maritime community of utilizing satellite techniques. The objective of this program is to improve ship productivity, safety and control through the use of satellite-based navigation and communication systems. The current efforts are addressing the technical feasibility of the approach by utilizing and adapting existing state-of-the-art hardware and systems to the shipboard, shoreside and space applications. A satellite constellation and system could provide excellent communications and surveillance coverage, allowing the maritime community to link up with a Marine Data Coordination Center (MDCC) and other distributed support and control facilities.

Significant economic benefits may be feasible based upon the increased productivity provided through improved course tracking, weather routing, and logistic support. Improved navigation and communication may also provide means for reducing damage and pollution at sea. Technical developments are proceeding and sea demonstrations of selected satellite-aided capabilities have been accomplished. The implementation of national or of international navigation and communication systems involves a variety of other agencies, activities and national interests. Inter-agency liaison, coordination, and cooperation is essential for the successful development of the system capabilities and the introduction of these services to the maritime community on a viable basis.

(h) Position Location And Aircraft Communication Experiment (PLACE)

PLACE is a NASA funded experiment to obtain engineering data and practical experience for determining the operational feasibility of an air traffic control (ATC) satellite system operating in the aeronautical L-Band (1535-1660 MHz). The principal experimental system elements of the PLACE system consist of the large aperture parabolic antenna and communication transponder of the NASA Advanced Technology Satellite 6 (ATS-6), appropriate aircraft transmitters and receivers, and a primary control center.

The PLACE experiment has two main objectives. The first is to demonstrate the feasibility of two-way communication between ground terminals and aircraft. Included in the plans for accomplishing this objective are:

- The use of the ATS-6 synchronous satellite for relaying all communications.
- The use of the aeronautical L-Band frequencies for the satellite-to-aircraft links.
- The application of multiple access techniques through a satellite for aircraft-to-ground communication.

The second objective of the PLACE program is to investigate the feasibility and evaluate the absolute and relative accuracies of several position location techniques using a single satellite. These techniques relay various signals from the aircraft through the satellite to the control center for data processing and position determination.

These objectives are to be achieved through the use of three types of experiments: ground-based engineering, ground-based simulation, and in-flight performance. Included in the ground-based engineering experiments are the determination of needs, link performance measurements, multiple-access performance tests, evaluation of power and frequency control techniques, determination of system communications capacity, and evaluation of the quality and ranging precision obtainable for the various techniques considered. The ground-based simulation experiments involve the use of noise loading, signal simulation and other tests for the determination of system performance parameters.

Additionally, the position location and tracking accuracy performance characteristics for both fixed and mobile simulated aircraft terminals will be determined. In-flight performance experiments are planned in close cooperation with the FAA and possibly other coexperimenters. These will include aircraft flights to determine the effects of multipath, the ionosphere, the noise environment, and the geographic location of the receiver on both the L-Band communications and position location links.

The combined NASA and FAA research effort is planned to address a variety of important operationally-oriented experiments including communication link utilization, multiple aircraft tracking, determination of capacity limitations, possible extensions of current concepts or system uses, and tests with both ground and cockpit terminals.

(i) Aeronautical Satellite System (AEROSAT)

AEROSAT is a joint international program of the aeronautical authorities in the United States, Canada, and Europe for the experimentation, evaluation, and demonstration of the use of satellites to provide improved communication and surveillance capability for oceanic air traffic control. As currently planned, the program would involve the efforts of at least 11 different countries, involve a total cost of somewhere in excess of $150M, and cover a span of almost ten years. The program has been in various stages of discussion and planning since the late 1960's and involves a number of technical, operational, economic, and institutional considerations.

Background

Oceanic traffic control responsibility is delegated to various countries under standards and procedures agreed to by the International Civil Aviation Organization (ICAO). At present, air traffic control and air carrier communications for oceanic flights are almost entirely dependent on high-frequency (HF) voice radio circuits. Shore-based extended-range VHF facilities provide communications for oceanic flights up to 200-400 miles from the coasts; beyond that range, communications between aircraft and the traffic control facilities is by HF and is generally relayed through communication stations operated by the countries with the assigned traffic control responsibilities. HF has well-known propagation deficiencies, the adverse effects of which are only somewhat improved by the recent change to single sideband operation. HF frequency assignments are made in terms of "families," each of which consists of several frequencies; only one frequency per family is used at a time, with the selection based on propagation conditions. A single HF family has the capacity to handle the air traffic control communications for about 50 aircraft. But the number of HF families is limited, and the growing amounts of oceanic traffic, with its tendency to peak at certain hours for eastbound and westbound flows, points to future communication saturation and corresponding traffic delays. Currently there are four HF families available in the North Atlantic with plans for a fifth; two HF families are available for the Central Eastern Pacific. No further frequencies are now available for use. It is noted that the point at which saturation becomes a serious problem was previously predicted to be in the mid-1970's. However, the recent slow-down in traffic growth and the greater use of larger capacity aircraft has delayed the critical period to the early-to-mid 1980's.

For oceanic flights, ground-based radar or beacon surveillance, as employed over the continental U.S., is non-existent since the radar/beacon systems have the same line-of-sight limitations as VHF. Rather, aircraft positions are first derived in the aircraft by on-board navigation equipment and are then reported every 10° of longitude or every 40-60 minutes of flight to ground communication stations for relay to the ATC facilities. In view of these rather rudimentary provisions, oceanic flights must be provided with rather large lateral and longitudinal separations. To achieve a substantial reduction of separation standards, without which the airlines would be faced with the possibility of excessive delays and unfavorable track assignments, requires some form of surveillance to guard against navigational inaccuracies or blunders. Again, the need for and timing of such surveillance is strongly dependent on the rate of future traffic growth.

A third aspect of the current situation relates to the growing practice of the countries concerned to recover the costs of providing these oceanic communication, meteorological, and air traffic services by means of enroute user charges imposed on aircraft operators and their passengers. On a world-wide basis, for example, Pan American in 1974 paid over $10M per year in enroute user charges, with future charges projected to reach over $50M per year by the end of the decade. Modernization and consolidation of oceanic communication and ATC facilities offers one of the few hopes for checking this cost escalation.

These problems have been studied over the past 10-15 years by the aeronautical authorities of the countries affected, both separately and in conjunction with ICAO panels. There is general agreement today that a system of satellites in geostationary orbit offers the optimum solution to the problems, will provide the voice and data communications and surveillance capabilities desired for system operation and modernization, and that an operational satellite system will be needed in the mid-1980's when the available HF families in the Atlantic and Pacific become saturated. Studies have indicated that such a satellite system will remove the technical objections to possible consolidation of oceanic communication and traffic control facilities. As noted, such consolidation could significantly reduce the future costs of providing communication and air traffic services.

AEROSAT is the planned forerunner of such an operational system. It is proposed for experimental, evaluation, and demonstration purposes only, and is intended to gather the information required to proceed with the technical design of a future operational system meeting ICAO-sponsored standards. To achieve an operational system by the mid-1980's, scientific data collected from the NASA-sponsored ATS satellite series must be supplemented with the broader systems experimentation, evaluation, and demonstration program of AEROSAT. With a decision now to proceed with this program, initial satellite launch would be in 1979, followed by several years of experimentation, demonstration, and data collection. This data would be used as the basis for establishing ICAO standards, normally a lengthy process since it involves international consideration, coordination, and approval. In parallel, and also on an international basis, the institutional arrangements and provisions for an operational system must be developed and approved. Lastly, production and implementation of the necessary avionics and other elements of an operational system would be initiated. These steps each take considerable time, and affect the scheduling of the AEROSAT operational system and whether or not this capability can be achieved by the mid-1980's.

Frequency Selection

A technical design problem should be highlighted, relating to the choice of the frequencies for the satellite-to-aircraft link. VHF would seem to be the first choice, but VHF frequencies are in great demand, and future use over the U.S. will require operating at 25 kHz spacing as opposed to the present 50 kHz separation. For surveillance purposes, propagation aspects of VHF offer some difficulties in obtaining ranging accuracies of less than a few miles. For these reasons, international aeronautical authorities selected L-band (at about 1600 MHz) for the satellite-to-aircraft link. ICAO adopted L-band and the U.S. stated it as a policy position in early 1971. Nevertheless, the airlines have reservations relating to this decision. They are concerned with the current lack of L-band avionics and the uncertain cost picture. Coupled with their experience and equipments operating at VHF, and being unconvinced of the need for independent surveillance, they would prefer that VHF be used.

Program Plans

Plans for an aeronautical satellite program have evolved from international discussions initiated in the mid-1960's. The present AEROSAT program is the direct result of the discussions begun in June 1971, when the European Space Research Organization (ESRO), acting for ten countries in Europe, made a proposal to the U.S. Department of Transportation (DOT) and the Federal Aviation Administration (FAA) for a joint program with partnership arrangements and cost-sharing provisions. In the discussions of the subsequent three years, the program evolved to the following general features:

- The program is directed to the Atlantic Ocean and involves the initial construction and development of two satellites. Additional flight units will be built, as necessary, to achieve two satellites in orbit.

- The ownership of the space segment, consisting of satellites and an associated ground tracking and control facility, will be shared equally by ESRO and a U.S. private sector co-owner, with Canada owning about six percent.

- The FAA would only be a user of the space segment and would achieve access to the satellite capability on a contractual basis from the U.S. co-owner for a five-year period.

- The principal document for implementing the program is a Memorandum of Understanding, or MOU, signed by the users: ESRO, Canada, and the United States.

- Under the provisions of the MOU, an AEROSAT Council representing aeronautical authorities would meet once or twice a year to review progress and to approve or make major decisions. The guiding principle of the Council, as well as of the entire joint program, is one of partnership, with approval of both FAA and ESRO being required on any substantive matter.

- The Council would establish a jointly-manned AEROSAT Coordination Office to handle the day-to-day aspects of the program for the users. This office would be concerned with the detailed planning and execution of a coordinated test and demonstration program and would provide liaison with the representatives of the owners on the use of the space segment.

- The space segment will be utilized in a test program to be coordinated among the users under terms established in the MOU. Generally, the users will each provide their own aircraft, avionics, and associated ground equipment. The test program will also permit each participant to use the space segment capability to conduct any special tests, evaluations and demonstrations that he deems necessary.

During 1971 and 1972 and based on ICAO recommendations and U.S. Government planning, the design for AEROSAT assumed that the channels linking the satellite and the aircraft would be at L-band. In early 1973, DOT/FAA approached Congress for approval to proceed, only to find that Congress was concerned with the strong opposition raised by the U.S. airlines and therefore, stipulated that the U.S. could sign the MOU only with approval and support of the airlines. The airline concerns were severalfold: (a) obtain assurances that participating countries would not attempt to recover costs of the experimental program through user charges imposed on aircraft operators, (b) limit the size of the program and ensure all possible cost economy considerations and (c) add a limited VHF capability for test and evaluation.

In view of U.S. incurred delays in reaching agreement on the program in 1971 and 1972, and under strong pressure from Europe to avoid further delays, the FAA defined a package of changes acceptable to the U.S. airlines and Congressional approval was obtained to discuss these changes with ESRO and Canada, which was done. The idea of a hybrid capability, which ran counter to ICAO recommendations, was not readily received by the European aeronautical authorities. By early 1974, all parties agreed on a revised Memorandum of Understanding incorporating the hybrid capability and certain other modifications. By August 2, 1974, representatives of the United States, Canada, and ESRO signed a Memorandum of Understanding (MOU) on a Joint Program of Experimentation and Evaluation using an Aeronautical Satellite Capability referred to as the Joint AEROSAT Evaluation Program. The MOU now calls for the construction and launch of two satellites, each with five L-band and two VHF channels. One full earth terminal, or ASET, will be located in Europe; while one full earth terminal, consisting of two similar sections each owned by U.S. and Canada, will be used on the western shores of the Atlantic.

Objectives

The overall objectives of the AEROSAT program are to:

- Provide a minimum cost, minimum configuration oceanic air traffic control system, using satellite technology as an element of the system, which will support the experimentation and evaluation necessary to determine the merit of proceeding into an operational system and to define the technical and operational characteristics of such a system.
- To use that minimum AEROSAT system configuration in evaluating competitive technical techniques and operational concepts.
- To develop the technical and operational characteristics of an optimum international operational system.

Specific program objectives in support of the three broad objectives are as follows:

- To bridge the gap in time and knowledge between the current experimental efforts, and an operational satellite capability anticipated after 1980. The initial capability must be an extension of the current experimental efforts and provide verification of system design: subsequently, it must demonstrate that it will be possible to attain the quality of service expected in an operational phase for air traffic control (ATC) and air carrier purposes.
- To provide experience in technical, operational, and managerial areas required in advance of establishing a fully operational capability.
- To evaluate the technical and operational performance of voice and data communications between ground and aircraft over various areas.
- To permit experimental evaluation of dependent and independent surveillance capabilities, and of navigational data derived by an aircraft utilizing ground and satellite transmissions.
- To contribute data to enable ICAO to develop its Standards and Recommended Practices (SARPS) for an operational capability.
- To continue the evaluation of the propagation characteristics of the chosen frequency bands if prior tests using the NASA ATS-5 and ATS-6 indicate the need for supplemental data.
- To evaluate alternative modulation techniques and technical and operational performance of voice and data communications.
- To assess the performance and characteristics of aircraft installations.
- To provide the U.S. oceanic airline companies with a limited number of avionics to provide them with first-hand knowledge of performance; to gather operational type experimental data; and to provide for sufficient aircraft operations to enhance evaluation of operational concepts and procedures.
- To explore the feasibility of reducing the number of oceanic air traffic control centers and thus reduce operating costs to the airline users and reduce the number of government owned and operated facilities.

System Description

The AEROSAT experimental and evaluation system will provide the communication functions necessary to evaluate performance of voice and data communications between ground and aircraft, to experiment with aircraft surveillance capabilities, and to develop operational standards and practices. A summary of the system configuration as currently envisioned is shown in Figures 4 and 5.

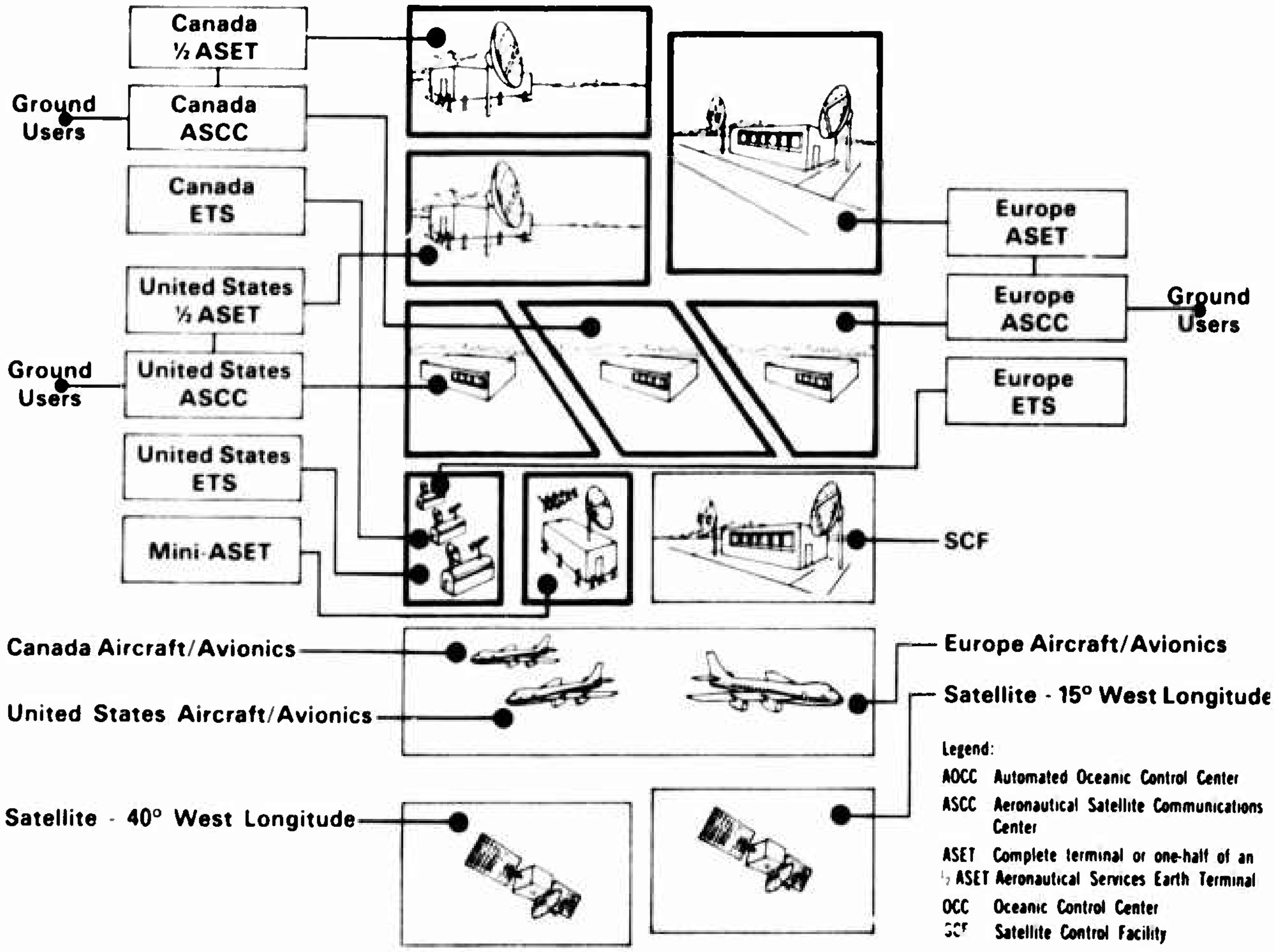
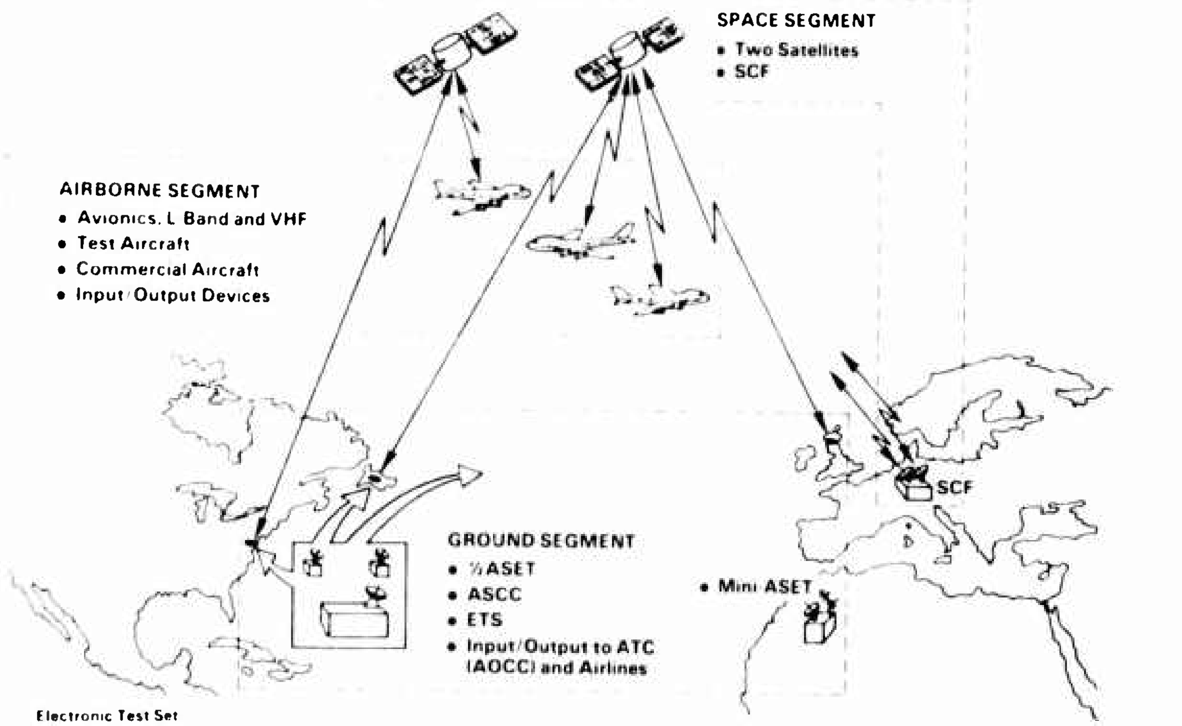


Figure 4 - AEROSAT System Summary



ETS Electronic Test Set
Mini ASET Small Remote Earth Terminal for L Band and VHF Accessing

Figure 5 - AEROSAT System Configuration

The AEROSAT system hardware includes two satellites and a Satellite Control Facility (SCF), one AeroSat Earth Terminal (ASET) located in North America, one in Europe, a few electronic test sets (ETS), L-band and VHF avionics, cooperating aircraft, and interfaces to user terminals (e.g., Automated Oceanic Control Centers and airline companies). The North American ASET will be shared by the U.S. and Canada.

The satellite facilities are shared by all participants, and the ground facilities and airborne systems are provided by individual participants.

Space Segment

The space communication facility is provided by two geostationary satellites that serve as radio relays between the ground and airborne users of the AEROSAT system. Communication evaluations can begin after one satellite is on station. A second satellite is required as part of the initial system to enable evaluation of surveillance techniques employing range measurements via two satellites as shown in Figure 6. The two satellites will be deployed to provide Atlantic Ocean coverage and will be stationed between 15 and 40 degrees west longitude. The physical arrangement of the satellites will be determined by the satellite contractor, based on satisfying the need for the specified number of channels and optimizing reliability within the payload weight limitations of the designated NASA launch vehicle. Each satellite will be of like design.

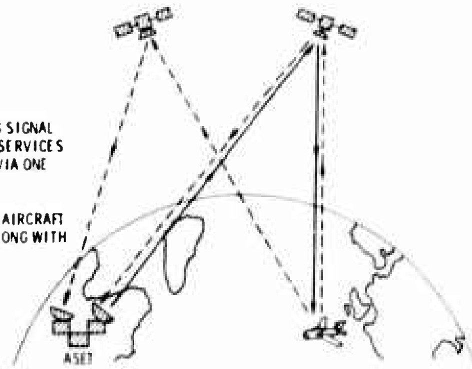


Figure 6 - Independent Surveillance Concept

The following specifications for a satellite as represented in Figure 7 are tentative and are used for illustrative purposes only. The general satellite characteristics are listed in Table IV. The tentative characteristics permit considerable latitude in the design of an acceptable solution.

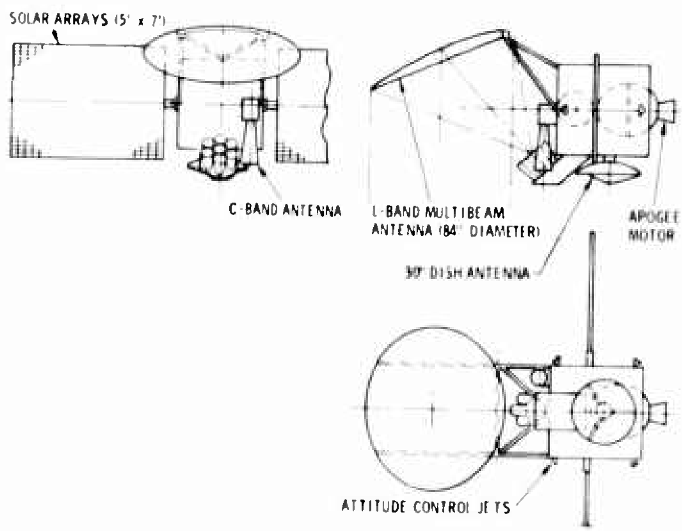


Figure 7 - Spacecraft Configuration

The specific parts of the spectrum to be used by the AEROSAT system are planned to be as follows:

- Ground-to-Satellite:
  5,000 MHz to 5,125 MHz (C-band)
- Satellite-to-Ground:
  5,125 MHz to 5,250 MHz (C-band)
- Satellite-to-Aircraft (Communication/Surveillance):
  1543.5 MHz to 1558.5 MHz (L-band)
  125.425 to 125.975 MHz (VHF)
- Aircraft-to-Satellite (Communication/Surveillance):
  1645 MHz to 1660 MHz (L-band)
  131.425 MHz to 131.975 MHz (VHF)
- Satellite-to-Aircraft (Experimental):
  1543.5 MHz to 1558.5 MHz (L-band), or
  1542.5 MHz to 1578.5 MHz, if required.

TABLE IV - SATELLITE GENERAL CHARACTERISTICS

SATELLITE WEIGHT	ON-ORBIT	1035 POUNDS
	LAUNCH	2000 POUNDS
SIZE	DIAMETER	7 FEET MAXIMUM
	HEIGHT	18 FEET WITH ANTENNA DEPLOYED
ELECTRICAL POWER	ARRAY POWER @ LAUNCH	1070 WATTS
	MINIMUM AFTER 5 YEARS	800 WATTS
	ECLIPSE	200 WATTS
DESIGN LIFE	5 YEARS	
CONFIGURATION	3 AXIS STABILIZED SATELLITE USING MOMENTUM WHEELS AND HOT GAS: DEPLOYABLE SOLAR ARRAY: APOGEE MOTOR FOR INJECTION	
STATION CHANGING CAPABILITY	40 DEGREES/WEEK	
ORBIT	SYNCHRONOUS/EQUATORIAL	
LAUNCH VEHICLE	THOR-DELTA 3914 WITH 8-FOOT DIAMETER FAIRING	
LAUNCH CAPABILITY	2000 POUNDS INTO SYNCHRONOUS TRANSFER ORBIT	

The satellites will be supported by a Satellite Control Facility (SCF) that will perform orbit determination functions and tend to stationkeeping, housekeeping, and necessary operational controls via a radio link between the control facility earth terminal and the satellite. The facility will monitor satellite operation in geostationary orbit and during positioning/repositioning maneuvers, command the satellites' orbits to the required accuracy, provide channel status information, provide antenna pointing information, and provide satellite position information. These functions will be performed by one control facility for the Atlantic region. The facility will maintain C-band communication links to each satellite for control and monitoring. The SCF, located in the coverage area, will include two antennas and C-band telemetry systems for monitoring and control of two satellites simultaneously. The computational facilities for orbit determination and control will also be contained in the SCF.

The function of the space segment is to perform the relay of communications between ground and aircraft, and between pairs of ground stations. The satellites communicate via C-band links to ground terminals and via L-band and VHF links to aircraft terminals. Initially, one satellite will be used to provide limited coverage of the Atlantic region. A second satellite will be added to increase the channel capacity to that required for full coverage and to enable independent surveillance by range measurements via two satellites. Three types of channels are provided:

(1) Communication channels for ground-to-air (forward channel), air-to-ground (return channel) and ground-to-ground voice or data messages. These channels will provide voice quality comparable to existing VHF (but over greater ranges) and data error rates of better than 10^{-5} at 1200 BPS with normal power or at 2400 BPS with twice normal power.

(2) Surveillance channels for two satellite two-way ranging. Surveillance channels will provide the capability to perform independent surveillance which consists of making aircraft position determinations in real time by an ASCC utilizing range measurements to aircraft and aircraft altitude communicated via satellite. These channels also have the capability of serving as air-to-ground or ground-to-air communications channels.

(3) Experimental channels for wide band two-way ranging and communication experiments using two satellites.

Upon ground command, the ground-to-air channels can be rearranged so as to provide one channel at twice the normal power and additional channels at their normal power level or greater. One forward communication channel and at least one return communications channel in coverage area A, indicated in Figure 8 (see next page), will remain operable when the experimental channel is activated in coverage area C. This capability will be provided on one of the satellites serving the Atlantic region as a minimum. The channels in coverage area B will have the same capability as those for area A. In area C, one ground-to-air and one air-to-ground experimental (wide band) channel will be provided. The coverage area D will be subdivided as convenient and included in areas B or C or both.

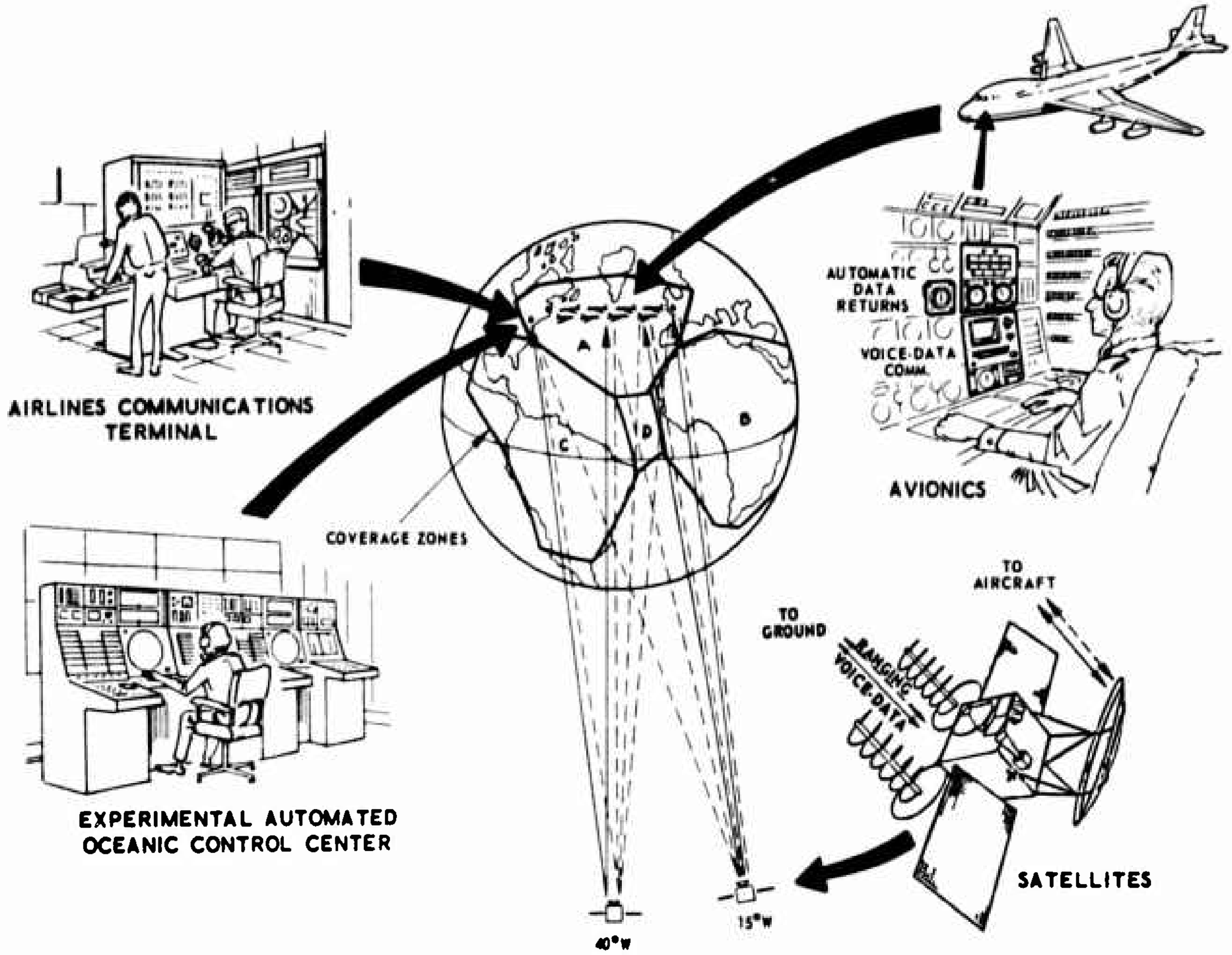


Figure 8 - Principal System Elements and Coverage Regions

An air-to-ground communication channel will be available in the overall Atlantic coverage area. It is desired that one ground-to-air communication channel and two air-to-ground communication channels be available in this area.

Two simplex ground-to-ground channels will be available for coordination during the AEROSAT program between ground stations located anywhere in that part of the earth surface from which the satellite can be seen with an elevation equal or greater than five degrees.

Ground Segment

Ground facilities will be provided by the users of the AEROSAT system to enable communications between ground stations and aircraft via the satellites as shown in Figure 9. Each organizational entity constituting a primary participant in the AEROSAT system will provide one or two ground facilities each, including an earth terminal, a control center, interface with user stations, and test systems.

The Aeronautical Services Earth Terminal (ASET) provides reception and transmission of communications on the satellite links, and interfaces with ground users via a control center. The Aeronautical Satellite Communications Center (ASCC) provides control of the communications service, supervising user access and distributing communication messages. Test and monitoring of satellite communications is provided by an Electronic Test Set (ETS) that includes receiving and transmitting capabilities similar to an aircraft avionics set.

Initially, two ASET's will be deployed, one in North America and one in Europe. The one in North America will be divided into two parts -- designated 1/2 ASET's; one will be located in Canada and the other in the U.S. Each 1/2 ASET serves a single satellite but when interconnected full service is provided. The 1/2 ASET consists of a single C-band antenna, receiver and transmitter sections, modems, ranging equipment, channel controls, and interconnection to its associated ASCC. L-band and VHF equipment will also be included to enable channel control and test.

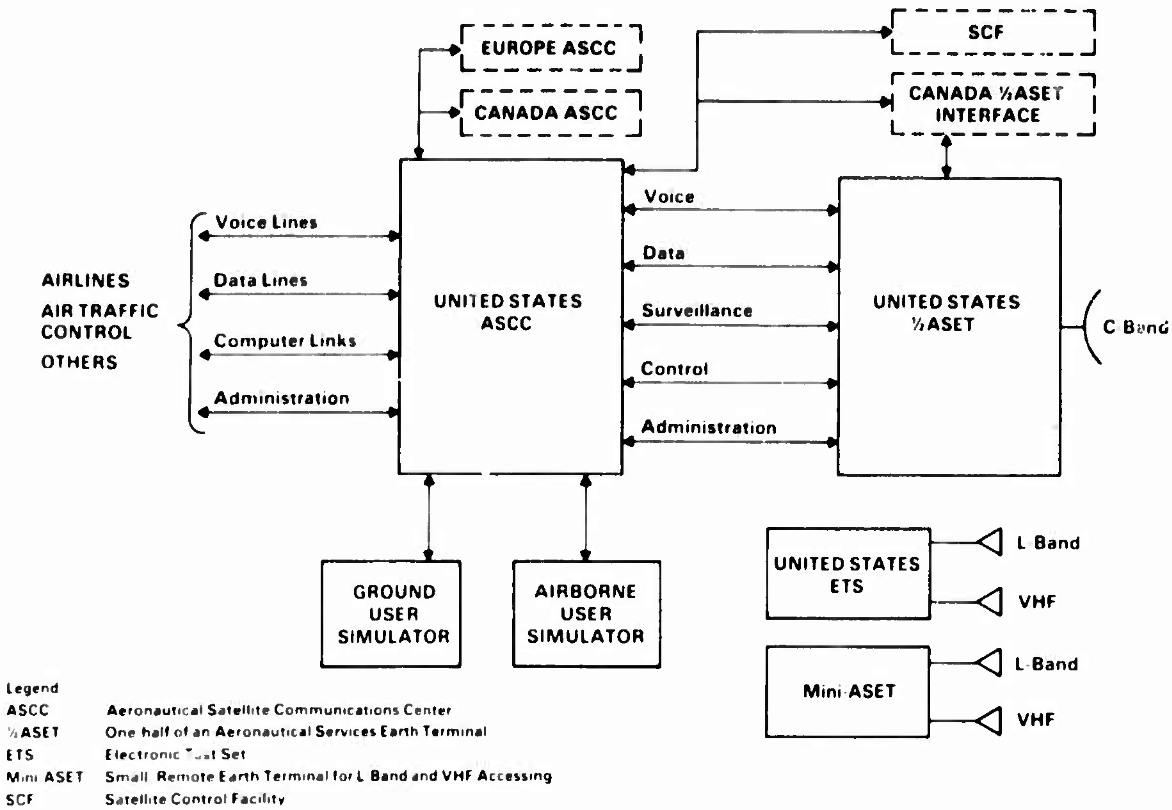


Figure 9 - U.S. AEROSAT Ground System

An Aeronautical Satellite Communications Center (ASCC) will be provided to supervise the use of communication services for all ground users desiring access through the ASET serving a given region. Each ASET will be associated with one ASCC and both will be located at the same facility. The ASCC hardware includes terminals to interconnect to users and satellite facilities, a computer, channel controls and switching, and control and test consoles.

The ASCC will enable test and evaluation of a variety of communication techniques applicable to AEROSAT communication service operating modes. The ASCC will provide the capability for communications channel switching, signaling, message processing, and test control to enable evaluation of the AEROSAT communications potential. The ASCC serves as the functional interface between the associated ASET and ground user terminals and coordinates related ground control activities. A major function of the ASCC is to perform the channel control needed to interconnect user stations over voice bandwidth communications channels for the exchange of voice or data messages. The ASCC will respond to channel requests in order of priority, provide logic to control, accept or send messages.

Electronic Test Sets (ETS) will provide system test and calibration functions. A few sets of equipment will be located on the ground or in aircraft to communicate with the satellites on L-band or VHF frequencies. The ETS hardware will include L-band communications equipment comparable to the avionics equipment, and instrumentation to measure system performance and enable system calibration.

Air Traffic Control Center (ATCC) users of the AEROSAT system will provide control facilities to interface with the system via the ASCC. Several ATCC user facilities may connect to one ASCC for access to the associated coverage area. The ATCC communications will include exchange of messages between controllers and aircraft for flight control and reporting, signaling for system access, aircraft polling instructions, and surveillance data. The U.S. ATC facilities interfacing with AEROSAT will include an experimental Automated Oceanic Control Center (AOCC) test bed. In addition, interfaces will be provided to existing operational Oceanic Control Centers (OCCs) for system evaluation by OCC controllers.

Other users besides Air Traffic Control Centers (ATCCs) will participate in the use and evaluation of the AEROSAT system as associate users. These other users will include airline company communication facilities, meteorological offices, special evaluation and experimental facilities, military facilities, etc. These communications will include message exchanges between ground and aircraft, automatic return of aircraft derived data, weather reports, etc. The associate users facilities will interface with the ASCC for access to the system.

Mini-ASETs will be provided to permit ground access to the system via L-band or VHF channels from isolated airline or ATC locations which have limited or difficult access to communications channels. The Mini-ASET will be similar to the ETS and operate in a manner equivalent to a ground-based airborne installation. It will be packaged in a transportable configuration and the basic VHF and L-band components will be obtained from the avionics development contracts.

Airborne Segment

Each AEROSAT user will arrange for obtaining and equipping test or cooperating aircraft to provide air terminals as required for evaluations and experiments. The AEROSAT avionics are designed with maximum flexibility to meet test objectives. Avionics system and subsystem configurations will allow technical and operational performance evaluation of a number of techniques for voice and data communications between ground and aircraft, experimental evaluation of surveillance concepts, evaluation of various access control techniques and a comparison of VHF and L-band frequencies for satellite communications. Table V provides a requirements summary.

The AEROSAT Test and Evaluation Avionics is a VHF and L-band system that will provide the necessary airborne hardware and software to permit a wide range of tests designed to provide experience in oceanic air traffic control (ATC) communication using satellites. As shown in the simplified block diagram of Figure 10, the minicomputer is seen to be the focal point of all digital signal and control paths, permitting a high degree of system flexibility through software design.

TABLE V - U.S. AVIONICS REQUIREMENTS SUMMARY

FUNCTIONS

- Provide Aircraft Terminals for Test and Evaluation
- Transmit and Receive L-Band and VHF Signals via Both Satellites
- Communicate Voice, Data, and Ranging Signals
- Provide Input/Output for Voice, Data, and Surveillance Tests
- Monitor Channel Status
- Display and Record Test Data
- Collect Data During Air Carrier Revenue Flights

DESIGN CRITERIA

- Installed in FAA and Airline Aircraft Participating in Tests
- L-Band and VHF Capabilities
- Provide Representative Configuration for Future Airline Developments
- Collect System Data Automatically

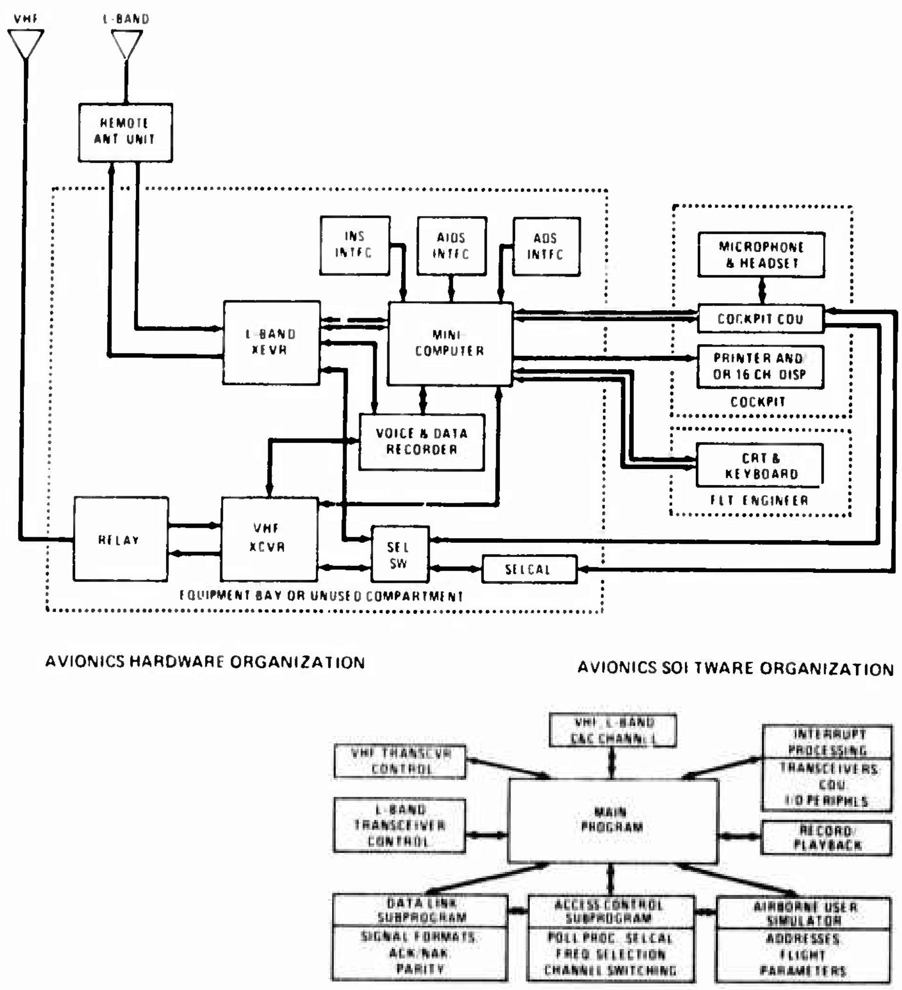


Figure 10 - U.S. Avionics Block Diagrams

(j) Advanced Air Traffic Management System Concepts (AATMS)

The system concepts of any air traffic management system can be viewed, in part, by considering the major subsystems involved in performing surveillance data acquisition, navigation, control data processing, and communications. There are, of course, many combinations of alternative concepts that might be applied in any future system. In performing concept formulation work for the AATMS study, the Boeing Company's commercial airplane group and the North American Rockwell Corporation's Autonetics Division conducted parallel tradeoff studies to independently arrive at recommended future system concepts. There are significant similarities as well as some basic differences, in the proposed concepts. For this discussion, the Boeing Company system concept is identified as System I and the Autonetics system concept as System II.

System I Concept

System I shown in Figure 11 uses satellites for both surveillance and en route communications. Communications in the terminal area, however, is by direct ground-air links using the same airborne equipment as that used for en route operations. Both L-Band digital data and voice communications are employed. Domestic en route and oceanic service is relayed between airplane and control centers by synchronous satellites. On and near the airfields, communication is direct between aircraft and the control facility. Voice capability is provided for nonstandard messages, backup, and VFR flight.

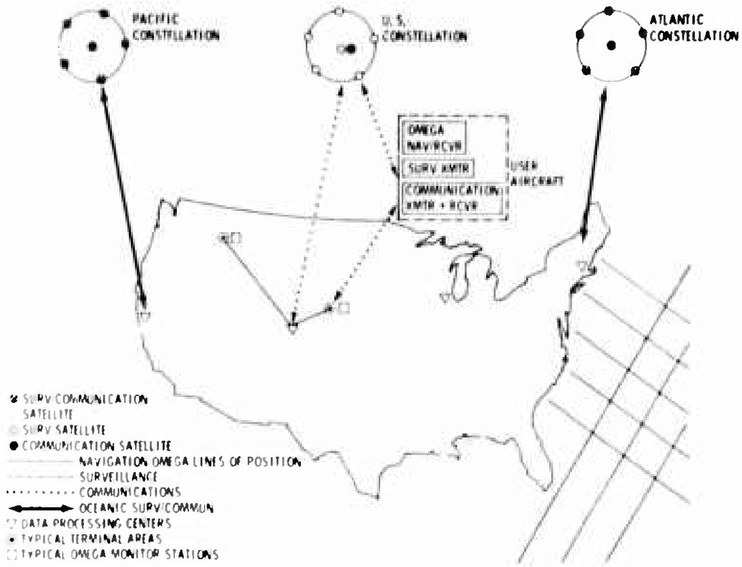


Figure 11 - AATMS System I Concept

The Omega Navigation System is proposed for use as the basic navigation environment in both the domestic airspaces and in oceanic areas. The differential mode would be employed in the continental United States (CONUS) to provide increased accuracy capabilities where needed for higher traffic densities. Precision navigation near the airfield is based on the coverage of the microwave landing system under development, as shown in Figure 12. Position, identity, and altitude data are obtained for surveillance purposes in all airspace regions from the satellite system. Each aircraft transmits its own unique ranging code once every second. The code is relayed simultaneously through four satellites to the appropriate control center. The control center calculates each aircraft's position based on the differences in time-of-arrival of the signal at the satellites.

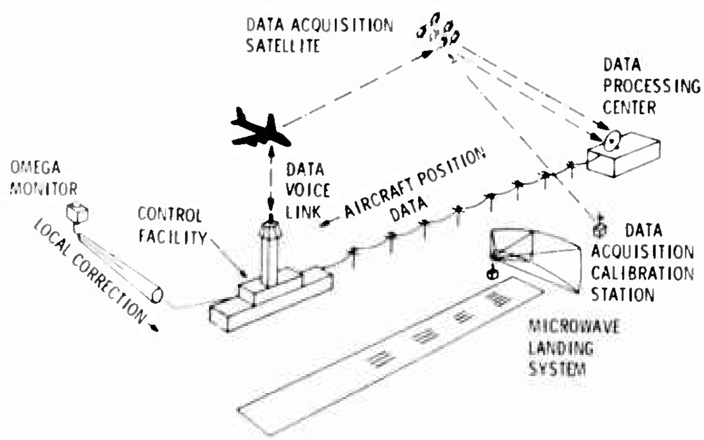


Figure 12 - System I Terminal Area Operation

The data acquisition subsystem can provide position data on about 60,000 instantaneous users through the combination of frequency, code, and random time-division multiplexing of the aircraft-emitted, free running beacon signals. User aircraft equipment required for operation in the system includes a 2,000-watt peak power L-band pulse transmitter (0.5 watts of average power) coupled to a top-mounted hemispherical coverage antenna. The beacon modulation format consists of a two million bits per second Pseudo-noise range code, a code length of 511 bits with 16 codes, and an integration time of 255 microseconds. The pulse repetition period varies from 1 to 1.3 seconds with a spacing of 50 microseconds. Frequency access is provided by use of six channels, with a channel bandwidth of four megahertz providing a capacity of 10,000 users per channel.

Data processing would be accomplished primarily with a highly integrated network of digital computers. The ground communication network would accommodate the effective transfer of information from computer to computer. The control data processing and display functions for en route control for the U.S. would be centralized in two data processing air route traffic control centers as shown in Figure 13 . Two additional centers, one on each coast of the U.S. would handle oceanic control in co-operation with foreign facilities. Terminal area control would be provided by separate centers located in each of the approximately 150 terminal hub areas expected to contain most of the larger airfield complexes.

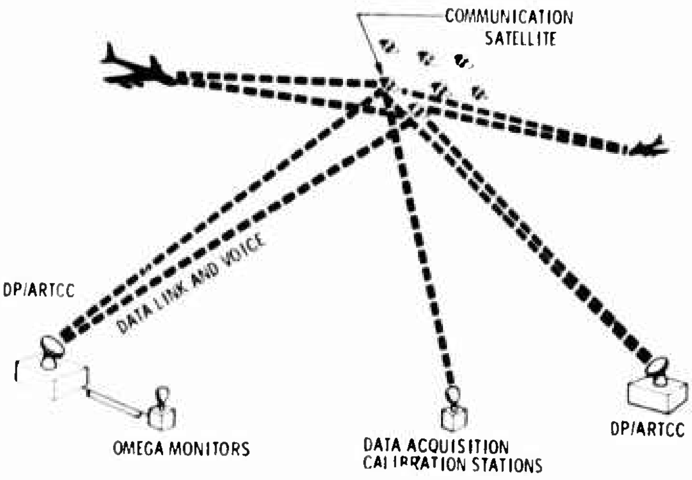


Figure 13 - System I En Route Operation

Three satellite constellations would be used to cover the domestic U.S. and two oceanic area needs. The oceanic communication and data acquisition would be provided by a six-satellite constellation over each ocean area. The CONUS configuration would use the same basic six-satellite configuration augmented with a seventh geostationary communication satellite to handle the domestic message load for en route traffic in the U.S.

Airborne navigation and guidance services would range from a basic radio area navigation capability for general aviation to four-dimensional navigation (3D with precise time) and guidance using combined radio-inertial data on the best equipped scheduled air carrier aircraft. The bulk of oceanic traffic would use combined hyperbolic-inertial navigation systems.

System I could be implemented to meet the demands for air traffic services projected for the 1990's. The system is structured to provide a high degree of independence among the three primary subsystems which perform the surveillance, navigation, and communication functions. Data acquisition is satellite based; navigation is land based, and communication uses a combination of land- and satellite-based elements.

System II Concept

System II shown in Figure 14 uses satellites to perform the surveillance function for all user aircraft. In addition, satellites are used to provide navigation and communication links for all en route aircraft. However, ground based antennas on towers are proposed to supply communication links and to supplement the navigation data for terminal area operations as shown in Figure 15.

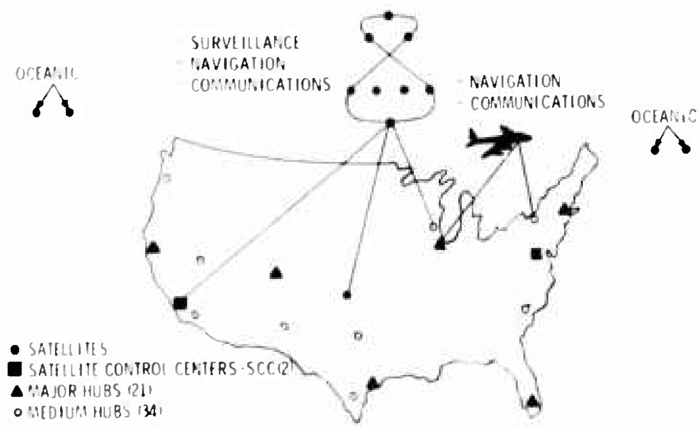


Figure 14 - AATMS System II Concept

A multifunction signal waveform would be used to permit integration of all surveillance, navigation, and communications equipment. In the System II concept, the waveform consists of a pulse triplet. Each aircraft is assigned an identification which consists of a Pseudo-noise (PN) code, a transmission frequency, and a particular spacing between each pulse. This concept utilizes 10 PN codes, 10 frequencies, and 100 possible spacings between the second and third pulses, which provides 10^6 possible identification combinations. Each aircraft has a unique surveillance identity and transmits its pulse triplet periodically to provide a surveillance signal.

Surveillance and control data processing for en route aircraft are accomplished at one of two national satellite control centers. Surveillance and control functions for terminal aircraft are performed at hub control centers located at each major and medium hub area. The smaller terminal areas are considered to be part of the en route control structure.

The two satellite control centers provide mutually supporting capabilities to perform surveillance data processing and provide conflict intervention control for en route aircraft. One satellite control center would be located in and have jurisdiction over en route operations in the Eastern United States, while the other center would be located and provide service in the Western United States. Services for operations in the adjacent oceanic areas would also be provided by these en route centers. A National Flow Control Center (NFCC) would also be collocated in one of the satellite control centers. The National Flow Control Center would provide real-time supervisory control over the operations of the entire air traffic management system and schedule the traffic flow of aircraft on a nationwide basis. Terminal area operations in major and medium hub areas would be managed by the individual terminal area hub control centers. Airfield landing guidance would be provided by the microwave landing system.

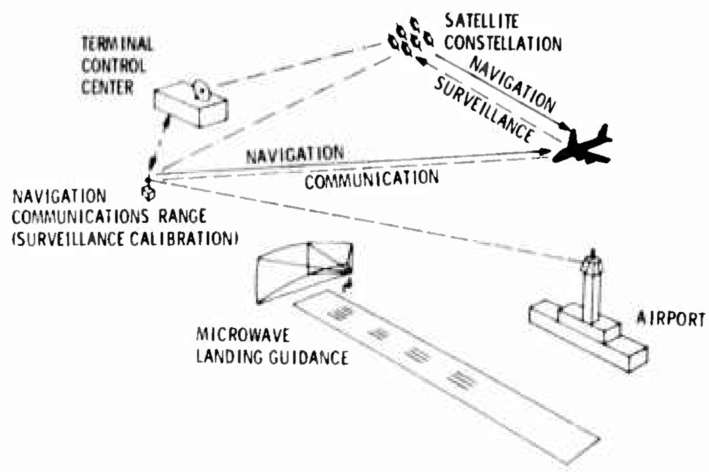


Figure 15 - System II Terminal Area Operation

The satellite constellation proposed in the System II concept, as depicted earlier, employs eight satellites in a highly inclined ($i=80^{o}$), eccentric ($e=0.25$) orbit. All the satellites are equipped with navigation transmitters and surveillance transceivers. In addition, two communication satellites in geostationary orbit over the equator contain communication transceivers to provide constant communication in the United States airspace. Each communication satellite is capable of performing the total communication relay function to protect against single point failure. Also, a pair of redundant satellites for each ocean area in geostationary orbits provide oceanic traffic with both digital and voice communications.

The surveillance subsystem obtains position data by measuring the difference between time of arrival of signals from aircraft, these signals being relayed to ground control centers via satellite. The times of arrival (TOA) of each of four signals are compared at the processing centers and three differential TOA's are derived. The processing centers can then calculate the latitude, longitude, and altitude of each aircraft using the differential TOA's and satellite position data.

The navigation technique proposed for the System II concept uses satellites for one-way ranging. Each satellite is assigned a time slot and, through the use of precise timing, maintains its synchronization relative to a master station. The user aircraft, also employing accurate timing, computationally synchronizes its clock through an iterative timing process. A range measurement is thus obtained each time a satellite transmits. By recurrently combining these measurements, aircraft position is determined.

The System II concept provides a high capacity integrated system approach to meeting the demands for air traffic management service in the late 1980's and beyond. A well-defined systems management approach would be required for system implementation and operation phase-in to achieve the potential benefits offered in the System II concept. The combination of automation, equipment redundancy, and an air-managed backup control concept can assure safety in the event of system failures. A summary of System II capabilities is shown in Table V. An indication of certain of the more significant similarities and differences between System I and II is presented in Table VI.

The candidate system concepts presented are two proposed system concepts among several possible future system approaches being evaluated by the Department of Transportation and FAA. The concept evaluation work now in progress involves further contractual efforts with industry and coordination with user groups as well as system definition studies by both the FAA and the Transportation Systems Center. This comparative evaluation is being carried out on both the system and subsystem levels, and includes the consideration of such factors as safety, capacity, cost, technological risk, growth capability, and ease of system transition.

TABLE VI
AATMS SYSTEM II CAPABILITIES

Function	Terminal	En Route	Oceanic
Navigation	Ground towers	Satellite	Satellite
Accuracy	30 ft	150 ft	250 ft
Update interval	4 sec	8 sec	8 sec
Coverage	Entire area	Entire U.S.	Atlantic, Pacific
Surveillance	Satellite	Satellite	Satellite
Accuracy	50 ft	250 ft	450 ft horiz., 250 vert.
Update interval	2 sec	8 sec	8 sec
Coverage	Entire area	Entire U.S.	Atlantic, Pacific
Communications			
Data rate	25 K BPS	25 K BPS	25 K BPS
Spectrum	L Band, 20 MHz C Band, 60MHz	L Band, 20MHz C Band, 60MHz	L Band, 20MHz C Band, 60MHz

TABLE VII
SYSTEMS I AND II
COMPARISON OF GROSS CHARACTERISTICS

SIMILARITIES

- HIGH LEVELS OF AUTOMATION
- GROUND-MANAGED
- EXTENSIVE PRE-PLANNING - HIGH DENSITY AIRSPACE
- SATELLITE SURVEILLANCE AND COMMUNICATION
- COOPERATIVE SYSTEM

DIFFERENCES

- SYSTEM I - DISSIMILAR NAVIGATION AND SURVEILLANCE
- SYSTEM II - INTEGRATED AVIONICS

(k) ASTRO-DABS

ASTRO-DABS is a system concept developed by the MITRE Corporation in response to DOT and FAA interest in satellite-based advanced air traffic management systems. This system concept is one of several which have addressed the air traffic control surveillance, reporting and navigation requirements of the CONUS in the 1980-90 and beyond time period. The concept has an important distinction which separates it from other satellite-based ATC systems; it is designed to be compatible with the evolving Discrete Address Beacon System (DABS) currently under development.

The ASTRO-DABS is an ATC satellite-relayed surveillance, navigation, air-to-air CAS and data link system concept. It has the potential for realizing the advantages of satellite systems in terms of coverage and accuracy, and to be cost competitive with alternative techniques. It incorporates features which largely circumvent the satellite system performance hazards resulting from hostile acts against the system. Additionally, ASTRO-DABS has the character of an evolutionary system in that it can evolve from the upgraded third generation ATC system.

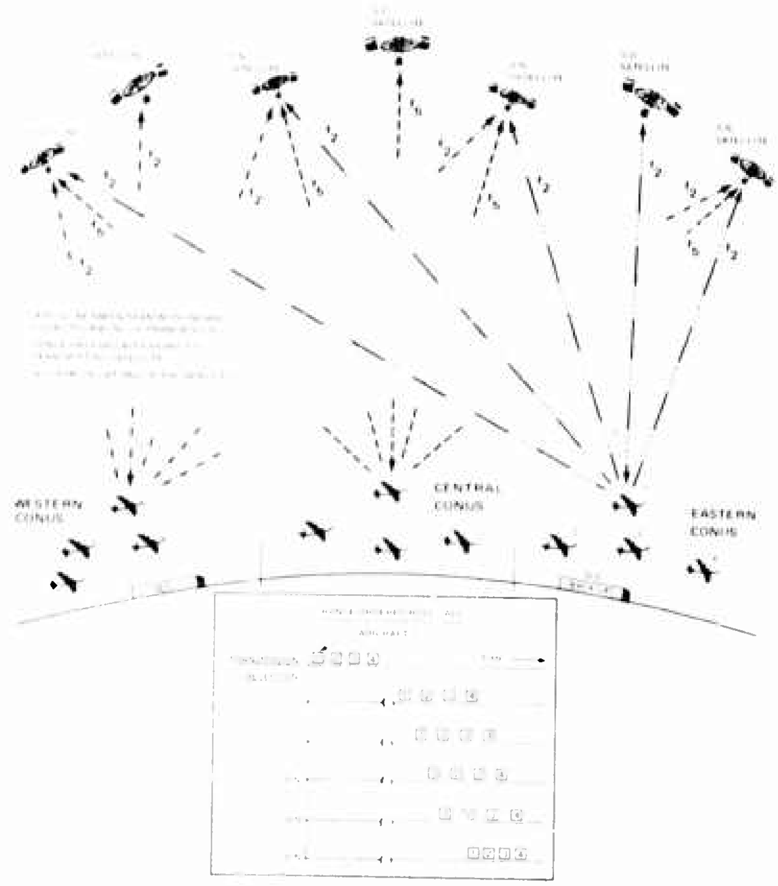


Figure 16 - ASTRO-DABS Surveillance Concept

The principal advantages of ASTRO-DABS are as follows:

- o Moderate cost avionics.
- o High capacity data link.
- o The surveillance acquisition and start-up after failure modes are very fast.
- o The normal surveillance tracking mode provides high accuracy and has essentially no system interference.
- o An accurate and moderate cost area navigation system is provided.
- o A three dimensional, highly accurate air-to-air CAS is provided whose cost impact is only on the CAS-equipped user.
- o A ground-based back-up is available in case of up-link (L-Band) jamming.
- o System capacity is high with about 4 million addresses accommodating 50,000 users with growth potential to 100,000 users.

Satellite System Characteristics

The characteristics of the satellite segment of the system are shown in Figure 16. ASTRO-DABS aircraft operate on three separate equal bandwidth (5.33 MHz) channels through the satellite relay. There are two aircraft-to-satellite channels (f_2, f_5) and one satellite-to-aircraft channel (f_1). Three large surveillance and data link (S/D) satellites each interrogate and receive responses from one-third of CONUS. At any instant there are at least six surveillance and navigation (S/N) satellites which receive responses from all of CONUS. Thus, every aircraft in CONUS receives interrogations from one satellite and has its reply received by at least seven satellites. The two frequency channels for aircraft-to-satellite transmissions prevent aircraft returns from adjacent sections of CONUS from garbling. By using a range-ordered polling technique only one aircraft in each one-third of CONUS answers at a time.

Position data for surveillance is obtained by differences in time of arrival measurements on the ground, as relayed through the seven satellites. Freedom from garble on the return link is ensured by (1) the range ordered polling, (2) the two reply frequency channels, and (3) satellite spatial selectivity (antenna directivity). Thus, a reliable system of high capacity can be configured.

The ASTRO-DABS navigation capability (ASTRO-NAV) is accomplished by providing navigation signals interlaced with surveillance data. This is accomplished by transmitting precisely timed pulses from the S/N satellites, with difference in time of arrival measurements being made in the aircraft. Since this function is only from satellite-to-aircraft and requires no reply, there is a limited threat from jamming. The principal jamming concern is on the aircraft-to-satellite link, where a moderate power transmitter and a high gain ground antenna can easily dominate the ASTRO-DABS signal entering the aircraft antennas.

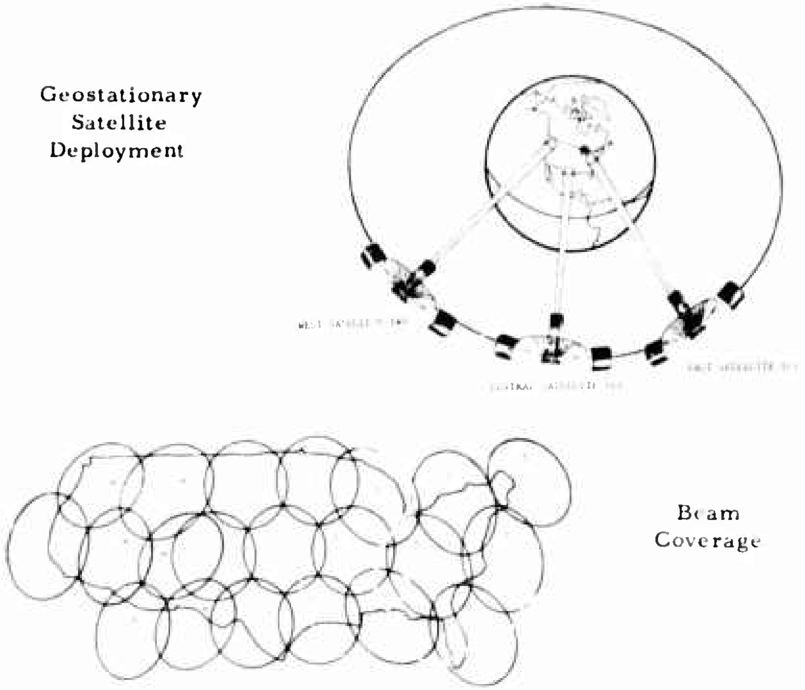


Figure 17 - Geostationary Satellite Deployment and Beam Coverage

Surveillance and Data Link Modes of Operation

Each of three S/D satellites located in a synchronous equatorial orbit has responsibility for 1/3 of CONUS. The transmitting satellites use 30 ft. diameter antennas, one common frequency channel (at L-Band) and have beam agility (seven beam positions). Thus, CONUS is divided into 21 areas of coverage as shown in Figure 17 Satellite transmissions are controlled by the satellite control center by C-Band transmissions from the ground to the satellite.

There are two basic surveillance and data link modes in the ASTRO-DABS system concept: the acquisition mode and the tracking mode. In the acquisition mode an aircraft will be acquired by the ASTRO-DABS system while still on the runway. When the equipment is activated on the ground, the aircraft will receive one acquisition interrogation within the next two seconds and one every two seconds thereafter until acquired. Acquisition mode transmissions are interleaved with tracking mode transmissions. However, the aircraft will be nearly always placed in the tracking mode within a six second period from the time it starts receiving interrogations from the satellite. Acquisition occurs with detection of the correct address on only two of the aircraft-to-SCC links. An indicator light informs the pilot that his aircraft has been acquired.

A normal acquisition mode message set is 1/8 second. Each message set is divided into 1/56 second intervals. The first message of each interval once again tells the aircraft within a satellite beam that this subset of the message is an acquisition mode message. The second message is 84 us long, carries two bits of aircraft acquisition address information and 54 bits of synchronization information. There are no other acquisition transmissions for the remainder of the 4.4 ms period. The reason for the separation of 4.4 ms between acquisition mode addresses is due to uncertainty in the location of aircraft to be acquired. The 4.4 ms dwell time allows aircraft located anywhere in the beam with a given acquisition address to respond before the next set of addressed aircraft respond. The procedure sequences through acquisition mode addresses in a similar manner and then repeats for each of the other seven beam positions.

Once acquired, aircraft entering the tracking mode are interrogated by an L-Band transmitting satellite, one at a time in a sequential manner. Thus, no more than three different aircraft are interrogated at any instant of time in the tracking mode throughout CONUS. Tracking mode interrogations are grouped into approximately 1/8 second message sets. Tracking mode interrogations can be either 47 us or 84 us long and each message set can contain a mix of short and long messages. The complete message set will normally be transmitted in one beam position. The first message of the set is a general synchronizing message (bit, word, mode) to all aircraft within a beam. The remaining messages of the set are individually aircraft-addressed with gaps left between messages equal to the length of the previously transmitted message. A tracking mode message set can contain in the order of 1000 discretely addressed messages. Data and address bits are 1.5 us long with each bit expanded in bandwidth by a four chip (3/8 us per chip) Barker sequence. This spreading code provides improved timing accuracy and a degree (6 db) of multipath protection.

Signal garbling in ASTRO-DABS is eliminated in the tracking mode by the range-ordered polling procedure. The interrogation period (time between interrogations to the same aircraft) can be bounded between 3.5 and 5.5 seconds. The interrogation procedure basically structures a network of hexagonal-cylindrical cells (Figure 18) whose longest dimension d_0 is 14 miles. The system interrogates all aircraft within a given cell in a time-ordered, range-ordered sequence monotonically increasing with respect to the transmitting satellite. An appropriate time gap is inserted and the procedure is repeated in the nearest occupied cell. All cells in a given beam are sequenced through in a similar manner. All beams are sequenced through repeating the above procedure until all aircraft in CONUS have been interrogated.

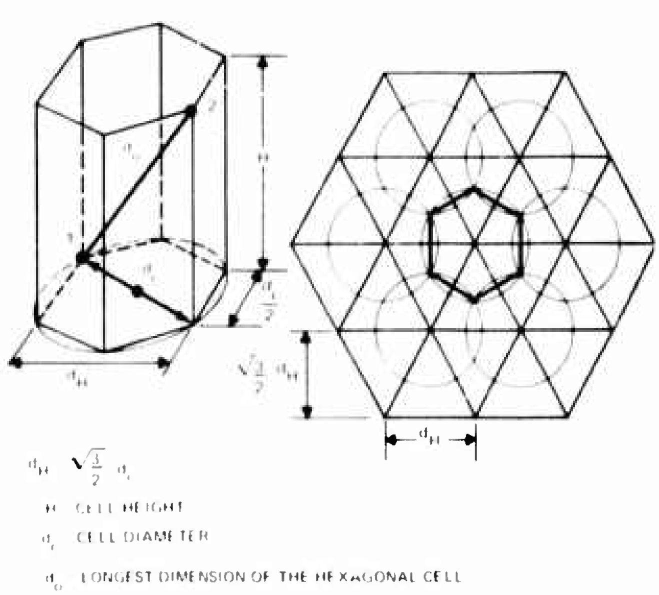


Figure 18 - Spatial Interrogation Geometry for Tracking

Replies from a given aircraft are received by the transmitting S/D satellite and by at least six S/N satellites. Although the beam traces from the seven beams of an S/N satellite in an inclined orbit change shape as the satellite's orbit position changes, it is believed that East Coast aircraft returns can be kept from garbling West Coast returns by beam separation, without resorting to a third up-link frequency. Two returns of adjacent regions such as from East and Central CONUS within a given beam will not garble since such returns are on different frequency channels.

ASTRO-NAV Operation

In ASTRO-DABS the aircraft surveillance operation alternates between an acquisition mode (for aircraft newly entering the system) and a tracking mode (for aircraft already in the system). Acquisition mode transmissions occur every two seconds over a 1/8 second interval that includes considerable time gaps between transmissions. ASTRO-NAV utilizes this gap time for navigation data transmissions plus an additional 1/8 second during which navigation timing pulses are transmitted. The interleaved transmission sequence is shown in Figure 19.

In the first 1/8 second the three S/D geostationary satellites (S1, S2, S3) transmit acquisition and navigation data sequentially over each of their seven beams preceded by appropriate preambles. The navigation data pulses (ND) contain satellite ephemerides, local time delays, altitude corrections, and localized navigation algorithm coefficients that can significantly reduce onboard computations. The next 1/8 second is used to transmit up to 7 timing pulse pairs (TP) throughout CONUS. The navigation preamble pulses (NP) alert aircraft in respective beams to forthcoming navigation data and provide clock references.

Since timing pulse transmissions are for all of CONUS and utilize the same frequency channel as other transmissions, suitable timing gaps must be introduced to preclude garbling in CONUS. Larger gaps are allotted for separating the timing pulses from the inclined satellites to allow for the greater range in propagation times.

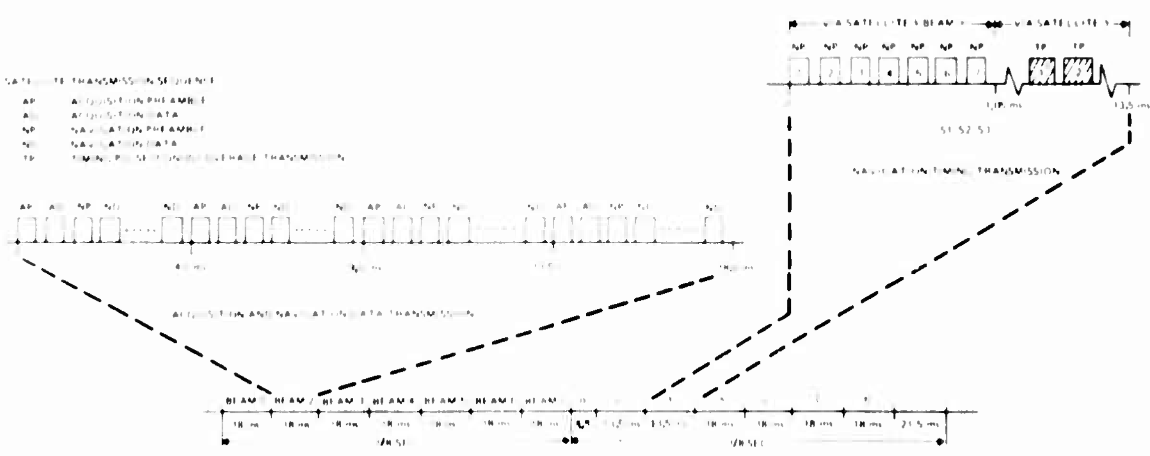


Figure 19 - ASTRO-NAV Timing and Data Transmission Sequence

The two timing pulses transmitted in time sequence are sent over separate frequency channels. The first is transmitted over the normal satellite-to-aircraft channel and is used by both aircraft and calibration stations. The second timing pulse is transmitted by one of the aircraft-to-satellite surveillance channels and is used by the calibration stations. This second transmission will not interfere with aircraft-to-satellite transmissions, since during the navigation mode of ASTRO-DABS, these are not scheduled. The dual frequency transmission is used by the calibration stations to correct for ionospheric propagation delay. Note that this information is used for surveillance as well as navigation.

The ASTRO-NAV receiver determines the relative time of arrival of six of the timing pulses with respect to the seventh received timing pulse. This information together with data received on the data link as to the relative time of transmission of the several timing pulses together with ephemeris and propagation data allows the ASTRO-NAV computer to compute the 3D information of the aircraft with accuracy and at moderate cost.

To minimize multipath effects in ASTRO-NAV, aircraft use only those transmissions from satellites whose local elevation angles are greater than 19°. Also, inclined satellites positioned south of the equator do not transmit timing pulses. The aircraft computer is informed as to which transmissions to use as a part of the navigation data sent by the appropriate geostationary satellite.

ASTRO-DABS Surveillance Backup

The requirement for a surveillance backup exists because of the threat of intentional and unintentional jamming. This threat exists particularly with aircraft-to-satellite transmissions since there is only moderate transmitter power available and essentially no aircraft antenna gain. For a relatively small investment, directive ground-based jammers could effectively jam at least part of the system.

Fundamental to the design of ASTRO-DABS is the structuring of an interrogation range-ordered algorithm which insures that aircraft responses to satellite interrogations do not garble anywhere in space. Since ASTRO-DABS transmissions propagate garble-free in space, one could receive such transmissions on the ground, and, if they contained position and identity information, utilize this data for tracking aircraft. Therefore, to obtain a surveillance backup, it is only necessary that each aircraft transmit identity and positional information such that the ground sites used for surveillance backup can obtain their information.

In ASTRO-DABS, aircraft can obtain positional information via ASTRO-NAV. To utilize this information for a surveillance backup, refer to Figure 20. Each aircraft determines and stores the time of arrivals of up to six timing pulses relative to the arrival of the first timing pulse. Upon reception of the ASTRO-DABS surveillance interrogation, an aircraft follows his usual ASTRO-DABS surveillance response with a second transmission on a separate L-Band frequency which provides his identity together with the set of stored relative arrival times. This second transmission, because it is keyed to the ASTRO-DABS interrogation, propagates through space garble-free and can be

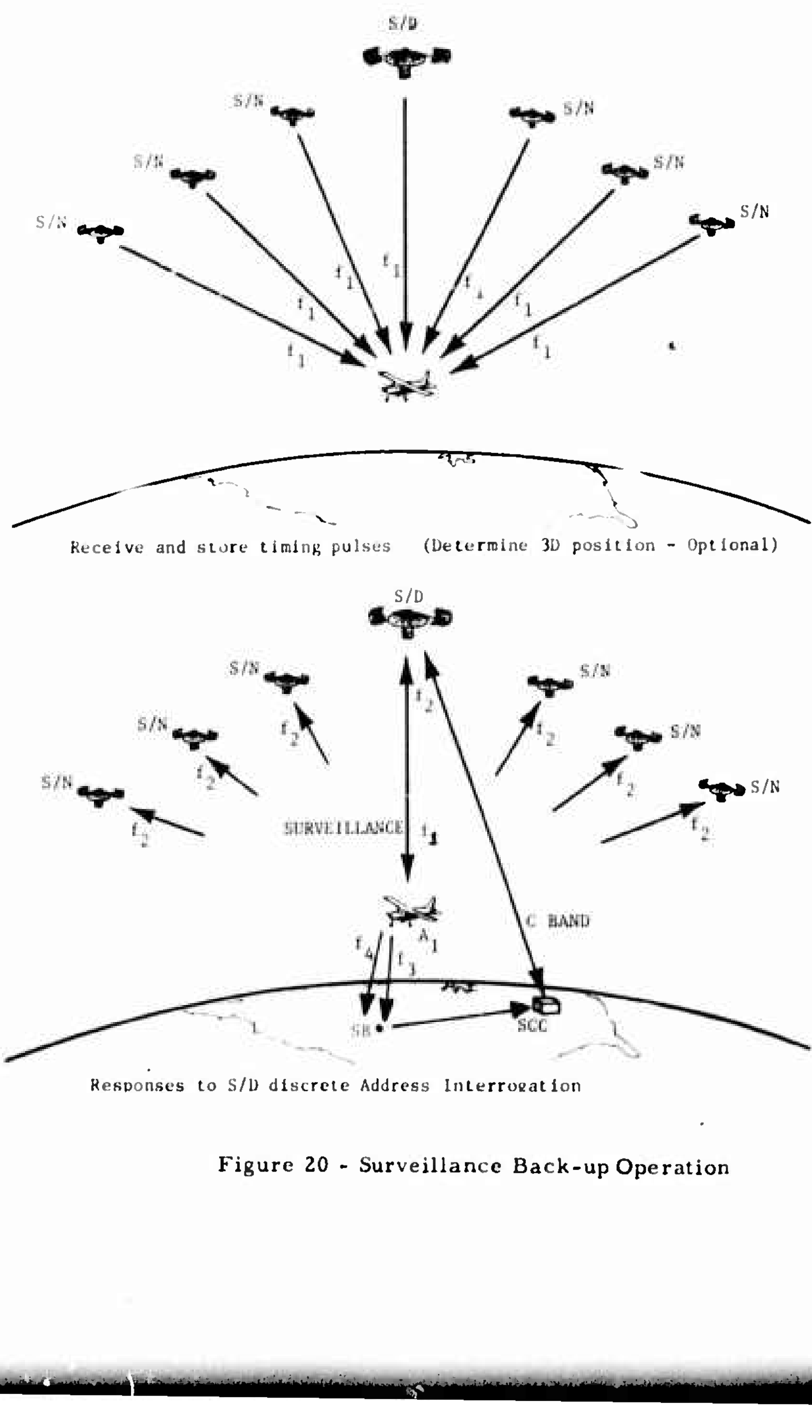


Figure 20 - Surveillance Back-up Operation

received by any equipped ground site. The ground site then relays this information to the satellite center which then determines the target's identity and position. Thus, the satellite control center obtains the three dimensional position and identity of all targets in the vicinity of the surveillance backup ground sites. This backup surveillance information, however, would only be transmitted in cases of emergency (e.g., under conditions of up-link jamming).

In any air-to-ground communication the possibility of aircraft shielding of signals has to be considered. In ASTRO-DABS two antennas are used on the aircraft to reduce the problem of shielding. The aircraft transmits his first surveillance backup transmission as previously discussed via a bottom mounted antenna; a second backup signal, via the top mounted antenna on a separate frequency channel, is transmitted immediately afterwards. Since this second transmission is also keyed to the ASTRO-DABS interrogation and is frequently separated from both the surveillance reply and the first backup transmission, it will also propagate garble free anywhere in space.

Air-To-Air Collision Avoidance System (CAS)

The ASTRO-DABS surveillance backup transmissions propagate garble free everywhere in space and therefore can be monitored by any aircraft to provide three dimensional garble free air-to-air CAS as depicted in Figure 21.

All CAS equipped aircraft would require top and bottom antennas, two receivers and an ASTRO-NAV capability. Once the ASTRO-NAV computer has been initialized, navigation update positional measurements can be made in milli-seconds. A CAS positional determination of a target is expected to result in a simple algorithmic extension of ASTRO-NAV since target positions of interest will be relatively close, i.e., within a few miles.

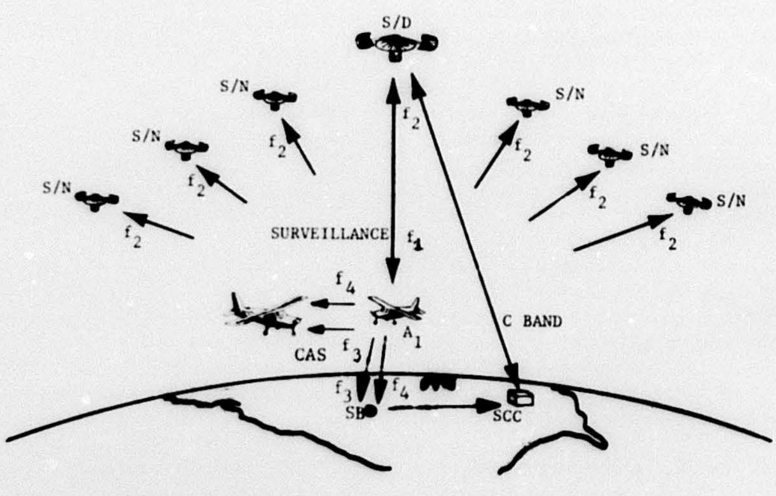


Figure 21 - Air-to-Air Collision Avoidance System Concept

Position accuracy in CAS is a function of receiver delay jitter, ephemeris accuracy, propagation delay, navigation and surveillance data update rates and target position computation delay. A one-sigma three-dimensional navigation measurement error of 150 feet anywhere in CONUS appears to be a realistic design goal. The overall three dimensional relative positional error in the CAS operation should be somewhat less than the navigation error. These accuracies appear consistent with the data requirements necessary for the development of track predictions and avoidance algorithms associated with an effective CAS implementation.

Evolving to ASTRO-DABS

Evolving to ASTRO-DABS could occur progressively in the four states as described in terms of frequency band occupancy as shown in Figure 22. State one illustrates the ATCRBS surveillance system as it exists today. The initiation of DABS service is initiated in state two. The launching of ASTRO-DABS satellites starts the operation of state three and state four is realized when CONUS surveillance is provided only by ASTRO-DABS.

In state three the compatible ATCRBS and DABS coexist with the satellite-relayed ASTRO-DABS which operates on different frequency channels. During this transition state, new avionics need not be purchased until normal replacement is required and the carriage of multiple equipment would not be necessary. To obtain consistent position information for all aircraft, ASTRO-DABS, DABS and ATCRBS have to be placed in a common coordinate system. This is achieved by placing ASTRO-DABS calibration stations at or near DABS and ATCRBS sites and treating each of these sites as aircraft so that they are periodically polled as part of the interrogation roll call. Thus, each DABS site and ATCRBS site is referred to the ASTRO-DABS coordinate system and the positional information of all aircraft in a given region is provided to the regional control facility in a consistent common-grid system.

The satellite control center, colocated with an enroute control center, is linked by satellite at C-band with 18 other enroute centers and five busy terminal centers located throughout CONUS. These centers are in turn linked via land lines to all other regional ATC facilities. The data received via the satellite can be transmitted anywhere in CONUS via an enroute or terminal relay.

All aircraft flying by instrument flight rules (IFR) are under control with a regional controller receiving surveillance information from ATCRBS, DABS and ASTRO-DABS. Aircraft flying by visual flight rules (VFR) in mixed (VFR and IFR) airspace and in regions of a DABS site are under automatic intermittent control by the Intermittent Positive Control (IPC) processor. To perform the automatic VFR control function properly, the IPC processor must know the positions of all aircraft (VFR and IFR) in its region of responsibility. ASTRO-DABS surveillance information would be utilized for all aspects of position data acquisition and would be coordinated with ATCRBS and DABS.

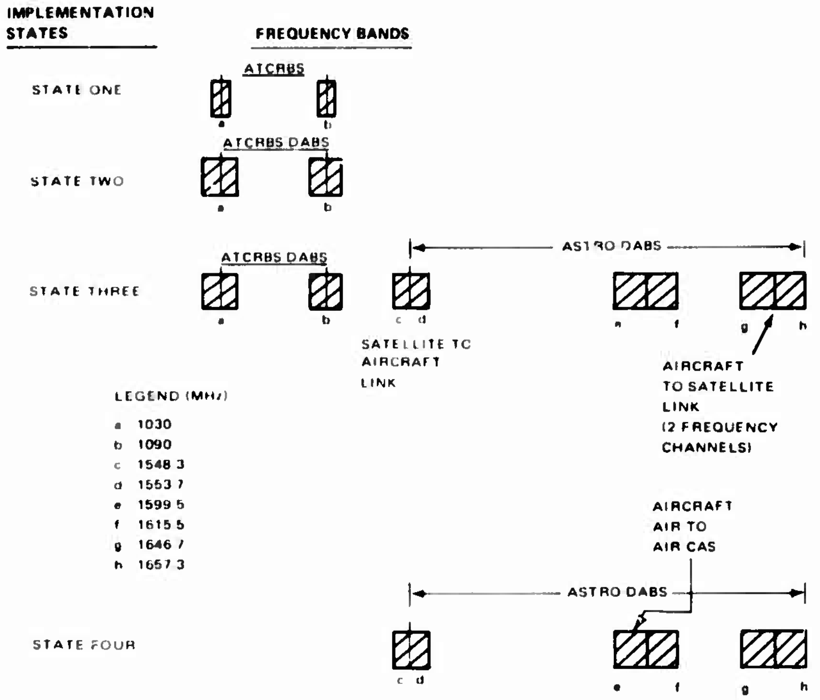


Figure 22 - Spectrum Occupancy Considerations of System Evolution

Aircraft which normally would be surveyed by a particular DABS site, but which are in the ASTRO-DABS system would have their surveillance information transformed to DABS coordinates and transmitted to an ATC facility and the DABS IPC computer. Since the accuracy of the ASTRO-DABS aircraft position data is expected to be greater than that of DABS, the use of ASTRO-DABS surveillance data should not degrade the operation of IPC. Thus, the DABS IPC computer would provide IPC service for all aircraft in its region. The IPC commands to DABS aircraft would be transmitted by the DABS data link while IPC commands to ASTRO-DABS aircraft would be transmitted via the satellite data link.

In states one and two, areas where ATCRBS exist are under control so that there is no data link to aircraft but only voice communication for control messages. During state three such areas will have both ATCRBS and ASTRO-DABS aircraft to control. Again, the improved accuracy of ASTRO-DABS surveillance information received by the controller will not degrade the control function since the transformed absolute accuracy of the ASTRO-DABS aircraft position will be better than the ATCRBS surveillance data.

In state four, the only surveillance system would be ASTRO-DABS. At that time, ASTRO-DABS surveillance services become universal and the IPC and CAS services are extended everywhere. State four will allow full advantage to be taken of the tracking system accuracy.

6. OPERATIONAL BENEFITS AND LIMITATIONS

The operational benefits to be derived from satellite-based systems for ATC are difficult to accurately guage for a variety of reasons but principally because of the lengthy phase-in periods normally associated with ATC system elements. The economic justification for this is not difficult to understand since a new system must coexist with that being replaced for an interval approximating the useful lifetime of the avionics for the old system. It is currently not clear when an operational satellite-based ATC system will be deployed, even though the AEROSAT evaluation program shows significant promise. The FAA and others have accomplished cost benefit analyses for selected satellite-based systems such as AEROSAT, AATMS and ASTRO-DABS, however the results are not definitive, partly due to uncertainties in the assumptions such as avionics costs, phase-in period, technology improvement factors, inflation rate, space segment costs, initial investment risk aspects and performance benefits. The analyses typically indicate a beneficial result for the satellite-based systems and in some cases very significant cost savings can accrue.

The enhanced communications capability and the improved surveillance accuracy gained from satellite systems can result in a system capacity improvement. In the case of the over-ocean application and possibly also for continental applications such as for CONUS, this may constitute a significant benefit. If the improved performance allows increased confidence and use to be made of automated ATC facilities, then an improvement in both system capacity and cost of system operation may be realized.

Certain limitations of satellite-based systems for CONUS ATC should be indicated. First, it is doubtful if the avionics costs associated with any satellite system will be competitive for some time with the low cost VOR/DME equipment currently employed by the general aviation user. The avionics required for satellite service may be considerably more costly than today's ATCRBS transponders because of the normal requirements for high transmitter power, coherent operation, near uniform upper hemispherical coverage aircraft antennas and avionics which limit the mutual interference between aircraft. Thus it does not appear that on a competitive basis, satellite techniques would replace the FAA's VOR/DME ground networks in the foreseeable future.

The susceptibility of a satellite-based ATC system to intentional or unintentional jamming and interference is an area of significant concern. There do not appear to be satisfactory mechanisms, technical or political, for dealing with this threat. It may be unrealistic to expect implementation of an ATC system which could be disrupted by the activities of a unfriendly country or a terrorist group. A number of strategies can be employed to minimize this hazard but it remains a serious problem which is receiving attention.

Satellite systems will prove themselves in the competitive ATC environment, if they provide a viable "best" solution for the requirements of the users. Unfortunately, this is a deceptively difficult task because we find in nearly all areas that:

(1) The user "requirements," needs or deficiencies are seldom clearly or comprehensively established, especially in the context of the costs and priorities involved in satisfying these requirements.

(2) There is no single acceptable criterion for a "best" solution, but rather a large number of interrelated criteria (such as coverage, accuracy, availability, costs, etc.) which must be considered on a comparative basis in evaluating candidate systems.

(3) The uncertainties in determining the actual expenditures associated with the research, development, implementation and operation of current and proposed systems is such that a high level of confidence cannot normally be given to system comparative cost evaluations and projections.

(4) Also in the cost area, it is very difficult, if not impossible, to quantify the benefits which may result from the introduction of a new or significantly improved capability.

(5) In many areas of technology, multiple approaches to particular problems are available and are found to provide sets of alternatives of essentially equal merit or value. This flexibility is desirable but compounds the system selection issue.

(6) Advocacy positions taken by various factions supporting particular alternatives and the competitive parochialism which may develop tends to confuse the decision process.

(7) The differing applications, requirements, coverage characteristics and modes of operation of the various user classes tend to favor special purpose systems, deployed in multiple to provide the desired service. However, advantages relating to cost and other common use features tend to favor a small number of system implementations.

This provides an indication of several of the uncertainties and constraints associated with investigating the technical, operational and economic factors which influence system benefit analyses. Additionally, it points out some of the difficulties in formulating the bases for a valid system selection process.

Possibly one of the greatest liabilities of satellite-based system for ATC is simply the fact that they are relatively new systems. The following brief quotation from Machiavelli's The Prince, written in 1513 may be of interest

> "It must be remembered that there is nothing more difficult to plan, more doubtful of success, more dangerous to manage than the creation of a new system ... for the initiator has the enmity of all who would profit by the preservation of the old institutions and merely luke-warm defenders in those who would gain by the new ones."

7. FUTURE TRENDS AND TECHNICAL CHALLENGE

The widespread interest and activity associated with the development of satellite-based air traffic control and position determination systems indicates that the potential benefits obtainable from these systems in a variety of applications are significant. The demonstrated capabilities of satellite systems such as those discussed earlier, and the experience gained from the large number of civil and military communication satellite systems which have been deployed, leads one to the conclusion that there exists a trend toward the use of space-based systems for many important applications. Since the developments in space systems, aircraft avionics, electron devices, and other areas of technology continue to proceed at a lively pace, it appears that continued exploitation of satellite-aided systems will continue in the future. The full implications of this activity are not clear but many of the functions and services which we have traditionally had fulfilled by ground-based systems may be accomplished in the future by systems configured with satellite-based elements. Further, these systems may accomplish a variety of functions with greater effectiveness and at lower cost.

The introduction of satellite systems for communications, air traffic control, navigation and other services could cause significant changes in the operation of many related large systems. Substantive operational changes of this character do not occur rapidly or without comprehensive pre-operational system demonstration and user community acceptance. Therefore, the introduction and integration of satellite-based elements into the ATC and related system frameworks on a domestic and international basis may be expected to occur in a gradual manner responsive to the combination of needs, technology, costs, benefits, alternatives available and other operative institutional and political factors.

In the ATC role, a principal technical, or techno-economic, challenge is to improve the productivity of the human elements within the system, increase system capacity and concurrently improve safety. This is an extremely difficult task but one of great importance if the projected growth in air traffic occurs and the current trend in labor costs continue. Satellite-aided systems provide certain capabilities which may assist in achieving these objectives. Two of the principal features of satellite-aided ATC systems are their pervasive coverage characteristics and improved position surveillance accuracy. If the higher quality surveillance allows an improvement in the control of air traffic through increased use of automated facilities, then a less labor intensive system may be feasible with attendant long term savings. Additionally, increased capacity may be obtainable and at reduced cost. However, the satellite-aided capability is but one element in the series of improvements which in combination can achieve the desired goal.

The technical challenges associated with the development of satellite-aided ATC systems fall into several areas. These include system technology, component development, fabrication or production, and cost. There remains the need for satellite-aided ATC systems to be configured which meet the current requirements of the users, allow for an evolutionary phase-in, are competitively viable and acceptable to the ATC community. Possibly the greatest technical challenge relates to the development of reliable, low cost user avionics which perform their assigned functions as specified and require low levels of maintenance. There have been impressive advances made by the defense agencies in this area during the past several years and with additional innovations, standardization and improvements in production techniques, continued reductions in the cost of complex avionics are inevitable. We are aware of the reductions in cost that have accompanied large scale production and competition in the small electronic calculator and computer fields. The benefits of large scale integration (LSI) of solid state devices snd similar technology improvements has possibly opened a new frontier for low cost avionics development.

Concerning the satellite-based systems under development or planned, the question naturally arises as to the possible duplication or proliferation in system development which may occur. Although the applications of the various systems are generally divergent, it is apparent that certain advantages can be gained through either the combining of system functions, or by consolidating independent systems into a single set of space platforms. Currently, satellite-based ATC systems are in the system concept, experimental or developmental stages and are somewhat removed from operational deployment. The differing applications, requirements, coverage characteristics and modes of operation of the various system techniques and user classes tend to justify the initial development of special purpose systems to provide the desired services. However, careful consideration should be given to the potential benefits of the operational deployment of a minimum number of common-use systems.-

8. ACKNOWLEDGEMENTS

The information presented is largely based upon published data for the systems of concern. To some extent, unpublished notes, personal communications, and discussions with a number of the principals involved in the development of satellite-based systems have been used in the preparation of this work. Although it is not possible to credit all whose efforts have materially advanced the development of the systems reported upon, it is appropriate to acknowledge the assistance, cooperation and contributions of several individuals and organizations. Particular recognition and appreciation is appropriate for the efforts of: Mr. Robert Maxwell and Captain William Simpson, U.S.N. of the Department of Transportation for their information on Advanced Air Traffic Management System Concepts; the Autonetics and Boeing Companies for material contained in their AATMS investigations, including several figures which have been used; Mr. David Isreal of the Federal Aviation Administration for information and assistance on the Aeronautical Satellite Program; Mr. Harry Fiegelson of the Maritime Satellite Communication/Navigation System; Mr. Leonard Schuchman of the MITRE Corporation for information and illustrations relating to the ASTRO-DABS system concept; Mr. Roger Easton of the Naval Research Laboratory for data on the TIMATION system concept; Dr. Richard Kershner of the Applied Physics Laboratory of Johns Hopkins University for information on TRANSIT and the Two-In-View system concept; Commander Jack Wilson, U.S.N., for information and assistance on the TIMATION, TRANSIT, expanded TRANSIT, and TRANSIT improvement programs; Colonel Jack Price, U.S.A.F., for information on System 621B and Colonel Burden Brentnall, U.S.A.F., for data on the NAVSTAR GPS Program. Responsibility for errors in fact or interpretation are solely those of the author.

BIBLIOGRAPHY

Air Traffic Control Advisory Committee, Final Report, Vols. 1 and 2 (Appendices), B. Alexander, Chairman, Department of Transportation, Washington, D.C., December 1969.

Advanced Air Traffic Management System Study Executive Summary, DOT-TSC-OST-75-1, Department of Transportation, Transportation Systems Center, Cambridge, Massachusetts, January 1975.

Advanced Air Traffic Management System - Final Report, C72-1206/201. Rockwell International Corporation, Anaheim, California, April 1973.

Advanced Air Traffic Management System Study Overview, Department of Transportation, Transportation Systems Center, Cambridge, Massachusetts, Estimated publication date: June 1975.

Advanced Air Traffic Management System Study Technical Summary, Department of Transportation, Transportation Systems Center, Cambridge, Massachusetts, Estimated publication date: March 1975.

Advanced Air Traffic Management - System B Summary Report, FAA-EM-73-10, MTR-6419 Series 1, MITRE Corporation, McLean, Virginia, June 1973.

An Advanced Air Traffic Management Concept Based on Extensions of the Upgraded Third Generation ATC System, Report FAA-EM-73-10A, MITRE Corporation, Department of Transportation, Federal Aviation Administration, Washington, D.C., March 1974.

Analysis of an Aeronautical Satellite Evaluation and Development System, Report No. FAA-RD-73-2577, Aerospace Corporation for DOT/FAA, Contract No. DOT-FA71WA-2577, January 1973.

Analysis of Alternative Oceanic Air Traffic Control Systems, Contract No. FAA-DOT-71-WA 2493, University of California, Berkeley, California, DOT/FAA, November 1971.

Application of Space Techniques Relating to Aviation Panel, ASTRA III Meeting, ICAO Document 8873, March 1970.

Application of Space Techniques Relating to Aviation Panel, ASTRA IV Meeting, ICAO Document 8929, January 22, 1971.

Avant, A. L., ASTRO-DABS Startup and Acquisition Performance, MTR-6475, MITRE Corporation, McLean, Virginia, July 31, 1973.

Beusch, J. U., et al., Final Report, Concept Formulation Studies of the Control Aspects of the Fourth Generation Air Traffic Control System, DOT/TSC-140, ATC-6, Lincoln Laboratory, Massachusetts Institute of Technology, Lexington, Massachusetts, September 15, 1971.

Beusch, J. U., et al. Interim Report, Concept Formulation Studies of the Control Aspects of the Fourth Generation Air Traffic Control System, DOT/TSC-140, ATC-2, Lincoln Laboratory, Massachusetts Institute of Technology, Lexington, Massachusetts, July 2, 1971.

Brentnall, Col. B., USAF, Status Report on DOD Navigation Satellites, Office of the Director of Defense Research and Engineering, Presented to the Radio Technical Commission for Aeronautics, Washington, D.C., November 21, 1974.

Concept for a Satellite-Based Advanced Air Traffic Management System (Vols. 1-10), Report No. DOT/TSC-OST-73-29, Rockwell International Corporation, Autonetics Division, Anaheim, California, 1974.

Diamond, P. D., The Potential of a System of Satellites as a Part of an Air Traffic Control System, AGARD Conference Proceedings No. 105, June 1972.

Ebert, Dr. P. M., A Geometric Interpretation of Geometric-Dilution-of-Precision (GDOP), MTR-6477, MITRE Corporation, McLean, Virginia, October 24, 1973.

Ebert, Dr. P. M., ASTRO-DABS Avionics Transponder Design, MTR-6476, MITRE Corporation, McLean, Virginia, August 24, 1973.

Engineering and Development Program Plan - Aeronautical Satellites (AEROSAT), FAA-ED-17-2, U.S. Department of Transportation, Federal Aviation Administration, Washington, D.C., September 1974.

Ekstrom, J. L., Evaluation of ASTRO-DABS Performance While An Aircraft Turns, MTR-6393, MITRE Corporation, McLean, Virginia, April 27, 1973.

Elrod, B. D., Aircraft Interrogation Scheduling with ASTRO-DABS, MTR-6368, MITRE Corporation, McLean, Virginia, March 30, 1973.

Elrod, B. D., and Fee, J. J., Design Considerations for ASTRO-DABS Receiver Satellites, MTR-6453, MITRE Corporation, McLean, Virginia, July 20, 1973.

Elrod, B. D., and Fee, J. J., Preliminary Analysis of the ASTRO-NAV Concept, MTR-6478, MITRE Corporation, McLean, Virginia, October 3, 1973.

Fee, J. J., A Receiver Satellite Constellation for ASTRO-DABS, MTR-6346, MITRE Corporation, McLean, Virginia, March 21, 1973.

Fee, J. J., Techniques for Reducing Ionospheric Propagation Delay Error in Surveillance and Navigation Systems, MTR-6637, MITRE Corporation, McLean, Virginia, April 1974.

Fourth Generation Air Traffic Control Study, Vols. 1-4, DOT-TSC-304-1, North American Rockwell Corporation, Autonetics Division, Anaheim, California, June 1972.

Frankel, F., System Concept Studies for Future Air Traffic Management, Department of Transportation, Transportation Systems Center, Cambridge, Massachusetts, Estimated date of publication: March 1975.

Goldfein, David, Target Acquisition of a Satellite-Based Surveillance System, DOT-TSC-OST-73-27, Department of Transportation, Transportation Systems Center, Cambridge, Massachusetts, Estimated publication date: January 1975.

Horn, R. A., Some DOD Programs Related to Civil Aviation Technology Needs, MTR-2871, MITRE Corporation, Bedford, Massachusetts, August 8, 1974.

Jenney, L. L., and Lawrence, K. A., Implications of Automation for Staffing an Advanced Air Traffic Management System, DOT-TSC-OST-74-32, The Planar Corporation, Alexandria, Virginia, June 1974.

Kirkwood, T. F., et al., Some User Benefits Achievable From An Advanced Air Traffic Management System, R-1320-DOT, The Rand Corporation, Santa Monica, California, July 1973.

Lee, H. B., Accuracy Limitations of Hyperbolic Multilateration Systems, DOT/TSC-241-3, TN-1973-11, Lincoln Laboratory, Massachusetts Institute of Technology, Lexington, Massachusetts, March 22, 1973.

Lee, H. B., Accuracy Limitations of Range-Range (Spherical) Multilateration Systems, DOT-TSC-RA-3-8-1, TN-1973-43, Lincoln Laboratory, Massachusetts Institute of Technology, Lexington, Massachusetts, October 11, 1973.

Lee, H. B., and Goode, B. B., Technical Assessment of Satellites for CONUS Air Traffic Control, Vol. III: Satellite to Aircraft Techniques, Project Report ATC-26, Vol. III, Lincoln Laboratory, M.I.T. (February 1974), DOT/TSC-RA-3-8-3.

Lee, H. B., and Wade, A. E., Improved Satellite Constellations for CONUS-ATC Coverage, DOT/TSC-RA-3-8, ATC-23, Lincoln Laboratory, Massachusetts Institute of Technology, Lexington, Massachusetts, May 1, 1974.

McDonald, K. D., A Survey of Satellite-Based Systems for Navigation, Position Surveillance, Traffic Control, and Collision Avoidance, Journal of the Institute of Navigation, Vol. 20, No. 4, pp. 301-320, (Winter 1973-4).

Mertes, F., and Jenney, L., Automation Applications in an Advanced Air Traffic Management System, DOT-TSC-OST-74-14.1, Vols. 1-10, T.R.W. Incorporated, McLean, Virginia, August 1974.

Orr, R. S., and Schneider, K. S., Technical Assessment of Satellites for CONUS Air Traffic Control, Vol. I: Coordinated Aircraft-to-Satellite Techniques, Project Report ATC-26, Vol. I, DOT/TSC-241-4, Lincoln Laboratory, M.I.T., January 31, 1974.

Orr, R. S., and Yates, R. D., On the Estimation of the Arrival Time of Pulse Signals in Gaussian Noise, DOT-TSC-RA-3-8, TN-1974-13, Lincoln Laboratory, Massachusetts Institute of Technology, Lexington, Massachusetts, August 26, 1974.

Reck, Robert H., and Hagopian, J. B., Air Traffic Demand Estimates for 1995, DOT-TSC-OST, 74-18, Department of Transportation, Transportation Systems Center, Cambridge, Massachusetts, December 1974.

Report of the Seventh Air Navigation Conference, ICAO Document 9004, AN-CONF/7, April 1972.

Otten, D. D., Study of a Navigation and Traffic Control Technique Employing Satellites, Vols. I-IV, Group Report No. 08710-6012-12000, TRW Systems, Redondo Beach, California, December 1967.

Scardina, Dr. John, Description and Cost of a Satellite-Based Oceanic ATC System, M73-75, MITRE Corporation, McLean, Virginia, August 1973.

Schneider, K. S., and Orr, R. S., Technical Assessment of Satellites for CONUS Air Traffic Control, Vol. II: Random Access Aircraft-to-Satellite Techniques, Project Report ATC-26, Vol. II, DOT/TSC-RA-3-8-2, Lincoln Laboratory, M.I.T., February 1974.

Schneider, K. S., Stiglitz, I. G., and Eckberg, A. E., Surveillance Aspects of the Advanced Air Traffic Management System, Project Report ATC-10, DOT/TSC-241-2, Lincoln Laboratory, M.I.T., June 22, 1972.

Schneider, K. S., et al., Surveillance Aspects of the Advanced Air Traffic Management System, DOT-TSC-241-2, ATC-10, Lincoln Laboratory, Massachusetts Institute of Technology, Lexington, Massachusetts, June 22, 1972.

Schuchman, L., ASTRO-DABS: An ATC Satellite Relayed Surveillance Navigation and Data Link Concept, MTR-6675, MITRE Corporation, McLean, Virginia, May 1974.

Schuchman, L., The ASTRO-DABS Concept, MTR-6287, MITRE Corporation, McLean, Virginia, January 19, 1973.

Schuchman, L., ASTRO-DABS Updated-1974, MTR-6770, MITRE Corporation, McLean, Virginia, October 1974.

Stiglitz, I. G., et al., Concept Formulation Studies of the Surveillance Aspects of the Fourth Generation Air Traffic Control System, DOT/TSC-241-ATC-7, Lincoln Laboratory, Massachusetts Institute of Technology, Lexington, Massachusetts, September 21, 1971.

Stiglitz, I. G., Technical Assessment of Satellites for CONUS Air Traffic Control, (4 Vols.), DOT/TSC-RA-3-5, ATC-26, Lincoln Laboratory, Massachusetts Institute of Technology, Lexington, Massachusetts, January 1974.

Strategic Control Algorithm Development, DOT-TSC-OST-74-3, (6 Vols.), Boeing Commercial Airplane Company, Seattle, Washington, December 1973.

Study and Concept Formulation of a Fourth Generation Air Traffic Control System, Vols. I-V, DOT/TSC-506-1, Boeing Company, Commercial Airplane Group, Renton, Washington, April 1972.

Weinberg, A., ASTRO-DABS Acquisition and Startup With Pulse Amplitude Modulation, MTR-6775, MITRE Corporation, McLean, Virginia, October 1974.

Pages 698, 699 and 700 are Blank.

EUROCONTROL DATA PROCESSING SYSTEMS

by

B.Kirstetter, G.Maignan, V.Vachiery

European Organization for the Safety of Air Navigation
EUROCONTROL
72, rue de la Loi
Bruxelles
1040
Belgique

TABLE OF CONTENTS

EUROCONTROL DATA PROCESSING SYSTEMS

by

B. Kirstetter, G. Maignan, V. Vachiéry

European Organisation for the Safety of Air Navigation
EUROCONTROL
72, rue de la Loi
Bruxelles
1040
Belgique

Summary

MADAP, KARLDAP and SHANDAP are Air Traffic Control data processing systems which are implemented in the Upper Airspace Centres of Maastricht, Karlsruhe and Shannon by the EUROCONTROL Agency. MADAP, operational since October 1973, and KARLDAP have a similar operational and technical concept and are designed for combined flight data and radar data processing as from the initial phase of operation. They are both large multicomputer systems with high computing and storage capacity. The size of the real-time software is of the order of some hundred thousand instructions. This paper gives a description of the functions and of the hard- and software structure of the two systems with special consideration of reliability, programming and implementation aspects.

SHANDAP is initially a radar data processing system to be followed by a flight data processing system which is being developed.

1. Introduction

The EUROCONTROL Agency is responsible for the implementation of various projects. Amongst these are the Data Processing Systems of the Maastricht, Karlsruhe and Shannon Upper Airspace Centres. The operation of each of the three above Centres will heavily depend upon rather advanced Data Processing Systems.

Electronic Data Processing is here essentially used for the automatic reception and handling of Flight Plan and Radar Data and for the presentation of the processed information. The three systems are known under the acronyms of :

- MADAP : Maastricht Automatic Data Processing and Display System
- KARLDAP : Karlsruhe Automatic Data Processing and Display System
- SHANDAP : Shannon Automatic Data Processing and Display System.

For operational, technical and historical reasons, MADAP and KARLDAP are quite similar. The Shannon problem is different and so is the SHANDAP solution.

2. MADAP and KARLDAP

2.1. General Aspects of Maastricht and Karlsruhe Centres.

The Maastricht Upper Airspace Centre is responsible for providing Air Traffic Control service in the Upper Airspace above Belgium, Netherlands, Luxembourg and Northern Germany. The area of responsibility of the Karlsruhe Upper Centre is the Upper Airspace above Southern Germany.
The operational objective is to achieve Air Traffic Control services for the whole area, that is to say for all flights, irrespective of their flight conditions. Both organisations (Maastricht and Karlsruhe) also provide for tight coordination with Air Defence.

This is achieved in a different way for each Centre :

- In the area of responsibility of the Maastricht Centre, military operations above BENELUX are controled by separate military ATC facilities. Only as far as Nothern Germany is concerned, these facilities are located in the Maastricht Centre itself.
- The concept of Air Traffic Control operations for Southern Germany (Karlsruhe Centre) is based upon the integration of the civil and military control of Upper Airspace traffic. The Karlsruhe Centre is organized in such a way that Civil and Military operations are controlled from the same control sectors.

The basic choice for the systems and subsystems for the Maastricht Centre was made in 1968, that is about three years earlier than those for the Karlsruhe Centre. The progress in technology plus the variations in the operational concepts explain the differences between the two projects, and in particular between the two main Data Processing Systems. As will be seen later, these differences are rather marginal from a technical point of view.
MADAP has been introduced into operation on the 1st October 1973. KARLDAP is foreseen to become operational by the beginning of 1976.

2.2. Basic Features of MADAP and KARLDAP.

2.2.1. General

MADAP and KARLDAP are characterized by a high degree of automation in the processing of flight plan and radar data.
Complete primary and secondary radar coverage is provided by remote radar stations from which radar data are transmitted in digital form via telephone lines. The air traffic situation is presented to the controllers on a purely synthetic display system.

2.2.2. Functions

The main functions of MADAP and KARLDAP are as follows :

- Flight Plan Processing
  - Reception and verification of flight data messages from a local permanent storage, or from a local manual input and neighbouring Air Traffic Service units.
  - Extraction of data relevant to flight prediction (route).
  - Flight plan activation and navigation (i.e. periodical computation of flight plan position).
  - Processing of Estim[illegible] of Arrival over reporting points.
  - Preparation of Flight [illegible] a for display to controllers either in the form of a printed flight prog[illegible]s strips or in the form of messages on an Electronic Data Display.
- Radar Data Processing
  - Reception and verification of radar information arriving in digital form over land lines from a number of remote radar stations.
  - Automatic tracking on secondary radar data. Initiation of tracks is automatic on SSR selected codes and manual on primary radar data.
  - Multiradar tracking : there may be as many tracks as there are radars seeing a particular aircraft. These tracks are combined in a multi-radar tracking logic to display a unique system track.
- Correlation of flight plan and radar data.
  - Association of the aircraft call-sign to its secondary radar code.
  - Measurement of track to flight plan position divergence and updating of flight plan data using track data.
- Display and input of data at operators positions
  - Storage and processing of information for display in various forms, and the routing of this information to the appropriate operator position at the correct time.
  - Provision of facilities for direct communication between operators and the processing system by means of manual input messages.
- Additional functions
  - Such as presentation of weather data, direct digital exchange of information with neighbouring Air Traffic Control Centres, recording of data for statistical, legal and reconfiguration purposes.

2.2.3. MADAP and KARLDAP major subsystems.

From a data processing point of view, the operational functions described above may be grouped into three categories :

I The handling of transmission and reception of digital data to and from neighbouring Air Traffic Control Units

II The processing, assembling, correlation and preparation for display of data of various nature.
This category may be subdivided into two parts :
- The general system functions e.g. : - Flight data processing
  - Radar data processing
  - Correlation of radar/flight plan data.

  Such general functions aim at the building of a general data bank describing the air traffic situation in the area considered.

- Specific functions which are "operator working position orientated", such as : - Filtering of data for display purposes
  - Interpretation of controller input orders
  - Updating and renewal of displayed data.

III The presentation of the information to operators and the introduction of data into the system from operators positions.

Taking the above into account, MADAP and KARLDAP consist logically of three major subsystems :

- a data communication and terminal system (D.C.T.S.) for the exchange of data with external units,
- a central computer complex (C.C.C.) subdivided into :
  - a main computer complex (MCC) handling the general functions of the system,
  - a peripheral computer complex (PCC) handling operators oriented functions.
- an operator display system including the input devices forming the man/machine interface.

2.2.4. Reliability aspects.

Maximum availability is of course one of the main design objectives of ATC data processing systems. This is even more true for MADAP and KARLDAP which have no conventional manual fall back possibilities.
The high reliability of MADAP and KARLDAP results from a combination of several features such as : - unit reliability
- hardware redundancy
- software design capable of handling reconfigurations
- possibility of operation in a down-graded mode.

2.2.5. Test and improvement facilities.

The complexity of the systems and the dynamic aspect of operational requirements necessitate that adjustments and evolution must be feasible without putting operation in danger. This leads to a strong requirement for facilities allowing the implementation of new functions or packages in a real environment. This must be done without interference to the proper operation of the system. As will be seen later, such facilities are introduced in MADAP and KARLDAP, by striking a balance between the amount of redundancy to be implemented and the complexity of switching over from one configuration to the other.

2.2.6. Programming.

The on-line programmes, used for the real-time processing of operational data, are run under real time supervisors. The organisation of the programmes is characterized by the following aspects :- modular structure of programs
- central data banks
- reconfiguration requirements.

- Real time supervisors

  The use of standard Operating Systems could not be envisaged for MADAP and KARLDAP real-time operation. It was therefore necessary to develop special real-time supervisors : RTSX and MARTOS.

- Modular structure of programmes.

In view of the complexity and volume of programmes required (250.000 instructions about), modular programming has been used. The operational and system functions have been divided into modules. Each module performs an elementary function or group of functions, such that the size of the corresponding group of routines is as far as possible of the order of 1.000 to 4.000 instructions. The modules can communicate only via the supervisor program. They have also access to a common data pool.
Each module contains the following :
- local data
- module control brick which is in charge of the interface with the supervisor
- processing bricks.

- Data banks.

The data banks contain information common to several modules. There are essentially two categories of data banks :

- the static data bank containing the environment description, i.e. the data which are not subject to real time changes such as sector boundaries, airports, reporting points, air-routes, aircraft characteristics.
- the dynamic data bank containing the real-time data describing the real-time air situation.

The dynamic data bank is core store resident.

- Reconfiguration aspects.

The necessity to reconfigure devices and start, stop and restart system with a minimum of perturbation to the operation has lead to a particular organisation of the system data base and to the implementation of special modules.

3. The MADAP System.

3.1. Size of the system.

The size of the MADAP system derives from the following initial design parameters :

Flight plan processing capacity	:	200 active flight plans
Radar data processing capacity	:	250 tracks
Radar inputs	:	up to 6 primary-secondary radar stations (average number 110 plots/second)
Controller operating positions	:	46

As will be described later, the actual implementation plan has lead to an organisation different from the one contemplated at the design stage.

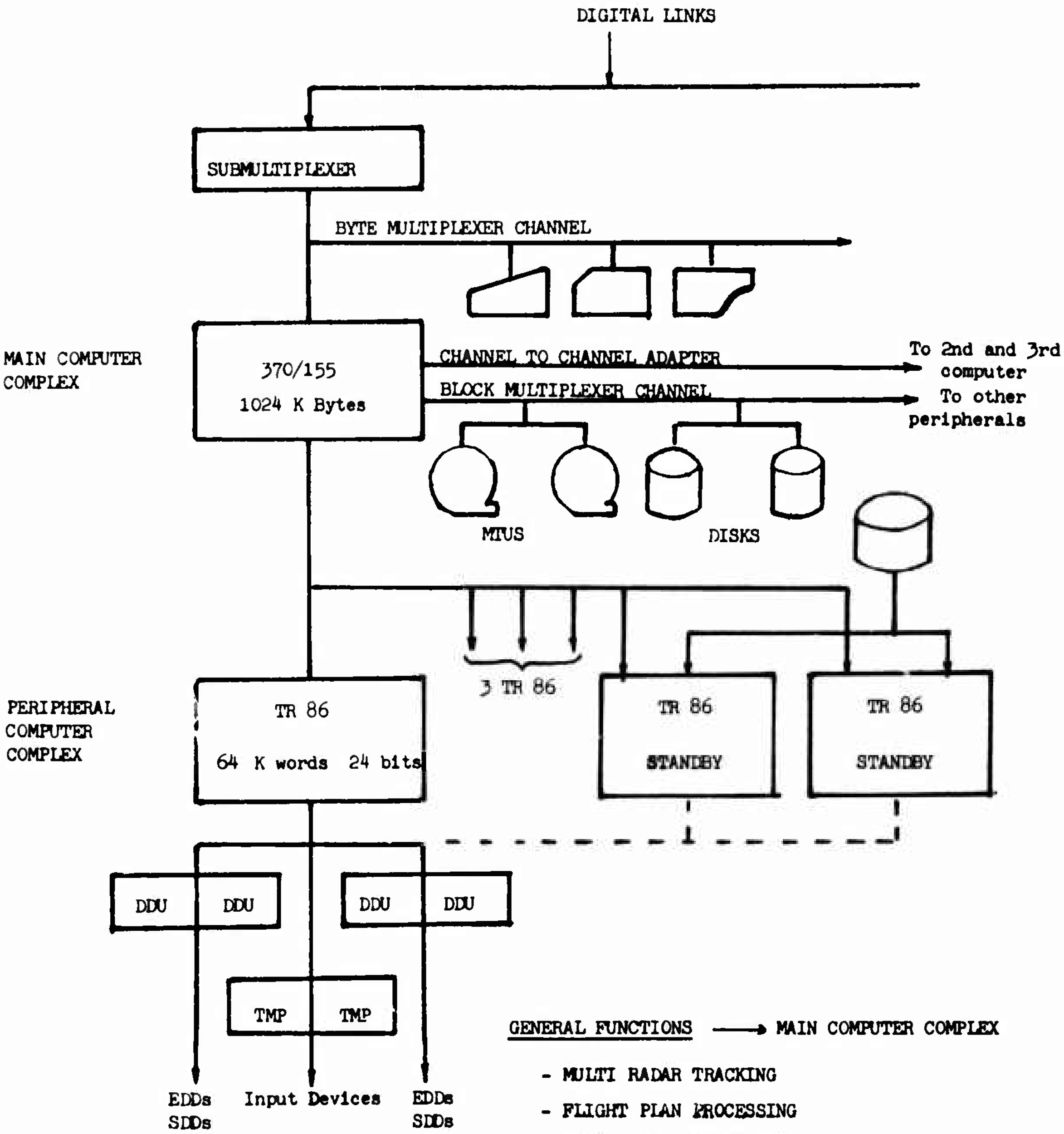
3.2. The Central Computer Complex.

The Central Computer Complex comprises :
- the Main Computer Complex
- the Peripheral Computer Complex.

3.2.1. The Main Computer Complex (MCC) is a triplicated IBM 370/155 system. The input of radar data (plots) to the system is executed by a special device : the sub-multiplexer, which ensures the interface between the radar lines and the Main Computers. Each IBM 370/155 has a core storage of 1024 K Bytes. Four disc control units each with three 2314 drives of 29 M bytes capacity can be connected via two channel switches and 2914 switches to any of the three main computers. Two magnetic tape units driving each three tapes can similarly be connected to any of the three main computers. Only one computer executes the on-line programmes. The second is then in a stand-by state whilst the other may be used for off-line jobs.
In the present situation, the on-line system makes use of two disc control units (2 x 3 discs) and one tape control unit (3 tapes).
The Main Computer Complex interfaces with the Peripheral Computer Complex via parallel data adapter IBM 2701.

3.2.2. The Peripheral Computer Complex comprises six AEG-Telefunken TR 86 computers. Four out of six are for on-line use, the two remaining being used as stand-by. Each TR 86 has a core sorage of 192 K Bytes (64 words of 24 bits).

MADAP SIMPLIFIED DIAGRAM



GENERAL FUNCTIONS ⟶ MAIN COMPUTER COMPLEX

- MULTI RADAR TRACKING
- FLIGHT PLAN PROCESSING
- RADAR/PLN CORRELATION

SECTOR FUNCTIONS ⟶ PERIPHERAL COMPUTERS

- DISPLAY FILTERING
- INPUT ORDER HANDLING
- UPDATING AND PICTURE RENEWAL

The main tasks of the peripheral computers are :
- formatting and display of data coming from the Main Computer
- interpretation of inputs from the controller working position

The core storage of the TR 86 is used as the picture repetition store of the displays.
Each TR 86 is connected to a magnetic tape unit.
Two disc units, each of 2 M bytes can be accessed by either of the two stand-by computers.

3.3. The Operator Input and Display System (ODS).

The ODS is subdivided into two parts : - the display sub-system
- the operator input sub-system and teleprinters.

3.3.1. Display sub-system.

There are two types of displays : the Synthetic Data Display (SDDs) and the Electronic Data Displays (EDDs). Both types of displays are driven by common equipments, the Display Drive Units (DDUs).

- SDDs are circular plan position indicators used to present the controller with the following data : - radar track positions and speed vectors
  - artificial afterglows
  - associated labels (call-sign, altitude, attitude)
  - radar plots
  - maps of air routes, sector boundaries, reporting points etc...)
- EDDs are rectangular displays used to present various flight data information (ETAs, flight levels, call-sign etc...)
- DDUs share the load of driving a certain number of displays of both type.

Apart from brilliance and brightness control, all the display functions are performed under computer programme control.

3.3.2. Operator input sub-system and teleprinters.

The following input/output devices are at the disposal of the controller :

- Display Control panel for selection of the types of data to be displayed on SDDs and for scale and centre settings.
- Rolling Ball used to drive a position designation marker on the displays.
- Touch Input Device consisting of an Electronic Data Display covered by a transparent overlay where touch wires are inbedded. This is used as a instantaneously programmable keyboard.
- Printers including strip-printers and hard copy printers
- Keyboards allowing the input of alphanumeric data for entering the flight plan.

3.4. Other sub-systems

3.4.1. The Digital Communication Terminal System (DCTS)

The DCTS is an autonomous front-end processor used for the exchange of ATS messages between MADAP and neighbouring Centres.
The DCTS is a dual computer complex comprising two CII Mitra 15 computers each with its own discs and tapes.
The two DCTS computers work in parallel. The inputs from the external lines are processed in parallel and are buffered on disk until completely retransmitted.

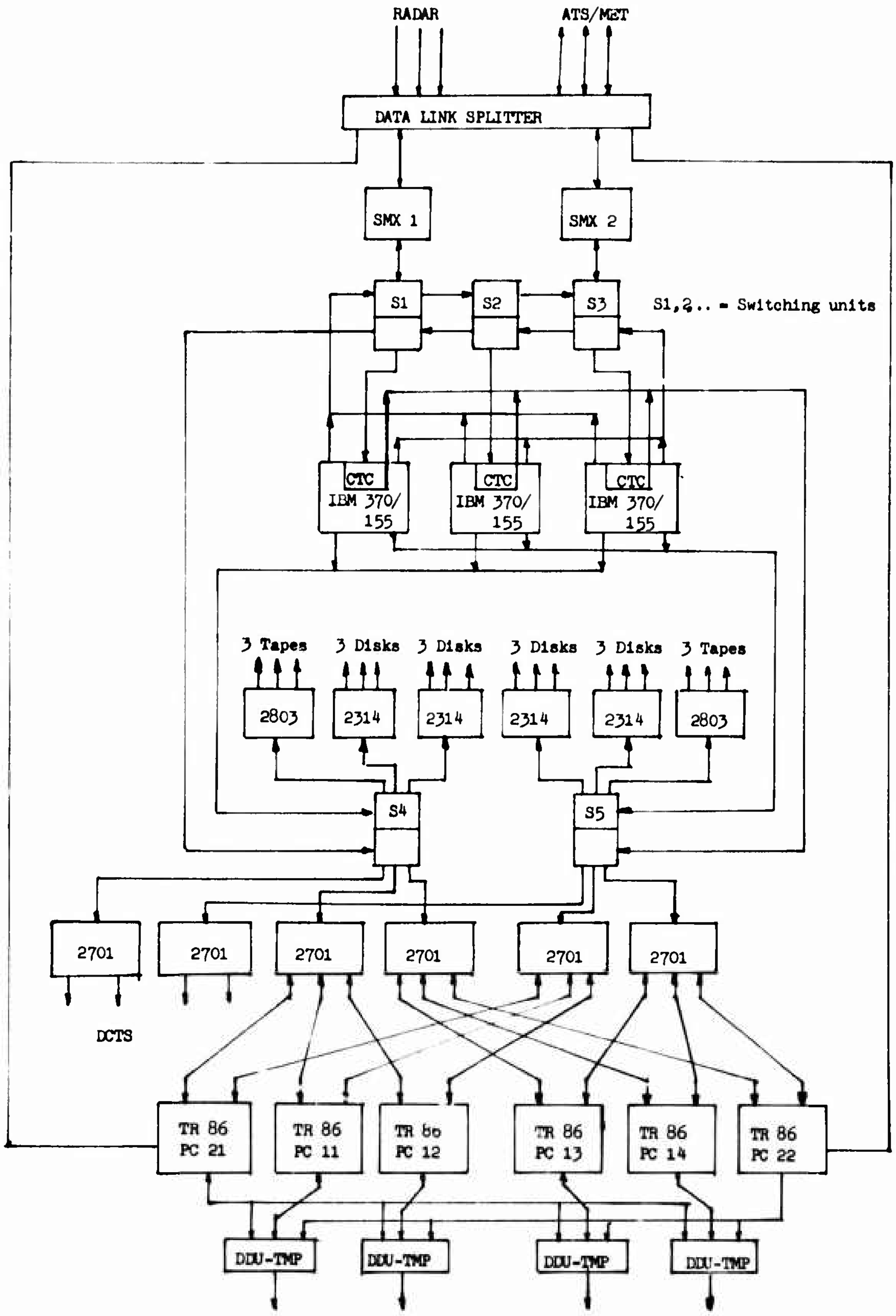
3.4.2. The Technical Monitoring and Control System (TMCS)

The functions of the TMCS are the following :

- centralize the monitoring and control of MADAP and ancillary systems (air-conditioning, power supply etc...).
- eventually, initiate the automatic switching and reconfiguration of the system.

These functions are performed under control of a dual processor system comprising two Telemecanique T 1600 computers.

MADAP CONFIGURATION



to SDDs, EDDs, I/O Units.

3.5. Modes of operation of the MADAP System

3.5.1. Normal state

The system is in a normal state as long as a "minimum string" of equipment is available for the system to perform the totality of its operational functions to the required capacity.
This minimum string is basically composed as follows :
- one sub-multiplexer (out of two),
- one main computer (out of three) with one disc control unit with discs and one magnetic tape control unit with three tape units,
- the number of peripheral computers (with their interface) which is necessary to drive the whole display system (two in the present state, three for the full extension of MADAP).

The hardware and software presently available permits that the system can be operated as follows :

- A complete operational MC-PC chain drives the operational system. Within this operational chain, it is necessary to perform the required peripheral unit switching in case of malfunctioning (SMX, disc, tapes etc...). Those actions are in principle automatic and software driven.
- A standby MC is in a waiting state. If necessary it will be switched into the new hardware configuration.
- Two standby PCs are ready to take over automatically the role of any operational PC.
- The third MC can be connected to any spare PC in order to constitute a complete test and development chain. This is of great importance for thorough testing of the new programmes.
- Alternatively, the third MC with its peripherals can be used as a computing centre allowing batch processing, time sharing operation etc.... The spare PCs can also be used as autonomous computer.

3.5.2. Reduced state

In order to provide, on one hand, an ultimate fall-back in case of total break down of the Main Computer Complex and on the other hand a possibility to bridge the gap between switch-over periods, a special software package has been developed. It is known as REP (Radar Emergency Processor) and constitutes a replacement for the Radar Data Processing and Display functions of the normal state.
This package is run in the Peripheral Computer Complex only. In such a mode of operation, radar data from the plot extractors are fed into one of the standby peripheral computers which takes over the reduced system functions.

These functions consist basically in performing mosaic filtering and radar to call-sign correlation in order to allow maintaining of a labelled radar picture to the controllers.
The REP is activated whenever a MC failure is detected. It takes then approximately five seconds before REP information is displayed on controller screens.
During that time, the restart of the MCC is attempted. When the MCC is restarted, an operator command will make REP inactive and within 8 seconds of that signal, the Peripheral Computer will be in the position to display again MC data.

3.6. The MADAP Software

3.6.1. Main Computer Operational Programme.

- Categories of programmes

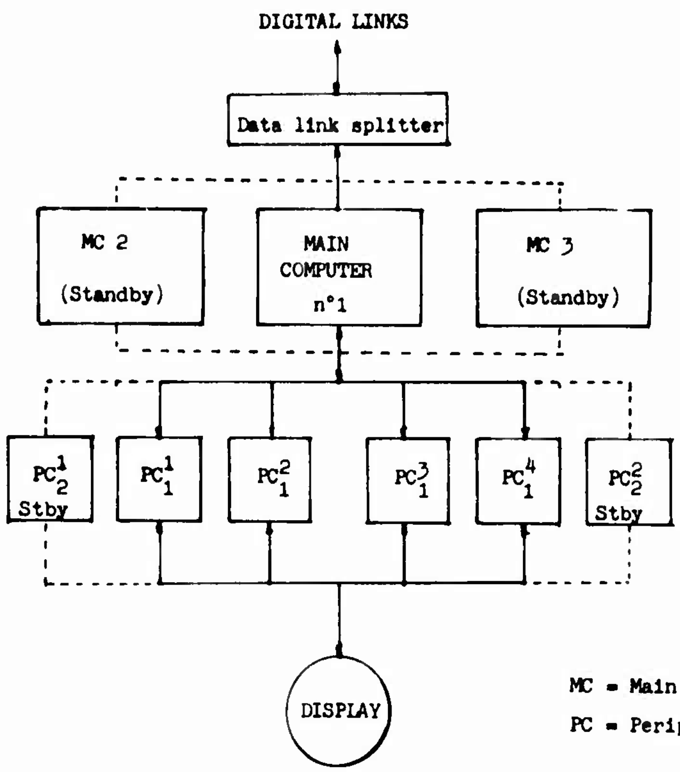
The MC operational programmes comprise :
- the MC on-line supervisor
- the System programmes
- the Application programmes.

MC on-line supervisor

The MC on-line supervisor , RTSX (real-time system executive) is responsible for controlling the activity of the machine. It is divided into five sections as detailed below :

MODES OF OPERATION OF MADAP

NORMAL STATE

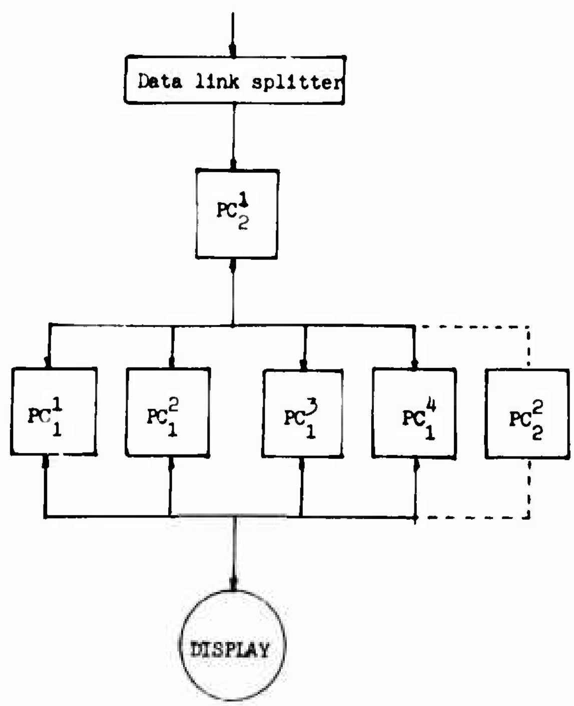


MC = Main Computer

PC = Peripheral Computer

REDUCED STATE

RADAR EMERGENCY PROCESSING



- Program Control

  All programmes are arranged in a set of loop submitted to time constraints. The Program Control section is responsible for selection and initiation of programmes and makes sure that all the different programs are driven in order to respond to their time constraints. All the loops are arranged in a level of priority, and the algorithm of selection of a program in a loop can be different on each loop and is application dependent.

- Interrupt Supervision

  Interrupts occur asynchronously in the system. They are handled with absolute priority over all user programs (System and Application). The routine driven as result of the interrupt can be application dependent. At the end of this routine, control may either return to the interrupt program or to another one.

- Input/output supervision

  This section controls all the I/O operations of the system. It deals with the queuing and initiation of requests.

- Storage management

  This section is responsible for :

  - dynamic allocation of storage for programmes transferred from discs
  - dynamic allocation of data storage for programs which need to construct messages
  - transmission of data from one programme to another by means of the input queue
  - dequeuing of data from input queue for processing by user programs.

- The System Services section executes services for user programs. These includes get time services, communication with the operator console for the output of messages and the interpretation and dispatching of various commands debugging facilities and data management.

- The System Programmes

  They represent about 160 K Bytes, 50% of which are non resident. The system programmes are composed of about 31 modules responsible for functions such as :
  - confidence checking
  - start-stop, restart
  - status recording
  - device reconfiguration
  - data exchange within the system.

- Application programmes

  They represent approximately 30 modules, totaling about 232 K Bytes. The majority is core resident. They are divided into the following groups :

  - Radar tracking

    The corresponding modules are in charge of the management of information supplied by several radars located at different points of the MADAP area. Three main steps appear in the treatment of radar information.

    - Plot elaboration

      The radar plots are elaborated at the radar site. A digital message is fed into MADAP, which contains for a plot, a position, a mode-code (if any) and other complementary information.

    - Local Tracks handling

      A local track is a succession of plots from one radar concerning one aircraft. The corresponding information which is called in MADAP the "local track block" contains mainly smoothed local position, speed and track quality.

- Common tracks handling local tracks and converted to system or "common tracks".

  Several local tracks corresponding the same aricraft seen by several radars are combined to a single common track, continuing smoothed system position, speed and track quality.
  The configuration of radar used with their plot filtering areas and track responsibility areas can be changed automatically in case of non availability of data from a particular radar station.

- Flight plan processing

  The flight plan processing modules form the basis for all the planning functions within the ATC system. The flight information arrives in the system through a set of ATC messages coming from adjacent centres via automatic links or from manual inputs on keyboards.
  The flight plan information is checked, processed and updated for communication to the display modules.

- Flight plan/radar track correlation

  Correlation is established by cross referencing common track blocks and flight plan data. After correlation, the aircraft call sign is inserted in the label of the radar track displayed. The flight plan position of an aircraft is periodically computed and checked against the actual position derived from radar data. The flight plan position is then updated from the radar position.

3.6.2. PC operational programs

The supervisor system for the peripheral computers (MARTOS) includes :

- program control
- core management
- program control sub-routines
- I/O routines.

The application modules are in charge of the composition and management of message presented to controllers on EDD's, TID's and SDD's. They are also responsible for controller/computer dialogue via the normal input device (TID, KYB, DCP, RLB and printers).
The function of display refreshing is a computer function.

3.6.3. Data banks

The data banks of the Main Computer Complex constitute in fact the system data base.

- The static data bank consists of a number of stores defining the environment of the system (sector boundaries, Airports, Reporting points, standard routes, aircraft characteristics). It is stored on discs and requires about 450 K Bytes.

- The Dynamic Data Bank contains the data describing the instantaneous image of the system. These data are in a number of different stores of fixed size. The principle stores are :
  - flight segment store
  - plan flight store
  - track flight store
  - route segment store

  The dynamic data bank of MADAP has at present a size of 200 K Byte. It is core resident.

- The Reconfiguration stores contain the data necessary for reconfiguration and restart action. The most important stores, which are kept on discs are :
  - the extended flight store
  - the minimum restart table
  - the system configuration table

  - The extended flight store

    Upon entry of a flight plan in the system, the data is checked for correctness and a record called record A is created for this flight. This record contains therefore practically unprocessed data. This is then processed by several modules in order to extract route data and break the flight data into flight segment, route segment etc... to prepare insertion into the various stores of the Dynamic Data Bank.

The result of the processing is record B, which, together with record A are stored on disc until the flight is activated. At activation time, record B is inserted into the Dynamic Data Bank and whenever its content is completed and updated, the same modifications are made to the EFS on disk.
The extended flight store contains therefore two types of information to allow system restart.
Normally record B will be used as it contains the most up-to-date information. This will be a "normal restart". In certain cases, it is however necessary to restart from "safer" data and then, record A is used. This will be an "Elementary System Restart".

- The Minimum restart tables contain tables such as :
  - call signs
  - display settings
  - code allocation
  - MET data
  - Time update message store

- The System configuration table contains all necessary data concerning system status for RTSX, application and system programs.

- The Peripheral Computer data banks

  As already mentioned, display refreshing is a computer function.

- The Picture Repetition Store which is an extension of the PC store contains all information sent to the display units.
  It has two roles : - Data filing to facilitate the storage and retrieval of any information displayed.
  - Refreshment memory to display data at a rate of 20 times per second.
  The PRS is divided in two parts : EDDs and SDDs. Each part is divided in blocks allocated to a specified screen.

- The Display Control Store allows correct management of the PRS (space management, retrieval, updating).

3.6.4. Start - restart problems

MADAP has basically three types of start and restart : the System Start
the System Re-start
the System Elementary restart.

The System Start uses no dynamic data and requires that the complete hardware configuration is specified. It must not be considered as a normal procedure for restoring service after a system break-down.

The System Re-start is the normal procedure after a hardware or software failure. Its activation can be automatic or manual. During a system restart, the complete data base is rebuilt in core, from the information stored on disc at the time of failure. The EFS records i.e. the flight plans are all reprocessed. All programs including the supervisor are reloaded. A typical time for restarting the system is 30-40 seconds, this time being measured from the automatic detection of an error until the first radar plots are visible to the controller.

System Elementary Restart : in case of incompatibility of the data recorded on disc to be used for a system restart, an elementary restart is made. In this case, the hardware configuration needs to be specified as in a system start. Only the code-call sign correspondance is used from the minimum restart tables and the EFS contents are checked for validity. Erroneous records are discarded for further processing.

3.7. Implementation problems.

MADAP was initially conceived as a complete system capable of providing Air Traffic Control over BENELUX and Germany. It later on appeared that it was necessary to organize Air Traffic Control in this area on a different basis and the zone of responsibility of MADAP was limited to BENELUX and Northern Germany.

This together with the necessity to transfer responsibilities into Maastricht on a progressive basis has led to a phased implementation as far as geographical extension is concerned, the sequence being the following :
- Belgium Upper Airspace as from 1st October 1973
- Northern Germany Civil Upper Airspace as from 1st April 1974
- Northern Germany OAT - 1st October 1975

followed by the Upper Airspace above Netherlands.

On the other hand, it has been found necessary also to introduce operational functions on a progressive basis. The implementation of new software packages do not necessarily coincide with geographical extension. The evolution is in the direction of enhanced reliability and flexibility as well as increased use of dynamic flight data displayed on EDD's for planning and coordination purpose and of introduction of automatic exchange of data with adjacent centres.

The size of the hardware and of the software (several hundred thousand instructions) gives an indication of the complexity of such a system. Several hundred man-years of effort have been spent since 1968 to achieve the present result.

The implementation of MADAP has required that solutions be found for a variety of problems such as :
- System architecture using a maximum of standard equipment
- Choice of adequate hardware and system software
- Proper definition of operational functions
- Programs detailed functional specifications
- Organisation of software production
- Software integration and testing
- Operational evaluation
- Introduction into operation
- Maintenance of hard- and software
- Introduction of hardware and software modifications.

4. The KARLDAP System.

4.1. Evolution from MADAP.

When designing the KARLDAP System due account was taken of the MADAP experience. Although MADAP was not operational yet at the initial design stage of KARLDAP, it was pretty obvious that in view of the operational target date for the Karlsruhe Centre, only rather limited variations to MADAP could be considered.
Those variations are mainly in the following fields :
- Method of inputting radar data
- Architecture of the display system with its consequences on the peripheral computer complex
- Technical solution for a Radar Emergency Processor
- Project Execution.

Other differences have been also introduced due to several factors such as variations in the operational requirements or necessity to phase the system implementation. This has had an influence mainly on the approach to software production.

- Method of inputting radar data.

  KARLDAP, like MADAP allows for two possible routings of radar data. At the radar line demodulator output, radar data is fed both to the normal state system and to the radar emergency processor. In MADAP, radar data is fed into the MC through the Submultiplexer specially developped which was also supposed initially to handle Input/Output for data links with adjacent centres. The later function is now one of a front-end processor for both MADAP and KARLDAP, and the radar data, in KARLDAP is fed into the system via the Peripheral Computer Complex were standard hardware/software can be used.

- Architecture of the Display system.

  It was decided to implement KARLDAP with an improved display system allowing better working conditions (accuracy, readability, ambiant lighting conditions).
  This led to a requirement for high picture refreshing rate (50 Hz) which together with considerations linked to display load capability resulted in a decentralized concept for the display system whereby each Synthetic Data Display has its own character generator and picture repetition store.
  For EDDs, however, the centralized version was maintained. This approach which allows to see each SDD as an individual peripheral permits a more flexible organization of the Peripheral Computer Complex whereby the processing load, already reduced compared with MADAP can be better shared between processors. This results in a reduced number of processors.

- Technical solution for a Radar Emergency processor

In KARLDAP terminology, the subsystem used for maintaining radar labelled display in case of central computer complex failure is known as the radar by-pass. This is a completely independant subsystem which can feed directly the display system. Rather than using a general purpose computer (as done in MADAP) it was found preferable to make use of a specialized processor which allows a form of tracking and multi radar mosaic display. This processor is in fact, in the first phase of operation of KARLDAP responsible for the radar data processing functions.

- Project execution

In view of the experience acquired with the MADAP system, EUROCONTROL has decided to assume direct overall responsibility for the implementation of the KARLDAP system. After the preparation of the general configuration of this system, the sub-systems were studied in depth and the equipment was chosen accordingly. Furthermore - and this is a factor of major importance in the use of MADAP "know-how" - Eurocontrol has assumed overall responsibility for the production and integration of the software.

From a more general point of view, the need to transfer the experience acquired progressively from one data processing system to another led EUROCONTROL to set up PADE (Programming and Analysis Division EUROCONTROL) which is attached to its Experimental Centre at Brétigny, France. It is this Division which, vis-à-vis the General Directorate of the Agency, is acting as the supplier of the software for KARLDAP. The System hardware is being supplied through contracts with various firms.

4.2. Stages of Implementation.

As is customary when initial projects are adapted to suit the realities of a situation, the Agency has decided to employ several stages in the implementation of the KARLDAP project, the first two of these being known as KARLDAP I and KARLDAP II.

In fact, where KARLDAP I is concerned, the scheduled date for the start of operational service (late 1975 or early 1976) has dictated the adoption of an intermediate stage called KARLDAP A.

These various stages differ from each other chiefly in the degree of complexity of the functions performed. One of the guiding principles in the design of the KARLDAP system as a whole was that the configuration and capacity of the equipment should be such as to facilitate the smooth transition from one stage to another where both the hardware and the software were concerned.

Obviously at the present juncture the three stages referred to above have not been defined with the same degree of detail. KARLDAP I is scheduled for 1978 and KARLDAP II has no fixed date for entry into service. The present paper therefore deals only with the KARLDAP I stage and its preliminary phase, KARLDAP A.

The basic parameters determining the size of the KARLDAP system are as follows :

- up to 300 active flight plans in the system
- up to 250 simultaneous radar tracks
- ability to drive up to 60 Synthetic Data Displays (SDDs) and 90 Electronic Data Displays (EDDs).

4.2.1. KARLDAP I - Functions

The main functions of KARLDAP I are :

- Reception and processing of radar information transmitted in digital form from a number of radar stations,
- Reception and processing of ATS data messages from adjacent ATC units or messages taken from a data bank,
- Direct exchange of digital data with adjacent control centres,
- Correlation of flight plan and radar data and updating of flight plan data from the radar data,
- Storage and processing of data for presentation in various forms (strip printout and display on screen) and distribution of data to the different working positions, as appropriate,
- Direct communication between controllers and the data processing system by means of manual input devices,
- Recording of data for statistical, legal or reconfiguration pruposes,
- "Off-line" operation for program development and staff training purposes.

4.2.2. KARLDAP A

As already mentioned, a preliminary stage for KARLDAP I has had to be planned. The KARLDAP A system differs from KARLDAP I in the following respects :

(i) radar data processing is carried out by a processing system which is separate from the computer complex.

(ii) there is no updating of flight data by means of radar data, the radar data/flight plan correlation being limited to associating the call-sign with the SSR codes.

(iii) the display system is limited to SDDs.

(iv) flight data are displayed by means of printed flight progress strips.

(v) automatic data input is limited to information received via the AFTN and MET information.

The hardware components used in the KARLDAP A and KARLDAP I systems, and the functions which they perform, are shown schematically.

4.3. System Structure

Similarly to MADAP, KARLDAP consists of three parts, viz. :

- the Data Communication and Terminal System (DCTS)
- the Central Computer Complex
- the Operator Display System (ODS).

Reliability is guaranteed by the provision made for system reconfiguration.

With regard to the computers, 100% redundancy is provided, and switching devices for switching the standard peripherals and the display units and special input/output units (ODS), are installed. These switching devices are controlled either by the software or manually.

Despite this redundancy and the feasibility (demonstrated by MADAP) of software capable of bringing about system reconfiguration in the shortest possible time, it has been felt necessary to provide a standby unit separate from the computer complex. This unit ensures that the radar data continues to be displayed in the event of a breakdown of the computer complex; it is referred to as the "Radar By-pass" and can handle up to four radar stations simultaneously using a "mosaic" display.

In the initial operational phase of the system the Radar By-pass acts as the radar data processing unit.

4.3.1. Data Communication and Terminal System.

This system will ensure the exchange of digital data between external bodies and the central computer complex. It forms part of the KARLDAP I system and is now in the definition stage. In function and organisation it is identical with the system being installed for MADAP.

As far as KARLDAP A is concerned, only limited functions are envisaged; these will be performed directly by the Computer Complex.

4.3.2. The Central Computer Complex.

The functions associated with automatic processing were, in the case of MADAP, classified under the two following headings :

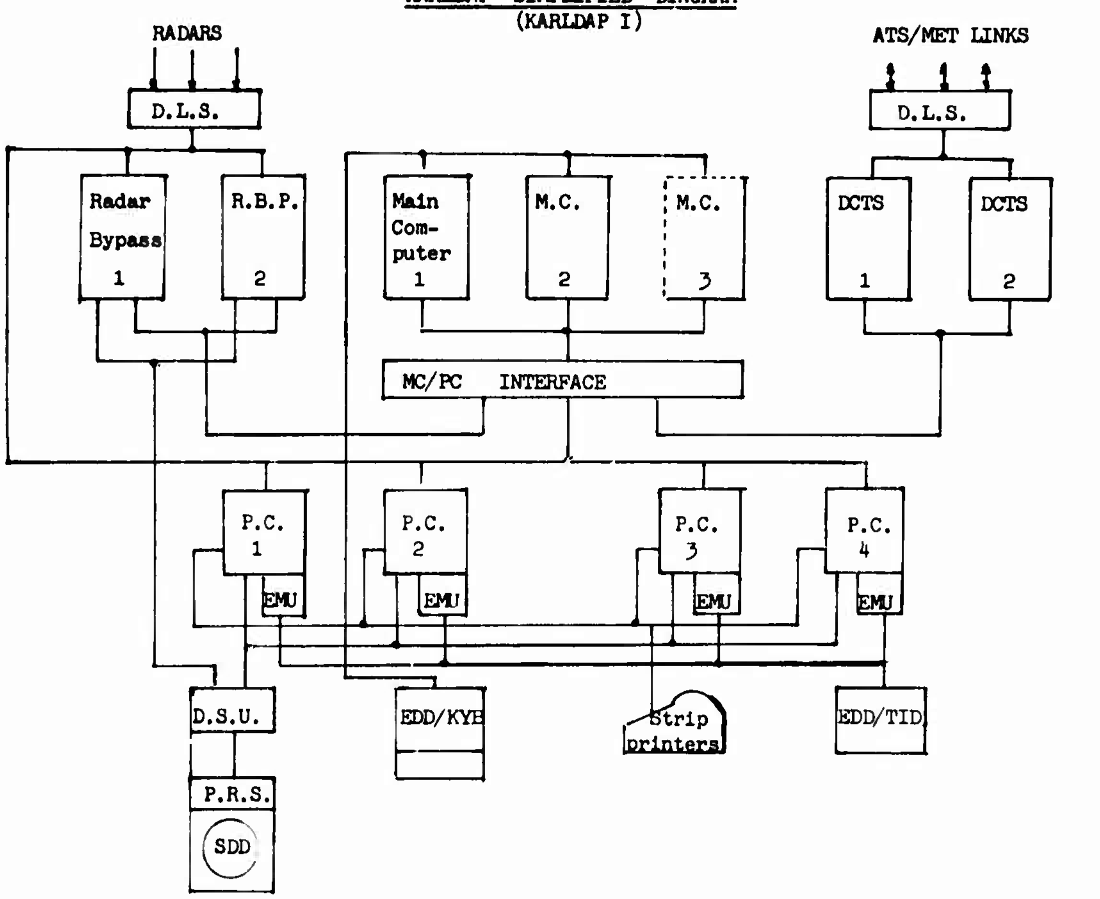
(a) General functions of the system, e.g. :

- Multiradar tracking
- Flight plan data processing
- Correlation of radar data/flight plan data, giving a general picture of the airspace situation in the area in question.

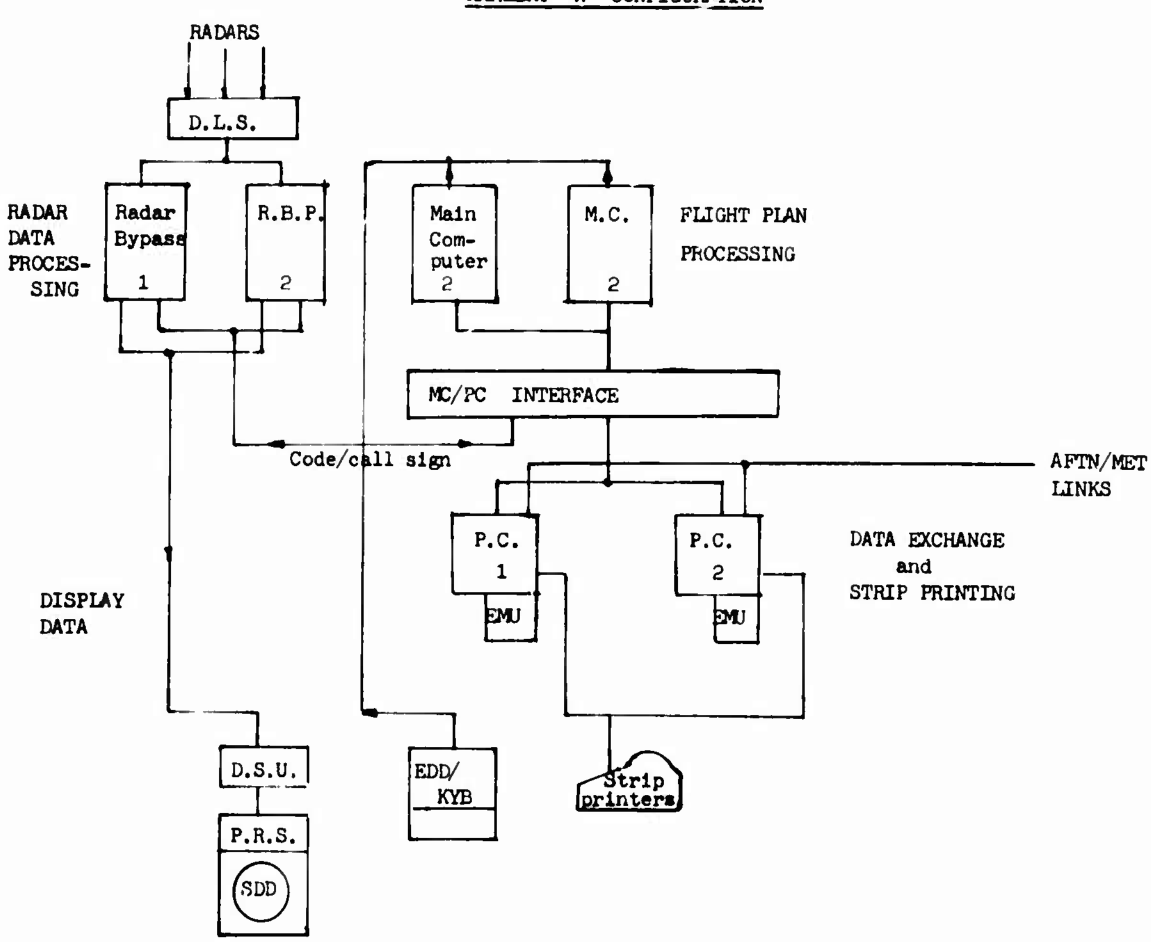
(b) Specific functions in conjunction with action taken at the working positions,e.g.:

- Filtering of data for display purposes
- Interpretation of controllers'input orders
- Updating and renewal of displayed data.

KARLDAP SIMPLIFIED DIAGRAM
(KARLDAP I)



KARLDAP A CONFIGURATION



This produced a two-tiered structure for the computer complex, namely a main computer complex and a peripheral computer complex.

Virtually the same idea is being used again for KARLDAP, apart from a few differences due either to improvements resulting from the experience gained with MADAP or to technological advances. The Central Computer Complex therefore consists of a main computer complex carrying out general functions, and a peripheral computer complex which performs functions specific to the working positions.

Main Computer Complex

The main computer complex comprises two IBM 370/158 computers, one of which is used for normal system operation, while the other acts as a standby unit ready to come into immediate operation in the event of a failure of the first computer. The structure of the computer complex and the connection to the peripheral computers have been designed so that a third computer can be added without difficulty if this becomes necessary for reasons connected with program development and system reliability.

Each computer is supplied in the following configuration :

- a central store of 1,024 K bytes
- a byte multiplexer channel
- four block multiplexer channels.

The secondary external stores consist of the following :

- six IBM 3333 discs in three pairs, each pair being connected to one of three IBM 3830 disc control units
- six IBM 3420 tape units in groups of three, each group being connected to one of two 3803 disc control units.

The connection to the other components of the system (i.e. peripheral computer complex and the DCTS) is made by means of four 2701 adapters.

All the devices incorporate a double access for switching from one computer to the other under the control of a program.

Two high-speed printers and two card punch/readers are also provided.

A switching system comprising a set of eight switching units, each of which consists of a matrix for connecting four central unit channels to four peripheral control units, makes it possible to obtain a wide variety of configurations, thus contributing to great operating flexibility and optimum availability. These switching units can be controlled either manually or by means of remote control.

Peripheral Computer Complex

The peripheral computer complex performs the following functions :

- Pre-processing of radar data.
- Routing of radar data to the main computer complex.
- Routing and formatting of messages originating in the main computer complex for display on the SDDs or tabular EDDs.
- Conversion of radar data for display on SDDs.
- Processing of manual input orders from the working positions.

The complex consists of a total of four identical TR 86 computers supplied by AEG/Telefunken, two of which are necessary for the operation of the system. The remaining two are intended in the first instance for standby operation in the event of failure of either or the first two computers, and secondly, to enable comprehensive program development and testing facilities to operate independently. Provision has been made for the complex to be expanded to a total of 6 peripheral computers, while the manner in which the complex is organized is such that the computers are completely interchangeable from the functional viewpoint. This makes for great flexibility in operation and increases computer availability.

The necessary connections are available to enable each computer to handle data appertaining to all radar stations and all working positions. During normal operation the load will be distributed equally.

Each computer is supplied in the following configuration :

- Central unit
- Store of 48K words each of 24 bits
- An extension storage unit of 16K words for refreshing the tabular displays.

The peripheral equipment consists of the following :

- 2 disc units
- 2 high-speed printers
- 4 magnetic tape units
- Interfaces with the operator/machine exchange and display system
- Tape readers and punches
- One punched card reader.

Radar data link splitter

The radar data link splitter (DLS) enables data to be routed to the peripheral computer complex and the independent radar processing system. The data is sent in serial form. This unit is of modular construction and can be extended to accommodate up to 24 separate radar data transmission lines enabling a total of 6 radar stations (primary/secondary) to be handled.

The Independent Radar Processing System

The independent MP 720 radar processing system is an improved version of equipment already developed for the Agency by T-VT (Thomson-CSF) for use in the initial Maastricht system (MINFAP) and for the SHANDAP radar data processing system.

The system planned for KARLDAP can provide a composite picture from four radar stations selected out of a possible six stations. Furthermore, it does not perform the picture renewal function.

The MP 720 can undertake the following :

- Secondary radar tracking
- Primary radar plot processing
- Radar track/flight plan correlation of the code/call-sign by automatic input from the computer complex, or by manual input
- Transmission of track, plot, and video map data to the display system.

The system is duplicated and the two MP 720s operate in parallel, the information from one of them being used for display purposes.

Testing facility

With a system of such complexity, an important aspect is the need for improvements and developments after the initial operational version has been brought into service.

Before the commissioning of the future operational system, the final phase of acceptance demands for practical purposes that a complete system which is representative of actual operation, but does not interfere with the current operational system from the performance or reliability viewpoints, should be available.

Consequently, the means for completely isolating such a system (consisting of Main Computer, Peripheral Computer and input/output displays) from the operational system have been incorporated. The appropriate hardware has already been ordered as far as the peripheral computers and the displays are concerned.

In the case of the main computer, provision has been made for a third computer and its peripherals to be added without difficulty.

4.3.3. Operator Display System (ODS)

Composition

The system is made up of the following sub-systems :

- synthetic dynamic data displays (SDDs)
- electronic data displays with touch input devices (EDDs + TIDs)
- tabular display stations with input keyboards
- page printers
- flight progress strip printers

4.4. KARLDAP Software

4.4.1. Method of development

As already stated, the KARLDAP system makes the maximum use of what has been achieved in connection with the MADAP project. However the operational characteristics of the two systems are somewhat different, and the requirements of the time schedule have precluded the possibility of a straightforward transfer of software. Moreover, the hardware configuration shows a number of improvements over that employed for MADAP.

The EUROCONTROL Programming and Analysis Division is responsible for the production and integration of the KARLDAP A and KARLDAP I software. For this purpose it has a program testing facility at Brétigny which consists of the following :

- An IBM 360/50 computer (which is compatible with the IBM 370/158)
- A TR 86 peripheral computer which is on temporary loan from the KARLDAP system
- A data display and input/output unit of each of the types used in KARLDAP.

Final programme integration will be carried out at the Karlsruhe Centre.

4.4.2. Development of KARLDAP A

In order to speed up the development of the system the following choices have been made :

(i) -the monitor for the main computer will be the Customer Information Control System, (CICS),

(ii) -the software language will be PL/1,

(iii)-the modules used with the TR 86 will be modules which have already been developed either for MADAP or for other systems.

4.4.3. Development of KARLDAP I

The time-table for implementing KARLDAP I allows MADAP developments to be used to the full, particularly where supervisors, reconfiguration procedures and multi-radar tracking are concerned.

Consequently, the work of the Programming and Analysis Division is mainly concerned with the application software, since the RTSX (Real Time Executive) supervisor and special executive modules (discs, interfaces, etc...) have been adapted directly from MADAP - the RTSX for the IBM 370/158s, and MARTOS (MADAP Real Time Operating System) for the TR 86 computers.

4.5. KARLDAP implementation

The KARLSRUHE UAC is scheduled to come into service with the KARLDAP A system at the begining of 1976. In order to facilitate software integration and the operational assessment of the system, the entire hardware required for KARLDAP A is scheduled for installation and connection by late 1974, begining 1975. The hardware required for KARLDAP I will also be available in 1975; however, the increased complexity of the system precludes the likelihood of this phase becoming operational before 1978.

5. The SHANDAP System

5.1. SHANDAP phases

SHANDAP (Shannon Automatic Data Processing and Display System) is a new Air Traffic Control system which the EUROCONTROL Agency is setting up in Ireland for the Upper Airspace Centre at Shannon.

Already in 1965, the increase of traffic in the area had led to a requirement for secondary radar coverage in order to supplement the conventional procedural based ATC - system.

Here, contrary to the classical approach whereby flight plan processing is introduced first if not simultaneously with radar data processing, it was felt necessary, due to the characteristics of air traffic in the area to implement first a radar data processing system to be followed latter on by a flight plan processing system. These are the two phases for the SHANDAP System :

SHANDAP I, the main component of which is the RDPS (Radar Data Processing System)

SHANDAP II, which incorporates the FDPS (Flight Data Processing System).

The first phase, for which the equipment is functionning on site on a evaluation basis should be introduced into operation in early 1975. It is designed essentially to provide the executive radar controller of the UAC Shannon with a labelled radar picture containing position and identity information.

The planned date for introduction into operational service of the second phase, SHANDAP II is early 1978. It will automate the processing and display of flight plan data, in a system where the FDPS and the RDPS are interfaced. It is extensible to meet the whole of Irish airspace requirements.

5.2. The Radar Data Processing System : SHANDAP I

The RDPS of SHANDAP I receives aircraft position and code information from a remote secondary radar station. A plan position labelled display is generated on Synthetic Data Displays as well as reduced plan information on Electronic Data Displays.

The equipment comprises two main parts :
- the radar subsystem
- the tabular display subsystem.

Both subsystems are duplicated and are independant from a reliability point of view.

Operators communicate with the system essentially with two operational messages :

- the Current Message (CM) which describes for one flight the basic flight plan data.

  This message will be on one hand displayed to the relevant controller and on the other hand used to establish code/call sign correspondance.

- The Modification Message (MM) is a proposal to correct the Current Message.

Furthermore the code/call sign correspondance is to be inserted manually.

5.2.1. The Radar Sub-system.

The radar information is received from the radar site through a broad-band microwave link over a distance of about 10 miles. Radar extraction is performed in the centre by the radar video extractor (EV 720). It is a hard-wired equipment and outputs digital plot messages.

The digital plot messages are then fed into a plot processor (MP 720) which may be considered as a flexible "fixed programme" special purpose computer. This equipment establishes tracks from plots. Once a track is initiated, the future extrapolated position is computed and compared with the actual next radar position. Three windows of increasing size are centered on the extrapolated position. The size of the window in which the next radar position falls is used as a quality factor to compute a smoothed track position.

The plot processor contains a data selector which constitutes the interface with the radar displays on one hand, and with the tabular part on the other hand.

It is through this data selector that the code/call sign correspondance is inserted and processed in order to deliver the Synthetic Data Display, with respect to the aircraft in the system, its position, call-sign and flight level (mode C).
The radar display equipment comprises 4 executive controller positions each with their synthetic character and symbol generator.

5.2.2. The Tabular Subsystem.

The tabular subsystem contains Electronic Designation and Input devices (EDID) in the form of tabular displays and keyboards driven by a multi-8 (M8) computer system (mini-computer manufactured by Telemecanique).
The tabular subsystem is used basically for three purposes :

- Insertion of code/call sign correspondance for track to call sign correlation to be executed in the plot processor.
- Dispatching of reduced flight plan data information to the relevant executive controllers.
- Management of the radar display label positioning.

5.3. The Flight Data Processing System : SHANDAP II.

5.3.1. Function

The Shandap II system is to implement the following additional functions, for a capacity of 200 active flight plans.

- Input of flight data

  The flight plan messages, routed via the Aeronautical Fixed Telecommunication Network (AFTN) or later via computer-to-computer links, will be inserted into the system at a flight data suite (FDS) under computer assistance.
  For recurring, regularly operating flights, Stored Flight Plans (SFPs) will be used. Flight activation will occur on the basis of an activation message by either manually or automatically input.

- Processing and management of flight data

  This includes :

  . route recognition
  . calculation of ETA and updating from the radar information
  . SSR code management.

- Printing of strips

  Provisions are made for extending the strip printing functions to Shannon and Dublin ACCs.

- In conjunction with RDPS

  . flight plan track correlation
  . automatic elaboration and transmission of the Current Message (CM) to the RDPS, thus allowing automatic code/call sign association
  . handling of Modification Messages (MM)
  . transmission to RDPS of updated ETAs, through the Modification Message (MM).

5.3.2. SHANDAP II Hardware

The SHANDAP II system is organised around a central computing complex which performs the general system functions as well as the interfacing with ground-ground data links, flight data input positions, flight data display and input subsystem and radar data processing system.

- Computing Complex

  The computing complex comprises two identical chains, one being the operational chain and the second one the standby chain.
  Each chain comprises :
  - an Iris 55 computer (manufactured by CII) with a memory capacity of 256 KBytes,
  - a control typewriter

- one disc control unit driving 2 disc units MD 50. A disc unit has a capacity of 50 M bytes.
- one tape control unit driving 2 magnetic tape units 72 330
- a high-speed printer
- a card reader.

The IRIS 55 is member of an upward compatible family (until 1024 K Bytes)

- Operational Peripherals :

The operational peripherals comprise the displays, keyboards and strip - printing units. These devices can be driven by a common control unit (T-VT 6000 microprocessor). In the initial configuration, two micro-processors share the task of driving a total of six printers and 7 display/keyboards.

The strip printers are special purpose equipment (SODERN) of the same type developed for Karldap. They are connected to the IRIS 55 through the control unit (T-VT 6000).

- Interface equipment

The interface with the AFTN network and with RDPS is ensured by standard CII devices.

5.3.3. SHANDAP II Software

- System Software

The system software includes :
. The Real Time Monitor
. The ASTRE Compiler
. Standard IRIS software
. System Support software

- The ASTRE Compiler

ASTRE is the high level language which the Agency has opted to use for the development of SHANDAP II programs. The ASTRE compiler is part of the standard IRIS 55 software.

- Standard IRIS Software

The standard operating system used for off-line operation is SIRIS 3. The standard software operated under SIRIS 3 includes
- Assemblers and compilers (ASSIRIS, Macro Assembler, MAGIRIS, FORTRAN, COBOL, etc...)
- Programming and testing aids.

- System Support Software

This covers the ability to simulate the operational inputs and outputs of the SHANDAP II system as well as test programmes allowing modules or groups of modules to be tested in a stand-alone mode.

- Application Software

The production of the application software is the direct responsibility of the EUROCONTROL Agency.
The production phases are rather conventional for such a project. At the end of the software system design period, the amount of application programs to be produced for the initial version of SHANDAP is estimated to be of the order of 160 K Bytes or about 40.000 object instructions.
This corresponds to an effort of the order of 40 man x years assuming the use of ASTRE as a high level language.

5.4. Implementation

In order to enable the Programming and Analysis Division of EUROCONTROL to develop the programs required, a computing chain has been installed at the EUROCONTROL Experimental Centre, Brétigny-sur-Orge, France, at the end of 1974.

After program integration, the final system integration is to take place at Shannon, with subsequent operational evaluation.

THE NETHERLANDS ATC AUTOMATION PROGRAM

by

J.S. SMIT+

Deputy Director ATS and Telecommunications,
Rijksluchtvaartdienst
Ministerie van Verkeer en Waterstaat
Plesmanweg 1
THE HAGUE, The Netherlands

TABLE OF CONTENTS

+ With a contribution by J. de Koning and H. Visscher and assistance from other members of the ATC system design team.

THE NETHERLANDS ATC AUTOMATION PROGRAM

by

J.S.Smit,
Deputy Director ATS and Telecommunications,
Rijksluchtvaartdienst
Ministerie van Verkeer en Waterstaat
Plesmanweg 1
THE HAGUE, The Netherlands

1. HISTORICAL SUMMARY

1.1. SATCO 1

The Netherlands were one of the first States in the world to start an ATC automation program. Back in 1956 Hollandse Signaalapparaten started the development of a laboratory model of a stripprinting system, which it demonstrated in 1957. Then the Netherlands Department of Civil Aviation placed an order for an operational stripprinting system for Amsterdam ACC. At that time, the main consideration of the Administration for ordering automation equipment was not the immediate need for such a system (it even was not to be expected that the first and simple stripprinting system would prove to be of noticeable help to the controllers). The overriding consideration was the conviction that the time was approaching at which ATC automation would become an indispensable tool to keep up with the traffic growth and, secondly, that the development and putting into operational use of such systems would be a difficult and therefore lengthy process. Also, it was considered that the initiative of Dutch industry was a valuable one and worth supporting, not only for national air traffic control purposes, but for the development of ATC in general.

Rather than starting an experimental development project the Administration chose for an operational system as this would bring it closer to reality and lead to a basis for further operational system development faster than any other approach. Of course, the very first stripprinting system did not play a vital rôle in the total ATC system in the sense that a system break-down would make ATC impracticable. Therefore, the risk of delays and mistakes could be taken. But on the other hand the pressure to make an operational system was more challenging, particularly as for such a pioneer system the entire world was watching.

The first system, SATCO phase 1, was delivered to the ACC in 1960 and put into operational usage early 1961. Compared with what we know and see to-day, it was a very simple system, but it <u>did</u> produce printed strips. The system contained many features which were never really used operationally, partly because of the lack of suitable input equipment, partly bacause of Amsterdam ACC still being on an "automation island". Among these features were the processing of revisions to basic flight plans (clearances, estimates, position reports), even the input of radar positions (be it that these had to be inserted by means of coordinates-inputs on a teleprinter which, of course, was not practicable) and the output of boundary estimates for adjacent centres.

Nobody will claim that SATCO 1 was a real help for the control staff in the sense that it created a decrease of the workload (maybe even the contrary was true) but SATCO 1 played an enormously important rôle in getting control staff used to the idea that a machine had a potential to take away certain burdens from them and, of equal if not greater importance, it taught system designers what mistakes had been made and what more had to be done to develop a truely helpful system.

1.2. SATCO 2

The development of SATCO 2 (again destined for operational use) had already started before SATCO 1 was operational. Though convinced that a further step had to be taken, particularly to cope with the "jet age", the Netherlands Administration took a risk in ordering phase 2 as this certainly would penetrate into the overall operational system. Radar data processing was still in a stage of development and not ready for inclusion into an operational ATC system. Therefore, SATCO 2 concentrated on flight data processing. The controller's tool would be a computer-driven flight progress board, the controller using a keyboard instead of his pencil. The major feature of this approach was, that - as the controller could only update his flight progress board by using the computer - the computer would "automatically" be informed of all changes to the flight data, which it could then by program rules use for the automatic updating of data of that flight displayed, or to be displayed, at other positions. In other words, one of the major features of SATCO 2 was the use of the data processing system for coördination betwee control positions.

A conflict search program (probably the first in the world) was included, also aimed at easing the coördination workload.

The operational computer program was based on the rules for procedural control. This, as later showed, was the cause of severe difficulties, which, at one stage, made it even doubtful whether SATCO 2 could ever be put into operational use.

In 1964 the system, for which the program specifications had been frozen in 1961, was installed at Amsterdam and the trial period started. But in the meantime, fundamental changes had taken place at the Amsterdam ACC as a result of the increasing rôle of radar in the ATC system. As in most centres, a mixture of procedural and radar control had emerged which required a very flexible way of applying the rules for separation and coördination. The SATCO 2 program was not built for this method of ATC and the trial period ended early 1965 in a very discouraging result: SATCO 2 was no good the way it was......

Much happened in 1965. Two parallel actions started, one aimed at using SATCO 2 equipment (particularly the duplicated central computer system with its automatic switchover facilities and duplicated mass memories) for an extended stripprinting system to replace SATCO 1. The other action was aimed at finding out why SATCO 2 had failed and what could be done to remedy. Both actions were succesful. In 1966 SATCO 1 was replaced by an extended stripprinting system including both ACC and APP. Shadow trials also proved that the - modified-flight progress boards, its use based on a computer program adapted to the plan-executive system. could play a rôle especially for inter-sector coördination and become of real assistance to the system.

In the meantime, a new ATC building was being built, in which a centre had to be housed suitable to meet the requirements of the vastly increasing traffic volume. A five-sector system was developed, based on the use of SATCO 2 with automatic flight progres boards as the operating tool for the plan-controller. Radar controllers were provided with printed strips providing clearance data. Approach control and the tower were equipped with printers and also an output link with the military ATC centre was foreseen. Control-assistants were equipped with printers and could, directly upon receipt by telephone, input boundary estimates, while outgoing estimate messages were printed on their printers for passing of the information by telephone to the adjacent centres.

As of consequence, the changeover process into the new centre not only meant a physical removal but also a fundamental change of the ATC system and its working methods. After several months of trials, shadow working, programme adaption, etc, the new centre with SATCO 2 was put into operational use on the evening of February 26, 1968.

SATCO 2 contains a duplicated computer-system with automatic switch-over, independently duplicated drum mass-memories, seven flight progress boards and 48 input-output printer channels. SATCO 2 serves the Amsterdam ACC, now including 9 sectors and a flight data section, Schiphol TWR and APP, Roterdam TWR/APP, and the automated military ATC system PHAROS. Also, flight data outputs are provided for the Airport Authority, airlines, costum authorities and for statistical purposes.

2. THE NEXT GENERATION

2.1. Basic approach

At the time the oparational use of SATCO 2 was well established, mid 1968, it was realized that a start had to be made with the successor to SATCO, which later was given the code name SARP. When approaching this problem the sessons learnt from the past experience were put to practice. Most inportant of these are

- the design of an ATC system requires a detailed operational plan covering the lead-time plus a certain period thereafter, in order to ensure that the system - when ready to be implemented - fits into the operational environment; such an operational plan must contain a detailed specification of the procedures and working method; statements of "what will be done" do not suffice, they should specify "how it will be done",
- the development of the operational environment (route structure, procedures etc) must be in accordance with that plan; if the necessity arises to deviate, these deviations must be verified with the plan to ensure that both reality and development remain matched,
- system designers, particularly the software staff, must work in an integrated team with the staff responsible for the design, implementation and execution of ATC procedures.

Taking advantage of the modest size of this Administration, system design and development responsibility was integrated with the procedure and executive responsibility, thus ensuring to a maximum degree the mutual evaluation of the impact that a future system will have on the present and the impact that unforeseen requirements of to-day will have on that future system.

It had been learnt (the hard way, but maybe the best way) that, once a goal has been set, one cannot unpunishedly deviate from it; but neither can one ignore newly arisen requirements and get away with it in the end. Therefore the development of an ATC system demands a constant careful watch both of the operational executive system ad well of the one being developed.

The operational plan developed for the seventies aimed at expansion of the terminal area with the holding stacks located at the TMA entry points, and the establishment - as far as possible - of uni-directional routes. The data processing system to serve ATC was to comprise radar data processing and a complete replacement of the SATCO system.

The overall plan resulted in a three-phase approach:

- phase 1 contained the re-organization of the route structure and TMA, still using the existing equipment,
- phase 2 contained the introduction of radar data processing for approach control (SARP 1), while keeping SATCO in use for flight data processing,
- finally phase 3 would introduce radar data processing for ACC and also comprise the replacement of SATCO; this would be realized by the SARP 2 system.

The above phased approach avoids, to the extant possible, that fundamental changes in the airspace organization (and hence the sectorization and operational working method) coïncide with the introduction of completely new equipment with the objective to prevent the control staff being faced with a twofold adaptation process at one and the same moment. An exception to this will be the introduction of an upper airspace sector at Maastricht as part of phase 3.

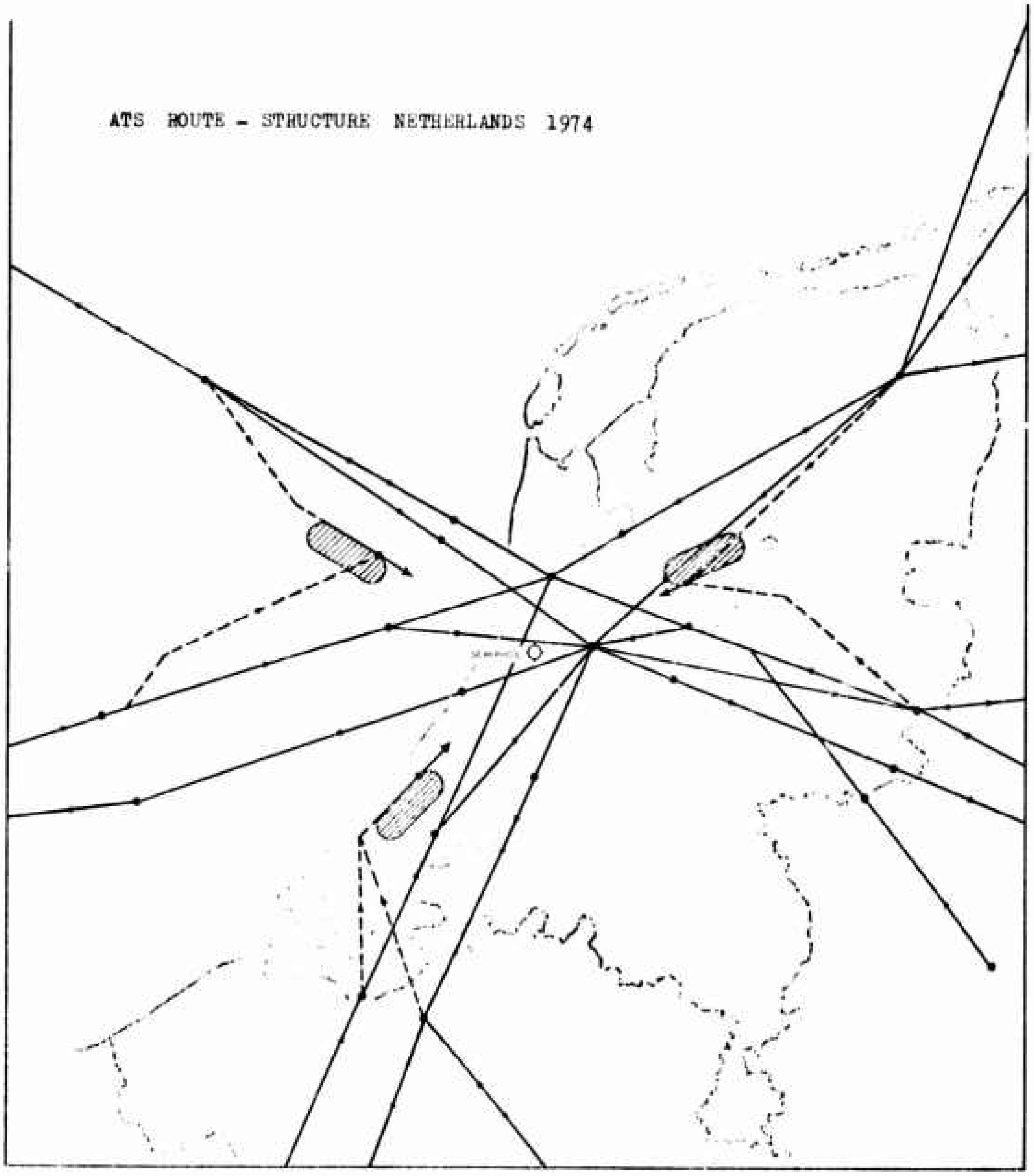
Phase 1 was, via a number of intermediate steps, completed in the beginning of 1974, when the so-called "three stack system" was introduced.

The SARP 1 system has been installed and is expected to be operational by the end of this year.

SARP 2 will be installed in the new wing to the ATC building in 1976.
Following the de-bugging and trial period, the operational familiarization will take place in 1977 and the system is expected to be operational in the beginning of 1978.

2.2. SARP 1

SARP 1 is a radar data processing system for APP Schiphol. Radar data are derived from a 10 cm TMA radar plus SSR; both primary and secundary radar data are digitized on-site by a duplicated extractor (both the radar and the extractor were manufactured by CSF), which feeds the



plot data into the (duplicated) computer system. The computer is linked with the SATCO system, from which flight data is received a.o. to enable callsign-SSR code correlation. The display sub-system comprises two-colour PPI's on which, by interscan-techniques, synthetic data is added to the raw video, thus producing a labelled display on which raw radar data and synthetic information can be easily distinguished by virtue of the different colours. In addition, each controller is equipped with an electronic data display (EDD) which provides him with additional flight data, basically coming from the SATCO system (arrival times, expected approach times,etc). The SARP computer and display systems were manufactured by Hollandse Signaalapparaten (SARP stands for Signaal Automatic Radar data Processing).

SARP 1 was designed on the pattern of the ARTS-III system of the FAA. There are, however, some marked differences. As already mentioned, the tracking process is applied to both primary and secundary plots. Also, an automatic calculation of expected approach times (EAT, the time the TMA may be entered) is introduced the facilitate the coördination between ACC and APP and to avoid TMA congestion. The latter facility is already practiced to-day by use of the SATCO system; the calculations will continue to be carried out by the SATCO computer and fed to SARP-1 by direct link.

The contents of the labels may vary according to the wishes of the controller. The maximum data shown is callsign, mode C read-out, current cleared FL, calculated ground speed, required ground speed, aircraft type, given heading.

Aircraft identification and label initiation is carried out automatically for 4-digit SSR codes. In cases were controller identification is required, his input is facilitated by lightpen.

2.3. SARP-2

SARP-2 is a combined flight data and radar data processing system, serving both Amsterdam ACC and Schiphol TWR/APP. In addition SARP-2 will be linked with the military ATC radar data processing system PHAROS, the Schiphol airport data processing system and will have input/output facilities at Rotterdam ATC, airlines, etc. Also, the system will be linked with the AFTN switching centre.

Two radarstations, both equipped with duplicated electronics, duplicated extractors and a back-up SSR antenne, will feed primary radar and SSR plot data to the central computer complex. One of these is the same radar station as used in SARP-1 (at Schiphol airport), the other is the 23-cm radarstation at Herwijnen, which is also used by the Maastricht UAC and the PHAROS system. In addition, digitized D/F data will be received from a four-station fixer network to facilitate plot identification in case of non-automatic identification.

The central computer complex (CCC) consists of two radar computers, two main computers (the same as in SARP-1) and three magnetic drum memories.

Cross-connections give maximum safety: as long as one of the duplicated elements is operative, the CCC will work.

The CCC directly feeds the EDD's and teletypewriters used in the system. Each EDD has its own refresh-store.

Plan view displays (PVD), showing synthetic data only, are driven by display computers (two PVD's on one display computer), which are linked with the CCC. In addition, the display computers are directly linked with the radar extractors, to ensure that, in the case of a total failure of the CCC, plot data is still received (radar by-pass). The display computers are organized in such a way that closely related operational positions are not dependent of the same display computer. To cater for emergency re-configuration spare PVD's, connection on-line, are provided for and, in addition, display computers can very rapidly be re-allocated.

The radar and display computers are mini-computers (type SMR-S) developed from the type SMR main-computer. The usage of computers of the same family throughout the system greatly eases interface- problems and is benificial for the software development and maintenance.

Sarp-2 is designed for controllers to work in day light conditions; the PVD's (50 cm diameter) have a repetition rate of 50 Hz. and an extremely high definition. The controller uses a light-pen not exclusively to identify the aircraft for which he wants to make a input, but also the most-used functions are displayed on the PVD, which enables the controller to make routine inputs by use of his light-pen only (a form of touch-display). In other words, although he may use a keyboard, he normally will only use the light-pen, thus avoiding the need to take his attention away from the PVD.

The EDD's have been designed for fast and easy operation; identification keys along the side and input keys at the bottom refer to data on the EDD, by which method a programmed "keyboard" is made available.

The combination of PDV's and EDD's and the concept of plan and executive controllers working side-by-side, is aimed at abandoning the use of paper strips.

The Amsterdam ACC will have 8 of these combined working position, plus three spares. In addition, there will be five flight data positions, equiped with EDD's and alpha-numeric keyboards, plus a flow control position equipped likewise. A few teletypewriters, positioned at strategic positions in the centre, are available for non-routine inputs.

Schiphol APP will have four combined working positions, plus one spare. Schiphol TWR will have four EDD's with input facilities.

Data Display

SARP-2 will integrate all ATC functions into one overall automated ATC system, aimed at providing improved working conditions (daylight) and further easing the controllers workload. Advantage is taken of modern programme techniques to minimimze the flight data shown to each controller without witholding any data.

For the presentation of stripinformation on EDD's a number of rules are laid down in order to determine precicely when and what type of information has to be presented at which working positions. Also an attempt has been made to keep the stripinformation for each ATC function as compressed as possible. This could be done due to the very nature of the EDD's on which, in a simple and fast way, extra information of any flight can be shown.

Unlike some other ATC systems, the position of all aircraft within a controller selected area will be shown on the PVD. For the presentation of these positions a number of special symbols will be used. Some of these symbols belong to a special category so called "under control symbols". As soon as an aircraft has been taken under control by an airtraffic controller the displayed position or under control symbol will be changed to the under control symbol that has been assigned to that specific control function.

Labels of flights will only appear automatically on the PVD's of those controllers which belong to a so called under control sequence for those flights. Using the route information from the flightplan of a newly entered flight in the system, a sequence will be made up of all ATC functions which will possibly get these flights under control. ATC functions which do not belong to the under control sequence can only be provided with label information on a manual input.

If it has not been possible to relate radar plot information with flightplan information, labels of these flights will be shown at predetermined times on the PVD's of controllers which belong to the under control sequence, of these flights. These labels will be positioned in a minitable which has a fixed position on the screen.

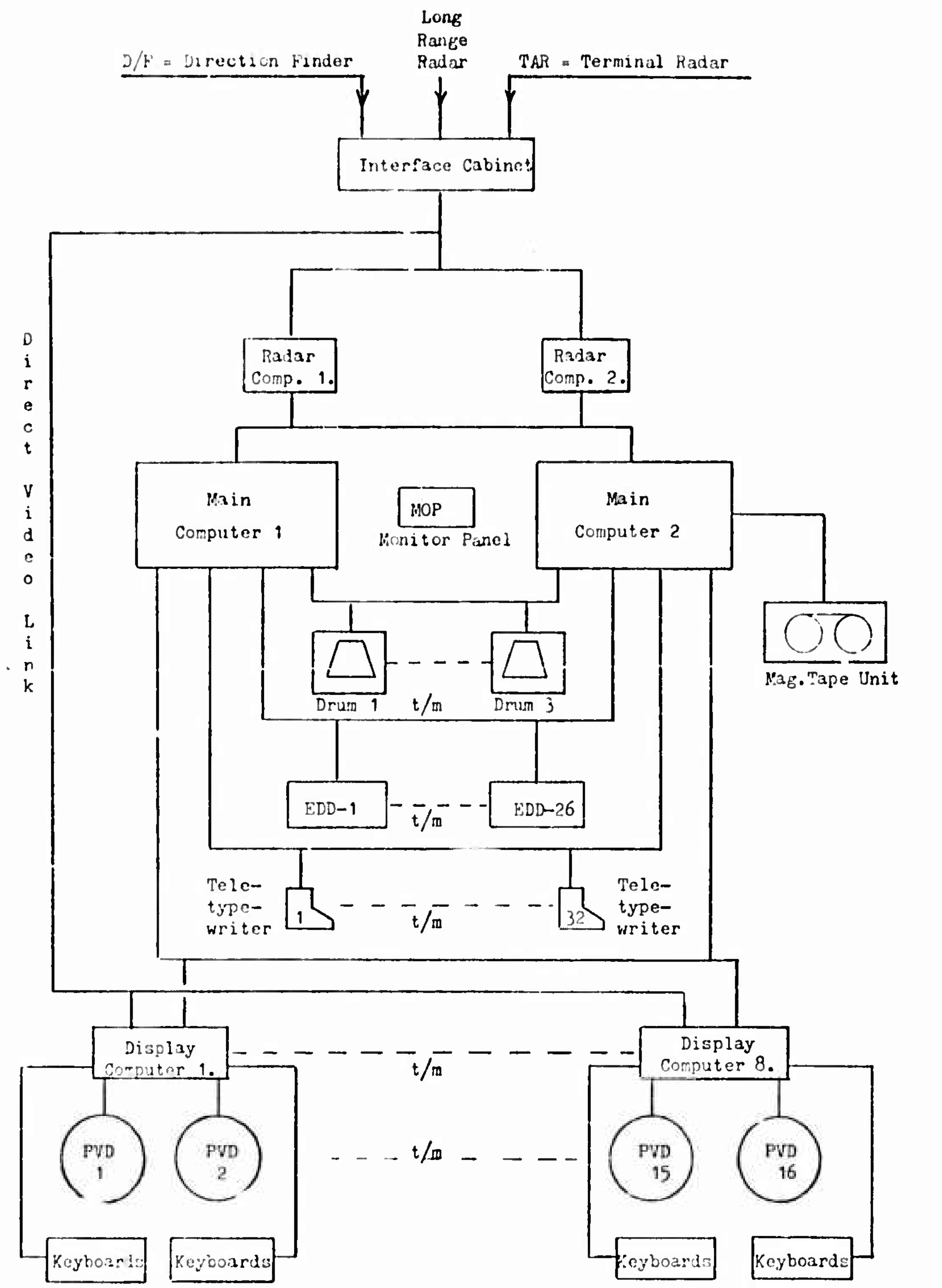


DIAGRAM OF THE SARP II SYSTEM

When a flight is entered in the system it will be assigned a category. This will be done by using the route information (especially the places of departure and destination) as entries to a matrix in which all flightcategories are assembled.

From \ To	Schiphol	airfield inside TMA	airfield outside TMA	Boundary exit point
Schiphol				
airfield inside TMA				
airfield outside TMA				
boundary entry point				

Each category is related to an under control sequence. As an example the under control sequence for a flight entering the control-area at a boundary entry point and going to the main landing runway of Schiphol is:

- en route control by an Entry Sector controller.
- Stack control by a Stack controller.
- Terminal area control by a Feeder and Director.
- Tower control by an Inbound controller and Ground controller.

From the route information will be determined which one out of five Entry Sector controllers and which one out of three Stack controllers will be concerned with this flight. These controller functions will be added to the under control sequence of the flight.

Once the under control sequence has been made up it can be used as an entry to the so called time/action diagram of that flight category. In this diagram have been laid down the moments on which information has to be distributed to all the controllers which are mentioned in the under control sequence. These moments can be determined by certain times (e.g. a given number of minutes before ETA Boundary RP) or by a for this flight significant action which has to be taken (e.g. after a SLOT input has been made to plan the fligt in the landingsequence for the main landingrunway).

Because for all flight the under control sequences and the time/action diagrams are known it is possible to state a few general rules for display and erasure of flightinformation.

As far as the EDD's are concerned flightstrips will be displayed at the moments mentioned in the time/action diagrams. Erasure of strips on a given EDD will be initiated by an under control input of a subsequent controller in the under control sequence. We have intentionally used the term subsequent controller instead of next controller in the under control sequence. This allows the flexibility to neglect one or more controllers in the under control sequence. The usefullness of this feature can be shown in the before mentioned example of the control of an inbound Schiphol flight. The task of the Feeder in the TMA control will be the guidance of flights from one of the three holding positions to a vector area which belongs to the landing runway in use. In certain combinations of holding positions involved and runway in use the distance from the holding position to the vector area could be so short that the flight will be handed over to the Director without involving the Feeder. Taking the flight under control by the Director implies that the stripinformation on the Stack controllers EDD and the Feeders EDD will be removed at the same time.

The genaral rules for display and erasure of label information on PVD's are even more simple than the EDD rules. Only those PVD controllers who belong to the under control sequence of a given flight will be provided with labelinformation during the time that the position symbol f that flight will be displayed on their PVD. Only in those cases that no relation between the aircraft position and the flightplan can be made, the controllers will get the labelinformation in a minitable at the predetermined times mentioned in the time/action diagram. In the latter case the labelinformation has to be removed from the screen by manual input.

These rather simple rules together with the time/action diagrams prevent controllers from being overloaded with irrelevant information.

Conflict search

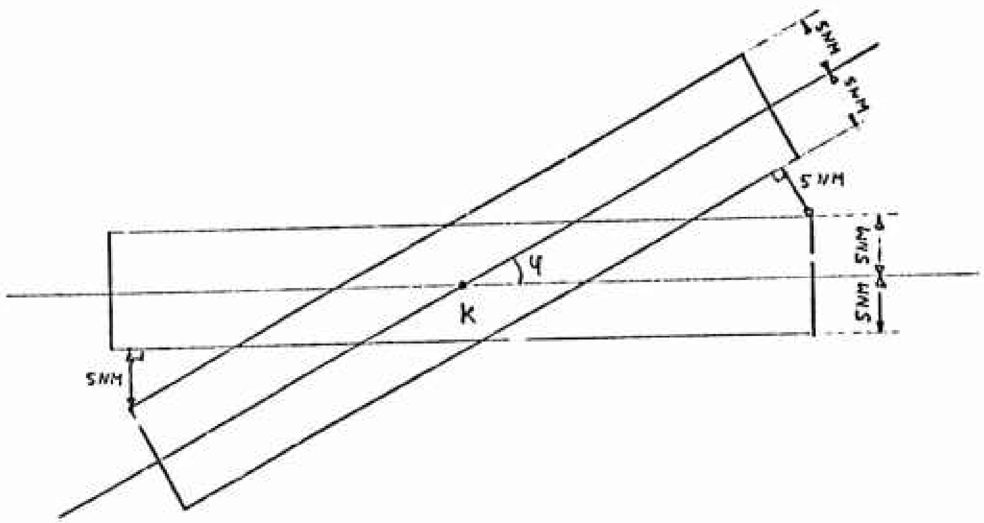
SARP-2 contains two conflict search programmes, one for boundary positions and one for aircraft transiting via de positions where the five ACC sectors join. Thus, the conflect search programmes assist in the inter-centre and inter-sector coördination.

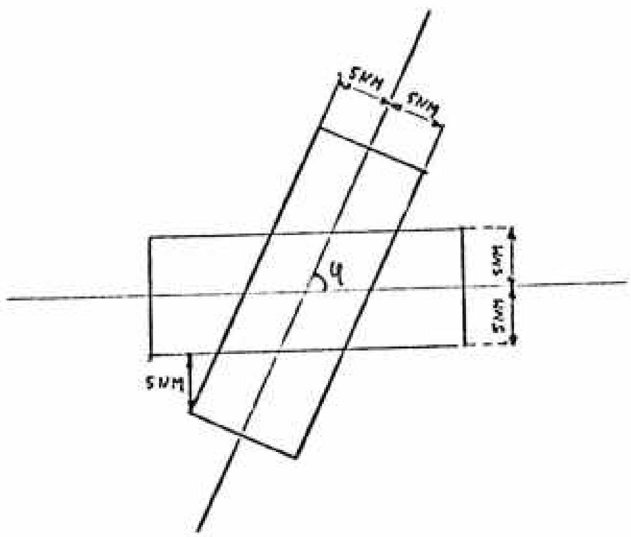
The separation minima applied in the latter case are five NM and 1000 feet.

However, track deviations of 5 NM are accounted for; consequently routes which are separated less than 15 NM may cause conflicts. The crossing point of two routes, or the point at which two routes are separated less than 15 NM is called a junction. At each junction of two routes, two conflict-blocks are defined with a width of 10 NM (5 NM on each side of the nominal route), the length depending on the angle between the tracks of the routes. A conflict exists when and as long as two or more aircraft, which are not separated vertically, are calculated to be present at the same time in the conflict blocks around a junction. Although the system does not contain a conflict resolution programma, it provides a number of facilities to assist the controller in finding a conflict-free path. For the plan controller, the aircraft involved in the conflict are shown on the EDD and, upon request, the programma will indicate conflict-free levels available within 3000 feet from the original level. The executive controller is provided with a means - by a simple light-pen input - to confine the display of labels to those aircraft involved in the conflict.

The conflict search programme at boundary positions applies time and height separation criteria with distance separation for same direction traffic. This conflict search programme is only for use by the plan controller; the assistance for conflict resolution is confined to the display of the conflicting aircraft and, of course, newly inserted re-clearances are subject to the conflict search programme.

The boundary conflict search programme may also be used for the allocation of RETD's (revised estimated time of departure) by the start-up controller of Schiphol TWR. In the case of fixed FL's at the boundary for upper airspace routes, the computer will automatically search for conflict-free boundary "slots" and, from these, calculate RETD's, which are displayed on the start-up controller's EDD. He will then choose the RETD which fits best in the ground traffic-pattern and input the one allocated. Then, the relevant sector display will be updated accordingly. This is just another example of using the programme to minimize the need for verbal coördination.





Implementation

Together with SARP 2 most associated systems and equipment will be renewed, such as the CCTV (with digital data display of certain MET data, such as RVR, received automatically) and the internal/external operational telephone system with facilities such as direct access break-in, conference switching, inter-centre-dialling etc. Much attention is also paid to console design, lighting and central monitoring of the technical systems.

The change-over process will be lengthy. Firstly, it is complicated by the use of parts of SARP-1 (in fact all SARP-1 equipment exept for the display sub-system) in SARP-2. This requires a phased built-up, as SARP-1 will continue to be in operational use during the change-over process. Secondly, the operational trials and familiarisation of control staff require an elaborate programme; it is intended to link SARP-2 with the radar simulator, which will make it possible to carry-out complete trial an training exercises with simulated traffic.

In addition to the operational system, an off-line test and programming system will be available, containing one main and one mini-computer, plus separate EDD and PVD displays and, of course, standard peripherals required for programming services.

3. FURTHER DEVELOPMENT AND CONCLUSION

Work is already in hand for a number of further developments. One of these is CAAS, computer assisted approach sequencing. Another is the adaptation of the present ATC radar simulator to the SARP-2 system to provide for adequate training facilities and minimize the need for on-the-job training. SARP-2 will also be made responsive to the ICAO specifications for ground-ground automated data interchange.

SARP-2 is the result of close collaboration between the Department of Civil Aviation (Directorate for ATS and Telecommunication), the Netherlands National Aerospace Laboratory and the Industry. It will provide a flexible and modern basis for air traffic control into the eighties and, we hope, prove to be a worthy successor to its poineer predecessor SATCO.

OVERVIEW OF US AIR TRAFFIC CONTROL SYSTEMS

by

Howard J.Kirshner

Technical Director, Air Transportation Systems Division
The MITRE Corporation
1820 Dolley Madison Blvd
McLean Va., 22101, USA.

TABLE OF CONTENTS

OVERVIEW OF U.S. AIR TRAFFIC CONTROL SYSTEM

Howard J. Kirshner
Technical Director Air Transportation Systems Div.
The MITRE Corporation
1820 Dolley Madison Blvd.
McLean, Va., 22101, USA

SUMMARY

The purpose of this paper is to provide a brief overview of the technical features of the systems used in U.S. air traffic control centers. Emphasis is placed on the recently modernized en route and terminal control systems.

INTRODUCTION

The U.S. air traffic control system is operated by the Federal Aviation Administration (FAA) of the Department of Transportation. Within the domestic United States, control is exercised from twenty-one FAA en route control centers and about 370 FAA control towers. A network of approximately 90 long-range primary and secondary surveillance radars provide data to the twenty-one en route centers, and about 140 of the towers have radar approach control facilities associated with them. The airways serviced by the domestic en route system encompass almost 200,000 miles of high-altitude routes and about 178,000 miles of low altitude routes.

Flight plan information flows on-line into the en route system from 315 Flight Stations serving general aviation aircraft, and on-line from military base operations offices. Pre-stored magnetic disc files provide the flight plans for scheduled commercial flights. This flight plan data is then automatically distributed to the responsible sectors within the concerned en route centers and to the concerned terminals.

All domestic U.S. en route centers[1] and sixty-one busy terminal radar approach control facilities are now equipped with modern electronic data processing and display equipment. The equipment and associated software provide controllers with an accurate representation of the current air traffic situation. Terminal and en route systems are netted together so that flight plan information is automatically transmitted by digital transmission between terminal and en route centers, between terminals, and between en route centers. Similarly, radar/beacon track handover between facilities is automatically accomplished via computer-to-computer digital communications. Additionally, a central flow control facility, which also has data processing support, is automatically fed flight planning information from the en route centers so that national flow patterns may be adjusted as needed.

ARTS (Automated Radar Terminal Equipment)

The sixty-one busy terminals are equipped with a configuration of data processing and display hardware and software called ARTS for Automated Radar Terminal System. ARTS' basic function is to superimpose, on the controller's radar display, alphanumeric data blocks containing aircraft identity, altitude, and ground speed. ARTS is used principally by the Approach/Departure function in the terminal control facility. A special daylight viewing version of the ARTS display is also used in the control tower cab.

Figure 1 shows an ARTS display. Both search radar and ATC radar beacon data are displayed on the Plan Position Indicator (PPI) in the normal manner. Alphanumeric data are superimposed on the PPI by positioning the beam to write alphanumerics during the dead time of each radar sweep (i.e., after the time required to display the maximum range of the radar).

A block diagram of the ARTS system is shown in Figure 2.

Beacon video, containing a digitally encoded identify and altitude message, and antenna azimuth are quantized by the "Data Acquisition Subsystem." The "Data Processing Subsystem" then determines the center azimuth of each beacon target reply train and performs an automatic tracking function. The automatic tracking associates aircraft flight number with particular beacon replies and positions an alphanumeric data block alongside its corresponding beacon video target reports. The tracking process also computes ground speed for each track to permit monitoring by controllers of speed control instructions or regulations.

Figure 3 shows the configuration of equipment used at Chicago's O'Hare Terminal. At this location, two radar/beacon systems feed the ARTS. In most locations, only one radar/beacon system is used, and only one input/output processor (computer) is required. Similarly, most sites use a lesser amount of memory (24,000 words) than the 40,000 words supplied to Chicago. Processor speed is approximately 210,000 instructions per second, and basic memory speed is 750 nanoseconds.

The UNIVAC Division of the Sperry-Rand Corporation is the prime contractor for ARTS, supplying the computers and computer programs; the Burroughs Corporation provides the "Data Acquisition" subsystem, and Texas Instruments, Inc. provides the "Data Entry and Display" subsystem.

[1] In 1975, twenty en route centers will be operational with both flight plan processing and radar data processing.

As of the end of 1974, sixteen centers will be completely operational with radar and flight data processing. Four are operational with flight data processing only. One center with radar processing only will be phased out.

ARTS has been designed as a modular system to provide for future additions of such system features as:

- Modular redundancy – the ability to sense failures and automatically replace failed units with spare units
- Radar Tracking – the ability to perform automatic tracking of search radar data to enhance beacon data in fades, and to track non-transponder-equipped aircraft
- Metering and Spacing – the provision of control instructions to the controller to permit him to sequence approaching aircraft to obtain more optimum utilization of a runway.

ARTS computers are interconnected with the en route systems computers so that flight information on arriving aircraft (e.g. time estimate at arrival fix) is automatically transmitted from the en route center to ARTS. Similarly, automatic target acquisition in the terminal is accomplished by transmission of track position and velocity when handover is to take place. Additionally, the en route center provides ARTS with a list of scheduled departures so that ARTS may automatically acquire departing aircraft. Upon acquisition of a departure target by ARTS, a departure message is automatically generated to activate the flight plan in the responsible en route center with the actual time of departure. Track handover is also accomplished by a transmission of position and velocity when control is transferred from the terminal to the en route center.

NAS EN ROUTE

The FAA has provided data processing and display equipment, and computer programs, in twenty domestic en route centers. This system is called "NAS – En Route – Stage A – or more familiarly – NAS (for National Airspace System). A block diagram of the system is shown in Figure 4.

Radar Processing and Display

As in the terminal, the principal control tool in en route centers is search radar and ATC radar/beacon. To provide the needed radar and beacon coverage over the large area covered by an en route center, some centers are scheduled to be serviced by as many as ten radars and radar/beacons. Thus, a major functions in the semi-automation of the en route system, or NAS, is the processing and display of search radar and radar/beacon data and the correlation of these data with aircraft identity. This is used to form a pictorial display of the air situation in each sector[1] of the en route center.

The pictorial display shows a digitally generated picture of search and beacon radar data; the identification of aircraft as well as their reported and assigned flight-levels; certain attention-getting information – such as that identifying a newly handed-over track; map information for geographic reference; and certain tabular information such as lists of departing aircraft and their assigned atlitude on lists of aircraft holding at fixes and their altitudes.

Typical NAS sector equipment is shown in Fig. 5.

Flight Plan Processing

Flight plan processing is a major function performed in the NAS en route system (but not in the ARTS system), and it requires a considerable amount of the system's capacity. Each pilot who intends to fly under instrument flight rules (IFR) must file a flight plan containing, among other items, the aircraft identity, type, speed, cruising altitude desired, departure time, and route of flight. This information is entered into the NAS en route system either from previously stored magnetic discs for most airline flights, or on-line from flight service stations or military operations offices for general aviation and military IFR flights.

The flight plan becomes active on receipt of a departure time from the terminal. This is automatic in the case of an ARTS terminal. If a flight is to traverse more than one center's area, the flight plan is automatically transmitted to the next center twenty to thirty minutes prior to its scheduled time of transfer to the next center.

The NAS computer program is designed to take the route of flight information from each flight plan and print, at each sector through which the flight will pass, at least one "flight progress strip." Based on the route information in the flight plan, as well as the departure time and speed data, a time estimate over key traffic control points in each sector is computed and printed on the "flight progress strip." The route of flight is also printed on the strip. Thus the "flight progress strip" may be used by the controller as a planning aid, since it indicates the future intention of each aircraft which will traverse his sector. The flight progress strip is also used by the controller as a handy means of recording changes to the flight route or altitude.

Within the computer program, the flight plan file is updated automatically by the automatic tracking programs so that future time estimates of position can be calculated, and wind or speed errors corrected.

After a strip has been printed at a sector, Controller Updating Equipment (CUE), which consists of a display and entry keyboard, is used to pass new information, such as updated time estimates, from control sector to control sector. The CUE can also be used for entry of data such as route revisions and altitude reports.

NAS Computer Program[2]

The operational program provides the set of computer instructions required to satisfy the air traffic control functions. Figure 6 is a simplified flow diagram of the system monitoring and control function which shows how the various elements of the computer program are related. This function deals with system control, system performance monitoring, and control of the real time subprograms.

[1] A control sector is a subdivision of airspace in an en route center. In future busy centers, there may be as many as 80 such sectors.

[2] The material on the NAS Computer Program was extracted from "Use of Computers in Air Traffic Control." Federal Aviation Administration, June, 1970.

As seen from the figure, this control function is composed of six major subprograms. During normal system operation, only three of these control subprograms are active: namely, the master operational control, confidence checking, and recovery data recording. The remainder are called in only when failures are detected or errors occur during the system operation.

The master operational control subprogram provides for the scheduling of the ATC functions and for the servicing of input/output data and initiates the required action in response to interrupts that are received. It also provides a capability of timing analysis which calculates time information relative to system performance. Further, this control subprogram provides the recording control for the legal and analysis data recording function.

The programming approach permits the operation of two or more compute modules working simultaneously on the task assignments (this operation we term multiprocessing). The scheduler assigns a set of tasks for one compute module and another set of tasks for the other compute module(s). These tasks are classified as either routine, which requires that they be operated on periodically; or high-priority, which requires the computer subsystem to operate on them within a specified minimum time following the request. It is the responsibility of the scheduler to insure that the routine tasks are sequenced in the correct order and that the high-priority tasks are completed within the required minimum time response.

The confidence checking subprogram dynamically monitors the quality of the data inputs and maintains a current record on the availability and quality of the connected equipments and data paths to the CCC.

The error analysis subprogram analyzes errors that are detected, determines their significance to the total operation, maintains a summary of the errors, and decides on the appropriate course of action. If an error has been classified as equipment failure, the faulty unit will be isolated from the rest of the system and control is passed over to the reconfiguration subprogram.

The reconfiguration subprogram accepts requests for reconfiguration from the error analysis subprogram or from a manual request. When the request is made as a result of a failure condition, the reconfiguration subprogram replaces the failed module with a redundant module of the same type and control is then passed over to the startover subprogram.

The startup/startover subprogram is entered during the initial system initiation and during a system restart due to a transient error or a reconfiguration. Its primary function is to establish initial conditions for the central computer complex and, during a system restart, to expand portions of the recovery data into a complete data base. Control is passed from this point back to the master operational control.

The recovery data recording subprogram provides for the recording of essential portions of the total data base used by the operational program to permit automatic recovery of the system operation following a failure. Types of information involved in recovery data would be such items as flight plan data, sector control responsibility, track positions, and sector/display console pairing.

ATC Functions: The ATC functions 1 to 4 shown in the diagram provide for the acceptance, storage, and processing of all flight data information and the automatic transfer of information between facilities. Functions 5, 9, and 10 generate computer responses to controller-generated requests and process information required for display, supervisory, and other positions. Functions 6 to 8 accommodate the processing of primary radar or beacon radar data and the automatic tracking of aircraft to maintain flight identification with the appropriate radar returns.

The adaptation function provides the required parameters for all of the above functions and is the means by which one common program can be used at different geographical locations.

Each of the ATC functional tasks is comprised of many individual subprograms. The modular programming technique, employed in the subprogram design, enables the agency to upgrade any of the operational functions by replacement with more advanced algorithms at a later date and with a minimum of effort. This modular programming design has also been extended to the data tables which have been centralized into a common pool of information independent of the main body of program instructions.

NAS Equipment[1]

The NAS En Route system which has been implemented provides both flight data processing (FDP) and radar data processing (RDP) functions. The system consists of several subsystems of equipment:

Radar Data Acquisition and Transfer
Communications
Central Computer Complex
Computer Software
Data Entry and Display
System Maintenance Monitoring

The Radar Data Acquisition and Transfer Subsystem (RDATS) consists of primary and beacon radars, Common Digitizers (CDs) provided by the Burroughs Corporation, Weather and Fixed Map Units (WFMUs) provided by Tasker, Inc., and links for transmitting the data to the en route center and to Air Defense Facilities for FAA/USAF joint-use radar sites.

The Central Computer Complex (CCC) consists of general purpose data processing equipment supplied by the IBM Corporation. One model of the CCC, the 9020A equipment, has been installed at 11 en route centers, and another model, the 9020D, is installed at the 9 largest en route centers. Both configurations are modular, having multiple compute, storage, and input/output elements.

The operational software, also provided by IBM, resides within the CCC. The CCC receives radar data (aircraft targets and weather) from the RDATS, flight plan information over teletype lines from Flight Service Stations and Military and Airlines operations; and computer-to-computer track and flight plan data via 2400 bit per second duplex lines from adjacent en route centers and from Automated Radar Terminal Systems (ARTS). The processed flight and radar data are provided to the Data Entry and Display Subsystem from the CCC for presentation to the controllers.

[1]The material on NAS equipment was extracted from "En Route Automation" L. G. Culhane, The MITRE Corp., Oct. 1972.

The Data Entry and Display Subsystem (DEDS) consists of a display channel, flight strip printers, input/output typewriters, and computer update equipment. The display channel includes digital data processing equipment and display consoles. The devices at each radar position include a 20-inch diameter plan view display (PVD), computer readout display (CRD), alphanumeric keyboard, trackball, and keys for selecting display data and for entering data. The digital computing portions of two different display channel equipments are provided by two companies: Raytheon Company and the IBM Corporation. The Raytheon system, called the Computer Display Channel (CDC), is a general/special purpose mix of equipment and software. The IBM system, identified as the Display Channel Complex (DCC), is comprised of a 9020E computer. Both systems are modular (up to 120 plan view displays) and have automatic reconfiguration capabilities. The display generating and PVD equipment for both the CDC and DCC was provided by the Raytheon Company and is identical for both systems. The DCC (IBM 9020E) is installed at the 5 largest centers, and the Raytheon CDC is installed at the remaining 15 en route centers.

The display channel equipment provides radar and selected flight data on the PVD and CRD. For each tracked aircraft, identity and other pertinent control data are displayed on the PVD. Radar data are displayed, with data from up to five previous scans automatically presented at a lower brightness than that of the current radar data. Weather contours, showing areas of heavy and severe weather, and map data are also displayed on the PVD.

The D and A sector positions are provided with Raytheon Computer Update Equipment (CUE), and a flight strip printer built by IBM. The CUE provides a means for the controller to interface with the CCC for receiving, entering, and updating flight plan data. It consists of a small cathode ray tube, data entry keys, and an alphanumeric keyboard. The flight strip printer outputs the strip data generated by the computer program.

The System Maintenance Monitoring Console (SMMC), manufactured by Electronics Laboratories, Inc., provides a continual display of status information to the Systems Engineer for all of the hardware subsystems, and allows the Systems Engineer to input hardware configuration changes.

OTHER U.S. EN ROUTE AND TERMINAL SYSTEMS

The FAA is currently in the process of procuring a low-cost version of ARTS called ARTS II. This system will be deployed at terminal facilities which have traffic of sufficient volume to qualify for a radar approach control facility, but are not busy enough to warrant the original ARTS equipment. Burroughs Corporation was awarded the contract for this equipment.

The Anchorage, Alaska en route center has very recently been equipped with a hybrid equipment system to meet its moderate capacity needs. This system uses a NAS Common Digitizer to "plot extract" long-range radar and beacon video, and to transmit the extracted video by narrow-band digital techniques to the center. The radar and beacon data are tracked by a UNIVAC ARTS computer, which also prepares this data and associated alphanumerics for display on Raytheon NAS Plan View Displays.

Another unique en route center in Great Falls, Montana has been in operation since 1963. It operates in a joint facility with the U.S. air defense SAGE system and makes use of SAGE data processing and displays for ATC purposes.

THE FUTURE

The systems described have evolved from earlier experimental, field trial, and operational U.S. ATC data processing systems. Currently, the United States has installed the largest netted ATC data processing system in the world. This system, whose backbone is the general purpose digital computer, provides the base from which more advanced generations of ATC data processing can evolve.

For example, the FAA and its contractors are developing computer program algorithms for eventual inclusion in NAS and ARTS which will:

- warn controllers of potential accidental violation of radar separation standards by controlled aircraft
- warn controllers of potential accidental descent below safe minima by landing aircraft
- detect potential conflicts between planned flight routes
- provide controllers with aircraft sequencing and spacing instructions to maximize airport approach route and runway utilization
- automatically provide routine control instructions to aircraft via digital data links
- automatically provide proximity and collision warning information to aircraft via digital data links

The current NAS and ARTS systems are, then, only the beginning of a new generation of automation and semi-automation in ATC leading to safer and more efficient air traffic control operations.

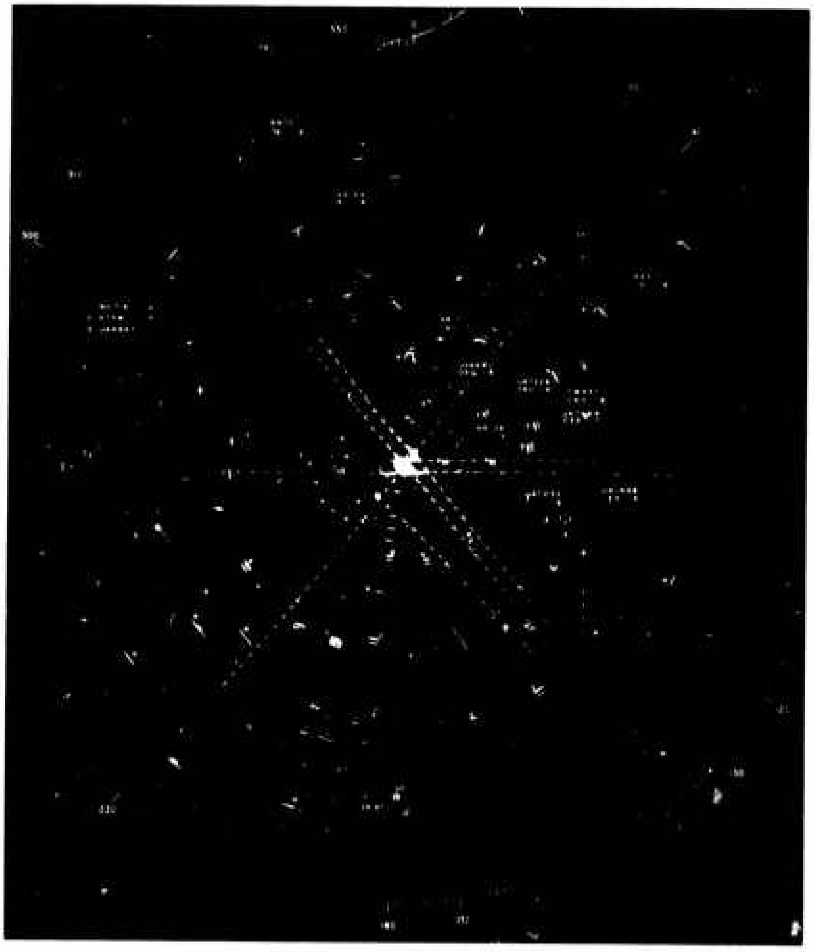


FIGURE 1
ARTS DISPLAY

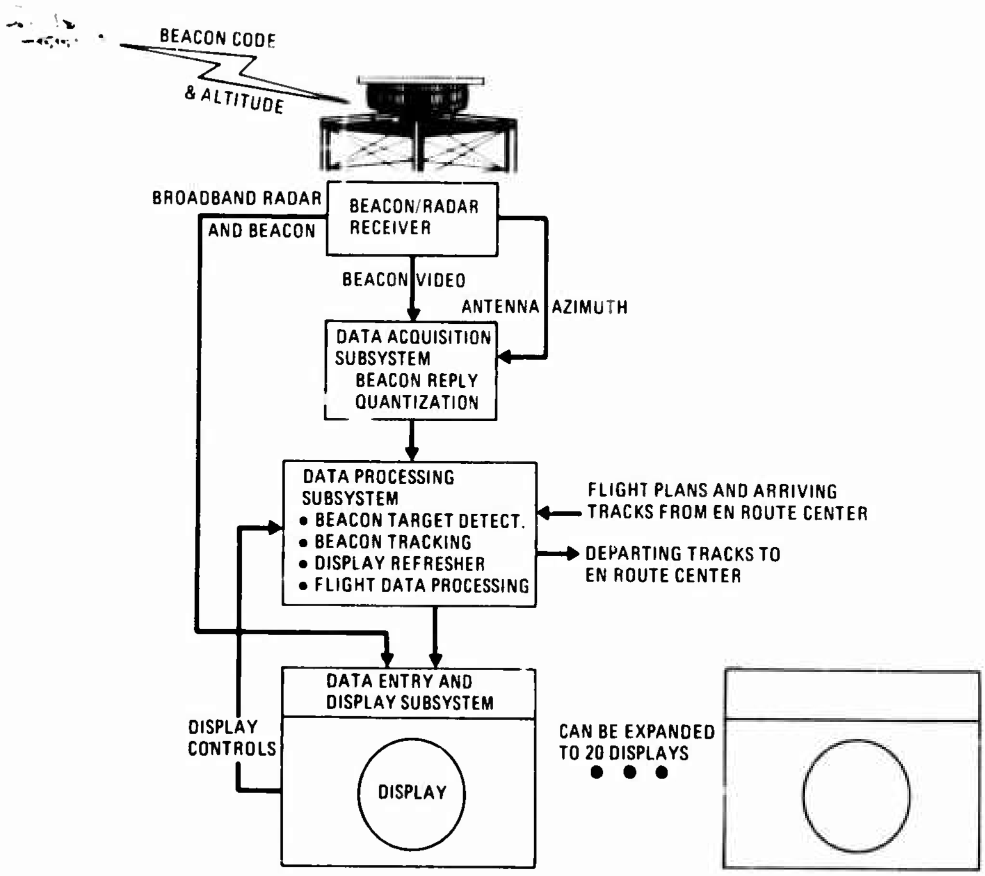


FIGURE 2
ARTS BLOCK DIAGRAM

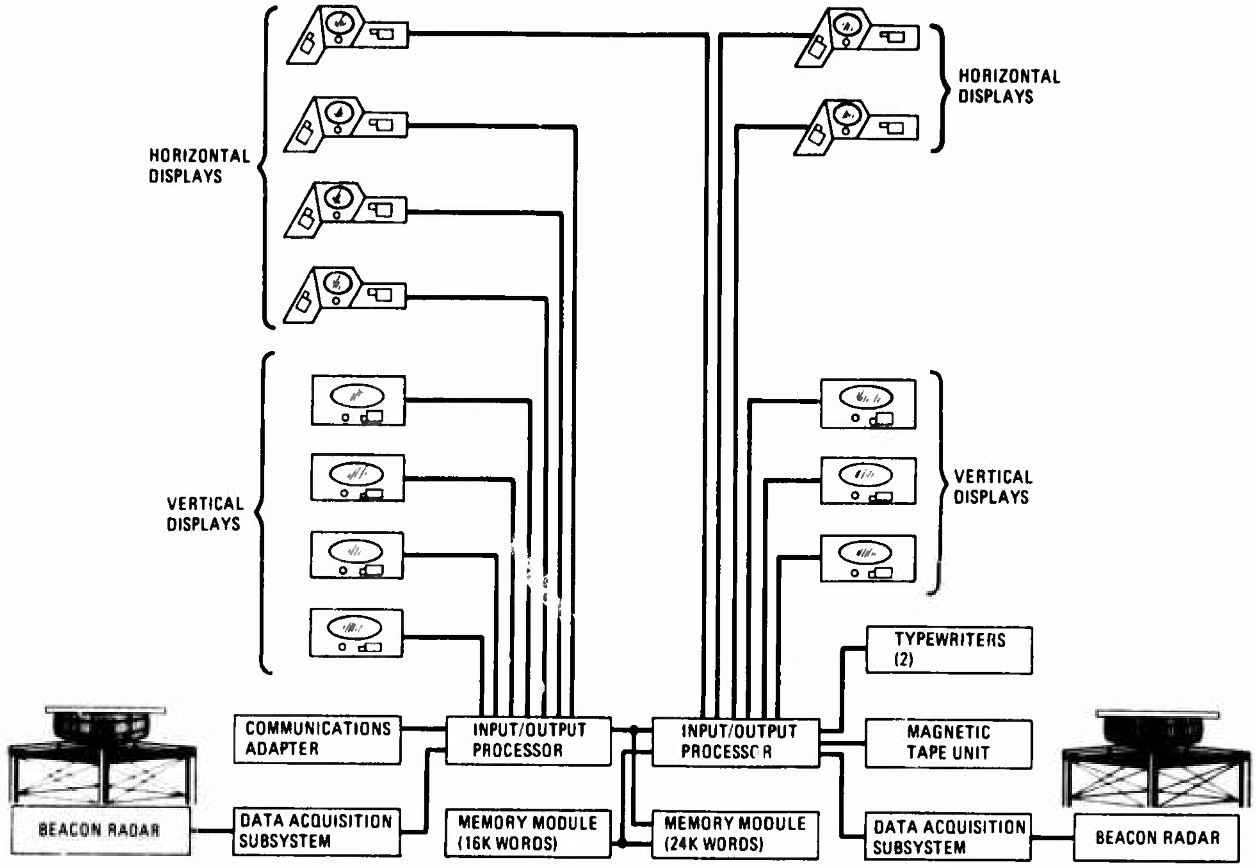


FIGURE 3
CHICAGO-O'HARE ARTS CONFIGURATION

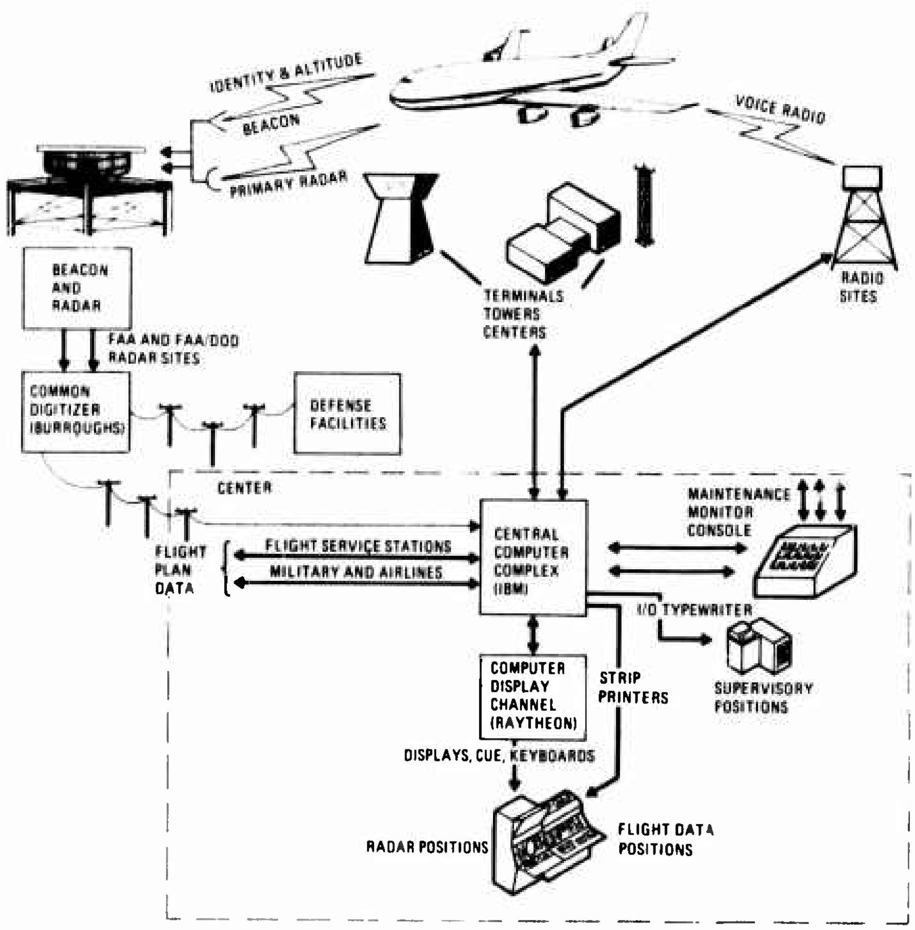


FIGURE 4
NAS EN ROUTE SYSTEM

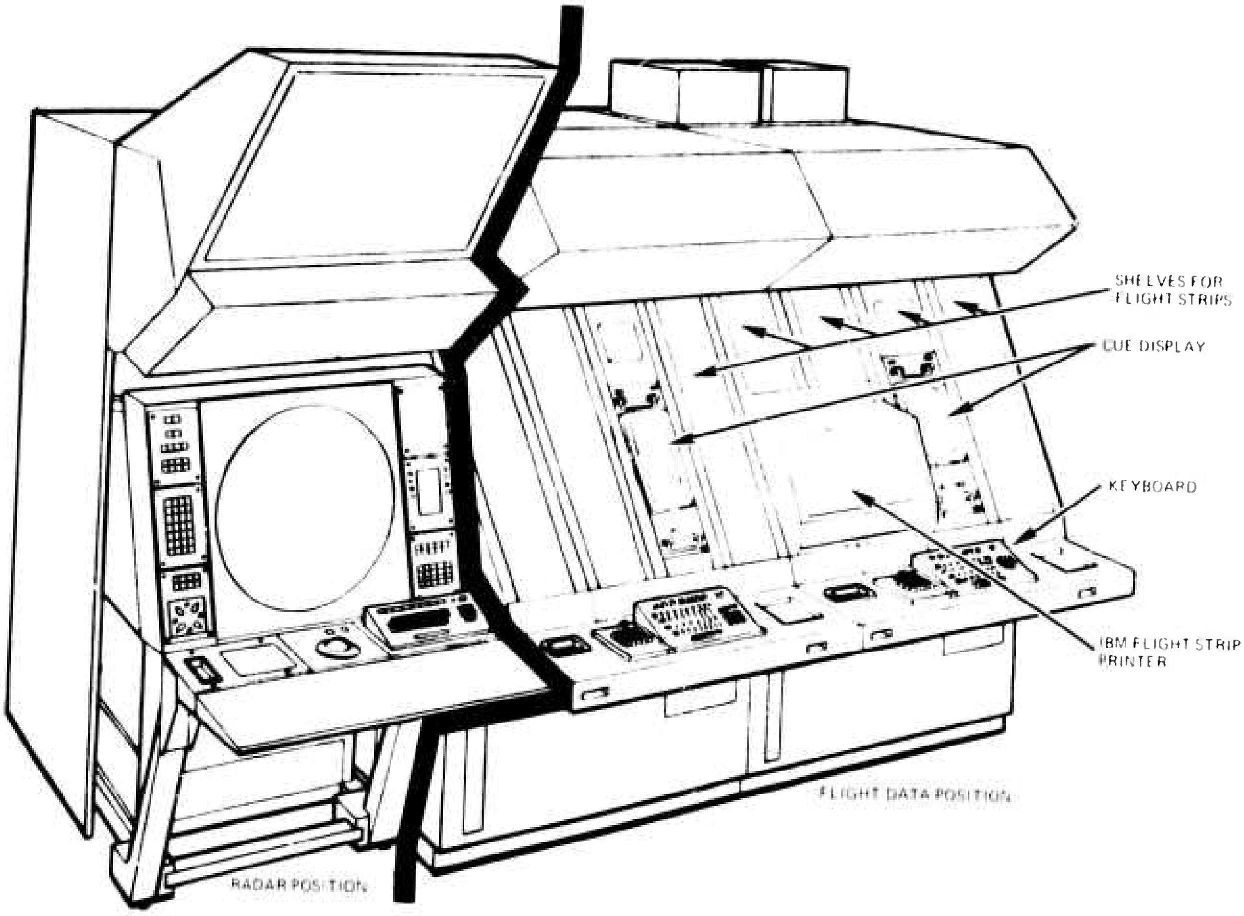


FIGURE 5
NAS SECTOR EQUIPMENT

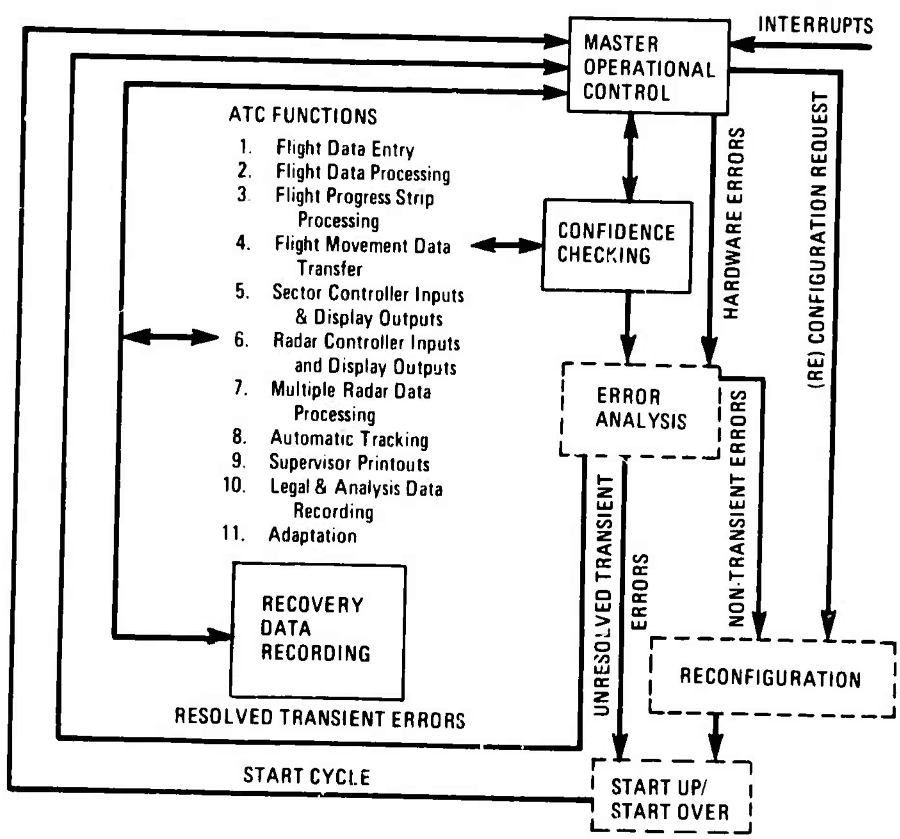


FIGURE 6
SYSTEM MONITORING AND CONTROL FUNCTION

BIBLIOGRAPHY

1. The MITRE Corporation, SATIN: SAGE Air Traffic Control Integration, SR-16, July, 1961.
2. Federal Aviation Agency, Report of Task Force on Air Traffic Control–Project Beacon, October, 1961.
3. Federal Aviation Agency, Design for the National Airspace Utilization System, June, 1962.
4. "An Application-Oriented Multiprocessing System," IBM Systems Journal, 6, 2 (1967).
5. Dept. of Transportation, Report of Department of Transportation Air Traffic Control Advisory Committee, December, 1969.
6. The MITRE Corporation, Genealogy of Terminal ATC Automation, M70-9, McLean, Virginia, 1970.
7. The MITRE Corporation, ATC Operational Computer Program Description, MTR 4429, 1970. (Also published by FAA, SPO-MD-309).
8. IEEE, "Federal Airways System," 1970.
9. "Automated Air Traffic Control," Sperry Rand Engineering Review, Sperry-Rand, Corporation, 24, 2 (1971).
10. The MITRE Corporation, Displays in the National Airspace System, MTP-349, 1971.
11. The MITRE Corporation, System Description–Automated Radar Terminal System (ARTS III), 1971. (Also published by FAA as SPO-MD-600.)
12. The MITRE Corporation, En Route System Description, MTR-1676, 1971. (Also published by FAA as SPO-MD-109.)
13. Raytheon Co., "Interactive Graphics Displays for Air Traffic Control," Electronic Progress, XIII, 3 (Fall 1971).
14. FAA, ATS Fact Book, December 1973.
15. The MITRE Corp., En Route Automation, Oct. 1972.
16. FAA, Use of Computers in Air Traffic Control, June 1970.

AGARD

NATO OTAN

7 RUE ANCELLE · 92200 NEUILLY-SUR-SEINE

FRANCE

Telephone 722.28.00 · Telex 61176

DISTRIBUTION OF UNCLASSIFIED AGARD PUBLICATIONS

AGARD does NOT hold stocks of AGARD publications at the above address for general distribution. Initial distribution of AGARD publications is made to AGARD Member Nations through the following National Distribution Centres. Further copies are sometimes available from these Centres, but if not may be purchased in Microfiche or Photocopy form from the Purchase Agencies listed below.

NATIONAL DISTRIBUTION CENTRES

BELGIUM
Coordonnateur AGARD – VSL
Etat-Major de la Force Aérienne
Caserne Prince Baudouin
Place Dailly, 1030 Bruxelles

CANADA
Defence Scientific Information Service
Department of National Defence
Ottawa, Ontario K1A OZ2

DENMARK
Danish Defence Research Board
Østerbrogades Kaserne
Copenhagen Ø

FRANCE
O.N.E.R.A. (Direction)
29 Avenue de la Division Leclerc
92 Châtillon sous Bagneux

GERMANY
Zentralstelle für Luft- und Raumfahrt-dokumentation und -information
D-8 München 86
Postfach 860880

GREECE
Hellenic Armed Forces Command
D Branch, Athens

ICELAND
Director of Aviation
c/o Flugrad
Reykjavik

ITALY
Aeronautica Militare
Ufficio del Delegato Nazionale all'AGARD
3, Piazzale Adenauer
Roma/EUR

LUXEMBOURG
See Belgium

NETHERLANDS
Netherlands Delegation to AGARD
National Aerospace Laboratory, NLR
P.O. Box 126
Delft

NORWAY
Norwegian Defence Research Establishment
Main Library
P.O. Box 25
N-2007 Kjeller

PORTUGAL
Direccao do Servico de Material
da Forca Aerea
Rua de Escola Politecnica 42
Lisboa
Attn: AGARD National Delegate

TURKEY
Department of Research and Development (ARGE)
Ministry of National Defence, Ankara

UNITED KINGDOM
Defence Research Information Centre
Station Square House
St. Mary Cray
Orpington, Kent BR5 3RE

UNITED STATES
National Aeronautics and Space Administration (NASA),
Langley Field, Virginia 23365
Attn: Report Distribution and Storage Unit

THE UNITED STATES NATIONAL DISTRIBUTION CENTRE (NASA) DOES NOT HOLD STOCKS OF AGARD PUBLICATIONS, AND APPLICATIONS FOR COPIES SHOULD BE MADE DIRECT TO THE NATIONAL TECHNICAL INFORMATION SERVICE (NTIS) AT THE ADDRESS BELOW.

PURCHASE AGENCIES

Microfiche or Photocopy
National Technical
Information Service (NTIS)
5285 Port Royal Road
Springfield
Virginia 22151, USA

Microfiche
ESRO/ELDO Space
Documentation Service
European Space
Research Organization
114, Avenue Charles de Gaulle
92200 Neuilly sur Seine, France

Microfiche
Technology Reports
Centre (DTI)
Station Square House
St. Mary Cray
Orpington, Kent BR5 3RF
England

Requests for microfiche or photocopies of AGARD documents should include the AGARD serial number, title, author or editor, and publication date. Requests to NTIS should include the NASA accession report number. Full bibliographical references and abstracts of AGARD publications are given in the following journals:

Scientific and Technical Aerospace Reports (STAR),
published by NASA Scientific and Technical
Information Facility
Post Office Box 8757
Baltimore/Washington International Airport
Maryland 21240, USA

Government Reports Announcements (GRA),
published by the National Technical
Information Services, Springfield
Virginia 22151, USA

Printed by Technical Editing and Reproduction Ltd
Harford House, 7–9 Charlotte St, London W1P 1HD