University of Bradford eThesis

SATELLITE MULTIPLE ACCESS PROTOCOLS FOR LAND MOBILE TERMINALS.

A study of the multiple access environment for
land mobile satellite terminals, including the
design, analysis and simulation of a suitable protocol
and the evaluation of its performance in a U.K. system.

Hector Tony FENECH.

submitted for the degree
of Doctor of Philosophy

University of Bradford
Postgraduate School of Studies in
Information Systems Engineering

1987

ABSTRACT.

Title: Satellite Multiple Access Protocols for Land Mobile Terminals.

Author: H.T. Fenech.

This thesis is a study of multiple access schemes for satellite land mobile systems that provide a domestic or regional service to a large number of small terminals.

Three orbit options are studied, namely the geostationary, elliptical (Molniya) and inclined circular orbits. These are investigated for various mobile applications and the choice of the Molniya orbit is justified for a U.K. system.

Frequency, Time and Code Division Multiple Access (FDMA, TDMA and CDMA) are studied and their relative merits in the mobile environment are highlighted. A hybrid TDMA/FDMA structure is suggested for a large system.

Reservation ALOHA schemes are appraised in a TDMA environment and an adaptive reservation multiple access protocol is proposed and analysed for a wide range of mobile communication traffic profiles. The system can cope with short and long data messages as well as voice calls.

Various protocol options are presented and a target system having 100,000 users is considered. Analyses are presented for the steady state of protocols employing pure and slotted ALOHA and for the stabilty of the slotted variant, while simulation techniques were employed to validate the steady state analysis of the slotted ALOHA protocol and to analyse the stability problem of the pure ALOHA version.

An innovative technique is put forward to integrate the reservation and the acquisition processes. It employs the geographical spread of the users to form part of the random delay in P-ALOHA.

Finally an economic feasibility study is performed for the space-segment. For costs of capital (r) less than 23 % the discounted payback period is less than the project's lifetime (10 years). At r = 8 % the payback period is about 5.6 years, while the internal-rate-of-return is 22.2 %. The net present value at the end of the projects lifetime is £M 70 at r = 8 %.

lil missieri u 'l ommi.

ACKNOWLEDGEMENTS

The author would like to express his appreciation to Prof. P.A. Watson and to Prof. J.G. Gardiner.

The author is also very grateful for the numerous informal discussions with and assistance from my collegues. These include J.P.V.P. Baptista, E.M.E.M. Carrapatoso, I.A. Glover, Miss A.C. Heathman, D.E. Inglesent and O. Turney (in alphabetical order).

This thesis would have taken much longer without the help of friends which transformed ideas into matter of this form. Claire Hart proof-read the drafts, Beate Marahrens who input most of the text files and last but not least Lesley A. Clow who proof-read the final version, performed most of the editing and was of great assistance on the financial aspects.

CONTENTS

Chapter 1.

INTRODUCTION.

1.1. Scope. 1

1.2. Evolution of Satellite Mobile Communications. . . 1

1.3. Multiple Access. 5

1.4. Performance Appraisal Tools. 8

1.5. Thesis Organisation. 9

Chapter 2.

ORBIT OPTIONS FOR SATELLITE MOBILE COMMUNICATION.

2.1. Orbital Communications Requirements. . . . 12

2.2. System Options. 13

2.3. The Geostationary Orbit. 15

2.3.1. Perturbations and Station Keeping. . . . 15

2.3.2. Application to Mobile Scenario. 17

2.3.3. Propagation and Multipath. 18

2.4. The Elliptical Orbit. 19

2.4.1. Orbit Characteristics. 19

2.4.2. Semiconductors in the Radiation Belt. . . . 20

2.4.3. Molniya Applications. 22

2.4.4. Variant Options of the Molniya Orbit. . . . 25

2.5. Inclined Circular Orbits. 25

2.5.1. Polar Orbits. 26

2.5.2. Efficient Constellations. 30

2.6. Conclusion. 31

Chapter 3.

MULTIPLE ACCESS SYSTEMS.

3.1. Introduction. 38
3.2. Frequency Division Multiple Access. 40
  3.2.1. Non-linear Amplifier. 42
  3.2.2. Degradations. 44
3.3. Time Division Multiple Access. 46
  3.3.1. System Efficiency. 48
  3.3.2. Synchronisation Methods. 49
3.4. Code Division Multiple Access. 52
  3.4.1. Direct Sequence System. 53
  3.4.2. Frequency Hopping System. 55
  3.4.3. Synchronisation. 57
3.5. Multiple Access Techniques for Mobile Systems. . 58
3.6. Conclusion. 64

Chapter 4.

TDMA MULTIPLE ACCESS PROTOCOLS.

4.1. Introduction. 75
  4.1.1. ALOHA Protocols. 78
  4.1.2. Reservation Schemes. 81
4.2. Analysis of the ALOHA Protocols. 87
  4.2.1. Steady State Analysis. 87
  4.2.2. Stability Considerations. 92

Chapter 5.

TDMA PROTOCOLS FOR THE MOBILE ENVIRONMENT.

5.1. Nature of Mobile Communication Traffic. . . . 103
  5.1.1. Voice Communications. 104

5.1.2. Data Communications. 105
5.2. Design of Protocols 107
5.2.1. Frame Structure. 108
5.2.2. The Reservation Subframe. 110
5.2.3. The Information Subframe. 113
5.3. Steady State Analysis of Performance. . . . 115
5.3.1. Assumptions. 115
5.3.2. Glossary. 116
5.3.3. The Reservation Subframe. 117
5.3.4. The Information Subframe. 120

Chapter 6.

PROTOCOL SIMULATION.

6.1. General Considerations. 128
6.2. Development of Simulation Models. 129
6.3. Review of Programming Languages and Software Tools. 133
6.4. Steady State Simulation. 136
6.4.1. Reservation Subframe. 141
6.4.2. Information Subframe. 145
6.5. Stability of the P-ALOHA Reservation Subframe. . 145

Chapter 7.

ASSESSMENT OF PROTOCOLS FOR THE LAND MOBILE ENVIRONMENT.

7.1. Framework. 152
7.2 Performance. 154
7.2.1. Traffic profile. 154
7.2.2. Frame Duration. 155
7.2.3. Reservation Subframe. 155
7.2.4. Information Subframe. 159

7.2.5. Steady State Simulation Results. 161
7.2.6. Dynamic Control of RSF/ISF Boundary. 162
7.2.7. Acquisition and Synchronisation. 165
7.2.8. Link Budgets. 168
7.3. Cost Models. 173
7.3.1. The Satellite. 174
7.3.2. Launch Cost. 178
7.3.3. Telemetry, Tracking and Control Station. 179
7.3.4. Revenue. 179
7.3.5. Economic Analysis. 180
7.3.6. Proposed Schedule. 181
7.3.7. Discounted Payback Period. 184
7.3.8. Internal-Rate-of-Return. 186

Chapter 8.

CONCLUSION.

8.1. Multiple Access Framework. 196
8.2. The Protocol. 197
8.2.1. Refinements. 199
8.3. Financial Feasibility Study. 200
8.4. Further Work. 201

Appendix A1.

PASCAL Source Code Producing the Sub-Satellite Trace. 204

Appendix A2.

The Average First Exit Time (FET) in an S-ALOHA Process. 208

Appendix A3.

Steady State Simulation of a Protocol with S-ALOHA. 211

Appendix A4.

Stability Simulation for the P-ALOHA Reservation Subframe. 219

Appendix A5.

Ephemeris of a Satellite in the Molniya Orbit during the Active Period. 223

References. 224

CHAPTER 1.

INTRODUCTION.

1.1. Scope.

This thesis investigates multiple access systems for land mobile satellite communications. The context of the study is that of a domestic or regional service for the U.K. or Europe serving a large population of small mobile terminals.

The work on this thesis has its origin in C.E.R.S. (Communication Engineering Research Satellite), when in 1982 a feasibility study for the payload was carried out by a consortium of universities: Bradford, Essex, Surrey, London, Loughborough, Manchester and Portsmouth Polytechnic, coordination by the Rutherford Appleton Laboratories. The object of this research program was to assess the application of the Molniya orbit for land mobile and business communications.

As a necessary part of the study, the characteristics and implications of the Molniya orbit were considered together with other orbital configurations in order to evaluate their impact and relative merits in a small terminal multiple access system.

1.2. Evolution of Satellite Mobile Communications.

Satellite mobile communications have their roots in point-to-point satellite systems. The first artificial satellite, SPUTNIK I, was launched in October 1957 by the Soviet Union. The satellite transmitted telemetry information for 21 days. The United States performed a similar experiment the following year using EXPLORER I. In December 1958, voice communication was performed via the U.S. satellite Score.

In October 1960, the U.S. Army launched COURIER IB, the first satellite with an active-repeater. This satellite received messages, stored them on magnetic tape and retransmitted them at a later stage in the orbit. It was also the first spacecraft to utilise solar cells.

The first transatlantic satellite communication occurred via TELSTAR I. Launched by the American Telephone and Telegraph Co. (A.T.&T.) in July 1962, it was put in an elliptical orbit with a period of 2 hours 40 minutes allowing simultaneous visibility across the Atlantic for only 10 min.

Up to 1963, all satellites were in non-synchronous orbits. Thus, earth stations had to employ steerable antennae. The reason for such orbits was mainly due to the limitations of the launch vehicles that could not boost satellites in orbits in excess of 10,000 km altitude. However, the concept of a geostationary orbit dates back to 1945 when Arthur C. Clark wrote his article in the "Wireless World". The first geostationary satellite was N.A.S.A.'s (National Aeronautics and Space Administration) SYNCOM II which was launched in July 1963; SYNCOM I being lost at orbit injection.

Following the formation of the INTELSAT organisation in 1964, the first commercial communication satellite was launched in April 1965. This was the Early Bird. Later, it was referred to as INTELSAT I, the first of a series of communication satellites . It is interesting to compare it to the latest version, INTELSAT VI. The Early Bird had a communication capacity of 240 voice circuits or one TV channel, a mass of 38.6 kg B.O.L. (Beginning of the Life-time) and had a life-time of 3.5 years. The INTELSAT VI has a capacity of about 40,000 voice circuits and two TV channels, a mass of 2004 kg B.O.L. and a life-time of 10 years.

Practically all western communication satellites lie in the geostationary orbit. The situation is somewhat different in the U.S.S.R. The first such Soviet satellite, MOLNIYA I, was launched in April 1965 to provide domestic voice and TV communications. It utilises a highly elliptical inclined orbit that makes geographical and economic sense for the Soviet Union. At present, it makes use of the MOLNIYA III-type satellites and employs a similar elliptical orbit. The first Soviet geostationary satellite was the RADUGA, launched in December 1975.

Mobile satellite communications in the civil commercial sphere started with the launch of MARISAT I in February 1976. This was the first commercial satellite intended for maritime communications. Initially the space-segment was operated by an U.S. consortium, MARISAT Joint Venture. In 1979 INMARSAT came into existence and started operating the system, providing maritime communications that are quasi-global (polar areas are not covered).

Before MARISAT, the space communications were used for point-to-point communications mainly for telephone and television services. The earth stations made use of huge antennae and very high power transmitters. This was an attempt to keep the spacecraft reliability high and the mass as low as possible. Mobile communications required a radical move from this philosophy. Early ship terminals made use of antennae that were 1.2 m in diameter. This has now been reduced to 85 to 90 cm while maintaining a minimum G/T of -4 dB/K. INTELSAT systems uses 15 to 30 m antennae and a G/T of 30 to 42 dB/K. (This may be a somewhat unfair comparison since the bandwidths involved are very different).

Prior to the advent of satellite communications, maritime radio made use of M.F., H.F. and V.H.F. However, this does not provide a reliable continuous global system. INMARSAT provides voice, telex and data services via three satellites (plus spares) located over the Atlantic, Pacific and Indian Oceans, thus practically covering the earth. The satellites currently used are a MARSAT, a MARECS and the M.C.S. (Maritime Communications Subsystem) on the INTELSAT V F-5, F-6 and F-7 satellites.

Today INMARSAT remains the only commercial mobile satellite system providing a service to some 4,000 users. Considerable attention is being given to land systems, whilst interest is growing once again in the aeronautical area, following the ill-fated AEROSAT programme in the 1970′s.

A study is being undertaken to explore the extension of the use of navigational satellites to provide a polar communication service (NAVSTAR) employing INMARSAT satellites. The British consortium between British Telecom International RACL/DTI-RAE and British Airways is investigating a pilot system to provide such a service on a few 747´s on the North Atlantic and African routes. There is interest from INMARSAT to diversify and also provide aeronautical and land services.

Several land systems have been proposed including the Scandinavian TRUKSAT, E.S.A.´s (European Space Agency) PROSAT which is supposed to provide all three mobile services, C.E.R.S. and perhaps the most promising one, M-SAT.

M-SAT has been under development by N.A.S.A. and the Canadian Department of Communications (D.O.C.). The system uses a geostationary satellite to cover the U.S.A. and Canada and several spacecraft configurations have been proposed. Multispot beam schemes are envisaged though various numbers of beams have been mentioned in literature. The most common figure for a first generation satellite seems to be five. A spacecraft antenna diameter of about 9 m has been proposed.

The system is designed to supplement the public switched telephone system especially in sparsely populated areas, which are rather large in the two countries and where terrestrial systems would not provide an economic realistic alternative. A number of gateways operating at S.H.F. interface the satellite system to the terrestrial network.

1.3. Multiple Access.

While literature abounds in descriptions of multiple access schemes for satellite communications, there is very little experience with accessing techniques from operational systems in the mobile field. This is particularly so for the mobile-to-base link where several users are competing for resources.

The situation for trunk communication satellite systems is relatively simple and straightforward. Perhaps the best known system is the scheme developed by COMSAT for INTELSAT called SPADE (INTE 85, FEHE 83a, STAL 85). In this system there is a TDMA (Time Division Multiple Access) reservation channel with a frame containing 50 slots. The slots are dedicated to users and that implies that the system is rather inflexible in terms of the number of users and unusable in mobile communications.

As mentioned earlier the only operational mobile system is INMARSAT. The problems encountered in the present maritime environment are considerably different from those in the land mobile scenario. The population sizes are much smaller than those forecast for land mobile systems. The difference may be as many as two orders of magnitude.

INMARSAT utilises a SCPC (Single Carrier per Channel) structure, (LIPK 77, DASI 84) where one channel per region is reserved for requests or call set-up. Thus if a ship wishes to initiate a call, it would generate a request message. This would include the shore station address, priority, ocean area, type of message, terrestrial network and the nature of request. All are contained in 39-bits of information.

All mobile terminals share the same request channel so that overlap of messages is possible. Since the channel is assumed to be lightly loaded such collision of messages is not very likely. If, however, a request is not acknowledged, the terminal waits for six seconds and reattempts. Once the message finally reaches the coast earth station, two possibilities exist:

1. All the working channels may be occupied and the ship earth station is instructed to wait.
2. The coast earth station sends an assignment message which would cause automatic retuning of the ship's equipment.

Once the assignment has been established then communication can proceed but at the end the working channel must be released. It is interesting to note that this is basically the automatic version of normal marine radio operation. The operator listens on the calling channel and if it is available he requests a working channel otherwise he waits three minutes and reattempts. Once a working frequency is assigned the ship operator will switch to the right channel and start his traffic.

Present day requirements put heavy demands on accessing protocols for mobile communications. Since computer networks place similar demands with respect to the population size, most new protocols for mobile communications tend to employ techniques originally developed for computer networks. Of particular importance is the ALOHA system. Developed in the early 1970's it formed the genesis of random access for data communications. It was originally developed for a terrestrial computer network but its ability to perform in the presence of large propagation delays made it very suitable for satellite communications.

ALOHA has the great attraction of simplicity. However its channel utilisation is far from satisfactory. This brought about the use of the ALOHA scheme as an ideal accessing technique for the reservation channel of reservation protocols. The classical one is that developed by Robert's (ROBE 73) which allows great flexibility by having a dynamically controlled portion of the system capacity being used as an ALOHA reservation channel.

1.4. Performance Appraisal Tools.

The classical method of analysing accessing protocols is, of course, by mathematical modelling. The immediate impression of such analysis is that of daunting complexity. Nevertheless mathematical models have been developed that provide a great deal of insight into the design and performance of systems. To a large extent most of these models are based on queuing theory. The field originated from the work of a Danish mathematician called A.K. Erlang who was investigating telephone switching systems. It now forms the basis of operational research.

In its generic form queuing theory provides models for random arrivals of customers to one or more servers. The service time may also be random. Various models are available with various combinations of probability distributions for the arrival rates, the service time and number of servers. The theory attempts to quantify quantities like the size of queues, the delay experienced by customers and the server utilisation.

Queuing theory is extremely useful in a very wide range of applications particularly when the steady-state performance is sought. However it soon becomes mathematically untractable when seeking transient performance of queuing models. This is one application where simulation can be very useful.

With the increasing computing speed and the development of suitable programming environments, simulation is becoming increasingly popular. Simulation involves the development of an algorithm that mimics the system. It normally includes mathematical modelling but a closed-form mathematical solution is not necessary. This provides a very powerful tool that can be used for experimentation with the system and to estimate the performance. Where a mathematical analysis is available, simulation can be used to complement and validate the analysis, thus increasing the figure of confidence.

The development of the algorithm requires a thorough understanding of the system. This invokes the designer to consider practical details that may have not received their due consideration in the formulation of the mathematical analysis. This results in a greater insight into the system and the elimination of some teething problems.

Unfortunately development of simulation software can be time consuming and running it can place heavy demands on the computer resources. Furthermore its application can be rather specific particularly when large systems are considered. This is a great drawback since it implies that the effort put into a specific project is not readily reusable.

1.5. Thesis Organisation.

This thesis identifies the problems of multiple access in mobile satellite systems and discusses the possibilities in the light that accurate forecasts on such a system are not available (ESA 86). Though data is required to evaluate the performance of such a protocol, sufficient flexibility has to be included so that the protocol can perform under other conditions.

Chapter two investigates the relative merits for mobile communications of the three orbit systems, i.e. the geostationary orbit, the elliptical (Molniya) orbit and a constellation of satellites in the inclined circular orbits.

The third chapter studies the three multiple access systems, i.e. frequency time and code division multiple access and analyses their respective merits in the mobile environment.

Multiple access protocols are investigated in chapter four. Various protocols are discussed and the ALOHA protocol is analysed under steady state conditions. The stability problem that is inherent in such schemes is also addressed.

In chapter five an accessing protocol suited for mobile communications is developed and analysed. A mathematical model is derived. The analysis is performed for a TDMA environment. It employs a reservation subframe and an information subframe. The former is an ALOHA channel while the latter is analysed by employing queuing theory.

Chapter six addresses simulation techniques and modelling. It includes a discussion on simulation programming languages. The software design of two discrete event simulation programs is provided. One of the programs is used to increase the confidence level of the slotted ALOHA reservation protocol while the other is used to obtain a stability measure for the pure ALOHA reservation channel.

System implications are studied in chapter seven. It develops the frame structure for a given user population and communication traffic profile. The multiple access protocol is applied to U.K. system and its performance is evaluated. Link budgets are provided and a financial feasibility study is conducted for the space-segment.

Finally chapter eight provides a conclusion and attempts to identify areas where further work is required.

CHAPTER 2.

ORBIT OPTIONS FOR SATELLITE MOBILE COMMUNICATION.

2.1. Orbital Communications Requirements.

In the more conventional use of communication satellites for trunk communication, the situation is such that the number of ground stations is small and the stations are fixed and provide processed traffic which tends to be heavy in nature. Satellites that occupy the same spatial coordinates relative to two communicating stations are highly desirable as this allows the earth stations to have very high gain antennae with very high pointing accuracies.

If there is relative motion of the satellite with respect to the earth stations, then antennae on earth would have to track the satellite. This introduces an added complication, especially if the antennae are large, though it could be accepted, particularly since the relative motion takes place in a predictable manner.

The mobile scenario is considerably different. The number of mobiles is very much larger and mobile equipment almost tends towards consumer products. In this situation a steerable beam is not desirable for two reasons:

1. The equipment required for mechanically steering the antenna is expensive and possibly a source of unreliability. With

electronically steered antennae the gain drops as the beam is steered away from its natural boresight. It also requires a rather complex control circuit and a phasing network. Customised VLSI could be the answer to the control circuitry and microstrip techniques for the phasing network and antennae (JAME 81).

2. With steerable antennae motion, the instantaneous coordinates of the satellites relative to the mobile have to be known. This makes it somewhat difficult in the case of a mobile. The elevations of the antenna may be roughly determined for a mobile expected to be in some restricted area. The azimuth is more difficult to determine. A ship is not very agile and therefore a steerable antenna system can be and is employed. Such systems include the INMARSAT system and the Skynet system with its terminals called Scot (PHIL 85, LAW 85). However, if we consider an automobile system, or hand-held or man-pack systems, then discontinuities in the direction of motion are not at all unlikely. Such conditions would give the tracking system serious problems.

2.2. System Options.

From the above, it can be seen that to have a mobile system in the full sense without putting limitations on agility, three possibilities exist:

1. The satellite is located at zenith so that as the mobile moves about, the satellite is always above the user. This scheme precludes linear polarisation since the mobile should be able to rotate about its zenith.
2. If the satellite is at a given elevation then an antenna with a pattern that is equivalent to an elevated lobe gyrated about the

zenith is required. As the azimuth of the mobile changes relative to the earth, then whatever the orientation, the satellite would always be within the antenna's pattern. Limitations on polarisation are eased.

3. The final option allows the satellite to move but is only used within a certain angle of mobile zenith. The mobile's antenna would have the main lobe at the zenith and the beam-width adjusted so that the desired portion of the orbit centred around the user's zenith is used. This system inherently requires more than one satellite. Only circular polarisation can be used.

Having stipulated these three options we will now see how they can be implemented. In this context it is convenient to identify three classes of orbits:

1. Geostationary orbit,
2. Elliptical orbit
3. Inclined circular orbit.

Very often the choice of orbits is not fully recognised by the communications engineer and the geostationary orbit is assumed.

Of course all orbits characteristics are governed by the same orbit mechanics. In fact the general case would be an inclined elliptical orbit. If the eccentricity is zero we get the inclined circular orbit and by putting the angle of inclination to zero and the appropriate altitude we get the geostationary orbit. To a first order of accuracy the period of the orbit is given by

$$T = 2\pi (A^3 / u)^{1/2} \tag{2.1}$$

where T is the period in seconds,

A is the semi-major axis in km and

u is the gravitational constant equal to $3.99 * 10^5\ km^3/s^2$.

This applies for all orbit types so that for a circular orbit the semi-major is equal to the radius. What determines the type of orbit is the injection point and velocity at which the satellite enters the orbit during its orbital transfer procedure.

2.3. The Geostationary Orbit.

The geostationary orbit represents a special case of the inclined circular orbit but because of its interesting features it will be discussed separately. Most communication satellites are situated on this orbit, which is circular and lies on the equatorial plane (theoretical angle of inclination = 0^o) with a period of one sidereal day which is 23 hours 56 minutes 4.091 seconds, the direction of rotation being prograde. This implies that the satellite sub-satellite point theoretically remains on the same latitude and longitude on earth.

2.3.1. Perturbations and Station Keeping.

The earth contributes towards the major force that determines the orbital mechanics. However, other celestial bodies also contribute to produce secondary effects. The equatorial plane is not in the same plane as the orbital plane of the moon or that of the earth round the sun. This causes a secular perturbation in the inclination of the orbit. The mean rate of change is around 0.85^o/year building up to a maximum of 14.67^o from zero in about 26.6 years (ISLE 74). Since the lunar effect is approximately three times the solar, the exact rate of change depends on the inclination of the lunar orbit to the equatorial plane.

The result of such a perturbation is that instead of having a fixed sub-satellite point on earth on the equator, a figure of '8' is described symmetrically about the equator with a lobe to the north and another to the south (figure 2.1), the larger the perturbation the larger the dimensions of the trace. Varying the period (or altitude) causes the sub-satellite trace not to have such a closed form.

The fact that the earth is an oblate spheroid causes a small change in the orbit period (theoretical mean radius: 42,164.1 km, corrected radius: 42,164.7 km, a change of 0.6 km). The earth's equator is also non-circular. This causes the satellite to drift east or west in the orbital plane. The exact amount of drift and the direction is dependent upon the satellite longitude. The drift is zero longitudes of about 80^{o} E, 160^{o} E, 110^{o} W and 10^{o} W.

It would be desirable to keep all these perturbations to a minimum but, unfortunately, the cost for this is quite high in terms of fuel mass, particularly, the N-S station keeping which is most demanding in this respect. The change of velocity capability would be of the order of 40 - 51 m/s/yr while the E-W corrections would be of the order of 0 - 2 m/s/yr (PRIT 84). The requirements for a 1000kg satellite with a 10 year life-time would be of the order of 300 kg of conventional hydrazine propellant. Fortunately, the E-W station keeping is more important so that in some cases N-S corrections may be dispensed of. This would reduce the propellant mass to about 70 kg.

2.3.2. Application to Mobile Scenario.

In the mobile context, for countries in the tropics, the geostationary orbit provides a zenithal service. In this situation, the diurnal figure of '8' excursion of the sub-satellite point is acceptable. Assuming a life-time of 10 years, the inclination increase is likely to be within 9°. This means that by setting the right initial inclination, the effective angle can be halved. The pointing accuracy of the mobile antenna is unlikely to be very high and therefore the N-S station keeping specifications can be relaxed. For such regions the geostationary orbit offers the best service. The probability area of where the satellite is situated is very small and symmetrical about the zenith, allowing the mobiles to use higher gain antennae.

As the observer moves away from the equator, the angle of elevation drops and the probability area of where the satellite is with respect to a moving mobile increases rapidly. At about 80° latitude the angle of elevation is zero. At this latitude, the satellite can be anywhere in the azimuth of a mobile observer referred to itself and just visible over the horizon. Beyond this latitude the satellite is no longer visible. The desired mobile antenna pattern would be the volume generated by rotating an elevated lobe about the zenith. Such an antenna pattern exists, for example, the dropping dipole and the microstrip disk with a peak gain of about 5 dBi each. However as these antennae are forced away from their broadside pattern, they tend to become less efficient and the gain drops.

2.3.3. Propagation and Multipath.

Another problem related to propagation is that horizontally polarised fields do not propagate well at low angles of elevation. At 0° elevation the electromagnetic wave will be virtually propagating parallel to an equipotential (earth) plane. As a result, the circular polarised signals become elliptical and eventually linear, representing a maximum signal loss of 3 dB. Horizontally polarised signals are severely attenuated to low elevations while vertically polarised signals are virtually unaffected. There are, of course, also problems with rain attenuation. These factors limit the angle of elevation to the minimum of 5° to 10° depending on frequency.

At low elevations, the problems of multipath and shadowing become more severe. Since the angle of incidence on buildings approaches the normal, the building need not be very tall to cause problems even if it is far away. Multipath problems of this type should be quite small since the reflected wave should be somewhat attenuated due to of the high grazing angle. This thesis has been confirmed for land mobiles in the results of the Phase 1 of the Prosat programme where figures of 10 to >17 dB have been quoted for elevations greater than 24° (JONG 86). The major problem in this environment is shadowing which could result in fades in the region of 20 dB in the urban areas. The fade duration can vary from milliseconds to several seconds depending on the vehicle speed.

In the marine environment reflections could originate from the sea surface. Since the grazing angle is so low, the signal strength of the reflection could be quite high particularly in calm waters. The

Prosat report quotes figures for the carrier-to-multipath ratio of 7 - 10 dB at elevations in the range 2^o to 16^o. The figures of merit for the terminals used range from -19 dB/K and -12 dB/K. It should be noted that the multipath performance improves with terminal gain. Aeronautical tests indicate a similar figure at all elevations. The figure of merit for both the land and aeronautical receiver was 24 dB/K.

2.4. The Elliptical Orbit.

2.4.1. Orbit Characteristics.

The inclined elliptical orbit is the general case. However, our main interest here is the Molniya orbit. The earth is the cause for first order effects,while the sun and the moon are the main bodies that cause second order effects such as secular perturbations in ellipticity and angle of inclination. The oblateness of the earth also plays a role. It causes the major axis to rotate in the orbital plane (apsidal rotation) and there is also rotation of the ascending node (nodal regression) (ELYA 67, PRIT 86).

The apsidal rotation becomes zero for an angle of inclination equal to acrcos(sqrt(1/5)) = 63.4^o (figure 2.2). For angles of inclination lying between this angle and its complement, the major axis rotates in the direction of the motion of the satellite. For other angles the motion of the major axis is in reverse. The apsidal rotation is at a maximum when the orbital plane coincides with the equatorial plane. The nodal regression is directly proportional to the cosine of the angle of inclination. These perturbations (rad/rev) are inversely proportional to the square of the latus rectum ($p = a\ (1 - e^2)$, p:

latus rectum, a: semimajor axis and e: eccentricity).

The perigee passes close to the atmosphere. At altitudes above 600 km aerodynamic drag should be quite small. Air drag causes the orbit to describe a convolute elliptical spiral which tends to a circle. At the high altitude of the atmosphere, the density of the air becomes very sensitive to the diurnal and annual time on earth, the location on earth, the rotation of the sun and solar activity. In fact at this altitude the density might vary through a few orders of magnitude.

The use of an inclination of 63.4° reduces station keeping fuel requirements. The apogee altitude is 39,000 km and the perigee is 1,000 km for 12 hour period (WATS 83, DOND 83). The perigee altitude is a compromise between payload mass and atmospheric drag. Figure 2.3 illustrates the sub-satellite trace for such an orbit. (The PASCAL program that produces the sub-satellite coordinates with time for a circular or an elliptical orbit is given in Appendix A1).

2.4.2. Semiconductors in the Radiation Belt.

The low altitude of the perigee implies that the satellite passes through the trapped-proton radiation belt (Van Allen belts). This spans over altitudes of 1,000 and 10,000 km. High energy protons affect the microelectronics and degrade the solar cells.

The two main effects of proton and other heavy ion radiation on semiconductors are the generation of extra carriers and the degradation of the semiconductor by, for example, changing silicon atoms to magnesium (PETE 81). Extra carriers cause drift in the operating conditions of the transistor and causes digital circuitry to change state - soft upsets. These problems increase with decreasing technology size. There is yet another problem. The extra carriers may cause parasitic PNPN junctions acting as SCRs to fire. This may cause heavy currents to flow resulting in damaging localised heating. Two options are available:

1. Sensing circuitry can be included which resets the supply when the supply current exceeds a certain limit.
2. The other option is to use technologies that inherently avoid such parasitics, for example SOS.

Soft upsets occur to some extent in all technologies, though bulky and current hungry technologies tend to suffer less. Latch-ups tend to be technology and lay-out dependent.

Besides these transient circuit malfunctions, there is also gradual deterioration of the chip performance due to nuclear reactions involving alpha particles and spallation. This damage is proportional to the integral of the radiation dose on the sensitive chip volume with time. Radiation doses can be reduced by aluminium shielding. The spacecraft is subjected to high radiation doses when it is close to the perigee. However, since in this region the spacecraft is moving very fast and the most of the on-board electronics need not be on the resulting damage is small.

The total absorbed dose per a 12 h orbit has been estimated (CRAI 83, COAK 84) to be about 10 rads per orbit or about 7 krads per year. The dose at the active part of the orbit is one or two order of magnitude lower than the given figures. These figures have been estimated using the Molniya I data. However, the accuracy of these data seem to be in question since for the geostationary orbit, the data predict a figure that is an order of magnitude lower than most other estimates. Large variances in the doses are also experienced depending on the solar activity and actual doses may vary by a factor of ten.

For a 12 h orbit, the likely net dose can be assumed to be 10 krads per year and if the on-board electronics is switched off when the satellite is below 40^{o} N the relevant dose is less than 1 krads per year.

2.4.3. Molniya Applications.

The Molniya orbit has been fully exploited by the Soviet Union. The apogee at 63.4^{o} inclination is convenient for the USSR whose land-mass lies between 35^{o} and 75^{o} latitude. Satellites in such an orbit can be launched from a higher latitude than that required for the geostationary orbit for which a launch site should be as close as possible to the equator. The Molniya orbit is clearly suited for countries situated in the high latitudes.

This, of course, also covers the polar regions which have often been considered commercially uninteresting. There is, however, much potential. Search and rescue operations need a particularly swift response since survival time in these regions is shorter than in other parts of the world. A hundred flights per week are expected to cross the Arctic region by the end of the decade. The scientific activity in Antarctica accounts for a population of 2,000 to 3,000 in summer. There are also other prospects. Estimates of mineral resources are promising though more advanced technologies are required. There is a large potential for fishing which could yield the greatest quantities for a single species (BERR 84).

The cost of a dedicated launch should be cheaper than a geostationary launch. Comparing the launch sequence to the geostationary maneuver an apogee kick motor is not required and only minor corrections are required after the perigee orbit transfer. If a coplanar transfer orbit sequence is adopted the launch mass can be double that of the geostationary, implying a cheaper launch.

The orbit characteristics have important implications on the communication link. Assuming that out of the 12 hour period only the four hours before and after the apogee are used, the total variation in longitude and latitude of the sub-satellite points is only about 1.5^{o} and 18^{o} respectively. To the observer located near the apogee sub-satellite point, the satellite travels an angular distance of about 23^{o}, so that a fairly narrow beam antenna pointing to the zenith can be used. The probability of having another satellite in the vicinity is quite low and so is the probability of causing interference to neighbouring systems. This implies relaxation on the antenna polar diagram which makes the antenna cheaper. In the mobile environment, circular polarisation would be very attractive since, as the mobile changes azimuth, the link remains unaffected. With linear polarisation the antenna would have to be aligned to the signal.

Four hours before and after the apogee, the satellite is 23,489 km away from the apogee sub-satellite point. This contributes a variation in the free-space path loss of about 4.4 dB. Since during the useful portion of the orbit the satellite is in motion relative to the terrestrial observer, the signal received suffers a Doppler shift.

This is at a maximum at the beginning and end of the 8 hour period and is equal to 8.1 ppm. (At 1.5 GHz this is equivalent to about 12 kHz.) This puts heavier demands on the carrier recovery circuit in the receiver. One way of reducing this problem on the down-link is to include compensation in the satellite local oscillator. The Doppler shift can be predicted from the orbit parameter and the carrier steered accordingly. This compensation can be achieved via the control channel.

The propagation delay is consequently also affected. At the apogee it is equal to 260 ms while at the edges of the 8 hour period it drops to 167 ms resulting in an apparently varying clock-rate at the mobile receiver. In the first half of the useful portion of the orbit, before the apogee, the satellite is moving away from the terrestrial observer and consequently the clock rate would seem to reduce while the opposite effect happens when the satellite passes the apogee. At the mobile end, it is not envisaged that this would create problems since the integrated effect over a two minute conversation using 64 kb/s is less than 62 bits. If lower bit rates are used the discrepancy is of course less.

However, at the base station if the traffic is too heavy, a continuous stream of bits is received and then the integrated effect accumulates over the whole useful side of the apogee. In the orbit described, the discrepancy in the number of bits amounts to 2.3 bits per kHz clock rate. The 12 hour orbit implies that there are two apogees per day and their sub-satellite points are displaced 180^{o} in longitude. The use of 8 hours out of the 12 hour period implies that three satellites are required for 24 hour coverage. This introduces the complexity of hand-over. As one satellite approaches the four-hour point after apogee, another is approaching the 4h point before apogee and hand-over of communications has to take place. This implies more stringent station keeping.

2.4.4. Variant Options of the Molniya Orbit.

Of course, other orbits inclined at 63.4° with different periods also enjoy small in-plane secular perturbation. The use of an 8h orbit would, of course, involve smaller altitudes with a gain in the link budget relative to the 12h orbit of about 2.3 dB. This comparison is made for path distances when the satellite is at the apogee and the observer is at the sub-satellite point. However, if the same ellipticity is maintained, the useful portion of orbit drops to 5 hours 20 minutes, implying that the number of satellites would have to be increased to five. It also means that the satellite also spends more time in the radiation belts.

Increasing the period to 24 hours would introduce a loss of about 4.1 dB in the link budget relative to the 12 hour orbit. The useful portion is extended to 16 hours so that two satellites would be adequate for a 24 hour service. The time the satellites spend in the radiation belt is also reduced.

2.5. Inclined Circular Orbits.

The orbit parameters of the previous two systems are rather rigid and do not allow the flexibility required for a truly global system. The use of a constellation of inclined orbiting satellites solves this problem at the cost of requiring several satellites that are moving relative to a fixed terrestrial observer. Such a constellation would normally operate at a lower altitude than that of the geostationary or Molniya counterparts. If the same flux density is to be maintained over a given area and we assume that the frequency and the antenna type remains unchanged, then no appreciable gain in the link budget is to be expected since as the altitude is decreased so that the free space propagation loss decreases, the antenna beam-width increases so that the antenna gain drops. Sub-synchronous orbits tend to be preferred because of the ease of station keeping. However, the availability of satellite navigation systems like NAVSTAR, the Global Positioning System may offer greater flexibility in this respect.

If such a system is to provide a 24 hour service, then several satellites are required. This is because the satellite precesses round the earth in the orbit plane and there is relative movement between the earth and the orbit plane. It is interesting to note that in such a scheme, 24 hour service at a specified location implies global coverage. This is, of course, an attractive feature since the cost of the extra satellites can be offset by the potential of much greater user population, assuming that appropriate international collaboration can be achieved.

2.5.1. Polar Orbits.

The simplest and possibly the most obvious way of achieving global 24 hour service is by using the polar orbit. The earth's surface can be divided into an even number of segments whereby each orbit passes over two diametrically opposite segments. For orbits at altitudes above one earth radius, three orbits are required for 24 hour service, each spaced by 60° in latitude with three satellites in each (nine in total). At this altitude, the angle of elevation is zero at the edge of coverage. At higher altitudes, two perpendicular orbits with six satellites in total would suffice but the geometry will not be as symmetrical as the nine satellite system.

Figure 2.4 shows the number of satellites required per segment pair with altitude for various beam-widths of the observers antenna centred round the zenith so that a beam-width of 180° would imply a zero degree angle of elevation. The number of segment pairs required would then be equal to the number of satellites per segment pair. Thus, referring to figure 2.4, the number of satellites required for a constellation at a given altitude would be the square of the number given by the plot. The constellation provides for observers with antennae which, having the right beam-width and pointing towards the zenith, will always see at least one satellite. Effectively the space is split into rings of squares rotating in longitudinal planes with a satellite covering each square.

The disadvantage here is that the polar regions tend to get better coverage than other areas on the earth's surface. Unfortunately, the radiation dose at the polar region is considerably lower than at the equator. This implies that switching off satellites at the poles does not greatly decrease the integrated dose of the on-board electronics.

It would be desirable to reduce the beam-width of the observer's antennae or conversely the size of the square per satellite as much as possible. This would increase the gain of both antennae at the satellite and the mobile end. It would also decrease the vulnerability of the system to interference. However, figure 2.4 shows how quickly the reduction of the beam-width would make the system impractical due to the number of satellites required. Thus, at 6,370 km altitude, the number of satellites required for 180° beam-width mobile antennae would be nine while for a 150° beam-width terrestrial antenna sixteen would be required. The received power for any link is proportional to

$$P_T \quad A_T(f)^2 \quad (R\,f)^{-2} \quad A_R\,(f)^2 \qquad 2.2$$

or

$$P_T \quad \theta_T^{-2} \quad (R\,f)^{-2} \quad \theta_R^{-2} \qquad 2.3$$

where P_T is the transmitter power,

A_T Is the effective antenna area at the transmitter,

f iS the frequency,

R is the distance between the transmitter and the receiver,

A_R is the effective antenna area at the receiver.

θ_T is the beam-width of the transmitter antenna and

θ_R is the beam-width of the receiver antenna.

The first two terms represent the transmitter power and the antenna gain so that together they represent the EIRP. The third term represents the free-space loss and the last term represents the receiver antenna gain.

Equations 2.2 and 2.3 can be applied to both the up-link and the down-link so that the transmitter power could either be the power of the HPA on the satellite or of the mobile. The received carrier power at the edge of cover relative to a similar situation in a geostationary system is plotted in figure 2.5 for various orbit altitudes. The transmitter power has been maintained constant so that the figures obtained can be applied to both the up-link and the down-link. The spacecraft antenna has a beam-width such that it covers a 120° arc on the earth's surface and the mobile antenna is such that, when it lies between two sub-satellite points (i.e. 60° away from either), it can just see both. Figure 2.5 shows that the received signal power is lower in polar orbit systems using the minimum of three satellite per great circle. Positive gain can only be achieved by involving more satellites.

In such a system the Doppler shift also plays a role. Figure 2.6 shows the Doppler shift for a system when two system parameters are specified. These are the altitude and the beam-width of the terrestrial antenna. It can be seen that the Doppler shift in such a scheme could be quite large especially for systems with wider beam-widths and lower altitudes. If we consider a system with a beam-width of 150°, then for altitudes of above 1,000 km the Doppler shift is less than 11 ppm. This is of a similar order, or better than the Molniya orbit.

The perturbances in a inclined circular orbit have been discussed in the sections 2.3.1 and 2.4.1. The nodal regression for a polar orbit is zero while the apsidal rotation for a circular orbit can be corrected by slightly changing the period. At altitudes less than about 600 km aerodynamic drag plays an important role, though the use of such low orbit is not envisaged. Needless to say there is also considerable saving in fuel for the transfer orbit.

2.5.2. Efficient Constellations.

A constellation of polar orbiting satellites is, however, not the most efficient way of achieving global 24 hour coverage, mainly because of the non-uniform spatial distribution. It has been shown that a minimum of five satellites in circular orbits, are required for continuous single coverage (WALK 73, WALK 77, BALL 80). The constellation consists of five orbits with an inclination of 43.7° and a satellite in each orbit. In such a system R, the great circle range, is 69.2° (see figure 2.7). This implies that for a terrestrial beam-width of 170° and 160° the minimum constellation altitude is 16,957 km and 27,142 km respectively.

Walker produced specifications for constellations involving various numbers of satellites with the possibility of having one to four satellites visible from anywhere on earth. In this context, only single satellite visibility is of importance. Multiple satellite visibility is useful for navigation. More recently Draim (DRAI 85) has shown that four satellites are enough for global coverage and three satellites would suffice for a hemisphere coverage. These two systems cater for single satellite visibility only. However, these constellations make use of orbits with altitudes of 100,000 km. This would imply that much smaller beam-widths are required at the satellite antenna and consequently larger antennae and more stringent pointing accuracies are required. The altitudes required would increase the launch cost (HADF 74) by a factor of about 2.3.

Figure 2.8 shows the minimum orbit altitude for a given number of satellites and the signal power relative to the geostationary link.

Signal powers are compared at the great circle edge of cover and for a minimum angle of elevation of 5^{o} for the lower orbit constellation. In general, the situation is similar to the polar system and the signal tends to be weaker, though there is actually some gain for large constellations.

It should be noted that whilst such a system would guarantee that there is at least one satellite within the antenna beam-width, there may be more than one, necessitating the need of avoiding interference from different satellites under multi-satellite visibility, such as assignment of different operating frequencies to the different satellites or code division multiplexing.

As in the Molniya case there is more than one satellite involved. The difficulties involved are worse than those of the Molniya since all satellites are operational at the same time and overlap of the coverage is possible. This introduces problems related to traffic management and routing.

2.6. Conclusion.

The three schemes have applications for different environments and the best contender will depend on the geographical constraints.

The geostationary orbit is the most attractive for low to middle latitudes depending on the lowest acceptable angle of elevation. For an minimum angle of elevation of 30^{o} the maximum latitude is about 50^{o}. One satellite will suffice for national services and three suitably positioned satellites provide quasi-global coverage as in the INMARSAT system. The polar regions are not covered. Communication is theoretically feasible up to 81° latitude but communication at the implied low angles of elevation is not attractive in the mobile sphere.

The main attraction of this orbit is, of course, the stationarity of the sub-satellite point, making the link performance considerably stable and communication comparatively easy. Because the position of the satellite is fixed fewer variables are involved in pointing the antenna. Towards the sub-satellite point an antenna with its main lobe centred around the zenith will produce an excellent link but in general this is not the case. Away from the equator a steerable or a low-gain antenna would have to be used. Practically all communication satellites occupy this orbit and most western launching facilities are geared for it.

The sub-synchronous Molniya orbit does not have a fixed sub-satellite point but is quasi-fixed for sufficient periods at places at about 55° latitude, but separated in longitude by 180°. This is more suited for countries around this latitude and may also be used up to the poles where increased activity is predicted. Operation in these areas is practically zenithal, greatly simplifying antenna design.

Due to the high ellipticity the sub-satellite trace is very non-linear with time and therefore this orbit is unsuitable for global coverage. Signals also suffer a Doppler shift which is not incurred in the geostationary orbit. The relative motion of the satellite to the earth implies that a multi-satellite system is required for 24 hour coverage. In fact three would suffice. Assuming a launch into an appropriate inclination, the Molniya launch should be cheaper considering that the velocity increments involved in the launch sequence are lower than those of the geostationary.

The inclined circular orbit constellation is inherently a true global coverage system. At least five satellites are required which is only two more than the Molniya system, but since it is a global system a greater market could be reached if political issues permit. The design of such a system is flexible so that number of satellites and altitudes can be traded for elevation angles.

The number of satellites involved, the problems of multi-satellite visibility at the mobile and routing of traffic between different satellites, together with possible political operational issues make such systems unattractive commercially. However, the number of satellites involved inherently improves survivability which could make them attractive in the military sphere.

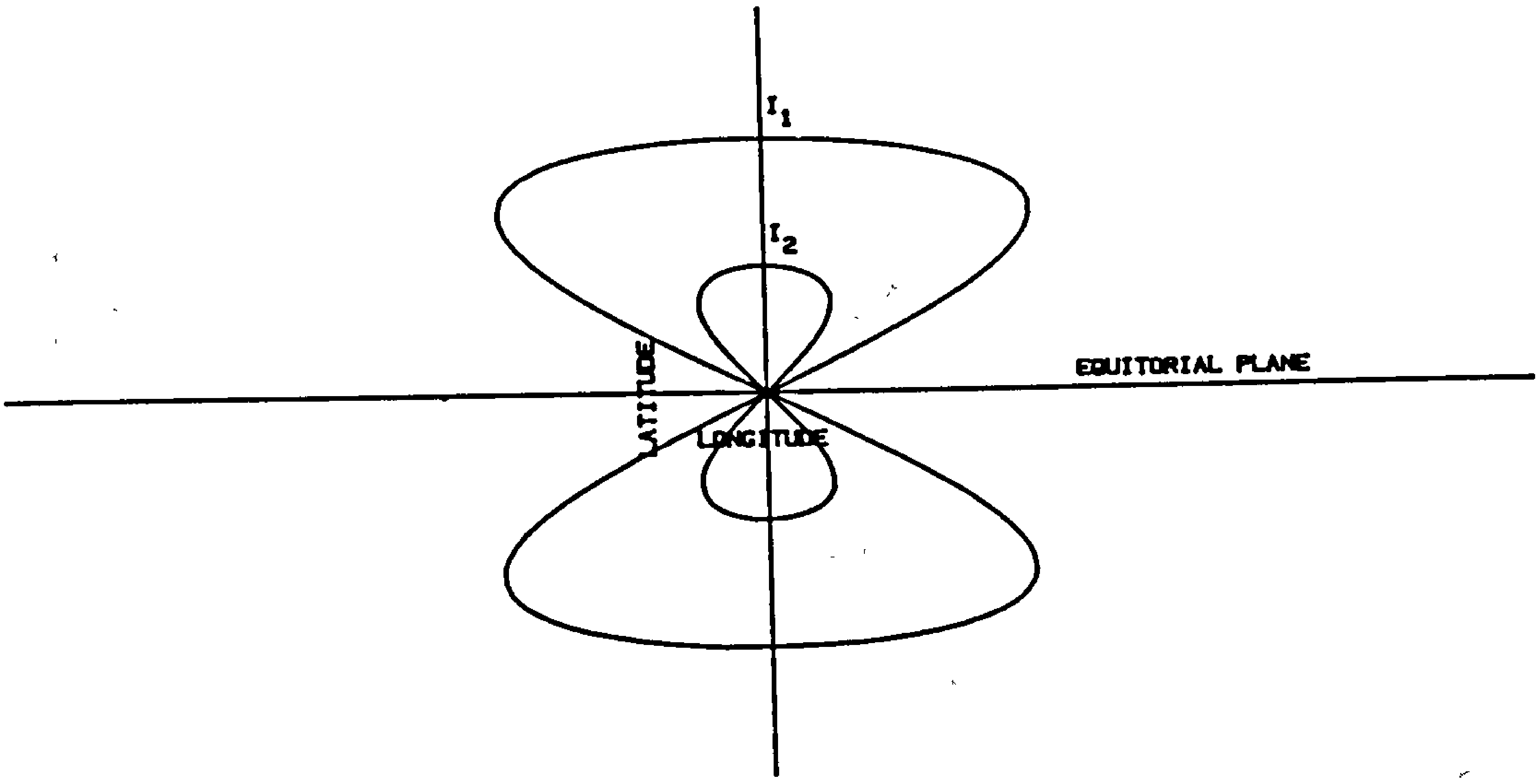


Figure 2.1. Figure of '8' trace for an inclined synchronous orbit where I_1 and I_2 are the angles of elevation with $I_1 > I_2$.

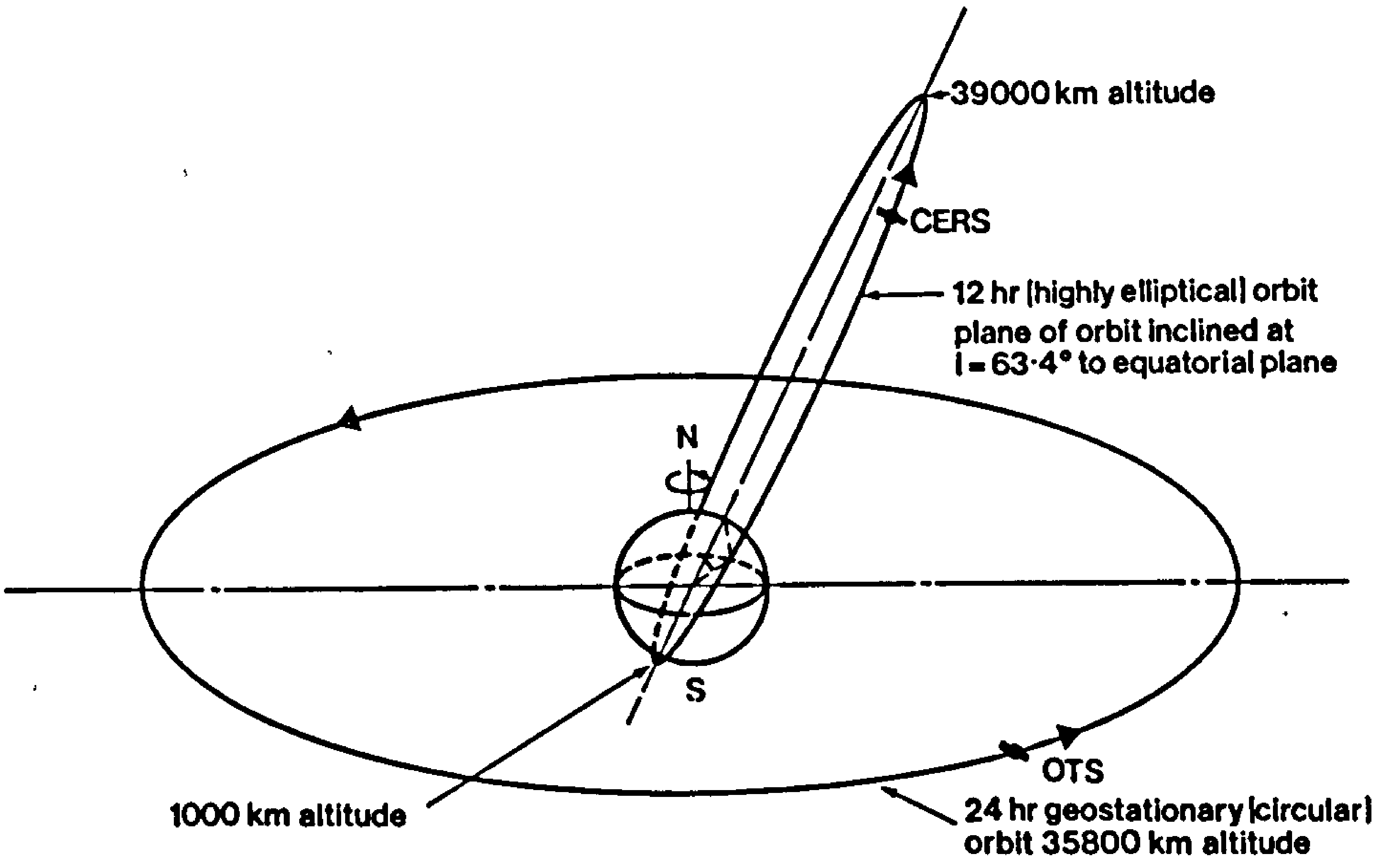


Figure 2.2. The CERS and the geostationary orbit as seen from a fixed point in space.

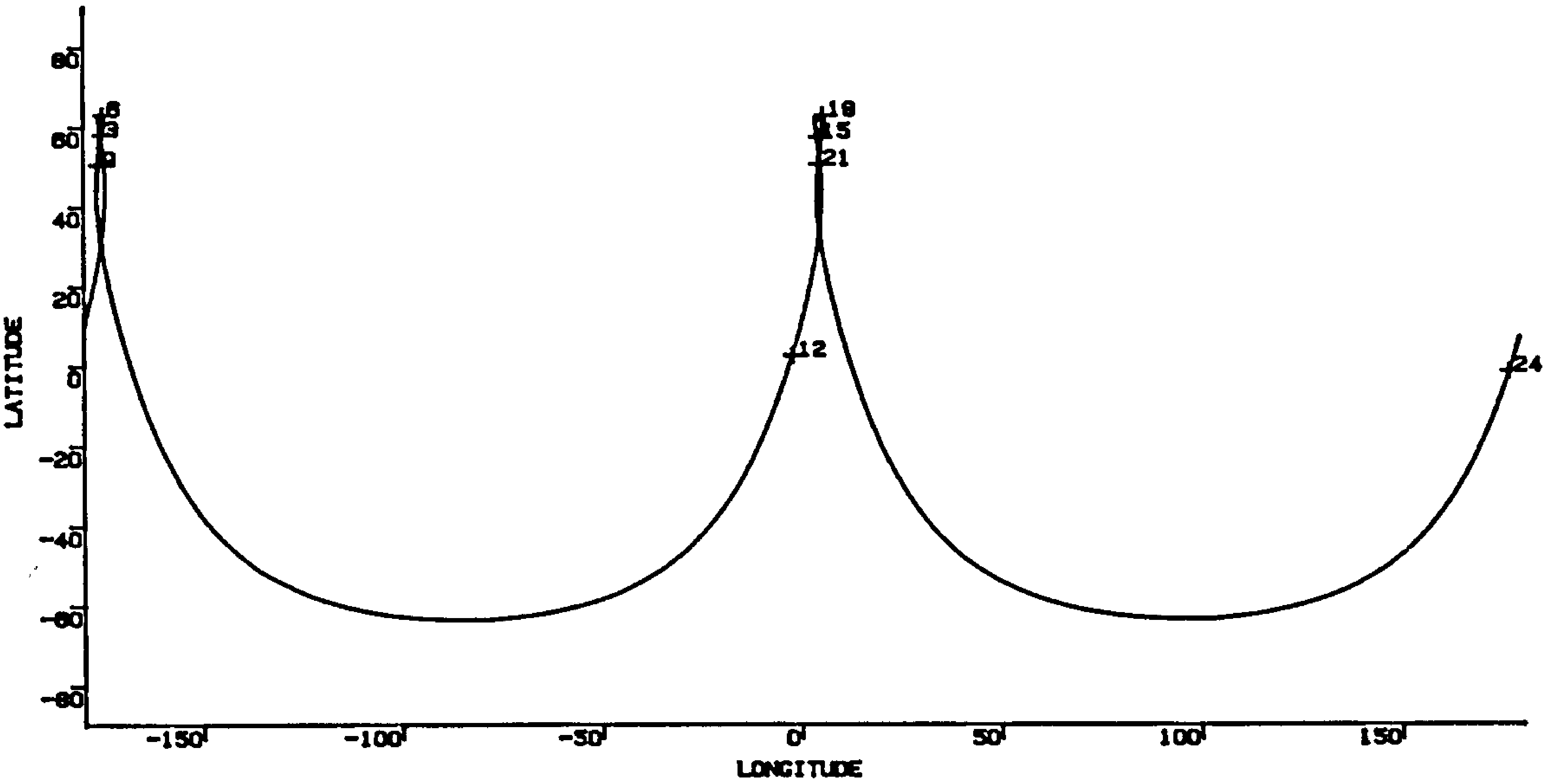


Figure 2.3. The sub-satellite trace over 24 hours of the Molniya orbit. Marks indicate progression in time (in hours).

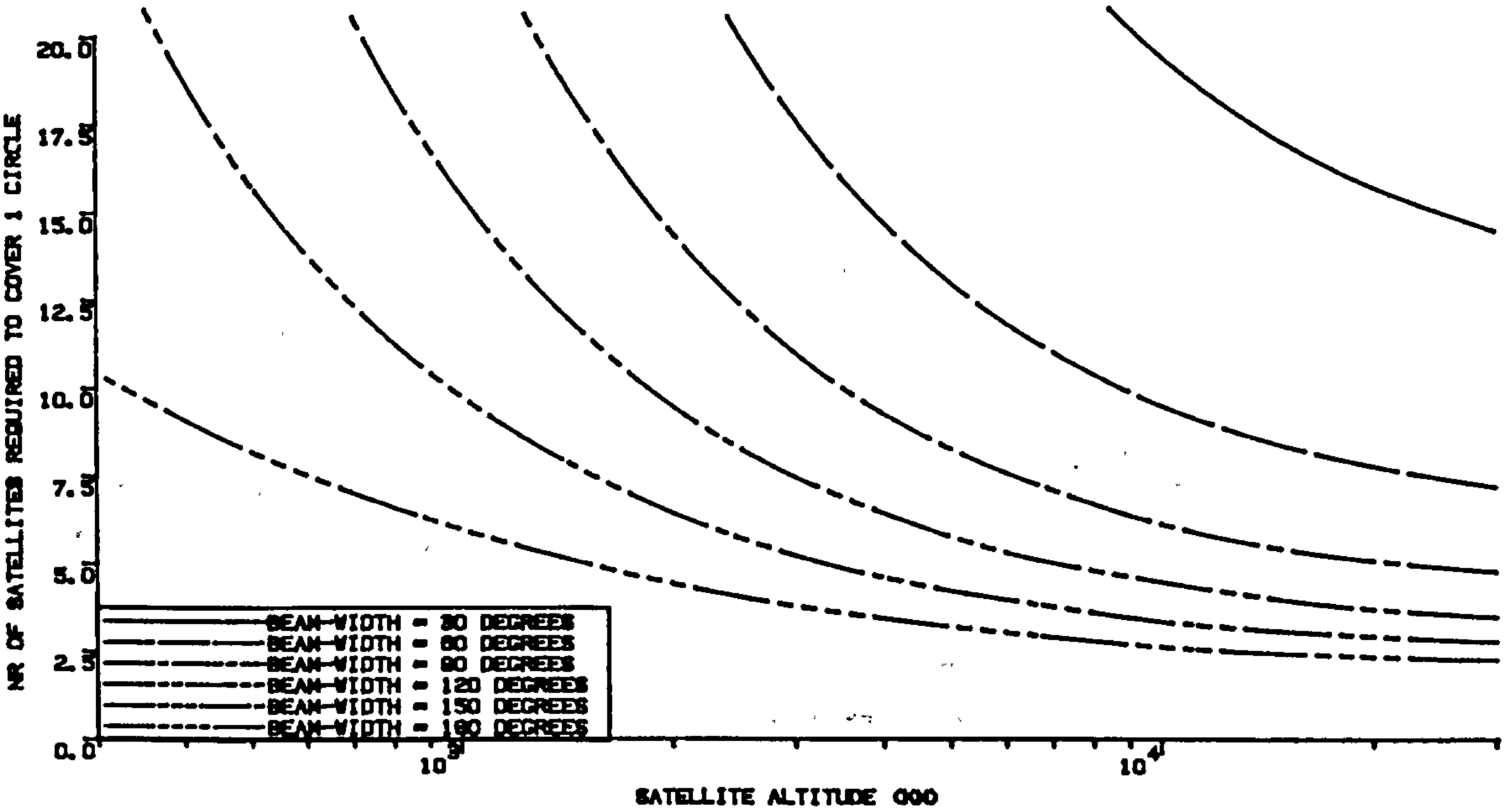


Figure 2.4. The number of satellites required per polar orbit for different beam-widths of the terrestrial antenna.

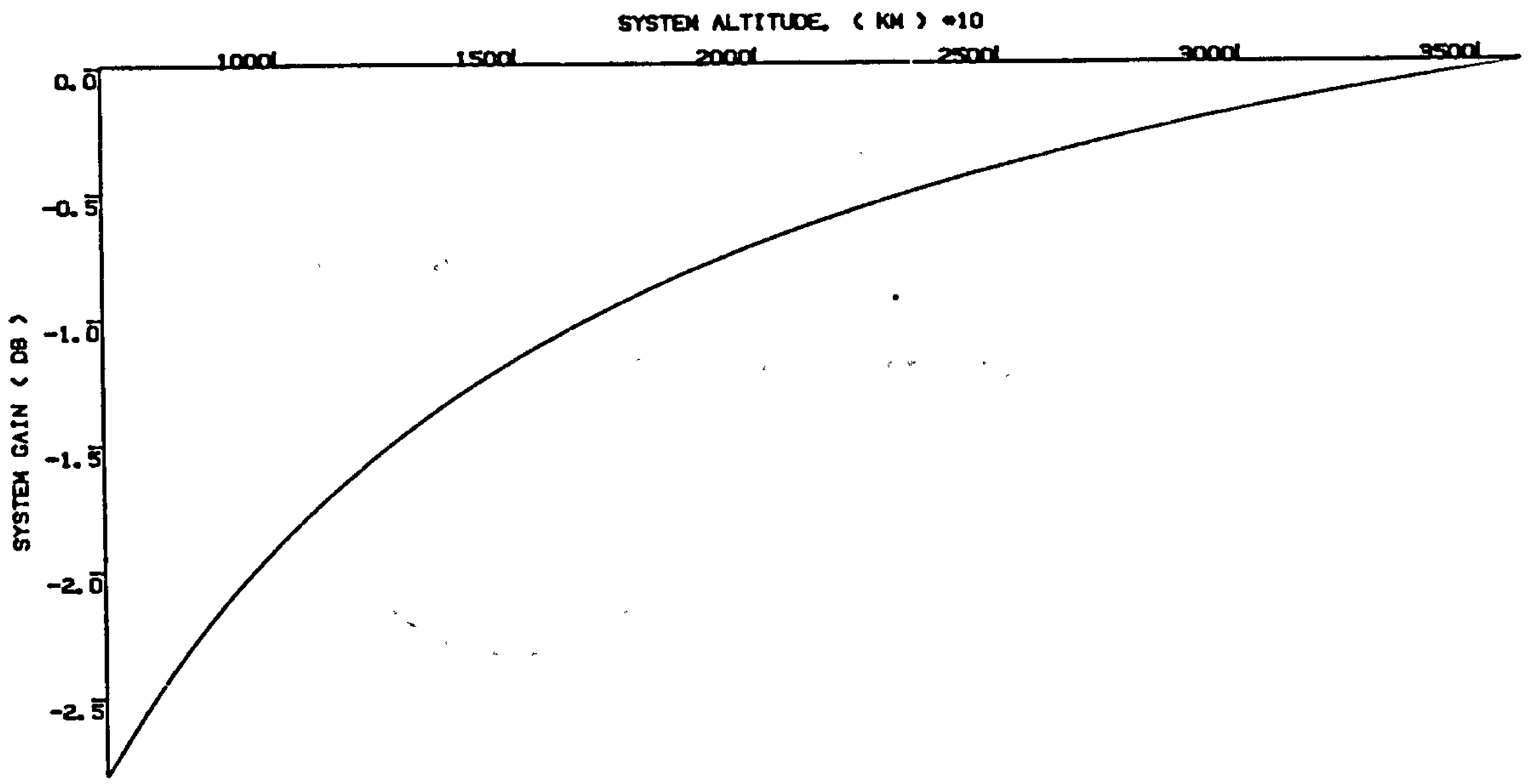


Figure 2.5. Comparison of the received power of a polar system and a geostationary system for various altitudes of the polar system.

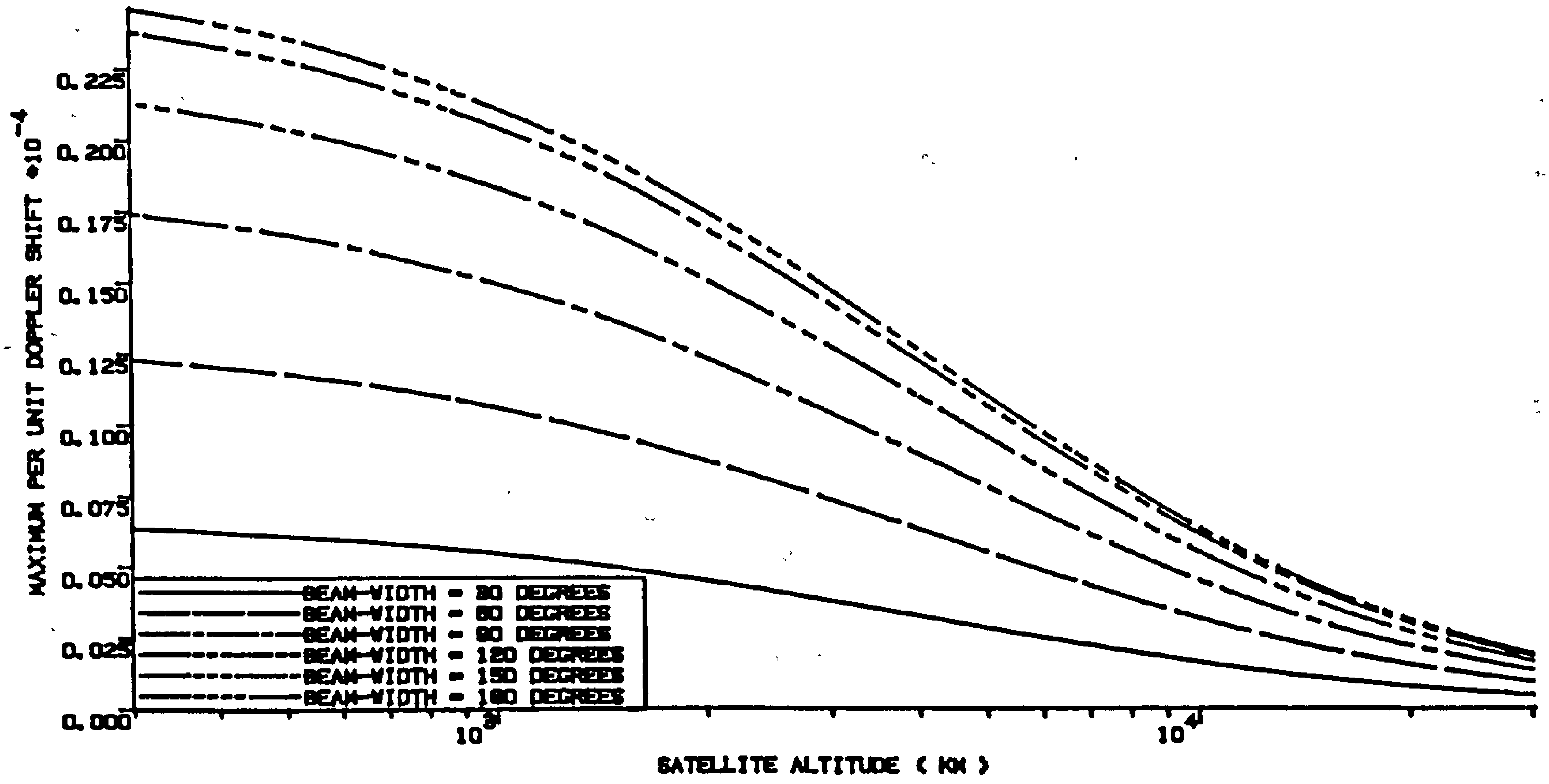


Figure 2.6. Maximum Doppler shift for a polar orbit system at different altitudes for various beam-widths of the terrestrial antenna.

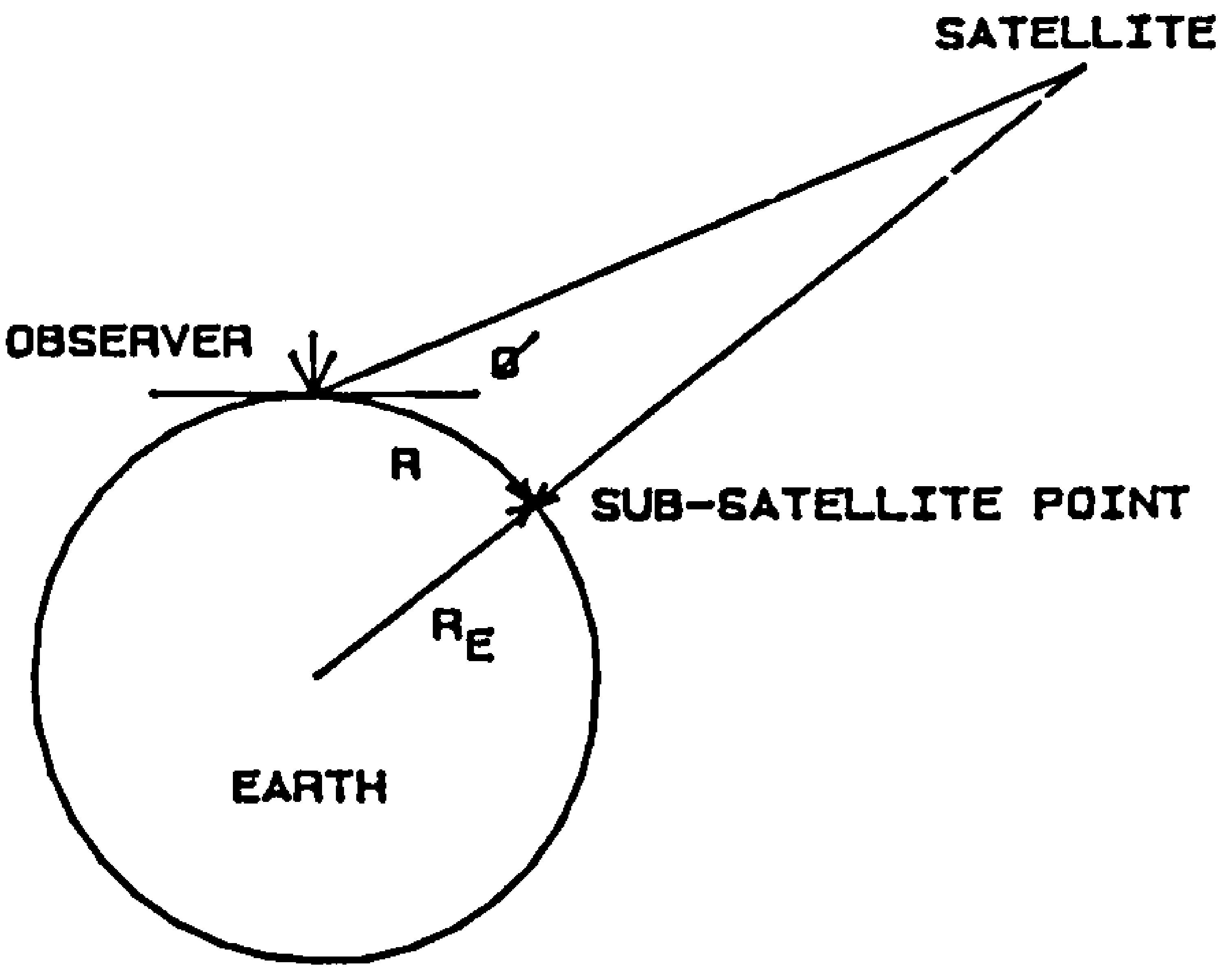


Figure 2.7. Observer to satellite plan-view where Ø is the angle of elevation of the observer, R_E is the earth's radius, H is the satellite altitude and R is the great circle range.

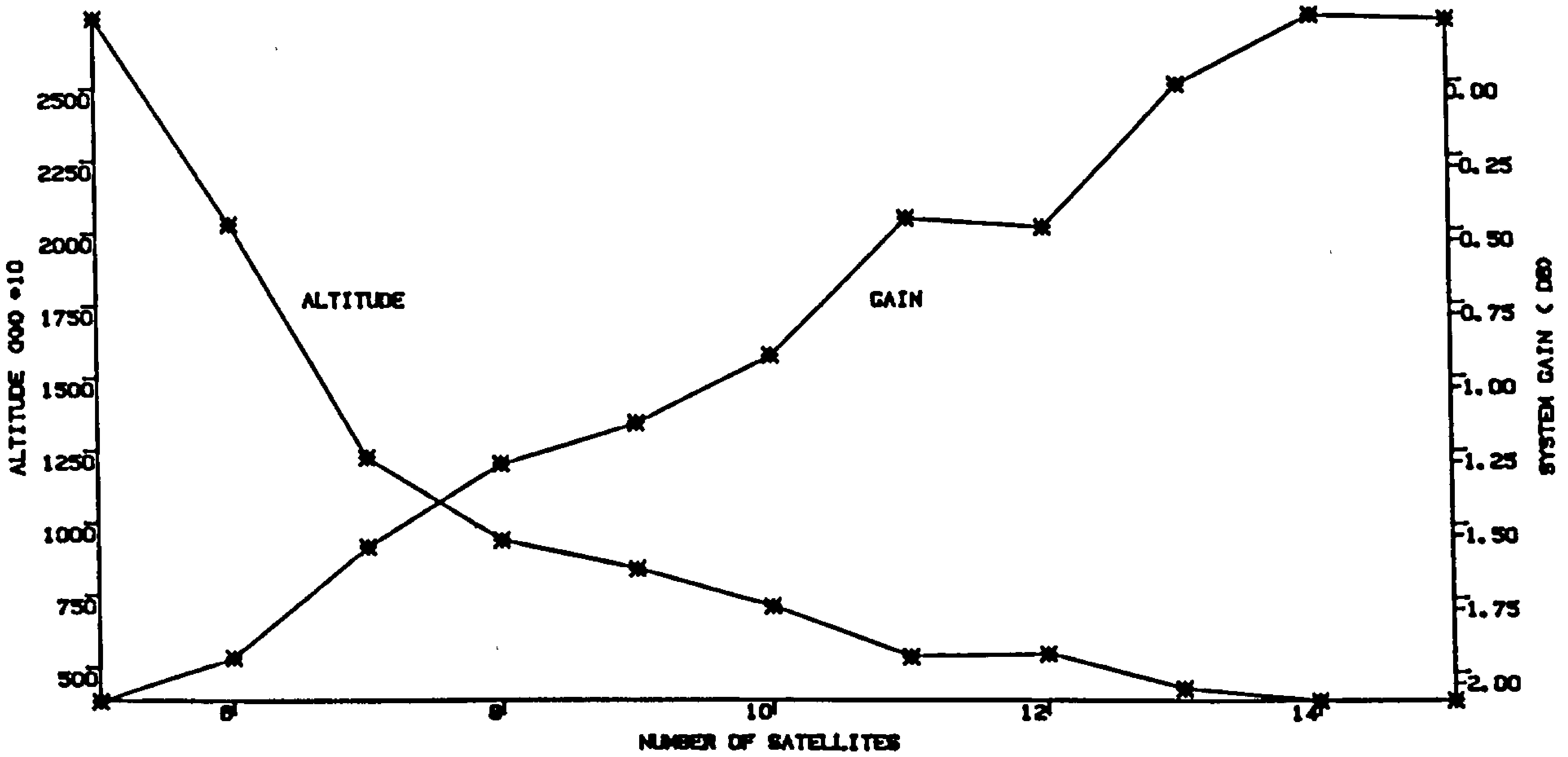


Figure 2.8. The minimum orbit altitude for a given number of satellites in a Walker constellation and the signal power relative to the geostationary link for an elevation of 5°.

CHAPTER 3.

MULTIPLE ACCESS SYSTEMS.

3.1. Introduction.

If we consider N users and consider a connection matrix C where

$$C = \begin{bmatrix} c_{11} & c_{12} & c_{13} & c_{14} & \cdots\cdots & c_{1N} \\ c_{21} & c_{22} & c_{23} & c_{24} & \cdots\cdots & c_{2N} \\ \vdots & \vdots & \vdots & \vdots & & \vdots \\ c_{N1} & c_{N2} & c_{N3} & c_{N4} & \cdots\cdots & c_{NN} \end{bmatrix}$$

so that $c_{ij} = 1$ means a connection from user i to user j. It will be obvious that $c_{ii} = 0$. For duplex operation c_{ij} and c_{ji} will both be equal to one, while for simplex only one is equal to one, the other is zero. The total number of connections in this matrix is then given by

$$L = \sum_{i=1}^{N} \sum_{j=1}^{N} c_{ij} = N (N - 1) \qquad 3.1$$

This gives an acceptable number of connections only when N is small. As N increases the number of connections will increase to the square of N. As connections are valuable resources, and as not all connections are normally required at the same time, L is usually less than N (N - 1) and the available links have to be shared.

Connections can take various forms. The important issue is that they must be mathematically mutually orthogonal, otherwise interference occurs. Connections can be physical wires and this is the case in classical telephony. However, in this context we are interested in RF connections. Three techniques can be identified:

1. Frequency Division Multiple Access (FDMA).

   In this scheme the total system bandwidth is split into channels. Thus the signal originating from different sources must satisfy the orthogonality condition

$$\int s_i(f)\ s_j(f)\ df = 1 \quad \text{for } i = j \qquad 3.2$$
$$= 0 \quad \text{for } i \neq j$$

2. Time Division Multiple Access (TDMA).

   In this scheme each user makes use of the system's full capacity but sharing is done in the time domain. The orthogonality condition in this case is

$$\int s_i(t)\ s_j(t)\ dt = 1 \quad \text{for } i = j \qquad 3.3$$
$$= 0 \quad \text{for } i \neq j$$

3. Code Division Multiple Access (CDMA).

   CDMA is also known as spread spectrum multiple access (SSMA). Users use the full spectrum continuously. However orthogonality is achieved by multiplying the signals from different users by respective pseudo-random sequences that exhibit low cross-correlation properties. The output from user i is of the form

$$f_i(t) = g_i(t)\ s_i(t) \qquad 3.4$$

where $g_i(t)$ is a pseudo-random sequence and

$$\int g_i(t)\, g_j(t)\, dt = 1 \quad \text{for } i = j \qquad 3.5$$

$$= 0 \quad \text{for } i \neq j$$

The above three methods of multiple access will determine the number of simultaneous connections or channels. In systems with large user populations as in the mobile environment, there are not enough channels to allow for N (N - 1) connections and protocols are required to establish the procedure a user must apply in order to obtain a link from the pool of links in the system. Such protocols are called multiple-access protocol or demand assignment protocols and will be discussed in the next chapter.

3.2. Frequency Division Multiple Access.

Orthogonality in the frequency domain leads to FDMA. The system bandwidth B Hz is split into L channels so that the channel bandwidth b Hz is B/L Hz. In practice b is less than B/L to allow for filtering and frequency instabilities. B/L becomes the channel spacing. Figure 3.1 illustrates the channel allocation in the INTELSAT systems (CCST 85, FEHE 83a): SCPC (Single Channel Per Carrier) and SPADE (Single channel per carrier PCM multi-Access Demand-assigned Equipment). Both systems have the same channel allocation except that the 160 kHz channel at the low frequency end is used for different applications.

Since the systems use full duplex, channels are paired as i-i′. Channels 1-1′ and 2-2′ are not used commercially. The lower and upper channel sets are separated by a pilot frequency. The modulation used is QPSK allowing a channel capacity of 64 kbps. The channel bandwidth is 38 kHz. For simplicity we will consider the upper channel set where there are 400 channels in 18 MHz, giving a channel spanning of 18000/400=45 kHz. The guard band is the difference between the channel bandwidth and spacing, i.e. 7 kHz. A simplified block-diagram of an earth station is shown in figure 3.2. The bandwidth of the terminal is effectively equal to the channel bandwidth so that it is basically a narrow-band system. The information signal is modulated using the output of a frequency synthesizer. The modulator output is then filtered. Modulation, particularly digital modulation, involves bandspreading so steps must be taken to ensure that the modulated output is still within the channel bandwidth.

Though filtering has been included in both block diagrams not all filtering is practical or can be performed in the desired fashion. For example, post-HPA filtering at the terminal at Ku-band (14/12 GHz) with say 38 kHz channel bandwidth, would involve Q-factors in the region of 370,000 (FEHE 83b). This would be the case for the INTELSAT FDMA system. On the mobile scene the situation will be even worse. Channel bandwidths as low as 2.4 kHz are proposed and though there are systems at UHF, L-band (1.5-2.7 GHz) and X-band (8/7 GHz) it will not take long until systems in the 30/20 GHz will have to be considered. The corresponding Q-factors would be 320,000, 600,000, 2,800,000, and 8,000,000 respectively.

The same applies for filtering at the input of the LNA. This means that the receiver cannot be made as selective as would be desirable and the output cannot be made as clean as desirable. Of course this makes channel selection easier, particularly on the transmitter side. The consequences of inadequate filtering are that the bandspreading due to the HPA non-linearity cannot be limited and the LNA may be driven into non-linearity by adjacent channels.

Since the transparent transponder deals with the full system bandwidth the filtering problem is considerably less. In the INTELSAT system the Q-factor is about 400 while in the MSAT system (MSAT 85) the up-link and the down-link bands are proposed to be 821-825 MHz and 866-870 MHz corresponding to Q-factors of 200.

3.2.1. Non-linear Amplifier.

The travelling wave tube (TWT) is the most popular and proved device for the HPA in the spacecraft, though transistors are beginning to wedge in, particularly in smaller systems. HPA's are critical components with respect to non-linearities. Typical TWT (INTE 80) characteristics are illustrated in figure 3.4. These non-linearities contribute two effects:

1. Amplitude modulation to amplitude modulation (AM/AM) conversion which is a non-linear output-input power characteristic.
2. Amplitude modulation to phase modulation (AM/PM) conversion which is a non-linear output phase-input power characteristic.

The output-input power characteristic shows that as the input power increases the output power also increases, up to a point beyond which this relationship does not hold; this point is called saturation. In practice this point is very broad and the "7/11" rule can be applied. Thus saturation can be redefined as the point from which an 11 dB decrease in the input power results in a 7 dB drop in the output power.

Two equivalent models can be used to study the non-linearities of the device. These treat the device as being bandlimited and in their simpler form the non-linearities are considered frequency independent.

1. Amplitude-Phase Model (SHIM 71).

If the input signal is

$$x(t) = r(t) \cos\{w_0 t + \psi(t)\} \qquad 3.6$$

The output of the amplitude-phase model yields an output

$$y(t) = A[r(t)] \cos\{w_0 t + \psi(t) + \phi[r(t)]\} \qquad 3.7$$

where A(r) is an odd function of r and $\phi(r)$ is an even function of r. The function A(r) represents the AM/AM conversion while the function $\phi(r)$ represents the AM/PM conversion. Various numbers of parameters have been proposed for the two functions, but probably the simplest is the two-parameter representation (SALE 81)

$$A(r) = a_A\, r \,/\, (1 + b_A\, r^2)$$

$$\phi(r) = a_B\, r^2 \,/\, (1 + b_B\, r^2) \qquad 3.8b$$

2. Quadrature Model (FUEN 73).

In this model, while the input is assumed to be of a similar form to the previous model, the output is the sum of two components in quadrature. The model is illustrated in figure 3.5. The in-phase and quadrature components are given by

$$p(t) = P[r(t)] \cos[w_0 t + \psi(t)] \qquad 3.9a$$

$$q(t) = -Q[r(t)] \sin[w_0 t + \psi(t)] \qquad 3.9b$$

It can be shown that this model is equivalent to the previous, since

$$P(r) = A(r) \cos[\phi(r)] \qquad 3.10a$$

$$Q(r) = A(r) \sin[\phi(r)] \qquad 3.10b$$

The functions P(r) and Q(r) can be represented by odd polynomials or by summations of the Bessel function of the first kind of order one. A simpler method is to use two-parameter formulae (SALE 81).

$$P(r) = a_p\, r \,/\, (1 + b_p\, r^2) \qquad 3.11a$$

$$Q(r) = a_q\, r^3 \,/\, (1 + b_q\, r^2)^2 \qquad 3.11b$$

3.2.2. Degradations.

The major problem in the context of a non-linear amplifier in an FDMA system is in the satellite HPA where several carriers are being handled by the same amplifier. Such non-linearities result in intermodulation products (IMP's) where new spectral lines are generated. Their frequencies are given by

$$f_{lmn} = l\, f_1 \pm m\, f_2 \pm n\, f_3 \pm \ldots \qquad 3.12$$

where f_{lmn} is the frequency of the IMP,

f_1, f_2, f_3 are the frequencies of the individual carriers and

l, m, n are the harmonic numbers.

This is the result of AM/AM conversion. AM/PM distortion causes intermodulation noise and intelligible crosstalk (BERM 70). The even order IMP's fall outside the bandwidth of interest and therefore can be filtered. However, some of the odd order products fall within the bandwidth of interest.

Another phenomenon is signal suppression (BHAR 81). Due to the non-linear gain, power outputs of carriers are affected by their relative magnitude. Thus if two signals of equal power are present at the input, the output will yield equal power of each. However if one is stronger, the power output of that one will increase while the other will decrease so that the strong signal appears to suppress the weaker one.

The amplifier non-linearity also causes spectral spreading so that the output spectrum is always wider than the input spectrum. If we consider wideband signal, spectral spreading can be considered as intermodulation on a continuous spectrum.

The AM-PM and AM-AM effects are not so much of a problem in constant amplitude modulation schemes. However, in practice, these are difficult to achieve and some amplitude modulation will always be present.

If we look at figure 3.4. we observe that at low input levels the amplifier is more linear. This is generally true for any amplifier. Thus, by operating the HPA below the saturation point, i.e. applying a backoff, the amplifier is rendered more linear and these degradations are suppressed. Typical backoff levels of 3-6 dB are applied. However, this implies a less efficient use of an important resource.

3.3. Time Division Multiple Access.

TDMA allows several users to access the full bandwidth sequentially. This involves the need for the organisation in the time domain so that messages from different users do not interfere with each other (that is to maintain orthogonality). The system lends itself to digital communications. A group of time slots are organised in a frame as indicated in figure 3.6. A continuous bit stream can be transmitted through a TDMA system by first packetisation, then compression. This involves transmission of a packet at a rate which is higher than the original bit rate. At the receiver end the packet is expanded and then concatenated to constitute the original bit stream at the original rate. In figure 3.7. it can be seen that slot 2 is used.

The system invariably introduces a delay. Compression involves reading into a buffer at the bit stream rate and reading out at the transmission rate. Expansion is required at the receiving end. The delay introduced is the sum of the packet duration and any processing delay. This is, of course, to be added to the propagation delay. In satellite communications the latter component is usually much larger than the others.

It is worthwhile noting that in this simplistic scheme the packet duration is equal to the slot duration. However, the transmission rate will be slightly higher than the product of the number of slots and the bit stream rate. It is higher because a preamble is transmitted per slot which is effectively extra to the data and contains no useful information that the receiving user is interested in.

This overhead is composed of four sections:

1. The guard time. This signifies an area of uncertainty where a burst is going to start. The actual duration is dependent on the timing and ranging inaccuracies in the system. It should also be long enough to allow any trailing effects of the previous burst to drop to sufficiently low levels.
2. Carrier recovery. This is a sequence of ones or an alternate pattern of bits. It allows the carrier recovery circuit to extract the coherent carrier that is essential in coherent demodulation. The use of incoherent or differential demodulation disposes the need of this sequence. However, this type of demodulation has an inferior performance to that of the coherent type. Differential demodulation is likely to be used on-board a satellite to keep complexity low. The use of Differentially Quadriphase Shift Keying (DQPSK) on the down-link is assumed in an ESA report (PENT 81). The degradation over the coherent alternative is about 2.4 dB.
3. Bit recovery. This function is accomplished through an alternating bit sequence which has a strong component at half the clock rate. This process can either be concurrent with the carrier recovery process or after it.
4. Unique word. This is a special binary word that signifies the start of the message. The size of the word should be such that the probability of an error is adequately low . This probability is given by

$$P\ (\text{UW error}) = \sum_{i=e+1}^{n} \binom{n}{i} p^{i} (1 - p)^{n-i} \qquad 3.13$$

where n is the number of bits in the unique word,

e is the number of allowed bits in error,

$\binom{n}{i}$ is the binomial coefficient and

p is the bit error rate.

A typical format is illustrated in figure 3.6.

3.3.1. System Efficiency.

The efficiency of a TDMA frame is the portion of the frame that is actually used to convey useful information.

If the guard time is T_g us,

the number of symbols in sequences for the carrier and bit clock recovery and the unique word is P,

the symbol duration T_s us,

the frame duration is T_f us

and the number of slots in a frame is K

then the efficiency is given by

$$E = \frac{1}{T_f} \{T_f - \sum_{i=1}^{K} (T_{gi} + T_s P_i)\}$$

$$= 1 - (1 / T_f) \sum_{i=1}^{K} (T_{gi} + T_s P_i) \qquad 3.14$$

Very often the slot structure within the frame is the same so that the efficiency simplifies to

$$E = 1 - K (T_g + T_s P) / T_f \qquad 3.15$$

It is obviously desirable to keep the efficiency as high as possible, which in turn implies that all overheads are to be kept to a minimum. However, it should also be clear that the frame efficiency is only one of the system parameters and that trying to keep a high figure will affect other system parameters.

It has already been mentioned that carrier recovery sequence can be totally disposed of, but this will degrade the signal-to-noise performance. The guard time can be reduced if the specifications of the timing circuitry are tighter and more sophisticated acquisition techniques are employed. This might not be desirable in the mobile environment. Similarly, for the price of hardware complexity, the bit clock recovery sequence can be made shorter while for a given bit error rate, shortening the unique word would reduce the reliability of the system.

3.3.2. Synchronisation Methods.

One of the factors that affect ranging errors is the uncertainty in the satellite position. The space region of a geostationary satellite is typically limited to a square with sides of 0.1°. The altitude typically is controlled to within ± 0.1 % due to the orbit ellipticity. This means that the satellite is positioned in a box 25 km x 25 km x 75 km. This is equivalent to a round-trip delay variation of 0.5 ms.

There is also a delay variation due to the geographical location of the different users. For a global beam system such as the INMARSAT system this factor contributes to a one-way delay variation of about 20 ms.

With frame lengths in the region of 0.1 to 100 ms it is not difficult to see that efforts in order to maintain a low uncertainty in the propagation delay are essential to achieve high frame efficiencies.

As mentioned earlier the guard time will take care of any discrepancy in the burst synchronisation.

The process of making sure that the bursts are maintained within the time slots can be split into two phases: the acquisition phase and synchronisation phase. Acquisition is the initial phase which involves obtaining an estimate of the delay with reference to the reference burst, that would result in transmitted bursts reaching the satellite at the appropriate location within the TDMA frame. The synchronisation phase maintains subsequent bursts within the allocated time slot to within the required tolerance.

The open-loop technique determines the initial delay referred to the reference burst by computing it with the ´a priori´ knowledge of the exact location of the earth station and the distance between the earth station and the satellite. The simplicity of this technique makes it very attractive in the mobile scenario. The distance of the satellite to a given geographical location can be obtained from the ranging data on the TT&C channel, but the exact geographical location of a mobile may be difficult to determine. Relaxing the accuracy of the mobile location leads to larger guard times.

The closed-loop technique requires a reference burst to be broadcast and relies on the ability of each station to transmit and listen to its own acquisition signal on its down-link. This signal can either be a burst of unmodulated carrier some 25 dB below the normal level, or a pseudo-random sequence. This ensures that these signals do not interfere with the other bursts by making sure that the acquisition signal is well below the normal burst.

This technique inherently assumes a transparent transponder, or at least a fixed delay within the satellite. The station can estimate the delay to be introduced from the difference of the reference burst and the acquisition signal.

If the reference burst is at the beginning of the frame, then its start referred to time at the satellite, is given by

$$R_s(n) = t_o + n\,T \qquad 3.16$$

where n is an integer,

t_o is a constant,

T is the frame duration,

subscript "s" signifies that the measurement is at the satellite, while "e" will be used for the earth.

At the earth's side the reference bursts will occur at

$$R_e(n)=t_o + n\,T + t_p(nT) \qquad 3.17$$

where $t_p(t)$ is the one way propagation delay which is dependent on the time and geographical location of the terrestrial station.

A burst transmitted from the station at $R_e(n)$ would reach the satellite at

$$A_s(n) = t_o + nT + t_p(nT) + t_p(nT + t_p(nT)) \qquad 3.18$$

If we assume that K is an integer such that

$$(K - 1)T \leq 2\ t_p(\max) + d \leq KT$$

where $t_p(\max)$ is the maximum one way propagation delay,

d is the extra delay incurred by the passage through the satellite and

K is an integer

At the satellite the frame start from the station is then referenced to

$$R_s(n + K) = t_o + (n + K)\ T \qquad 3.19$$

so that the delay to be introduced at the station is

$$T_d = R_s(n + K) - A_s(n) \qquad 3.20$$

3.4. Code Division Multiple Access.

Spread spectrum signals use the full system bandwidth continuously. This implies that the transmission bandwidth is larger than the baseband bandwidth. A conventional modulation method that fulfils this is wideband FM where, if the modulation index is large enough there are advantages in noise and interference performances. Wideband FM in itself however, does not cater for orthogonality. If two signals are frequency modulated over the same bandwidth, then detection cannot differentiate between the two signals.

Spread spectrum techniques achieve orthogonality by virtue of the auto and cross-correlation properties of pseudo-noise (PN) sequences. These are pseudo-random binary sequences with noise-like characteristics. Traditionally spread spectrum methods have been used by the military because of their anti-jamming,low detectability and encryptic properties which are the result of high system processing gain. This is the ratio of the transmission bandwidth to the information bandwidth so that

$$[G_p] = 10 \lg(B_c/B_i) \qquad 3.21$$

where G_p is the system processing gain,

B_c is the transmission bandwidth and

B_i is the information bandwidth.

[] indicates that the units are in dB.

Typical figures for the processing gain are in the range of 20 to 60 dB (BHAR 81).

The interference or jamming margin is given by

$$[M_j] = [G_p] - [(S / N)_o] - [L] \qquad 3.22$$

where L is the system implementation loss, typically 1 to 3 dB, and

$(S / N)_o$ is the required signal-to-noise at the information output.

Interfering signals not exceeding M_j dB above the wanted signal would not affect the desired performance.

Two spread spectrum techniques will be discussed: the direct sequence system (DS); and the frequency hopping (FH) system.

3.4.1. Direct Sequence System.

Two methods are available for DS system. The signal can either be modulated in the normal fashion and the modulator output is then multiplied by the PN sequence or, the baseband signal is first multiplied by the PN sequence and then modulated on the carrier. The bit rate of the PN sequence is called the chip rate. This avoids confusion with the information bit rate.

At the receiver the reverse process takes place. The RF signal is either multiplied by the PN sequence and then demodulated to extract the information or, demodulated and then multiplied by the sequence. Figure 3.8 shows the block diagrams for receivers and transmitters using these methods.

The PN sequence at either end of the link must be the same and in synchronisation. Looking at block diagrams in figure 3.8a and 3.8b, if the modulated signal at the receiver $S_1(t)$ and the PN sequence is $g_1(t)$ then the RF output is

$$g_1(t)\ S_1(t) \tag{3.23}$$

At the receiver input, besides the desired input, other components appear and these are

$$\sum_{i=2}^{n} g_i(t)\ S_i(t) + S'(t) + N \tag{3.24}$$

The summation represents the net result of signals from other n-1 users in the system. The second term represents interfering signals not generated by the n users and N represents the noise.

Multiplying the input by the synchronised version of the PN sequence $g_1(t)$ will yield.

$$g_1^{\,2}(t)\ S_1(t) + g_1(t) \sum_{i=2}^{n} g_i(t)\ S_i(t) + g_1(t)\ S'(t) + g_1(t)\ N \tag{3.25}$$

When integrated over the sequence length the first term will yield the wanted signal because of the autocorrelation properties of the PN sequence. The second term should produce a zero result because of the low cross correlation between different codes. The last two terms will produce unwanted outputs, but these will be attenuated due to the filtering of the bandpass filter.

Looking at the power spectra at various points of the system, figure 3.9a shows a typical spectrum of a digitally modulated signal. After multiplying by the spreading function, the PN sequence, the spectrum assumes the shape of $(\sin(x)/x)^2$ with the main lobe bounded within $2*R_c$, R_c being the chip rate. This is illustrated in figure 3.9b. The input spectrum at the receiver is shown in figure 3.9c where besides the desired spectrum shown in figure 3.9b there are also contributions from other users using other PN sequences, thermal noise and spot frequency interference.

After correlation the wideband interference from other users is eliminated due to low cross-correlation, the spot-frequency interference is spread so that only a small portion will pass through the receiver bandpass filter while only the thermal noise within the bandpass is admitted. The diagrams illustrate that as the processing gain increases the system becomes more interference resilient. The power density of the spread version becomes smaller so that the contribution of spot frequency interference after correlation is also smaller.

3.4.2. Frequency Hopping System.

The heart of this method is a frequency synthesizer that is controlled by a PN sequence. This means that the carrier frequency hops within the system bandwidth. The frequency separation between the discrete frequencies is f. If the PN-code generation has n-stages then the system bandwidth is

$$B_{rf} = (2^n - 1)\ f \qquad 3.26$$

As indicated in figure 3.10 the envelope of a FH system is rectangular.

If the information bandwidth is B_i, then the processing gain in such a scheme is given by

$$G_p = B_{rf} / B_i = (2^n - 1)\ f / B_i \qquad 3.27$$

If $f = B_i$ then

$$G_p = 2^n - 1 \qquad 3.28$$

The block diagrams for the FH transmitter and receiver are shown in figure 3.11. In the transmitter section the signal is modulated in a conventional fashion and then it is mixed via a wideband mixer by a local oscillation which is a frequency synthesizer controlled by the code generator.

At the receiver the signal is first dehopped. It is multiplied by the same sequence of frequencies as in the transmitter. This is locally generated by another synthesizer controlled by a PN-code generator which is synchronised to the transmitter's code. It is then demodulated in the conventional manner.

The chip rate for an FH system need not be larger than the information bit. However, in the presence of interference, it is effectively sampled for a chip duration. As the chip rate increases the chip smaller, the interference is spread over a larger band and the adverse effect is reduced.

At low chip rates interference will be coherent while at high chip rates interference appears to be noise-like. It is difficult to operate at high chip rates because of the performance of frequency synthesizers.

FH systems with several chips per bit are commonly referred to as fast FH, FFH. Due to the high switching rate it is very difficult to maintain phase continuity at the synthesizer output. This rules out coherent detection. This limitation is imposed due to the switching performance of synthesizers. At chip rates corresponding to a bit or more per chip it is easier to maintain phase continuity.

Another problem is receiver non-linearity, which may cause intermodulation products to fall within the passband. Such effects are, of course, periodic and only occur at a particular frequency (or subrange of frequencies) of the total range of frequencies generated by the frequency synthesizer.

3.4.3. Synchronisation.

Perfect timing is not essential in CDMA but synchronisation between the transmitter and receiver PN sequence is essential. Initial synchronisation can be achieved by a sliding correlator at the receiver. In this system the chip rate of the local PN sequence is varied so that the two PN sequences shift in phase relative to each other until synchronisation is achieved. In order to speed up the acquisition process, there might be an overlayed PN sequence of a much shorter length which is easier to lock to. This may then be used as a stepping stone to lock to the full length sequence.

Correlators normally employ tapped delay-line matched filter techniques. These can be implemented in digital electronics, charged-coupled devices (CCD) and surface acoustic wave (SAW) devices. The latter technology offers the simplest implementation above 25 MHz and is the only choice above 60 MHz but at low frequencies the digital technology is expected to dominate. The future of CCD appears to be less optimistic (RAPP 84).

It should be noted that CDMA exhibits a strong capture property. Once acquisition has occurred, the correlator can lock onto the packet. Knowing the length of the PN sequence, the approximate position of the correlation pulse can be determined so that the correlator output can be gated through a narrow time window. This provides discimination against users employing the same code if the codes are sufficiently out of phase. This is, of course, important in a multiple access environment. It also provides protection against multipath which can be a problem in terrestrial systems and when operating at low angles of elevation in a satellite system.

3.5. Multiple Access Techniques for Mobile Systems.

Most conventional systems basically employ FDMA. The capacity of such a system can be given in terms of the number of simultaneous links the system can provide. Thus, assuming a digital system where the system bandwidth is split into identical channels, the capacity can be expressed as

$$C = \frac{B_s\, k}{R_i\, b_n\, b_g} \tag{3.29}$$

where B_s is the available system bandwidth,

k is the ideal spectral efficiency of a digital M-ary modulation scheme so that $k = \log_2 M$,

R_i is the bit rate at the user,

b_n is a factor that allows for the noise bandwidth (due to non-ideal filtering) and is typically 1.2 and

b_g is is another factor that allows for the guard band and is also typical 1.2.

Given the available system bandwidth and the user bit rate, the only variables that remain are dependent on the modulation scheme and filtering. The spectral efficiency is a parameter that is common to all other multiple access schemes, so that it is not important in comparing multiple access schemes. BPSK will therefore be assumed with a spectral efficiency of 1 b/s/Hz.

If we normalise the capacity, then

$$c = \frac{C\, R_i}{B_s} = \frac{k}{b_n\, b_g} \tag{3.30}$$

Under ideal conditions, the normalised capacity of the FDMA system would be equal to k, in this case unity. However, assuming the typical values, its value becomes 0.69 and the only way to improve on this figure (barring changing the modulation scheme) is improving the filters. This tends to be expensive and is consequently not desirable particularly in the mobile environment.

TDMA systems follow the trend of utilising digital circuitry which through integration is becoming increasing cheap. This makes such schemes suitable for the application we are considering.

For a given system bandwidth, the maximum transmission rate is given by

$$R_s = \frac{B_s k}{b_n} \tag{3.31}$$

so that the capacity of a TDMA system can be expressed as

$$C = \frac{R_s E}{R_i} = \frac{B_s k E}{b_n R_i} \tag{3.32}$$

where E is the frame efficiency. This was defined in equations 3.14 and 3.15. However, for comparison purposes the preambles need not be considered. If short single packet messages are considered, as in a paging system, the preamble will constitute a large overhead. The situation is the same for other multiple access scheme since the preamble in this case is really an overhead of the inherent burst mode of operation. If packetised data is transmitted then the preamble becomes an overhead of the packetisation that is required to operate in a TDMA environment. However, with increasing frame durations larger packets can be accommodated and this overhead decreases. The guard time is unaffected so that in the limit this will become the only limit.

The frame efficiency of the INTELSAT system is about 95 %. Considering that larger frame durations can be used in a mobile system because of the decreasing cost of memory devices, a better figure is expected. However, timing and ranging inaccuracies are expected to be larger for a mobile system. This calls for bigger guard times and results in a lower frame efficiency. The figure of 95 % will therefore be assumed for a mobile system.

The normalised capacity is given by

$$c = \frac{k\,E}{b_n} \qquad 3.33$$

Using the assumed values, the normalised capacity for a TDMA system is 0.79 showing some improvement over the FDMA system. It should be noted that in this analysis the guard band to adjacent systems has not been included and if this is included, the normalised capacity decreases to 0.66 which is similar to the FDMA figure.

If we consider the power-limited situation, the link equation is given as

$$[E_b/N_o] = [EIRP]_{sat} - [BO] - [L_s] + [G/T] - [k] - [R] - [M] \qquad 3.34$$

where E_b/N_o is the ratio of energy per transmitted bit to the noise density,

$EIRP_{sat}$ is the effective isotropic radiated power at saturation level,

BO is the backoff,

L_s is free space propagation loss,

G/T is the figure of merit of the receiver,

k is Boltzmann's constant and is equal to -228.6 dBW/K/Hz,

R is the link transmission bit rate and

M is the link margin.

In comparing FDMA and TDMA systems, we assume the same bit error performance so that E_b/N_o is maintained constant. We also assume that that the same maximum power, figure of merit and link margin are used in both systems. The only remaining variables therefore are the backoff level and the transmission rate.

In a FDMA system a backoff of 3-6 dB is required on the down-link but the transmission rate is considerably lower than that of a TDMA system in particular when considering a large user population system. A TDMA system does not require any backoff. The quotient of the backoff and the transmission rate is therefore more likely to be larger for an FDMA system than a TDMA system. This indicates that a FDMA system is more likely to operate in the bandwidth limited condition than the TDMA system. It should be noted that more power may be available on the up-link since the HPA is operated in the burst mode.

CDMA is relatively new as an access scheme and there is less commercial experience in this field, possibly because it has been considered to be spectrally and power inefficient (VITE 85, SHIM 84). A direct capacity comparison of CDMA with FDMA and TDMA is not easy because CDMA cannot be analysed under power-limited and bandwidth-limited conditions in isolation.

In a CDMA system the received noise is the sum of the thermal noise and the interference due to C - 1 users. The number of orthogonal sequences and the sequence length for a given n-stage pseudorandom chip generator is given in figure 3.12. Since the interference from the other users is noise-like due to orthogonal coding or random phasing, their powers can be added and the total spectral density is

$$N_o' = N_o + (C - 1) E_c \qquad 3.35$$

where E_c is the energy per received pseudo-random chip.

The effective ratio of signal energy per chip to the noise density is thus given as

$$E_c/N_o' = \frac{E_c/N_o}{1 + (C - 1) E_c/N_o} \qquad 3.36$$

However, the processing gain relates the energy per chip to the energy per bit (equation 3.21) and the spread bandwidth to the user bit rate so that

$$G_p = E_b/E_c = B_s/R_i \qquad 3.37$$

and equation 3.36 can be rewritten as

$$E_b/N_o' = \frac{E_b/N_o}{1 + (C - 1) E_b/N_o /G_p}$$

or

$$\frac{E_b/N_o}{1 + (C - 1) E_b/N_o R_i/B_s} \qquad 3.38$$

where we have assumed that the signal is spread to the system bandwidth.

If we further assume that the user population is large then

$$C = \frac{B_s}{R_i E_b/N_o'} \qquad 3.39a$$

This result can also be obtained from the definition of the jamming margin given in equation 3.22. Since this margin gives the ratio of the jamming power to the signal power, it also represents the number of concurrent users the system can support.

The normalised capacity becomes

$$c = \frac{1}{E_b/N_o'} \qquad 3.40a$$

If a value of E_b/N_o' of 10 dB is assumed the normalised capacity is only 0.1 which compares adversely with FDMA and TDMA and thus the conclusion of low efficiency of spectrum use. It should be noted that the capacity given in expression 3.39a represents the number of simultaneous users. It has been shown that CDMA can offer significant advantages if the traffic load from each user is very low (COST 59, UTLA 78).

If we assume that each user is actually transmitting for an average fraction of the time, a, equation 3.39a becomes

$$C = \frac{B_s}{a\ R_i\ E_b/N_o'} \qquad 3.39b$$

and the normalised capacity becomes

$$c = \frac{1}{a\ E_b/N_o'} \qquad 3.40b$$

This indicates that CDMA provides a very good multiple access scheme without the need of accessing protocols where the traffic load from each user is very small.

3.6. Conclusion.

The channelisation properties of TDMA and FDMA are very similar and the number of users that the system can support can, of course, be improved by accessing protocols. Since TDMA systems lend themselves to digital techniques, they appear to be very attractive because such processing utilises integrated circuitry, that subject to production volume, offers great system flexibility at a low cost. However, the high transmission rate required in TDMA system is a serious disadvantage. High powers are required and processing at high bit rates becomes expensive.

It appears that for a large system, the optimal scheme for the mobile up-link is a hybrid, one in which several TDMA systems operate concurrently within a FDMA system. This brings the transmission bit rate down by a factor equal to the number of TDMA systems employed, to frequencies that make processing possible at sufficiently low costs.

In general, it is more difficult to implement on-board regeneration with a FDMA than a TDMA one because this involves a number of regenerators equal to the number of channels and this is a severe drawback. On-board regeneration allows the up-link and the down-link to be optiminsed independently so that the the lowest possible power output and the lowest possible figure of merit of the mobile terminal can be used. A hybrid system could vastly ease this disadvantage by drastically reducing the number of regenerators required.

If on-board regeneration is employed then it seems logical to take the system a step further and utilise a TDM down-link. The incoming packets from the the mobile up-link are concatenated together or multiplexed on the down-link. This means that there is considerable saving in spectrum since it is no longer operating in the burst mode.

Guard times can be dispensed with and the preamble information is required less frequently. The fact that the down-link operates in the continuous mode rather than the burst mode also leads to a simpler terminal receiver design that helps to keep the terminal cost low.

Similarly the connection matrix for a multi-beam system employing FDMA becomes very large making implementation more difficult. Again a hybrid system would ease this problem.

If the number of TDMA systems is sufficiently small, separate HPA's can be employed on the spacecraft to eliminate the need of backoff and this could keep the power per TDMA system low enough to utilise semiconductor devices.

CDMA could provide an excellent scheme for a small system operating on an overlay with other systems sharing the spectrum. If the system is sufficiently small than the apparent rise in the noise floor could be low enough not to adversely affect the existing narrowband systems. Control would obviously be required to limit the interference from such a system. The main obstacle of such a system is that there seems no incentive for present spectrum users to share their allocations unless this is done as an extention to their own system.

Equations 3.39b and 3.40b show that the number of users in a CDMA system can be enhanced if the traffic load per user is sufficiently low without the requirement of any accessing protocols. Though the need of a protocol is not a large price to pay for better utilisation of resources, CDMA can offer a system that provides a service to a large user population with no network control if the user traffic has a low duty cycle. The relaxed requirement on synchronisation is also an advantage.

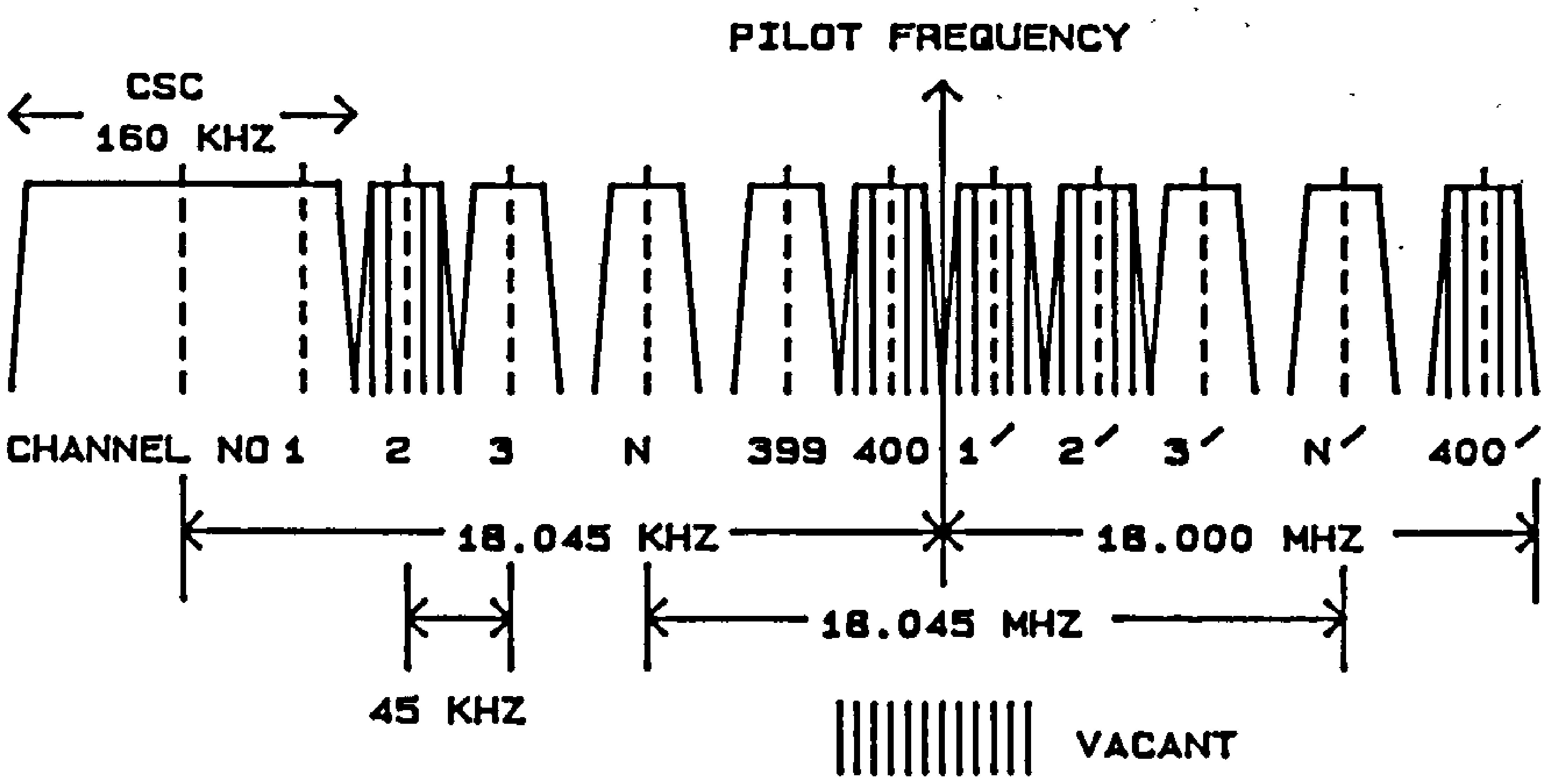


Figure 3.1. INTELSAT SCPC multichannel frequency allocation for a 36 MHz transponder channel.

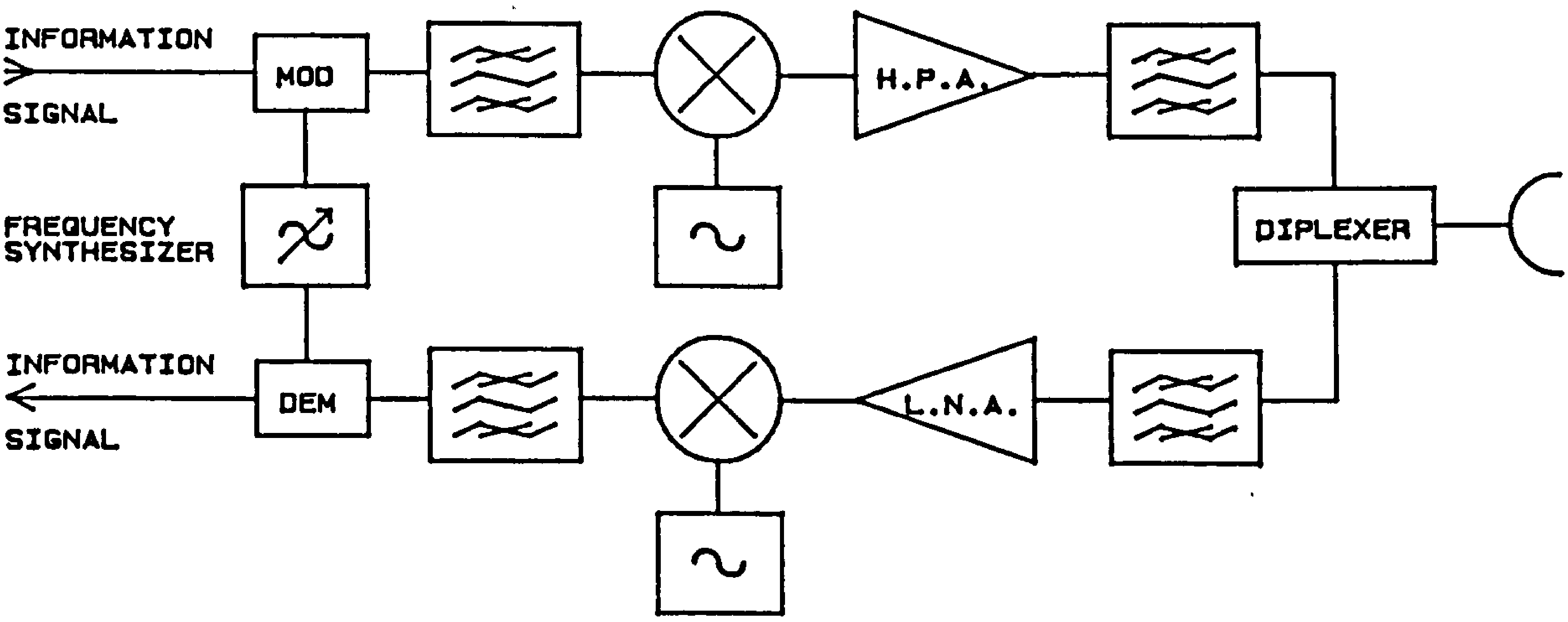


Figure 3.2. Earth terminal.

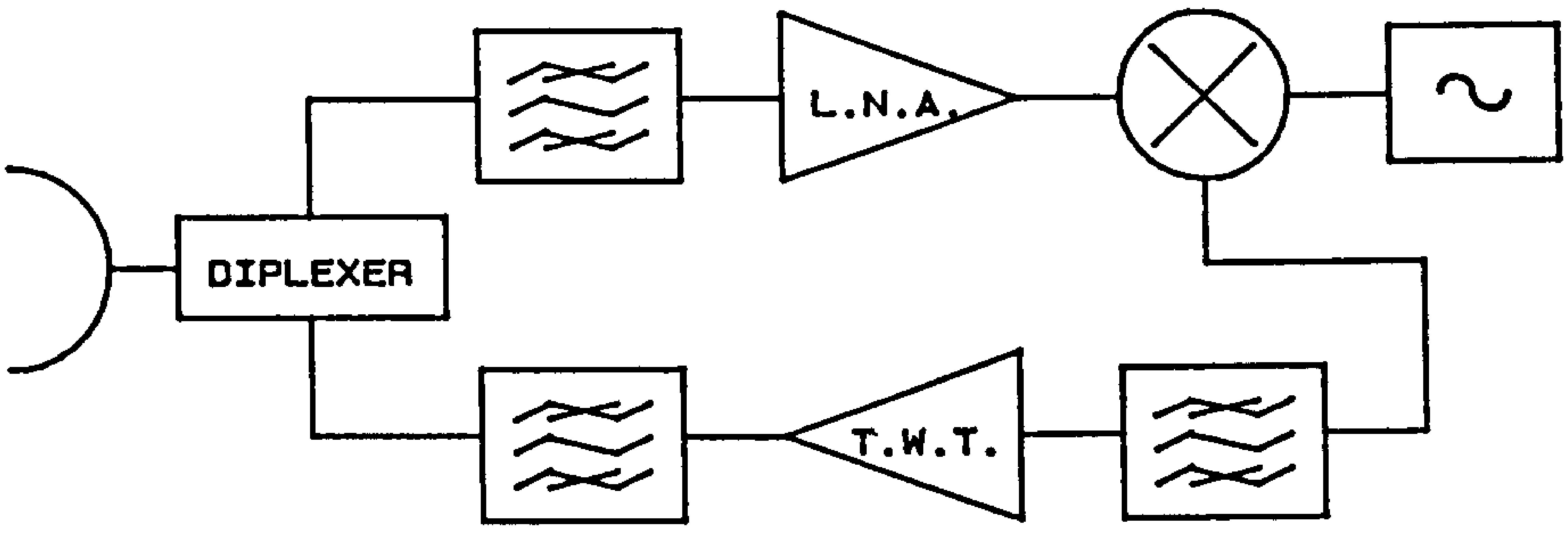


Figure 3.3. Transparent transponder.

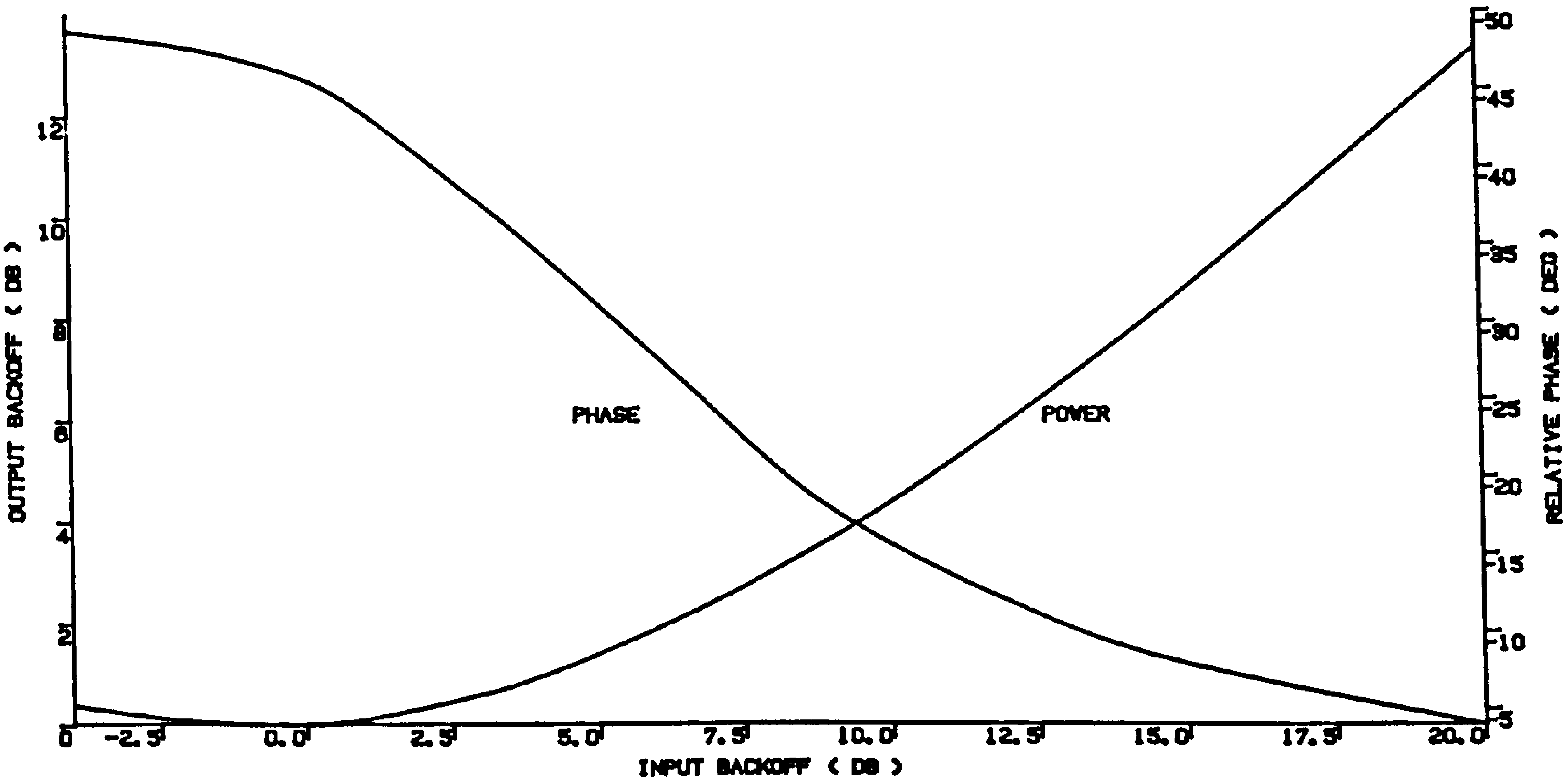


Figure 3.4. TWT characteristics. (a) power output - power input characteristic. (b) output phase - input power characteristic.

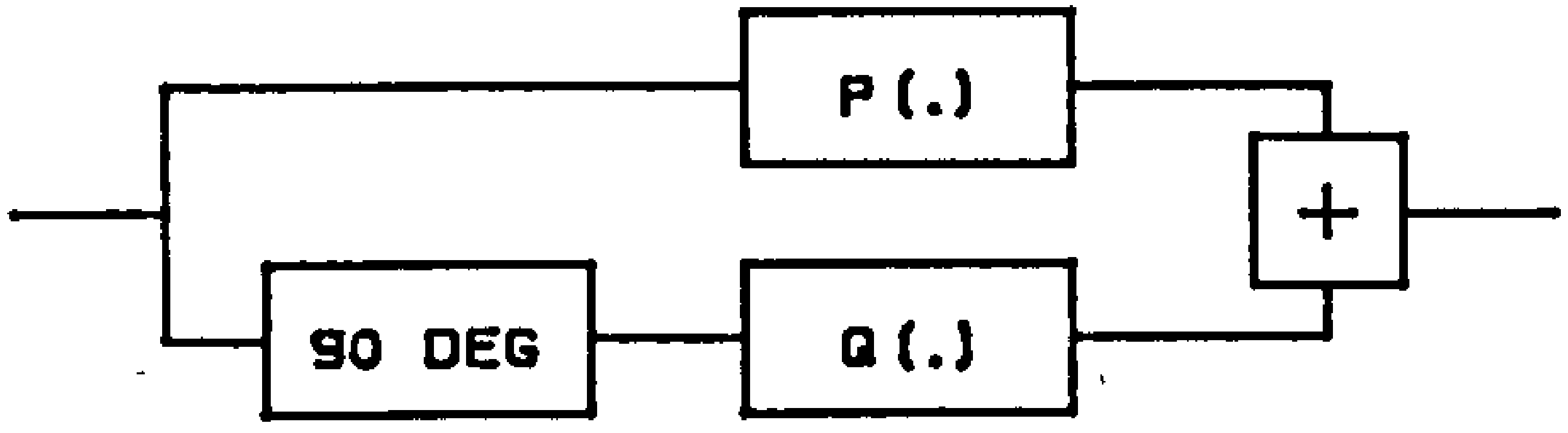


Figure 3.5. Quadrature non-linear model of a power amplifier.

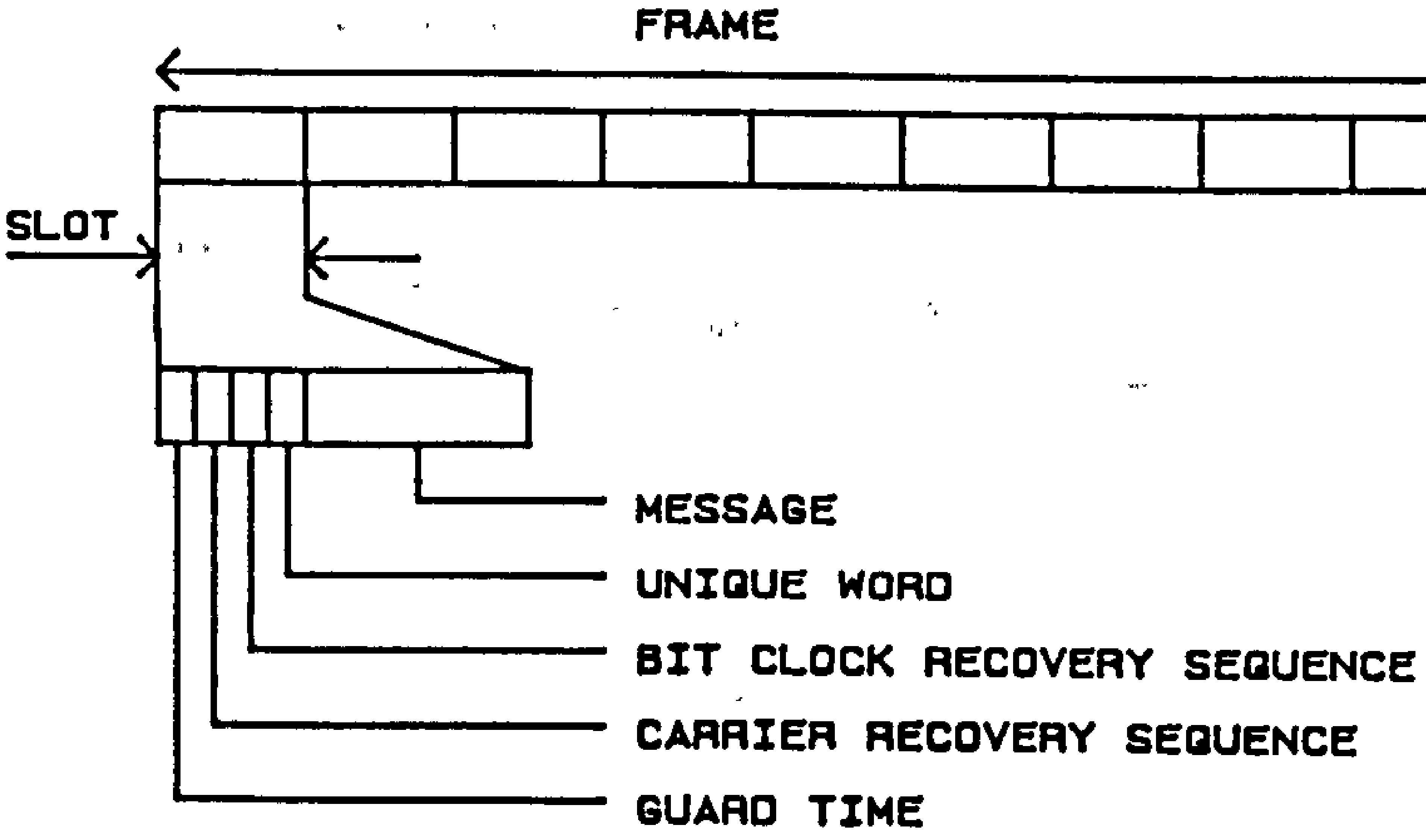


Figure 3.6. A TDMA frame format.

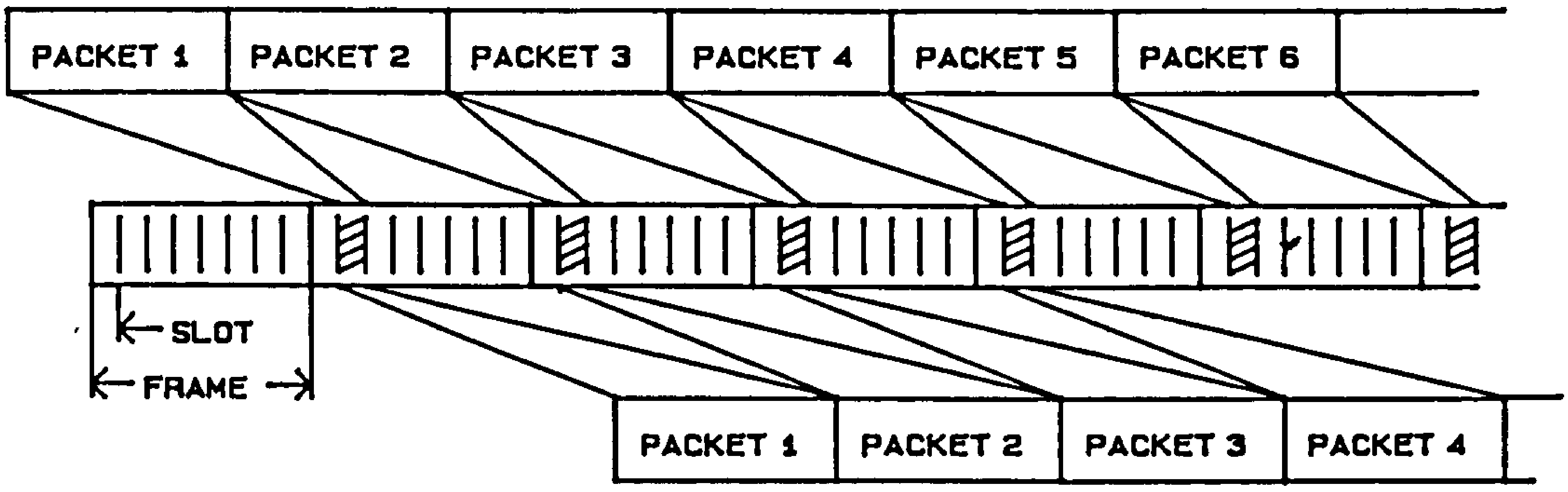


Figure 3.7. Transmission of a continuous bit stream in a TDMA system.

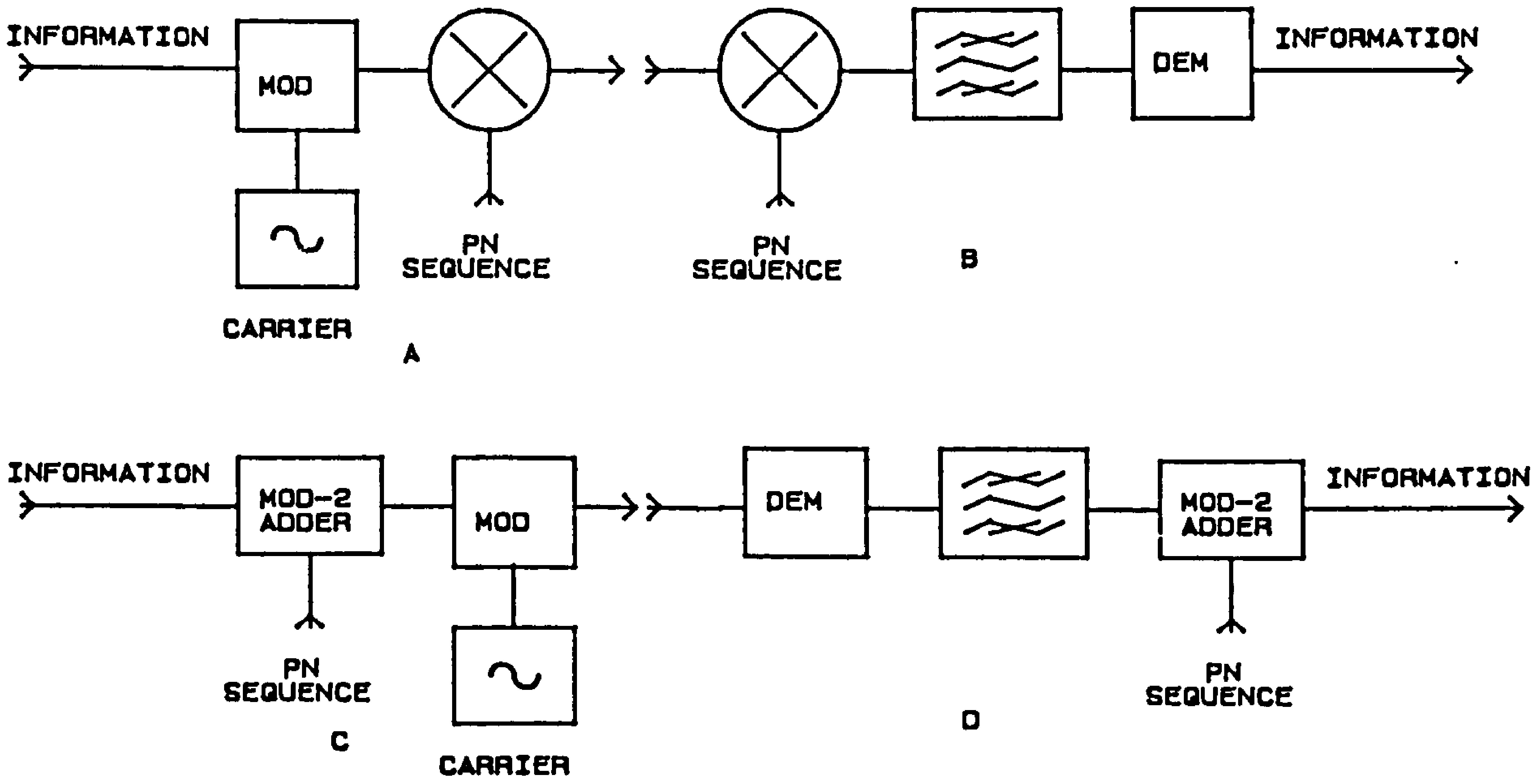


Figure 3.8. Direct sequence technique. (a) and (c) are transmitters while (b) and (d) are receivers.

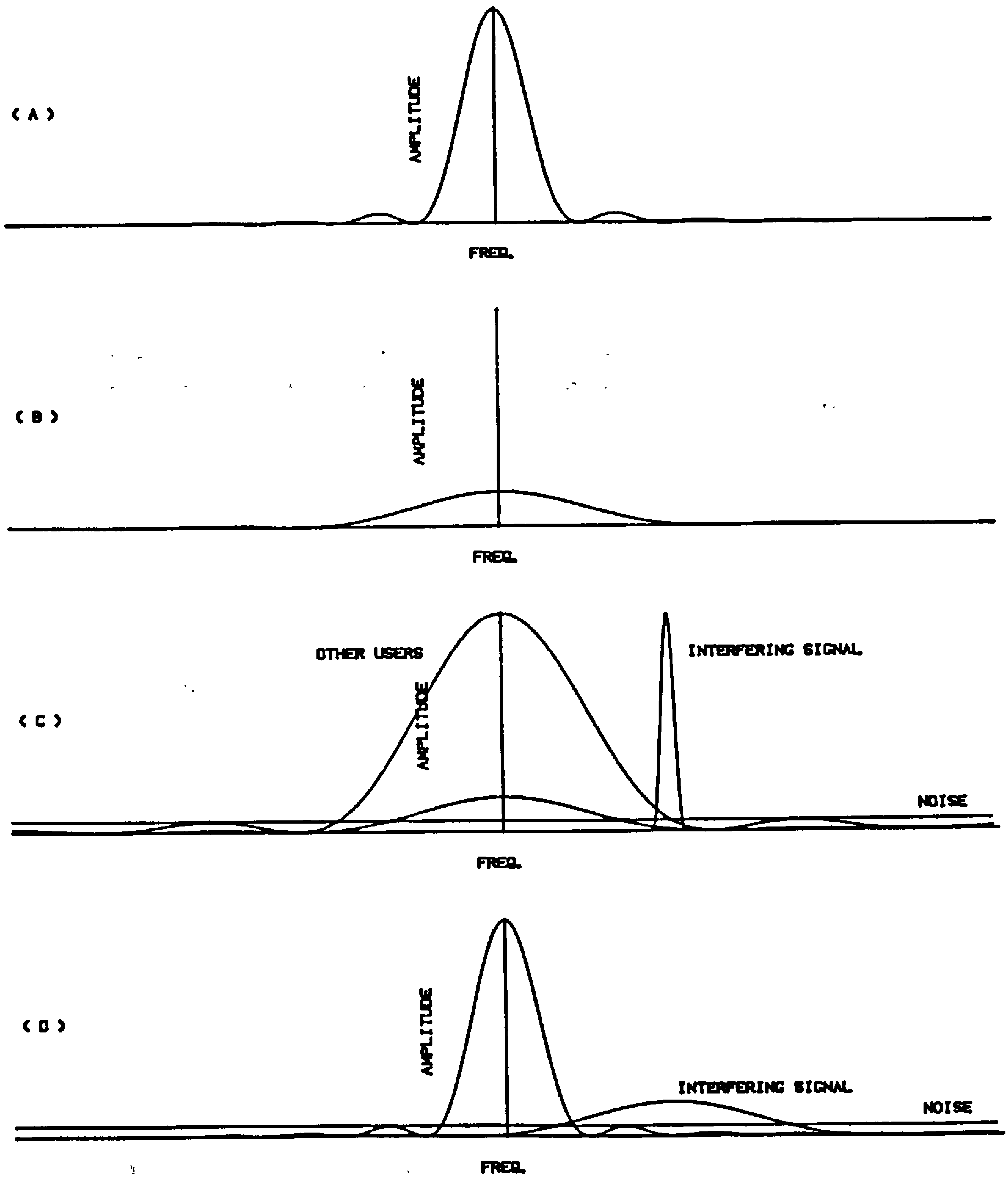


Figure 3.9. Various power spectra in a DS system. (a) power spectrum of a conventionally digitally modulated signal, (b) spread spectrum signal, (c) power spectrum at the receiver input showing contribution from the wanted signal, unwanted signals from other users, thermal noise and a spot frequency interference, (d) post-correlator power spectrum.

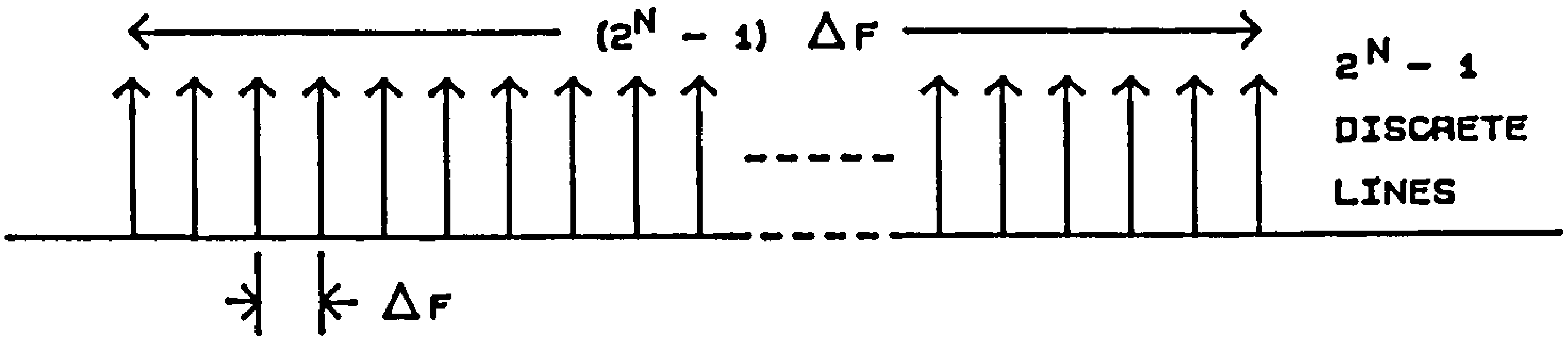


Figure 3.10. Spectrum of a FH system with a n-stage PN code generator and a frequency increment of f Hz.

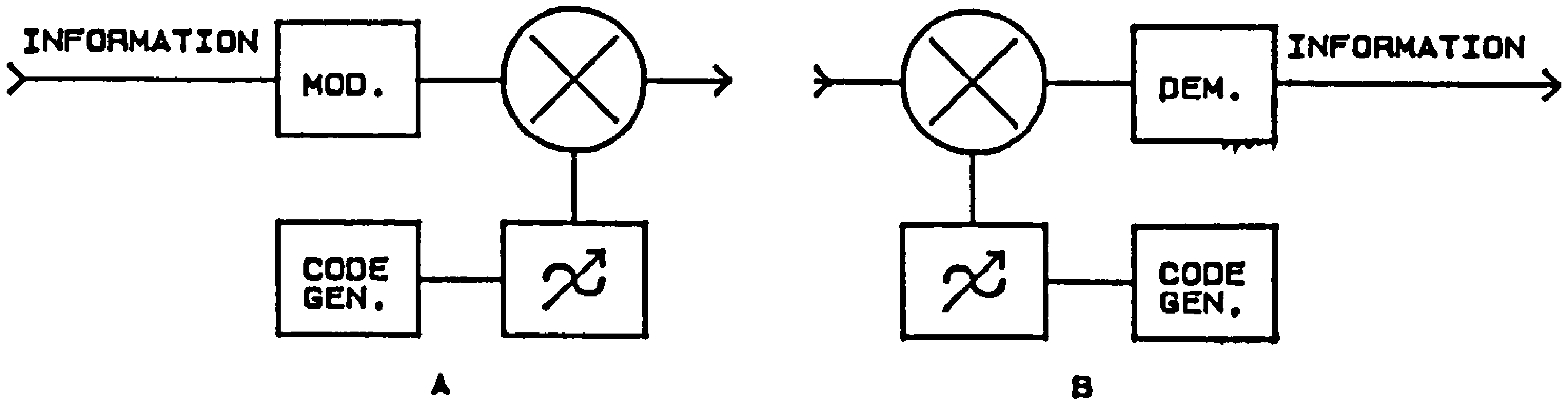


Figure 3.11. Block diagrams for FH transmitters (a) and (b) receivers. while (b) and (d) are receivers.

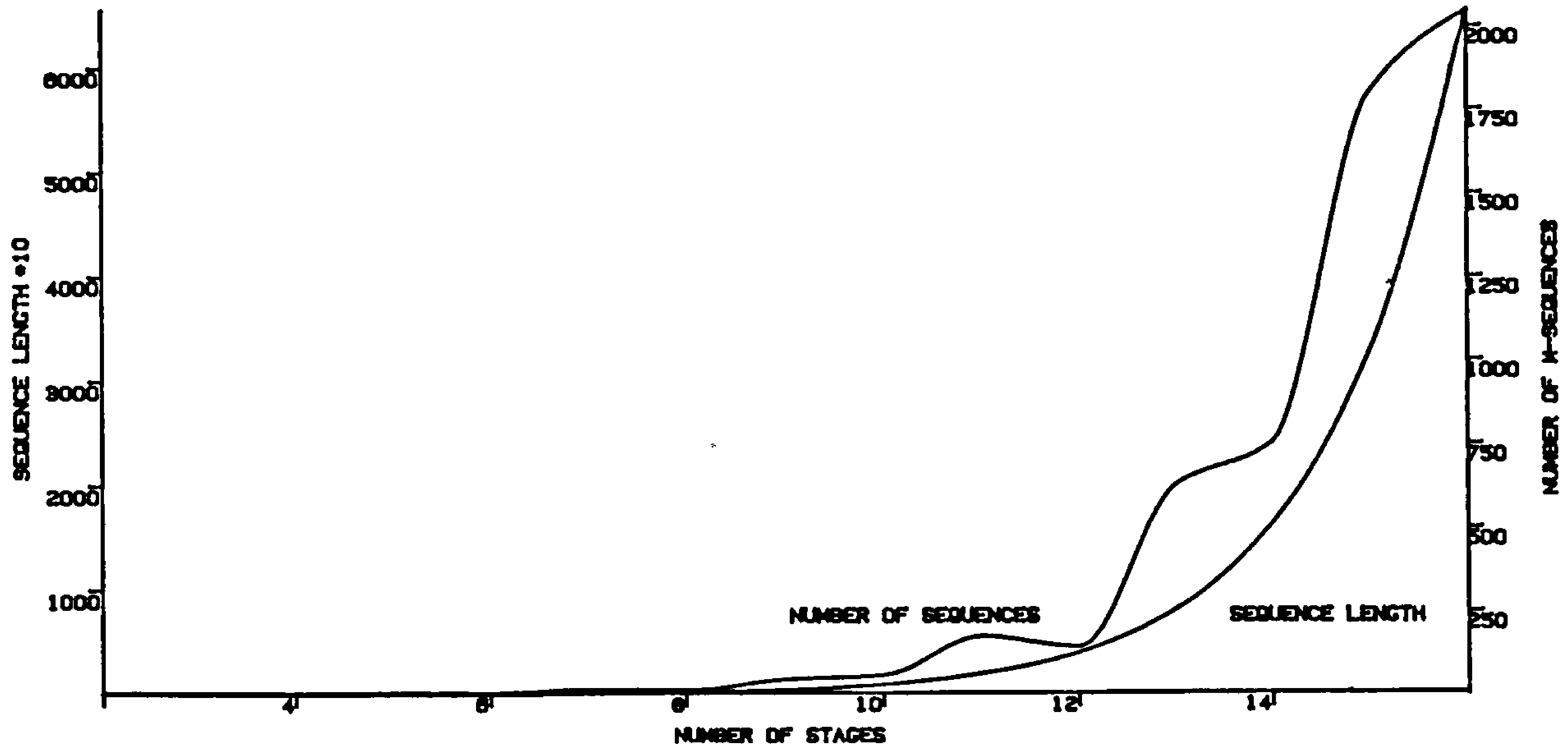


Figure 3.12. Plot showing the variations of the sequence length and the number of distinct m-sequences with n form an n-stage generator.

CHAPTER 4.

TDMA MULTIPLE ACCESS PROTOCOLS.

4.1. Introduction.

The fundamental multiple access schemes discussed so far provide a number of orthogonal links generated by time or frequency channelisation or codes. This forms the backbone of a multi-user communication system. If the number of users is equal or less than the number of links provided, then the ´channels´ could be preassigned. The term ´Channels´ here is used in a broad sense to mean a frequency channel, a time slot or communication over a code.

In a preassigned system, user i will always use ´channel´ i to transmit, and all other users listen to channel i to get messages from user i. Such a system can be usable where the traffic per circuit is close to one erlang. In general this is not the case and particularly so in the mobile scenario where the average load per user lies in the range of 0.008 to 0.03 erlang (WATS 83).

Such a scheme would have a system utilisation in the same region and the number of users would be severely limited. For thin route circuits preassigned schemes provide inefficient solutions: they are inefficient in the utilisation of the space segment resources and of the spectrum available. However, they are fast. Whenever a user wants to transmit, a channel is available and consequently there are no further delays.

With frequency spectrum being a limited resource under such heavy demand and with the space segment representing a significant portion of the system cost, it is unrealistic to think of a mobile system employing a preassigned scheme. The alternative is a demand-assignment multiple access (DAMA). As the name implies, ´channels´ are no longer dedicated but are all put in a pool and the user uses ´channels´ when required and if available. When no longer required the ´channels´ return to the pool.

If traffic intensity per user is low enough, the number of users can exceed the number of ´channels´ and this is generally the case. This means that under busy conditions the demand for ´channels´ may exceed the availability and therefore the user may have to wait until a ´channel´ is returned to the pool. This delay is one of the most important parameters in DAMA and the performance of such systems can be described by the delay-load curve. A more traditional method that has its roots in telephony is to study the circuit load vs probability of blocking performance. The latter is the probability that on application by a user for a ´channel´ from the pool, the ´channel´ is not made available.

A number of multiple access protocols have evolved mainly for computer networks. Some of these are unsuitable in the context of satellite mobile communication. In polling (LAM 79, LAM 83), users are asked sequentially whether they have anything to transmit. If so, the user will acknowledge the request and the user can start transmitting.

This is not suitable because the average walk time between users is dependent on the propagation delay while the average cycle time is dependent on the average walk time and the number of users. Since in satellite communications the propagation delay is in in the region of 250 ms and in a mobile system the population size is large, the delay performance would be poor.

A refinement of the basic polling scheme is adaptive polling. This is particularly suited for a lightly loaded network and makes use of the binary tree search to identify users wanting to use the network. If there is only one user requesting attention then the number of enquiries can drop from N, the total number of users, to one. This results in an improvement in the delay performance as the spectrum ultilisation falls.

Carrier Sense Multiple Access (CSMA) (KLEI 75b, TANE 81) is becoming popular in local area networks where the propagation delay is low. In such systems, the user checks that the channel is idle before attempting to transmit. This is a major departure from the polling scheme in that the users are active and do not wait for enquiries from the system controller.

CSMA takes several forms. In the 1-persistent form, the user monitors the channel and if idle it will transmit. If the channel is busy then it waits until the end of the present message and then commences transmission. Since more than one user might have monitored the channel during the transmission of a message, collisions are possible. In such a case the users involved randomise a delay and start again.

In non-persistent CSMA when users sense a busy channel they do not wait for the end of the present message. Instead, they randomise a delay upon sensing a busy channel. The p-persistent CSMA utilises a slotted channel i.e. there is a fixed time grid. A user senses the channel, and if the present slot is busy then it waits for the next. When a free slot is sensed, it will then transmit its message with a probability of p and defer transmission to the next slot with a probability of 1 - p. When collisions occur the same algorithm is applied.

Due to the sensing mechanism the propagation delay should be very small compared to the packet-size. This is because when a channel is sensed as idle there might be a packet in transit and a collision would occur if another user starts transmitting. This makes CSMA unappropriate for satellite communications.

4.1.1. ALOHA Protocols.

The ALOHA protocol originated from a packet broadcasting radio data network that was implemented in the University of Hawaii (ABRA 70, BIND 75a). The system went into operation in 1971. It was designed to enable seven campuses on four islands to access the main computer via radio links. Two 100 kHz UHF channels were used. This was the genesis of random access techniques.

This first system made use of what is now called pure ALOHA. Users transmit packets when they become ready. If single users access the channel one at a time then messages go through. However, this need not be the situation and if two or more users attempt to access the channel then there is a collision. This means that no message goes through and the users will then generate a random delay and try again.

The process is repeated till the messages go through. The random delay reduces the probability of another collision.

It should be noted that terrestrial systems often utilise the capture effect. This occurs when the receiver can discriminate a strong signal from a weak one, so that the strong signal can be received without error. This results in a improved performance. However this improvement cannot be exploited in a satellite system where the signals are likely to be of similar levels. Further analysis will be restricted to non-capture schemes.

ALOHA systems are simple to implement. No synchronisation is required and strictly speaking packets do not have to be of equal size. The price for this is efficiency. Assuming equally sized packets and an infinite population then the through-put is maximum when the channel traffic is 0.5 i.e. on average there is a demand for half the slots available. In this condition the through-put is 1/(2e) i.e. 18.4 % (ABRA 70). This means that out of the total channel capacity only this fraction is used. In conventional terminology the channel traffic corresponds to the channel erlang load.

If the ALOHA protocol is operated in a synchronised fashion (ROBE 72) the through-put is doubled. In this case there is a rigid time structure, so that packets can only be transmitted in the given time slots. Partial collisions are thus eliminated pushing the through-put to 36.7 % at a traffic load of unity. The price for this improved through-put is system complexity. A system clock is now required to inform all the users of the instances when they can start sending a packet. Due to the time structure this protocol is called slotted-

ALOHA (S-ALOHA).

In this context a collision can be defined in terms of orthogonality. A message does not suffer a collision if it is orthogonal to all other messages. Since, as has been mentioned before, orthogonality can be achieved through frequency and coding, it is interesting to note that one can conceive an ALOHA system operating in these domains. In a frequency system a message can either be transmitted in a randomly chosen channel out of a number of a set or at a randomly chosen frequency in the operating bandwidth. The two systems are analogous to S-ALOHA and P-ALOHA respectively. Similarly randomly choosing a code out of a set could also represent a S-ALOHA system.

It should be noted that in the ALOHA protocols instability is possible. Referring to figure 4.1, the through-put curve shows that when the traffic load is low, most of the traffic is due to actual first attempts, so that as the traffic increases the through-put increases. At the same time the probability of a collision is also increasing so that the portion of the traffic from the backlogged stations - those involved in multiattempts - also increases. Beyond the peak the backlogged traffic is so large that it is occupying a large portion of the channel capacity. Any new station becoming ready will only increase the number of backlogged users. Under such situation the protocol breaks down and external control is required.

This effect can also be seen in the delay curves. In the delay vs through-put the delay increases gradually until the maximum through-put is reached. Beyond this point the delay shoots up. The change in gradient can also be seen in the delay vs load curves though it may be less obvious in this case.

The instability problem makes it necessary to include some form of control. The conventional technique is to control the retransmission random delay. By increasing this delay, the collision probability can be reduced. However this will lead to longer delays.

Another simple method is to reduce the number of users trying to access the system. This reduction in number can either take place in an arbitrary fashion in a homogeneous population or according to the users priority. In a homogeneous population the binary tree approach can be used so that the active population is repeatedly halved till an acceptable through-put is achieved. The active population is then scanned so that averaged with time, all the users are included in the active population. This becomes similar to the multi-access tree protocol suggested by Capertanakis (CAPE 79).

The Urn protocol limits the number of contending users statistically. Ready users transmit with a probability that is dependent on the number of ready users so as to give an average traffic load of unity. One way of achieving this is by introducing a small overload so that a log of the number of ready users can be maintained. The Urn protocol has a performance that approaches the ALOHA on light traffic loads and tends towards the TDMA on heavier loads.

4.1.2. Reservation Schemes.

The protocols discussed so far have their roots in computer networks and are best suited for a message service with a bursty traffic. Most of them suffer from low through-puts which can be tolerated when dealing with say a host computer serving many interactive terminals. In such a case the message may only be a few characters long and to keep system complexity low such schemes may be useful.

The major problem arises in attempting to transmit speech. The user may access the channel and maintain it until the end of the conversation which effectively the implies that the whole conversation becomes a single packet or message. When packetisation is employed, these protocols have no way of efficiently guaranteeing that the delays between packets are always bounded so that undesirable pauses are not introduced.

Reservation protocols offer the solution. It can support packet communications where the delay from packet to packet can be guaranteed without operating inefficiently. This is achieved by splitting the system capacity in two, so that there is a reservation section and a message section. The partition between the two can be dynamically controlled.

In this scheme a ready user will first issue a reservation request (RR). This is done using the reservation channel (or subframe). Once this is accepted, it is queued and when its turn arrives, the ready user can start transmitting the packet.

The idea of explicit reservations on a separate subchannel was first suggested by Roberts (ROBE 73). This system uses a TDMA time structure so that there is a reservation subframe (RSF) and a information subframe (ISF). The slots in the RSF are further split into smaller slots or subslots (figure 4.2). Users use the RSF using the S-ALOHA protocol. The queuing discipline used is FIFO according to the order of reception of the RR's.

On start-up, and every time the queue length is zero, there cannot be any message packets being transmitted, and therefore the frame is wholly a RSF to take in RR's. As soon as a RR is accepted, the ISF is created. The RSF shrinks so that space for the ISF is allowed while maintaining the same frame size, though strictly speaking this is not essential.

Queue length information can either be provided in a distributed fashion or in a centralised fashion. One suggestion is that every packet leader contains such information, so that when a user becomes operational in the system the information can be obtained by listening in. Alternatively, it can be provided by a central controller.

In this scheme once a user gains access to the system, it will use as many subsequent slots as it has applied for in the RR. In this format it is only able to support one voice link at any time. However, it can easily be modified to support more. If each user only gains access to one slot every frame for as many frames as required, then the number of simplex links available is equal to the maximum number of slots in the ISF.

Though Roberts suggests the ALOHA protocol for the RSF, other protocols discussed in the previous section are usable. The important issue to bear in mind is that though the RSF is an overhead, under heavy traffic load it is a very small portion of the capacity. The through-put in the ISF is always close to unity so that the aggregate through-put should be very high indeed. Looking at the RSF as an overhead and considering its size, the system through-put is mainly determined by the size of the ISF relative to the frame size. It is thus not always worthwhile squeezing the most out of the RSF if this implies cost in terms of system complexity.

Similar schemes employing implicit reservations have been developed for data communications. Crowther et al (CROW 73) employed a TDMA system where users acquire slots by contention and a successful slot entitles the user for that slot in the next frame. Control is required to ensure that users do not monopolise slots. This scheme operates well even if the number of users is large or dynamically varying as long as the traffic capacity does not exceed the system capacity.

Binder's technique (BIND 75b) uses dedicated slots but the slots of users with no traffic become available to the rest. A user can reclaim its slot by forcing a collision which would stop all other users from occupying the slot. This system can only operate when the number of users is equal or less than the number of slots.

An out-growth of Roberts' protocol is referred to as priority-orientated demand assignment (PODA) (JABO 78, TOBA 84) and was designed to suit a general purpose packet satellite network. This has been originally implemented in SATNET, a prototype satellite network. The basic structure is unchanged but a more sophisticated scheduling algorithm is used. This takes into account the priority class and delay constraints of the user.

Besides explicit RR's in the RSF, implicit RR's can also be piggybacked into the leader of a packet. Two types of traffic are distinguished when reservations are made: block message or datagram traffic and stream message traffic. Though the frame duration is fixed the portions of the capacity assigned for reservation, datagrams ans streams are dynamically controlled. Only one RR is required for each block message or stream message. In the case of stream messages the interpacket or stream interval is specified so that voice can be supported.

The burst size is variable and depends on factors such as fairness, the length of the RR queue and priority. The variable burst size allows for several messages to be transmitted as one datagram burst thus reducing overheads.

Reservation can operate in one of two modes: fixed assignment and contention. If the number of users is sufficiently small a slot in each RSF is assigned to a user and the protocol is called fixed PODA or FPODA. As the number of users increases such a system becomes impractical and the RSF is accessed on contention basis. This is called contention PODA or CPODA.

Probably the most well known operational DAMA system is called SPADE (Single-channel per carrier PCM multiple-Access Demand-assigned Equipment). This system was designed by COMSAT laboratories for INTELSAT-IV and later satellites (FEHE 83a).

The frequency allocation has been shown in figure 2.1. The 160 kHz channel is the common signalling channel (CSC). Access to the CSC is in the form of a fixed assignment TDMA. The frame is 50 ms long and contains 50 slots. One slot is used as a reference burst and another for test operations, leaving 48 slots usable by 48 SPADE terminals (INTE 85, FEHE 83a, STAL 85). Note that the CSC occupies a very small portion of the system bandwidth: 160 kHz in 36 MHz of the available transponder bandwidth.

Each station maintains a log of the usage of channels. If user A is to communicate to B it randomly selects an idle channel. It requests that channel via its slot in the TDMA system in the CSC. If the channel and station B are available, B acknowledges A on its slot in the CSC and the circuit is established. This procedure takes at least 0.5 s so that in the meantime the chosen channel may not remain idle. In this case the algorithm is repeated. When the call is completed the disconnect information is provided in one of the slots belonging to one of the stations and the channels are released on the return to the pool. SPADE makes use of a distributed control system.

SPADE was not intended for a large user population. In fact its main aim is to supplement the international communication system in order to provide circuits between countries whose traffic does not justify the usage of a whole channel continuously.

A proposed experimental system specifically intended for mobile communications is MSAT-X. The protocol is proposed by Jet Propulsion Laboratories and is known as an Adaptive Mobile Protocol (AMAP) (LI 84). The system is similar to Roberts but is based on a FDMA system. Some channels are reserved for RR´s and the ALOHA scheme is used. The number of channels used for RR´s is dynamically controlled according to the traffic load.

4.2. Analysis of the ALOHA Protocols.

4.2.1. Steady State Analysis.

The traffic rate (G) is the mean number of short messages generated per message duration. It includes messages that are newly generated and messages that have previously been generated but have been involved in a collision and are now undergoing subsequent attempts. It can assume values bounded by zero and infinity. The mean number of short messages per message going through the system is called the through-put (S). Under equilibrium, this is also the mean number of new messages that are generated. The through-put is absolutely bounded by unity and it is also known that it cannot exceed the traffic rate. The difference between the two arise from the extra load due to the messages involved in collisions. Thus under very low load conditions, the through-put is approximately equal to the traffic rate. The discrepancy increases as the load increases. In fact, with increasing load the traffic rate increases monotonically while the through-put increases to a maximum and then drops.

In a P-ALOHA system, there is no message synchronisation and partial collision is possible. This implies that for a message to be uncorrupted, no other message should start anywhere in that duration and anywhere in a message duration before the start of the desired message (see figure 4.3). Having assumed a Poisson message generation process with a mean input of G per message duration, the probability of k message arrivals in two message durations is

$$P[k] = \frac{2G^{k} \; e^{-2G}}{k!} \qquad 4.1$$

Thus, the probability that $k=0$ is e^{-2G}. If this is not the case the message will be corrupted and the probability of a retransmission, whether it is new or a reattempt, is given by

$$1 - e^{-2G}$$

The mean number of retransmissions per message duration is

$$G (1 - e^{-2G})$$

so that since the traffic rate is made up of the rate of generation of new messages and of the retransmission load, we have

$$G = S + G (1 - e^{-2G})$$

or

$$S = G e^{-2G} \qquad 4.2$$

This equation (ABRA 70) yields a maximum through-put of 0.5/e (18.39%) at a traffic rate of 0.5. This indicates a very low channel utilisation which is normally undesirable. Since in this system the message synchronisation is greatly relaxed and since it is only a subsystem, the effect on the total system must be kept in perspective.

The assumption that the message generation rate is a Poisson process implies that there is a finite probability that within a packet duration more than one message is generated. This leads to the possibility of collision of packets originating from the same user. This is obviously not an accurate representation of the practical system. However, it can be shown (SANT 80) that if the distribution function is modified to exclude this, result 4.2 is still valid if the traffic per user is small and if no user dominates. Since our interest here lies mainly in large homogeneous user populations this result will be assumed.

For the S-ALOHA, partial collision is not possible due to message synchronisation into subslots. The probability that a packet from user i is not involved in a collision is given by

$$\prod_{j=1, j \neq i}^{N} (1 - G_j) \qquad 4.3$$

G_i can also be interpreted as the probability that user j generates a message and cannot exceed unity.

The probability that a message generated by i is successful in getting through the system is the probability that user i generates a message while all other users are idle. This is given by:

$$S_i = G_i \prod_{j=1, j \neq i}^{N} (1 - G_j) \qquad 4.4$$

Since we are dealing with a homogeneous population then

$$S_i = S / N \qquad 4.5$$

and

$$G_i = G / N \qquad 4.6$$

Note that though G_i can be interpreted as a probability, G is not, and can be greater than one. Rewriting equation 4.4 with the definitions in equations 4.5 and 4.6 we get

$$S = G\,(1 - G/N)^{N-1} \qquad 4.7$$

and if N tends to infinity then

$$S = G\,e^{-G} \qquad 4.8$$

The maximum through-put occurs at a traffic rate of unity and is equal to $e^{-1} = 36.79\%$.

Result 4.5 can also be obtained by assuming a uniform population where users become busy with a probability G/N. Since each user can either be idle or busy, a binomial distribution can be applied so that the probability that there are r busy users is

$$P[r] = \frac{N!}{r!\,(N - r)!} \left(\frac{G}{N}\right)^{r} \left(1 - \frac{G}{N}\right)^{N-r} \qquad 4.9$$

The probability that there is only one busy user is then given by

$$P[1] = G\left(1 - \frac{G}{N}\right)^{N-1} \qquad 4.10$$

Equations 4.2 and 4.8 are depicted in figure 4.1. In the first portion of both plots, from the origin to the peak, as the traffic load increases so does the through-put. Their values are initially very close, so the gradient is initially unity. As the traffic rate increases the number of collisions also increases causing the discrepancy to increase and the gradient to drop. Beyond the peak a considerable portion of the traffic load is due to reattempts, so that finally the backlogged users maintain themselves backlogged.

Unfortunately there is no way out of this situation without external intervention. Two possibilities exist:

1. The number of users can be reduced. This can be implemented via priority so that a global message can be issued allowing users with certain priority messages to remain active in the system. The others are asked to go dormant.
2. The alternative is to decrease the traffic rate. In this protocol this can be achieved through the control of R. Since R is dynamically controlled depending on the ratio of the loads of the SCM's and long messages, the latter method is preferred in this case.

An estimate of the traffic load can be performed by a central controller from the number of subslots which are idle, busy and those involved in collisions. A distributed measure is also possible by users keeping track of the number of collisions they have incurred.

Result 4.8 is true if all messages entering the system were independent. This is true for the first attempts but is not generally true for the reattempts, unless the retransmission delay distribution extends to infinity. Roberts (ROBE 73) developed an empirically derived approximation which is

$$S = C\,G\,e^{-G} \tag{4.11}$$

where $C = (K - 1)/K$,

K being the retransmission randomisation period in slots.

In satellite communications, K is large since it includes a long propagation delay and C is close to unity. This assumption is made for the rest of the analysis.

4.2.2. Stability Considerations.

In an ALOHA system (pure or slotted), the delivered through-put is the probability of a successful transmission. The users that had a collision and are still awaiting retransmission are said to be backlogged. The probability of a successful transmission can therefore be given as a transmission from the ready users while there are no retransmissions from the backlogged users, or a transmission from the backlogged users while there are no transmissions from the ready users. Assuming an infinite population model this probability is given for the S-ALOHA and P-ALOHA respectively by

$$S_{out} = S\,e^{-S}\,(1-p)^{n} + n\,p\,(1-p)^{n-1}\,e^{-S} \qquad 4.12a$$

$$S_{out} = S\,e^{-2S}\,(1-2p)^{n} + n\,p\,(1-2p)^{n-1}\,e^{-2S} \qquad 4.12b$$

where n is the number of backlogged users while p is the probability of a transmission from a backlogged station.

The expression given in equation 4.12 is very accurate even for finite N if $L \ll 1$ and if we substitute S by $(N - n)L$ (KLEI 76), where L is the message generation rate per user.

Figure 4.4 shows the curves for equations 4.12. This is done in figure 4.4a for S-ALOHA and figure 4.4b for P-ALOHA for various capacities. (The actual capacity is given by the quotient R * A and the frame duration, where the frame is 100 ms long and A is assumed to be constant and equal to 4). In the S-ALOHA case, load lines are included for a constant population size of 100,000 (thus the common intercept point at 0) and for various traffic profiles per user (thus the various intercept points at A, B and C). The load line gives the input rate of the non-backlogged users versus the backlog. If we consider load line BO and the curve at R = 2000, there are three intercept points, indicating three points of equilibrium.

Point X is a stable equilibrium point with a low backlog. In this situation if the system is operating to the left of the point due to some statistical fluctuation, this would cause the backlog to decrease causing an increase in the generation rate from the non-backlogged users and the operating point moves back to the right, to the equilibrium point. If the fluctuation moves the operating point to the right, the backlog increases and so does its traffic and the system reverts to equilibrium. Thus this is a stable point. Similar reasoning can be applied to the high backlog situation (the intercept close to 0), so that this also corresponds to another stable point. Here the delay is considerably higher.

At point Y, an excursion to the left corresponds to a lower backlog. The generation rate from the non-backlogged traffic increases, shifting the operating point further to the left until ultimately it reaches point X. If the excursion were to the right, the opposite effect occurs, so that the increasing backlog moves the operating point further to the left towards the stable point close to point O. Point Y is therefore an unstable equilibrium point. Thus the area to the right of point Y represents an unsafe region.

It has been shown that a system with three intercept points is potentially unstable and similarly it can be shown that a system with one intercept is unconditionally stable but this requires clarification. Looking at the R = 2000 curve, all the load lines cross the curve once implying that the system is stable. The backlog is rather low, indicating that this is a desirable system at least from this aspect. Referring to the R = 40 curve and the AO load line there is also one intercept at O. The stable position is such that all the users are backlogged and the system is consistently useless! This is obvious since the through-put at A exceeds the optimal value considerably and the channel is overloaded.

The same reasoning can be extended to the P-ALOHA situation. This seems to indicate that systems with a large R, i.e. a large capacity are required for stability. This is true but the price for attaining this stability is spectral efficiency. A simple capacity calculation shows that for the given traffic profile and population size, and assuming optimal through-puts, the values of R should be 2 and 3 for S-ALOHA and P-ALOHA respectively. Putting R = 2000 is definitely an overkill!

Assuming operation with realistic spectral efficiencies, the system must be operated in a bistable condition. However, if the probability of reaching the detrimental region is sufficiently low, then the system could still be acceptable. Taking into account the daily traffic load distribution, if the probability that the high backlog stable point is reached once a day is very small, then the system is acceptable.

The bistable nature of ALOHA protocols has been studied by Carleial and Hellman (CARL 75) and Kleinrock (KLEI 75a). The first paper analyses the stability conditions for both types, giving qualitative guidelines concerning stability. Kleinrock's analysis is limited to the S-ALOHA and introduces a stability measure called FET (average First Exit Time). This is the average time in slots required for an empty system to reach the unsafe region, which is the area to the right of the middle intercept point. Appendix A2 gives a FORTRAN 77 program for the evaluation of the FET. This is performed by determining the backlog at the central intercept point, so that the unsafe region is identified and a Markovian model can be employed to determine the average length of time it takes to reach such a state.

The system is modelled as a discrete parameter Markov chain. This is because the number of backlogged stations and the time variable are both discrete and are modelled by the state of the chain and the parameter respectively. The analysis is based on the theory of first entrance times of Markov chains with stationary transition probabilities (PARZ 62).

Consider a Markovian chain N^t with one-step transition probabilities p_{ij}, defined as follows:

$$p_{ij} = 0 \qquad j < i - 1 \qquad 4.13a$$

$$= P[n = 0]\ P[r = 1] \qquad j = i - 1 \qquad 4.13b$$

$$= P[n = 0]\ P[r \neq 1] + P[n = 1]\ P[r = 0] \qquad j = i \qquad 4.13c$$

$$= P[n = 1]\ P[r \geq 1] \qquad j = i + 1 \qquad 4.13d$$

$$= P[n = j - i] \qquad j > i + 1 \qquad 4.13e$$

where $P[n = i]$ is the probability that i new packets are generated,

$P[r = i]$ is the probability of i reattempts, etc.

The one-step transition probability gives the probability of moving from state i to state j in one time increment. The set of probabilities can be compactly specified in matrix form and satisfies the following two properties.

$$0 \le p_{ij} \le 1 \text{ and } \sum p_{ij} = 1$$

Equation 4.13a follows from the fact that in such a system the backlog cannot decrement by more than one in one time interval. The system essentially has a single server.

For the infinite population model, for S-ALOHA these probabilities become

$$\begin{aligned}
p_{ij} &= 0 && j < i - 1 \\
&= e^{-S}\, i\, p\, (1 - p)^{i-1} && j = i - 1 \\
&= e^{-S}\, (1 - i\, p\, (1 - p)^{i-1} && \\
&\quad + S\, e^{-S}\, (1 - p)^{i} && j = i \\
&= S\, e^{-S}\, (1 - (1 - p)^{i}) && j = i + 1 \\
&= S^{j-i}\, e^{-S} / (j - i)! && j > i + 1 \qquad 4.14
\end{aligned}$$

The state space for N^t is the set of cardinal numbers {0, 1, 2, 3, ..., M}. In this case M is infinity. We can partition this set in a safe set $\{0, 1, 2, \ldots, n_c\}$ and an unsafe set $\{n_c+1, n_c+2, n_c+3, \ldots, M\}$. The state space can now be modified to $\{0, 1, 2, \ldots, n_c, n_u\}$, where n_u is an absorbing state. The new transition probabilities can be defined as:

$$
\begin{array}{lll}
p_{ij}' = p_{ij} & i,j = 0,\ 1,\ 2,\ \ldots,\ n_c & \\
\quad = \sum_{l=n_c+1}^{M} p_{il} & i = 0,\ 1,\ 2,\ \ldots,\ n_c;\ j = n_u & \\
\quad = 0 & i = n_u;\ j = 0,\ 1,\ 2,\ \ldots,\ n_c & \\
\quad = 1 & i = j = n_u & 4.15
\end{array}
$$

Defining the random variable T_i as the number of transitions required from state i to reach the unsafe absorbing state n_u, we can say that such a change of state can either occur directly or via some intermediate state T_j. This is given as:

$$
\begin{array}{lll}
T_i = 1 & \text{with a probability of } p_{in_u}' & \\
\quad = 1 + T_j & \text{with a probability of } p_{ij} & 4.16
\end{array}
$$

so that the mean value of T_i is

$$\overline{T}_i = 1 + \sum_{j=0}^{n_c} p_{ij}\, \overline{T}_j \qquad i = 0,\ 1,\ 2,\ \ldots,n_c \qquad 4.17$$

Now $\overline{T}_0$ is the number of transitions required for an empty system to reach the unsafe state which according to our definition is the same as the FET. Solving equation 4.17 is by no means a trivial task especially for large values of n_c. This equation forms a set of n_c+1 linear equations. The classical method would therefore be the application of the Gauss elimination technique using computers. This is quite demanding in terms of computer time and memory space. Noting that in equation 4.13a $p_{ij} = 0$ for $j < i - 1$ allows us to utilise an algorithm (KLEI 75a) that uses each component of the probability matrix only once.

This saves considerable storage space though the amount of computational effort is similar to the other method. In the algorithm produced in appendix A3 further rearranging of the equations was required and scaling factors were introduced in order to minimise problems associated with propagation of errors, overflow and underflow. The program functioned satisfactory within the area of interest when the Cray X-MP/48 was used. (This machine has a word size of 64 bits.)

A similar analysis for the stability of the P-ALOHA case is exceedingly difficult. In fact, very few analytical results regarding the behaviour of P-ALOHA exist. The corresponding stochastic process model is a continuous parameter semi-Markov process. In this case, message synchronisation is not employed and the time spent in a particular state is dependent on that state and the next state in the process. This duration is independent of all previously occupied states and the lengths of stay in those states.

The distribution function of the time spent in a state cannot be assumed to be exponential because of the preemptive nature of the system and consequently the process becomes a semi-Markov one. This makes the analysis very difficult and cumbersome to pursue (CARL 75, FERG 77, HAYE 86).

Perhaps the most rigorous analysis of P-ALOHA is due to Tsybakov and Bakirov (TSYB 84). In their paper, an accurate mathematical model is presented and the stability issue is also addressed. New packets arrive independently at individual stations with the possibility of queuing. The times of transmission are determined a renewal process. Unfortunately this paper does not deal with the bistable condition and defines stability in terms of queue length at stations being bounded under all conditions.

It is thus evident that an analytical approach to the characterisation of the stability of P-ALOHA is not readily tractable and consequently simulation was adopted as a practical solution. A simulation program is developed and discussed in chapter 6.

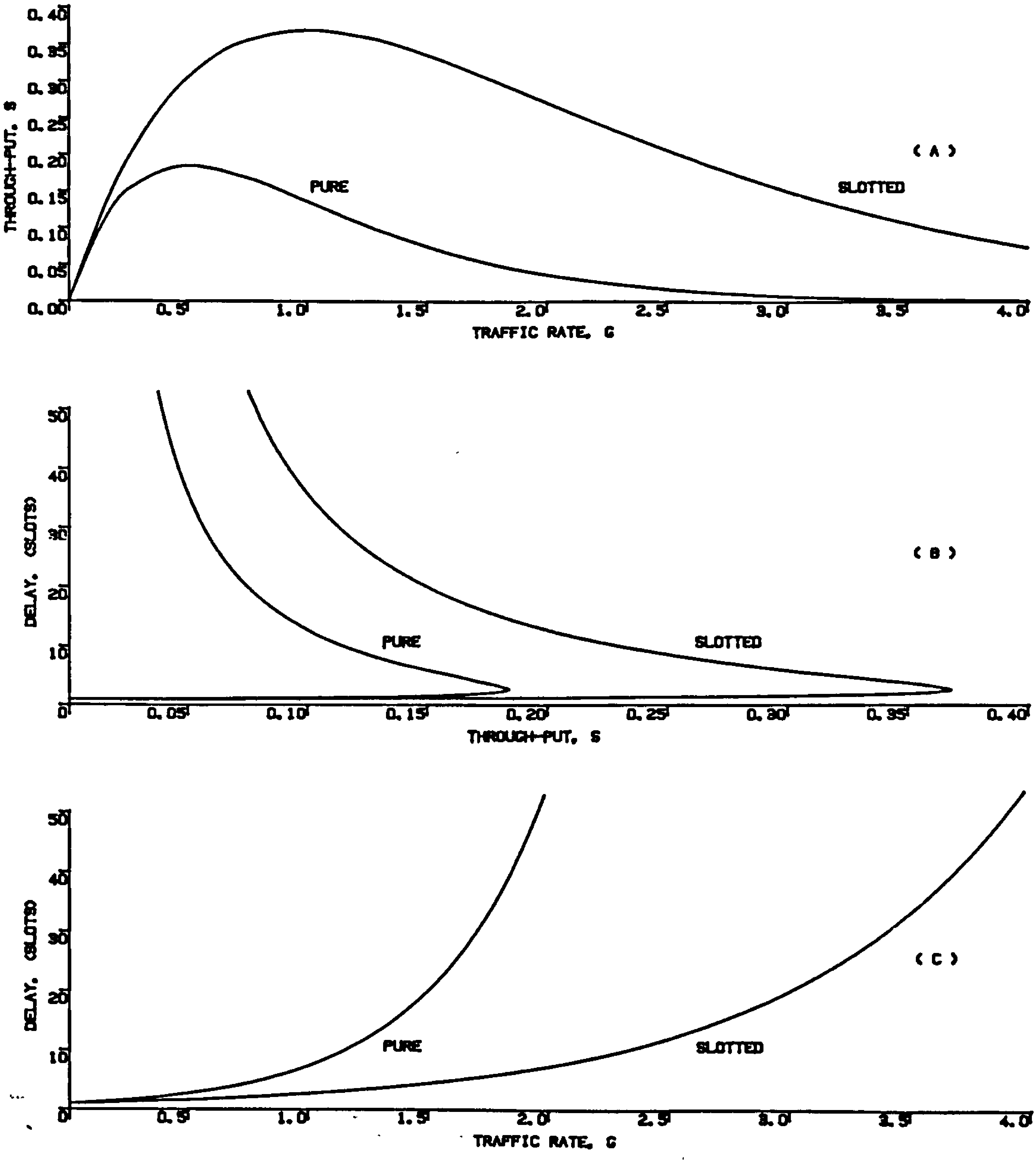


Figure 4.1. Through-put and delay curves for the Pure and Slotted ALOHA.

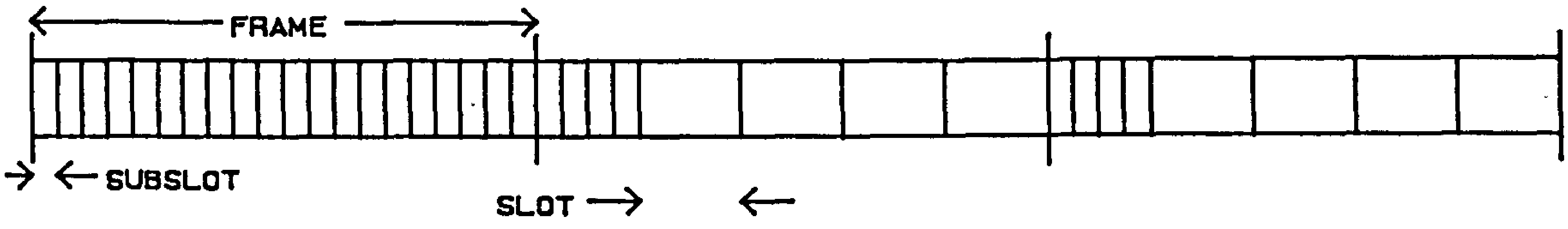


Figure 4.2. Reservation system channel division.

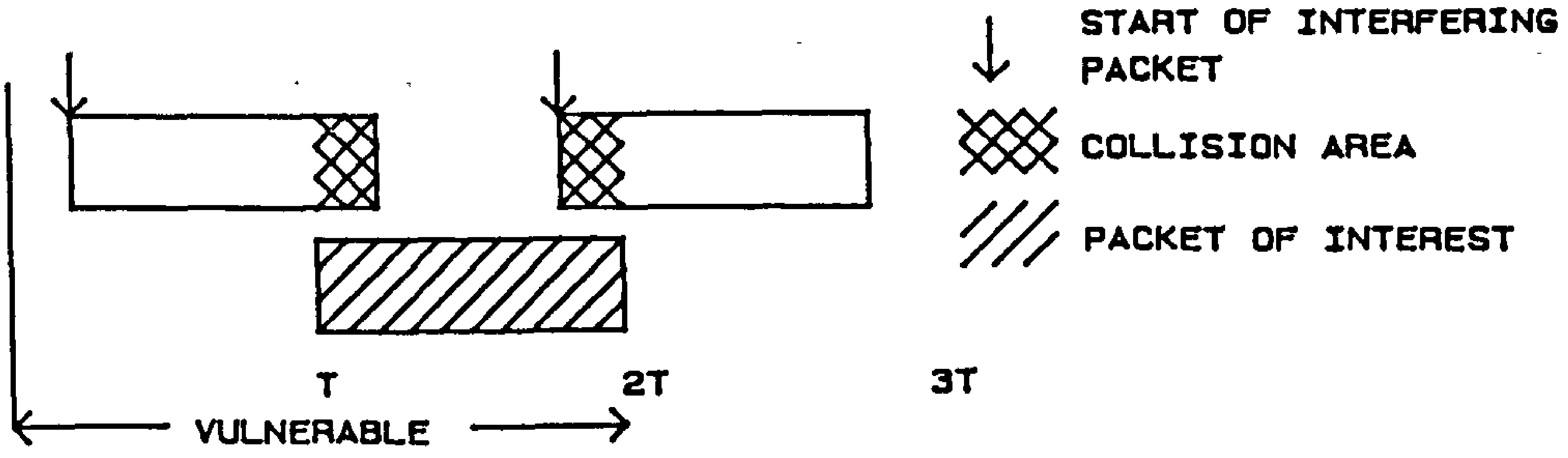


Figure 4.3. Vulnerable period for hatched message.

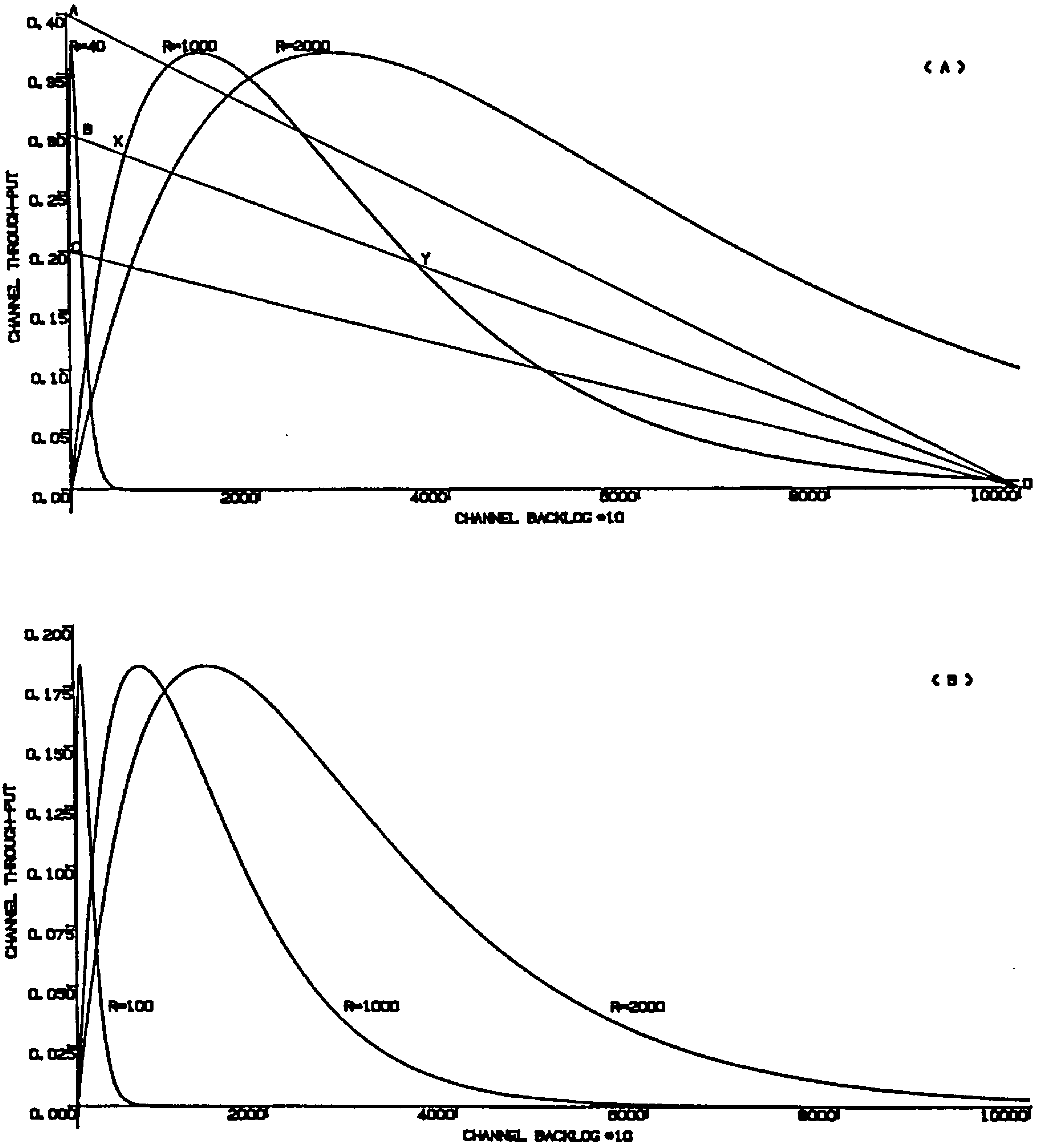


Figure 4.4. Backlog versus channel through-put performance for various channel capacities for (A) S-ALOHA and (B) P-ALOHA. Load lines are also shown for the S-ALOHA.

CHAPTER 5.

TDMA PROTOCOLS FOR THE MOBILE ENVIRONMENT.

5.1. Nature of Mobile Communication Traffic.

Mobile communication traffic is characterised by its bursty profile and by the number of users involved. As far as access protocols are involved this presents a similar situation to computer accessing protocols. The common situation is a host computer with a large number of terminals, all trying to access the host. The bulk of the load on the network is messages of only a few tens of characters. In satellite communications there is, of course the additional problem of the long propagation delay. This affects both computer systems and mobile communications, though the problem is more severe with speech.

Calls originating from mobiles tend to be relatively short (though longer than the situation just described) and are becoming increasingly varied in nature. Conventional communication systems deal with speech only but this is changing very rapidly, data and messages are claiming a considerable share of the market. Much effort has been put into providing data communications on to what are essentially voice systems. This is particularly true for terrestrial systems where it seems unlikely that data rates in excess of 4.8 kb/s will be feasible.

The situation is further complicated by the fact that there is very little operational experience in the field of mobile satellite systems. The only existing commercial system is INMARSAT. However, new systems may be significantly different thus making the available data of little use. This is particularly so when one considers the trends to smaller data-type terminals and the trend towards a multiservice system. INMARSAT is also investigating an aeronautical service and in Phase I of ESA's PROSAT, all three services have been considered.

A recent workshop at ESTEC (ESA 86) has indicated large variances in market surveys and expectations of such systems. It is unlikely that land mobile satellite systems will compete with the terrestrial counterparts particularly in Europe where the cellular systems offer a more cost effective solution. It can be envisaged that satellite systems will compliment land-based systems and that economics dictate that a large user population is required. This may be achieved by satellite systems providing a combined service to the land, maritime and aeronautical mobile communities to increase the market penetration.

5.1.1. Voice Communications.

The largest demand of market still lies in voice communications. Though this is certainly not the most efficient way of communicating information, it is very popular and it does not seem to be drastically different in the future. In terrestrial systems, most schemes use 12.5 kHz channels though there is a move towards 5 kHz. Analogue modulation is still very popular. The audio information can be filtered. The telephony standard is 300-3400 Hz. In some applications like military and amateur radio the bandwidth is further limited to 2.4 kHz. This does not severely affect intelligibility but makes it difficult to recognise voices.

The statistics of the call duration depend on the type of user. This can be categorised in three ways:

1. Public safety: This includes the police, fire, ambulance, highway maintenance, forestry conservation, rescue etc.
2. Dispatch service: A central dispatcher communicates with a fixed number of vehicles in a fleet. Examples are taxis, repair trucks, delivery vehicles, etc.
3. Radiotelephone: This is just a mobile extension of the conventional telephone network.

It is generally accepted that voice call have durations that obey an exponential distribution. Results presented in a paper by da Silva Curiel (DASI 84) confirm this for INMARSAT. The average message length for radiotelephone calls is no different from normal telephone usage. This is about 3 min. For dispatch service it is about 15 s (DUFF 80). Statistics for public safety messages lie somewhere in between.

5.1.2. Data Communications.

Data messages can be of various forms. A complete message can be contained in a single packet with about 100 data bits. This would be the case for paging messages or possibly short coded messages (SCM). SCM's provide an efficient way of communicating brief messages. A dedicated set of messages could be devised for a particular user environment. For the police, pressing a button could send the message "Car number N needs help". An address for taxis could be keyed in a given format at the base station and appear in the vehicle on a CRT or LCD display. In some cases voice messages can be synthesised at the receiver end.

Longer messages could require several packets and these could have a definite length (where the number of packets is known at the beginning of the transmission) or an indefinite length. The possibility to communicate still pictures is very attractive in several areas. Police cars could access a data base of suspect pictures at the base station and could obtain pictures in a car.

Fire appliances could also access a similar data base and obtain plans of buildings on fire while it is en route, possibly indicating areas in the building with particular hazards. In the commercial field a sales engineer could ask for small drawings from his car while on site.

Telex service also falls into this category and is one of the services supported by INMARSAT, besides other optional facilities like high speed (56 kb/s) and low speed (2.4 kb/s) data communications and facsimile transmissions (DASI 84). Telex is normally transmitted at 50 bauds. It does not, however, make sense to transmit at this rate over a channel of higher capacity. Buffers could be used to interface standard telex equipment to higher capacity channels. For most applications a hard copy is not required so that electronic displays could be employed. A typical text message is 25 words long (RADI 82) which a equivalent to approximately 900 bits (6 characters per word, 6 bits per character). The probability of character error is 0.01 %.

Field tests for facsimile have been performed in the maritime environment (CCIR 24). These tests were conducted at a transmission rate of 4.8 kb/s using two-dimensional coding with a line density of 3.85 lines/mm. The CCITT SG XIV test documents were used and the document size was ISO A4. At high signal level condition (no retransmissions due to erroneous frames), the transmission time for document numbers 2, 6, 8 and 5 were 16, 22, 23 and 38 s which give an average of 120 kbits. Though this system uses a HDLC protocol, channels with error rates of 10^{-6} can be used for error free transmission.

5.2. Design of Protocols.

As has been stated previously the difficulty in accessing protocol is related to the population size, and the variety and the bursty nature of the messages encountered in mobile communication. In contrast with terrestrial system the capital outlay of satellite systems forbids dedication. A satellite systems would have to cater for as large a portion of the communication traffic profile as is possible in order to make it financially attractive and feasible.

It is evident that the main problems are with the mobile-to-base link and consequently this will be treated in detail. In chapter 3 the various multiple access systems have been discussed. CDMA has been put forward as a system that is well suited for a simple system where minimal network control is desired and can support a large population if the traffic load from each user is very small. It can also be overlayed with other operating systems sharing the spectrum with very little interference to the existing systems. CDMA systems appear to be attractive for services such as paging and other message orientated services. Of course, its properties of low detectablity, anti-jamming and encryption make it very attractive in the military sphere.

A large system offering a wide range of services can make more efficient use of the resources, namely spectrum and the available power, by employing TDMA or FDMA or a hybrid of the two. Chapter 3 went into considerable detail to show the advantages of a TDMA/FDMA hybrid system.

For simplicity of the analysis, the protocol will be assumed to operate in a TDMA environment though the application of the protocol to a FDMA or a hybrid system is also possible. Though conventional satellites normally only employ a frequency translating transponder the analysis will be also performed for systems that include more on-board processing.

5.2.1. Frame Structure.

The protocol employs reservation techniques similar to Robert's protocol (ROBE 73) and its more sophisticated version Contention Priority Oriented Demand Assignment (CPODA) (JACO 78). None of these protocols have been specifically developed for the mobile environment. Therefore, a simple protocol will be designed and analysed to cater for SCM, data files and speech in the mobile system with a large population of similar small users.

The reservation channel can, of course, occupy a spread spectrum channel. However, in order to keep the hardware complexity of the mobile terminal low, this has not been employed and using a reservation subframe has been considered to be a more elegant solution, particularly when maximum flexibility is desired. As the reservation traffic increases, the traffic in the information subframe would be subjected to an apparent rise in the noise floor. This can render the system unusable especially if voice is considered. Of course, limits can be imposed on the reservation channel traffic but this reduces the system flexibility.

The analysis will also be extended to investigate stability. The basic frame structure is illustrated in figure 5.1. The frame is t_F s long and is divided into F slots. F and t_F are fixed. The frame is split into two subframes: the reservation subframe (RSF) and the information subframe (ISF) with R and (F-R) slots respectively. Two possibilities exist for the RSF: Pure ALOHA (P-ALOHA) and Slotted ALOHA (S-ALOHA). In the P-ALOHA, the subframe duration (equivalent to R slots) is used continuously and the only boundaries are the beginning and end of the RSF. Inter-slot boundaries within the RSF do not exist. In the S-ALOHA, since the slot capacity for RSF purposes is quite high, each slot in the RSF is further subdivided into A subslots, so that there are A*R subslots in a RSF. A is fixed while R is dynamically controlled in both cases.

Messages transmitted over the RSF are either SCM or reservation requests (RR). RR's would include identity numbers of the originating and destination stations; a code which gives the priority of a message; the type of message (i.e. RR or SCM) and the size of the message in terms of packets required in the case of a finite length data message.

Besides the bits concerned with the core information, a preamble and some parity bits will also be included. Since the accessing protocol in the RSF does not distinguish between a SCM and RR we shall use a short message to mean either. Typical formats are shown in figure 5.2. The extra bits available in RR´s can be used as parity bits so as to improve the error performance.

5.2.2. The Reservation Subframe.

The proposed accessing protocol in the RSF is ALOHA and both the pure and the slotted variants will be considered. This protocol is very well suited for this application where the propagation delay is long, the population size is large and the traffic profile is bursty. The S-ALOHA differs from the pure ALOHA in that packets are transmitted so as to fit a given time slot, in this case a subslot. In the other scheme the only timing constraint is that the message must lie within the confines of the RSF. No other message synchronisation is required. This makes the hardware simpler but the price is that under optimum conditions twice the capacity is required.

The protocol for a mobile to send a short message is given below in a PASCAL-like algorithm:

```
BEGIN
   ACK := FALSE;
   StartOfFrame := FALSE;
   SystemSaturated := FALSE;  { Indication to user }
   Timer := 0;
      { Timer time units:
           P-ALOHA: continuous,
           S-ALOHA: subslot duration. }
   TimeOutTimer := 0;
   Compose ShortMessage;
   REPEAT
      IF (NOT BufferFull) OR (Message = SCM) THEN
      BEGIN
```

```
          SystemSaturated := FALSE;
          Generate X;
             { P-ALOHA: Uniformly distributed random real,
                        $0 <= X < (A*R*t_{subslot})$,

                S-ALOHA: Uniformly distributed random integer,
                        $0 < X <= (A*R)$. }
          Wait for StartOfFrame;
          Start SubSlotCounter;
             WHILE SubSlotCounter < X DO;
             Transmit ShortMessage;
             Start TimeOutTimer;
          WHILE (NOT ACK) AND (TimeOutTimer <= TimeOut) DO
             Wait for ACK;
       END
       ELSE
          SystemSaturated := TRUE;
    UNTIL ACK;
    Transmission Of ShortMessage Completed;
 END.
```

The mobile first composes a short message. An attempt for a RR is only initiated if the system is not saturated. The global Boolean message **BufferFull** from the controller indicates the state. The user is made aware of such a state through SystemSaturated.

A random delay is required for both ALOHA systems. For the P-ALOHA, the delay is in the continuous time domain, whereas for the S-ALOHA, the delay is in a discrete time domain with the subslot duration as the time unit. In both cases the random delay is spread over a RSF duration less a subslot duration. Only uniform distributions are considered to ensure that delay does not extend to the ISF causing data corruption.

Having randomly generated a value of the delay for the appropriate system, it then uses this delay to select the epoch in the RSF of the next frame when the short message is to be transmitted. The delay is referred to the beginning of the RSF. This means that a mobile has to plan one propagation delay in advance. The message will be acknowledged by the controller if, and only if, only one user made use of that message duration. If more than one user used the message duration (wholly or partly), all the messages involved will be garbled and no acknowledgements are transmitted.

After the message transmission is initiated the mobile waits for a given time, termed time-out period. If no acknowledgement is received within this period, the retransmission of the message at a randomly selected epoch in the next available frame is repeated. The whole process is repeated until finally an acknowledgement is received.

The delay incurred by the process per attempt is heavily dependent on the time-out period, which in turn is dependent on the propagation delay. In a satellite whose processing power is frequency translation only the time-out period must be at least two round trip delays (about 500 ms). However, if error correction and collision detection can be installed aboard then the time-out period can in practice be halved.

The level of on-board complexity is crucial. Typical down-time figures for the space sector is 0.001 %, while in the maritime environment a typical figure for the coast earth station (equivalent to the network controller) is 0.05 % (CCIR 18). Once collision detection is available on-board it seems logical to strip the incoming messages of the preamble and guard-time and concatenate the messages together so that the down-link is a continuous bit stream.

In effect some receiver complexity has been transferred to the spacecraft. The terrestrial receiver now operates on a continuous bit stream, not in the burst mode. There is also the advantage that less capacity is required on the down-link. For example in the INMARSAT TDMA telex channels, a slot is 79.1 ms or 379 bits long and the information is 12 six-bit characters (LIPK 77). Such on-board concatenation would reduce the down-link capacity requirement by a maximum of about 81 %. In practice such a reduction could not be achieved since some synchronising information is still required, though not as frequently.

5.2.3. The Information Subframe.

Some of the short messages being accepted will be RR´s. A PASCAL-like algorithm that the controller will use to process these RR´s is given below:

```
VAR RR : ARRAY [Addressor,Addressee] OF StationID;
    Buffer : ARRAY [0..(BufferMax-1)] OF RR;
PROCEDURE PutRR;
BEGIN
   Buffer[In] := RR;
   Transmit ACK to RR[Addressor];
   In := (I + BufferMax - 1) MOD BufferMax;
   Counter := Counter + 1;
   IF Counter > BufferMax THEN
   BEGIN
      BufferFull := TRUE;
      Transmit BufferFull to All;
   END;
END;

PROCEDURE GetRR;
BEGIN
   I := F - R + 1  {First slot in ISF.}
   WHILE (I < F) AND (Counter > 0) DO
      I := I + 1;
      IF NOT SlotBusy[I] THEN
      BEGIN
         Assign Slot[I] to Buffer[Out,Addressor]
                      and Buffer[Out,Addressee];
         SlotBusy[I] := TRUE;
         Out := (Out + BufferMax - 1) MOD BufferMax;
         IF Counter = BufferMax THEN
            Transmit NOT BufferFull To All;
         Counter := Counter - 1;
      END;
```

```
      END;
   END;

   BEGIN
      Initialise Buffer;
      Counter := 0;
      In := 0;
      Out := 0;
      IF BufferFull THEN
         Transmit BufferFull To All Once Per Frame
      ELSE
      WHILE RSF DO
      IF InRR THEN
         PutRR;
      IF Counter > 0 THEN
         GetRR;
   END.
```

At this stage the incoming RR's are those that have gone through the ALOHA process in the RSF. The method used is that these RR's are buffered and as soon as a slot is made available the slot is assigned to the addressor and addressee. The slot is maintained for as many frames as are required to complete the message, transmitting a packet per frame. This assumes simplex or half-duplex operation. For full-duplex operation, two slots will have to be assigned. The operation of the input process (ALOHA), the buffering action and the output process is illustrated in figure 5.3. The available number of slots in the ISF is equivalent to the number of servers when the system is viewed as a queuing system. This will be discussed further in the analysis.

Being a practical system the buffer size must be finite. The buffer space per customer is small. All that needs to be stored is two identification numbers and possibly some indication of the size of message to be communicated in the case of a finite length message. This means that the buffer size can be quite large.

When the buffer is full a global message is sent indicating the situation. Upon reception of such a message, the users will stop initiating RR´s. Due to the long propagation delay, this global message might be in transit while RR´s are still coming in. For this purpose RR´s are only acknowledged if the buffer is not full. When the buffer is full a global message `BufferFull = TRUE´ is sent to all mobiles. As soon as buffer space becomes available mobiles will be notified of the condition so that they can resume transmission of RR´s.

The queuing discipline as described in the algorithm is a First-In-First-Out (FIFO). This will be the general case and would constitute the bulk of the traffic load so it will be analysed in detail. If the system is to cater for SOS messages, then priority is to be included, and although the load due to such messages is very small there are important delay constraints.

5.3. Steady State Analysis of Performance.

This analysis will be limited to the steady state conditions. Some of the stability analysis of S-ALOHA have been addressed in the previous chapter. In the next chapter an algorithm is developed to investigate the stability of P-ALOHA. The stability of the protocol to be developed are studied further in chapter 7.

5.3.1. Assumptions.

The following assumptions will be made regarding the system:

1. The number of users is very large so that it can be assumed to be infinite. In practice a population size in excess of 10,000 is

typical.

2. All users are identical and indistinguishable.
3. Each mobile will only hold one active message.
4. The channel traffic is independent over the RSF.
5. No retransmission occurs due to noise corrupted packets.
6. Processing delay is negligible.
7. The output process of ALOHA reservation process forms a Poisson input process to the queue.
8. A FIFO queuing discipline is assumed in the general case.
9. The long message length distribution is exponentially distributed.

5.3.2. Glossary.

ceil(x) : the function returning the integer just greater than or equal to x,

D_p : the equivalent propagation delay in slots,

F : number of slots in a frame,

G : channel traffic rate (per short message duration),

G_i, G_j : traffic rate due to user i,j (per short message duration),

G_R : traffic rate for RR's (per short message duration),

G_S : traffic rate for SCM's (per short message duration),

K : the spread in the ALOHA random delay,

L_R : message rate for RR's (RR/s/mobile),

L_S : message rate for SCM's (SCM/s/mobile),

M : mean long message duration (s),

N : mobile population size,

R : number of slots in a RSF, i.e. F-R in an ISF,

S : channel through-put (per short message duration),

S_i, S_j : through-put due to user i,j (per short message duration),

S_R : RR through-put (per short message duration),

S_S : SCM through-put (per short message duration),

t_F : frame duration,

t_o : time-out period,

t_p : propagation delay,

trunc(x) : the function returning the integer just less than or equal to x,

Z : channel capacity in slots/s i.e. F/t_F,

5.3.3. The Reservation Subframe.

Here we shall assume the previously achieved results for the through-put of ALOHA systems, i.e. for P-ALOHA

$$S = G e^{-2G} \tag{5.1}$$

for S-ALOHA

$$S = G e^{-G} \tag{5.2}$$

In a conventional ALOHA system, the retransmission probability, p would given by

$$p = \{D_p + (K + 1)/2\}^{-1} \tag{5.3}$$

However, in this context where the RSF is interleaved with the ISF, it has found that the system is insensitive to the distribution of the random delay and only sensitive to the mean. In a reservation system, R is effectively controlling the ALOHA channel capacity, which can be be given by RA/t_F. Thus in the system being considered p can be defined as

$$p = \{I\ R\ A + R\ A/2\}^{-1} \qquad 5.4$$

where $I = \mathrm{ceil}(t_p/t_F)$

Proceeding with the analysis, considering that the traffic is made up of two components, i.e. RR's and SCM's, then we can write

$$S = S_R + S_S \qquad 5.5$$

$$G = G_R + G_S \qquad 5.6$$

and

$$S_R = NL_R t_F/(RA) \qquad 5.7a$$

$$S_S = NL_S t_F/(RA) \qquad 5.7b$$

$$S_S = N(L_R + L_S)t_F/(RA) \qquad 5.7c$$

Equations 5.7 can be interpreted as the messages generated by the whole population (e.g. NL_R) per frame per number of subslots (R*A) in a RSF.

S/G is the probability that a transmitted message is successful so that the number of attempts required to transmit a message is G/S. This can, of course, be evaluated from equations 5.1 and 5.2 for the P-ALOHA and S-ALOHA respectively. However to maintain generality the quotient will be used.

The first delay incurred in the system is that a message can be generated anywhere between two starts of frames, so that the expected delay to the next start of frame is

$$E[w] = F/2 \qquad 5.8$$

We will assume that the propagation delay is greater than the frame duration. This will always be the case if the system is to cater for speech communications. If the collision detector is on-board than the delay for the first attempt is

$$D_1 = D_p + R \qquad 5.9a$$

This provides for the worst case where the last subslot in the RSF is used.

Defining

$$I = \text{ceil}((D_p + R)/F) \qquad 5.10a$$

then the delay per subsequent attempt is

$$D_A = IF \qquad 5.11a$$

In order not to incur any extra delay the value of the time-out period is assumed to be

$$t_p + R\, t_F / F \leq t_o \leq I\, t_F \qquad 5.12a$$

The time-out can, of course, exceed the upper limit, in which case

$$D_A = F \text{ ceil } (t_o / t_F) \qquad \text{for } t_o > I\, t_F$$

The corresponding equations for a system with a terrestrial collision detector are

$$J = \text{ceil}(D_p/F)$$

$$D_1 = D_p + R + JF \qquad 5.9b$$

$$I = \mathrm{ceil}((2D_p + R)/F) \qquad 5.10b$$

$$D_A = IF \qquad 5.11b$$

$$2\,t_p + R\,t_F\,/\,F \leq t_o \leq I\,t_F \qquad 5.12b$$

Equations 5.9a to 5.11a essentially differ from 5.9b to 5.11b by D_p. This is because the return link from the terrestrial controller is assumed to operate at a different frequency.

The total delay incurred in transmitting a short message thus is given in slot time units by

$$D_S = E[w] + D_1 + (G/S - 1)D_A \qquad 5.14$$

This is illustrated in figure 5.4.

5.3.4. The Information Subframe.

The ALOHA process for the RR´s form the input process for the ISF. If the ALOHA process is in equilibrium, then the arrival rate per second to the buffer can be be assumed to have a Poisson distribution with variance and mean given by

$$y = S_R\,A\,R\,/\,t_F \qquad 5.15$$

The length of the long message distribution is assumed to have a negative exponential distribution. Since the mean message length is M s, the mean service rate is $u = 1\,/\,M\ s^{-1}$.

Assuming the stated distributions for the input and the output, the system can be described by a M/M/c/K/M queuing system. The first two M´s indicate the Markovian property of the input and output distribution. There are c parallel servers and the system has a finite capacity of K. It should be noted that the capacity includes the number of servers as well as any storage space. The system also operates on a finite population of M.

The queuing model can be simplified by making a few assumptions. The population size though finite is large (10,000 to 1,000,000) and can be assumed to be infinite. Similar reasoning can simplify the system to an infinite capacity system. Obviously the buffer space has to be finite but it can be shown that the buffer space can be large enough so as to enable the infinite capacity model to be employed.

To support the application of an infinite capacity model, let us assume a finite capacity queuing model and find the probability that the buffer is full. By manipulation of the standard equations for a M/M/c/K system, the probability that the buffer is full when the utilisation factor is less than unity is given by

$$p_K = \frac{(y/cu)^B}{\sum_{i=0}^{c-1} \frac{c!}{i!} \left(\frac{y}{u}\right)^{i-c} + \frac{1 - (y/cu)^{B+1}}{1 - y/cu}} \qquad 5.16$$

The constraint on the utilisation factor is imposed in order to ensure stationarity.

Figure 5.5 shows a plot of equation 5.16 for a utilisation factor of 0.99 for values of c equal to 5, 50 and 500. When the buffer is full, causing RR's to be rejected and when the probability of a collision in the ALOHA subframe is very small, the probability that the buffer is full gives the grade of service. The plot shows that for large buffers this probability is insensitive to the number of channels. It can be concluded that for realistic and practical buffer sizes, the probability that the buffer is full can be sufficiently small. Under this condition we may use the infinite capacity model and assume a M/M/c queuing model.

Assuming a first-in-first-out discipline, we can apply Little's formula to obtain the time spent in the queue. This is given by

$$D_q = L_q / y \tag{5.17}$$

where

$$L = \frac{(y / u)^c}{c! \, [1 - y/(cu)]^2} \cdot \frac{y}{(c\ u)} \cdot P_o \tag{5.17a}$$

and

$$P_o = \left[\sum_{i=0}^{c-1} \frac{1}{i!} \left(\frac{y}{u}\right)^i + \frac{1}{c!} \left(\frac{y}{u}\right)^c \left(\frac{cu}{cu - y}\right) \right]^{-1} \tag{5.17b}$$

If the ISF controller is satellite-borne then we define

$$B = \text{ceil}(D_q / t_F) \tag{5.18}$$

$$C = \text{ceil}(t_p / t_F) \tag{5.19}$$

Assuming that when a slot is available the users are notified before the beginning of the next ISF, then the delay to the start of the ISF in which the first packet is to be transmitted is given by

$$D_2 = B\ F + C\ F \tag{5.20a}$$

The time diagram in figure 5.6 illustrates this.

The assigned slot could be any one of the slots in the ISF. They are all equally likely so that the choice of slot follows a uniform distribution of zero to F - R (but subject to availability). The expected position of the beginning of the assigned slot in slot time units is given by

$$E[f] = (F - R - 1)/2 \qquad 5.21$$

The corresponding equations for a system with a terrestrial controller are the following

$$E = \mathrm{ceil}[(D_P + D_Q)/F] \qquad 5.22$$

$$D_2 = (E + C)F \qquad 5.20b$$

The total delay incurred by a long message is thus given by the sum of the delay required to get a RR through, the delay involved in queuing and notification of the users involved and the expected delay in a frame from the beginning of that frame to the beginning of the assigned slot (see figure 5.6). This is given by

$$D_L = D_S + D_2 + E[f] \qquad 5.23$$

The through-put for the ISF is given by the quotient of the slots used and the slots available

$$S_L = \frac{N\, L_R\, M\ /\ t_F}{(F - R)\ /\ t_F} \qquad 5.24$$

The same result could have been derived from queuing theory which appears as the traffic intensity or utilisation factor. It should be obvious that the operating value for this factor has to be a compromise. It has to be less than unity if a steady state condition is to exist. Theoretically it could be arbitrarily close to unity which would be desirable to maximise resource utilisation. In this case, this would indicate a more efficient use of the allocated spectrum. However, as the through-put approaches unity the delay will start to increase rapidly.

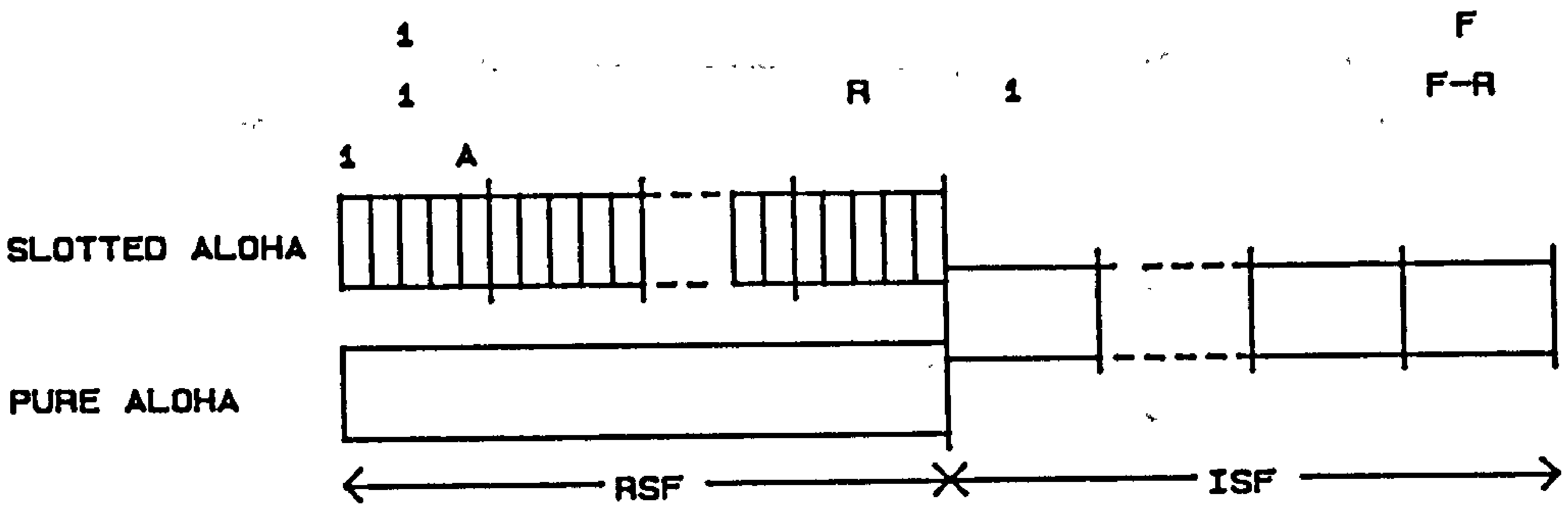


Figure 5.1. The frame structure.

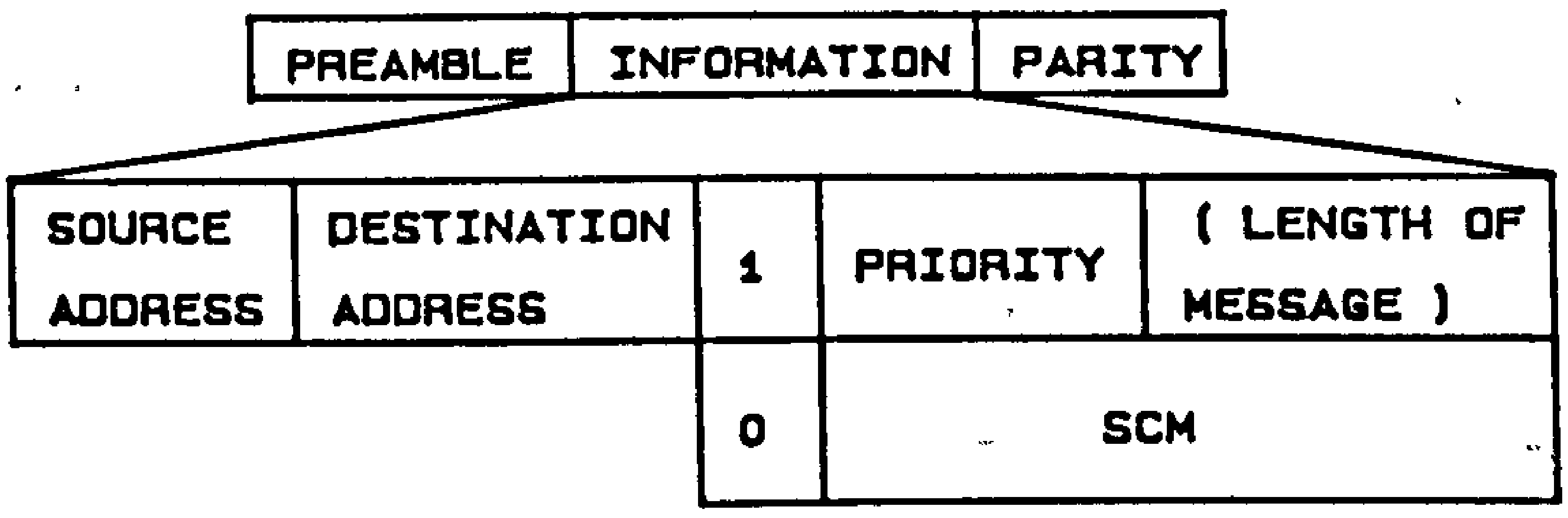


Figure 5.2. Short message formats: (A) RR and (B) SCM.

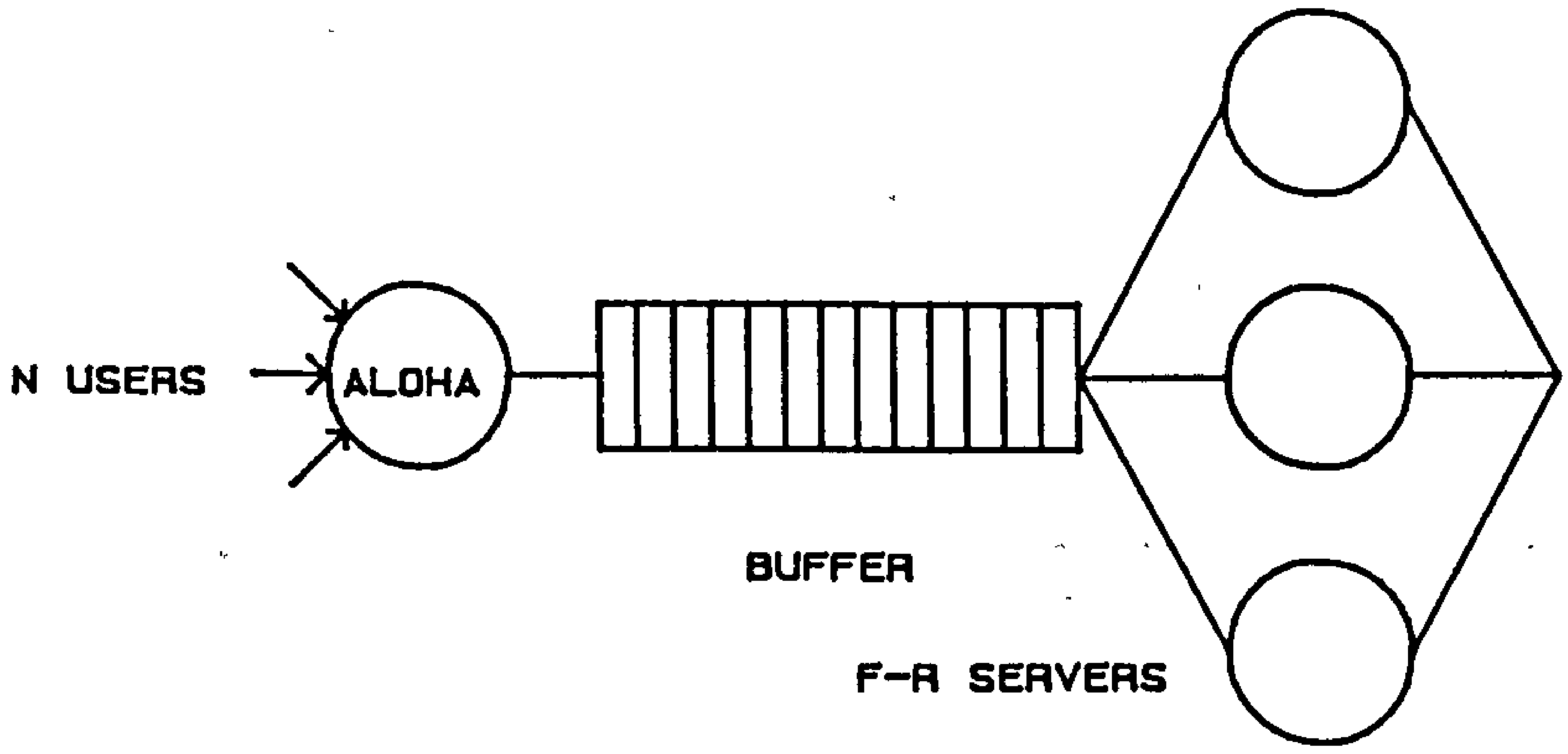


Figure 5.3. An illustration of the operation of the input process, the buffering action and the output process for the ISF.

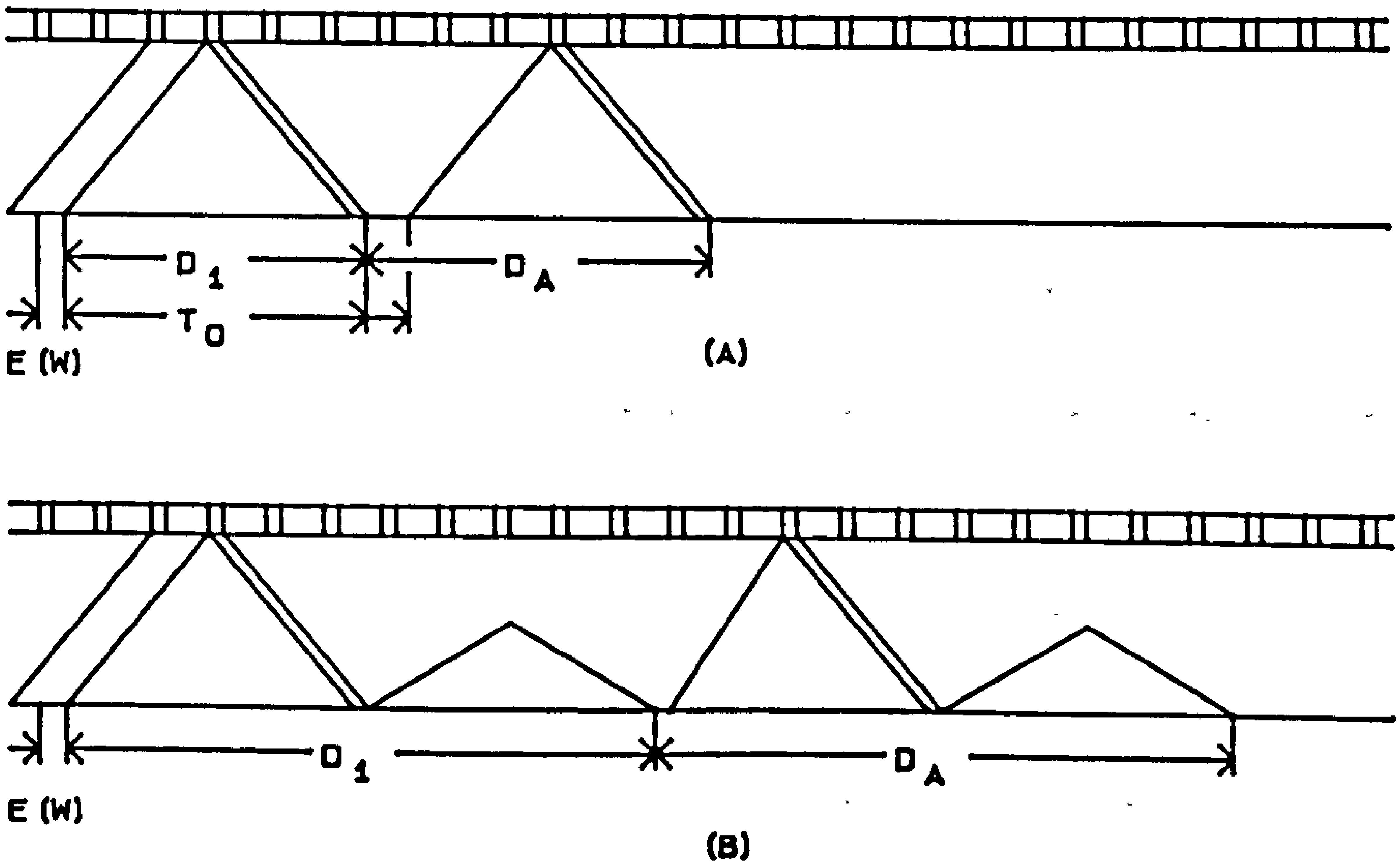


Figure 5.4. Illustration of the delays incurred in transmitting a short message (A) for an on-board collision detector and (B) for a terrestrial collision detector.

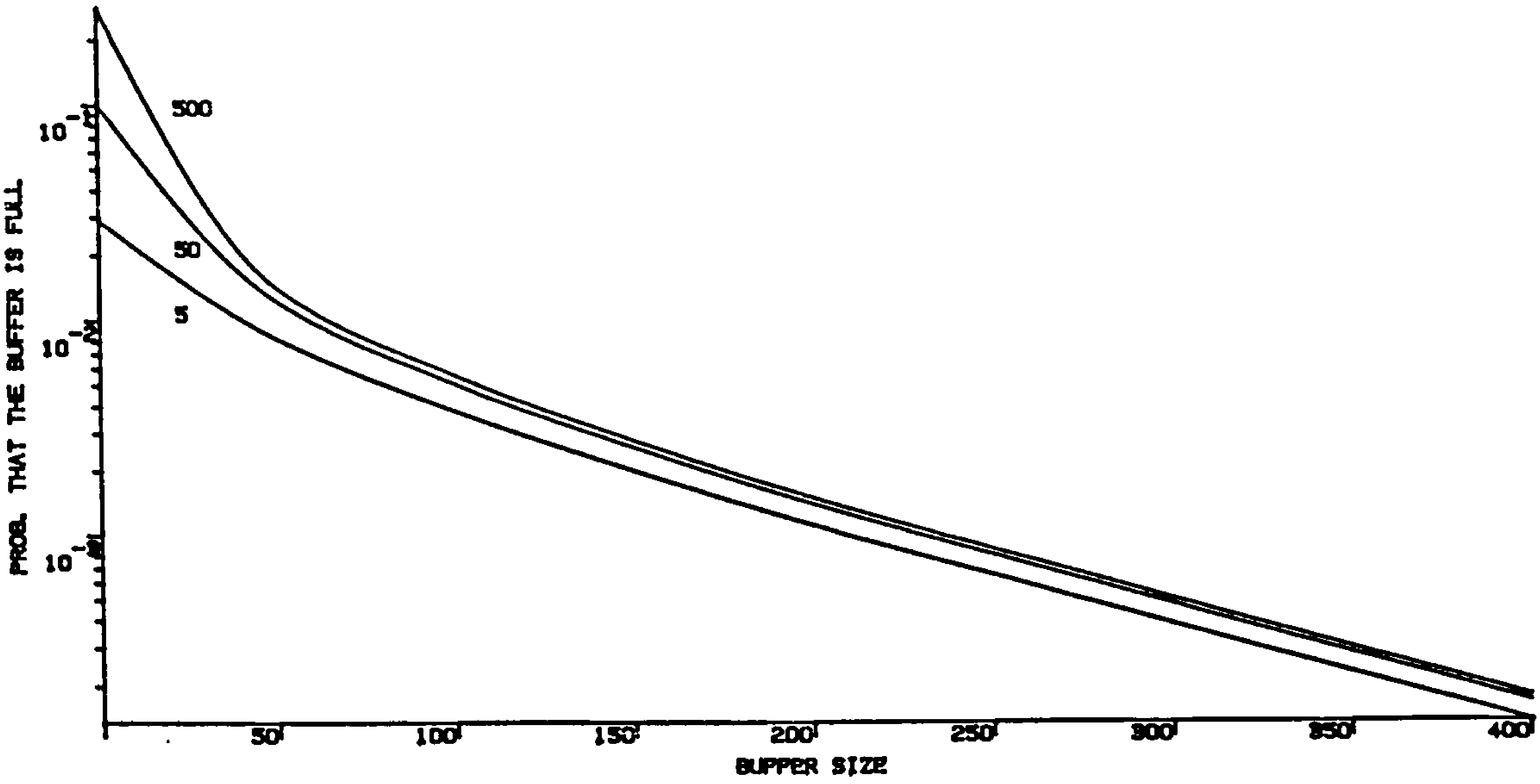


Figure 5.5. The probability that the buffer is full for a range of buffer sizes given the number of servers for an M/M/c/K queuing system.

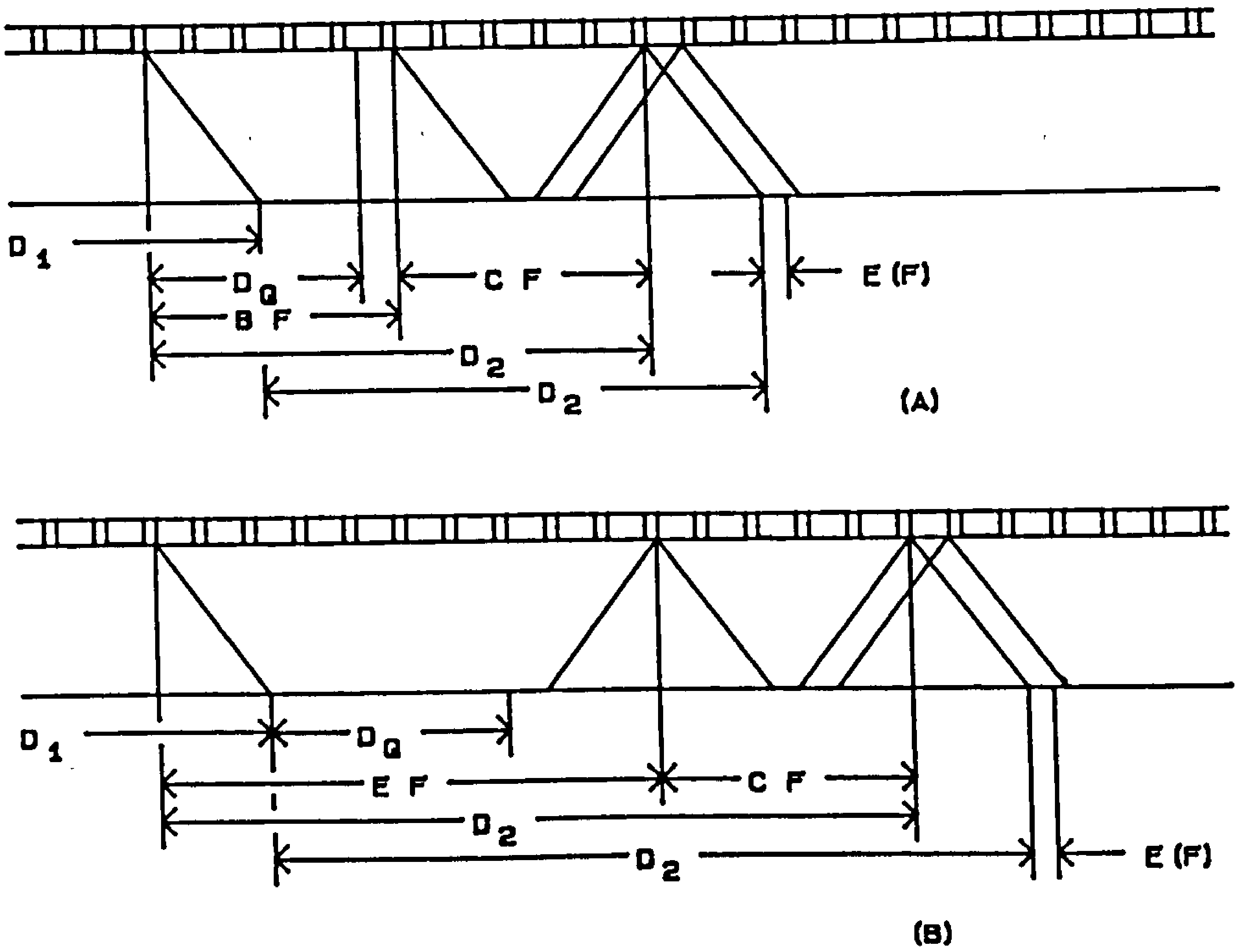


Figure 5.6. Illustration of the delays incurred in transmitting a long message after a RR has been accepted (A) for a spacecraft-borne controller and (B) for a terrestrial controller.

CHAPTER 6.

PROTOCOL SIMULATION.

6.1. General Considerations.

In many applications, actual real life experimentation can be costly and risky. Increasing computer speeds and the development of suitable programming environments have made simulation a more attractive tool to the engineer and it is being applied in more and more fields. Simulation allows the analyst to experiment with a model of a real system. The simulation model describes the operation of the system in terms of individual events of the individual elements in the system.

The development of the simulation model requires a thorough understanding of the problem, so that the engineer is forced to consider details that may not have received the required attention. These may lead to modifications in the model which are obviously cheaper to implement then than in the final system. It will also save time by possibly reducing teething problems.

Simulation should be considered (SHAN 75) when one or more of the following conditions are satisfied:

1. A complete mathematical formulation is not possible.
2. When the analytical methods are complex and simulation provides a simpler and a more cost-effective method.

3. When the mathematical analysis involved is beyond the ability of the available personnel.
4. Simulation runs may improve the confidence of a design before an expensive system is implemented.
5. Time compression is possible in simulation, so that results can be obtained about an experiment that would take a long time in the real world.

The aim of the simulation is crucial in the model development phase. If the overall system behaviour is sought, then too much detail may require too much computer time without extracting any further information and making it too expensive to run. On the other hand if a particular detail is required the model should be striped of all unrelated material and a more specific model could be used to investigate the aspect in question. Three factors must be considered when opting for simulation techniques: computer time, development time and flexibility.

In contrast to what has been said previously, since one cannot "see" how the simulation actually operates, it is easy to overlook hidden limitations such as the model assumptions and the compromises of the computer implementation Very often results are accepted without challenge because of the confidence put in computer results.

6.2. Development of Simulation Models.

There are three steps in the development of a simulation algorithm:

1. Formulate the model: In general, this will include mathematical modelling. In a continuous system this would be set of differential or difference equations representing the dynamic

behaviour of the system. In a discrete event system, a set of equations are required that determine the conditions of the next epoch given the past history of events.

2. Design the experiment: The objectives of the simulation are set. This will enable the engineer to develop an experiment at the right level of complexity that produces the results being sought. The right level of complexity is extremely important in order to maximise the efficiency of the resources, human and computer. There will obviously be some overlap between this step and the previous one.
3. Develop the software: The previous step basically provides the algorithm to the simulation program, so that a program in a suitable language can be written.

Simulations can be continuous or discrete. The former simulates a system in which the system parameters (or attributes) vary smoothly with time. An example of such a system would be chemical process or the corruption of signal waveforms by noise. In communication networks, we are mainly interested in discrete event simulation. This is one in which the system parameters change in a stepwise discrete fashion. All the simulations performed in this thesis are of this type.

Discrete event simulations can be clock-driven or event-driven. In a clock-driven simulation, the clock is advanced at regular instances irrespective of the activity within the system and the system parameters are updated with every time step (figure 6.1a). In general, the time step is chosen so that it is small enough to approximate the system. In a pre-emptive queuing model, for example, the time increment should be such that it is small compared to the service time. In a non-pre-emptive situation the time increment may not be so critical. There is a small subset of systems in which there is inherent synchronisation. For example, products leaving a production process at a constant rate and entering a warehouse. Another example in the context of this thesis is the S-ALOHA.

In an event-driven system, the clock is advanced by a variable amount determined by the occurrence of events (figure 6.1b). In a queuing model, instead of incrementing the clock in small increments and updating the system parameters when most of them are unchanged for a large number of time steps, the clock is advanced to the next occurrence of an event. In this case an arrival of a customer to the queue or a departure of a customer from a server. The sequence of occurrences of events is stored in an event-list.

Once a model has been developed data is required for it to operate on. This can be made available via a data base representing real life data or it can generated within the algorithm according to some probabilistic model. Real life data can be very useful in developing models to explain past events, for example, understanding a failure or a natural phenomenon. Monte Carlo techniques can be used for statistical data generation. This involves assigning probabilities to all the possible outcomes to establish an outcome interval. This is a linear array which extends from zero to unity. The difference between two consequtive components represent the probability that an event identified by one of the indices occurs. A random number within the same interval is then generated and depending on where it falls in the array an event is assumed to have occurred statistically. This technique will be used extensively in the simulations to be described in this chapter.

The model in itself has little use unless the simulation process is tapped at instances and certain parameters are logged for analysis and reporting (possibly in graphical form) at the end of the run. Three types of data can be identified: timing, resource utilisation and historical. Timing data may give information on time required to complete a task and can be constituted of several more basic time measurements. In the simulation of the ALOHA reservation protocol, the delay for a long message would fall under this classification as it is the result of the delay of the ALOHA process and the queuing process of the ISF. Resource utilisation data would, for example, contain the typical queuing parameters and can be very useful in the model validation phase. Historical data is used to give chronological step-by-step information of how a parameter changed its value during the course of the run.

The final task in the model construction is its validation. A modular approach to the model design can greatly ease this exercise. In this case testing each module is a much easier task than testing the system as a whole. Performance bounds can be used to increase the figure of confidence in the software. For example, whatever the load to a queuing system is, the utilisation can never exceed unity. Over the run duration, the total number of arrivals must be equal to the number of customers in the system plus the number of customers that left the system. Under certain conditions, the system may revert to another system whose performance is known and therefore provides a reference point. However, it should be mentioned that exhaustive validation is often difficult to arrange.

6.3. Review of Programming Languages and Software Tools.

Any algorithmic programming language can be used and about 75% of simulation software is written in general purpose languages like FORTRAN and PASCAL (MITT 84). However, these languages do not provide the best programming environments with the result that simulation algorithms will take longer to implement. Moreover since most simulation languages are based on one of these general purpose languages they may take longer to run. A considerable number of simulation languages exist.

Most simulation algorithms share some common features. These include the generation of random numbers from various distributions, system clocks, event-lists where appropriate, recording of data, statistical analysis of data, formatting of output data and detection and reports of logical inconsistencies. Simulation languages have been developed to provide such a programming environment.

The common classification is in terms of the class of problems they were intended for. Thus for continuous systems languages like CSMP (Continuous System Modelling Program) and DYNAMO exist. The former simulates models that can be formulated in terms of differential equations while the latter caters for models represented mathematically by difference equations. The other classification is the discrete simulation languages. This includes languages like GPSS (General-Purpose System Simulator), SIMULA and DEMOS. Some languages like SIMSCRIPT, GASP (General Activity Simulation Program), and SLAM allow for systems that combine continuous and discrete modelling.

There are distinct differences between languages developed in the U.S.A. and Europe; the U.S. languages tend to be based on FORTRAN while the European ones are based on ALGOL (TOCH 65). Here we will only have a brief look at GPSS, SIMSCRIPT and SIMULA.

GPSS

GPSS is one of the oldest and most well used simulation languages. The first version dates back to 1961. It is a process-orientated discrete event language. Its model consists mainly of dynamic or activity-type entities moving in time and acted upon by a set of system entities. The dynamic entities are called transactions and it is described by a set of parameters. These can be created and destroyed in the simulation process.

System entities that act on one transaction at a time are called a facility while those that cater for more than one transaction are called storages. Thus a single server is a facility while a multi-server unit is a storage. The actual statements that determine the logical processing are called blocks.

GPSS automatically maintains a statistical record on the system entities. At the end of a simulation run, GPSS produces a standard fixed-format output which includes the utilisation of facilities and storages, and queuing statistics.

SIMSCRIPT

Originally this language was designed for event-orientated discrete simulations but it was later extended to include continuous and combined models. It has its origins in FORTRAN. A SIMSCRIPT program describes a system in terms of entities (permanent and temporary), attributes of these entities and sets (of entities). For example, the existence of a telephone call may be represented by a temporary entity while the entire system would obviously constitute a permanent one. The temporary entity would then have attributes that include the time when the call started and its duration. Sets are a collection of entities so that the number of calls waiting for a line may be represented by a set.

SIMSCRIPT provides a record of the occurrence of events but it does not provide automatic statistical analysis and reporting. These are available through optional functions.

SIMULA

SIMULA is a process-orientated discrete event language that is based on ALGOL and consequently it is strongly typed. There is little interest in or availability of ALGOL in the U.S.A. This has restricted its use mainly to Europe. SIMULA, particularly the present version, SIMULA 67, is an elegant and powerful simulation language.

This language views the system as a collection of processes. The dynamic behaviour of the system is described by the actions and interactions of the processes. Each process has a set of attributes attached to it. The most important of these is the time attribute. There is also a user defined routine that describes the respective process. There are four possible states in which a process can exist: active, suspended, passive and terminated. A process becomes active when triggered to action by use of the time attribute.

Unfortunately, some of these languages do not tend to be very portable and compilers may only be available for specific machines. On the mainframe at the University of Bradford, SIMULA is available. However, being a rather specialised language, there is very little support available and it was considered more time-effective and convenient to use a popular language like PASCAL which is strongly typed like SIMULA, highly structured and supported very well.

Entities can easily be implemented as procedures of one of the set of data structures available and though in standard PASCAL there are no random number generators, most implementations include a uniformly distributed random number generator for the range zero to one. This can very easily be translated to the common distributions (SHAN 75), particularly the ones required in this field, i.e. the Poisson and exponential distributions.

The number of variables required for the report is small (a maximum of about seven) so that data gathering is a rather simple task. The possibility of employing graphics was considered to be a great asset in the report generation phase.

6.4. Steady State Simulation.

The steady state simulation of the protocol discussed in the previous chapter had two functions in the protocol design: simulation was employed in the development of the analytical analysis and finally it provided a valuable cross-check for the results. Since the original analysis was that for the S-ALOHA system and the P-ALOHA analysis evolved from it, the steady state simulation was only performed for the S-ALOHA case.

Due to the inherent rigid time grid structure of this protocol, the simulation lends itself to a clock-driven discrete-event situation. The most straightforward technique would be to treat each terminal as a separate entity. In PASCAL the data structure, record, would represent a very elegant and simple format where the fields represent the attributes of the user. A suitable type declaration is given below:

```
TYPE
   StatusType=(Busy,Collided,Idle);
   ModeType=(SCM,LM);
   Terminal:RECORD
      Status:StateType;
      Mode:ModeType;
      Delay,Length:INTEGER;
   END;
```

The record contains four fields:

1. State is of an enumerated type and can either be Busy, indicating that the terminal has a new message to transmit, or Collided , which indicates that the terminal has been involved in a collision and is in the retransmission state, or finally Idle indicating that the terminal is idle with no messages to transmit.
2. Mode , which provides information on the message type if there is one to be transmitted i.e. whether it is a short coded message, SCM, or a long message, LM, which is a long indefinite message like transmission of a data file or speech.
3. The field Delay is used in conjunction with State = Collided. It does not exist if State = Busy, or if State = Idle. It represents the delay to the next reattempt and is decremented when the clock is incremented. Thus if State = Collided and Delay is decremented

to zero there is another retransmission.

4. If and only if Mode = LM then information on the length of the message to be sent is required. This is stored in Length and is decremented every time the clock is advanced. Since long messages are packetised the type of Length can be an INTEGER.

A more formal and rigorous way for expressing the above entity makes use of variant records to ensure that the field **Delay** only exists when Status = Collided and Length when Mode = LM. This becomes rather clumsy due to the fact that records in PASCAL are only allowed one variant part. The type declaration can be expressed as:

```
TYPE
   StatusType = (Busy,Collided,Idle);
   ModeType = (LM,SCM);
   Terminal = RECORD
      Status : RECORD
         CASE StatusValue : StatusType OF
            Busy : ();
            Collided : (Delay : INTEGER);
            Idle : ();
         END;
      Mode : RECORD
         CASE ModeValue : ModeType OF
            SCM : ();
            LM : (Length : INTEGER);
         END;
   END;
```

Under this declaration only if **Terminal**.Status.StatusValue = Collided, does the field Terminal.Status.Delay exist. Similarly the field Terminal.Mode.Length exists only if Terminal.Mode.ModeValue = LM.

The whole population can then be represented by an array of such records. The program would proceed by commencing at the subframe and the clock incremented in subslot time units. During this period, Idle terminal may become busy indicating a transmission and for terminals in the Collided state, if Delay = 0, there is also a retransmission.

The next step would be to update the terminals' status. If there is only one transmission, appropriate action is taken depending on whether the message was a SCM or a long message. In the latter case the message length and the buffer to the ISF are updated. If there were collisions the status and the delays of the terminals are set accordingly. At the end of the RSF the report section is updated.

For the ISF, available slots are assigned to any buffered reservation requests. The ISF can be represented by an array of integers. A zero value for an array component indicates that the slot is available and can be loaded by the number of packets required for the first buffered long message. All non-zero components are decremented by one representing a transmission of a packet. At the end of the ISF the report section is again updated.

This approach is simple and clear because the simulation model is very close to what happens in the real world. However, since in the systems we are trying to simulate the number of users is so large, such an approach would be very costly in terms of computer time and memory space. Besides we are not really interested in particular information regarding the users. Our main interest is the channel utilisation.

To make the software more realisable, a probabilistic modelling approach is employed and the whole population is treated as a single entity. The algorithm used for the reservation subframe is similar to to that used by O'Reilly (OREI 84) to simulate a CSMA/CD (Carrier Sense Multiple Access with Collision Detection) channel.

The main difference between the two protocols is that in ALOHA, transmission of a packet is independent of the state of the channel. The ALOHA system used is slightly different from the standard version because it is interleaved with the ISF and also because it caters for SCM's and RR's, the latter forming the input process to the ISF. This provides an efficient means of simulating the protocol in a large population (10^3 to 10^5) environment. Simpler discrete-event modelling techniques proved to run extremely slowly as the population size increased.

The source code for the simulation program in PASCAL is given in appendix A3. Throughout the discussion, the identifiers are printed in boldface. A list of the functions and procedures that were developed in this software are given below:

FUNCTION Random, is an external function from the PASCAL library PASCLIB and generates a uniformly distributed random variable in the half-open interval (0,1] and has a period of 2^{98-1},

FUNCTION Ceil returns the ceil function of a real number,

FUNCTION NegExp returns a random variable with a negative exponential distribution with a mean of Theta.

PROCEDURE RSF, to be described in detail later,

PROCEDURE Adjust, to be described in detail later,

PROCEDURE ISF, to be described in detail later,

PROCEDURE Report, generates a file, Data, that contains the input and the output data of the simulation run,

PROCEDURE ReadData, reads the input data and assigns corespondind values to respective variables,

PROCEDURE Initialise, initialises variables.

The simulation software is logically split into two sections: the RSF and the ISF. The program calls the two procedures, RSF and ISF, alternately for a number of runs (**XTimes**) incrementing the system clock every time, and the process is then repeated for a number (NrOfSteps) of increments in the population size. The main body of the program is given below:

```
BEGIN
   ReadData;
   Report(SystemSpec);
   Initialise;
   FOR I:=1 TO NrOfSteps DO
   BEGIN
      NrOfStx:=ROUND(I/NrOfSteps*NoOfStx);
      Adjust;
      FOR N:=1 TO XTimes DO
         BEGIN
            RSF(R,AlohaRate/R,SCMRate/(SCMRate+LMRate),C,CT,K,KA,KB);
            ISF(QInput,F-R,MesLen);
         END;
      Report(RunData);
   END;
END.
```

6.4.1. Reservation Subframe.

Defining the three dynamic variables as:

K_A being the number of backlogged stations due to one collision,

K_B being the number of backlogged stations due to more than one collision, and

C the average number of collisions incurred by users which have been involved in more than one collision,

then the number of collisions due the backlogged users can be expressed as

$$C_T = K_A + K_B\ C \qquad 6.1$$

The total number of backlogged stations is therefore, given by

$$K = K_A + K_B \qquad 6.2$$

The probability of i new arrivals in a slot from the non-backlogged users is given using the Poisson distribution by

$$P[i] = \frac{(N - K)!}{(N - K - i)!\ i!} (1 - e^{-y/RA})^{i} (e^{-y/RA})^{n-K-i} \qquad 6.3$$

$$0 \leq i \leq N - K$$

y being the mean short message generation rate per frame per user.

The probability of i arrivals from the backlogged users is given in terms of the probabilities of messages being generated from the backlogged groups K_A and K_B. This is expressed as:

$$q_i = \sum_{j=0}^{i} q_j^{(A)} q_{i-j}^{(B)} \qquad 0 \leq i \leq K \qquad 6.4$$

where

$$q_i^{(A)} = \frac{K_A!}{(K_A - i)!\ i!} (1 - e^{-1/v})^{i} (e^{-1/v})^{K_A-i} \qquad 6.5$$

$$0 \leq i \leq K_A$$

and

$$q_i^{(B)} = \frac{K_B!}{(K_B - i)!\ i!} (1 - e^{-1/v})^{i} (e^{-1/v})^{K_B-i} \qquad 6.6$$

$$0 \leq i \leq K_B$$

where v is the probability of a retransmission from a backlogged station.

We now have all the system dynamic variables to calculate the three states of the channel, i.e. idle, busy or involved in a collision. A slot is idle if there are no transmissions form either the backlogged or the non-backlogged users. This is given by

$$p_I = p_0 \, q_0 \qquad 6.7$$

A slot is busy if, and only if, there is one attempt. Since an attempt could arise from either the backlogged or the non-backlogged groups, it can be expressed as the probability that there is one transmission from the backlogged group while none from the non-backlogged, or vice versa, an attempt from the non-backlogged and none from the backlogged. This can be expressed as

$$p_G = p_1 \, q_0 + p_0 \, q_1 \qquad 6.8$$

The final probability can be put simply as the probability that the previous two states do not arise, i.e.

$$p_C = 1 - p_I - p_G \qquad 6.9$$

We can now produce an outcome interval as indicated in figure 6.2 and use the Monte Carlo technique to obtain an outcome. This implies generating a uniformly distributed random number in the interval (0,1) and depending on where it falls on the outcome interval, the outcome is assumed. If the outcome is a success, a probabilistic decision is made to check whether the event was a SCM or a RR, and the report variables are updated.

Given that there was a successful transmission, then the probability that it originated from a backlogged station can be given using Bayes rule as

$$G = p_0 \, q_1 \,/\, p_G \qquad 6.10$$

Another Monte Carlo process is then involved to determine this outcome and if in the affirmative, the dynamic variables: K, K_A, K_B and C_T, are adjusted accordingly.

The next step is to determine the number of retransmissions involved in a collision, given that a collision has occurred. The probability that i or less retransmissions were involved in a collision is given by s_i, where:

$$s_0 = q_0 \, (1 - p_0 - p_1) \,/\, p_C \qquad 6.11a$$

$$s_1 = s_0 + q_1 \, (1 - p_0) \,/\, p_C \qquad 6.11b$$

$$s_i = s_{i-1} + q_i \,/\, p_C \qquad 1 < i \leq K \qquad 6.11c$$

An outcome interval can then be established and a Monte Carlo process is involved so that if the random number lies between s_{i-1} and s_i, the number of retransmissions involved in a collision is i.

Interchanging p_i and q_i in the set of equations 6.11 will yield the number of first attempts involved in a collision. It should be noted that in both cases the probabilities of large values of i will tend to be clustered together towards unity. Since the intervals and consequently the probabilities become increasingly small, these higher values can be ignored. The number of steps on the outcome interval can either be truncated by actually limiting the number element (IndexLimit) in the probability arrays or by not exceeding a certain probability barrier (**Trunk**). The arrays (S and U) of the two probability sets are generated in procedure GenProbs.

6.4.2. Information Subframe.

The successful number of RR´s from the RSF form the input to the queue. The ISF is represented by an array (**Frame**). All the non-zero elements of the array are decremented by one. If there are any zero elements and the buffer is not empty, then the element is loaded by an integer with a negative exponential with a mean equal to the number of slots required to transmit a long message.

As described, there would be a transient condition in the occupancy of the frame at the beginning of each step in the simulation. This would increase the errors in the simulation. In order to keep this to a minimum a procedure **Adjust** has been included. This has the function of initialising the ISF with the expected occupancy.

6.5. Stability of the P-ALOHA Reservation Subframe.

A P-ALOHA system is an asynchronous system and though such a system could be modelled as a clock-driven discrete-event system, an event-driven model was used. For a clock-driven model, the simulation clock is incremented in steps such that the probability of two arrivals within the step duration is sufficiently small. However, in this case this is rather inefficient particularly when considering that very long runs are envisaged. Computer effort is wasted processing large numbers of small steps when the probability of events is very small. In an event-driven model the clock is incremented to the next event without any undue processing. An event-driven PASCAL program for this simulation is given in appendix A4. A list of the functions and procedures that were developed in this software are given below:

FUNCTION Random, is an external function from the PASCAL library PASCLIB and generates a uniformly distributed random variable in the half-open interval (0,1] and has a period of 2^{98-1},

PROCEDURE SetRandom, is also an external procedure from the library PASCLIB and sets the seeds for the random number generator,

FUNCTION Poisson, returns a random variable with a Poisson distribution and a mean, Mean,

FUNCTION Bernoulli, returns the number of successes out of a number, Trials, of Bernoulli trials where the probability of success is P,

PROCEDURE ArrivTimes, to be described laterm,

Time in this algorithm is in a continuous domain using units in terms of the subslot duration so that messages are one unit long. If we assume a Poisson message generation process then a random variable, New, can be used to represent the number of messages generated in a frame duration. The number of retransmissions in the same duration also needs to be estimated. If the probability of a reattempt is given by P and there are Re backlogged terminals, then we can assume a Binomial distribution for number of retransmissions.

Function Bernoulli accepts the value parameters P and Trials as the probability of success and the number of trials respectively. An array ProbArr can then be constructed to represent the outcome interval so that the difference between two adjacent components give the probability that there has been a number of retransmissions equal to the index of the larger component.

Next we need to establish a data structure that would contain information as to when events happen and whether the message is a new message or a retransmitted one. This is elegantly implemented by having an array of records, the type being called EpochArray. The record type EpochType, has two fields: EventTime, which represents the arrival time and the message type **Message** which is an enumerated type with two possible options **Tryl** for the first attempt and **TryN** for retransmissions. The type declaration is given below:

```
TYPE
   MessType = (Tryl,TryN);
   EpochType = RECORD
                 EventTime:REAL;
                 Message:MessType;
               END;
   EpochArray = ARRAY[0..100] OF EpochType;
```

A procedure called **ArrivTimes** accepts the value parameters New and Re, and the variable parameter **EpochArray**, and updates the variable parameter by the new event-list for the next RSF. It sets the field Message of the first New components of Epoch to be Tryl and the rest as being TryN. Obviously this is not accurate since the relative order of new and backlogged transmissions is random. However the randomisation is performed later.

The field EventTime of each component is then assigned a value equal to a uniformly distributed random number in the range zero to the number of subslots in the RSF less one. Note that messages are not allowed to arrive in the RSF during the last subslot since this would cause them to extend beyond the RSF boundary and cause interference with the first ISF slot. The components of array Epoch are then arranged in increasing order according to the order of the field EventTime as well. As the order of random number generation is also random, the ordering of field **EventTime** will randomise the order of the field Message.

Once the arrival times and the the message type have been determined for an RSF the algorithm proceeds by updating the reporting variables and the main algorithm variable **Back**. The algorithm is a forward one in the sense that it draws conclusions based on the present conditions in relation to the next set of conditions. A Boolean variable, Corrupted, is used to indicate whether the packet starting at the present epoch will terminate before the previous one has started. In this case the present packet is destined to be corrupted. This is done in a section of the main body given below:

```
N:=1 TO Arrivals DO
   BEGIN
      Jump:=Epoch[N+1].EventTime-Epoch[N].EventTime;
      IF Jump > 1
      THEN
         BEGIN
            IF NOT Corrupted
            THEN
               BEGIN
                  Succ:=Succ+1;
                  IF Epoch[N].Message=TryN
                  THEN
                     Back:=Back-1;
               END
            ELSE
               Corrupted:=FALSE;
         END
      ELSE  (* Jump < 1 *)
         BEGIN
            Corrupted:=TRUE;
            IF Epoch[N].Message=Try1
            THEN
               Back:=Back+1;
         END;
   END;
```

Commencing from the beginning of the RSF the first packet starts off uncorrupted. Thus if the next epoch is more than one unit ahead (noting that the clock units are in subslot time units) then the message goes through and the variable, Succ, is incremented by one. This represents the number of successful transmissions. If a message originated from a backlogged terminal the backlog (using variable Back) is decremented by one. Another possibility exists where the next event is still more than one time unit away but the packet is corrupted. In this case, Corrupted becomes FALSE .

If the event is less than one time unit ahead, then the variable Corrupted becomes (or remains) TRUE and if the message resulted from a fresh attempt, then the backlog is increased by one.

This process is repeated for the number of arrivals (Arrivals). At the beginning of each loop the new values for New and Re are estimated and the outer loop is either repeated for a given number of times (Max) or until the critical value (Nc) of the backlog is reached. The value used was the value of the backlog at the central intercept of the load line on the through-put/backlog curve (equation 3.12).
Since it was envisaged that the run-time of this program is very long, the initial conditions are entered in a intermediate file called NOTE. This allows the program to be run in concatenate runs using the final conditions of the last run as the new conditions of the new run. The data required are the final backlog, the maximum backlog attained, the time in terms of the frame duration, the number of successful transmissions and two seeds for the random generators. (The respective identifiers are Back, BackMax, Time, Succ, Seed1 and Seed2. Initially all variables are set to zero and at the end of each run

NOTE is updated. The seeds are also changed to ensure that the random number of subsequent runs are independent.

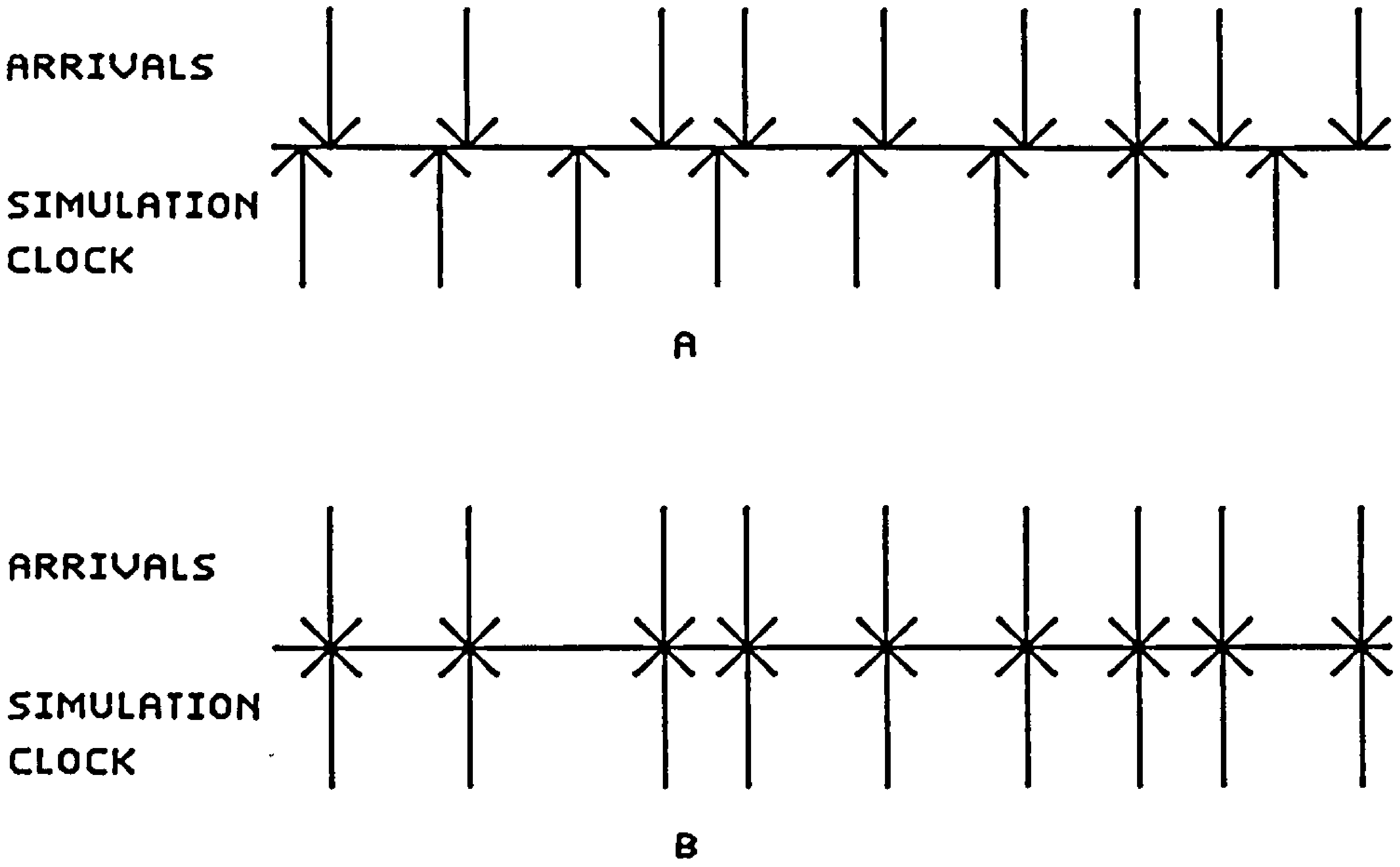


Figure 6.1. Illustration of (A) a clock-driven and (B) an event-driven discrete event simulation.

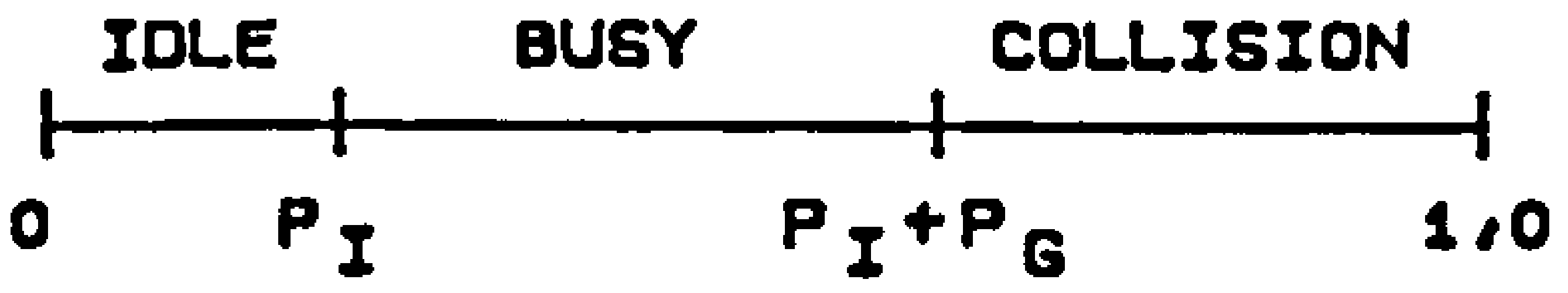


Figure 6.2. Outcome interval to determine whether a slot is involved in a collision, is idle or is involved in a successful transmission.

CHAPTER 7.

ASSESSMENT OF PROTOCOLS FOR THE LAND MOBILE ENVIRONMENT.

7.1. Framework.

The protocol will now be assessed in the U.K. context. Since this country is approximately covered by the latitudes 50^{o} N to 60^{o} N the best orbit appears to be the (12 h period) Molniya orbit. This permits quasi-zenithal operation of the terminals which is free from shadowing and multipath. If a geostationary orbit is employed the angle of elevation for the mobile antenna would vary from about 33^{o} in the south to the south to about 22^{o} in the north. This makes such a system unattrative for mobile use.

If the active part of the Molniya orbit is 8 h long, the system would require three satelltes. For reliability reasons, an operational system would have at least one spare in flight. The cost of the extra satellites is offset by the fact that the launch cost is about half that of the geostationary orbit. Since more than one satellite is involved, handover of traffic has to occur. As one satellite enters service and the other exits there has to be a smooth transition. Such an occurence only happens once every 8 h and can be controlled from either earth or via an inter-satellite link.

In such a system a satellite can be viewed from anywhere in the country at elevations larger than about 74° (see figure 7.1). This can be calculated from the theoretical 8 h ephemeris centred around the perigee and is given in appendix A5. The longitudes 8° W and 2° E bound the British Isles. The apogee is thus centred on 3° W. A single circular footprint is assumed. This implies that the beamwidth requirement of the spacecraft antenna is determined at the edges of the 8 h period and is equal to 2.7°. However, this assumes perfect pointing accuracy.

The Molniya orbit also has advantages for a European system. If we consider Europe to be bounded by the 36° N and 70° N lines of latitude, the angles of elevation at the respective locations are approximately 48° and 10°. In a similar system employing the Molniya orbit, the angle of elevation will always be in excess of 57°.

The performance of the system can be enhanced by employing a multispot antennae at the spacecraft. This would improve the spectral efficiency because frequency re-use is possible and the bandwidth of each spot can be decreased since each spot is now serving a portion of the population. Smaller beamwidths also imply higher gains. This can lead to smaller antennae at the terminal side. In a U.S. geostationary system, MSAT, six beams have been suggested though as many as 100 have been suggested for future generations (WEBE 83).

Unfortunately this also has the implication of larger space-borne antennae. Normally such antennae employ a single reflector with a complex feed strutures. This arrangement inevitably leads to poorer efficiencies. For the system coverage considered here, it seems unlikely that a multispot system is used because attempts to obtain very small beamwidths are thwarted by the spacecraft attitude-control precision limitations and by antenna reflector imperfections.

It should be pointed out that in a Molniya orbit, area of coverage is dependent on the altitude if the beamwidth is maintained constant. This is normally the case if a simple antenna structure is to be used. Complicated phased array antennae are required in order to attain a constant foot-print pattern. It is unlikely that this is employed in a simple system. A multispot system is not attractive with a Molniya orbit because the individual spots would not only need to have adjustable beamwidths but would also have to be steered to maintain constant foot-print centres with altitude.

Besides the spacecraft limitation problem there are also routing complications. Since the users are mobile, the system is not aware in which foot-print the user is. This leads to more complicated call set-up procedures and overhead traffic.

7.2. Performance.

7.2.1. Traffic profile.

The following assumptions will be made regarding the traffic profile:

1. The size of the mobile user population is 100,000.
2. The mean generation rate for long messages is three messages a day per user. Since 10 % of them occur during the busy hour then the figure used will be 0.3 messages per hour per user (IIDA 80).
3. The mean generation rate for short messages is one message per hour per user during the busy hour with an average of ten such calls a day per user.

4. The mean holding time for a long meesage is 2 min.
5. The mean waiting time is 5 s for a short message and 10 s for a long message.

7.2.2. Frame Duration.

The first decision concerns the length of the frame. Long frames are desirable since they improve the frame efficiency. If the preamble and the guard time remain constant then the slots can be made longer. As the frame duration increases the frame-to-frame coherence of the carrier deteriorates. This introduces additional phase noise at the demodulator imposing heavier demands on a coherent demodulator. The required buffering space also becomes larger, though with the falling price of semiconductor memories this is not too much a problem. Recommendation G.114 (CCIR 14) provides guidelines regarding one-way propagation delays for telephony. It states that subjective tests indicate that one-way delays above 400 ms are unacceptable, and care has to be exercised when dealing with delays in the range 150 to 400 ms, particularly above 300 ms, with regard to echo effects.

Taking all these factors into account a frame duration of 100 ms will be assumed for this system.

7.2.3. Reservation Subframe.

Assuming that the steady state operating point lies between the origin and the peak of the G-S curve, G/S will be less or equal to e^{1} for both P-ALOHA and S-ALOHA. Thus, rewriting equation 5.14 in time units for a system with an on-board contention detector, we get

$$D_S \leq t_p + [0.5 + R/F]t_F + (e - 1)\ t_F\ \mathrm{ceil}(t_p/t_F + R/F) \qquad 7.1$$

If the contention detector is on earth then equation 5.14 becomes

$$D_S \leq t_p + [0.5 + R/F + \mathrm{ceil}(t_p/t_F)]t_F$$

$$+ (e - 1)\ \mathrm{ceil}(2\ t_p/t_F + R/F)\ t_F \qquad 7.2$$

In both situations the following relationship must hold:

$$0 < R/F \leq 1 \qquad 7.3$$

Thus, the absloute upper limit is unity. This situation corresponds to a traffic load made up solely of SCM's. The corresponding values for the maximum delay are 0.92 and 1.73 s for systems with an on-board contention detector and a terrestrial one respectively.

Using the data in section 7.2.1 on the traffic profile assumptions, the minimum number of channels required for the ISF is given when the utilisation factor is unity. (This value would also produce infinite queuing delays so that in practice more channels are required. Here it suffices as a lower limit for F-R.) This gives the absolute minimum number of channels as

$$\frac{L_R * N}{u} = \frac{0.3 * 10^5 * 2 * 60}{60 * 60} = 1{,}000 \text{ channels} \qquad 7.4$$

From the steady state conditions, the value of R for such a traffic profile and a reasonable quiescent operating point, would be four, giving a value of R/F equal to 0.004. This would produce figures for the delays of 0.82 and 1.63 s for the two possibilities.

It might appear odd that in this formula the delay increases with R/F since the quotient indicates the portion of the system capacity dedicated to the RSF. Assuming that G is constant, then as R increases, increasing the capacity of the ALOHA channel, the traffic increases proportionally. However, in this variant of ALOHA, R also represents the period of time over which a short message can arrive into the system and the randomised delay of the ALOHA process, so that R affects the initial delay for the transmission in a RSF for a first attempt and the delay of subsequent retransmissions. Thus maintaining a constant G/S while increasing R/F implies an increase in the traffic load, an increase in the initial delay and an increase in the retransmission delay.

In an operating system, increasing R/F would result in a drop in G/S, bearing in mind that G and S are defined as the load and through-put respectively of the ALOHA channel per subslot. Thus for a given population size and a generation rate, G and S would decrease as R/F increases, provided the peak is not exceeded. However the above calculation provides an aid to estimate the upper bounds of the delay.

Both figures for the delays are well within the system specifications so on this account there seems to be no justification for having an on-board contention detector. A terrestrial one is thus assumed. This will improve the system reliability and ease billing.

Assuming the specified profile and a R value of four, the FET (First Exit Time) value is $8.87*10^{21}$ slots or about $6.4*10^{14}$ days! This implies that the probability requiring a control mechanism is very small, particularly when taking into account the diurnal variation of the load on such a system.

Figure 7.2 shows a plot of the FET for a range of subslots in the RSF. It should be noted that the units used for the FET is actual time (not slots) This implies that as the number of subslots increase, the effective capacity of the ALOHA channel also increases. This has a time scaling effect on the FET. However since the FET increases at a higher rate than the scaling effect, large gradients result anyway.

As mentioned earlier, a simulation approach was employed to obtain such a value for the P-ALOHA. The algorithm is discussed in chapter 4. As might be perceived, with such large numbers, very long CPU time is required and a large sample size is not possible. A series of concatenated runs was performed with the total equivalent length of the simulation being about 10 days. The quiescent point was chosen at R = 6. Throughout this run the backlog never exceeded 23 when the central intercept occurs at about 62. The corresponding figures for R = 10 are 11 and 177 respectively.

Figure 7.3 shows the maximum backlog over a simulation run equivalent to 24 hours for R = 5, 6 and 8. It can be seen that the larger the size of of the RSF the faster the maximum backlog is reached. At R = 5 the system approaches the maximum after 24 hours while at R = 6 it is reached after 10 hours and at R = 8 it is reached after 8 hours. This is true in general though the process is accelarated by the time scaling effect associated with the increase in capacity.

Intuitively, the FET depends on the distance between the first and the central intercept point and the height of the "hill" between them.

Figure 7.4 indicates the backlog at the first stable point and the central equilibrium point for the P-ALOHA and S-ALOHA under the assumed profile conditions. The way this protocol is structured, R affects both the distance between the points and the size of the "hill".

For a given load, figure 7.4 shows how the distance between the two points increases for a rising value of R. The fact that increasing R causes S to drop implies that the "hill" also becomes larger. Since the maximum size of the "hill" in S-ALOHA is inherently higher than that of the P-ALOHA, and since as indicated in figure 7.4 the width between the two points is larger for S-ALOHA than that of the P-ALOHA, the stability performance of the S-ALOHA will always be superior than the P-ALOHA for a given channel capacity. However the FET increases very steeply for increasing values of R and this is of course due to the double effect of R, which makes it a powerful parameter for controlling the stability.

7.2.4. Information Subframe.

The expression for the delay incurred with long messages was given in equation 5.23 as

$$D_L = D_S + D_2 + E[f] \qquad 7.5$$

The values for the short message (RR) delay were obtained in the previous section. The conservative value for a terrestrial contention detector will be assumed here. This is given as 1.73 s.

E[f] was given in equation 5.21 as $(F - R - 1)/2$. Since R is normally much smaller than F, then we assume E[f] to be equal to F/2 slot time units or $t_F/2$ s. Equation 7.5 thus becomes

$$D_L = 1.78 + D_2$$

or putting the specification for the long message delay,

$$D_2 < 8.22 \text{ s} \qquad 7.6$$

Putting this condition in equations 5.20a and 5.20b, we get the maximum tolerated queuing delays being 7.92 s and 7.67 s respectively for systems with and without a space-borne controller. The difference does not seem enough to justify a space-borne ISF controller and the penalty in terms of spectrum is only marginal.

The waiting time for a M/M/c queue is given in equation 5.17. Rearranging this equation, making appropriate approximations using Stirling's formula for the value of the factorial of large numbers and approximating the series to an exponential, we can obtain figure 7.5. A more useful form of figure 7.5 is given in figure 7.6. For the required waiting time, the number of channels reqiured is about 1,010. The grade of service and the ultilisation of the slots in the ISF are about 1 % and 99 % respectively.

From the given assumptions there is a total of 1,050 slots in a frame so that with 1,040 slots in an ISF, the slot utilisation is 96 % and the grade of service is improved approximately by a factor of two. Under these conditions the queuing delay drops to about 0.4 s so that the total delay for a long message is now down to 2.2 s. The queuing length is given by Little's formula and amounts to 3.3. The buffering facility is also required to accept the RR output form the ALOHA process before processing them for the ISF and this will provide some further improvements in the grade of service.

7.2.5. Steady State Simulation Results.

Figure 7.7 shows the results of the simulation described in the previous chapter with 1,010 slots in the ISF. The output of the program produces four plots and the end of each plot corresponds to the condition at the assummed profile.

The top left figure illustrates the delay for short and long messages. This demonstrates that the delay performance is within the set constraints. The maximum short message delay is about 0.8 s and the long message delay is about 1.4 s. For an ISF with 1,040 slots, the delay performance of long messages is considerably smoother. This is as expected since by increasing the number of ISF slots the utilisation drops. A direct consequence of this is the movement towards the flat portion of the delay-utilisation curve. The SCM performance is close enough to the estimated theoretical value while the long message delay is considerably lower.

The reason for this can be found from figure 7.5. The operation point on this curve lies in a region where the gradient is tending toward infinity, so large errors are to be expected. The simulation result would consequently become very dependent on the number of runs.

The next figure on the right shows the G-S performance for the two message types and for total effect on the channel. Again the correlation between these results and theory is satisfactory.

The bottom left figure indicates the ISF through-put or the utilisation factor versus the ALOHA channel load. The relationship is fairly linear but shows signs of saturation at high loads when the utilisation factor approaches unity. It should be noted that saturation commences when there is a sharp change in gradient in the delay performance.

The final plot demonstrates the population generation rate against the through-put. Under stability these should be equal and therefore a linear curve with unity gradient is expected. The simulation result is close to this. This figure was included as an indication of the performance of the simulation.

7.2.6. Dynamic Control of RSF/ISF Boundary.

In order to obtain total flexibility, the RSF/ISF boundary is dynamically controlled. This means that the system will be able to cater for profiles ranging from traffic profiles that are solely composed of SCM - this would reduce the ISF to nothing - to profiles that consist only of very long messages with minimal RR overhead.

Through the spread of profiles, certain conditions are to be observed:

1. The utilisation factor of the ISF is to be less than unity so as to maintain stability of the M/M/c queue.
2. The ALOHA process is to be kept close to the stable operating point with low backlog and delay. The RSF is therefore operated below the maximum through-put at values of S that allow sufficiently high FET's. This ensures a control over the delay and stability.
3. A minimum size of the RSF is to be maintained. This ensures that the system never disables the input process which would terminate the capability of handling emergency and high priority messages.

A measure of the RSF state can be obtained by counting the number of idle, busy and collision subslots over a number of frames. The theoretical probabilities of these events at the peak of the G-S curve are 0.368, 0.368, 0.264, and 0.368, 0.184 and 0.448 respectively for S-ALOHA and P-ALOHA. These figures can be used to ensure that the "steady state" of the system is on the low backlog side of the peak. Beyond the maximum the collision probability increases while the idle probability decreases.

Once the general area of the operating point has been located, R can be adjusted so that the required FET value can be assumed. For S-ALOHA, the FET at $R = 4$ is $8.87*10^{21}$ slots. Assuming S remains constant, the FET will improve as R increase. Since the FET at $R = 4$ would also be adequate at the maximum value of R, i.e. $R = F$, the algorithm can be simplified by just aiming to maintain an S value equal to or less than that at $R = 4$. At $R = 1,014$ the equivalent value of the FET is $1.6*10^{19}$ days, which will be perfectly adequate!

Using the counted numbers normalised over the number of subslots and comparing them to the theoretical values, and comparing the current value of S to the to the target value of S, a decision can be made as to which direction the boundary is to be pushed. On the left of the maximum of the G-S curve, the approximation that the probability of a collision is small is made. Beyond the peak the load on the system is estimated from the probability of an idle slot using the relationship

$$P[idle] = \exp(-G) \qquad 7.7a$$

so that

$$G = -\ln(P[idle]) \quad 7.7b$$

A suggested PASCAL algorithm for a S-ALOHA system is given below:

```
PROGRAM BoundaryControl;
CONST
   NoOfFrames=20;
   A=4;
   RMin=1;
   PrIdle=0.368;
   PrBusy=0.368;
   TargetS=0.226; (* The value of S at R = 4 at the assumed load. *)
   TargetG=0.307; (* Evaluated from the G-S relationship for S-ALOHA
                     at S = 0.226. *)
VAR
   SlotAvail,StopRR : BOOLEAN;
   NoOfBusy,NoOfColl,NoOfIdle,NoOfSlots,R : INTEGER;
   Idle,Busy,Coll,G,PrColl : REAL;
BEGIN
   PrColl := 1 - PrIdle - PrBusy;
   NoOfSubSlots := NoOfFrames*R*A;
   Idle := NoOfIdle/(NoOfSubSlots);
   Busy := NoOfBusy/(NoOfSubSlots);
   Coll := NoOfColl/(NoOfSubSlots);
   IF (Coll < PrColl) THEN
      BEGIN
         R := CEIL(R * Busy / TargetS);
         IF R < RMin THEN
            R := RMin;
      END
   ELSE
      BEGIN
         G := -LOG(Idle);
         IF SlotAvail
            R := CEIL(R * G / TargetG)
         ELSE
            StopRR = TRUE;
      END;
END.
```

It should be noted that the values for the **NoOfFrames**, **RMin**, **TargetS** and consequently **TargetG** are only suggested values. **RMin** can be increased and **NoOfFrames** can be varied. The Boolean variable StopRR ensures that the ISF is operated at utilisations in the sub-unity range. It is a global message that shuts down the input process to the ISF.

7.2.7. Acquisition and Synchronisation.

In TDMA systems the object is to transmit packets that reach the satellite within a fixed time grid structure. It would be desirable if this could be done accurately, since this would mean that the time slots could be filled more completely with information. However, this is not possible and some guard-time has to be included as a tolerance to allow for inaccuracies. These arise from the inaccuracy in the range between the user and the satellite. It is therefore convenient to identify two sources of such inaccuracies, namely the uncertainty in the position of the satellite and the geographical location of the user.

The inaccuracy of position of the satellite can be greatly diminished if the ranging data is broadcast at periodic intervals. The mobile location remains a problem. Besides the fact that the mobile itself does not know its location, it is also moving. In the U.K. situation, the worst variation in the range, assuming that the satellite position is known, is when the satellite is at the edge of its service. The variation in range is 253 km or a time variation of 0.843 ms for users at the north of the country and at the sub-satellite point.

Assuming telemetry information is made available often enough then the position of the satellite can be assumed as known. Errors due to variation in the propagation delay are in the nanosecond region while effects due to doppler can be compensated for.

Two methods of acquisition are considered as feasible:

1. The first method makes use of classical acquisition techniques.

Low power (about 25 dB down from the normal traffic power levels) bursts of unmodulated carrier or PN modulated signals can be used. The power density of both these signals are sufficiently low so as not to interfere with the rest of the traffic through the channel. The PN version may be preferred since several sequences may be used to reduce the probability of locking to other uers' signals.

In order to avoid the power levels building up, it is suggested that this procedure only occurrs when the mobile is switched on. By estimating the difference between the received reference burst and the anticipated position, a log can be maintained of the correction required.

2. The second method adopts a more integrated approach and makes use of the RSF in the P-ALOHA mode. The use of P-ALOHA would decrease the RSF through-put and possibly increase the delay though this can be corrected by increasing the size of the RSF. It should not be a problem to extend the RSF to achieve a similar delay and stability performance as that of the S-ALOHA. Under the assumed mix of message types, the RSF is only a small portion of the frame, so that while the RSF efficiency drops, the total efficiency is only marginally affected.

   If the reference burst is timed such that if a user at the shortest range would transmit a short message, it would arrive at the satellite at the first subslot in the RSF. The shortest range corresponds to the sub-satellite point. As a mobile moves away from this point, the short message would arrive later on in the RSF. The RSF needs to be at least 0.843 ms plus a short message duration for a system with an on-board collision detector or approximately twice that amount otherwise. When the message gets through via the ALOHA process, the satellite would acknowledge and provide information about the position of the message in the RSF. This data is then used to make the required corrections in future transmissions. Assuming a 100 ms frame duration, the lengths of the RSF's is equal to 10 and 20 slots for the respective systems.

This scheme makes use of the geographical spread of the users in order to even out the load on the subslots. It would work well in a ideal situation where there is a uniform population density. In a practical case where there are urban concentrations together with regional population density variations, messages would cluster together at points in the RSF.

The system can be modified by extending the RSF beyond the stated minimum, so that each user does not always aim for the first subslot, but can randomly choose from a number of sublots and possibly use some sort of a distribution that corresponds to the inverse north-south population density distribution, that occurs in the U.K.

In a European system, the largest variation occurs at the apogee and is equal to 822 km or 2.74 ms. This would make the system less efficient since the required RSF sizes would increase to about 29 and 57 for the respective systems.

Synchronisation for long messages can be achieved by the satellite acknowledging each packet and notifying the user of the error, i.e how far in the guard time the packet started.

7.2.8. Link Budgets.

In order to decouple the up-link from the down-link a regenerative transponder is assumed. In a transparent transponder, there is an adding effect of the up-link noise on to the down-link. The use of regenerative transponder allows both links to be optimised independently and efficiently without the need of increasing power from the HPA or improving the figure of merit of the receiving system.

We will consider the mobile up-link first since in the situation the main constraint in the link budget is the EIRP from the mobile. This will influence the space-craft antenna size. Once this has been established, then for the down-link budget the satellite HPA power can be fixed for the given mobile G/T. Table 7.1 provides the proposed link budgets for the basic TDMA system assuming the following system parameters:

1. A bandwidth allocation of about 6 MHz at L-band, 3 MHz for the up-link and 3 MHz for the down-link. The frequency allocation is assumed to be similar to the CERS project, i.e a mobile link with an up-link frequency of 1,650 to 1,653 MHz and a down-link frequency of 1,550 to 1,553 MHz.
2. A maximum HPA output of 20 W at the mobile.
3. A LNA equivalent noise temperature of 120 K (noise figure = 1.5 dB) for both satellite and mobiles.
4. A noise temperature of 300 K for the satellite-borne antenna and a negligible noise temperature for the mobile antenna.
5. Antennae aperture efficiency of 80 % for the spacecraft antenna and 60 % for the mobile antenna (to allow for beam shaping).
6. Maximum antennae diameters of 1.0 m and 3.5 m for the mobile and

satellite reflectors.

7. A bit error rate of 10^{-4} with no coding.
8. Modulation scheme used is QPSK with spectral efficiency of 1.68 b/s/Hz.
9. Information rate of 4.8 kb/s.
10. A link margin of 1 dB is assumed (FENE 86). This includes losses due due to noise temperature degradation, polarisation coupling losses and effects of a wet radome.

Though a 20 watt HPA for the mobile is feasible, it would be desirable if lower power is used. This can be accomplished on the up-link by using a hybrid scheme. The full system bandwidth can be split into, for example four, so that the transmission rate drops to a quarter. A low transmission rate is also desirable since it eases problems associated with acquisition and synchronisation. Such a system would effectively become a hybrid TDMA/FDMA system and the frame can be looked upon as changing from the one dimensional domain to the two dimensional domain with time and frequency being the dimensions. All the protocol analyses remain unchanged since the capacity is unchanged. A link budget for such a system is given in table 7.2.

On the down-link, several possibilities exist. Such a hybrid system could also be implemented. Separate HPA's per TDMA system could be used aboard the satellite. This would increase the reliability of the satellite HPA system. The total power could either be the same as that of the TDMA system or individually equal to the HPA of the TDMA system. The former makes the HPA system of satellite cheaper while leaving the mobile LNA unaltered. The latter increases the power consumption aboard the satellite, thus increasing the cost, but the

Table 7.1a. Proposed Mobile Up-Link Budget at L-Band for a TDMA system.

Frequency : 1,650-1,653 MHz

Mobile Terminal			
HPA power output	13.0		dBW
Coupling and feeder losses	-1.0		dB
Antenna gain (1.0 m diameter)	22.5		dBi
EIRP		34.5	dBW
Free space attenuation		188.7	dB
Satellite			
Antenna gain (3.5 m diameter) at 3 dB edge	31.7		dBi
Total coupling losses	1.0		dB
Receiving system equivalent noise temperature (526 K) (CCIR 73)	27.2		dB
G/T		3.5	dB/K
Boltzmann's constant		-228.6	dBJ/K
C/N_o		77.9	dBHz
Link margin		1.0	dB
Transmission rate (5.0 Mb/s)		67.0	dBHz
Modem implementation loss		1.5	dB
E_b/N_o		8.4	dB

Table 7.1b. Proposed Mobile Down-Link Budget at L-Band for a TDMA system.

Frequency : 1550 - 1553 MHz

Satellite:			
HPA power output (5.8 W)	7.6		dBW
Coupling and feeder losses	-1.0		dB
Antenna gain (3.5 m diameter) at 3 dB edge	31.1		dBi
EIRP		37.7	dBW
Free space attenuation		188.1	dB
Mobile Terminal:			
Antenna gain (1.0 m diameter)	21.9		dBi
Total coupling losses	1.0		dB
Receiving system equivalent noise temperature (226 K)	23.5		dB
G/T		-2.6	dB/K
Boltzmann's constant		-228.6	dBJ/K
C/N_o		75.6	dBHz
Link margin		1.0	dB
Transmission rate (5.0 Mb/s)		67.0	dBHz
Modem implementation loss		-1.5	dB
E_b/N_o		8.4	dB

Table 7.2. Proposed Mobile Up-Link Budget at L-Band for a TDMA/FDMA system.

Frequency : 1,650-1,653 MHz

Mobile Terminal			
HPA power output	7.0		dBW
Coupling and feeder losses	-1.0		dB
Antenna gain (1.0 m diameter)	22.5		dBi
EIRP		28.5	dBW
Free space attenuation		188.7	dB
Satellite			
G/T (as table 7.1a.)		3.5	dB/K
Boltzmann's constant		-228.6	dBJ/K
C/N_o		71.9	dBHz
Link margin		1.0	dB
Transmission rate			
(5.0 Mb/s / 4 = 1.25 Mb/s)		61.0	dBHz
Modem implementation loss		1.5	dB
E_b/N_o		8.4	dB

LNA of the mobile can survive with a higher noise temperature. Of course, besides these extreme ends of the spectrum, intermediate options exist. Since incoming packets at the satellite can be efficiently concatenated together, a TDM down-link was adopted.

7.3. Cost Models.

A financial viability study will now be conducted in order to assess how attractive such a multiple access system is. This will be limited to the space-segment where three main components are identified: the spacecraft itself, the launch vehicle, and the Telemetry, Tracking and Control (TT&C) station.

The spacecraft cost is normally related to mass or primary power. Some models give cost-estimates for subsystems in the satellite and some also include "learning curves" to take into account multi-satellite systems. All these models are empirically or quasi-empirically derived and consequently the figures they yield can only be used as estimates. The actual cost would differ due to changes in technology, unforeseen departures from the mission model and inflation.

The launch cost for most launch vehicles is related to the mass of the satellite. The actual physical dimensions would dictate which vehicle is to be used. This is mainly the case for expendible vehicles. For the S.T.S. (Space Transportation System, commonly known the Shuttle), the cost is dependent on the mass and the length.

Most of the models do not produce a price in the current value of sterling but in U.S. dollars at the value of the year when the model was developed. The method of adjustment to 1985 sterling can either be performed by converting the cost to sterling at the historic rate and using the U.K. indices for the consumer prices to adjust to the present value of sterling, or by inflating the cost by the U.S. indices for consumer prices and translating to sterling at the present rate. The latter method is perferred in this application since the aerospace industry in the West is mainly U.S. dominated. Information regarding the indices for the consumer prices and the rates of exchange for both countries is obtained from the International Financial Statistics (IMF 84, IMF 86). Similar information for the U.K. can obtained from U.K. Government Publications (CSO 86, BSO 86).

Consumer prices were used for their close approximation to the actual inflation rate over the period considered, thus yielding an accurate cost estimate with only a few variables.

7.3.1. The Satellite.

Several models exist (HADF 74, PRIT 79). Pritchard (PRIT 84) has developed a method for estimating the mass and power of the individual subsystems of the satellite. Sandrin (SAND 84) gives a similar model in the appendix to his paper. It was felt that this is more suitable since it provides the desired detail in the system break-down and has been developed for the frequency range of interest. The possibility of either leasing satellite capacity or employing a subsystem similar to the INTELSAT MCS has not been explored. This is because satellites in the Molniya orbit are practically non-existent in the West.

Sandrins model gives the mass of the main subsystems; namely the reflector and its supporting structure, the electronics package and the prime-power subsystem.

The mass of the antenna reflector structure and its supporting structure is modelled as

$$M_A = 0.785\ (1 + K_s)\ K_r\ D^2 \qquad \text{kg} \qquad 7.8$$

$K_r = 0.979$ kg/m^2,

$K_s = 0.3$,

D is the diameter of the antenna in metres.

The mass of an 3.5 m antenna reflector structure is thus given as 12.24 kg.

The mass of the electronic package includes the radiating feed elements, LNA's, diplexers, HPA's, filters and converters, interconnecting cables and electronics, thermal control, and integration structure. This is modelled from data on the INTELSAT MCS package and is given as

$$M_E = 2.2\ N_B\ (12.3 - 2.67\ f + 0.4\ P_{RF}) \qquad \text{kg} \qquad 7.9$$

$$0.8 < f < 1.7 \text{ GHz}$$

where N_B is the required number of beams,

P_{RF} is the required saturated R.F. per amplifier in watts and

f is the frequency in GHz.

For a single beam system operating in L-band with an R.F. power of 5.8 watts, the mass of the electronic package is 23.35 kg.

The mass of prime power subsystem is given by

$$M_{PWR} = K_{PP} P_{SC} \qquad \text{kg} \qquad 7.10$$

where P_{SC} is the total spacecraft prime-power requirement,

K_{PP} is the mass-power efficiency assumed to be 0.091 kg/W. This is probably a conservative estimate since the actual mass-power efficiency for the solar array is closer to 0.033 kg/W (PRIT 84). The assumed figure is for a three-axis spacecraft with full eclipse power in a geostationary orbit and this figure therefore also includes the contribution of the secondary storage cells, although theoretically such batteries are not required for the Molniya orbit. The figure used is thus assumed in order to keep all options open. If we assume that the HPA produces 5.8 watts at 30 % efficiency and that another 10 watts d.c. is required for the receiver and housekeeping electronics, the prime-power requirement is 29.33 W. Putting this in equation 7.10, the mass for the prime-power subsystem is 2.67 kg.

The total mass is 38.26 kg. The mass of the support subsystems is approximately equal to that of the communications and primary power (PRIT 84) while the wet mass for the considered life-time is about 50 % more than the dry mass. The brings the wet mass of the spacecraft at the beginning of the life-time to 114.78 kg.

Using Hadfield's formula, which is given in 1971 U.S. dollars, the payload cost is given by

$$C_p = 0.0440 \text{ K } (n \text{ } M_S)^{2/3} \qquad \$M \text{ (1971)} \qquad 7.11a$$

M_S is the satellite dry mass in kg.

n is the number of satellites.

The power to which it is raised corrects the figure to a learning-curve. For example the cost of a batch of four satellites would be less than four times the cost of manufacture of the single satellite.

K allows for the payload sophistication which includes life-time and environmental protection against adverse natural and man-made effects. Values ranging from 3.2 to 14 have been used. Obtaining a reliable value is difficult because it would require too much detailed information. Hadfield mentions that the value of 5 is probably typical. Iida uses the value of 4. Using these values as a typical range, 5 would signify a high level of complexity and 4 a low level, so we can use the value of 4.5 representing an average level.

To adjust the estimate to 1985 sterling we use the following factor

$$\frac{\text{1985 U.S. index for consumer prices}}{\text{1971 U.S. index for consumer prices} * \text{1985 exchange rate}}$$

$$= \frac{130.55}{49.1 * 1.2964} = 2.051$$

The figures (IMF 84, IMF 86) used are averaged over the four quarters of the year. It should be noted that the use of the rate of exchange makes the cost estimate rather sensitive to fluctuations in the exchange rate.

Equation 7.11a then becomes

$$C_p = 0.406 \; (n \; M_s)^{2/3} \qquad \text{£M (1985)} \qquad 7.11b$$

This would yield a cost estimate of £M 24.16 for four units: three operational units and a spare.

7.3.2. Launch Cost.

In order to maintain flexibilty, a generalised model is used. Here the model by Hadfield is used mainly because it caters for launches in elliptical orbits and because it includes a good mix of vehicles. However, unfortunately the Ariane and the S.T.S. are not included.

Since the spacecraft under consideration are not particularly heavy, a multisatellite launch would be cheaper. The total mass involved could be handled by most of the vechicles except the smallest. The adjusted cost estimate for a launch in which the initial parking orbit and the Hohmann transfer ellipse are coplanar, is given by

$$C_L = 0.0902\ r\ (n_S\ M_S)^{2/3}\ (1 + \frac{H}{13050}) \qquad \text{£M (1985)} \qquad 7.12$$

where r is the number of launches,

n_S is the number of satellites per launch,

H is the apogee altitude in km.

This indicates the cheapest launch since any other launch would require fuel for correcting the angle of inclination. For a launch from the Eastern Test Range which is at latitude of 57^o the fuel required for this would reduce the dry mass to between 75 and 86 % depending on the fuel used. The fuel required would make a shared launch with a geostationary satellite rather inefficient and therefore unlikely.

If we assume that all four satellites are to be launched together and that the apogee altitude is 40,000 km, the launch cost would be £M 21.82. On top of this a launch insurance premium of 10 % can be added so that the total launch estimate becomes £M 24.00.

7.3.3. Telemetry, Tracking and Control Station.

As mentioned earlier the cost of the hardware of a TT&C station is in the region of £M 10. Pritchard (PRIT 79) models such stations as normal earth stations. Asuming R.F redundancy the adjusted formula is given as

$$C_{TTC} = 9473 \, (ERIP)^{0.3} \qquad 7.13$$

where EIRP is in watts.

A typical TT&C station has an EIRP of about 100 dBW at C-band. Using this formula yields a cost estimate of £M 9.5. The station life-time normally exceeds that of the satellites so that replacement is not likely. However, it has to be manned 24hr a day, 365 days a year. Assuming a shift of two and and a five shift system at an annual salary of £ 20,000 and a 100 % overhead rate, the yearly manning cost is £ 400,000. To this the maintenance cost can be added. If we assume that the station has a life-time that greatly exceeds that of the project, then we will assume a flat rate throughout the system life-time of 10 years. This is assumed to be 1 % of the hardware cost per annum. The total running cost of the TT&C station is £ 324,000 per year.

7.3.4. Revenue.

The system tariffs will be assumed to be similar to those of the British Telecom Cellphone system. We will assume that the satellite system caters for national coverage only. There are three types of tariffs:

1. A system connection charge of £ 60. This is charged only once for the life-time of the mobile. It is assumed that the life-time of the mobile is 5 years.
2. An annual subscription charge of £ 300. In the Cellphone system this is a quarterly fee of £ 75. Since the analysis will be made on a yearly basis, an annual fee is used.
3. In the Cellphone system, the price of a call charge varies from 9 p to 35 p for the first minute, depending on location and time of the day. The charge for subsequent minutes is half that of the first. To simplify the analysis an average figure of 22 p will be assumed for first minute and 11 p for minutes thereafter. If we assume a negative exponential distribution for the call duration and that the meam is 2 min, the average call charge would be

   $$22 * (1 - e^{-1/2}) + 11 * e^{-1/2} = 15.3 \text{ p}$$

   A flat call charge equal to this amount will be assumed. The cost for a short coded message will be assumed to be 5 p.

Assuming the traffic profile in section 7.2.1, so that each user makes three long calls and ten SCM's a day, five days a week, 52 weeks a year, then the annual revenue for calls per user is £ 249.34. This brings the total revenue per user to £ 549.34 per year plus £ 60 connection fee for the mobile's life-time, assumed to be five years.

7.3.5. Economic Analysis.

Two types of analysis will be performed: one makes use of the discounted payback period and the other makes use of the net present value and the internal rate of return (ATAT 77, MERR 74, BREA 85). Both methods implicitly assume a time dependence of money since a pound spent this year is worth more than a pound spent next year. It is normally assumed that the value of money can be expressed as

$$M(t) = M(0)\ (1 + r)^{-t} \qquad 7.14$$

where M(t) is the present value of M(0) money invested t years ago,

r is the annual rate of change of worth of money and is usually referred to the cost of money or capital or simply the interest rate. The latter brings about the concept of opportunity cost because it gives the investor a measure of how attractive the project is. If the cost of capital is higher than the normal interest rate then the project will yield more compared to banking opportunities. This does not include inflation but merely provides a method to compare investments.

It is normal to assume that r is constant over the project's lifetime. This is because though the cost of money does varies with time due to ecomonic and political pressures, it is quite impossible to predict such changes. Typical values tend to be employed for various types of projects depending on the size of the project and the risk involved. For such a project 8 % seems to be common and this figure will be used in examples. However, the exact figure is not important in the two methods to be employed.

7.3.6. Proposed Schedule.

If we assume a three year schedule before the operational period then the capital required at the beginning of each year is given in table

7.3. The amount required for the spacecraft production is approximately divided equally over three years. The fourth year represents a typical operational year where the only expenditure required is for the running of the ground station.

Table 7.3. Expenditures in £M for the three pre-operational years and for the first operational year.

Year	0	1	2	3	4
Design, manufacture and testing	8.05	8.05	8.06		
Launch				24.00	
Commissioning of ground station			9.5		
Running of ground station				0.72	0.72
	8.05	8.05	17.56	24.72	0.72

Various profiles can be assumed for capacity fill over the expected life-time of the project. We will assume a linear relationship between capacity fill and time, starting from zero at the beginning of the first year to 100 % at the end of the tenth year (CCST 80). Another possibility could be to assume an initial capacity and an annual growth rate. Figure 7.8 illustrates both profiles assuming a 20 % initial capacity fill and 25 % annual growth rate, obviously limiting the growth to 100 %. The financial impact of such profiles is analysed in figure 7.9.

If we assume the linear profile than the percentage capacity fill and consequently the percentage revenue from the calls and the discounted revenue is given in table 7.4.

Table 7.4. Revenue from calls.

Year	Annual Average Revenue (%)	Annual Discounted Revenue (%) (r=0.08)	Revenue from Connection Fee (%)	Discounted Revenue from Connection Fee (r=0.08)
0	5	5	5	5
1	15	13.889	10	9.259
2	25	21.434	10	8.573
3	35	27.784	10	7.938
4	45	33.076	10	7.350
5	55	37.432	15	10.209
6	65	40.961	20	12.603
7	75	43.762	20	11.670
8	85	45.923	20	10.805
9	95	47.524	20	10.005
		316.785		93.412

The total of the third column gives the present value factor for the calls revenue for such a profile over 10 years. In the above calculations the revenue is averaged over a year. This means that although the profile is called linear in actual practice it is assumed to be stepped with the height of the step equal to the revenue at the middle of the year as illustrated in figure 7.8. The equivalent factor for the revenue from the connection fees is calculated in the last two columns. However, in reality, the results of columns three and five should be discounted taking into account the three operational years, but since the figures are only used to compare different capacity fill models, this is not essential.

Figure 7.9 indicates the present value factor per unit for the calls revenue for a family of profiles where the annual growth is the independent variable and the initial capacity fill is a parameter. The present value factor over the given range of the plot varies from 1 to 7.3, so that the figure obtained by the linear profile is quite conservative and dispenses of another two variables in the analysis.

7.3.7. Discounted Payback Period.

The payback period essentially asks the question : having made a particular investment, how long must we wait to recover the money invested? In its most basic form it ignores the dependence of money with time and any cash flows that occur after the payback period has expired. This is quite an over-simplification. However, it is very popular and very frequently used since the underlying concepts are easy to grasp and since it provides a simple screening device - inaccurate as it may be - of the speed at which invested funds are returned to the project.

The discounted payback period recognises that a pound today is worth more than a pound tomorrow, although it still carries the limitations associated with the payback listed above. In this section, the discounted payback period will be analysed over the range of r from zero to 23 %. It should be noted that when r = 0 the discounted payback period is no longer disounted and reverts to the basic form.

The present value of any project is found by discounting at the project's cost of capital all future net cash flows to their present value equivalent. The net present value (NPV) of the project at year T of its lifetime is defined as

$$NPV(T) = \sum_{t=0}^{T} [R(t) - E(t)] [1 + r]^{-t} \qquad 7.15$$

where R(t) is the receipts for year t,

E(t) is the expenditure for year t,

r is the cost of capital and

R(t) - E(t) is the net cash flow for year t.

Assuming $r = 8$ %, figure 7.10 shows the NPV throughout the life-time of the project. It should be observed that in the initial phase the NPV is negative and in the final phase it is positive, i.e. the project starts earning money. By the end of the project the NPV is around £M 70. The discounted payback period at 8 % is given by the point where the NPV is zero. In this case it is approximately about 8.6 years and since during the three preoperational years there is no income, it implies that it takes 5.6 years of service to recover the initial investment.

Figure 7.11 indicates the discounted payback period for various costs of capital. The period is expressed in terms of years of service, i.e. year zero is the end of the three years of development and manufacture. The discounted payback period is approximately equal to the expected life-time of the project at the maximum considered value of the cost of capital of 23 %. This implies that within this range of cost of capital there will always be a positive NPV at the end of the project. The shape of the plot is, of course, indicative of the structure of the project. There is a heavy investment in the initial phase when the buying power of the currency is high. On the other hand, the income is greatly weighted at the end of the project where the currency is weakest. This implies that a high cost of capital extends the discouted payback period.

7.3.8. Internal-Rate-of-Return.

The internal-rate-of-return (IRR) technique views the situation slightly differently. The IRR assumes the cost of capital to be a variable and defines it as the cost of capital that would make the NPV at the end of the life-time of the project equal to zero or expressed mathematically

$$0 = \sum_{t=0}^{T} [R(t) - E(t)] [1 + IRR]^{-t} \qquad 7.16$$

where T is now the life-time of the project.

It is called internal because the rate is solely based upon the cash flow and there are no external influences. This seems to be a more recognised technique than the previous method because it appraises the project over its entire life-time. However care should be taken when the cash flow profile alternates between positive and negative values since this will not yield a unique value for the IRR. Fortunately, this does not apply in this case because the bulk of the expenditure occurs at the beginning of the project.

The NPV at the end of the life-time of the project is plotted against the cost of capital in figure 7.12. This indicates that the IRR for this project is about 22.2 %, so that as long as the cost of capital remains below this figure the NPV at the end of the project will be positive. Since the value of the cost of capital is expected to be well below this value of the IRR, the project appears attractive from this aspect because it guarantees a positive NPV at the end of the project.

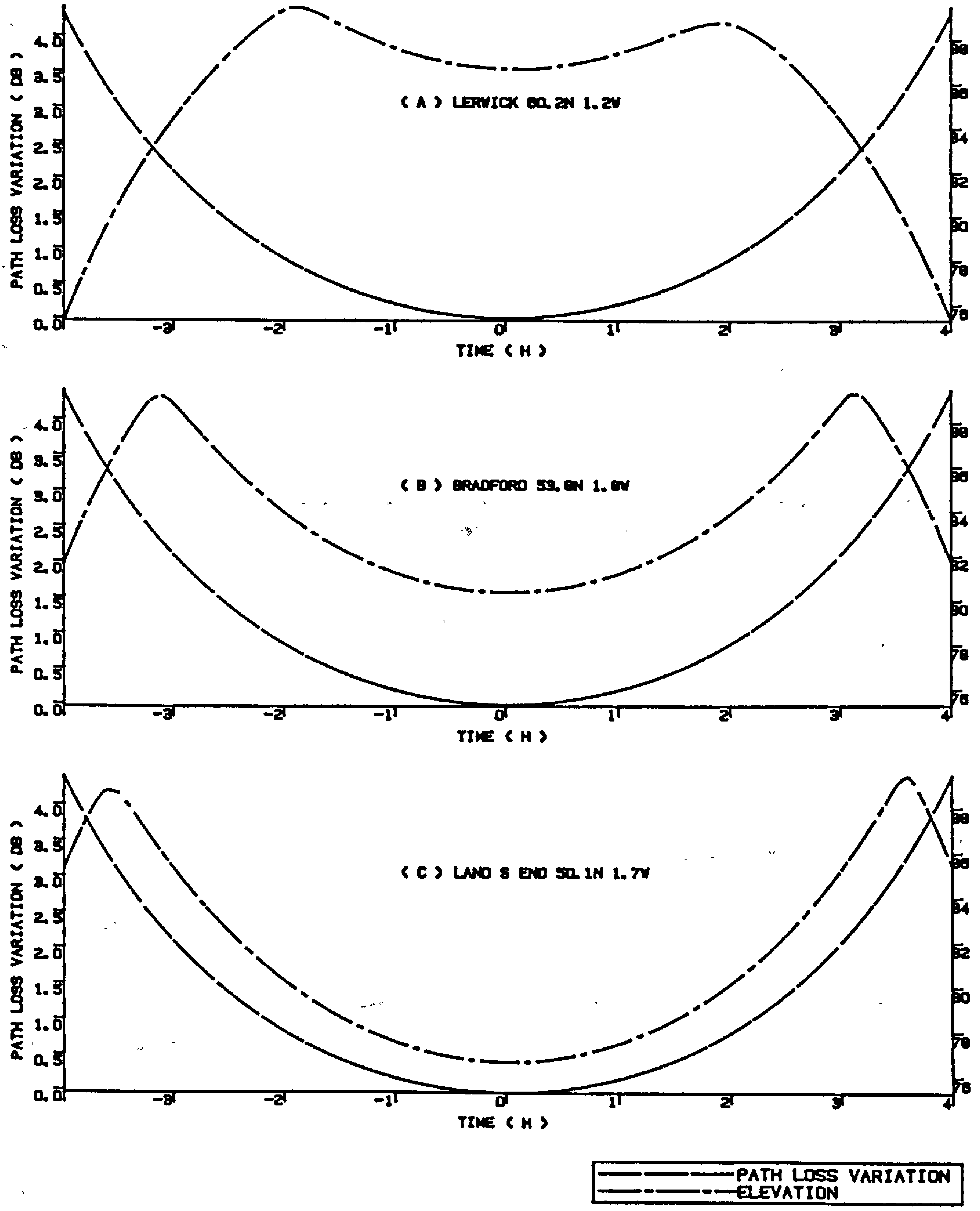


Figure 7.1. Variation of signal strength and elevation over the 8 h active service of the satellite for Lerwick, Bradford and Land's End. The satellite is in a Molniya orbit with the apogee positioned at E 3°.

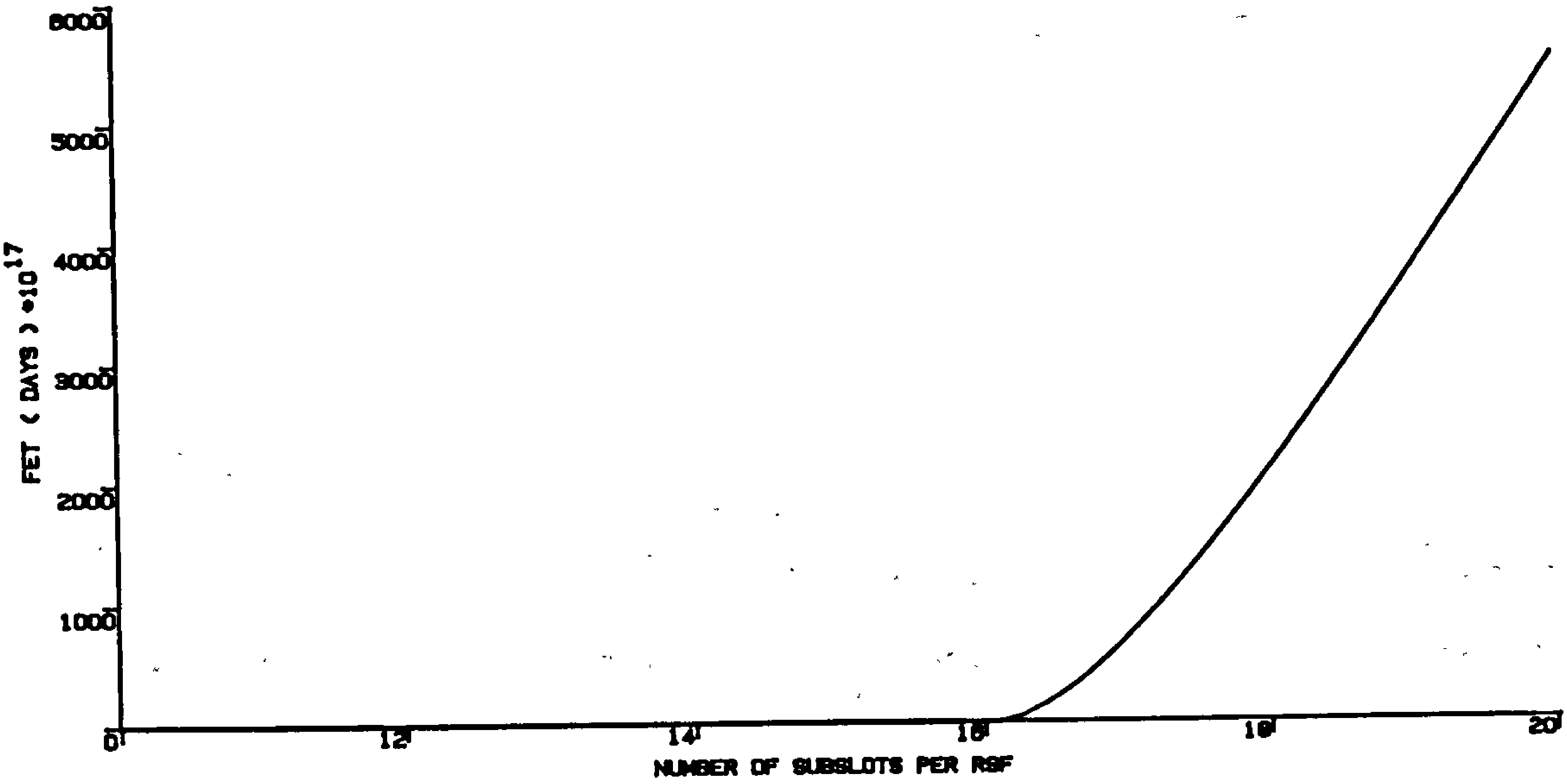


Figure 7.2. The FET of the S-ALOHA RSF for various numbers of subslots in the RSF.

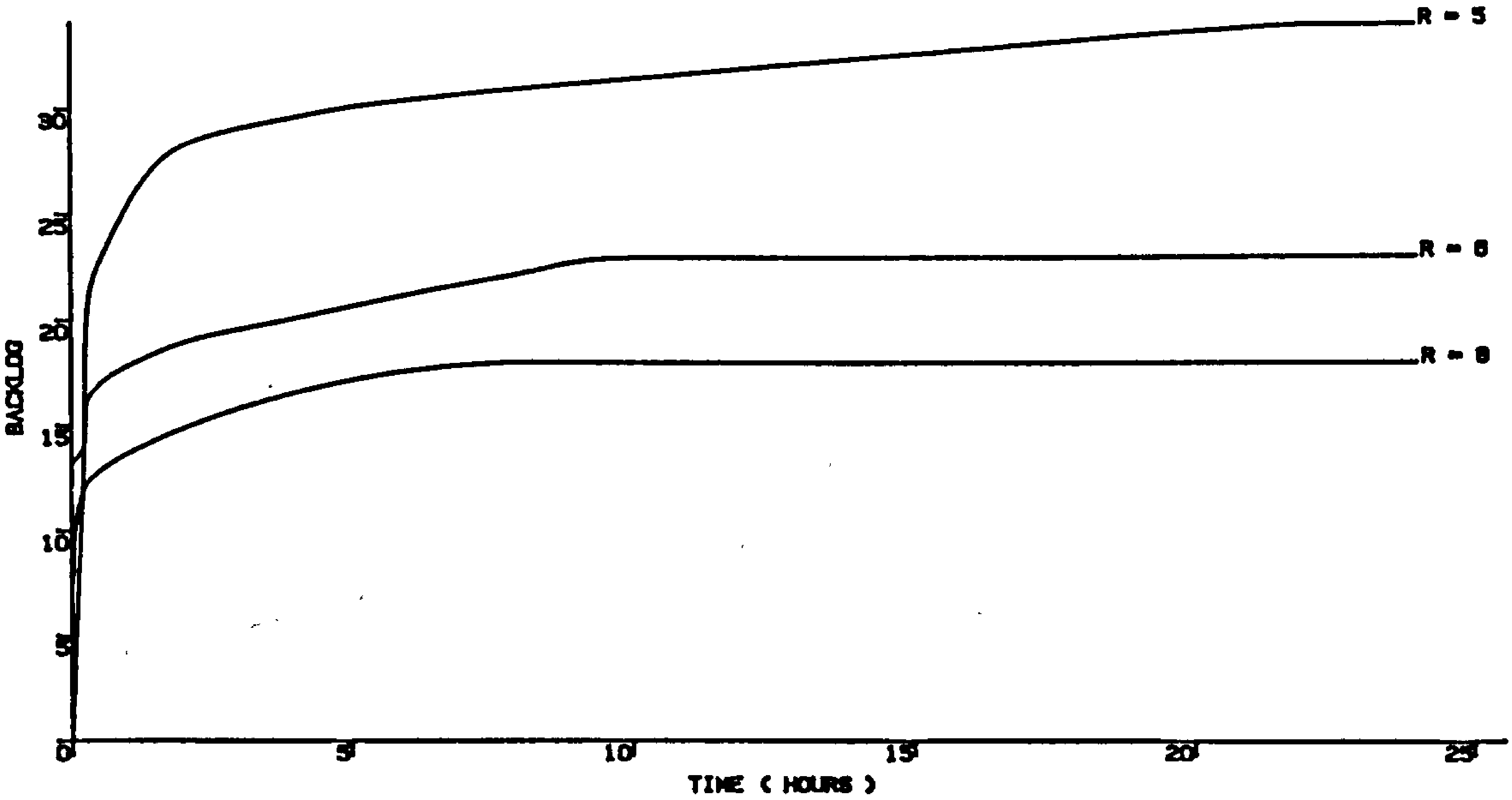


Figure 7.3. The maximum backlog with time of a P-ALOHA RSF over a simulation run.

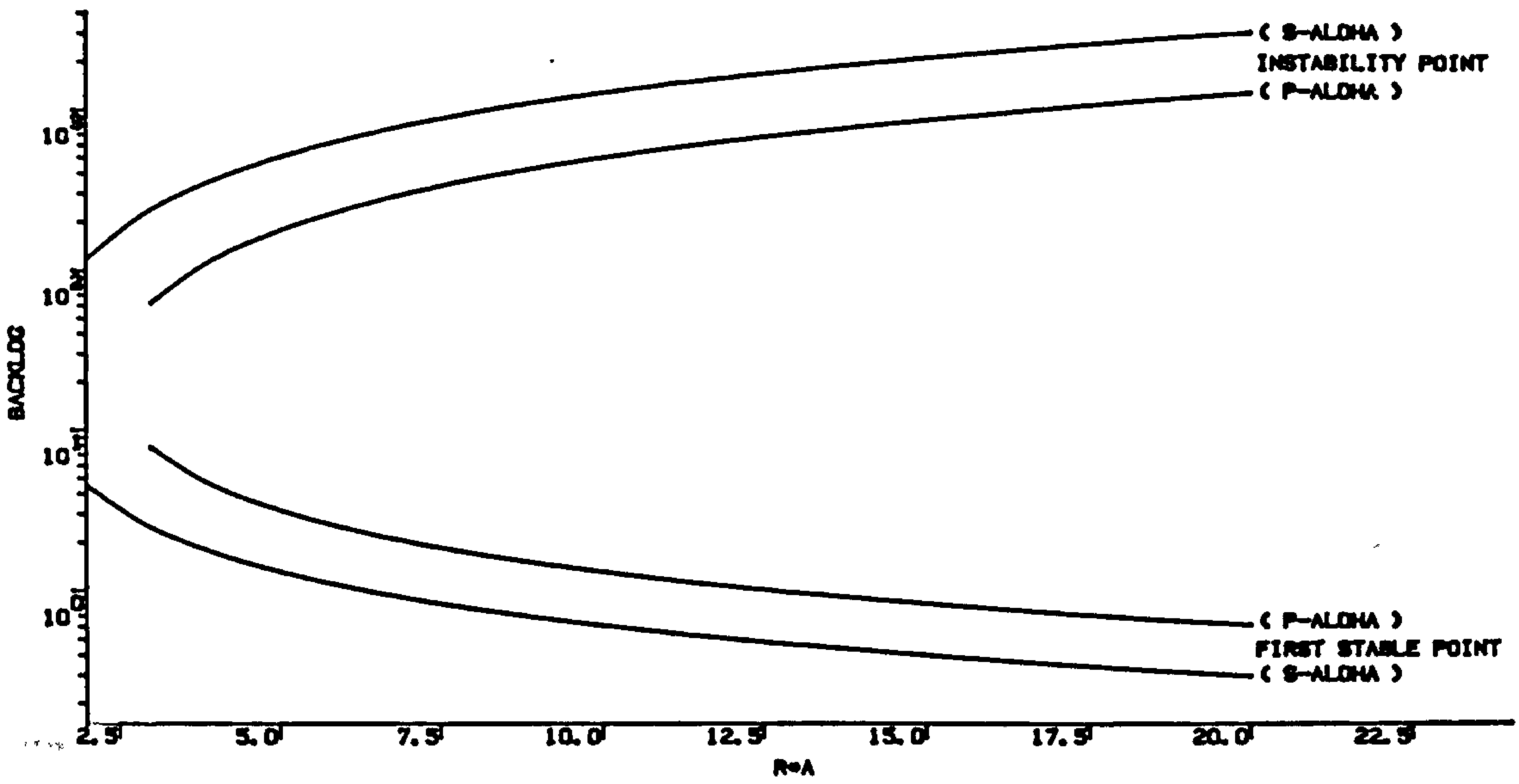


Figure 7.4. The first and second intercept points of the G-S curves and the load line for S-ALOHA and P-ALOHA.

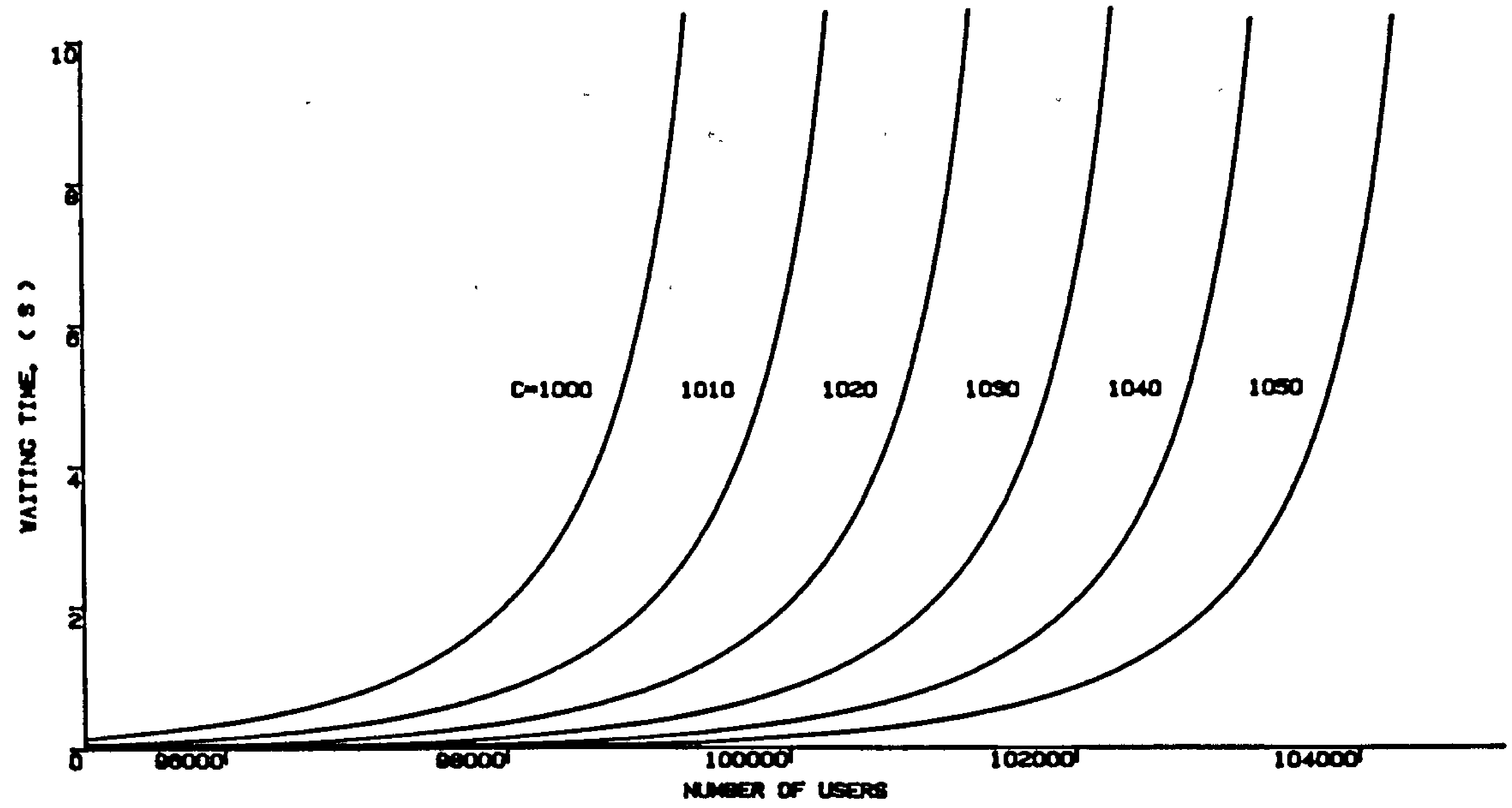


Figure 7.5. The waiting time versus number of users with number of servers as a parameter for a M/M/c queue.

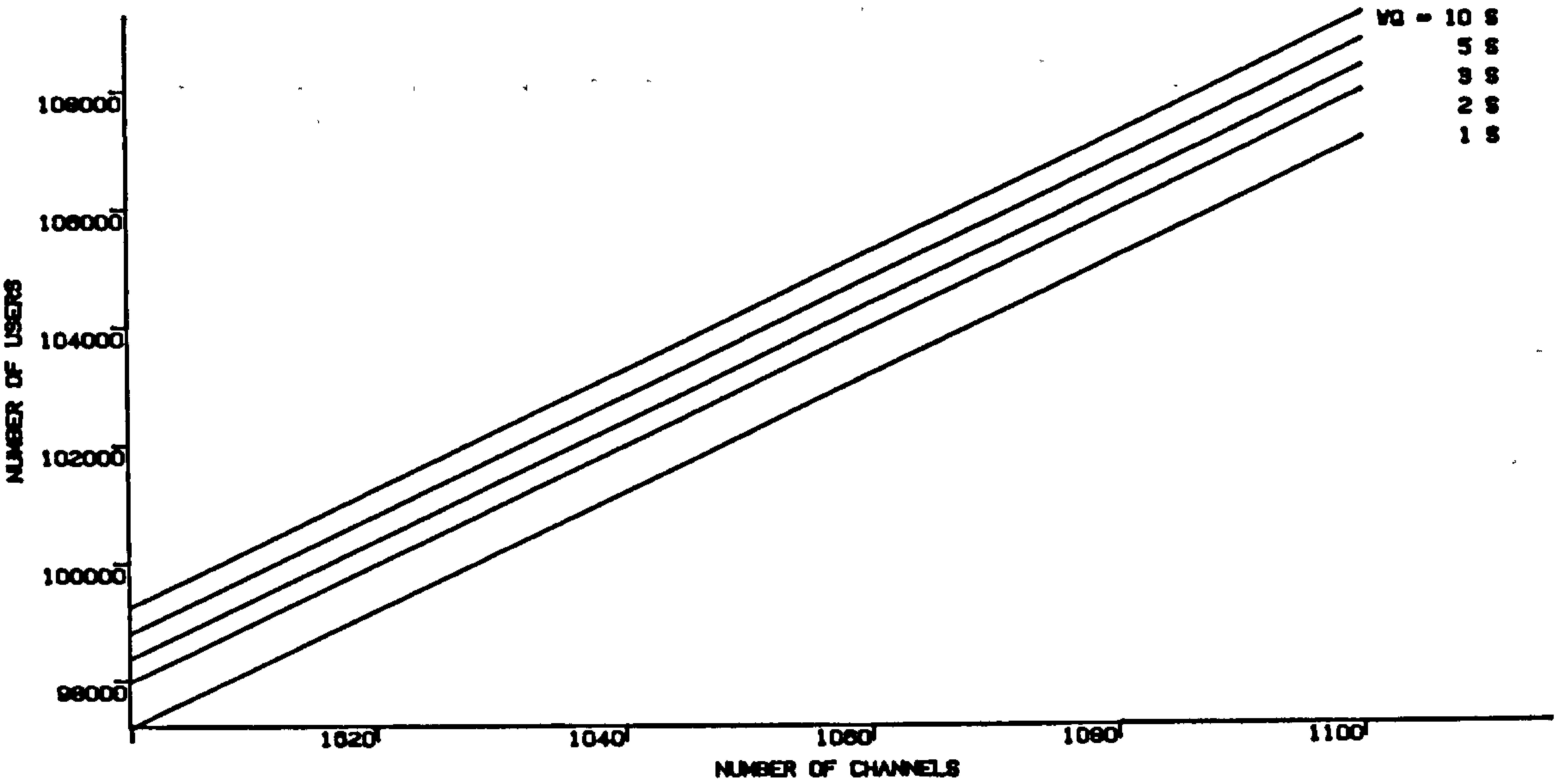
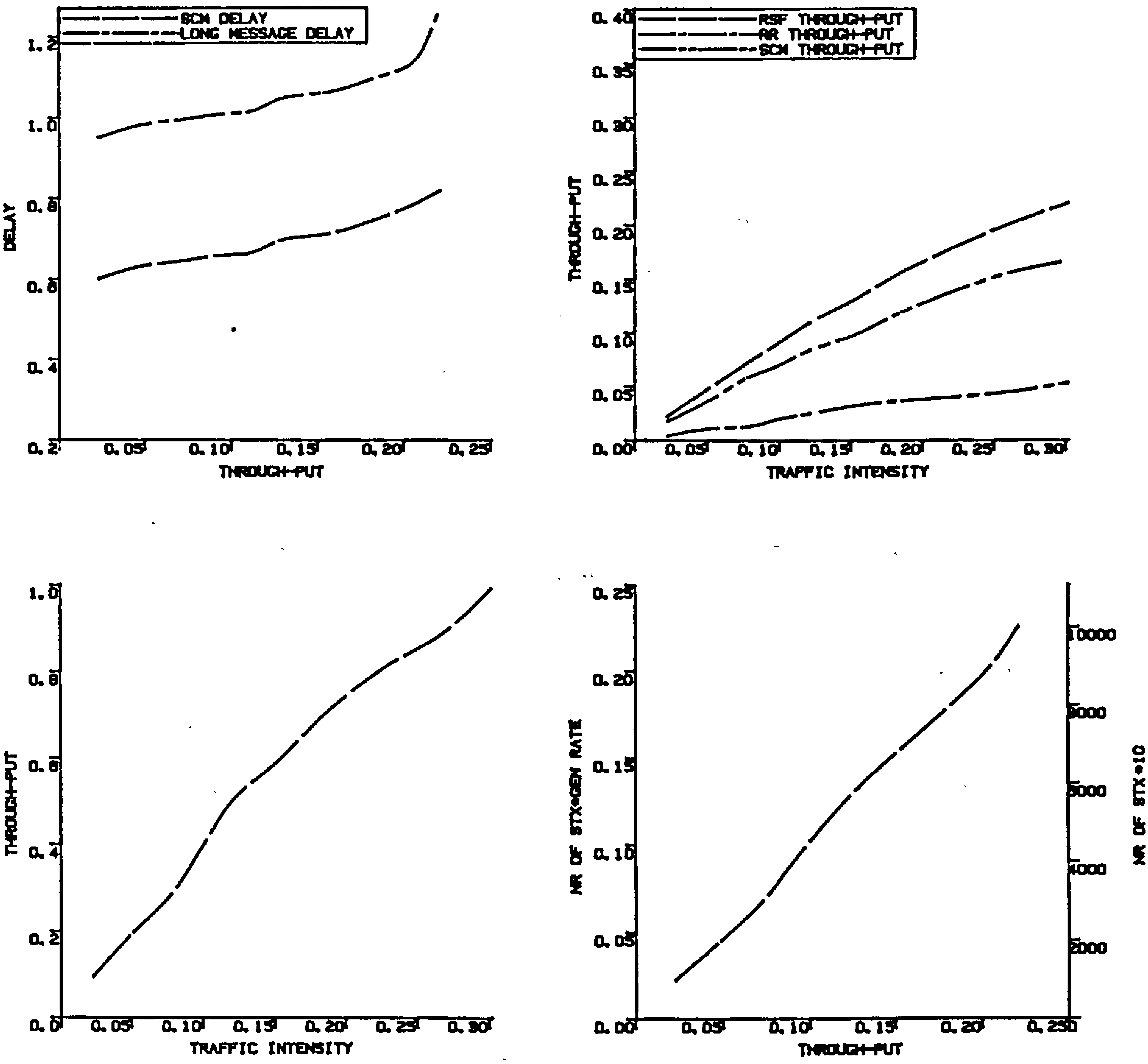


Figure 7.6. Number of users versus number of channels with delay as a parameter for a M/M/C queue.



SHORT MESSAGE GENERATION RATE (PER HR) : 1.000
LONG MESSAGE GENERATION RATE (PER HR) : 0.300
MEAN MESSAGE LENGTH (IN SECONDS) : 120
PROPAGATION DELAY (IN MILLISECONDS) : 270.0
FRAME DURATION (IN MILLISECONDS) : 100.0
NUMBER OF SLOTS IN A FRAME : 1014
NUMBER OF SLOTS IN A RSF : 4
NUMBER OF SUBSLOTS IN A RSF SLOT : 4

Figure 7.7. Simulation results for a system with terrestrial controllers.

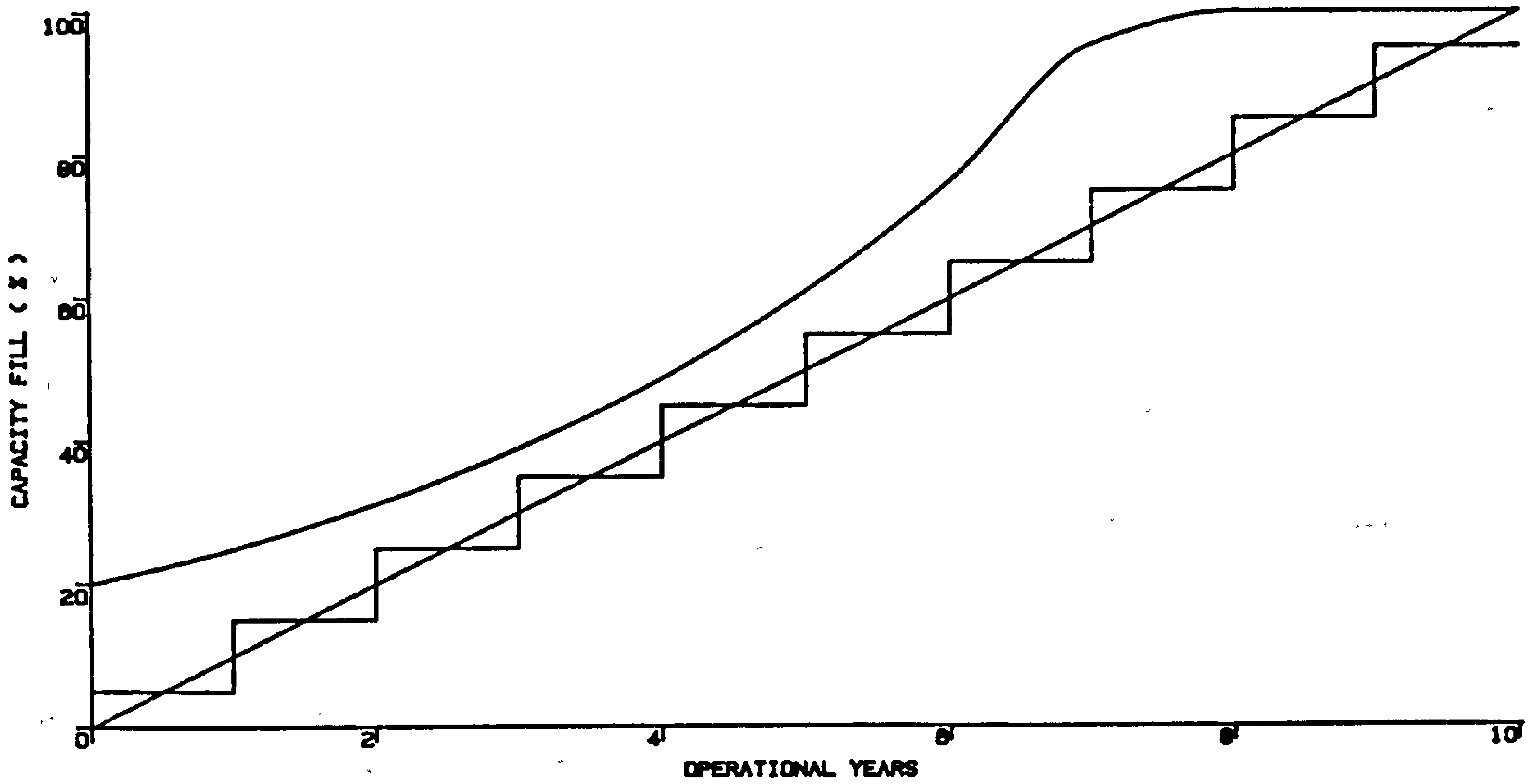


Figure 7.8. Two theoretical capacity fill profiles: one assuming a 20 % fill over the first year increasing annually at a growth rate of 25 %, the other assuming a linear fill and indicating annual averaging.

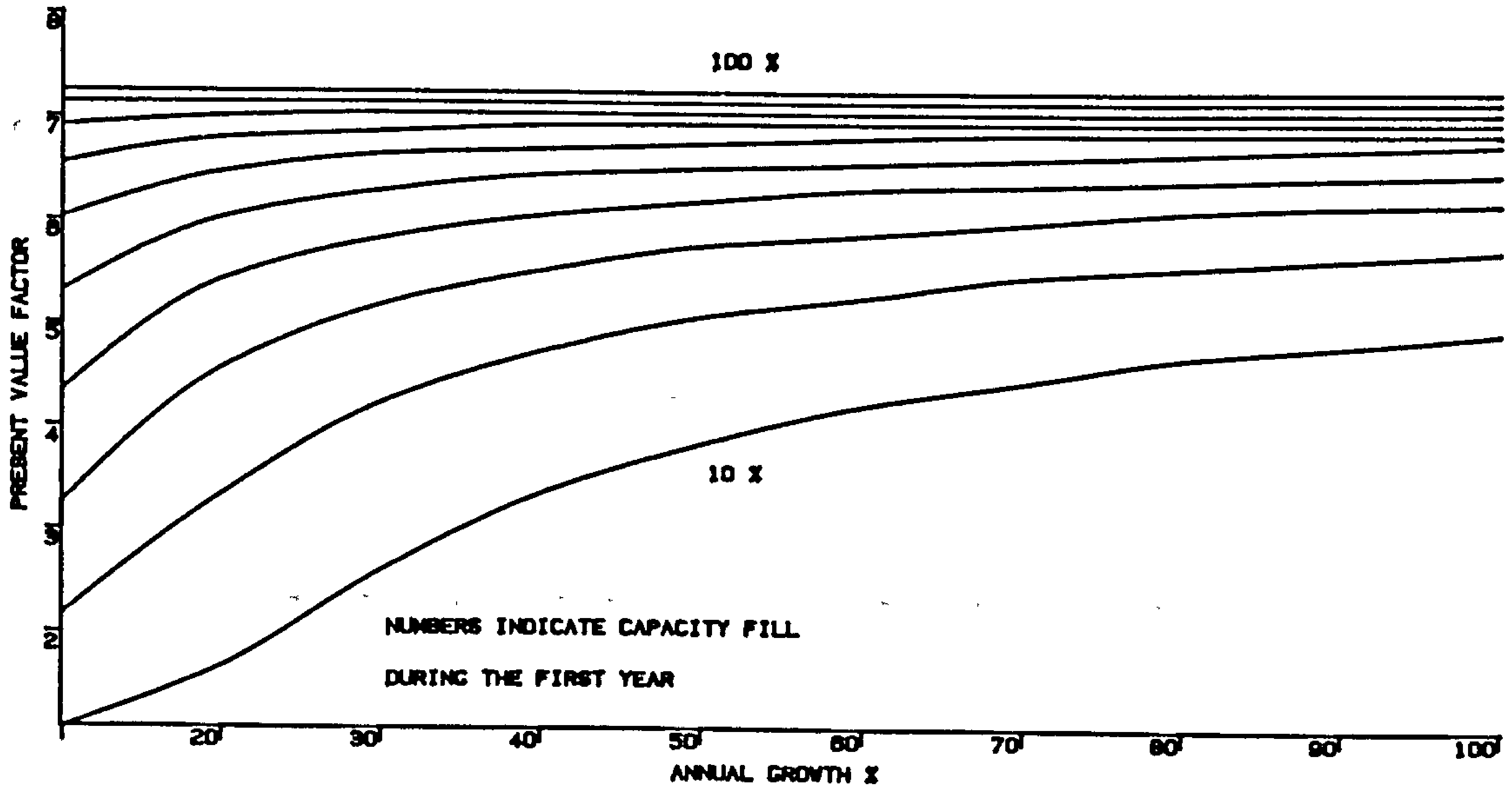


Figure 7.9. The present value factor for various capacity fill profiles with various initial capacity fills and various growth rates.

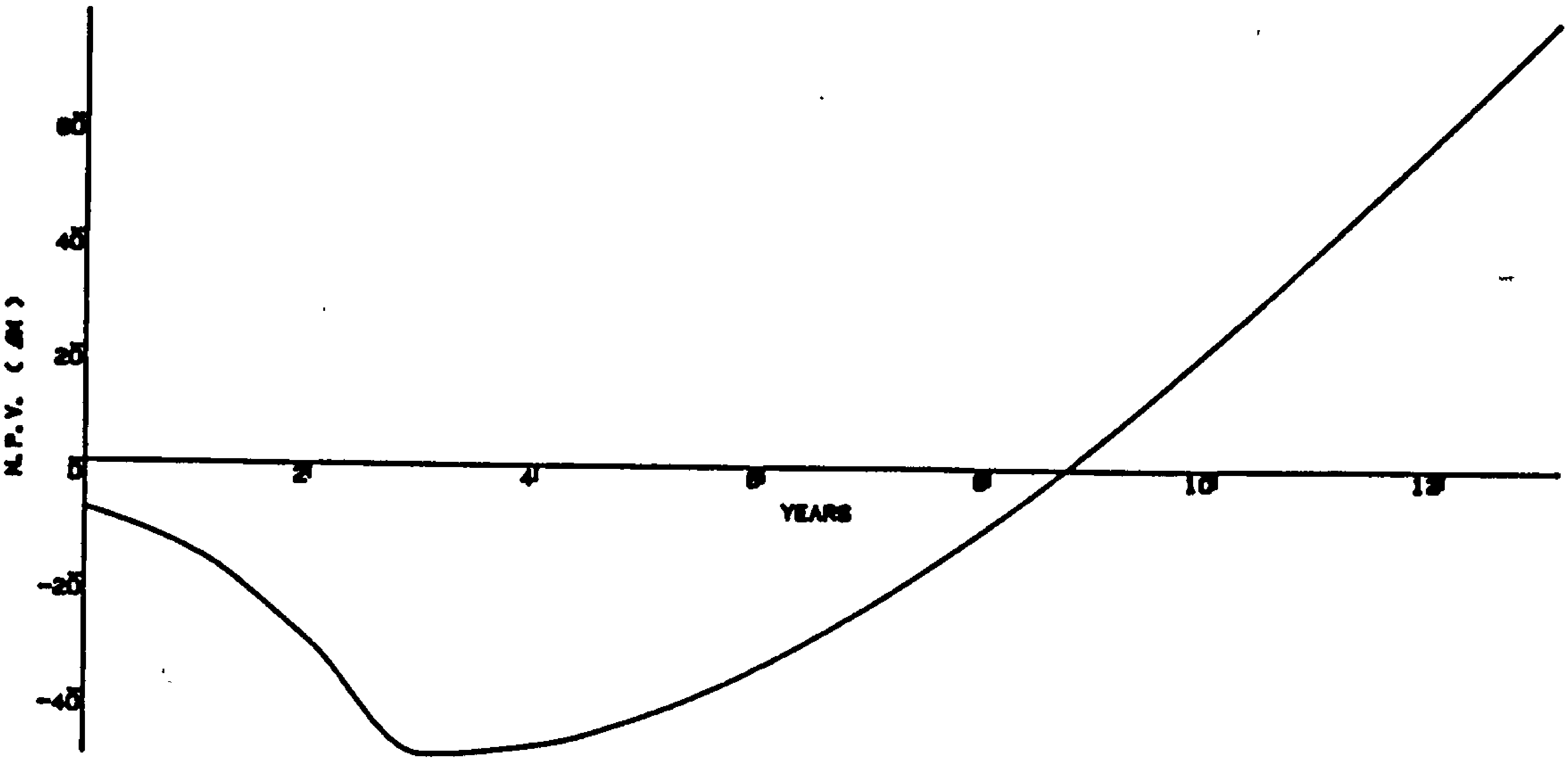


Figure 7.10. The NPV throughout the project life-time for a cost of capital of 8 %.

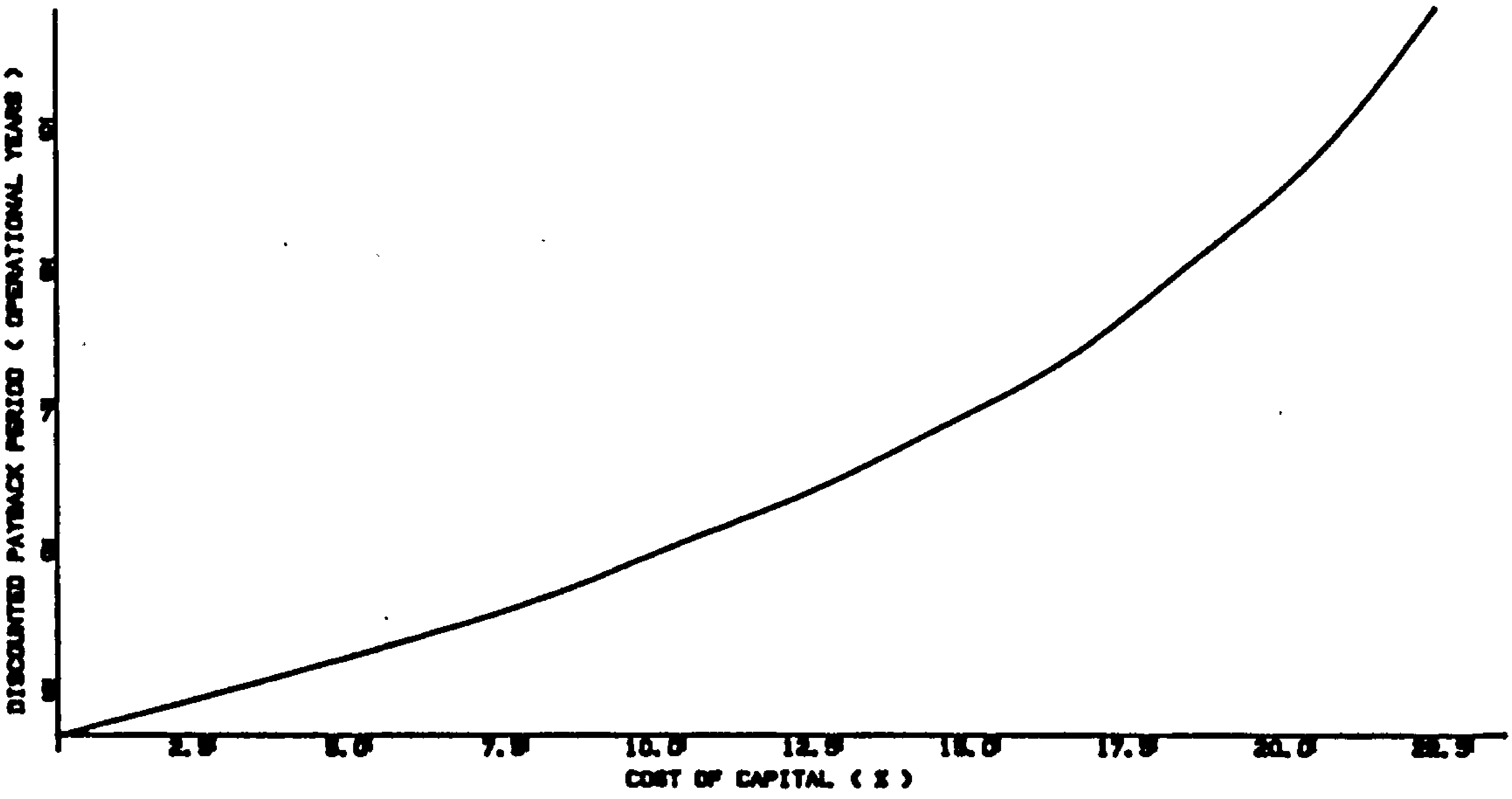


Figure 7.11. The discounted payback period in terms of operational years for a range of values for the cost of capital.

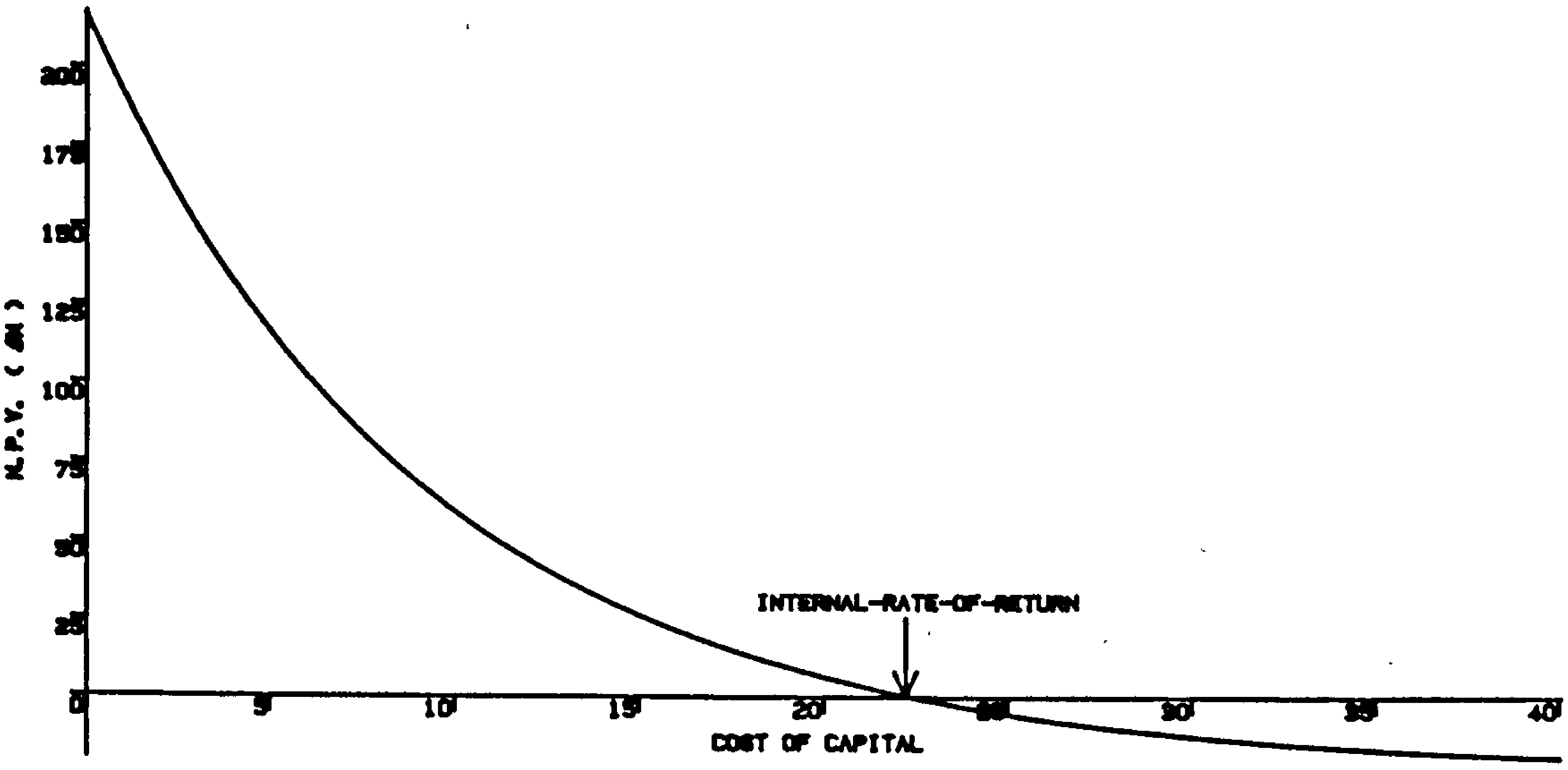


Figure 7.12. The NPV at the end of the project for various values of the cost of capital. The intersection on the x-axis indicates the IRR.

CHAPTER 8.

CONCLUSION.

8.1. Multiple Access Framework.

The objective of this thesis was to study, analyse and assess multiple access systems for a population of land mobile terminals operating in a satellite system to cover the U.K. Interest was focussed on exploiting the use of the Molniya orbit, which is ideally suited for a mobile system operating in a country at high latitudes like the U.K., or even Europe. Employing such an orbit implies operating at high angles of elevation which greatly simplifies the antenna design and reduces problems associated with multipath and shadowing.

The geostationary orbit and constellations of satellites at lower altitudes in inclined orbits have also been investigated but were considered less attractive in this context. The former requires steerable or inefficient antennae and operation at low angles of elevation while the latter does not offer the simplest solution in this context. The geostationary orbit is best suited for point-to-point communications or mobile communications within an area centred around the equator. A constellation of satellites truly offers global coverage.

FDMA, TDMA and CDMA have been considered for the multiple access system, particularly for the mobile up-link. CDMA has advantages in military application and in civilian spheres where network control is to be kept to a minimum and the traffic from each user has a very low duty cycle. For optimal resource utilisation, FDMA and TDMA have to be considered.

TDMA is attractive since it lends itself to digital techniques which can be implemented easily using integrated circuit (possibly VLSI) technology. This can allow for great flexibility at potentially low cost. However, the burst mode of operation imposes hefty requirements for the mobile HPA. FDMA has the advantage of requiring lower power when compared to the burst mode TDMA but processing at RF is not considered desirable.

A hybrid TDMA/FDMA system is recommended for the mobile-to-satellite link where a number of TDMA systems operate within a FDMA structure. This reduces the transmission rates and therefore the RF power required from the mobile HPA while benefiting from digital techniques at lower frequencies and consequently lower cost and possibly lower power consumption.

8.2. The Protocol.

A reservation technique is used to ensure efficient use of the space-segment resources and the available spectrum. The frame is split into two subframes: the reservation subframe and the information subframe. Though the total capacity is fixed, the proportions assigned to each subframe are dynamically controlled. This ensures that the protocol performs well under a large variety of traffic profiles.

Besides reservation requests, the reservation subframe can also handle short coded messages. This is a short (about 100 bits) complete entity of data. Different message sets can of course be devised for different applications.

The reservation subframe makes use of ALOHA techniques. Both the pure and the slotted variants have been considered. ALOHA was chosen for its ability to perform in environments where the propagation delay is long, the user population is large and where the traffic profile is bursty. They provide a fair accessing technique and the implementation is relatively simple. The fact that their through-put is low, is counteracted by the fact that the reservation subframe normally constitutes a small overhead.

The assumed traffic model consists of a population of 100,000 users, each generating a SCM every hour, and a long message approximately every three hours. The average length of a long message is 2 min. Four and six slots are required for the RSF for S-ALOHA and P-ALOHA respectively, and because of the size of the short messages each of these slots are further subdivided into four subslots. The frame duration is 100 ms.

Under these conditions, the SCM delays are 0.82 and 1.63 s for systems with space-borne and terrestrial collision detectors respectively. Both these delays are acceptable and consequently a terrestrial collision detector is preferred to keep the satellite complexity low.

The ALOHA subframe operates in the bistable state in order to maintain high spectral efficiency. The FET (average First Exit Time) criterion was used as a stability measure in the bistable situation. The FET of the S-ALOHA system was calculated to be $6.4*10^{14}$ days. This implies that taking the diurnal distribution of the traffic load on the system, the probability of requiring a control mechanism is very small. The same concept was employed for the P-ALOHA. However, a simulation approach was adopted. Obviously simulating such lengths of time is not practical and a simulation run equivalent to 10 days was performed. With the given assumptions, the peak backlog reached 23 while instability is considered to start at a backlog of 62.

From the assumed information and transmission rate, a total of 1,050 slots are available. If 1,040 slots are assigned to the ISF, the total delays drops to 2.2 s for a terrestrial ISF controller. Again the terrestrial version performs adequately well and is chosen.

The steady state results for protocols employing both ALOHA versions were derived from analytical analyses. Clock driven simulation techniques were found to be a useful method of validating the mathematical model. The stability of the slotted ALOHA protocol in the bistable state has been studied by employing a discrete parameter Markov chain model. However, a readily tractable analytical model for the pure ALOHA protocol is not available and an event driven simulation model was used.

8.2.1. Refinements.

The P-ALOHA version of the RSF requires minimal acquisition effort since the two processes of acquisition and access can be integrated together by employing the delay from the geographical spread of user as part of the random delays in the ALOHA process. The acknowledgment will then also include information regarding the position of the received packet in the subframe so that the users are provided with local ranging data.

For a U.K. system, the RSF would have to be at least 10 or 20 slots long depending on whether the collision detector is space-borne or terrestrial. A 10 slot long RSF would fit with a 1,040 ISF so that the delay figures remain unchanged. Using a terrestrial detector, either requires marginally more capacity (5.088 Mb/s instead of 5.0 Mb/s) or there would be some degradation in the long message delay since 1,030 slots would be available for the ISF. This would rise to 2.6 s.

As long as the mix of the load is biased towards long messages rather than short coded messages, P-ALOHA offers a simple and effective system. Efficiency and stability can be attained by extending the length of the subframe. Slotted ALOHA would be preferred if the mix were weighted towards short messages where the aggregate through-put tends towards the subframe through-put. With the assumed taffic mix, P-ALOHA is the most likely candidate.

In order to optimise the mobile link budgets, a regenerative transponder is assumed at the satellite. This minimises the RF power requirements and the figure of merit of the mobile terminal.

8.3. Financial Feasibility Study.

The inputs to the cost analysis originated from models available in literature over a number of years. The costs have been updated using the index for consumer prices. The analysis was performed using two techniques:

1. The discounted payback period is dependent on the cost of capital and the analysis was performed for a range in the cost of money of 0 to 23 %. For the considered range of the cost of capital, the payback period is within the project's life-time, although at the high end of the cost of capital, it is approximately equal to the project's life-time, indicating a high financial risk at this point. However, since such values are not very realistic, such a risk is highly unlikely. At the typical cost of capital of 8 % the payback period is about 5.6 operational years, which is reasonable.

2. The internal-rate-of-return is estimated to be 22.2 %. Since the cost of capital is not expected to be that high, this result is favourable. At a more typical value of the cost of capital of 8 %, the net present value of the project is about £M 70 at the end of the project's life-time.

This indicates that with both techniques yielding a positive result with a considerable margin of safety, the risks involved are less crucial especially when considering that only 10 % of the U.K. population is expected not to be covered by a cellular system by 1990 while 50 % of the European land mass will not be covered by the year 2000.

8.4. Further Work.

The proposed protocol performs within the established specifications for the assumed typical communication traffic profile. Mobile satellite systems are not likely to compete with terrestrial systems and therefore will be employed to augment terrestrial systems (FENE 85). It is not very clear how this will affect the satellite traffic.

The demand for a land mobile satellite system is obvious from a recent workshop at ESA/ESTEC (ESA 86). However, very few market surveys distinguish between the portion of traffic that can only be handled by satellites. A development of a traffic model is required that includes this fact.

The tolerance of a Molniya satellite position with respect to the expected position must be accurately determined in order to reduce timing uncertainties and guard times. Depending on the accuracy, the need and the frequency of producing telemetry information can be determined.

The technique of integrating the reservation and acquisition processes requires detailed information on the population distribution, taking into account regional and urban variations. The feasibility of such a technique depends on the population distribution relative to the (moving) sub-satellite point. The north-south variation in the population density is more important than that in the east-west direction because of the shape of the country.

A truly mobile system implies that the terminal should be physically small and light and the limiting component is normally the antenna. One factor that influences this is frequency and on this account there is a trend towards experiments at higher frequencies.

An examination of the International Table of Frequency Allocations reveals that in Europe there are no primary or secondary allocations to land mobile satellite service. Further work is therefore required on the regulatory side to identify the likely demand, user needs and spectrum requirements. The I.T.U. (International Telecommunications

Union) will be holding a W.A.R.C. (World Administrative Radio Conference) in 1987. At W.A.R.C. (MOBILE) 87 the possibility of using the presently unused aeronautical mobile satellite bands at L-band for land mobile will be discussed. No regulatory consideration is being given to higher frequencies at present.

A multi-spot Molniya system is attractive because of frequency re-use and the possibility of higher power density from the satellite. However, this poses some challenging problems to the antenna designer where varying altitudes would require controlling the beamwidths of each foot-print and also steering them to maintain the same foot-print centres.

APPENDIX A1.

PASCAL SOURCE CODE PRODUCING THE SUB-SATELLITE TRACE.

```
PROGRAM SUBSATTRACE (INPUT/,OUTPUT,Data); (* This program requires the
orbital data to be entered.  The orbit can  be  an  inclined  circular
orbit or an elliptical orbit.  The output from file, Data, is the sub-
satellite coordinates every 2 min.  This interval is controlled by  S.
*)

CONST
   S=720;
   R=6370;          (* Earth's radius, Km *)
   TeH=23;          (* Hours of sidereal day *)
   TeM=56;          (* Minutes of sidereal day *)
   TeS=4.09;        (* Seconds of sidereal day *)
   G=3.99E5;        (* Gravitational constant, Km^3/s^2 *)
TYPE
   TrigF=(SinA,CosA,TanA);
   DataType=(Lon,Lat);
   ArrayType=ARRAY[DataType,0..S] OF REAL;
VAR
   Typ:CHAR;
   I,J:INTEGER;
   AreaInc,AreaTotal,A,Alfa,Alt,Alt1,B,Beta,C,Dem,E,Gamma
   ,Inc,K,NewArea,Nom,Pi,Phi,Psi,Ra,T,Te,Theta,We,Ws:REAL;
   OutDat:ArrayType;
   Data:TEXT;

FUNCTION InvTri (Tri:TrigF;Nom,Dem:REAL):REAL;
   VAR
      Angle,X:REAL;
   BEGIN
      X:=Nom/Dem;
      CASE Tri OF
         SinA :
            Angle:=ARCTAN(1/SQRT(1/X/X-1));
         CosA :
            Angle:=ARCTAN(SQRT(1/X/X-1));
         TanA :
            Angle:=ARCTAN(X);
      END;
      IF Angle<0
      THEN
```

```
         Angle:=-Angle;
      IF Dem<0
      THEN
         BEGIN
            IF Nom>0
            THEN (* Pi/2 TO Pi *)
               Angle:=Pi-Angle
            ELSE          (* Pi TO 3*Pi/2 *)
               Angle:=Angle-Pi;
         END
      ELSE
      IF Nom<0
      THEN    (* 3*Pi/2 TO 2*Pi *)
         Angle:=-Angle;
      InvTri:=Angle;
   END;

PROCEDURE Order;
   VAR
      I,J:INTEGER;
      Dummy:ArrayType;
   BEGIN
      I:=0;
      WHILE OutDat[Lon,I]>=0 DO
         I:=I+1;
      J:=I;
      FOR I:=J TO S DO
         BEGIN
            Dummy[Lon,I-J]:=OutDat[Lon,I];
            Dummy[Lat,I-J]:=OutDat[Lat,I];
         END;
      FOR I:=1 TO J DO
         BEGIN
            Dummy[Lon,S-J+I]:=OutDat[Lon,I-1];
            Dummy[Lat,S-J+I]:=OutDat[Lat,I-1];
         END;
      OutDat:=Dummy;
   END;

FUNCTION NewAngle (Theta1,Theta2,NewArea:REAL):REAL;
   CONST
      Accuray=0.001;
   VAR
      Found:BOOLEAN;
      Area,Phi,Theta,LimitLo,LimitHi:REAL;

   FUNCTION CalArea(A,B,Theta:REAL):REAL;
      VAR
         X:REAL;
      BEGIN
         Phi:=InvTri(TanA,B*SIN(Theta),A*COS(Theta));
         X:=A*A*COS(Theta)*COS(Theta)+B*B*SIN(Theta)*SIN(Theta);
         CalArea:=A/2*(SQRT(X*(1-B/A*B/A))*SIN(Phi)+B*Theta)
      END;
   BEGIN
```

```
      LimitLo:=(1-Accuray)*NewArea;
      LimitHi:=(1+Accuray)*NewArea;
      Theta:=Theta1;
      Found:=FALSE;
      WHILE NOT Found DO
         BEGIN
            Theta:=(Theta1+Theta2)/2;
            Area:=CalArea(A,B,Theta);
            IF (Area>LimitLo) AND (Area<LimitHi)
            THEN
               BEGIN
                  NewAngle:=Theta;
                  Found:=TRUE;
               END
            ELSE
            IF Area<NewArea
            THEN
               Theta1:=Theta
            ELSE
               Theta2:=Theta;
         END;
   END;

BEGIN
   REWRITE(Data);
   Pi:=4*ARCTAN(1);
   Te:=TeS+60*(TeM+60*TeH);
WRITE('           WHAT KIND OF ORBIT, C(IRCULAR) OR E(LLIPTICAL) : ');
   READLN; READ(Typ); WRITELN;
   IF Typ='C'
   THEN
      BEGIN
WRITE('                   AlTITUDE OF SATELLITE ABOVE EARTH (KM) : ');
         READ(Alt); WRITELN;
         A:=Alt+R;
         B:=A;
         E:=0;
      END
   ELSE
      BEGIN
WRITE('         AlTITUDE OF SATELLITE ABOVE EARTH AT APOGEE (KM) : ');
         READ(Alt); WRITELN;
WRITE('        AlTITUDE OF SATELLITE ABOVE EARTH AT PERIGEE (KM) : ');
         READ(Alt1); WRITELN;
         A:=(Alt+Alt1+2*R)/2;
         E:=1-(Alt1+R)/A;
         B:=A*SQRT(1-E*E);
      END;
WRITE('    ORBITAL INCLINATION TO THE EQUITORIAL PLANE (DEGREES) : ');
   READ(Inc); WRITELN;
   Inc:=Inc*Pi/180;
WRITE('                   LONGITUDE OF APOGEE OR CREST (DEGREES) : ');
   READ(Gamma); WRITELN;
   Gamma:=Gamma*Pi/180;
   Ws:=SQRT(G/EXP(3*LN(A)));
```

```
  We:=2*Pi/Te;
  C:=A*E;
  BEGIN
     AreaTotal:=Pi*A*B;
     AreaInc:=AreaTotal*Ws/2/Pi*Te/S;
     NewArea:=-AreaInc;
     Theta:=0;
     FOR I:=0 TO S DO
        BEGIN
           NewArea:=NewArea+AreaInc;
           T:=Te*I/S;
           IF NewArea>AreaTotal
           THEN
              BEGIN
                 NewArea:=NewArea-AreaTotal;
                 Theta:=0;
              END;
           IF NewArea=0
           THEN
              Theta:=0
           ELSE
              Theta:=NewAngle(Theta,2*Pi,NewArea);
           Ra:=SQRT(A*A*COS(Theta)*COS(Theta)
               +B*B*SIN(Theta)*SIN(Theta));
           Phi:=InvTri(TanA,B/A*SIN(Theta),COS(Theta));
           Psi:=InvTri(TanA,Ra*SIN(Phi),(Ra*COS(Phi)+C));
           Alfa:=InvTri(TanA,SIN(Psi),COS(Psi)*COS(Inc))-We*T;
           Alfa:=Alfa+Gamma;
           IF Alfa<-Pi
           THEN Alfa:=2*Pi+Alfa;
           IF Alfa>Pi
           THEN Alfa:=2*Pi-Alfa;
           Beta:=InvTri(SinA,COS(Psi)*SIN(Inc),1);
           OutDat[Lon,I]:=Alfa*180/Pi;
           OutDat[Lat,I]:=Beta*180/Pi;
        END;
  END;
  Order;
  FOR I:=0 TO S DO
     WRITELN(Data,OutDat[Lon,I]:10:3,OutDat[Lat,I]:10:3);
END.
```

APPENDIX A2.

THE AVERAGE FIRST EXIT TIME (FET) IN A S-ALOHA PROCESS.

The Alogorithm is discussed in chapter 4, section 4.2.2.

```
      PROGRAM STABLE
C     THIS PROGRAM PRODUCES THE FET OF AN S-ALOHA SYSTEM USING
C     KLEINROCK'S TECHNIQUE. THE INPUT IS THE NUMBER OF SUBSLOTS IN A
C     FRAME. IT USES NAG ROUTINE C05ADF AND THEREFORE SHOULD BE
C     LINKED WITH NAG. THE TRAFFIC PROFILE PARAMETERS ARE ENTERED AS
C     CONSTANTS.
C     CAUTION!! THIS PROGRAM MAY RUN INTO PROBLEMS ASSOCIATED WITH
C     PROPAGATION OF ROUNDING OFF ERRORS!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!
C     THE USE OF SCALING (VARIABLE C) POSTPONES THIS PROBLEM TO M=20
C     ON THE CYBER  180-830.
C
      IMPLICIT DOUBLE PRECISION(A-Z)
      INTEGER I,K,L,IFAIL
      DOUBLE PRECISION A,R,S
      DOUBLE PRECISION PII,E(0:1000),F(0:1000),SPE,SPF,PK
      REAL FN,FMIN,FMAX,EPS,ETA,M,N,POP,TF
      COMMON A,S,C
      EXTERNAL FN
      PRINT*,'PLEASE ENTER THE NUMBER OF SUBSLOTS IN THE RSF (R*A)'
      READ(*,*)M
C     TRAFFIC PROFILE PARAMETERS
      POP=1E5
      S=1.3
      TF=0.1
C
      A=1.0/(3.0*M+(M+1.0)/2.0)
      S=S/(3600*M)*POP*TF
      EPS=0.5
      ETA=0
      IFAIL=0
      C=1.0E200
      FMIN=0.0
      FMAX=(1.0-A)*(1.0-S)/A
      CALL C05ADF(FMIN,FMAX,EPS,ETA,FN,N,IFAIL)
      PRINT*,'THE BACKLOG AT THE FIRST INTERCEPT IS ',N
      FMIN=FMAX
      FMAX=1.0E3
      CALL C05ADF(FMIN,FMAX,EPS,ETA,FN,N,IFAIL)
      PRINT*,'THE BACKLOG AT THE SECOND INTERCEPT IS ',N
      I=NINT(N)
      E(I)=1.0
      PK=P(I,I-1)
      PII=P(I,I)
      E(I-1)=(C-PII)/PK
      F(I)=0
```

```
      F(I-1)=-1.0*C/PK
      DO 100, K=I-1,1,-1
         SPE=E(K)
         SPF=F(K)-1.0
         DO 200, L=I,K,-1
            PK=P(K,L)
            SPE=SPE-PK*E(L)/C
            SPF=SPF-PK*F(L)/C
200      CONTINUE
         PK=P(K,K-1)
         E(K-1)=SPE*C/PK
         F(K-1)=SPF*C/PK
100   CONTINUE
      SPE=-E(0)
      SPF=F(0)-1.0
      DO 300, L=0,I
         PK=P(0,L)
         SPE=SPE+PK*E(L)/C
         SPF=SPF-PK*F(L)/C
300   CONTINUE
      TI=SPF/SPE
      T=E(0)*TI+F(0)
      PRINT*,'THE FET VALUE IS ',REAL(T)
      STOP
      END
C
      DOUBLE PRECISION FUNCTION P(I,J)
      IMPLICIT DOUBLE PRECISION(A-Z)
      INTEGER I,J
      COMMON A,S,C
      IF (J.LE.(I-2)) THEN
         P=0
      ELSE
         IF (J.EQ.(I-1)) THEN
            P=I*A*((1.0-A)**(I-1))*DEXP(-S)*C
         ELSE
            IF (J.EQ.I) THEN
               X=((1.0-A)**I)*S*DEXP(-S)
               P=(X+(1.0-I*A*((1.0-A)**(I-1)))*DEXP(-S))*C
            ELSE
               IF (J.EQ.(I+1)) THEN
                  P=C*S*DEXP(-S)*(1.0-((1.0-A)**I))
               ELSE
                  IF (J.GE.(I+2)) THEN
                     X=DEXP(-S)*C
                     DO 100, K=1,J-I
                        X=X*S/K
100                  CONTINUE
                     P=X
                  ENDIF
               ENDIF
            ENDIF
         ENDIF
      ENDIF
      RETURN
```

```
      END
C
      REAL FUNCTION FN(X)
      REAL FN,X
      DOUBLE PRECISION A,S,C
      COMMON A,S,C
      FN=DEXP(-S)*(((1.0-A)**X)*S+X*A*((1.0-A)**(X-1)))-S
      RETURN
      END
```

APPENDIX A3.

STEADY STATE SIMULATION OF A PROTOCOL WITH S-ALOHA.

(Discussed in chapter 6, section 6.4.)

```
PROGRAM CERS5 (INPUT/,OUTPUT,Data);
CONST
   BufMax = 1000;
   IndexLimit = 30;
   NrOfSteps = 10;
   SystemSpec = TRUE;
   RunData = FALSE;
TYPE
   IndexRange = -1..IndexLimit;
   ProbArr = ARRAY [IndexRange] OF REAL;
VAR
   BufAve,BufLen,BusyAve
   ,F,I,LMCounter,MesLen,N,NrOfStx
   ,NrOfSub,NoOfStx,K,KA,KB,QInput
   ,R,RSFCounter,SCMCounter,Succ,XTimes : INTEGER;
   AlohaRate,C,CT,Load,LMRate
   ,PropDel,R1,SCMRate,TF : REAL;
   Frame : ARRAY [1..2000] OF INTEGER;
   P,Q,S,U : ProbArr;
   Data : TEXT;

FUNCTION Random : REAL; EXTERN;

FUNCTION Ceil(X:REAL) : INTEGER;
   BEGIN
      IF X = TRUNC(X) THEN
         Ceil := TRUNC(X)
      ELSE
         Ceil := TRUNC(X + 1);
      END;

   BEGIN
      NegExp:=-Theta*LN(Random);
   END;

PROCEDURE RSF(R:INTEGER; Gamma,SCMProb:REAL;
   VAR C,CT:REAL; VAR K,KA,KB:INTEGER);
   CONST
      Trunk = 0.90;
   VAR
      KAB,K1,K2,M,N : INTEGER;
```

```
   RLimit,SLimit : IndexRange;
   G,PC,PG,PI,V : REAL;

FUNCTION Prob(N,I:INTEGER; A:REAL) : REAL; (* RSF *)
   VAR
      J : INTEGER;
      P : REAL;

   BEGIN
      P:=1;
      FOR J:=(N-I+1) TO N DO
         P:=P*J;
      IF I>1
      THEN
         FOR J:=2 TO I DO
            P:=P/J;
      Prob:=P*EXP(I*LN(1-EXP(-A)))*EXP((-A)*(N-I));
   END;

FUNCTION ProbQ (I : INTEGER) : REAL; (* RSF *)
   (* Calculation of eq 3 *)
   VAR
      Q,QA,QB : REAL;
      J : INTEGER;

   BEGIN
      Q:=0;
      FOR J:=0 TO I DO
         IF (KA>=J) AND (KB>=(I-J))
         THEN
            BEGIN
               QA:=Prob(KA,J,V);                        (* EQ 6.5 *)
               QB:=Prob(KB,I-J,V);                      (* EQ 6.6 *)
                Q:=Q+QA*QB;
            END;
      ProbQ := Q;
   END;

FUNCTION Search (* RSF *)
   (K : REAL; Y : ProbArr; Limit : IndexRange) : INTEGER;
   VAR
      J : INTEGER;

   BEGIN
      J:=0;
      WHILE (K>Y[J]) AND (J<Limit) DO
         J:=J+1;
      Search:=J;
   END;

PROCEDURE GenProbs; (* RSF *)
   VAR
      REnough,Senough : BOOLEAN;
      QK : INTEGER;
      I : IndexRange;
```

```
  BEGIN
     QK:=NrOfStx-K;
     FOR I:=0 TO 1 DO
        BEGIN
           P[I]:=Prob(QK,I,Gamma);
           Q[I]:=ProbQ(I);
        END;
     PI:=P[0]*Q[0];                                       (* EQ 6.7 *)
     PG:=P[1]*Q[0]+P[0]*Q[1];                             (* EQ 6.8 *)
     PC:=1-PI-PG;                                         (* EQ 6.9 *)
     S[0]:=Q[0]*(1-P[0]-P[1])/PC;                        (* EQ 6.11a *)
     U[0]:=P[0]*(1-Q[0]-Q[1])/PC;
     S[1]:=S[0]+Q[1]*(1-P[0])/PC;                        (* EQ 6.11b *)
     U[1]:=U[0]+P[1]*(1-Q[0])/PC;
     REnough:=FALSE;
     Senough:=FALSE;
     RLimit:=IndexLimit;
     SLimit:=IndexLimit;
     I:=1;
     WHILE (I<IndexLimit) DO
        BEGIN
           I:=I+1;
           IF NOT Senough
           THEN
              BEGIN
                 Q[I]:=ProbQ(I);
                 S[I]:=S[I-1]+Q[I]/PC;                   (* EQ 6.11c *)
                 IF (S[I]>Trunk)
                 THEN
                    BEGIN
                       Senough:=TRUE;
                       SLimit:=I;
                    END;
              END;
           IF NOT REnough
           THEN
              BEGIN
                 P[I]:=Prob(QK,I,Gamma);
                 U[I]:=U[I-1]+P[I]/PC;
                 IF (U[I]>Trunk)
                 THEN
                    BEGIN
                       REnough:=TRUE;
                       RLimit:=I;
                    END;
              END;
        END;
  END;

BEGIN
  V:=1/(Ceil(PropDel/TF)*R*NrOfSub+R*NrOfSub/2);
  QInput:=0;
  FOR N:=1 TO R DO
     BEGIN
```

```
RSFCounter:=RSFCounter+1;
FOR M:=1 TO NrOfSub DO
   BEGIN
      GenProbs;
      R1:=Random;
      IF R1<PG
      THEN
         BEGIN
            Succ:=Succ+1;
            IF Random<SCMProb
            THEN
               SCMCounter:=SCMCounter+1
            ELSE
            IF BufLen<BufMax
            THEN
               BEGIN
                  BufLen:=BufLen+1;
                  QInput:=QInput+1;
               END
            ELSE
               WRITELN('Buffer overflow');
            IF K>0
            THEN
               BEGIN
                  G:=P[0]*Q[1]/PG;           (* EQ 6.10 *)
                  IF Random<G
                  THEN
                     BEGIN
                        IF Random<(KA/K)
                        THEN
                           BEGIN
                              KA:=KA-1;
                              CT:=CT-1;
                           END
                        ELSE
                           BEGIN
                              KB:=KB-1;
                              CT:=CT-C;
                           END;
                     END;
               END;
         END
      ELSE
      IF R1<(PG+PC)
      THEN
         BEGIN
            K2:=Search(Random,U,RLimit);
            (* Number of new arrivals
            involved in a collision *)
            Load:=Load+K2;
            IF K>0
            THEN
               BEGIN
                  K1:=Search(Random,S,SLimit);
                  (* Number of retransmissions
```

```
                              involved in a collision *)
                              KAB:=ROUND(KA/K*K1);
                              Load:=Load+K1;
                           END
                        ELSE
                           BEGIN
                              K1:=0;
                              KAB:=0;
                           END;
                        KA:=KA-KAB+K2;
                        KB:=KB+KAB;
                        CT:=CT+K1+K2;
                     END;
                  IF KB<>0
                  THEN
                     C:=(CT-KA)/KB;
                  K:=KA+KB;
               END;
         END;
         LMCounter:=LMCounter+QInput;
   END; (* RSF *)

PROCEDURE Adjust;
   VAR
      EBusy,Cnt,J : INTEGER;

   BEGIN
      EBusy:=ROUND(NrOfStx*LMRate*MesLen);
      IF EBusy>(F-R)
      THEN
         BEGIN
            EBusy:=F-R;
            WRITELN('SYSTEM APPROACHING SATURATION');
         END;
      Cnt:=0;
      FOR J:=1 TO (F-R) DO
         IF Frame[J]<>0
         THEN
            Cnt:=Cnt+1;
      Cnt:=Ebusy-Cnt;
      J:=0;
      WHILE Cnt>0 DO
         BEGIN
            J:=J+1;
            IF Frame[J]=0
            THEN
               BEGIN
                  Frame[J]:=ROUND(NegExp(MesLen));
                  Cnt:=Cnt-1;
               END;
            END;
   END;

PROCEDURE ISF (InRate,M,MesLen:INTEGER);
   VAR
```

```
      Busy,J : INTEGER;

   BEGIN
      Busy:=0;
      FOR J:=1 TO M DO
         BEGIN
            IF Frame[J]>0
            THEN
               BEGIN
                  Frame[J]:=Frame[J]-1;
                  Busy:=Busy+1;
               END
            ELSE
            IF BufLen>0
            THEN
               BEGIN
                  BufLen:=BufLen-1;
                  Frame[J]:=ROUND(NegExp(MesLen));
                  Busy:=Busy+1;
               END;
         END;
      BufAve:=BufAve+BufLen;
      BusyAve:=BusyAve+Busy;
   END; (* ISF *)

PROCEDURE Report (Initial:BOOLEAN);
   VAR
      I : INTEGER;
      Delay,GT,QDelay,SA,SL,SR,SS : REAL;
   BEGIN
      IF Initial
      THEN
         BEGIN
            WRITE(Data,'SHORT MESSAGE GENERATION ');
            WRITELN(Data,'RATE (PER HR) :',SCMRate:6:3);
            WRITE(Data,' LonG MESSAGE GENERATION ');
            WRITELN(Data,'RATE (PER HR) :',LMRate:6:3);
            WRITE(Data,'      MEAN MESSAGE LENGTH ');
            WRITELN(Data,'(IN SECONDS) :',MesLen:6);
            WRITE(Data,'   PROPAGATION DELAY ');
            WRITELN(Data,'(IN MILLISECONDS) :',PropDel:6:1);
            WRITE(Data,'      FRAME DURATION ');
            WRITELN(Data,'(IN MILLISECONDS) :',TF:6:1);
            WRITE(Data,'            NUMBER OF ');
            WRITELN(Data,'SLOTS IN A FRAME :',F:6);
            WRITE(Data,'              NUMBER OF');
            WRITELN(Data,' SLOTS IN A RSF :',R:6);
            WRITE(Data,'      NUMBER OF SUBSLOTS ');
            WRITELN(Data,'IN A RSF SLOT :',NrOfSub:6);
         END
      ELSE
         BEGIN
            SS:=SCMCounter/(RSFCounter*NrOfSub);
            SR:=LMCounter/(RSFCounter*NrOfSub);
            SA:=SS+SR;
```

```
          IF XTimes*F>RSFCounter
          THEN
             SL:=BusyAve/(XTimes*F-RSFCounter)
          ELSE
             SL:=0;
          GT:=NrOfStx*AlohaRate/R;
          I:=Ceil(2*PropDel/TF+R/F);
          IF (Succ <> 0) AND (SCMCounter <> 0) THEN
             Delay:=I*TF*(Load/Succ+1)
          ELSE
             Delay:=0;
          IF LMCounter<>0
          THEN
             QDelay:=(I*(Load/Succ+1)
             +Ceil(BufAve/LMCounter)   (* Little's result *)
             +Ceil(PropDel/TF)
             +0.5*(1-R/F))*TF
          ELSE
             QDelay:=Delay;
          Load:=(Load+Succ)/(RSFCounter*NrOfSub);
          WRITE(Data,NrOfStx:7,GT:7:4,Delay:12);
          WRITE(Data,SA:7:4,SR:7:4,SS:7:4,Load:7:4);
          WRITELN(Data,SL:7:4,QDelay:12);
          BufAve:=0;
          BusyAve:=0;
          Load:=0;
          LMCounter:=0;
          RSFCounter:=0;
          SCMCounter:=0;
          Succ:=0;
       END;
   END;

PROCEDURE ReadData;
   BEGIN
      WRITE('   NUMBER OF STATIONS TO BE CONSIDERED :');
      READ(NoOfStx);WRITELN;
      WRITE('SHORT MESSAGE GENERATION RATE (PER HR) :');
      READ(SCMRate);WRITELN;
      WRITE(' LonG MESSAGE GENERATION RATE (PER HR) :');
      READ(LMRate);WRITELN;
      WRITE('      MEAN MESSAGE LENGTH (IN SECONDS) :');
      READ(MesLen);WRITELN;
      WRITE('   PROPAGATION DELAY (IN MILLISECONDS) :');
      READ(PropDel);WRITELN;
      WRITE('      FRAME DURATION (IN MILLISECONDS) :');
      READ(TF);WRITELN;
      WRITE('           .   NUMBER OF SLOTS IN A FRAME :');
      READ(F);WRITELN;
      WRITE('                 NUMBER OF SLOTS IN A RSF :');
      READ(R);WRITELN;
      WRITE('      NUMBER OF SUBSLOTS IN A RSF SLOT :');
      READ(NrOfSub);WRITELN;
      WRITE(' NUMBER OF STEPS IN THE SIMULATION RUN :');
      WRITELN(NrOfSteps:6);
```

```
      WRITE('                    NUMBER OF RUNS PER STEP :');
      READ(XTimes);WRITELN;
   END;

PROCEDURE Initialise;
   VAR
      N : INTEGER;

   BEGIN
      U[-1]:=0;                S[-1]:=0;
      Succ:=0;                 Load:=0;
      KA:=01;                  KB:=01;
      K:=KA+KB;                C:=00;
      CT:=KA+KB*C;             BufLen:=0;
      LMCounter:=0;            QInput:=0;
      SCMCounter:=0;           RSFCounter:=0;
      BufAve:=0;                BusyAve:=0;
      TF:=TF*1E-3;
      FOR N:=1 TO F DO
         Frame[N]:=0;
      PropDel:=PropDel*1E-3;
      MesLen:=ROUND(MesLen/TF);
      SCMRate:=SCMRate*TF/3600;
      LMRate:=LMRate*TF/3600;
      AlohaRate:=(SCMRate+LMRate)/NrOfSub;
   END;

BEGIN
   REWRITE(Data);
   ReadData;
   Report(SystemSpec);
   Initialise;
   FOR I:=1 TO NrOfSteps DO
      BEGIN
         WRITELN(I);
         NrOfStx:=ROUND(-I/NrOfSteps*NoOfStx);
         Adjust;
         FOR N:=1 TO XTimes DO
            BEGIN
               RSF(R,AlohaRate/R,SCMRate/(SCMRate+LMRate)
                  ,C,CT,K,KA,KB);
               ISF(QInput,F-R,MesLen);
            END;
         Report(RunData);
      END;
END.
```

APPENDIX A4.

STABILITY SIMULATION FOR THE P-ALOHA RESERVATION SUBFRAME.

(Discussed in chapter 6, section 6.5.)

```
PROGRAM Stability (INPUT/,OUTPUT,Data,Note);

CONST
   Max = 1000;
   Tf = 0.1;
   Rate = 1.3;
   Pop = 1E5;
   A = 8;
   R = 3;
TYPE
   MessType = (Try1,TryN);
   EpochType =
   RECORD
      EventTime:REAL;
      Message:MessType;
   END;
   EpochArray = ARRAY[0..100] OF EpochType;
VAR
   Corrupted : BOOLEAN;
   Arrivals,Back,BackMax,N,Nc,New,Re,Seed1,Seed2,X : 0..1000;
   K,Succ,Time : INTEGER;
   Jump,P,S,Y : REAL;
   Epoch : EpochArray;
   Note,Data : TEXT;

FUNCTION Random : REAL; EXTERN;

PROCEDURE SetRandom (Seed1,Seed2 : INTEGER); EXTERN;

FUNCTION Poisson (Mean : REAL) : INTEGER;
   VAR
      X : INTEGER;
      A,S : REAL;

   BEGIN
      X := 0;
      A := EXP(-Mean);
      S := 1;
      S := S*Ran;
      WHILE S-A>=0 DO
         BEGIN
```

```
         X := X+1;
         S := S*Ran;
      END;
   Poisson := X;
END;

FUNCTION Bernoulli (P:REAL; Trials:INTEGER):INTEGER;
   VAR
      N,R : INTEGER;
      Nx,X : REAL;
      ProbArr : ARRAY[-1..100] OF REAL;

   BEGIN
      IF Trials <> 0
      THEN
         BEGIN
            FOR R := 0 TO Trials-1 DO
               BEGIN
                  ProbArr[-1] := 0;
                  Nx := EXP(LN(P)*R)*EXP(LN(1-P)*(Trials-R));
                  FOR N := Trials DOWNTO Trials-R+1 DO
                     Nx := Nx*N;
                  FOR N := 2 TO R DO
                     Nx := Nx/N;
                  ProbArr[R] := Nx+ProbArr[R-1];
               END;
            ProbArr[Trials] := 1;
            X := Ran;
            N := 0;
            WHILE X>ProbArr[N] DO
               N := N+1;
            Bernoulli := N;
         END
      ELSE
         Bernoulli := 0;
   END;

PROCEDURE ArrivTimes(New,RE,Arrivals:INTEGER; VAR Epoch:EpochArray);
   VAR
      I,J,N : INTEGER;
      X : REAL;
      Try : EpochType;

   BEGIN
      FOR N := 1 TO New DO
         Epoch[N].Message := Try1;
      FOR N := New+1 TO Arrivals DO
         Epoch[N].Message := TryN;
      FOR N := 1 TO Arrivals DO
         Epoch[N].EventTime := Ran*(R*A-1);
      FOR N := 1 TO Arrivals DO
         FOR J := N+1 TO Arrivals DO
            IF Epoch[N].EventTime>Epoch[J].EventTime
            THEN
               BEGIN
```

```
                 Try := Epoch[J];
                 Epoch[J] := Epoch[N];
                 Epoch[N] := Try;
              END;
        Epoch[Arrivals+1].EventTime := R*A;
     END;

  BEGIN
     RESET(Data);
     RESET(Note);
     P := 1/(3.5);
     S := Rate/(3600)*Tf*Pop;
     WHILE (R<>X) DO
        READLN(Data,X,Y,Y);
     Nc := ROUND(Y);
     WRITELN(R,Nc);
     READLN(Note,Time,Back,BackMax,Time,Succ,Seed1,Seed2);
     SetRandom(Seed1,Seed2);
     Seed1 := Seed1+1;
     Seed2 := Seed2+1;
     K := 0;
     Corrupted := FALSE;
     WHILE (Back<>Nc) AND (K<Max) DO
        BEGIN
           K := K+1;
           Time := Time+1;
           IF Time MOD 5000 = 0
           THEN
              WRITELN(Time/R*Tf/3600);
           New := Poisson(S);
           Re := Bernoulli(P,Back);
           Arrivals := New+Re;
           ArrivTimes(New,Re,Arrivals,Epoch);
           FOR N := 1 TO Arrivals DO
              BEGIN
                 Jump := Epoch[N+1].EventTime-Epoch[N].EventTime;
                 IF Jump > 1
                 THEN
                    BEGIN
                       IF NOT Corrupted
                       THEN
                          BEGIN
                             Succ := Succ+1;
                             IF Epoch[N].Message=TryN
                             THEN
                                Back := Back-1;
                          END
                       ELSE
                          Corrupted := FALSE;
                    END
                 ELSE  (* Jump < 1 *)
                    BEGIN
                       Corrupted := TRUE;
                       IF Epoch[N].Message=Try1
                       THEN
```

```
                        Back := Back+1;
                  END;
            END;
         IF Back>BackMax
         THEN
            BackMax := Back;
      END;
   REWRITE(Note);
   WRITELN(Time,Back,BackMax,Time,Succ,Seed1,Seed2);
   WRITELN(Note,Time,Back,BackMax,Time,Succ,Seed1,Seed2);
   WRITELN(Time,Succ/Time);
END.
```

APPENDIX A5.

EPHEMERIS OF A SATELLITE IN THE MOLNIYA ORBIT DURING THE ACTIVE PERIOD

Time relative to apogee (h)	Sub-satellite point longitude (°)	latitude (°)	Altitude (km)
0.0	3.000	63.400	39,000
0.2	3.138	63.366	38,965
0.4	3.271	63.262	38,859
0.6	3.393	63.089	38,683
0.8	3.501	62.846	38,116
1.0	3.589	62.531	37,723
1.2	3.653	62.142	37,000
1.4	3.691	61.679	39,258
1.6	3.700	61.137	36,717
1.8	3.680	60.513	36,100
2.0	3.629	59.803	35,404
2.2	3.549	59.000	34,628
2.4	3.441	58.099	33,770
2.6	3.308	57.089	31,793
2.8	3.156	55.961	30,669
3.0	2.898	54.700	29,448
3.2	2.814	53.288	29,448
3.4	2.643	51.703	28,125
3.6	2.488	49.915	26,696
3.8	2.364	47.885	25,153
4.0	2.294	45.559	23,489

REFERENCES

ABRA 70 N. Abramson, "The ALOHA System - Another Alternative for Computer Communications", Proc. of the Fall Joint Computer Conference, AFIPS Press, 1970.

ATAT 77 American Telephone and Telegraph Company, "Engineering Economy", McGraw-Hill Book Company, 1977.

BALL 80 A.H. Ballard, "Rosette Constellation of Earth Satellites", IEEE Trans. on Aerospace and Electronic Systems, Vol. AES-16, No. 5, Sept. 1980.

BERM 70 A.L. Berman and C.E. Mahle, "Nonlinear Phase Shift in Travelling- Wave Tubes as Applied to Multiple Access Communications Satellites", IEEE Trans. on Communications Tech., Vol. COM-18, No. 1, Feb. 1970.

BERR 84 C.T. Berretta, "Extension of Mobile Satellite Commmunications to the Polar Regions", Space Communications and Broadcast, Vol. 2, Part 1, Mar. 1984.

BHAR 81 V.K. Bhargava, D. Haccoun, R. Matyas and P.P. Nuspl, "Digital Communications by Satellites", John Wiley & Sons, 1981.

BIND 75a R. Binder, N. Abramson, F. Kuo, A. Okinaka and D. Wax, "ALOHA Packet Broadcasting - A retrospect", National Computer Conference, 1975.

BIND 75b R. Binder, "A Dynamic Packet Switching System for Satellite Broadcast Channels",Proc. of the International Conference on Communications, 1975.

BREA 85 R. Breadley and S. Myers, "Principles of Corporate Finance" McGraw-Hill International Book Company, 1985.

BSO 86 Business Statistics Office, "Price Index Numbers for Current Cost Accounting, Monthly Supplement", Her Majesty's Stationary Office, MM17, Mar. 1986.

CAPE 79 J.I. Capetanakis, "Generalised TDMA: The Multi-Accessing Tree Protocol", IEEE Trans. on Communications, Vol. COM-27, No. 10, Oct. 1979.

CARL 75 A.B. Carleial and M.E. Hellman, "Bistable Behaviour of ALOHA-Type Systems", IEEE Trans. on Communications, Vol. COM-23, No. 4, Apr. 1975.

CCIR 14 CCIR, "Mean One-Way Propagation Time", "Fixed-Satellite Service", Vol. IV, Recommendation G.114, XVth Plenary Assembly, Geneva, 1982.

CCIR 18 CCIR, "Availability of Communications Circuits in the Maritime Mobile- Satellite Service", Mobile Services, Vol. VIII, Report 918, XVth Plenary Assembly, Geneva, 1982.

CCIR 24 CCIR, "Field Test of Digital Facsimile for Maritime Mobile-Satellite Service", Mobile Services, Vol. VIII, Report 924, XVth Plenary Assembly, Geneva, 1982.

CCIR 73 CCIR, "Characteristics of Ground Receiving Equipment for Broadcasting-Satellite Systems", "Broadcasting - Satellite Service (Sound and Television)", Vol. X and XI - Part 2, XVth Plenary Assembly, Geneva, 1982.

CCST 80 Consultive Committee on Satellite Telecommunications, "British Telecom view of Satellite Communication Markets", Document No.: 482, 10 Oct. 1980.

CCST 85 Consultive Committee on Satellite Telecommunications, "SPADE System Specification", INTELSAT Earth Station Standards, (IESS), Document No.: IESS-304, CCST 562, Sept. 1985.

COAK 84 F.P. Coakley, "On-Board Processor(s) for CERS", IEE Colloquium on ´CERS - Communication Engineering Research Satellite´, Digest No. 1984/39.

COST 59 J.P. Costas, "Poisson, Shannon, and the Radio Amateur", Proc. IRE, Vol. 47, Dec. 1959.

CRAI 83 K.H. Craig, "Radiation Dose Estimates for the CERS Orbit", Internal Menorandum, Ref. No. CERS/WP/5, Rutherford Appleton Laboratories, 10 May 1983.

CROW 73 W. Crowther, R. Rettberg, D. Walden, S. Ornstein and F. Heart, "A System for Broadcast Communication: Reservation-Aloha", Proc. Sixth Hawaii International Conference on System Science, 1973.

CSO 86 Central Statistical Office, "Economic Trends, Annual Supplement", Her Majesty´s Stationary Office, 1986.

DRAI 85 J.E. Draim, "Three- and Four-Satellite Continuous-Coverage Constellation", Journal on Guidance, Vol. 8, No. 6, Nov.-Dec. 1985.

DASI 84 Da Silva Curiel, "Operational Experience with the INMARSAT System", IEEE International Conference on Communications 1984.

DOND 83 P. Dondl, "LOOPUS Opens a New Dimension in Satellite Communications", NTZ Archiv (Germany), Vol. 5, No. 12, Dec. 1983.

DUFF 80 W.G. Duff, "A Handbook on Mobile Communications", Don White Consultants, Inc., Virginia, 1980.

ELYA 67 P.E. El´Yasberg, ´Introduction to the Theory of Flight of Artificial Earth Satellites", translated from Russian, Israel Program for Scientific Translations, Jerusalem, 1967.

ESA 86 ESA, "Land-Mobile Services by Satellite", Proc. of an ESA Workshop, ESA/ESTEC, Doc. No. SP-259, The Netherlands, 3-4 June 1986.

FEHE 83a K. Feher, "Digital Communications: Satellite/Earth Station Engineering", Praitice-Hall Inc. Englewood-Cliffs, N.J., 1983.

FEHE 83b K. Feher, "SCPC Satcom Systems for Voice and Data Services", Telecommunications, Vol. 17, No. 6, June 1983.

FENE 85 H.T. Fenech, M.D. Button, E. Aboje and J.G. Gardiner, "Satellite Augmentation of Terrestrial Cellular Radio Systems", Third International Conference on Land Mobile Radio, IERE, University of Cambridge, Dec. 1985.

FENE 86 H.T. Fenech and J.G. Gardiner, "Link Margin Consideration for T-SAT", T-SAT Technical Note, Ref. No. TN 1003 BRA.

FERG 77 M.J. Ferguson, "An Approximate Analysis of Delay for Fixed ang Variable Length Packets in an Unslotted ALOHA Channel", IEEE Trans. on Communications, Vol. COM-25, No. 7, July 1977.

FUEN 73 J.C. Fuenzalida, O. Shimbo and W.L. Cook, "Time-Domain Analysis of Intermodulation Effects Caused by Nonlinear Amplifiers", COMSAT Tech. Rep. Vol. 3, Spring 1973.

HADF 74 B.M. Hadfield, "Satellite-System Cost Estimation", IEEE Trans. on Communications, Vol. COM-22, No. 10, Oct. 1974.

HAYE 86 J.F. Hayes, "Modeling and Analysis of Computer Communications Networks", Plenum Press, New York, 1986.

IIDA 80 T. Iida, "Satellite Systems for Land Mobile Communications", Journal of the Radio Research Laboratories, Japan, Vol. 27, Nos. 122/123, Mar./July 1980.

IMF 84 International Monetary Fund, "International Financial Statistics", 1984 Yearbook.

IMF 86 International Monetary Fund, "International Financial Statistics", Vol. XXXIX, No. 2, Feb. 1986.

INTE 80 "INTELSAT TDMA/DSI System Specification (TDMA/DSI Traffic Terminals)", INTELSAT Doc. No. BG-42-65E B/6/80, INTELSAT, Washinton, D.C., 26 June 1980.

INTE 85 "SPADE System Specification", INTELSAT Earth Station Standards, Doc. IESS - 304, CCST Library 562, Sept. 1985.

ISLE 74 W.C. Isley and K.I. Duck, ´Propulsion Requirements for Communication Satellites,´ in P.L. Bargellini, Ed., Communicatons Satellite Technology, Cambridge, MA: MIT Press 1974.

JACO 78 I.M. Jacobs, R. Binder and E.V. Hoversten, "General Purpose Packet Satellite Networks", Proc. IEEE, Vol. 60, Nov. 1978.

JAME 81 J.R. James, P.S. Hall and C. Wood, "Mircostrip Antenna : Theory and Design", IEE, 1981.

JONG 86 A. Jongejans, A. Dissanayake, N. Hart, H. Haugli, C. Loisy and R. Rogard, "PROSAT Phase 1 Report", European Space Research and Technology Centre, Noordwijk, The Netherlands, Report No.: ESA STR-216, May 1986.

KLEI 75a L. Kleinrock, "Packet Switching in a Multiaccess Broadcast Channel: Performance Evaluation", IEEE Trans. on Communications, Vol. COM-23, No. 4, Apr. 1975.

KLEI 75b L. Kleinrock and F.A. Tobagi, "Packet Switching in Radio Channels, Part 1, : Carrier Sense Multiple Access Modes and their Throughput-Delay Characteristics", IEEE Trans. on Communications, Vol. COM-23, No. 12, Dec. 1975.

KLEI 76 L. Kleinrock, "Queuing Theory, Volume II: Computer Applications", John Wiley and Sons, 1976.

LAM 79 S.S. Lam, "Satellite Packet Communication - Multiple Access Protocols and Performance", IEEE Trans. on Communications, Vol. COM-27, No. 10, Oct. 1979.

LAM 83 S.S. Lam, "Multiple Access Protocols", Tutorial Principles of Communication and Network Protocols, editor Lam, S.S., IEEE Computer Society Press 1984.

LAW 85 P.A. Law, "The Royal Navy's Shipborne Satcom Terminal (SCOT) Enhancement Programme", Colloqium on "Satellite Communications for the Mobile User", IEE, Digest No.: 1985/43, 25 Apr. 1985.

LI 84 O.K.V. Li, and Y. Tsun-Yee, "Adaptive Mobile Access Protocol (AMAP) for the Message Service of a Land Mobile Satellite Experiment (MSAT-X)", IEEE Journal of Selected Areas in Communications, Vol. SAC-2, No. 4, July 1984.

LIPK 77 D.W. Lipke, D.W. Swearingen, J.F. Parker, E.E. Steinbrecher, T.O. Calvit and H. Dodel, " MARISAT - A Maritime Satellite Communications System", Comsat Technical Review, Vol. 7, No. 2, Fall 1977.

MERR 74 A.J. Merrett and A. Sykes, "The Finance and Analysis of Capital Projects", Longman Group Limited, 1974

MITT 84 S.S. Mittra, "Discrete System Simulation Concepts", Simulation, Sept. 1984.

MSAT 85 "FCC Releases Notice of Proposed Rulemaking for Mobile Satellite Service", MSAT-X Quarterly, Jet Propulsion Laboratory, No. 2, Feb. 1985.

OREI 84 P.J.P. O'Reilly and J.L. Hammond Jr., "An Efficient Simulation Technique for Performance Studies of Local Networks", IEEE Journal on Selected Areas in Communications, Vol. SAC-2, No. 1, Jan. 1984.

PARZ 62 E. Parzen, "Stochastic Processes", Holden-Day Inc., San Fansisco, 1962.

PENT 81 M. Pent, M. Elia, L. Lo Prest and V. Castellani, "Performance Analysis of Uplinks for Regenerative Satellite Repeaters", ESA Report ESA CR (P) 1429, Jan. 1981.

PETE 81 E. Petersen, "Soft Errors Due to Protons in the Radiation Belt", IEEE Trans on Nuc. Sci., Vol. NS-28, No. 6, Dec. 1981.

PHIL 85 R.O. Phillips, "INMARSAT Review", Colloqium on "Satellite Communications for the Mobile User", IEE, Digest No.: 1985/43, 25 Apr. 1985.

PRIT 79 W.L. Pritchard, "Economics of Satellite Communications Systems", XXX Congress, International Astronautical Federation, Munich, F.R.G., 17-22 Sept. 1979.

PRIT 84 W.L. Pritchard, "Estimating the Mass and Power of Communications Satellites", International Journal of Satellite Communications, Vol. 2, No. 2, Apr.-June 1984.

PRIT 86 W.L. Pritchard, "Satellite Communication Systems Engineering", Prentice-Hall Inc., 1986.

RADI 82 "Request for Proposals for a System Definition Study for Mobile Data Communications", Radio Studies Group, Research and Development Section, Directorate of Telecommunications, Home Office, DT/82 49/122/1, Jan. 1982.

RAPP 84 S.S. Rappaport and D.M. Grieco, "Spread-Spectrum Signal Acquisition: Methods and Technology", IEEE Communication Magazine, Vol. 22, No. 6, June 1984.

ROBE 72 L.G. Roberts, "ALOHA Packet System With and Without Slots and Capture", ARPANET Satellite System Note 8, NIC Document No. 11290, Stanford Research Institute, June 26, 1972.

ROBE 73 L.G. Roberts, "Dynamic Allocation of Satellite Capacity Through Packet Reservation", AFIPS 1973, National Computer Conference Proc., Vol. 42, 1973.

SALE 81 A.A.M. Saleh, "Frequency-Independent and Frequency-Dependent Nonlinear Models of TWT Amplifiers", IEEE Trans. on Communications, Vol. Com-29, No. 11, Nov. 1981.

SAND 84 W.A. Sandrin, "Land-Mobile Satellite Start-Up Systems", Comsat Technical Review, Vol. 14, No. 1, Spring 1984.

SANT 80 D. Sant, "Throughput of Unslotted ALOHA Channels with Arbitrary Packet Interarrival Time Distribution", IEEE Trans. on Communications, Vol. COM-28, No. 8, Aug. 1980.

SHAN 75 R.E. Shannon, "System Simulation: the Art and Science", Prentice-Hall Inc., 1975.

SHIM 71 O. Shimbo, "Effects of Intermodulation AM-PM Conversion and Additive Noise in a Multicarrier TWT System", Proc. IEEE, Vol. 59, No. 2, Feb. 1971.

SHIM 84 M. Shimada, T. Iida, K. Iwasaki and S. Shimoseko, "Millimeter Wave Personal Radio Using an Advanced Communications Satellite", IEEE Communications Magazine, Vol. 22, No. 9, Sept. 1984.

STAL 85 W. Stallings, "Data and Computer Communications", Macmillan Publishing Co., 1985.

TANE 81 A.S. Tanenbaum, "Computer Networks", Prentice-Hall International, 1981.

TOBA 84 F.A. Tobagi, R. Binder, B. Leiner, "Packet Radio and Satellite Networks", IEEE Communication Magazine, Vol. 22, No. 11, Nov. 1984.

TOCH 65 K.D. Tocher,"A Review of Simulation Languages", Operational Research Quarterly, 16(2), pp189-217.

TSYB 84 B.S. Tsybakov and V.L. Bakirov, "Stability of Asynchronous ALOHA System", Problems of Information Transmission, USA, Vol. 20, No. 1, 1984, translated from Problemy Peredachi Informatsii, Vol. 20, No. 1, Jan.-Mar. 1984.

ULTA 78 W.F. Ultaut, "Spread Spectrum, Principles and Possible Applications to Spectrum Utilization and Allocations", ITU Telecomunication Journal, Vol. 45, Jan. 1978.

VITE 85 A.J. Viterbi, "When Not to Spread Spectrum-Sequel", IEEE Communications Society Magazine, Vol. 23, No. 4, April 1985.

WALK 73 J.G. Walker, "Continuous Whole Earth Coverage by Circular Orbit Satellite", "Satellite Systems for Mobile Communications and Surveillance", IEE Conference 1973.

WALK 77 J.G. Walker, "Continuous Whole-Earth Coverage by Circular-Orbit Satellite Patterns", Royal Aircraft Establishment, Tech. Rep. 77044 (UDC 629.195:527), Mar. 1977.

WATS 83 P.A. Watson, J.G. Gardiner, and E.C. Bell, "Proposal for a Land Mobile Satellite Communications Experiment", Postgraduate School of Electrical and Electronic Engineering, University of Bradford, Report No. 347, Dec. 1983.

WEBE 83 W.J. Weber and F. Naderi, "NASA Mobile Satellite Experiment (MSAT-X)", Proc. of National Electronics Conference, Chicago, 1983.