Example solutions and hints to Revision questions

CHAPTER 1

1) What are the main components of a mobile satellite service? Outline the role of each component.

The main components of a mobile satellite system are illustrated in figure 1 below (Sketch a schematic including all major components).

Figure 1 (A sketch of figure 1.2)

Telecommunication services: The system offers communication services to a variety of mobile users within a predefined *service area*. Services include paging, telephony, data exchange, etc. The capacity and capability of an MSS is governed by the power and bandwidth constraints of the satellite's service link. Service area can span a country, a region consisting of a number of countries or continents, or the entire world.

Gateways: Users and service applications embedded within the fixed network are accessed through large fixed stations called *gateways* which carry aggregated volumes of traffic over the requisite terrestrial interfaces. The fixed stations are capable of high EIRP transmissions, comprise sensitive receivers and operate over stable radio links.

Radio frequency links: Radio links provide connectivity between a gateway and the mobile user population. The gateway-satellite radio link is called *feeder link* and the satellite-mobile user is known as *service link.* The gateways are sited to ensure unobstructed view of operational satellite(s) and thus feeder links are stable.

User terminals: The user population communicates with each other and with users in the fixed network over the servicing satellite on small portable units capable of supporting the desired data rate. User terminals' range from personal units servicing individuals to units mounted on ships, aircrafts and trucks capable of supporting several communication channels simultaneously. Signals received at mobiles fluctuate widely due to variations in a mobile's path profile and multi-path reception.

Space segment: Depending on the service- type and target area, the *space segment* consists of one or more geostationary or non-geostationary satellites. Inter-satellite links can bypass the need of terrestrial connectivity either and can be applied to geostationary or non-geostationary constellations. Telemetry and control ground stations, used for monitoring and controlling satellites, constitute a part of the space segment.

To reduce mobile terminal size and sensitivity and improve spatial frequency reuse, MSS satellites transmit high power in the service link and generate a large number of narrow spot beams. For example, existing L-band geostationary satellites can generate up to 12 kW DC power, with effective isotropic radiated power in excess of 67 dBW communicating via >300 spot beams.

A geostationary satellite remains almost fixed with respect to the Earth illuminating the service area with a *static* footprint that enables a simple network topology, whereas, for a non-geostationary earth orbit (NGEO) space segment network topology becomes complex due to non-stationary satellites and propagation conditions become harsher (except for a highly-elliptical orbit system), but because of a reduction in path length the propagation time reduces and handsets are able to transmit lower power in comparison to geostationary satellite.

A *satellite control center (SCC)* manages the functioning and maintenance of the satellite(s).

A *network control center (NCC)* controls the functioning of the network which includes provision of services to users, monitoring and control of traffic flow, radio resource allocation, etc. In a centralized radio resource management system, the network control center (NCC) manages the call set-up and release, radio resources and signaling. On receiving a call request, the radio resource manager assigns the desired radio resource for the duration of the call. In a distributed system the task is distributed to gateways.

The *Mobility management system* is responsible to manage the mobility of the user population. The home location register (HLR) and visitor location register (VLR) are responsible for mobility management to ensure that users can travel anywhere within the service area with seamless connectivity. The mobile switching center (MSC) manages call switching between the fixed network and the mobile network. A business management center (BMC) is responsible for billing, activations, customer support, etc.

A call in progress must be re-routed to maintain connectivity as the connected satellite or spot beam moves out of visibility necessitating *handover(s)* to a rising satellite or spot beam.

A call may be routed to the final destination through inter-satellite links, ground-satellite hops, or terrestrially. Figures 2(a) and 2(b) below respectively illustrate source-destination connectivity in a non-geostationary satellite system deploying a terrestrial backbone and inter-satellite links. For example, the Iridium system deploys inter-satellite links whereas the ICO system uses ground routing.

Figure 2(a) and 2(b) [Sketch of figures 1.3(a) and 1. 3(b) here]

A *business management center* supports commercial aspects that include billing, customer relations and management, traffic data analysis, strategic planning, etc.

An *operation management* entity oversees the operation of the entire network to ensure desired quality of service, compliance to network reliability goals, provide data for network planning, forecasting, etc.

2) Table 1.2 gives a comparison of various technical parameters of several MSS systems. Explain the reason for different sensitivity requirement of the gateways (specified as G/T; a higher value implies a more sensitive receiver) in view of the similarity in G/T of the user terminal; Compare the total space segment capacity of each system assuming that 3 regional geostationary systems are required for world-wide coverage. Comment on your results.

Reasons for different sensitivity requirement of gateways

This question can best be answered by comparing the link budget of the each system, a topic dealt in chapter 3. Here we provide a subjective analysis.

We observe that the gateway G/T range between 24 dB/K for a LEO system (Iridium) and 37 dB/K for a GEO system. The parameter G/T is the ratio of a receiver's antenna gain (G) to its system noise temperature (T) and hence represents the capability of a receiver to enhance the wanted signal over and above the inherent noise of the receiver - the receiver sensitivity therefore, increases as its G/T increases.

The user terminals are small in size transmitting low powers and hence the signals received at the gateway are low, necessitating a sensitive gateway receiver and hence the large G/T of the gateways. Given a user terminal EIRP, the parameters influencing the received signal quality at a gateway are:

- Path loss in uplink (Earth-space, i.e., mobile-satellite) and downlink (space-Earth, i.e., satellite-gateway)
- Spacecraft receiver G/T
- Spacecraft EIRP towards gateway

- Gateway G/T
- Link margins in both parts of the radio link
- Interference levels in both parts of the radio link
- Miscellaneous loses

In the absence of radio link parameters we highlight possible reasons for the G/T differences.

Considering the satellite-gateway path, we observe that the range and hence the path loss for a geostationary satellite is the highest – compare ~780 km altitude of Iridium satellites which operate in low earth orbit (LEO) with 35786 km of a GEO satellite that equates to a one-way difference of about 33.2 dB (20 log 35786/780) in path loss at the sub-satellite point (taken for simplicity). Assuming similar order of EIRP from satellite the signal level from a GEO satellite would be 33.2 dB lower, necessitating a 33.2 dB more sensitive receiver for the same order of signal quality to a LEO gateway.

We note that the difference in G/T of a GEO system and Iridium system is (37-24.5) dB or 12.5 dB rather than 33.2 dB. The remaining difference of 20.7 dB could be attributed to differences in link parameters such as link margin and satellite EIRP.

Comparing the ICO system at an orbital altitude of 10000 with Iridium system we note that path loss difference is about 22.1 dB, whereas the gateway G/T difference is 2.1 dB (i.e., 26.6-24.5). It is assumed that the remaining difference of 20 dB is attributed to differences in parameters such as link margin and satellite EIRP.

Comparing the ICO system at an orbital altitude of 10000 with a GEO system we note that path loss difference is about 11.0 dB, whereas the gateway G/T difference is 10.5 dB (i.e., 37.0-26.6). The remaining difference of 0.5 dB is attributed to differences in link parameters.

Note: The high G/T difference of Iridium system with ICO and GEO systems appears to be due to the large link margin built into Iridium system.

A comparison of total space segment capacity of each system

A comparison of the global space segment capacity is listed in the table below. It is assumed that the quoted values in table 1.2 are specified for power limited satellites i.e. the capacity per satellite is based only on available power on each satellite (i.e. spectrum is not limiting the capacity).

Global capacity = N*C

Where
N = Total number of satellites
C = Number of channels per satellite

System	**No of satellites**	**Capacity per satellite (RF channels)**	**Theoretical global capacity (RF channels)**	**Capacity relative to Globalstar %**
Iridium	66	1100	72600	63.0
Globalstar	48	2400	115200	100.0
ICO	10	4500	45000	39.1
GEO	3	16000	48000	41.7

Assumption: Power limited system

Comments:

1. LEO systems offer the largest theoretical global capacity, with Globalstar system providing the highest. The capacity of Iridium, ICO and GEO systems respectively relative to Globalstar is 63%, 39.1% and 41.7%.
2. In a power-limited satellite, the capacity is bound by the available effective isotropic radiated power (EIRP) in the forward direction. The capacity (in dB) is given as:

(E_{sat}-E_{ch}) dB

Where

E_{sat} = available satellite EIRP (dBW) in the forward direction

E_{ch} = EIRP per channel (dBW).

E_{sat} is conditioned by the prevalent spacecraft technology and system goal (e.g. cost, throughput, anticipated traffic flow).

E_{ch} is conditioned mainly by link margin, modulation/coding scheme and receiver sensitivity (G/T)

The required receive energy per bit to noise power spectral density (Eb/No) depends on the modulation and coding scheme. The magnitude depends on the receiver G/T and available forward link satellite EIRP and user terminal EIRP in the return direction in conjunction with respective path and miscellaneous losses and link margin. The link margin is based on the desired link reliability that depends on fading characteristics of the channel, interference, and miscellaneous system impairments. These parameters depend on system goal, features and assumptions used in the respective radio link design.

3. We note that all the systems have utilized variants of QPSK scheme and voice encoder in the range 2.4-4.8 kbps.
4. All the systems utilize FDM and three of the four systems utilize TDMA. Globalstar system utilizes a CDMA scheme.
5. The details in the table are insufficient to make a fair assessment of the spectral efficiency. For example, the number of voice channels per RF channel in not available in the table.

3) Explain the difference between architectures of non-geostationary satellite system which provide non-real-time communication service and real-time communication service.

In a non-geostationary earth orbit (NGEO) satellite system that provides real-time communication service, a call in progress must be re-routed to maintain connectivity when the connected satellite or spot beam moves out of visibility necessitating *handover(s)* to a rising satellite or spot beam. The coverage must therefore be seamless.

A call may be routed to the final destination over inter-satellite links, or via one or more ground-satellite hops, or terrestrially. Figures 3(a) and 3(b) below illustrate source-destination connectivity deploying a terrestrial backbone and inter-satellite links respectively. For example, the Iridium system deploys inter-satellite links whereas the ICO system uses terrestrial routing.

Figure 3 (a) and 3(b) Sketch a schematic capturing the main elements of figures 1.3(a) and 1.3(b))

For non-interactive or store and forward services, a discontinuous satellite coverage is acceptable. The message is stored in a satellite or ground station buffer and delivered to the destination over single or multiple satellite hops within a specified time limit. Figures 4(a) and 4(b) below represent the main entities of store and forward system with satellite and ground-based buffers, respectively. In figure 4(a), the message M received at a satellite at t=0, is transmitted to the destination at time t_2. In figure 4(b), transmissions from a mobile van, received at a fixed site 'A' is stored and transmitted when a satellite mutually visible with the destination N becomes available.

Figures 4(a) and 4(b) (Sketch a schematic capturing the main elements of figures 1.4(a) and 1.4(b))

Although constellation of such a system need not provide a seamless connectivity in the service area, the design must ensure that messages are delivered within the stipulated time delay.

4) What is the rationale for using low or medium earth orbit in preference to the geostationary orbit for the provision of hand-held service?

The arguments in favor of low and medium earth orbit are:

- Low and medium earth orbit systems offer low transmission delay that are comparable to terrestrial system
- Since path loss is lower in comparison to a geostationary orbit system it reduces the transmission power needs of user terminals which facilitates hand-held services as these terminals must operate with restricted transmit power in compliance to safe radiation limits and lower power reduces user terminal cost and increases time between battery recharge.

LEO systems have the lowest transmission delay, provide true global coverage but the space segment and the network is complex. MEO systems have intermediate transmission delay; provide true global coverage with a moderately complex space segment and network. GEO systems exhibit the highest transmission delay with coverage limited to ~± 76° latitude; but the network and space segment arrangement are simple and the system is supported by mature technology

5) The architecture of a satellite system is influenced by a number of technical considerations in addition to the service requirements. State these considerations. Briefly explain the role of each in system design.

Service requirements form the basis of the MSS network architecture. There may be several solutions and alternatives but the outcome is influenced by real-world constraints that include technical and non-technical matters such as regional priorities and business goals. Keeping within confines of the question, technical considerations in formulating system architecture and their respective roles are summarized as follows.

Radio frequency link budget

The constraints imposed by the service-link link budget impose the most significant limitations to the system capacity, user throughput and user terminal size and capability.

<u>Forward link</u>

Satellite EIRP and user's receiver sensitivity limitations: The permissible effective transmitted power of a satellite and the low sensitivity of user terminals (UT) bound the user throughput. The highest available satellite EIRP cannot fully offset the radio link impairments and the low UT sensitivity. Countermeasures to mitigate the problem include – use of robust transmission format that minimize the impact of impairments such as fading, increase in receiver sensitivity traded-off with terminal weight, size and cost, and terrestrial retransmissions in severely disadvantaged locations like city centers and tunnels.

<u>Return link</u>

UT EIRP and satellite receiver sensitivity restrictions: UT EIRP and satellite receiver sensitivity limit return link throughput. An UT's EIRP is limited due to small antenna gain for mounting space, weight and cost considerations, particularly when dealing with portable or mobile UTs; for hand-held units EIRP is further limited in order to maintain compliance to radiation safety standards. Satellite receiver sensitivity is limited by the receive antenna gain of a satellite and Earth's thermal temperature of 290^{o}K.

<u>Interference</u>

In an interference limited radio environment, the available throughput is further constrained as useful power is lost to overcome interference.

Spectrum

In a spectrum limited system the throughput is limited by available bandwidth.

Propagation impairments and physical environment:

In a mobile satellite system, the physical medium around a mobile terminal and the receiver's antenna characteristics set a boundary on throughput. An area surrounded by obstructions reduces the received signal power by shadowing and hence the throughput; while a reduction in antenna gain increases pick-up of multipath noise and reduces the receiver sensitivity, both features reduce a receiver's throughput.

The size and hence gain of the mobile antennas depend on the mounting space; for example, a hand-held terminal uses a low-gain omnidirectional antenna whereas a ship-borne terminal deploys a high gain (e.g. 80 cm parabolic dish) resulting in a more sensitive terminal with a capability for a significantly higher throughput.

Impairments for narrow-band communication at bandwidths up to ~100 kHz in L-band are relatively benign in maritime and aeronautical channels at mid-high elevation angles. Aeronautical channels become dispersive beyond ~100 kHz when traversing over a quiet sea.

Impairments are significantly high in land mobile channels, tending to get worse in shadowed areas and as antenna gain is reduced. Intermittent deep signal fades tend to break radio link causing discontinuity to real time services, while multipath manifests as extraneous noise. Countermeasures include provisioning higher link margin, robust modulation and coding and ARQ schemes. As an example, measurements demonstrate that L-band signals received by a hand-held unit can undergo peak-peak fluctuations in excess of 10 dB in ten second segments, with a gradual variation of the mean level, while for the vehicle-mounted system the fluctuations are of a similar level with stable but gradual change in the mean value in between. Figure 5 below illustrates a typical sample of received signals for hand held and on a stationary vehicle.

Figure 5 A sketch here showing variations on hand-held and vehicle-mounted antennas (Sketch approximations of figures 1.5(a) and (b))

Doppler is introduced by relative motion between satellite and mobile; it affects aeronautical channels and NGEO satellite systems in particular. Countermeasure includes open and close loop frequency correction arrangements.

Orbit

The visibility statistics from the service area is determined by the orbital characteristics.

Orbits are categorized by altitude, inclination and eccentricity. Altitude determines a satellite's field of view (coverage) – the area increases with altitude; Inclination influences the minimum-maximum latitudes of coverage; the two extremes of orbital inclination are - equatorial, (inclination = 0°) and polar (inclination = 90°). An equatorial orbit satellite would cover a belt around the equator, whereas a polar satellite would cover a belt orthogonal to the equator thus covering the full Earth over a period (depending on the orbital period) due to the Earth's west-east rotation. Eccentricity determines the shape of an orbit. Satellites in a circular orbit travel at a uniform velocity to provide an unbiased temporal coverage, whereas satellites in an elliptical orbit travel at variable velocities and hence dwell longer over regions where orbital velocity is low.

Tolerable delay in data delivery

Tolerable delay in a system is application-dependent. Whereas a delay of more than ~400 ms causes annoyance in a conversation, delays of minutes and hours are acceptable for e-mail delivery. A non-real-time service can tolerate a break in communication link by data recovery techniques such as store and forward; in an interactive service an uninterrupted connection must be maintained. The delay specification influences various features of a mobile satellite communication system. For example, in a delay-tolerant system an NGEO satellite system can scale down the constellation size such that visibility statistics is restricted to the tolerable delay.

Handover

Satellite location is static for a geostationary orbit (GEO) system. The boundaries of spot beams in a GEO are fuzzy, extending several tens of kilometers which promotes a graceful degradation in signal quality for slow-moving vehicles. Thus, slow-moving mobiles need not handover a call to the next beam or satellite, as users spend a considerable period in this fuzzy zone where signals degrade gradually that allows the user to adjust. Nevertheless handover is necessary for fast moving mobiles such as aircrafts or fast trains.

In non-geostationary satellite systems on the other hand, satellites are non-stationary and hence a user is likely to communicate through different beams and/or satellites during a call, making handover a necessity.

Moreover, when inter-satellite links are used, handover between satellites becomes essential due to dynamic path geometry.

Handover is a component of mobility management, described next.

Mobility management

The function of mobility management is to locate a called mobile, route the call and once established maintain it, meeting the quality-of-service criteria.

In derivatives of the GSM system, each mobile is registered in a database, called home location register (HLR); if the mobile migrates outside the home territory, the mobile registers itself with the visiting system's visitor location register (VLR). The VLR conveys the location of each visitor to its HLR. Whenever a call is addressed to a mobile, the mobile switching center (MSC) interrogates the mobile's HLR to obtain the mobile's location and then routes the call. Figure 6(a) below portrays a routing scheme in a NGEO system with satellite-ground hops, while Figure 6(b) represents a scheme over an inter-satellite link. Various alternatives are feasible and hence the routing strategy must be optimized carefully.

Figures 6(a) and 6(b) (sketch of figures 1.3(a) and (b) schematic).

In an IP enabled network prior to a connection handover at the RF (i.e., the physical layer) the IP address must migrate to a new attachment. The Internet engineering task force (IETF) has recommended mobility management schemes such as Mobile Internet Protocol version 6 (MIPv6) which have been adapted for satellite IP networks.

Satellite access

In an MSS environment thousands of users share satellite resources and therefore high satellite access efficiency is paramount. Demand-assigned single channel per carrier (SCPC) frequency division multiple access (FDMA) or FDM/time division multiple access (TDMA) schemes, where a pool of channels is shared by all users in time or packet(s) basis, and code division multiple access (CDMA) offer effective solution. The channel pool can be managed by either a central or a distributed architecture. In a central architecture, a pool is managed centrally, whereas in a distributed architecture, separate pools are assigned to each participating fixed station for self-management.

In a code division multiple access (CDMA) scheme an RF channel is shared by all the users each using a unique code. This scheme offers advantage in terms of interference and multipath mitigation, and facilitates soft handover.

Data traffic exhibits a variety of behavior - ranging from sporadic bursts to continuous streams and dedicated channels waste resources when the traffic flow is intermittent; consequently the accessing schemes are matched to the anticipated traffic characteristics. Common accessing schemes used for data communications include Aloha, slotted Aloha, Reservation Aloha, Time Division Multiple Access, etc.

Spectrum management

Frequencies are allocated by the Radio Regulations (RR) and managed by the local/regional regulatory regime taking into consideration engineering, commercial and political factors. From an operator's perspective, spectrum management includes selection of an appropriate frequency band, obtaining clearance from the regulatory authorities, and managing its usage efficiently.

At present, a majority of MSS systems operate in L (~1.5 GHz) and S (2 GHz) bands; a few in the K_u band, and a beginning towards use of K_a band. L and S bands are suitable for communication-on-the-move because of relatively benign propagation attributes and mature technology in these bands. Since these bands have very limited allocation (~40 MHz) shared by several operators, they are congested and hence unsuitable for broadband communication (> ~1 Mbps). Therefore some operators have preferred K_a band for broadband on portable terminals with directive antenna.

Due to heavy usage of the MSS spectrum, interference management is an important consideration in the planning and operation of mobile satellite systems. A certain level of interference is budgeted in the radio link design to enable inter and intra system frequency sharing. To minimize the probability of unwarranted interference, operators follow a strict spectrum monitoring regime and procedures to manage harmful interference. Techniques to maximize spectrum efficiency include spatial frequency reuse by spot beams, efficient radio transmission formats and judicious radio resource management.

Radio link reliability

Techniques for improving radio link reliability include use of robust modulation and forward error correcting codes (which govern link margin), fade countermeasures, store and forward technique to support communication in deep fades and adaptive power or code rate control.

6) Compare the characteristics of satellite and terrestrial mobile systems. Explain the reasons for a growing convergence between these systems.

Mobile satellite systems are used extensively on thin routes and in contingency on disruption to the terrestrial infrastructure.
The characteristics of satellite and terrestrial systems are compared in the table below.

Satellite	**Terrestrial**
Wide area coverage - typically thousands of kilometers	Relatively lower coverage - typically hundreds of kilometers; coverage can become disjointed over large regions.
Seamless roaming over wide areas is a natural system feature	Roaming over wide areas encompassing several countries involves several operators and possibly disparate systems.
Handsets resemble cell-phones; Large terminals (0.3-1 m parabolic antenna) provide broad-band services up to several tens of Mbps.	Terminals are small and attractively packaged with a variety of applications including broadband.
Terminal costs are relatively high	Terminal costs are low, consistent with consumer market.
Usage costs is high	Usage cost is relatively low
Suited for wide area coverage and thin routes (e.g. traffic density <0.1 Erlang/sq. km).	Suited for urban and suburban environment; uneconomic on thin routes (e.g. Traffic density <1 Erlang/sq. km).
Ideally suited in aeronautical/land/maritime environments to provide coverage over vast	Generally operate in land environment; limited coverage is possible in aeronautical (over land) and maritime

geographical span.	environments (close to shore)
Service include voice and data with a throughput up to ~ 500 kbps in L band and up to 50 Mbps in K_a band	Services include voice and data up to several Mbps. Higher throughput of several tens of Mbps available in first releases of 4th generation systems
Serve niche market - ships, aircrafts, trucks, international travelers and businessmen, cellular extension, tourism, media, etc.	Serve consumer market for social and business needs in populated areas; coverage limited to coastal areas in maritime environments; air coverage over land is available in certain parts of the developed world; air coverage on thin routes is available via MSS links
Frequency is reused spatially at distances of hundreds of kilometers.	Frequency is reused at distances of ~100 m – tens of km.
Handover between spot beams or satellites are not always essential for geostationary orbit systems but essential for non-geostationary systems.	Handover is frequent and necessary in all cellular systems.

7) What are the strengths and limitations of a mobile satellite service?

The strengths of a mobile satellite services are:

- Wide area coverage that can span continents and indeed the entire world, including the poles
- Unbiased coverage in service regions
- Broadband speeds up to 50 Mbps is available (year 2014)
- Service is available over vast expanses as soon as a satellite system is commissioned for operation
- Ideal for thin routes and adjunct applications such as distress signaling, search and rescue, first responders, etc. e.g., when the terrestrial infrastructure has failed, video streaming from remote locations

The limitations of the service are summarized as follows.

Limitations	**Comments**
Expensive infrastructure, call and terminal costs	Terminal and call costs continue to reduce; infrastructure costs reduce for systems derived from terrestrial mobile network.
Terminals are large compared to terrestrial systems	Hand-held units resemble cell-phones; A significant size reduction in recent years; Terminal size is not crucial in many MSS applications (e.g. ship-borne, railway and aeronautical)
User interface is complex	The limitation applies to specialist equipment such as ship-borne, aeronautical and large land terminals; nevertheless, end-users see a simple interface similar to terrestrial systems
A general lack of awareness of technology.	Heightened awareness in recent years
Systems susceptible to local interference.	Resolution of local terrestrial interference arduous
Routing arrangements can be complex and time consuming.	
Concern about unauthorized bypass of a country’s network.	Practical solutions feasible
Service limited to thin route	Terrestrial retransmissions using an ancillary terrestrial

	component (ATC) allow extension to populated areas
Service unreliable in areas susceptible to shadowing e.g. urban and suburban locations	An ATC improves reliability

8) Briefly outline the factors likely to influence the evolution of mobile satellite systems.

There are numerous technical and commercial factors which are likely to influence the evolution of mobile satellite systems

User needs

The urge of people to remain in contact under all circumstances has been instrumental in the success of mobile communication technology. Trends in the terrestrial mobile and fixed services are likely to be a precursor to those in the mobile satellite systems. Escalating data demands for both personal and business applications is likely to fuel demand for broadband systems. Example usages include internet access to database on the move or at remote locations, remote teleworking, remote access to e-mails, and image transfer from mobiles such as ships, etc.

MSS data rates have increased from a few kbps in the first-generation to 50 Mbps in the current generation; numbers of successful commercial MSS ventures have increased in recent years; and a number of MSS applications are identical to those of terrestrial mobile systems.

Location-based applications are an example of value added MSS service.

Low user terminal cost and usage charge are vital for service penetration.

Exploitation of untapped market

Vast, sparsely populated areas throughout the world remain unserved by fixed or cellular systems, either because the service is uneconomic in these regions or a lack of infrastructure. Satellite systems are ideally suited to fill such gaps.

User awareness

Recent growth trend is due to heightening awareness of MSS's potential in niche areas such as remote reporting, first responder, long-haul aeronautical, maritime and railway, environments, etc.

Impact of competing technologies

The growth of competing technologies such as mobile FSS, interactive broadcast technologies and expansion in coverage of terrestrial mobile system can adversely affect the growth rate. Synergistic solutions and system architecture are therefore a way forward in the evolution of MSS.

Advances in network architecture

Reuse of terrestrial mobile system technologies allows economies of scale for the MSS and facilitates a tighter integration with terrestrial mobile systems permitting a larger share of user base. Synergistic solutions are therefore a primary theme of modern MSS networks.

A mobile satellite system comprising an ancillary transmission component (ATC) architecture promises a seamless extension of the system to heavily disadvantaged locations such as city centers and within buildings.

Alleviation of spectrum shortage

Additional spectrum in L and S bands offers an unconstrained growth potential in the hand-held and land-mobile on-the-move sectors whereas frequency bands including and above K_a band offers growth potential for next generation broad-band services (hundreds of Mbps) on larger terminals installed on ships, aircrafts and used for portable land applications.

Standardization

While standardization allows economies of scale, a large number of MSS standards lead to unwarranted market fragmentation. Examples of prevalent standards include GMR, GMR 2 (an evolved version of GMR), GMR-1 3G, DVB RCS+M, etc.

Advances in technology

- Satellites with regenerative transponders and on-board computing**:** Regenerative transponders lower mobile terminal EIRP requirements, allow size and cost reduction and mitigate effects of interference. The ability of regenerative transponders to decouple up and down links enables optimization of up and down links separately. Software reconfigurable regenerative transponders can minimize risk of technology obsolescence.
- On-board processing can incorporate advanced space-based features such as individual call-routing, various network functions, sophisticated beam-forming techniques, etc.
- Inter-satellite links simplify ground connectivity.
- Improvement in satellite launch techniques leads to cost reduction.
- Improved satellite-production techniques**:** Long production cycle of satellites results in technology obsolescence and cost overrun, particularly for large constellations. There are benefits in applying mass-production techniques similar to those of the automotive industry to satellite manufacturing.
- Advancements in user terminal technology: Low-cost, cellular-integrated satellite phones affordable by individuals are vital to uptake of MSS. Phenomenal advances in VLSI, packaging, battery technology, etc. enable affordable dual- and multi-mode hand-held and portable MSS terminals.
- Technology improvements in areas of modulation, channel coding, voice-coding, compression, access technology, system architecture, network management, etc. are vital for MSS.
- Since mobile satellite services provide lower throughput than wired systems, evolution in robust spectrally-efficient transmission technology is essential as demonstrated in terrestrial mobile technology.
- Increase in the service link effective isotropic radiated power (EIRP) reduces the size and cost of user terminals and increases broadband speed in the forward direction, while improvement in spacecraft receiver sensitivity (G/T) increases the return link transmission rate. Improvements in spacecraft antenna and power-amplifier are enabling technologies in this respect.

CHAPTER 2

1) Define and explain the significance of orbital parameters in an MSS context.

Satellites orbit the Earth on well-defined paths specified precisely by a set of six orbital parameters. The parameters are used to pin-point a satellite's location in space. The knowledge of satellite position and geometry is necessary to track a satellite and estimate radio link parameters for design. Hence the parameters determine the characteristic of space segment. Figures 1(a) and 1(b) show the parameters.

Figure 1(a) [figure 2.3 (b])

Figure (1b) [figure 2.3 (a)].

1. The *Semi-major axis, a,* describes the size of an elliptical or a circular orbit (figure 1(a)). Semi-major axis is used in defining an elliptical orbit. In a circular orbit it reverts to orbital radius. It relates to a satellites altitude.
2. *Eccentricity, e,* represents the shape of an orbit. Eccentricity of a circular orbit is 1; an orbit's elongation increases with an increase in its eccentricity.
3. *Inclination, i,* describes the orientation of an orbit with respect to the equatorial plane (figure 1(b)). It is the angle between the plane of the orbit and the equatorial plane. The orbit crosses the equatorial plane at two points called ascending and descending nodes. At the *ascending node* a satellite crosses the equator from the southern hemisphere to the northern and at the *descending node* a satellite crosses the equator when moving from the northern hemisphere to the southern.
4. *Right ascension,* Ω, shows the orientation of an orbit with respect to the X-axis of the coordinate system. It is the angle between the X-axis and the *line of nodes* which refers to a line joining ascending and descending nodes.
5. The *argument of perigee,*ω, describes the orientation of an elliptical orbit's perigee with respect to the line of nodes measured as the angle between the line of nodes and the perigee.
6. Time t_p is the time elapsed since a satellite has passed a reference point in an orbit, usually the perigee; where the reference time is known as Epoch.

The eccentricity determines the shape of the orbit. Inclination determines the latitude limits of the coverage area. Altitude (semi-major axis) determines the propagation delay and relates to transmit and receive signal powers in radio link design.

The reliability of MSS radio link improves with an increase in the elevation angle and therefore provision of services at high elevation angles is desirable. At an inclination of 63.4° an eccentric, elliptical orbit provides a quasi-stationary, high-elevation coverage at high latitudes.

Low altitude orbits reduce propagation delay and the power requirements on satellite and user terminals but satellites in such orbits are non-stationary leading to a relatively complex space segment coupled with large variations in signal power and complex network architecture. Geostationary satellites provide stable radio link with a relatively simple network and a coverage extending nearly 1/3rd of the Earth below ± ~76° latitude. Whereas 10-80 are necessary for low or medium altitude satellites only 3 geostationary satellites can cover the Earth.

Sun-synchronous orbit maintain a constant relationship with the Sun and hence by using appropriate orbital parameters sun eclipse on a satellite can be eliminated - a useful attribute for satellites' power-generation sub-system as the need of a secondary power is eliminated.

2) Figure 2.7 (c) represents contour plots of azimuth angle as a function of latitude and longitude relative to the sub-satellite point of a geostationary satellite. Estimate the azimuth of a geostationary satellite

located at 15°E when the user is located at: (i) 55°E/20°N; (ii) 75°E/80°S; (iii) 35°W/40°N; (iv) 55°W/40°S.

An accurate estimate can be made using equation A.4 (see Appendix). The results are:

User (i): 247.8° ; User (ii): Satellite is below horizon; User (iii): 118.3°; User (iv): 76.8°

3) **What are the various types of perturbations on satellites in Earth orbits? How do such extraneous forces affect satellite motion?**

Perturbations are mainly caused by:

- non-uniform gravitation of the Earth;
- gravitational effects of the Sun and the Moon;
- atmospheric drag (at low orbital altitudes);
- to a lesser degree, solar radiation pressure.

Non-uniform gravitation of the Earth
The gravitational force around the Earth varies due to the non-uniform distribution of the Earth's mass; furthermore the Earth's shape is slightly ellipsoidal as the polar radius is ~21 km shorter than the equatorial radius. A cross-section of the Earth shows a semi-major axis approximately along the line 165°E and 345°E (15°W) and a semi-minor axis approximately along the line 75°E and 255°E (105°W). Therefore, gravitational force is no longer directed towards the geocentre, but towards the centre of the ellipsoid. Non-uniform gravitational fields cause the following effects to an Earth orbit:

- Precession of perigee in orbital plane;
- Precession of orbital plane around the Earth's north-south axis;
- Perturbing force in a direction along the orbit.

The first two effects are most noticeable in low Earth orbit (LEO) and medium Earth orbit (MEO) satellites whereas the last mentioned affects GEO satellites noticeably.

Gravitational effects of the Sun and Moon
Gravitational effects of the Sun and the Moon are negligible in comparison to the Earth's gravitational effects for satellites in low and medium altitude orbits, but these forces affect satellites in geostationary orbit. The main effect of the gravitational pull by the Moon is a change in the inclination of the orbit between ~0.48^0/year and ~0.67^0/year with a period of about 18.6 years. For example, a minimum occurred in the year 1997. The change is caused by variation in the inclination of the Moon itself. The yearly change in inclination due to the Sun is ~0.27^0, which is steady for practical purposes. The net effect of these two forces causes the inclination of a geostationary satellite to vary between ~0.75^0 and ~0.94^0 each year. These two forces act in the same direction on an average and hence the resultant change in inclination is the sum of these.

A component of the force due to the Earth's ellipsoidal shape is in a direction opposite to the force of the Sun and the Moon. These forces cancel each other at an inclination of about 7.5^0. Consequently if the inclination of a geostationary satellite is left uncorrected then the inclination of the satellite oscillates around an inclination of 7.5^0 with a maximum of 15^0 and a period of 53 years. In practice, inclination is corrected regularly by firing thrusters on satellites in an orbital manoeuvre called north-south station keeping.

Solar radiation pressure
Solar radiation pressure affect large space structures such as large geostationary satellites often deployed for MSSs. The net effect is an increase in the eccentricity together with disturbance along the north–south axis of the satellite that necessitates periodic corrections.

Atmospheric drag
Atmospheric drag is caused by friction to a satellite's body, in the upper parts of the Earth's atmosphere. Therefore, satellites in low earth orbits suffer the largest atmospheric drag. Below ~180 km, the friction causes excessive heat on a satellite's surface such that satellites burn out. From this perspective this altitude is considered as the lower limit of space. Atmospheric drag is directly related to the surface area and mass of

a satellite and it becomes noticeable below ~750 km.

Miscellaneous disturbances
A number of miscellaneous forces act on a satellite, which must be compensated to maintain the desired orientation of a satellite. Such disturbances are caused by the effects of the Earth's magnetic field, impact of meteorites, self-generated torque, etc.

4) Outline the principle, characteristics and advantages of a sun-synchronous orbit.

The disturbance caused to satellite due to the Earth's non-uniform gravitational field causes a rotation of the orbital plane around north-south axis. The rate of precession, Ω, is given as:

$\Omega = 9.95\,[R/a]^{3.5}\,(\cos(i) / (1-e^2)^2)\ °/\text{day}$

Where, R = the mean equatorial radius (~6378 km), a = the semi-major axis, i = the inclination and e = the eccentricity

If Ω is made equal to the rate at which the Earth rotates around the Sun i.e. 0.986° per day, the relationship of the orbit with respect to the sun remains constant and the orbit becomes sun-synchronous.

Satellites in this orbit rise at the same time over any location on the Earth – a feature that benefits acquisition and tracking of satellites. When used for Earth resource survey the satellite views the Earth at the same level of illumination each day, which makes detection of changes easier.

By orienting the orbit suitably it is possible to eliminate eclipse on a satellite. This simplifies the power supply sub-system of a spacecraft and reduces the weight of satellite by eliminating the need of secondary power source for eclipse operation.

5) Determine the inclination of a sun-synchronous circular orbit of 750 km altitude (Earth radius = 6378 km).

The rate of change Ω of the orbital plane is given by equation (2.5) as

$\Omega = 9.95\,[R/a]^{3.5}\,(\cos(i) / (1-e^2)^2)\ °/\text{day}$

Where, R = the mean equatorial radius (6378 km), a = the semi-major axis, i = the inclination and e = the eccentricity

Or, $i = \cos^{-1}\left(\frac{\Omega}{9.95}\frac{(1-e^2)^2}{\left(\frac{R}{a}\right)^{3.5}}\right)°$

For a sun-synchronous orbit, Ω = 0.986°/day

Given, a = 750 km, e = 0 [xxx check]

Substituting values in the equation for estimating the inclination,

$i = = \cos^{-1}\left(\frac{0.986}{9.95}\frac{1}{\left(\frac{6378}{750}\right)^{3.5}}\right)$

Then, inclination = 89.99°

6) How are the parameters of an eccentric orbit adjusted to provide service to high latitude region above 81° N? Suggest limitations of this approach.

The perigee of an orbit rotates in the orbital plane due to the Earth's non-uniform gravitational field. The rotation ceases at an inclination of 63.4° (or 116.6°) and moreover field-of-view of satellites in such an orbit extends above 81° N, beyond which geostationary satellites are invisible (below horizon). Inclination of either 63.4° or 116.6° can be selected as appropriate for the service region. A highly eccentric orbit with a large semi-major axis and apogee over the service area results in a large dwell time of satellites over the service area, recalling that satellite velocity reduces with altitude. Examples of this class of orbit are Tundra orbit that has a period of 24 hours and Molniya orbit that has a period of 12 hours.

The limitation of satellite systems using highly eccentric orbit arises due to non-stationary nature of satellites and therefore for seamless coverage a number of satellites are required, necessitating satellite handover. For example, for a 24 hour orbital period, 3 satellites can cover the service area each serving 8 hours. Furthermore, because of their high inclination, true global coverage is not feasible from such orbits

7) What is the rate of change of argument of perigee of an elliptical orbit with the following characteristics:

Eccentricity = 0.15; Semi-major axis = 10000 km; Inclination = 45°? (Earth radius = 6378 km)

Plot a graph of (rate of change of) argument of perigee versus inclination ranging between 5° and 90°. Comment on the results.

The rate of change of argument of perigee, ω, of an elliptical orbit is given as,

$\omega = 4.97\ [R/a]^{3.5}\ (5\cos^2(i) - 1)/\ (1 - e^2)^2\ °\ /\text{day}$

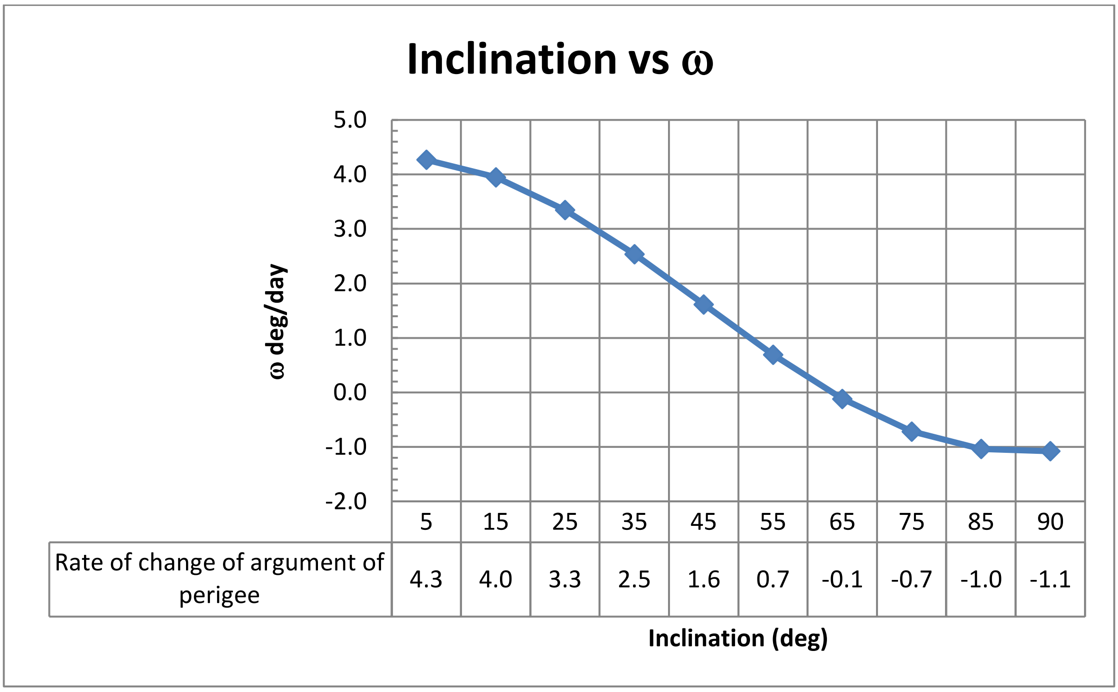
Where R = the mean equatorial radius (~6378 km), a = the semi-major axis, i = the inclination and e = the eccentricity.

Substituting,

$\omega = 4.97\ [6378/10000]^{3.5}\ (5\cos^2(45) - 1)/\ (1 - (0.15)^2)^2\ °/\text{day}$ (xxx calculate)

Or, ω = 1.617 °/day

A plot of rate of change of argument of perigee versus inclination ranging between 5° and 90° is listed in the figure below.



	5	15	25	35	45	55	65	75	85	90
Rate of change of argument of perigee	4.3	4.0	3.3	2.5	1.6	0.7	-0.1	-0.7	-1.0	-1.1

Comments:

- As expected, ω = 0 at i = 63.4 deg (see zero crossing)
- ω is positive below inclination of 63.4 deg and negative above.
- Variations range between +4.3 deg/day and -1.1 deg/day

8) Outline the concept of sun eclipse at the geostationary orbit? What is the impact of such eclipses on a spacecraft? Suggest an orbital configuration in which sun eclipses can be eliminated,

Figure 1(a) below shows the apparent motion of the ecliptic plane with respect to the Earth's equatorial plane.

Figure 1a: [Sketch figure 2.16 (a)] Apparent motion of ecliptic plane with respect to the Earth's equatorial plane over a year

At spring and autumn equinox the equatorial and ecliptic planes coincide and therefore the Earth lies in the same line as the Sun and the geostationary arc. The Earth's shadow subtends an angle of about 17.4° at the altitude of the geostationary orbit (GEO) and the Sun is eclipsed on days when the shadow intercepts the GEO during intervals when a GEO satellite lies behind the Earth.

Figure 1(b) below shows the Earth's shadow on the geostationary arc progressively from the first day of the eclipse when the shadow begins to graze the GEO to the last day when it grazes past the orbit.

Figure 1(b) [Sketch figure 2.16 (b)]: Eclipse duration in minutes on a geostationary satellite (left side of the figure) and the Earth's shadow on geostationary arc on days 1, 22 and 44 of the eclipse (right side of the figure)

The duration of eclipse increases progressively, maximizing on equinox, then subsides as the Earth's shadow moves away from the geostationary arc. On the day of the equinox, the Earth and the GEO are on the same line and hence the eclipse duration is the highest, reaching around 69.4 minutes, the time for a satellite to traverse 17.4°. Sun is shown moving relative to a fixed Earth and directions represent the apparent motion of the sun relative to the equator around the spring and autumn equinox.

The peak of the eclipse coincides with the satellite's midnight when the satellite is directly behind the Earth. Therefore, by selecting the satellite to the west of the coverage region, the onset of the eclipse can be delayed past the region's midnight when the traffic carried by a satellite is usually low and hence the spacecraft battery's capacity requirement is reduced. For example, setting the satellite 15° to the west would set back the eclipse to 01:00 am.

Impact of eclipse on satellite
Satellites generate power for their functioning from solar cells. In absence of sunlight the satellites rely on rechargeable batteries. Such batteries increase the weight and hence launch cost of satellites. Regular charge/discharge cycles reduce the lifetime of batteries. There is thus also a degree of risk to the mission. Besides, satellites undergo a thermal shock as they move in and out of eclipse where temperatures variation can range from -180°C in eclipsed condition to +60°C on exiting an eclipse, necessitating a stringent thermal control sub-system.

9) What are the types of constellations used for mobile satellite services? Outline the benefits and limitations of each.

Depending on the service area and communication requirements, MSS utilise low, medium, highly elliptical, geostationary or hybrid orbit.

Taking atmospheric drag as the lower limit and the first Van Allen radiation belt as the upper limit, the approximate altitude of low Earth orbit (LEO) is between ~750 and ~1,500 km. The altitude of a medium Earth orbit (MEO) lies in the range ~10,000 to ~12,000 km, a span of low radiation between the first and second Van Allen belts.

Benefits and limitations of each orbit is summarised in the table below

Orbit	**Advantages**	**Limitations**
Geostationary Orbit	Well developed and proven technology; Insignificant range related signal strength variation due to negligible ground-satellite range variation during calls; Interference calculations are straightforward due to stable and simple geometric relationships; Low Doppler; Supports hand-held and broad-band service; Coverage is available to most populated areas of world (but see the limitations); Only three satellites can provide near world-wide coverage; A single satellite is adequate for regional coverage.	Coverage unreliable between ±76° to ±81° latitude, and unavailable beyond ±81° latitudes; Poor service link reliability at mid-to-high latitudes, particularly for land mobile satellite services; Large propagation delay (~240 ms one-way) - affects voice and time-sensitive data protocols; Large path loss; Spectrum efficiency lower than LEO and MEO; High launch cost; In-orbit back-up satellite increases system cost disproportionately.
Highly elliptical orbit (HEO)	Reliable service links possible at mid-to-high latitudes; Lower launch cost than GEO; Distributed space segment architecture allows partial service in case of a satellite failure; 1:1 satellite redundancy is not required.	Inefficient for global coverage; Propagation delay can be higher than GEO system; Doppler effect quite significant; Handover is essential; Satellites pass Van Allen radiation belts regularly to the detriment of electronic components.
MEO	Offers true global coverage capability; Lower path loss than GEO; Medium propagation delay (~55-80 ms transmitter to receiver; Enables efficient use of spectrum; Distributed space segment architecture allows partial service in case of a satellite failure; 1:1 satellite redundancy is not required;	Large number of satellites necessary (10-12); Receive signal strength is variable due to variability in range and elevation angle; Doppler effect significant; Complex network architecture: e.g., handover, intersatellite links, dynamic satellite resource management, routing etc.; Tends to increase orbital debris because of need of large number of satellites per system; Relatively long time necessary for constellation deployment; Space segment maintenance is complex due to large number of satellites and distributed network

		architecture – higher number of satellite replacement than GEO but less than LEO
Low Earth Orbit (LEO)	Offers true global coverage capability; Lowest path loss; Lowest propagation delay (~ 3.5 to 15 ms one-way from transmitter to receiver) comparable to optical fibre system time delay; Efficient use of spectrum; Distributed space segment architecture allows partial service in case of a satellite failure; 1:1 satellite redundancy is not required;; Possibility of including position determination as a value-added service.	Large number of satellites necessary; Signal strength is variable due to variability in range and elevation angle; Doppler frequency shift is significant (highest); Complex space segment and network architecture due to: high hand-over rate, inter-satellite links, need of a dynamic satellite resource management, complex routing, etc.; Possibility of very large number of eclipses resulting in a large number of charge/discharge cycles; Rate of depositing orbital debris is the highest but it is relatively easy to remove debris from LEO; A busy space segment maintenance is involved - satellite replacement rate is more than in GEO or MEO.
Hybrid orbit	Utilises the best combination of orbits for the required service requirements.	Suited for unusual coverage requirements, e.g. to cover equatorial region with parts of high latitude region.

10) Which constellations would you select to provide mobile satellite service for: (i) a seamless global coverage; (ii) an equatorial coverage within ±30° latitude; (iii) a regional coverage to serve equatorial and mid-latitude regions? Justify your choice.

(i) Seamless global coverage

Geostationary and highly elliptical orbits are discarded since they cannot provide a seamless true global coverage. That leaves the option to use either a low or medium earth orbit. The choice between these would depend on factors such as cost in manufacturing and deploying the constellation, the target propagation delay, and type of service. Other factors which influence the choice include frequency availability, constellation maintenance and launch cost, etc.

(ii) Equatorial coverage within ±30° latitude

This type of coverage can be covered by an equatorial constellation where satellites follow each other to provide seamless or intermittent connectivity, as necessary. The altitude of the orbit may be either low or medium earth orbit and will be determined by the field of view which should ensure coverage within ±30° latitude.

A geostationary being an equatorial orbit is also suitable, although its coverage can only extend reliably up to ~± 76° latitude.

(iii) Regional coverage to serve equatorial and mid-latitude regions

Regional coverage can be provided by a geostationary orbit which gives good coverage up to the mid-latitude. A geostationary satellite system has the advantage of a simple network topology and static radio link which is more robust than those of non-geostationary satellite systems. Non-geostationary satellite systems may be unnecessarily complicated and wasteful as satellites will remain switched off in those parts of orbit which are outside the region. Nevertheless, if the service requirements warrant the use of low or medium earth orbit, the constellation can be designed according to service needs e.g., interactive or non-interactive. Walker and Beste provide methods for optimising medium/low earth orbit constellation for regional coverage (see book for reference).

11) Wha**t are the differences in the radio link design of geostationary and non-geostationary satellite from an orbital perspective? Suggest a technique to improve reliability of non-geostationary radio links.**

[Note: These topics are dealt in detail later in the book (e.g., see chapter 3 for Radio link analysis).]

In a geostationary satellite system for practical purpose the path length and elevation angle remain invariant for the duration of a user session. For a non-geostationary satellite on the contrary the path length and elevation angle vary rapidly due to motion of satellite relative to the user that may affect a call's performance. Maximum variations in these parameters occur in a LEO satellite system. Since the extent of shadowing is influenced by elevation angle, depending on user location and orbital parameters, non-geostationary satellite systems experience variable propagation impairments. Such variations are in addition to those caused by motion of a mobile and thus radio link design is more demanding for non-geostationary satellite systems.

Doppler frequency shift is of the order of a few hundreds of Hertz for geostationary satellite, whereas for non-geostationary satellites the variations are of the order of tens of kHz. Managing the Doppler shift in geostationary satellite stems is relatively straightforward. For slow moving vehicles the compensation can be done by the fixed network and with receiver involvement limited to aeronautical environment; however for non-geostationary satellite systems compensation at the receiver is also necessary.

In addition to the use of robust modulation and coding, reliability of radio link of non-geostationary satellite systems can improve by path and time diversity. In path diversity the signals are received from two satellites. It has been experimentally verified that probability of both paths fading simultaneously reduces when satellites are sufficiently apart and the signals from both paths can be combined to provide a more reliable radio link. Time diversity is based on the premises that if signals are separated in time e.g. by repetition then probability of simultaneous fading reduces provided time separation is adequate. In interleaving the adjacent bits of a message are separated in time (i.e. interleaved) to minimise corruption of a message by short error bursts.

Another approach is to increase the elevation angle in the service area to minimise corruption from ground clutter. For example, highly elliptical orbits inclined at 63.4° can cover high latitude locations at high elevation angle. Note that the coverage area shrinks as the minimum elevation angle is increased for circular orbits.

.

CHAPTER 3

1) Demonstrate with an example, the principle of co-channel frequency reuse by multiple spot beam satellites.

Figure 1 [sketch figure 3.2(a)]: A model for estimating interference

Figure 1 shows a model for estimating interference caused to a mobile or received on a satellite. The offset angle θ between the wanted and interfering source (another spot beam within the present context) determines the discrimination between the wanted and the interfering signals dependent on the receiver antenna gain at this offset; for a feeder link, the mobile station is replaced by a fixed earth station. The geometry is reversed when considering interference at a satellite receiver.

Figure 2 [sketch figure 3.2(b)] A model for estimating the interference between spot beams of the same satellite

Figure 2 illustrates a model for estimating the interference between spot beams of the same satellite; here the isolation between the wanted and interfering carrier is provided by the spacecraft antenna pattern. For example, a carrier transmitted at an EIRP of P dBW towards the center of coverage of spot beam 2 is received as interference -18 dB lower at the location M at the edge of spot beam 1. For a co-channel reuse and assuming that the same carrier type is reused in both beams, the carrier to interference ratio (C/I) is:

(P-3) - (P-18) dBW

i.e., 15 dB

More generally, interfering carrier power level in the *forward* link (space to Earth) is the sum of interference from all the interfering sources (i.e. co-channel frequency re-users):

$$\mathrm{I} = \sum_{k=0}^{n} I_{ek}\,(\theta)\mathrm{P}_{ik}\; I_{sk}(\varphi)(I_{pk})(\Delta\mathrm{c}_k) \qquad (1)$$

Where, I_{ek} (θ) = wanted ground station antenna gain towards k^{th} interfering source, P_{ik} = satellite transmitter output of k^{th} interfering carrier, I_{sk} (φ) = k^{th} interfering satellite antenna gain towards wanted ground station, l_{pk} = path loss of the k^{th} interfering carrier to wanted ground station and Δc_k = fraction of k^{th} interfering carrier power captured within wanted receiver's bandwidth.

For the *return* link the interference occurs at the satellite and hence the total interference I is estimated at the wanted satellite's input. The mobile interfering source may operate from anywhere within its service area – be it a spot beam on the same satellite or another satellite. For planning purposes it is usual to position the interfering source at the worst interfering location. Equation (1) is applied for the interference at the satellite receiver.

$$\mathrm{Ir} = \sum_{k=0}^{n} I_{ek}\,(\theta)\mathrm{P}_{ik}\; I_{sk}(\varphi)(I_{pk})(\Delta\mathrm{c}_k) \qquad (2)$$

Where, I_{ek} (θ) = wanted satellite antenna gain towards k^{th} interfering source, P_{ik} = transmitter output of kth interfering source, I_{sk} (φ) = k^{th} interfering source antenna gain towards wanted satellite, l_{pk} = path loss of the k^{th} interfering carrier to wanted satellite and Δc_k = fraction of k^{th} interfering carrier power captured within wanted receiver's bandwidth.

2) Estimate the spectrum required for a 350 spot beam satellite system capable of supporting seven-beam clusters when,

(i) Traffic is distributed uniformly such that each spot beam requires ten, 200 kHz channels.

(ii) Three clusters require 20 channels each, with traffic in other clusters uniformly distributed as stated in (i). State assumptions.
(Hint: Frequencies cannot be reused within a cluster)

(i) Number of clusters = 350/7 = 50
Spectrum required in each spot beam = 10x200 = 2 MHz
Spectrum required in each cluster = 7x2 = 14 MHz
Since spectrum in each cluster is reusable in others, total spectrum required = 14 MHz

(ii) Spectrum required in each 'hot' spot beam = 20x200 = 4 MHz
Spectrum required in a hot cluster = 7x4 = 28 MHz
Spectrum required in nominally loaded clusters =14 MHz, shared with hot clusters
Spectrum required by the system is then bound by the requirements of the hot clusters = 28 MHz
Comment: If the system has negotiated a total of 28 MHz of MSS spectrum, 14 MHz would remain unutilised in 47 out of 50 clusters. A more efficient approach would be to negotiate separately the additional spectrum for the hot spot beams, allowing other systems to utilise the top-up spectrum on a non-interfering basis.

3) State the difference between forward and return link interference analysis of a mobile satellite system. Demonstrate the difference algorithmically.

We will assume here that the interference analysis of interest here pertains to the service link as it is more problematic of the two links because of the wide beam-width of the user antennas that make them more susceptible to interference. We will address geostationary satellite systems. The concept can be extended to non-geostationary satellite systems by repeating the exercise at in time-steps to cover one constellation repetition cycle.

In the forward service link the interfering source resides in the space segment – i.e., spot beams on the same satellite when considering intra-satellite (spot beam) reuse or different satellite(s) when considering inter-satellite reuse. In either case the position of the interfering source remains fixed. The interference is estimated using equation (1) above in answer to question 1. The iso-interference contours are, therefore, clearly defined and remain unchanged.

In the return direction the interference occurs at wanted satellite's input by co-channel transmissions from interfering mobiles, as given by equation (2) above in answer to question 1. The offending mobile, can reside at any location within the offending beam, resulting in a temporal uncertainty in interference estimate. A probability distribution can thus be associated to interference based on the probability of presence of the interfering mobile at each location. For example, when considering a land-mobile system, the probability of interference from sea-covered areas is negligible and hence portions of the interfering beam in such regions can be discarded from the analysis. Similarly the wanted mobile can also be located probabilistically at any location within the wanted beam and the net interference can be expressed probabilistically using the joint probability of presence of wanted and interfering mobiles.

The algorithmic statements for the *forward direction inter-satellite interference* can be summarized as follows:

1. Calculate the offset angle θ between wanted and each interfering satellite (Figure 1).
2. Estimate wanted ground station antenna gain towards interfering satellite(s).
3. Estimate satellite antenna gain towards the wanted ground station for each interfering source.
4. Calculate power (I) of each interfering carrier received within the pre-detection bandwidth of wanted receiver.
5. Calculate wanted carrier power (C).

6. Calculate C/I_t, where I_t is the sum of the interfering power from all the sources. As an approximate rule. 20% of the total noise is budgeted for interference, which includes inter and intra-system interference.

The algorithm for the *forward direction inter-spot beam (i.e., intra-satellite)* interference can be summarized as follows:

1. Obtain the offset angle θ from each interfering beam towards wanted mobile (Figure 2, question 1).
2. Estimate antenna gain towards wanted mobile from each interfering beam.
3. Calculate power (i_k) of each interfering carrier received within the pre-detection bandwidth of wanted receiver taking into consideration antenna gain towards wanted mobile.
4. Calculate wanted carrier power (C).
5. Calculate C/I_t, where I_t is the sum of the interfering power from all the sources. As an approximate rule. 20% of the total noise is budgeted for interference, which includes inter and intra-system interference.

The algorithmic statements for the *return direction inter-satellite interference* are summarized as follows:

1. Calculate the offset angle θ between wanted and each interfering satellite (Figure 2, question 1).
2. Estimate antenna gain of interfering mobile at various judiciously chosen locations in its service area towards wanted satellite.
3. Estimate mobile antenna gain towards the wanted satellite for each location selected in step 2.
4. Calculate power (i_k) of each interfering carrier received at wanted satellite within the pre-detection bandwidth of wanted receiver.
5. Calculate wanted carrier power (C).
6. Calculate C/I_t, at each location where I_t is the sum of the interfering power from all the interfering carriers
7. Obtain C/I_t statistics

The algorithm for the return direction *intra-satellite* interference is as follows:

1. Obtain the offset angle θ_k from each location of each interfering beam.
2. Estimate gain of wanted beam towards each point selected in step 1.
3. Calculate power (i_k) of each interfering carrier from each beam and each location received within the pre-detection bandwidth of wanted receiver taking into consideration antenna gain towards wanted mobile.
4. Calculate wanted carrier power (C).
5. Calculate C/I_t, where I_t is the sum of the interfering power from all the sources.

4) Describe the differences in interference analysis of geostationary and non-geostationary satellite systems. Outline the salient features of interference model in each case.

For a geostationary satellite the path profile varies only due to a mobile's motion. The range and elevation angle remains invariant for practical purposes. For a static mobile, the path profile remains unchanged.

Figure 1, above shows the geometry of the interference model. The mobile, pointed at the wanted satellite, picks up interference from adjacent satellite offset at an angle θ. The interference level is calculated by summing interference power from all the interferers using standard link equations. Equations (1) and (2) above quantify the interference calculations.

The interference calculations for non-geostationary satellite must additionally take into consideration the motion of the satellite itself, which results in temporal variations in elevation and offset angles, path loss variations, and more dynamic propagation and interference environments as satellites rise and set. Moreover, handover between satellites or beams causes discontinuity in interference levels, due to abrupt changes in path profile and transmission frequency. Interference is therefore more involved than those of geostationary systems and is often calculated through computer simulations. Figure 3 below depicts a representation of a simulation model.

Figure 3: Entities of a simulation model for estimating interference in the service link of a non-geostationary satellite system (A sketch of figure 3.4)

The simulation model for interference analysis in the service link of an NGSO system comprises the following components:

- **Constellation,** comprising orbital parameters of each satellite with appropriate phase relationship between satellites
- **Each satellite's antenna pattern including side-lobe performance where necessary**
- **Service region**
- **communication traffic model**
- **mobile Earth station antenna pattern mask** including its side-lobe performance;
- **carrier parameters** (frequency, EIRP, voice activity, modulation scheme, pre-detection bandwidth and C/I tolerance, power control);
- **propagation model**
- **simulation granularity** in terms of time increments and number of Earth points
- **statistical data processing** (e.g., cumulative distribution of C/I versus geographical location)
- **numerical and graphical representation** of results

5) Suggest a methodology for long term spectrum forecast of a regional mobile satellite system. State its limitations.

Figure 4 below illustrates a flow chart of a generic methodology, which can be used to derive long-term MSS spectrum forecasts of a chosen service for a country, region or the world. The model can be modified on the basis of available or estimated data.

Figure 4 (Sketch figure 3.7)
A schematic of methodology for estimating long term spectrum forecast in a given region.

Spectrum requirements are categorized by the operational environment because of differences in traffic requirements and propagation conditions. For each environment and service, the target population is estimated; traffic penetration within the population gives an estimate of market size, which is weighed for loss to competition, other service offerings such as VSAT or terrestrial mobile systems, etc. The extent of penetration is influenced by the existing infrastructure, affordability of the target population, their attitude towards acceptance of new technology, exposure/experience with similar technologies and social trends such as the way people communicate during work or during leisure, internet penetration, etc. It is then necessary to estimate traffic generated by each terminal in the busiest hour, the holding time of each call for circuit-mode services and the average message length for packet-mode services. The busiest hour may be staggered if the service carries a mix of social and business traffic due to differences in their usage characteristics. The methodology is useful for early planning of a service, but its accuracy is sensitive to assumptions.

The total traffic carried for the circuit-mode traffic is,

$E = T_a N$ *Erlangs*

Where T_a = average traffic per terminal during busy hour and N = number of terminals. The total traffic for the packet mode is

P = [M + C] N Kbytes

Where M = average message length per user (Kbyte) during busy hour and C = coding overheads (Kbyte).

The total spectrum for each service is estimated individually and summed, taking into consideration the modulation efficiency, coding overheads, grade of service (for circuit mode) and permissible delay with packet retransmissions (packet mode), including network overheads, such as signaling, network test and support channels.

6) What are the propagation impairments common to all types of satellite communication systems? Discuss the implications of each on system design.

Satellite links are affected by the intervening medium causing changes to signal level, polarization, phase and noise contamination. These frequency-dependent effects are caused by the troposphere and the ionosphere. The troposphere consists of the first few tens of kilometers of the atmosphere characterized by clouds, rain and fogs; ionosphere is an ionized region which extends between ~ 80 and 1,000 km around the Earth.

The main sources of degradation in the troposphere are gaseous absorption in the atmosphere, absorption and scattering due to fog, cloud and rain, signal fluctuations due to atmospheric turbulence and depolarization due to rain. Tropospheric effects for MSS systems operating at the L-band are negligible compared to the shadowing and multipath loss.

In the ionosphere, signals undergo a variety of impairments including changes in polarization and rapid signal fluctuations known as scintillation. Scintillation causes rapid change in amplitude, phase angle and angle of arrival of signals and modifications to its time coherence properties.

In addition to these impairments, the received signals become contaminated by noise from extraterrestrial, man-made sources and water particles.

Tropospheric impairments

Gaseous absorption increases with frequency and peaks around 22.2 GHz due to water vapor absorption and near 60 GHz due to oxygen (see ITU-R Rec. 390-4). The absorption depends on temperature, pressure and humidity of the atmosphere as well as the elevation angle of the satellite. Absorption reduces with a reduction in the humidity and an increase in elevation angle. For example, in the frequency range of 1–18 GHz, one way gaseous absorption for 100% humidity at the zenith varies approximately in the range ~0.03–0.5 dB, which increases to ~0.35–5.7 dB at an elevation of 5°. These values become critical when propagation margins are low, as in MSS links and at higher frequencies.

Attenuation by hydrometers refers to attenuation caused by water particles existing in the atmosphere, such as fog, cloud, rain and ice, out of which rain produces the most significant attenuation through scattering and absorption mechanisms. Representative link margins for 99.50–99.95% link reliability are respectively of the order of 3–20 dB at 20 GHz and 6-30 dB at 30 GHz for the continental climate region of the USA; margins for 99.95 link reliability are too demanding for MSS links and therefore a reliability target of 99.5% is a realistic target, particularly so in MSS radio links because of the additional losses incurred due to shadowing and multipath.

To meet the specified link reliability throughout a year, it is essential to convert the annual p% rain-attenuation statistics to the worst month of the year, as certain months are the wettest (e.g. the monsoon season in the Indian subcontinent). Techniques for this type of scaling are well documented, for example, ITU-R Recommendation 581 provides a method of converting the worst-month statistics to annual statistics.

However, it has been observed that there is a notable yearly difference in statistics; variations in excess of 20% rms are possible.

Site diversity offers a solution to mitigate the effect of heavy attenuation at higher frequencies (i.e. generally, Ka band and above). It has been observed that intense rain cells have dimensions of the order of a few kilometers. Thus, if fixed sites are spaced several kilometers apart, it is unlikely that severe fading occurs simultaneously at both sites. By selecting or combining signals from two sites, it is therefore possible to reduce demands on spacecraft power and improve link reliability.

Although considerable measured rain attenuation data exist for geostationary satellite systems for fixed locations, much less statistics is available for non-geostationary satellites where elevation angle varies dynamically. Moreover only a limited work has been conducted to establish the effects of vehicular motion on rain fade statistics.

When a dual-polarized radio wave travels through rain or ice, some power from one polarization gets coupled to the orthogonal component as noise due to the anisotropic behavior of the traversing medium, causing impairments to the orthogonal component. Degradation is measured as cross-polar discrimination (XPD) or cross-polar isolation (XPI). It can be shown that XPD degrades with decrease in frequency at a given co-polar attenuation; and increase in co-polar attenuation at a given frequency.

Depolarization caused by ice occurs without accompanying co-polar attenuation; however, its magnitude is about 25 dB and therefore not very significant.

Attenuation due to clouds up to frequencies of 30 GHz is relatively insignificant for fixed links, but at higher frequencies, attenuation from clouds of high water content (e.g. cumulonimbus) becomes significant, for example, at 100 GHz the attenuation can be 4–5 dB, increasing up to 8 dB at 150 GHz.

Small-scale refractive index variations of troposphere (and ionosphere) cause signals to arrive at the receiver via different paths causing rapid signal variations, called scintillation, due to random phases and amplitudes of the multipath signal components. Tropospheric scintillation depends on the season, the local climate, and frequency and elevation angle. Its magnitude increases with frequency and a reduction in elevation angle. Degradation caused by tropospheric scintillation becomes noticeable above 10 GHz and is significant at K_a band and above. Scintillation can be accompanied with rain. Typical values of scintillation in K_a band are reported for a mid-latitude location to be of the order of 0.2–0.3 dB peak-to-peak level in winter, 1 dB in clear conditions of summer and 2–6 dB in some types of cloud while fade rates range from 0.5 to over 10 Hz.

A large number of propagation measurements have been conducted since 1970, mostly applicable to fixed geostationary satellite systems. Due to continuous changes in elevation angle of non-geostationary satellites, these results are not directly applicable to systems deploying non-geostationary satellite constellations but can be utilized by appropriate transformation. For radio link design it is necessary to combine attenuation due to path loss, rain, scintillation, gaseous absorption, shadowing and multipath as a function of elevation angle (when considering non-geostationary satellite). Feeder links use large antennas and are sited to avoid blockage, so shadowing loss and multipath variations are minimal. Figure 5 illustrates a hypothetical example of elevation angle variation for a LEO and a MEO satellite pass compared to the static elevation of a GEO satellite. The path loss in non-geostationary systems cases decreases as the satellite rises until the satellite is closest to the ground station, increasing again as the satellite sets and since link margin is a function of elevation angle the available link margin should increase in a system designed with a fixed worst-case margin (typically, the lowest elevation angle). It follows that the worst (‘promised’) link reliability depends on the minimum operational elevation angle; this rather obvious conclusion has considerable influence on the size and cost of a non-geostationary satellite constellation.

Figure 5 [Sketch Figure 3.10 (a)] A hypothetical example of elevation angle variation for LEO, MEO, GEO satellites

Ionospheric effects

The ionosphere causes rotation of radio wave polarization, propagation delay, refraction, variations in angle of arrival of signals, absorption, dispersion and scintillation. These effects are affected by the total electron content of the ionospheric path of signal and small scale irregularities (localized random patches) in the ionosphere.

For mobile satellite communications operating in 1–2 GHz band, polarization rotation, also known as the Faraday Effect and scintillation are dominant. The Faraday Effect is caused by the interaction of electromagnetic waves with Earth's magnetic field in the ionosphere. Circular polarized waves are consequently unaffected by Faraday Effect and therefore MSS service links use circularly polarized waves, as this eliminates the need for polarization tracking, whereas feeder links can use linear (or circular) polarization as they can incorporate polarization tracking without significant cost impact. The effect is generally predictable and hence can be compensated by rotating the polarization of transmitted waves (or receiver antenna) in an opposite sense.

Peak-to-peak signal-level fluctuations of up to 20 dB can occur at ~1.5 GHz. Hence, it is not surprising that scintillation cause outage to L and S band MSS links for several hours in a year, but fortunately, the onset of severe scintillation events is late in the evening, when satellite usage is quite low. Personal communication systems, which often operate close to threshold, are particularly susceptible. Even feeder earth stations in the equatorial regions can suffer occasional outage due to scintillation.

Ionosphere irregularities occur due to certain solar, geomagnetic and upper atmospheric conditions. Hence scintillation is affected by sunspot activity, position of the sun relative to location and hence time of day, latitude of location, season and magnetic activity.
The severity of scintillation is defined in a number of ways. A commonly used measure is known as the S_4 scintillation index, defined as the standard deviation of received power divided by the mean value of the received signal power

$S_4 = \sigma / \mu$

The fading period is variable ranging 1–10s in the GHz frequency range of MSS interest.

Due to large regional variability, the best estimate of scintillation is obtained from measured local data. In the absence of measurements, an appropriate model can be used (for example, see ITU, 1992). Some useful conclusions of ITU-R recommendations are summarized as follows:

- Measurements taken at frequency f can be scaled to another frequency according to the dependence, $f^{-1.5}$.
- Instantaneous fluctuation of a scintillation event can be approximated by the Nakagami probability density function, for estimating cumulative fade distribution.
- Power spectral density of scintillation events varies widely due to variations in drift velocity of refractive index irregularities; power spectra density slopes of f^{-1}–f^{-6} have been reported; a value of f^{-3} can be used as an approximation in absence of real data.
- Scintillation index s_4^2 varies as 1/cos(i) up to i ~ 70^0 where i is the zenith angle, and for lower values of i, variation is between cos(i) and $(\cos(i))^{1/2}$
- Seasonal and longitudinal dependence can be approximated as,

  $S_4 \propto \exp(-\beta/W)$

  Where β is seasonal and longitudinal dependent parameter and W is a location and day of year dependent constant.

- Measurements demonstrate that in an equatorial region (Hong-Kong) 1 dB peak to peak variation can occur for up to 5% of time in periods of high sun activity at 4GHz.
- The probability of simultaneous occurrence of ionospheric scintillation and rain fading in equatorial regions is relatively high especially in years of high sun spot activity resulting in differences to statistics. The occurrence probability of such events must be considered in the design of high reliability radio links, such as essential for critical safety applications.

7) The signal received on a mobile comprises a number of components, the magnitude of each depending on the local environment and mobile category – i.e., land, maritime and aeronautical. What are the factors that affect the magnitude of the components in each category?

Signals received on a mobile comprise various components, their presence depending on the surrounding medium:

- a direct or shadowed path
- diffracted components
- specular components caused by reflections from metallic or smooth objects such as the body of a car or a smooth sea surface
- diffused components caused by reflection and scattering of objects around a mobile.

The resultant signal r(t) at a receiver can be represented as

$$r(t) = \sigma(x)a(t) + s(t) + d(t) \quad (1)$$

The magnitude of each component is strongly influenced by the local environment around a mobile. Since such components vary randomly they are characterized statistically for practical purposes.

Factors that affect the magnitude of the components

The magnitude and behavior of each component depend on a number of factors:

- User terminal antenna characteristics: Extraneous components are picked up from the antenna and therefore wider the beamwidth larger is the pick-up from multiple paths.
- Speed of travel: The rate at which the signals fluctuate depends on the velocity of the vehicle. In addition frequencies are also affected by Doppler due to relative vehicular and satellite motion.
- User cooperation: In a co-operative environment the use attempts to get the best view of satellite and hence the probability of obtaining a direct unshadowed path is high.
- Elevation angle: In general higher elevation angel has lower probability of man-made or natural obstructions on the communication path.
- Environment dependence – there is a strong dependency on the physical environment around a mobile; the effect depends on the type of object (e.g. building, tree, sea surface, etc.,) their RF absorption and reflection properties and geometry with respect to a mobile.
- RF bandwidth – certain types of environments cause dispersion of radio signals

Depending on the interpretation of the propagation mechanism, a number of variants of statistical models have been proposed..

Land mobile channels

In an urban environment, the direct path remains shadowed for a considerable period by buildings and therefore the diffracted or scattered components dominate. The obstructions reduce for suburban areas and open areas such as rural and motorway. However, shadowing by trees begins to dominate in suburban and rural areas. There are vast differences in topography between geographically separate regions, countries, cities, etc. and therefore a general characterization of land mobile channel is problematic. And hence various interpretations to characterize the signal components have been proposed.

Maritime channel

The consolidated conclusions from investigations on maritime channels are that signal fades depend on elevation angle, sea conditions including wave height, slope and wind, receiver antenna characteristics (beam-width, side lobe and axial ratio), height of antenna above the sea and the structure of the ship. Other influencing factors include antenna pointing accuracy and polarization mismatch.

The magnitude of the specular component reduces with an increase in wave height and as satellite elevation angle and the transmission frequency increase; diffuse components dominate under rough conditions; fade depth depends on amplitude and phase difference between the direct and indirect waves.

The diffused component of the signal depends on the slope distribution of the sea-surface facets, the effects of which are RF dependent. Sea slopes range between 0.04 and 0.07 for rough sea conditions.

When specular reflection coefficient of the sea, rms height of the sea waves, RF, elevation angle and antenna gain pattern are known, it is possible to estimate the magnitude of specular components theoretically.

Similarly, with the knowledge of average scattering cross-section per unit area of sea surface, scattering angles, receiver antenna pattern and height of antenna above sea level, the magnitude of diffused components can be estimated.
Fade caused by specular components Fs reduces with an increase in wave height until the magnitude becomes negligible beyond a wave height of 2 m, while fade caused by diffused components increases linearly until saturation occurs. Fade depth reduces at a given elevation with an increase in antenna gain due to the progressive rejection of the multipath by high gain antennas. A similar effect is seen for a given antenna type, as the satellite elevation is increased – in this case multipath rejection increases with antenna elevation (i.e. as the antenna pointing steers away from the sea surface).

Aeronautical channel
The magnitude and characteristics of specular and diffused components depend on type and characteristics of the reflecting or scattering medium below the aircraft. Measurements demonstrate that the magnitude of multipath components is elevation angle-dependent; and fading is frequency selective, due to path delay associated with multipath components.
When the direct signal received at time t is $s_d(t)$, the reflected component can be represented as $s_r(t-\tau)$ where τ is the time delay of the reflected (or scattered) component whose magnitude depends on the geometry of the path, the reflection coefficient of the Earth's surface and the aircraft's antenna gain. The reflection coefficient depends on characteristics of the Earth below the aircraft - if the surface is sea, then the state of the sea and when land, whether the terrain comprises forest, built-up areas, desert, etc. Signal measurements on aircraft show that reflections and scattering from land is significantly lower than from the sea surface; furthermore, signals have elevation angle dependence over sea but not over land; as already mentioned, most studies have therefore concentrated on propagation behavior in flight paths over sea.

8) Explain the significance of each component of the received signal on system design differentiating between mobile terminal types where necessary.

Depending on the environment surrounding a mobile signals arriving at its antenna comprise direct and indirect components. The indirect signals contaminate the wanted signal causing random amplitude and phase fluctuations. The fluctuations, known as multipath fading, are usually flat but in an aeronautical mobile the indirect signals can be delayed resulting in frequency-selective fading. In a system using diversity more than one direct signal is received and combined constructively.

Shadowing and multipath influence system design in a number of ways:

- blockage affects link reliability;
- multipath causes errors in digital transmission;
- frequency-selective fading causes inter-symbol interference when signal bandwidth exceeds the coherence bandwidth of fading;
- Doppler frequency shift and jitter appear as noise causing detection errors;

The minimum elevation angle and the number of satellites simultaneously visible from user terminals set a bound on the radio link reliability of a system. The minimum elevation angle is directly related to the available link margin; the number of satellites simultaneously visible from a location and their relative separation provide an estimate of the diversity gain.

In a majority of MSS systems a margin is built in the radio link design to contend shadowing. The margin depends on the desired link reliability. Time and/or spatial diversity provide effective countermeasures. Unlike terrestrial systems where receivers can operate in absence of a direct signal (i.e. with indirect signals only), MSS receivers cannot do so, as the indirect signal level is too low.

The indirect signal may be specular when the signal is reflected off a smooth surface or diffused when received by scattering from objects such as trees in a land environment or sea surface in a maritime or aeronautical environment.

To select a MSS transmission scheme it is necessary to characterize the propagation channel for quantitative trade-off analysis. Various models have been developed, characterized as statistical, empirical and deterministic depending on the data generation and modelling approach. The relevant parameters include amplitude and phase characteristics of the specular and diffuse components; elevation angle dependence; and time-delay characteristics. Since the reception of extraneous signals depends on antenna radiation pattern, it is included it in the channel modelling process.

Consider the land mobile channel as an example. A number of variants of statistical models have been proposed, depending on the physical interpretation of the propagation mechanism illustrated in the examples below.

- When a direct line of sight is available, the probability distribution of the signal amplitude is Ricean, with the assumption that the component a(t) is constant; in-phase and quadrature components of the diffused signals are independent of each other and normally distributed with zero mean. In the absence of a direct component, the signal is assumed Rayleigh distributed with a log-normal mean; scattered signals are also assumed as shadowed.
- In a variation to the approach, only the direct path is assumed shadowed.
- In another variation scattered signals is specified as a composite of clear and shadowed components.
- In the environment-dependent approach, the probability distribution is a composite of a number of probability distribution. A Ricean distribution is assumed for the fraction of time when a signal is unshadowed and a Rayleigh distribution with log-normally distributed mean for the period when the signal is shadowed. The inclusion of environment dependence provides a powerful technique for quantifying system performance in a mixed environment.
- A further refinement of the environment-dependent approach is to represent channel behavior as a combination of a Markov process and a statistical model, where the state of a channel is described by a Markovian process and the signal variations within each state are modeled by a statistical model appropriate for the environment. The Markov process allows the signal to assume one of the M states with a probability, which only depends on its previous state. A number of authors have used this approach; their solutions differ in the number of states of M and the chosen probability distribution of signal within each state.

In a diversity scheme multiple direct signal paths are available. In a time diversity system, the information is dispersed in time with the assumption that the signal quality would improve after a suitably spaced time. In a spatial diversity system the signals are transmitted from separate satellites with the assumption that probability of simultaneous fading of both signals is lower than a single transmission. Both the schemes can be combined to provide more robustness. In both schemes the signal are combined to improve the signal quality.

9) Develop the transmission equation and explain its role in radio link design.

Transmission equation forms the basis of radio link design and optimization. It encapsulates the relationship between transmitted power, path loss, propagation loss, interference effects, modulation, coding, multiple access, received signal, diversity improvement and signal quality requirements, allowing simple trade-offs. More generally, it relates the received RF power at the destination to the RF power transmitted by the source taking into consideration the intervening distance, frequency and link margin to recover propagation losses.
Consider an isotropic radiator transmitting P_t watts. Since power from the radiator emanates equally in all directions, the received power flux density (PFD) at a distance d is given as,

$P_F = \frac{P_t}{4\pi d^2}$ W/m^2
For a directional antenna of gain G_t the PFD is given as,

$P_F = \frac{G_t P_t}{4\pi d^2}$ W/m^2

The quantity P_tG_t is known as effective isotropic radiated power ($eirp_t$), therefore

$P_F = \frac{eirp_t}{4\pi d^2}$ W/m^2 (1)
Expressing the equation logarithmically,
$P_F = (P_t + G_t - 10log4\pi - 20\log(d))$ dBW/m^2
Power received at the receiver input, P_r is then,
$P_r = P_F A_e$ W (2)
For a circular aperture A_e, the effective aperture area of the receiver antenna,
$A_e = \frac{\eta\pi D^2}{4}$ m^2 (3)
Where,
D = antenna diameter; η = RF collection efficiency
Substituting,
$P_r = \frac{eirp_t}{16d^2}\eta D^2$ W (4)
Since in practice antenna gain is available (rather than its area) let us express equation (2) in terms of generic antenna gain,

$$G_r = \frac{4\pi\eta A}{\lambda^2}$$

$or, A_e = \frac{G_r\lambda^2}{4\pi}$m^2 (5)
Substituting for P_F (see equation 1) and A_e (see equation 5) in equation (2) and including a margin l_m, to contend losses reduces the received power by $1/l_m$
$P_r = \frac{eirp_t}{4\pi d^2}\frac{G_r\lambda^2}{4\pi}\frac{1}{l_m}$ W (6)
Expressing logarithmically,
$P_r = eirp_t + G_r - 20\log\left(\frac{4\pi d}{\lambda}\right) - l_m$ dBW (7)
l_m represents a link margin to contend propagation and other impairments.

10) Explain the significance of various components of the transmission equation in a system context.

The significance of various components of transmission equation applied to the service link is summarised in the table below. In the forward link, satellite is the transmitting source and in the return direction it is the user terminal.

Transmission Equation	Parameter	Description	Context
$P_r = P_s + G_s + G_r - 20\log(4\pi D/\lambda) - l_m$	P_r	Received power level (dBW)	- Fundamental equation that determines viability and capability of a radio link - Depends on transmitter EIRP, link losses and receive antenna gain - Determines signal quality (E_b/N_o)
	P_s	Transmitter power (dBW)	*User terminal* - RF radiation effects on human health set an upper bound on handheld user terminal - Permissible size, cost, battery power set a bound on directional user terminal
	G_s	Transmitter antenna gain (dB)	*Satellite:* High gain with multi-spot beams is necessary to sustain MSS service link and enhance frequency reuse *User terminal:* - Omni-directional/low gain antennas are

			necessary for handheld and small mobile terminals - High gain antennas are necessary for portable, broadband systems - High gain tracking antennas are necessary for broadband fixed and on-move user terminals such as those installed on ships and aircrafts
$EIRP = P_s + G_s$	EIRP	Effective Isotropic Radiated Power (dBW)	A measure of effective power transmitted. *Satellite EIRP* Determines the throughput capability of forward radio link and permissible link margin Maximum limited by technical and commercial constraints *User terminal EIRP* Determines the throughput capability of return radio link and permissible link margin Maximum limited by safe radiation limit, cost, size and DC power source constraints
	G_d	Receiver antenna gain (dB)	Satellite - Antenna gain must be high enough to receive low power transmissions of user terminals User terminal - Antenna gain of user terminals govern the receiver sensitivity and hence user data rate
	$20 \log (4\pi D/\lambda)$	Free space loss (dB)	λ = Wavelength; D = distance between transmitter and receiver in same unit asλ. This loss is attributed to spreading of transmitted signal in free space.
	l_m	A margin included to counter various link losses dB)	Loss are caused by numerous factors e.g. shadowing, multipath, interference, modem imperfections, etc.;

CHAPTER 4

Q1.

(a) What are the channel degradations in a mobile satellite link which affect the performance of the modulation sub-system?

The main sources of degradation affecting the performance of modulation sub-systems in mobile communication links are:

- signal fades caused by environment and velocity dependent multipath, characterized by slow (< ~ 1 Hz) and rapid (~ tens of Hz) signal fluctuations and amplitude fluctuations (fraction of a dB to several dB peak to peak);
- signal fades caused by tropospheric effects, applicable at frequencies > ~10 GHz;
- signal phase fluctuations caused by local oscillators, multipath and Doppler jitter;
- very low carrier to thermal noise ratio due to radio link constraints;
- non-linearity arising in various system components such as those caused in mobile earth stations that use class-C amplifiers, feeder earth station and satellite owner amplifiers, etc.;
- large frequency changes attributed to Doppler associated with MEO and LEO that have to be adequately compensated.

(b) Outline the problems in a coherent demodulator associated to such degradations.

In coherent demodulation schemes, the carrier is recovered prior to demodulation. Carrier recovery circuits are susceptible to thermal noise, fading, and other impairments outlined above and therefore care has to be exercised in mobile earth stations. Furthermore, carrier recovery becomes more difficult as radio frequency or data rate increase in thermal noise limited, fading links.

Demodulation also requires symbol clock recovery for extracting the transmitted bit stream. The requirement becomes more stringent for higher level modulation schemes where it is essential to synchronize time, phase and amplitude of both I and Q channels. Again, the problem worsens as radio frequency and symbol rate increase in presence of channel impairments. From this viewpoint, schemes which have less stringent synchronization requirements are preferred when robust performance is important such as for signaling, although this advantage has to be traded-off against the lower spectral efficiency in comparison.

Q2

(a) Which modulation schemes are generally preferred in mobile satellite systems?

Modulation schemes which have less stringent synchronization requirements are preferred when robust performance is important such as for signaling, although this advantage has to be traded-off against the lower spectral efficiency in comparison to spectrally-efficient high level modulation schemes.

In general, constant envelope modulation schemes are preferred, as they offer a more robust performance. Of the constant envelope modulation schemes, FSK schemes are suitable for low bit rate transmissions, due to demodulator hardware simplicity desirable for applications like paging; they have a lower spectral efficiency than other schemes and are therefore unsuitable for higher bit rate transmissions. PSK modulation schemes have a near constant envelope, but exhibit discontinuity in phase, whereas constant phase modulation (CPM) schemes have a constant envelope with a gradual change in phase, which results in better side-lobe performance. Non-constant envelope modulation schemes were not used for MSS previously, but due to the pressing need of better spectral efficiency, multilevel schemes such as quadrature amplitude modulation (QAM) together with powerful convolution codes were introduced in wideband MSS systems in recent systems.

Common digital modulation schemes used in MSS are:

- binary phase shift keying (BPSK) and its variants such as aviation-BPSK (or symmetric BPSK);
- quadrature phase shift keying (QPSK) and its variants such as offset-QPSK (O-QPSK), aviation-QPSK, minimum shift keying schemes e.g. Gaussian minimum shift keying (GMSK);
- multi-level frequency shift keying and its variants.
- spread spectrum modulation in conjunction with CDMA;
- coded orthogonal frequency division multiplexing (COFDM) modulation scheme (used in radio broadcast systems);
- 16-QAM, 16-APSK and 32-APSK modulation represents the current preference for broadband communication on basis of standardization work in progress and evidenced by the techniques used in advanced operational MSS system. 16-APSK and 32-APSK are new entrants, yet to be established in MSS.

(b) Briefly outline the principle of each modulation scheme with the help of constellation diagrams.

Figure 1(a)-1(d) below shows constellation diagrams of BPSK, QPSK, 8-PSK and 16-QAM schemes. In a BPSK scheme, two states spaced 180° in phase are permitted; in a QPSK scheme the separation between the closest points is 90°; and in an 8-PSK scheme the separation reduces to 45°. In a 16-QAM scheme a state is decided by a combination of amplitude and phase.

In an amplitude phase shift keying (APSK) scheme the permissible phases are arranged in concentric circular rings at discrete angular distances to each other.

As the distance between the permissible states reduces, the signal becomes more susceptible to noise and hence although higher level schemes are more spectrally efficient, their power efficiency reduces.

Figure 1 (Draw a sketch of figure 4.7)

Figure 2 shows a simplified block schematic of a typical QPSK modulator, as an example of a hardware implementation. The input digital stream is divided into I and Q components. Each stream is fed into a multiplier, the other input of which is an I or Q carrier as necessary. The I and Q components are then summed to accomplish a QPSK signal; at the same time, undesired components produced in multiplication are cancelled out.

Figure 2 [A sketch of figure 4.5(a)] Hardware implementation of a QPSK modulator

(c) State the reasons to prefer these modulation schemes over others.

The modulation scheme should be such that channel impairments have minimal impact on the received signal; the scheme is spectral and power efficient; and demodulation of the signals is robust in presence of noise and channel impairments.

In general, constant envelope digital modulation schemes are preferred, as they offer a relatively robust performance in MSS channels. Of the constant envelope modulation schemes, FSK schemes are suitable for low bit rate transmissions where demodulator simplicity is paramount to minimize cost but they have lower spectral efficiency compared to others. PSK modulation schemes have a near constant envelope, but exhibit discontinuity in phase, whereas constant phase modulation (CPM) schemes have a constant envelope with a gradual change in phase, which results in better spectral efficiency (i.e. due to low spectral side lobes). Traditionally non-constant envelope modulation schemes were not used for MSS, but due to the pressing need of higher spectral efficiency, multilevel schemes such as quadrature amplitude modulation (QAM) supported with powerful convolution codes were introduced in wideband MSS systems.

The chosen modulation schemes are thus based on a trade-off between robust demodulator performance in presence of channel impairments, spectral efficiency and constrains imposed by power limited service links.

Q3

(a) Differentiate between coherent and non-coherent demodulation.

In coherent demodulation schemes, the carrier is recovered prior to demodulation. Carrier recovery circuits are utilized for recovering the amplitude and phase of the received carrier and utilized for demodulation. However such circuits are susceptible to thermal noise and fading therefore care has to be exercised in their design at the mobile earth stations. Furthermore, carrier recovery becomes difficult as radio frequency or data rate increase in thermal noise limited, fading links. More generally, carrier recovery can be problematic in the presence of multipath, low carrier-to-noise ratio and Doppler variations. Noise introduces phase error in the recovered carrier and timing jitters in the recovered bit stream.
Non-coherent demodulation schemes process the received signal without attempting to recover the carrier and as such are simpler, capable of rapid synchronization and are potentially robust in presence of channel impairments. Nevertheless, coherent detection of PSK signals gives a better power efficiency than non-coherent detection schemes in thermal noise limited conditions.

Additional information

As to which of these schemes are superior in a fading, interference-prone, Doppler affected environment – depends on impairment factors such as fading, Doppler shift, bit duration, interference, amongst others. However, in many operational systems designed to operate in low-Doppler, Ricean faded, geostationary environment, coherent demodulation schemes is common (e.g. see Feher, Kamilo, "A comparison between coherent and non-coherent mobile systems in large Doppler shift, delay spread, and C/I environment", JPL, Proceedings of the Third International Mobile Satellite Conference (IMSC 1993) p. 485-490 (N94-22735 06-32).
Liu, Jian; Kim, Junghwan; Kwatra, S. C.; Stevens, Grady H., 'Comparative study on the performance of power and bandwidth efficient modulations in LMSS under fading and interference', MILCOM '91 - IEEE Military Communications Conference, McLean, VA, Nov. 4-7, 1991, Conference Record. Vol. 1, Institute of Electrical and Electronics Engineers, Inc., 1991, New York, p. 257-261 (A92-54751 23-32).
Feher (1993) compares performance and implementation complexity of coherent and non-coherent QPSK and GMSK modulation/demodulation techniques in a mobile satellite system environment, including large Doppler shift, delay spread, and low C/I, and demonstrates that for large f_dT_b products, where f_d is the Doppler shift and T_b is the bit duration, non-coherent systems exhibit lower BER floor than their coherent counterparts. For significant delay spreads, e.g., greater than $0.4T_b$ and low C/I, coherent systems outperform non-coherent systems. The synchronization time of coherent systems was found to be longer than that of non-coherent systems.
Liu, Jian, et al, 1991 investigate error performance of various power and bandwidth efficient modulations for land mobile satellite systems under multipath fading and interferences. Computer simulations show that the performance of 16-QAM with differential detection was as good as that of 16-PSK with coherent detection and 3 dB better than that of 16-PSK with differential detection, although it degraded by about 4.5 dB as compared to 16-QAM with coherent detection under an additive white Gaussian noise (AWGN) channel.

(b) Suggest at least two techniques of carrier recovery in a coherent demodulator.

Examples of carrier recovery techniques include:

- $(C)^M$ method, where C is the carrier signal and M denotes number of symbols
- Costas loop [or, Decision feedback loop]

In the $(C)^M$ method, carrier is recovered by passing the received carrier, f_c, through a circuit which raises the carrier level to the power of M to give a carrier at a frequency of Mf_c, followed by a divide by M circuit.
The Costas loop, named after its inventor, consists of a two-phase lock loop (PLL) which uses a common loop filter and voltage control oscillator (VCO). The VCO frequency is kept synchronized to the carrier frequency through a feedback loop comprising a multiplier, whose two inputs are outputs of an in-phase (I) and a quadrature phase (Q) comparators. The multiplier output provides a corrective voltage to maintain the VCO synchronized to the carrier (see Figure 3).

Figure 3 (Sketch Figure 4.3 here) Schematic of Costas Loop method of carrier recovery.

(c) With the help of block schematics describe the functioning of QPSK modulator and demodulator.

Figures 4 and 5 respectively show block schematic of a typical QPSK modulator and demodulator, respectively.
At the transmitter, the input digital stream is divided into I and Q components. Each stream is fed into a multiplier, the other input of which is I or Q carrier as necessary. The I and Q components are then summed to accomplish a QPSK signal; at the same time, undesired components produced in multiplication are cancelled.

The signal is divided into two paths at the receiver, each of which is fed into the respective I or Q channel multiplier, the other input of which consists of the recovered carrier. The output of each multiplier is low pass filtered and fed in to an analogue to digital converter synchronised to the incoming bit stream through the timing recovery circuit resulting in the recovered bit stream. BPSK modems operate in the same manner as QPSK but only with one arm.

Figure 4 [Figure 4.5 (a)] Schematic of a QPSK modulator
Figure 5 [Figure 4.5 (b)] Schematic of a QPSK demodulator

Q4

a) What are the strengths and limitations of OFDM applied to a broad-band mobile satellite system?

Digital modulation schemes suffer in quality in a mobile environment. For broad-band systems the fading can be frequency selective causing inter-symbol interference. The quality worsens as transmission bit rate is increased to a level where the bandwidth approaches and exceeds the coherence bandwidth.
The OFDM modulation system overcomes the problem of frequency-selective fade by partitioning the incoming data in to segments, which are each coded, modulated and frequency division multiplexed with carrier frequencies orthogonal to each other. The segmentation results in transmission of a fraction of information in each RF carrier. The consequent reduction in bandwidth drops the bandwidth to less than the coherence bandwidth and hence eliminates inter-symbol interference. The system is used primarily for mobile broadcast system.

The main limitation in the use of OFDM arises in L and S band systems is due to the limited available bandwidth of (shared) ~40 MHz in these bands, making it difficult to transmit wideband data where the frequency-selective fades become noticeable (except for aeronautical channels - see answer to the next question) . The technique can be applied at higher frequency bands where there is adequate signal bandwidth for transmissions of true broad-band (tens of Mbps) when frequency-selective fades can set a bound on achievable BER.

b) What are the limitations in the use of the scheme in MSS systems, noting that the scheme has been widely proposed for mobile satellite broadcast systems?

There is no fundamental limitation in using the technique for MSS system except the availability of enough bandwidth where OFDM can provide an advantage. The OFDM does not offer noticeable advantage to the L and S band satellite systems because the available bandwidth is limited due to congestion in the band. The coherence bandwidth of MSS channels is generally more than the bandwidth of the signals in use in a majority of cases.
Aeronautical MSS channels are an exception where the frequency-selective fades begin to impact transmission in excess of ~100 kHz bandwidth. An OFDM scheme is an option in this environment when transmissions requiring bandwidths in excess of 100 kHz are required with medium gain antenna which exhibit relatively low discrimination from ground.

c) Explain the process of OFDM generation and reception.

Generation
Figure 6 shows the conceptual building units of a COFDM transmitter. Serial input data are converted into a

parallel stream each of which modulates a carrier from an orthogonal set which are finally summed to constitute a composite COFDM signal.

Figure 6 (Sketch of Figure 4.15) Conceptual building units of a Coded OFDM transmitter

In practice, the boxed part is implemented in software by fast Fourier transform digital signal processing (DSP) chips. Mathematically, the composite signal can be expressed as

$$S_s(t) = \frac{1}{N}\sum_{n=0}^{N-1} A_n(t)e^{j[\omega_n t + \Phi_n(t)]} \qquad (1)$$

where $A_n(t)$ and $\phi_n(t)$ are the amplitude and phase of the n^{th} carrier and $\omega_n = \omega_0 + n\Delta\omega$, ω_0 being the angular frequency of carrier 1.
The digital equivalent of the signal at instant kT can be represented as,

$$S_s(kT) = \frac{1}{N}\sum_{n=0}^{N-1} A_n e^{j\Phi_n} e^{j(n\Delta\omega)Kt} \qquad (2)$$

For $\Delta f = 1/NT = \tau$, equation (2) becomes equivalent to an inverse Fourier transform where s(kT) is the time domain representation of the signal. A DSP chip performs an inverse transform on the incoming signal to give sampled time domain signals which are converted to an analogue signal for transmission. To facilitate Fast Fourier Transform (FFT), the number of carriers N is made equal to 2^n.

Reception
A reverse process is applied at the receiver. The received signal is synchronized, digitized and Fourier transformed to the frequency domain to provide individual carriers with the desired amplitude and phase. A practical difficulty is that of carrier synchronization. One solution is to use a coarse synchronization, followed by a precise synchronization. A coarse synchronization can be achieved by switching off all carriers regularly for a short duration; an amplitude detector can be used to provide a synchronization pulse when carriers are switched on. Fine synchronization can be achieved by transmitting a reference signal which can be correlated at the receiver with a replica to achieve an accurate synchronization in time and frequency. Synchronization can be made robust at the expense of power efficiency by introducing guard intervals around each symbol, during which trivial or redundant data is transmitted. Guard intervals offer an additional advantage of reducing the effects of echo or co-channel interference, when the delay of interfering signals is small compared to the symbol period. A guard band of the order of 25% of the symbol period has been observed to be a reasonable compromise for terrestrial environments in the UK.

d) Outline the transmission and reception methods of direct-sequence and frequency-hopped spread spectrum systems.

Figures 7 and 8 illustrate the principle of the direct sequence spread spectrum scheme. The message is modulated using a nominal modulation scheme such as QPSK; the modulated signal is spread by a spreading function (code); the spread signal is up-converted, amplified and transmitted. At the receiver, the down-converted signal is correlated with a replica of the transmitted code. A correlation peak is obtained when codes match; the resultant signal is band pass filtered and demodulated to obtain the data stream. The transmitted signals are channel coded to improve performance.

Figure 7 (Sketch of figure 4.17a): Principle of direct sequence spread spectrum transmitter.
Figure 8 (Sketch of figure 4.17b): Principle of direct sequence spread spectrum receiver.

In a frequency-hopped spread spectrum system, the signal is modulated channel coded and transmitted such that the transmission frequency is altered in a pseudo-random sequence. A pseudo-random chip generator changes the frequency of a synthesizer used for up-conversion. At the receiver, the signal is down-converted by a synthesizer, synchronized to the transmitted signal frequency and demodulated/decoded using an appropriate technique. The transmission and reception methods are illustrated in Figures 8 and 9.

Figure 8 [Figure 4.18(a)] Principle of frequency-hopped spread spectrum scheme transmitter
Figure 9 [Figure 4.18(b)] Principle of frequency-hopped spread spectrum scheme receiver

Q5

a) **Differentiate between convolutional and block codes. Which of the two methods should be preferred for a narrowband mobile satellite system for applications such as messaging, machine-machine communication, etc.?**

Block codes operate on groups of bits organized as blocks, i.e. information bits are assembled as blocks before coding (see Figure 10).

Figure 10: A block coder schematic (sketch figure 4.19(a)).

Convolution codes are formed by convolving information bits with the impulse response of a shift register encoder (see figure 11, next question).

The narrowband systems of the type mentioned should require only a few kilobytes of intermittent data transfer. A block coder with an ARQ scheme provides a reliable solution. Convolution coders require the shift registers to be cleared at the end of each message by feeding a string of zero, since this incurs overheads convolution coders are inefficient for coding short messages.

b) Outline the principle of a typical convolution coder.

Convolution codes are formed by convolving information bits with the impulse response of a shift register encoder (see figure 11).

Impulse response of the encoder is defined as the encoder response when a single 1 followed by 0s are entered in the encoder. Figure 11 depicts a conceptual diagram of a non-recursive convolution coder. An input bit progressively moves down a shift register on arrival of each subsequent bit and dropped at the end of the final stage. The output of certain stages of shift register are combined (denoting code memory) in v exclusive OR adders such that v output bits are produced for each incoming bit.

Figure 11: A non-recursive convolution coder (sketch figure 4.19(b)).

c) Outline the principle of interleaving and state at least two scenarios where such a scheme can be applied advantageously.

Interleaving is an effective countermeasure for correcting errors caused by noise bursts. Information bits are dispersed in time such that consecutive bits of a message are shifted apart such that an error burst can cause an error only to one or a few bits which can be error corrected by an underlying coding scheme. Provided that the separation, called *interleaving depth*, is greater than the duration of a noise burst, the technique mitigates the impact of the error burst. Thus the technique requires an accurate characterization of the propagation channel for an effective implementation.
It is implemented by arranging shift registers as a matrix of x rows and y columns. The incoming data stream fills the register by row while transmitting it column-wise after coding each column. The separation between adjacent bits in a column is y bits thus contending noise burst of at least (y-1) bit duration increasing to (ny-1) bits if n error correction code is applied to the columns. The process introduces a delay in data transfer since transmission and reception can only occur after the interleaver has filled up

The performance of conventional coding techniques degrades in presence of noise bursts lasting longer than the error correcting ability of the code. In MSS links signals fluctuate widely due to multipath, ionospheric and tropospheric scintillation, fading due to rain, etc. An interleaver with sufficient depth (chosen on basis of channel behavior) provides an effective countermeasure.

d) Suggest at least two scenarios where code concatenation can be used advantageously.

In addition to error bursts MSS links suffer random errors due to low carrier to noise ratio, Ricean fading, etc. Their performance can be improved by cascading two or more codes – one for correcting random errors and the other to contend error bursts. This coding arrangement, called *concatenation*, can provide high coding gain with moderate complexity; typically, a block code such as the RS code (outer code) is cascaded with a convolution code (inner code).
Consider a motorway route. The majority of a typical route is likely to exhibit a Ricean probability distribution where a convolution code would suffice. However, in certain sections of the motorway such as junctions, tree-lined areas, etc. signals may fluctuate, causing error bursts. A concatenated scheme comprising a convolution coder for contending Ricean fading on clear parts of the route with an interleaver to account for other sections should provide an effective countermeasure.

Another scenario where concatenation could be applied effectively would be a railway channel where the channel is well-behaved with a periodic dip in the signal due to cable trellises.

Q6

a) **Explain the functioning of turbo coder and decoder**.

The answer is available on pages 200-201 of the book

b) Compare the main attributes of turbo and LDPC codes.

Turbo code (TC) and Low Density Parity Check (LDPC) code provide near optimum performance when block size is large (e.g., thousands of bits) and can be implemented on DSP chips at a reasonable cost. Both the codes utilise simple encoders which can produce long code words with good distance properties. While turbo coder uses a combination of an interleaver and convolution encoders, LDPC coder generates a parity-check matrix that has a low density of ones in relation to the block size. Both the schemes utilise low-complexity iterative decoding to produce near-optimum performance. The choice therefore depends on the requirements while addressing issues such as channel characteristics, target code rate and permissible encoding complexity; some general observations for thermal noise limited links are as follows:

- LPDC can lower the error floor compared to TC when block size is large. However TC offers a lower complexity at very low coding rate; TC was therefore preferred in the 3GPP2 standard as well as DVB-SH and SES SDR satellite multimedia broadcast standards.
- LDPC perform better than punctured TC at higher coding rates.
- TC can be generally encoded faster than LDPC codes, but there are exceptions.
- LPDC can achieve higher throughputs than TC with parallel decoding architectures.

Q7

a) Which accessing schemes are best suited for transmission of: (i) continuous streams such as voice (ii) Bursty traffic such as a bank transaction. State reasons for your choice.

The following accessing schemes are well suited for transmission of continuous streams:

(i) Frequency division multiple access with single channel per carrier, TDMA and reservation protocols are suitable for continuous streams. Such schemes provide a dedicated or virtually dedicated circuit for the duration of transmission that is essential for voice communication.

(ii) Aloha and slotted Aloha are suited for bursty traffic. Such schemes do not hold up resources in absence of data transmission and are thus suited for communication needing sporadic bursts of activity.

b) Explain, with reasons, the factors that influence efficiency of multiple accessing schemes for MSS.

The following factors influence the efficiency of multiple accessing schemes.

- RF interference: A certain amount of interference is permitted in radio link design (termed ‘intentional interference’) to facilitate spectrum reuse. In this context, schemes which offer higher interference rejection capability are advantageous.
- Voice and data activity: The average occupancy of transmission is about 40%. Similarly, certain types of ‘continuous’ data exhibit pauses. Therefore average interference power reduces when carrier suppression is used during speech and data pauses.
- Variations in traffic mix: MSS systems provide wide-area coverage which cover areas with different types of communications, for example, data usage may be prevalent in developed countries of the service area. Moreover the mix can change and evolve over time. Therefore, a multiple access scheme that allows for traffic mix and adaptability increases net system throughput.
- Propagation effects: The MSS propagation environment is characterized by multipath and shadowing. Accessing schemes which offer resistance to multipath and fading (including frequency-selective fading) impairments can increase utilization of radio resources.

However, it is observed that spectrum efficiency has to be considered in conjunction with others e.g. modulation, coding, inter/intra-system interference allowance.

Q8

a) Outline the principles of accessing schemes used in MSS.

All three basic multiple access schemes – frequency division multiple access (FDMA), time division multiple access (TDMA) and code division multiple access (CDMA) - are used in MSS for *circuit-mode* transmissions. A variety of accessing schemes - often called 'protocols' - are used for *packet-mode* transmissions, which are often characterized by bursty traffic (i.e., sporadic traffic bursts) so that dedicated channels are wasteful of resources.
MSS systems combine a number of multiple accessing schemes to match conflicting requirements caused by different transmissions characteristics coupled with practical considerations. For example, a request for a channel assignment is sent effectively through a short data burst, whereas voice communication requires a dedicated connection. Thus, an effective solution would to combine a data access protocol such as Aloha for the request channel with a demand assigned scheme such as FDMA or CDMA that allocates a dedicated circuit for each call.

Frequency Division Multiple Access (FDMA),
In the FDMA scheme, the available spectrum is partitioned into a number of segments. Each station is allowed to transmit in one or more segments according to communication needs. The concept is illustrated in Figure 12.

Figure 12 [A sketch of Figure 4.24] A DA/SCPC/FDMA frequency assignment scheme, portraying channels of various types and their occupancy with time

When spectrum blocks are pre-assigned, the scheme is known as *fixed assigned* (FA) and when the channels are allocated dynamically in response to requests, the scheme is a *demand assigned* (DA) FDMA. In the latter case, when each segment consists only of a single channel, the FDMA scheme is known as a DA single channel per carrier (SCPC) FDMA scheme (or DA/SCPC/FDMA in shorthand) . DA schemes are better suited when traffic requirement per user is low typical of MSS links. Fixed assignment is suitable when communication demand is high or for continuous broadcasts (e.g. transmission of bulletin board). In an MSS environment, fixed assignments are used for communication between gateways, network broadcasts and signaling, such as call initiation.

Time Division Multiple Access (TDMA)

In a TDMA scheme, users access the satellite in non-overlapping time bursts, as illustrated in Figure 13. The reference burst consists of a carrier bit recovery (CBR) field for carrier and timing recovery, a unique word (UW) field for burst synchronisation and a control (C) field for station identification. The reference burst is followed by non-overlapped time bursts that carry traffic. The sequence is repetitive; where the reference burst synchronises each frame. Each station transmits within its designated time slot.

Figure 13 [A sketch of figure 4.25] A TDMA scheme, comprising a reference burst, traffic burst encapsulated in repetitive frames

The time slots may either be fixed assigned or demand assigned. Demand assigned TDMA is better suited than fixed assigned TDMA for MSS. TDMA network management is complex due to the need to maintain the earth stations in time synchronization. Synchronization of mobile terminals in fading conditions and Doppler coupled with a low carrier-to-noise environment requires care.

Time division duplex (TDD) is a variant of TDMA. In a TDD scheme, a time slot is used for both transmission and reception, in effect doubling the slot capacity. Time delay in each direction must be low to make the scheme effective. This is achievable in LEO systems where time delays are of the order of a few tens of milliseconds.

Code Division Multiple Access (CDMA)

In CDMA, all users access the satellite without restriction within the full band. All users can co-exist simultaneously when spread spectrum modulation is used, since each user is assigned a unique code, which has low cross-correlation with codes used by others. The user is able to extract transmissions destined for it by correlating the received signal with a replica of the code used at the transmitter. The correlation function peaks when the codes match, whereas all other transmissions are rejected.

The direct sequence CDMA scheme is affected by the so-called *near–far* problem, which occurs when interfering transmissions are received at a higher level than the wanted transmissions, resulting in excessive BER or even a loss of signal. The effect, however, tends to be lower in frequency-hopped schemes, as the probability of co-channel interference is low. The mechanism of self-interference in these two schemes differs. In the direct sequence scheme, at any instant interference from all the active users is present and noise-like, whereas in the frequency-hopped system interference power spectral density is momentarily high (coherent) when the frequency of the interferer falls within the band of the wanted carrier. When the hopping rate is much smaller than the information rate, the interference is coherent, but intermittent; on the other hand, when the hopping rate is high, the interference tends to become noise-like. One of the issues that needs care in a frequency-hopped system is the necessity of maintaining the carrier phase noise within tolerance.

Data Access Protocols

Several types of data traffic tend to be intermittent and therefore dedicated channel assignment is inefficient. They are characterized by high activity but sporadic bursts which would be wasteful if a dedicated circuit were assigned. Satellite resource efficiency can be improved if channel assignments are made on per message or packet basis, taking into consideration the permissible time-delay. Schemes range from random access on one end of the scale to reservation TDMA on the other. Random access schemes do not require network coordination and are therefore simple; typical applications are registration of a mobile to the network or a channel assignment request. The most commonly used random schemes in MSS are Aloha, Slotted Aloha and to a lesser extent Reservation Slotted Aloha. In the Aloha scheme, the user transmits a data packet whenever necessary; in Slotted Aloha the packet transmission is constrained within a time slot; this requires network time synchronization; in Reservation Aloha, a user continues transmission for as long as it has reserved the channel, as negotiated with the network; at the end of transmission, the slot is open to contention. Reservation TDMA is useful when large amounts of data must be transferred.

b) Compare the efficiencies of the classical multiple accessing schemes as applied to MSS for power and bandwidth limited links, taking advantage of voice activity and spatial reusability. State assumptions in the quantitative analysis.

Spectral efficiency is defined as the bits/s per RF cycle, i.e. bits/s/Hz.

The spectral efficiency η_s of a spread spectrum system is given as:

$$\eta_s = [C/(N_0W_s)]/ (Eb/N_0')[1 + (C/N_0W_s)(M - 1)/M] \quad (1)$$

Where C = Total received carrier power from M earth stations, M = number of participating earth stations, N_0 = Total thermal noise power density, W_s = RF bandwidth = $1/T_s$, where T_s = chip duration, E_b = Received energy per bit of information= C/T_b, where T_b = information bit duration, $N_0' = N_0+I_o$ where N_0 = User noise power density, I_o = interference power spectral density

When interference from other systems exists, the total carrier power C in the above equation must be increased to (1+K)C, where K represents interference from other system.

Consider some refinements:

- V**oice activity** in each direction in a typical telephone conversation is about 40% of call duration. By using voice-activated transmission, therefore, the interference level is reduced by a factor of 2.5, increasing the CDMA capacity by about the same factor.

- In a system deploying s**pot beams**, the isolation between beams where frequencies are reusable is significantly lower in the CDMA than in the FDMA scheme. It is feasible to reuse the same frequencies in adjacent beams. This is possible because contributions to N_0' from users in adjacent spot beams can readily be absorbed in a spread spectrum system and interference from spot beams further away becomes progressively lower.

- In FDMA or TDMA systems, **frequency reuse by polarization discrimination** is difficult due to the poor polarization discrimination possible in mobile antennas and polarization reversal caused in a multipath. The CDMA system offers the potential of frequency reuse in an opposite polarization. An increase in capacity of about 60% is possible with a modest antenna isolation of 6 dB.

Introducing the improvements made possible by the factors above, the total noise received can be modified as follows:

$N_0 + I_0 = N_0 + a\rho v\ (N-1)\ (E_bR_b)/W_s$ (2)

Where, a = 1/number of spot beams (uniformly distributed traffic distribution), ρ = polarisation isolation factor, v = voice activity factor, N= number of users, E_b and R_b are information energy per bit and bit rate respectively and W_s = occupied RF bandwidth

The efficiency η_s is given as,

$\eta_s = (C/N_0W_s) / v[E_b/(N_0 + I_0)](1 + a\rho\ C/N_0W_s)$ (3)

An improved CDMA technique called band-limited-quasi-synchronous CDMA (BLQS-CDMA) tackles the problem of self-noise to offer the advantages of CDMA with efficiencies comparable to orthogonal accessing schemes, i.e., TDMA and CDMA. The scheme requires chip clock and carrier frequency synchronization using the information embedded in the CDMA signal structure in the form of a master code and Nyquist pulse shaping of signature chips, giving spectral compactness without compromising on detection performance. The spectral efficiency of QS-CDMA is r bits/chip, where r = coding rate.

In a thermal noise-dominated link, equation (3) also applies to FDMA and TDMA. For these schemes, C is the total EIRP of the satellite and Bt the total RF bandwidth.

(a) Power limited case
$\eta_p = (C/N_oB_t)/(E_b/N_0)$, when $MR_b/B_t < G_l\ r\ \log_2(m)$ (4)
(b) Bandwidth limited case
$\eta_b = G_l\ r\ \log_2(m)$, when $MR_b/B_t > G_l\ r\ \log_2(m)$ (5)
Where G_l = guard band loss for FDMA and guard time/preamble loss for TDMA, r = code rate, m = PSK constellation dimension, R_b= user information rate, M=number of participating users.

Again, extending the formulation to a spot beam MSS environment and including voice activity advantage, the efficiency for each case is given as,

(a) Power limited case
$\eta_p = \alpha\ (C/N_oB_t)/v(E_b/N_0)$, when $NR_b/B_t < G_l\ r\ \log_2(m)$ (6)
(b) Bandwidth limited case
$\eta_p = \alpha\ G_l\ r\ \log_2(m)$, when $NR_b/B_t > G_l\ r\ \log_2(m)$ (7)
Where
α = frequency reuse factor
v = voice activity factor

Given the applicable system parameters, it is possible to make a like for like comparison. An example has been cited in the book it has been demonstrated that given the following parameters,

c) Compare the applicability of CDMA and TDMA for a medium bit rate mobile satellite system and suggest your own preference. State the assumptions in making the comparison and in your assessment.

Selection of a multiple access scheme for MSS requires a numerous considerations, according to the emphasis placed by the system designers. For example, the TDMA scheme was chosen in preference to the CDMA scheme for the proposed ICO, whereas Globalstar designers preferred a CDMA system. Another study conducted in Italy comparing CDMA and TDMA concluded that neither was distinctly superior to the other (Priscoli and Muratore, 1996, see book for reference). The authors took two specific systems, the

mobile satellite business network (MSBN) and a satellite extension of the global system for mobile communications (GSM), and compared radio and network aspects of the network.
We assume a mobile system for communication on the move. L and S bands are suitable for this type of service. We present the rationale used by ICO for selection of TDMA in favor of CDMA.

- Wideband measurements show that frequency selective fades affects bandwidth beyond 5 MHz and hence the fade resistance advantage in favor of CDMA is not realizable within the narrowband available (This assumption is based on the limited spectrum allocation in L and S bands): .
- On the basis of an earlier investigation, TDMA was considered more efficient (Meidan, 1994, see book for reference).
- TDMA can benefit from satellite diversity and soft handover, as much as CDMA.
- Satellite links are power limited, and self-interference in CDMA links reduces link margins as the system approaches full capacity; the use of orthogonal CDMA increases capacity, but due to differential path delay achieving signal orthogonality is difficult when satellite diversity is used.
- In TDMA, use of diplexer can be avoided if transmit and receive bursts are arranged to occur in different time slots, whereas diplexers cannot be avoided in CDMA due to the need for continuous transmission and reception. Diplexers tend to add losses in the front end, thereby reducing a receiver's sensitivity.
- The inaccuracy introduced in power control loop due to propagation delay is likely to affect significantly the capacity of an asynchronous CDMA.
- Large interference into a CDMA signal can cause outage to all users when such interference occurs.
- Due to wideband characteristics of CDMA, the probability of interference outage in CDMA is higher than in narrowband TDMA, where interference can be counteracted by reassigning the affected call to an interference-free channel.
- TDMA can manage non-uniform traffic distribution in a spot beam environment better, as it allows the peak capacity in a beam to be increased and switched between spot beams.
- TDMA disadvantage vis-à-vis CDMA in terms of return link power can be reduced by using non-linear amplifiers at the mobile whereas CDMA requires linear amplifiers.
- Spectrum sharing between CDMA systems can be problematic, as wideband transmissions from one system will affect the capacity of the other, whereas in TDMA the band segmentation traditionally used has minimal impact on the capacity of other operators.

CHAPTER 5

1) Briefly describe the hardware entities and functions of a gateway.

Please see section 5.2 of the book.

2) Outline the characteristics of various types of antennas used in MSS user terminals and assess their suitability for various types of mobile platforms and environment.

Antennas used in L-band mobile terminals, together with their characteristics, for L-band are summarized in table 1 below. The use of K_u and K_a band enables smaller antennas in the 'high-gain' category. In an operational MSS, the antenna gain and side-lobe performance can be specified generically allowing manufacturers to optimize their terminal in the manner best suited to them. Thus antenna gain is specified in terms of EIRP (transmit) and G/T (receive), while antenna side-lobe performance is defined as a generic mask.

Category	Antenna tracking	Gain (dBi)	G/T (dB/K)	Antenna type	Suitability
High gain	Step-track	15-21	-10 to -4	Parabolic dish on ships or micro-strip array for land portables with mechanical or electronic steering.	High data rate service for ships and land portables
Medium gain	Fixed or tracking	4-15	-23 to -10	Short backfire, phased array, helical, micro-strip	Voice and medium speed data for land portables, aircraft, and ships.
Low gain	Fixed	0-4	-30 to -23	Dipole (drooping), quadrifilar, microstrip	Hand-held voice service, low data rate applications on ships, aircrafts, etc.

Table 1: Front-end characteristics of L-band user terminals

Common mobile antennas include crossed dipole antennas, helical antennas of various types, microstrip patch antennas, phased arrays and parabolic dishes. Parabolic dishes are generally used where space is not a major consideration, such as on large ships. Phased array antennas are used where minimal aerodynamic drag, reliability and high tracking speed is necessary. They are suitable for aeronautical and high-speed land vehicle and railway installations. Crossed dipole, helical and patch antennas are used for land mobile and maritime communications.

The *crossed dipole* is made of two half-wavelength antennas placed at right angles to each other and fed with equal amplitude signals, which are $\pi/2$ apart in phase necessitating a power divider and a 90° phase shifter for feeding the two dipoles. The antenna transmits in circular polarization with a near omni-directional pattern in the azimuth plane and an elevation angle pattern which has a maximum in the Z-axis or 90° elevation. By bending the dipoles and adjusting their distance from the ground plane, it is possible to tilt the boresight angle in any desired direction for optimizing the dipole to operate at any specific elevation. Figure 1 is a diagram of such a dipole. The bandwidth of this type of antenna is relatively narrow.

Figure 1 [sketch of Figure 5.3]: A crossed-dipole antenna.

The *helical antenna* consists of a number of turns of wire wrapped around a dielectric material or in air, mounted over a ground plane as shown in Figure 2

Figure 2 [A sketch of figure 5.4]: A helical antenna.

The antenna operates with circularly polarized waves in the Z-axis direction (at right angles to the ground plane) and exhibits a wide bandwidth of about 200% and a medium gain. The gain and beam width of the antenna are proportional to the number of turns of the helix. A quadrifilar antenna comprises four helical antennas spaced equally around a cylinder. Typically, the ground plane has three times the diameter of the cylinder. The antenna exhibits a much wider bandwidth than a single helical antenna. The four helixes are fed with equal amplitude signal with phase shifts of 0, 90, 180 and 270°, which make the arrangement relatively complex.

Antennas etched on microstrip, known as *patch antennas*, are useful when a low profile is essential, e.g. for vehicle mounting. Figure 3 depicts a single circular patch antenna for producing circular polarization.

Figure 3 [A sketch of Figure 5.5] A circular patch antenna

The patch can be excited in basic or higher order mode from two feed points. The resonance frequency of the patch varies inversely with the radius of the circular patch and the relative dielectric constant of the substrate. For a substrate of dielectric constant 3, the patch diameter is 5 and 10 cms at 1 and 2 GHz, respectively. Higher order excitation can be used to maximize gain at any given elevation angle. Microstrip technology is suitable for mass production and offers a potentially low-cost solution for personal communications.

Phased arrays comprise an antenna array in which the amplitude and phases of exciting signals are varied electronically to steer the main lobe. The boresight of a far field pattern depends on the amplitude and phases of excitation, which may be changed in response to a tracking error signal to affect an enormously agile and reliable tracking system.

The design of *land mobile antennas* is the most challenging as they must be compact, low profile and low cost. In general, antenna gain, multipath, blockage, polarization characteristics and antenna noise temperature of low profile antennas degrade as the elevation angle is reduced. Therefore, antenna gains are over-designed at higher elevation angles, or alternatively, their gain is maximized in the elevation angle range where the operational satellite is likely to appear within the service area, a condition which suits a regional geostationary system well.

The gain of a low-profile antenna in the direction of the satellite depends on the effective projected area in the direction of the satellite and therefore varies as $\sin(\eta)$, where η is the elevation angle. In array antennas, commonly used in medium–gain terminals, electrical boundary conditions do not support transmission of circularly polarized waves close to the horizon. Furthermore, the antenna axial ratio and gain degrade at low elevation angles. For example, a low-profile antenna of two wavelengths in size has a maximum gain limit of 7 dBi and an axial ratio of 11 dB at an elevation of 15°.

Antenna performance is sensitive to vehicle structure and the height of the antenna above the ground plane. Reflection and diffraction, which depend on the curvature and effective area of the ground plane on the vehicle roof, cause ripples in the elevation antenna pattern. The magnitude of the ripple decreases as the elevation angle is increased and is more pronounced for large beam width antennas.

The antenna noise temperature of low-profile antennas depends on the elevation angle and environment around the antenna. Measured antenna temperature at elevation angles of 30–60° range between 30 and 50° K when there is a clear line-of-sight, which increases to 50–85°K in the presence of the wooded skyline (Milne, 1995, see the reference in book). The system noise temperature is ≈ 200° K when diplexer and LNA units are mounted close to the antenna or made an integral part of the antenna assembly.

Portable antennas require maximum gain in the broadside direction without necessarily needing to have a low profile, whereas vehicle-mounted antennas must be low profile, compact and preferably tracking. Microstrip radiators offer a low-profile solution. It is possible to achieve near omni-directional patterns and gains of 2–6 dBi from a single microstrip radiator, while medium gain can be achieved by an array of radiators.

Hand-held terminals require an omni-directional pattern in the azimuth within the elevation angle range of interest, e.g. 5–90° for geostationary satellites and 30–90° for a non-geostationary satellite system and typical gains of 0–1 dB with an axial ratio < 5 dB. Quadrifilar helix and patch antennas can meet such requirements.
The interaction between head and antenna is an important consideration in antennas used in hand-held terminals. Studies demonstrate 10 dB or more attenuation due to obstruction from the head and so the antennas are designed to operate above head.

3) Discuss the tracking options for mobile platforms installed on ships and passenger aircrafts.

Ships are characterized by slow motion and large size amenable to V-SAT size antennas and, as such, tracking systems can be simple with slow response when operating with geostationary systems. Aircrafts are fast-moving, the mounting space is more restrictive than ships and need of a low profile to minimize aerodynamic friction. The tracking systems should therefore be agile.

Either mechanical or electronic steering with phased array is used in aircrafts. Phased arrays offer a low-profile solution giving high steering speed, adaptability to different environments, reliability and lower power consumption than their mechanical counterparts. Phase shifters used in phased arrays are lossy and therefore increase the antenna noise temperature which has to be compensated by increasing their size. An increase in the number of phase shifters such as PIN diodes tends to increase the insertion loss and cost. Low-loss mechanically steered antennas hold an advantage in respect of cost. Typically such antennas comprise a planar microstrip array which is mechanically steered in azimuth by low-profile stepper motors. For regional geostationary satellite systems, the elevation pattern can be shaped to the desired elevation angle range.

Step-track is a simple method where the tracking receiver maximizes the received signal level by stepping towards it using small exploratory steps. The technique is unreliable in the presence of signal fluctuations and is characterize by a low response time and hence suitable for ships and can operate with geostationary satellite systems and MEO satellite systems.

Open-loop techniques are not susceptible to signal fluctuations as they derive error signals from external sensors such as a flux-gate compass and gyrocompass or a vehicle's inertial navigation system and mathematically predicted satellite position from the given location. The error signals are derived by taking the difference between the current and previous estimated position of the satellite coupled with knowledge of the current vehicle position. In a moving vehicle the tracking system compensates for the vehicle motion in addition to the satellite motion.

Open-loop tracking systems generally require some form of closed-loop tracking for initial acquisition. Combining both tracking methods provides a higher level of accuracy and reliability.

The open-loop technique is established in aeronautical systems, as inertial sensors onboard aircraft are available, except in some smaller planes and helicopters where they are difficult to interface. In such cases, (low-cost) three-dimensional sensors are necessary.

4) With the help of block schematic describe the main features of : (a) handheld UT (b) large ship earth station (c) aeronautical earth station

(a) Handheld UTs are described in sub-section 5.3.2
(b) Features of large ship earth station are discussed in sub-section 5.3.3.1, with an example in part (a) of the sub-section
(c) Features of aeronautical earth stations are covered in subsections 5.3.3.3

5) Discuss the trade-offs in designing a portable multimedia UT of the type described in this chapter

We will consider an L-band portable UT of the type suitable for the BGAN network. The UT antenna size and amplifiers of such a portable multimedia UT are chosen so that the EIRP, G/T, and linearity requirements of the system are satisfied within the cost target that has to be constrained by the limited production volume anticipated for the service. The UT includes capabilities to decode dynamically changing modulation-coding schemes; manage the associated radio resource management functions; and support power saving features e.g. sleep mode. Furthermore, the UT should incorporate standard interfaces like Bluetooth, USB, LAN and WLAN; position location functions using GPS and/or other navigations systems; efficient voice processing algorithm; and robust synchronization and acquisition techniques to counteract channel impairments.

The antenna size is limited to 10-12 cms for consistency with the target UT size with a need to support both transmit and receive bands. A single patch antenna provides about 7.7dBi gain and results in a G/T of approximately –18.5dB/K. This is adequate to support BGAN's adaptable air interface that supports dynamic power variation of the order of 15 dB (Satellite EIRP ranges ~25-40 dBW towards the mobile for the BGAN system). The EIRP of the terminal is estimated as 10 dBW to support the target throughput of a few hundred kbps so that the high power amplifier can be sized at around 2 watts.

Since the production volumes are low in comparison to cellular phone, development of new chip-set are not economically viable and hence the design has to be based on commercially available chip with a provision for system-on-chip or ASIC design. Thus the hardware would comprise an (analogue) RF part with a digital baseband. To enable indoor operation there should be a provision to attach an external antenna. A GPS (or equivalent) receiver is integrated in the receiver to support various network functions; additionally the location feature is necessary to support the legal requirement imposed in many countries and the system includes provision for open-loop timing and location information. Accurate location information also facilitates location-based billing arrangement. However, it is recognised that due to the limited link budget of GPS in comparison to the communication per se (BGAN in this example) and possible obstructions around a user, the location accuracy may be compromised in unfavourable conditions and this may impact initial acquisition of the receiver when a mobile is switched on.

To comply the linearity requirements, efficiency of power amplifiers has to be compromised resulting in higher heat dissipation. Since cooling surface is limited, efficient heat dissipation is, therefore, essential. Features to minimise power drainage in the UT include - avoiding unnecessary transmission and reception when there is nothing to receive and to use sleep mode wherein all but the bare minimum subsystems are switched off. With typical battery capacity of the order of a few Ah the average drainage should be expected to be of the order of a few tens of mA and hence the stand-by time would be 24-48 hours.

The software should be able to take into consideration the propagation delay of the order of 250 ms and include features to log-on with minimum exchanges. To support IP suite including the internet and applications such as VoIP and multicast, performance enhancement proxies and quality of service settings the UT must also include an efficient application software platform. Since the system is compatible with 3G core network, the terminal supports USIM card so as to support roaming between terrestrial and satellite networks.

CHAPTER 6

1) Outline the generic system requirements of an MSS satellite. How do these translate into payload specifications?

Major system requirements that drive the payload architecture are listed below.

Operational frequency: At present, L and S bands are prevalent; K_u band secondary allocations is used sparingly for MSS and for a majority of FSS mobile very small aperture terminal (MVSAT) systems; K_a band MSS-MVSAT systems are a recent introduction;

Deployment of enhanced spot beam technology is mandatory to facilitate efficient use of radio resources, provide desired forward link EIRP and return link G/T with on-board capabilities to alter beam shape, size and position by telecommand in response to needs of the mission;

Spacecraft service link EIRP and G/T determine user capability;

Efficient spectrum utilization is mandatory to mitigate congestion;

Dynamic distribution of spacecraft resources, i.e., EIRP and spectrum across the service area is essential to respond to changes in traffic pattern;

Full eclipse operation is desirable for LEO, MEO and international GEO satellite systems where traffic load extends past midnight; regional GEO system may not operate at full power during an eclipse as traffic tends to subsides by the time an eclipse occurs.

Propagation delay: For low-delay terrestrial-like services or a fault-tolerant/survivable space segment architecture, low range and space-segment redundancy is necessary - LEO or MEO systems are preferred in this respect.

These generic needs translate to the following payload requirements:

Transponder

- Available spacecraft EIRP must be high enough to serve present and future traffic up to the end of spacecraft life; this necessitates efficient high power amplifiers with redundancy, narrow reconfigurable/steerable spot beam antennas and a power generation system to comply.
- Flexible real-time power and bandwidth allocation between spot beams to permit allocation of space segment resources dynamically (i.e. traffic driven).
- Benefits of onboard processing technology should be assessed.
- Suitability of intersatellite links to enhance network connectivity should be assessed

Antenna system

- Spot beam technology is essential in the service links to enhance frequency reuse, reduce mobile EIRP by enhancing spacecraft G/T and make effective use of spacecraft transmitter power
- Only a few large fixed Earth stations are necessary in the feeder links and therefore a single or a few spot beams are enough to meet capacity and frequency coordination requirements in the feeder radio link.

General

- Technology amenable to mass production is essential for supporting large non-geostationary earth orbit (NGEO) constellations.
- Higher integration of payload components such as output combiner with antennas, band pass filters with LNA, solid state power amplifier stacks, etc. is essential to reduce weight and power drainage, facilitate mass production, improve reliability, etc.
- High DC power to meet the large power requirements of the payload, necessitating large and efficient

solar arrays and power system.

Eclipse operation is a bus requirement but is included here as it is critical to LEO and MEO MSS systems. The number of eclipses is large for LEO and MEO satellites due to their orbital geometry, unless the orbital geometry minimizes or eliminates eclipses. Spacecraft batteries undergo a considerable number of charge/discharge cycles due to eclipse, which reduces the lifetime of the batteries and hence useful life of satellite.

2) Compare the advantage and disadvantage of a transparent and regenerative transponder. Develop the relationship between the uplink and downlink carrier to noise ratio and demonstrate the trade-offs between the two parts of the link stating assumptions.

Advantages of regenerative transponder compared to transparent transponder	Disadvantages of regenerative transponder compared to transparent transponder	Comments
Better interference resistance Higher capacity Flexible and dynamic routing at message level Better network interconnectivity Reduces terminal size Permits optimisation of up and down links independently Provides better access security	Increases payload complexity Heavier payload Larger power drainage Difficult to change signal format, accessing schemes, modem/codec etc. with current technology	Extent of disadvantages anticipated to reduce as technology matures. Reconfigurable payload allows flexibility in changing waveform

Further observations:

In a regenerative transponder, the received signal is regenerated to baseband at the satellite thereby decoupling the noise components in the up and down links; the total BER (derived in a later paragraph) is the sum of bit error probability of up and down link resulting in a signal quality improvement compared to the transparent transponder. In a conventional transponder, the uplink and downlink noise power get added at the user terminal (consider forward link) and therefore the bit error probability is multiplicative.

When the up-link (i.e., feeder link) and down-link (service link) carrier to noise ratios are of the same order of magnitude, a regenerative transponder shows a clear advantage. This is the case for MSS return link where mobile terminals transmit very low EIRP levels. The advantage is relatively less significant in the forward link but nevertheless between 2 and 5 dB improvements are feasible. The advantage of regenerative transponder increases significantly when non-linearity of the satellite channel is included; the resulting advantage in the link margin can be traded off against interference and hence regenerative transponders increase system capacity in an interference dominated environment.

The decoupling property of regenerative transponders separates the impact of fading in up and down links. Links can be optimised separately to give several decibels power advantage and hence an increase in system capacity.

It has been estimated that, under average operating conditions, improvements of the order of 8 dB are possible by regenerative transponders over conventional transponders in the presence of severe uplink rain fading. Under realistic fading conditions, improvements of the order of 10–12 dB have been demonstrated.

A regenerative transponder with baseband processing permits reformatting of data and therefore the uplink and downlink multiple access schemes can be different and optimized to suit each part. For example, in a

TDMA system operating through a conventional transponder, the forward link burst rate is limited by the mobile earth station's G/T and demodulator performance. A regenerative transponder eliminates such a limitation by allowing the downlink accessing scheme to be better matched to mobile Earth station capability, e.g. by grouping channels into a manageable set of TDM streams at a rate commensurate with the mobile earth station's capability, while leaving the feeder station the flexibility to transmit in TDMA.

A regenerative transponder offers the possibility of traffic routing, switching and onboard processing. Note that a satellite with an on-board-processing satellite need not be regenerative. For example an on-board processor can operate an RF or IF switch in a TDMA systems to provide coarse interconnectivity. In a regenerative satellite the switching is instead performed at message and packet level. Messages/packets from destined for a given beam, can be combined and routed to the appropriate beam to improve resource utilization and network connectivity.

Radio resource management can be transferred on-board, allowing a dynamic and flexible sharing of resources to improve resource productivity.

Relationship between uplink and downlink carrier to noise ratio

<u>Transparent transponder</u>

The system comprises an uplink Earth station, u, transmitting signals to an Earth station, d through a transparent transponder T. The following noise components are received at the destination Earth station, d:

N_u = uplink noise power at satellite receiver input

N_{ui} = uplink interference noise power at satellite receiver input

N_{sim} = satellite intermodulation noise power referred to satellite receiver input

N_{di} = downlink interference noise power at Earth station d receiver input

N_d = Earth station d's system noise power at receiver input

N_t = Total noise power at Earth station d receiver input

C_s = carrier power received at satellite receiver input

C_d = carrier power received at Earth station d receiver input

G = gain from satellite receiver input up to Earth station u receiver input

$G = g_s (1/l_d).g_d$

Where

g_s = gain of transponder from satellite receiver input up to satellite antenna output

l_d = downlink loss

g_d = Earth station d gain up to its receiver input

$N_t = N_d + N_{di} + G (N_u + N_{ui} + N_{sim})$

$C_d/N_t = C_d/ \{(N_d + N_{di}) + G (N_u + N_{ui} + N_{sim})\}$

Or,

$N_t/C_d = \{(N_d + N_{di}) + G (N_u + N_{ui} + N_{sim})\}/ C_d$

$N_t/C_d = N_d/ C_d + N_{di}/ C_d + GN_u/C_d + GN_{ui}/C_d + GN_{sim}/C_d)$

Noting that, $G/ C_d = C_s$

$$(C_d|N_t)^{-1} = (C_d|N_d)^{-1} + (C_d|N_{di})^{-1} + (C_u|N_u)^{-1} + (C_u|N_{ui})^{-1} + (C_u|N_{sim})^{-1}$$

<u>Regenerative transponder</u>

In a regenerative transponder the concept of total noise at Earth station d receiver input cn be replaced by probability of error. The demodulation at the satellite receiver decouples the noise in the uplink. The demodulated steam with errors (probability P_u) associated to the demodulation is re-modulated and transmitted. There is thus a probability P_uP_d that uplink error is 'corrected' during the demodulation at Earth station, d. The net probability of error P_t is therefore,

$P_t = P_u + P_d - P_uP_d$

The term P_uP_d is negligible in practice and therefore,

$P_t \cong P_u + P_d$

Since bit error rate is an estimate of probability,

$BER_t \cong BER_u + BER_d$

Trade-off between the up and down link

The trade-off refers to transparent and regenerative transponders. Figure 1 compares behaviour of uplink and downlink carrier to noise ratios for these two classes of transponder for the same link quality assuming a linear channel without coding.

Figure 1 (Figure 6.3(a) sketch here)

Relationship between uplink and downlink carrier to noise ratio for conventional and regenerative transponders

Figure 2 (Figure 6.3(b) sketch here)
Relationship between BER and E_b/N_0 for conventional and regenerative transponders

When the up-link and down-link carrier to noise ratios are of the same order of magnitude, a regenerative transponder shows a clear advantage. This is generally the case for MSS return link when mobile terminals can transmit very low EIRP levels. Note that the advantage is reduced as uplink carrier to noise is increased which would imply that the advantage is not notably significant in the forward link. The advantage of regenerative transponder is significantly increased when non-linearity of the satellite channel is included; non-linearity is introduced primarily at the power amplification stages of earth station and satellite. Between 2 and 5 dB gains are possible even in the presence of high up link carrier to noise ratio density (see figure 6.3(b) of the book). The advantage in the link margin can be traded off against interference, in an interference limited situation to increase system capacity (see figure 6.4 of the book).

To reiterate the decoupling property of regenerative transponders gives resistance to uplink and downlink signal fading. Further, up and down links can be matched separately giving an advantage of several decibels in power resulting in an increase in overall system capacity.

3) Illustrate with a block schematic, the core protocols of the RSM-A architecture and its relationship with other functional components of an MSS system.

Figure 3 below presents an implementation of RSM-A architecture – showing the distribution of IP network, packet control protocol, and packet lower layer functions across satellite, Network Operations and Control Center (NOCC), and terminal segments with respect to the core RSM-A functions. The core RSM-A functions relate to physical and link packet networking; the ground based NOCC provides addressing, QoS signalling, and multicast support for making network level decisions which are communicated over IP protocols to the end user networks designated as subnet 1 and subnet 2.

Figure 3: (Sketch of figure 6. 9) RSM-A core functions and a system instance.

The regenerative transponder of the satellite SPACEWAY® 3 is based on the RSM-A standard for K_a band. The satellite includes features such as as spot beam forming, dynamic beam mapping, and adaptive resource control. The regenerative payload using on-board packet processing provides full-mesh, single-hop connectivity between two or more terminals without a terrestrial support. The on-board processor routes IP packets to their destination and provides advanced broadband IP networking services including support for quality of service needed for voice and video applications and on-board bandwidth-on-demand for IP traffic.

4) Plot a graph to illustrate the variation in the antenna diameter as a function of altitude in the range 700-1500 km and 10 000 and 14 000 k m, given:
Antenna efficiency = 60%
E_b/N_o = 5 dB-Hz and 10 dB-Hz

We make the following assumptions:
User terminal - handheld omnidirectional
EIRP, E_h = 0.5 W (-3 dBW)
Supported data rate, r = 9.6 kbps
Satellite receiver noise temperature, T = 300°K

Altitude and satellite antenna pattern determine a satellite's field of view. Figure 4 depicts the geometry.

Figure 4 (A sketch of Figure 6.13): Geometry for estimating spacecraft antenna.

The carrier to noise power density ratio (C/N) received at the edge of coverage by a user terminal utilizing an omnidirectional antenna is obtained using equation (6.2) of the book, derived as follows:

Referring to figure 4, power flux density (PFD) at satellite altitude H from hand-held terminal using omni-directional antenna and positioned at beam centre is,

$PFD = E_h/4\pi H^2$ (1)

The signal power received on satellite by an aperture antenna of diameter D is,

$C = PFD.A_{eff}$ (2a)
$A_{eff} = \eta A$ (2b)

Where,
η = antenna efficiency at the given frequency
A = Antenna aperture area
$= \pi D^2/4$

Substituting in (2),

$C = (E_h/4\pi H^2)\ \eta\ (\pi D^2/4)$ (3)
For a user terminal placed at the edge of coverage, where gain = ½ gain at beam centre
$C = (\eta/32)E_h(D/H)^2$
Or,
$D = H \text{ sqrt } [(32C/E_h\eta)]$ (4)

When E_b/N_o, N_o, and transmission bit rate, r, are known, C can be obtained as follows:

$C = rE_b$, $N_0=kT$
Since, $C/N_o = r(E_b/N_o)$
Or $C = kT\ (rE_b/N_o)$ (5)

Substituting (5) in (4)
$D = H \text{ sqrt}[32kT\ (rE_b/N_o)/E_h\eta)]$

Returning to the numerical problem, the given values are:

H = 700-1500 km and 10000-14000 km

$\eta = 0.6$

E_b/N_o = 5 dB-Hz and 10 dB-Hz

$C = E_b.r$, where r=bit rate

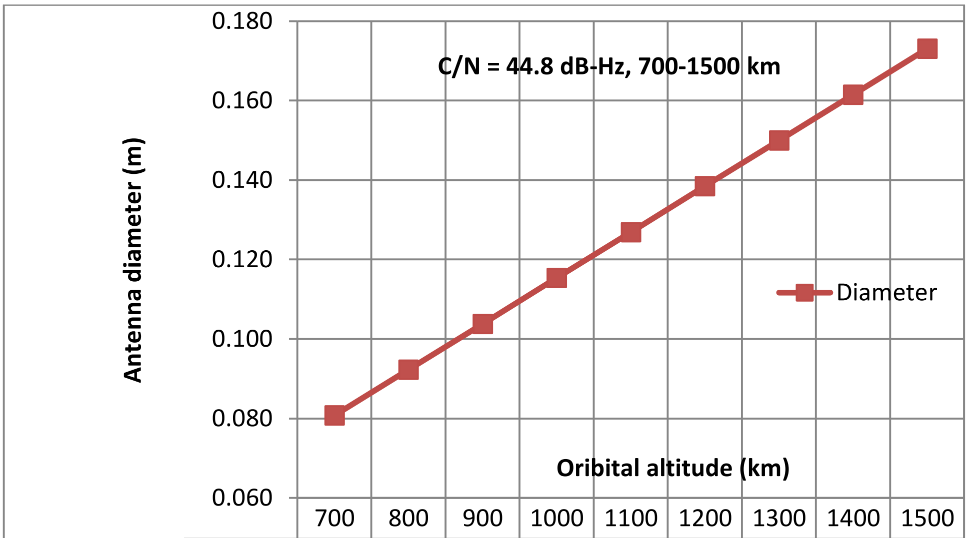
Typical value of r for hand-held is 2.4-9.6 kbps, we will assume 9.6 kbps [The student can use other values, consistent with hand-held user terminal capability]

C/N_0 (dB-Hz) is, therefore, E_b/N_o + 10log(9600), i.e.,

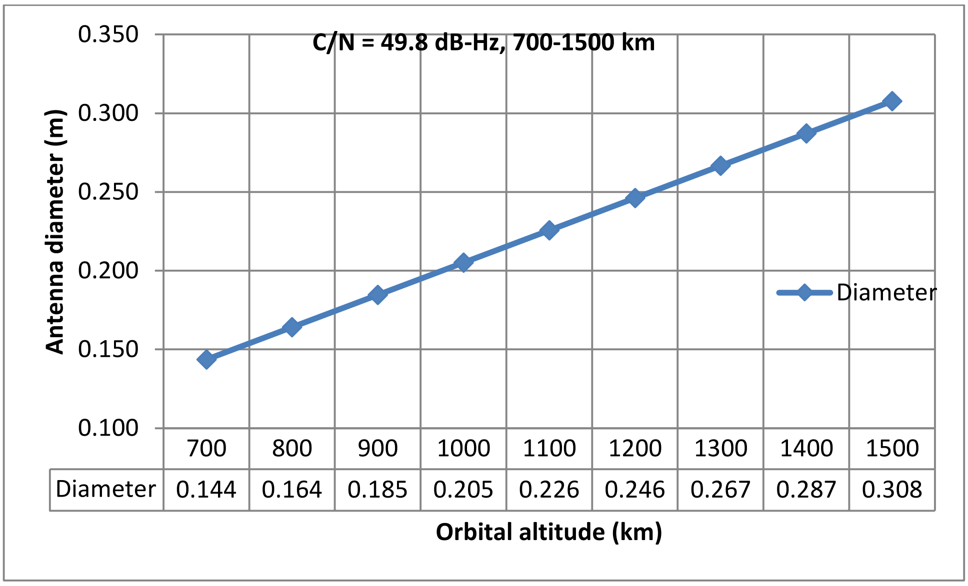
C/N_0 = 44.8 dB and 49.8 dB

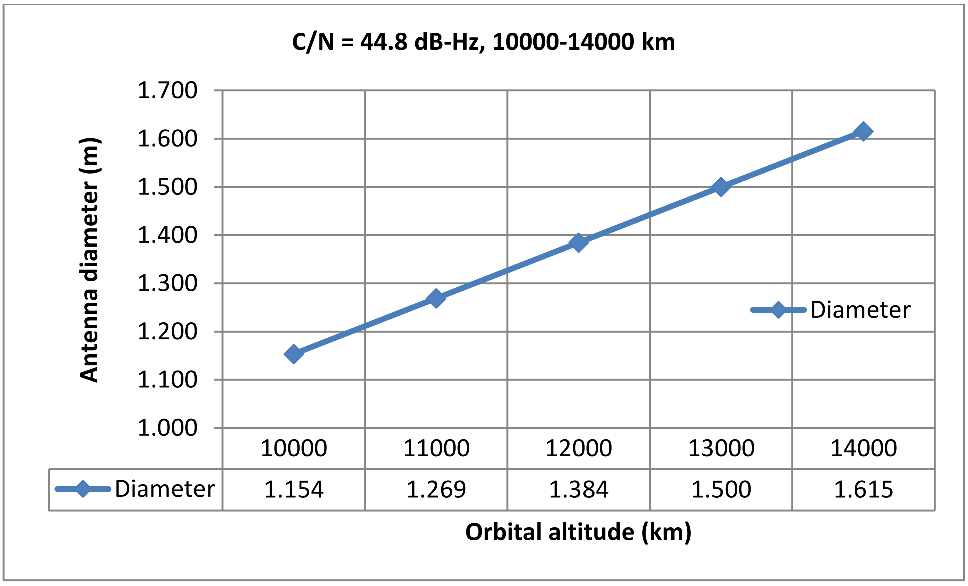
Typical value of E_h for handheld = 0.5 W (- 3dBW)

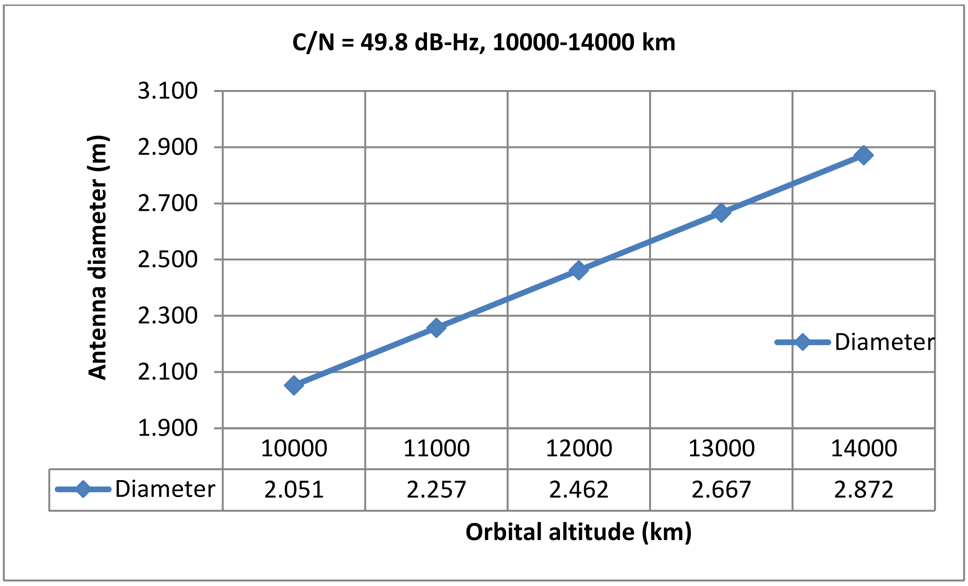
Substituting, the following plots were produced in a spreadsheet.



	700	800	900	1000	1100	1200	1300	1400	1500
Diameter	0.081	0.092	0.104	0.115	0.127	0.138	0.150	0.162	0.173


C/N = 49.8 dB-Hz, 700-1500 km
Antenna diameter (m)
0.350
0.300
0.250
0.200
0.150
0.100
Diameter
700 800 900 1000 1100 1200 1300 1400 1500
Diameter 0.144 0.164 0.185 0.205 0.226 0.246 0.267 0.287 0.308
Orbital altitude (km)


C/N = 44.8 dB-Hz, 10000-14000 km
Antenna diameter (m)
1.700
1.600
1.500
1.400
1.300
1.200
1.100
1.000
Diameter
10000 11000 12000 13000 14000
Diameter 1.154 1.269 1.384 1.500 1.615
Orbital altitude (km)



5) Compare (i) the launch sequence of Iridium and Globalstar systems and (ii) their manufacturing methodology for mass production satellites

Launch sequence comparison with data available in the book.

[Requires further research to fill in the missing information, e.g. (Globalstar-2, 2013)]

	Iridium	**Globalstar** [see website referenced below for Globalstar second generation mission profile (Globalstar-2,2013)]
Launch vehicle	Delta II	Three launch vehicles - First two launches on Delta II; next three on Zenit; final three on Soyuz
Number of satellites per launch	Groups of three	Delta II: Groups of four Zenit: Groups of 12 Soyuz: Groups of four
Jettisoning altitude (km) - first satellite	638	Delta II: 1250 Zenit: 920 Soyuz: 920
Time to jettison first satellite after launch (s)	3130 (remaining follow at 200 s interval)	Not available (N/A)
First radio contact	~1 h 40 m	N/A
Number of tracking stations	Four (Hawaii, Yellowknife and Iqualit in the Northwest territory of Canada, and Snjoholt in Iceland)	N/A
Drift time for satellite	Several days	N/A

separation		
Deployment sequence and time	Solar array (to avoid battery exhaustion), followed by communication antennas (ground and inter-satellite communication); 3 hours into mission secondary antennas deployed; switch from secondary to main antennas in fifth orbit	Delta 2: First four satellites, held in a canister, are placed in the same orbital plane. The initial few maneuvers are initiated by each satellite autonomously using onboard computers; maneuvers include extension of magnetometer boom, acquisition of the Sun and the Earth, stabilization to avoid tumbling, and deployment of solar arrays to avoid battery depletion. Zenit: Satellites are held in canisters (as with the Delta launch), and ejected within 4 s in rapid succession. Initial maneuvers are identical to Delta launch. Satellites are at a lower orbit than Delta launch; travel faster to give only 10–12 minutes of visibility from ground stations. Only six satellites can be placed in each plane, the satellites are grouped and injected in three separate planes. Satellites experience different radiation and thermal conditions at 920 km to those experienced at 1,400 km, for which they are optimized. Hence, satellites must be moved to the higher altitude as soon as possible. Soyuz: N/A
Initial check out	First two days: Satellites' primary antennas and modems are checked out, batteries recharged, software upgraded if required, and feeder link and inter-satellite link performance checked out	Delta 2: A preliminary health checkout of each spacecraft is made to ensure that the vital satellite functions, such as attitude control and propulsion, behave normally. Zenit and Soyuz: N/A
Manoeuver for final orbit	About 48 hours into the mission, satellites begin ascending to the final altitude (780 km); low-thrust electro-thermal hydrazine thrusters are fired.	Delta 2: Within a few hours, each satellite is commanded to fire thrusters to jettison itself to its final altitude of 1,400 km. The satellite injected last is boosted first to minimize the risk of collision. Zenit: Satellite altitude (from the initial) is altered in groups of two or three to minimize the workload on ground controllers. Satellites are allowed to orbit until they reach their respective orbital plane, when they are jettisoned to their final altitude of 1,400 km. Soyuz: N/A
Arrival at final orbit	About two weeks	Delta 2: A few hours Zenit and Soyuz: N/A
Tests at final location	L-band communication links activation and tests including subscriber call handling, call handover between beams and satellites, paging, call-forwarding followed by beta testing by average users.	Assumption: Similar to Iridium

The information on constellation launch and deployment as outlined in the book demonstrates the launch strategies applied for these two systems. A detailed mission analysis for non-geostationary satellite systems additionally involves**,** health-monitoring, orbital maintenance, replacement , and end-of life management strategies (e.g., Cornia, et al, 1999)

.

Manufacturing methodology comparison

Satellites of both the systems were mass-produced using a production line. Iridium satellites followed production technique similar to those used in automobile assembly line; the production aspect was kept in mind from the outset of spacecraft design. Globalstar satellite followed a production line enabled through standardized interfaces and by designating personnel to follow through the production right.

Iridium satellites were assembled by Motorola such that a satellite was produced every five days at the peak of activity in 1998, with up to 10 satellites being assembled in a day. The production time of each satellite was about 35 days. The production aspect was included from the design phase with the assumption that 80–90% of cost, quality and production time would be influenced by satellite design. The production was managed in entirety i.e. from component construction through to spacecraft delivery and on to the launch site. Components were produced in quantities needed only for the current assembly – this 'lean production' is used by Japanese car manufacturers. The technique allows continuous refinement of the assembly process. Other techniques involved were - use of off-the-shelf components, use of easily pluggable and interchangeable circuit boards, and rigorous low-level tests to minimize system-level problems.

Before the start of production, extensive electrical, software, acoustic and vibration tests were performed on two full-scale satellite models to minimize the tests on production runs. Experience gained through assembly of the first batch of satellites was used for refining further production runs. The production line operated with about 50 engineers working in three shifts continuously at 15 different stations. Each satellite required about 160 printed boards which were assembled in one factory, sent to the next one where they were assembled and tested as subsystems, and then sent to the final assembly line where they were integrated as satellite units.

The first generation *Globalstar* constellation of 48 satellites was manufactured by Loral Space and Communication. The company designated design engineers to remain assigned to a product through to its full production run. Mechanical and electrical interfaces were standardized to allow quick changes of failed subsystems. Each module of the modular design was tested thoroughly before being assembled as a spacecraft. In the production line, satellites were moved from one assembly point to another and tested at each 'functional island'. The satellites were always moved forward. In the case of a defective satellite, it was removed from the production line, and therefore did not affect the production flow.

References

1) Stefania Cornara, Theresa W. Beech, Miguel Belló-Mora, Antonio Martinez de Aragon, (1999) '*Satellite constellation launch, deployment, replacement and end-of-life strategies*', 13th Annual AIAA/USU Conference on Small Satellites, SSC99-X-1, 19 pages.
(Accessed March, 2014)

2) *Globalstar-2* (*2013)*, Globalstar second generation mission profile and Soyuz launch sequence is available here (accessed March, 2014):

http://www.russianspaceweb.com/globalstar2.html

CHAPTER 7

1) With the help of a block schematic, discuss various network functions and components of an MSS system, including the interaction between the functional entities.

Air interface

Consider an MSS network as an extension of the terrestrially available services. In this role, the satellite network transports terrestrial network's core functionalities transparently. A generic protocol architecture comprises a *network independent* layer that deals with radio frequency transmission which pertains to the satellite network dealing with parameters such as coding, modulation, multiple access and encapsulation of the incoming stream, etc. for transport in the satellite network and a *network dependent* layer which deals with transport of the terrestrial services. Since a majority of MSS systems reuse terrestrial core network, it is usual to deal primarily with the network dependent layers when defining the satellite air interface.

The *air interface* comprises optimized feeder and service radio links and associated transport functionalities to connect the land earth stations and the user community via a transparent or regenerative transponder. A number of radio transmission designs have evolved and a number of these have been standardized and recognized by the ITU.

Figure 1 illustrates the connectivity architecture of an MSS system.

Figure 1 [A sketch of Figure 7.1]

Service link (SL) provides radio connection between the users and the satellite constellation; mobile-to-mobile communication is routed through a mobile–mobile service link (MMSL); intersatellite links (ISL) are used for routing calls in space in preference to terrestrial routing; feeder links (FL) connect gateway to the satellite constellation; intra-gateway links (GLN), inter-gateway links (GGL) and network management links (NML) are essential for exchanging information between system entities to support network management; and terrestrial links (TL) connect an MSS system to terrestrial network(s). Feeder and service radio links carry user-services and provide the desired mobility; whereas other communication links transport supporting network functions independently, and hence utilize any suitable cost-effective and reliable transport facility such as a private transport network, PSTN lines, leased circuits, spare space segment capacity of the network, etc.

Network components, functions and interaction

Figure 2 depicts the primary functional entities of a commercial MSS, segmented broadly by their physical association.

Figure 2 [A sketch of Figure 7.12]

Functional entities of a commercial MSS segmented broadly by their physical association.

Functions that are essential to establish communication include call handling, switching, routing, mobility management, and users' profile management. Numerous associated functions are necessary for successful management of the network. The *network management* (NM) functions deals with the management of the network to ensure proper functioning of the network. They involve management of radio resource, quality of service, network traffic flow, collection of call-data records and dispatch of data to the *business management* (BM) system and *mobility manager* (MM), traffic trend analysis for radio resource management, fault finding/diagnosis, fraud detection and secure privacy management. The NM also oversees the management of the *space segment.* NM also deals with monitoring radio spectrum , dispatch of call data records to the business management system, interaction with the mobility management system for call set-up, etc., network traffic trend analysis to assist radio resource management, fault finding/diagnosis and fraud detection. Notice that many management functions have commonality with terrestrial mobile system and thus benefit by adaptation from terrestrial system.

Mobile services represent various user services on offer including support of the associated protocols. Current portfolio include, voice, data, facsimile, paging, message delivery, emergency calls; and supplementary services such as call transfer, call forwarding, call waiting, call hold, conference calls, etc. The *space segment* comprising one or more satellites provides connectivity between the fixed and the mobile segments. Some of the functions shown as terrestrial can be either wholly or partially performed in space when regenerative transponders are used e.g. call routing. *Constellation management* involves Telemetry Tracking and Telecommand (TT&C) functions of satellite health monitoring, ephemeris generation, spacecraft orbit raising, orbital adjustment in case of a spacecraft failure, launch support during initial deployment or replacement of failed spacecraft, etc. The *business management system* constitutes the company's business hub responsible for customer billing, interfacing with the gateways and the network management system to obtain call records, updating profile of existing subscribers and introduction of new subscribers. The *mobility management* (MM) system maintains user location in its database and interacts with the NM and gateways for call connection, authentication, etc. In a GSM based network the MM consists of Visitor Location Register (VLR) and Home Location Register (HLR) with associated protocols for data exchange. In a packet switched network, a route is established for each packet, such as by a pre-calculated routing table. In an IP network, mobility management involves mobile IP functionality.

Gateways of a network can either belong to the network provider such as in ICO system or owned by different operators as in Globalstar system. The *network planning centre* oversees network management, capacity and business trend, specific events of interest and develops strategies for changes to the network such as expansion of capacity, introduction of new services, redeployment of satellites, etc.

2) Suggest the desirable features of an MSS service link to support: (a) hand-held services (b) broadband services on nomadic user terminals

Hand-held service

To counteract low G/T of hand-held receivers, typically about -24 dB/K, and severe channel impairments, high power satellite transmissions are essential. This is achieved by using narrow spot beams and an array of high power amplifiers whose outputs are combined before transmission.

In the return link, terminal size/weight, battery capacity and health risks from radiation limits the user terminal EIRP and consequently the throughput. Low power transmission from hand held terminals results in extremely low received signal levels at satellites. The problem is mitigated by improving satellite G/T and/or by path loss reduction by deploying low altitude satellites. Hence, invariably, geostationary systems use high G/T spot beams and the other approach is to use high G/T spot beams in low or medium Earth orbits to additionally reduce path loss.

The service transmission format must use powerful coding and robust modulation scheme such as QPSK in conjunction with high link margin and a complementary multiple access scheme.

Broadband services on nomadic user terminals

Nomadic terminals are fixed and hence in most instances the user has the freedom to install it with unrestricted view of the satellite. Furthermore, the impact of multipath is reduced at high elevation angles due to a static environment and to some extent the relatively high directivity of user antenna. The Doppler associated is absent which simplifies the demand on acquisition and carrier recovery circuits.

In the forward direction high EIRP is necessary to support broadband services with sufficient link margin, although lower than that for a hand-held system. It is necessary to deploy narrow spot beam to achieve the desired EIRP and additionally improve spatial reuse. Transmission format should be spectrally efficient such as by using high level PSK scheme backed up with powerful codes.

In the return direction, the terminal EIRP is constrained by the limits on antenna and power amplifier, battery capacity and spurious/intermodulation generation. High G/T satellite receiver is essential leading to the necessity of narrow spot beams. The air interface should be robust and matched to the channel characteristics.

3) A prospective MSS operator intends to provide a seamless service aiming to address the mass-market that covers a wide region encompassing large uninhabited open areas, motorways, a large railway network and several densely populated metropolitan cities. Suggest an efficient network architecture, stating the rationale for the selection of each element of the network.

System requirements of the MSS are summarised as follows:

Coverage: The system is operated by a single operator and requires a wide area coverage rather than a global system

Service area: Judging by the given market, i.e., large uninhabited open areas, motorways, a large railway network and several densely populated metropolitan cities, it is assumed that the polar region is excluded. The service area may therefore comprise either: primarily high latitude region (e.g. >50°N) or equatorial region (e.g. <50°N) or a mix.
Services: The services may be narrow-band (e.g. voice) or/and broadband (e.g. Internet, including streaming).
User community: Land-based
User Terminals: A variety of user terminals are envisaged - hand-held; nomadic; vehicular; and train-mounted
Inter-network connectivity: Connectivity is required to terrestrial Mobile service core network (e.g. 3-G) including the Internet.
Intra-system mobility: A seamless coverage with handover is necessary.

Proposed network architecture

Referring to figure 1 of the previous question, service link (SL) provides radio connection between the user and the satellite constellation. The related issues are addressed later. Mobile-to-mobile communication is not necessary. Intersatellite links (ISL) can be used if a non-geostationary constellation is chosen if selected based on a techo-economic trade-off. The feeder links (FL) will depend on the chosen air interface and this aspect is discussed below. Other links include intra-gateway links (GLN), inter-gateway links (GGL) and network management links (NML) for exchanging information between system entities to support network management. Terrestrial links (TL) are used to connect an MSS system to terrestrial network(s).

Referring to figure 2 of the previous question, the functional entities of the network shall include the following systems - constellation management, business management, network manager (NM), mobility management (MM) and a network planning centre.

In view of the mobility requirement, the MM system will maintain user location in its database and interact with the NM and gateways for call connection and user authentication, user profile, etc. Assuming a GSM based system, the MM will utilize the Visitor Location Register (VLR) and Home Location Register (HLR) with associated protocols for data exchange. For the packet switched part of the network, a route is established for each packet, for example, using a pre-calculated routing table. Assuming an IP network, mobility management involves mobile IP functionality; the operator can choose from various available protocols to ensure a smooth handover between spot beams.

SPACE SEGMENT

The service area and service type determine the space segment architecture.

Primary service region <50°N

A geostationary earth orbit (*GEO) space segment* is suitable for regions <50°N due to numerous advantages such as simple network architecture, stable radio links, robust and well-proven technology.
LEO and MEO systems are discarded in the first instance for the following reasons:

- LEO or MEO systems could be inefficient for regional systems due to utilisation of only a part of an orbit while needing several satellites to ensure continuous coverage. Nevertheless, a detailed techno-economic trade-off is required before a final decision can be taken.
- LEO and MEO systems tend to have expensive and complex network architecture.
- Terrestrial-like propagation delay is not necessary.
- Coverage over polar region is not necessary.

Primary service region >50°N

A HEO system provides high elevation coverage over high latitude regions for a reliable radio link. In contrast, satellites in GEO system appear at low elevation angle which necessitates unrealistic link margins. LEO and MEO systems can be designed to provide robust links with high elevation angle and diversity. As mentioned above, a detailed trade-off analysis is recommended.

Frequency band

An important consideration is the operational frequency band. In this respect L band is preferred because it can support on-move communications reasonably well and technology is mature; a second choice is S-band. A detailed investigation regarding availability of bandwidth in L band is necessary; this includes a regulatory review in addition to ensuring that the desired bandwidth is available in presence of incumbent operators. A preliminary estimate of the required bandwidth is mandatory.

USER SEGMENT
User community is land-based using a variety of terminals: hand-held, nomadic, vehicular (cars, buses, trucks, etc.) and terminals mounted on railway trains.
The air interface design including link budget for each mobile class provides an estimate of satellite EIRP and bandwidth. Selection of a suitable air interface is necessary. It can be selected from the various designs recommended by the ITU (see below).
Detailed specification for each terminal type becomes essential can be drawn up as the requirements are crystallised.

FEEDER LINK and TERRESTRIAL INTERFACE
The feeder link frequency band and selection of an appropriate air interface is essential. Various air interface standards have been recommended by the ITU and the operator can select an appropriate one taking into consideration propriety, applicable patent and intellectual rights. Some standards also specify the space segment.
Interface with the terrestrial network is over a 3-G core network of the operator's choice, allowing maximum reuse of terrestrial technology. The interface includes circuit and packet mode transmissions.

4) **In comparison to a technology driven, a market-led approach in developing the requirements of an MSS reduces the financial risks to the system. Describe the process in evolving a market-led system development process including the influences and constraints which must be taken into consideration.**

Due to the large number of inter and intra-system dependencies, MSS system development can be visualized as a structured top-down multi-layered iterative optimization, wherein system synthesis cascades down progressively to lower layers iteratively, as design implications are better understood and real-world constraints applied. The concept of such a market-driven system development is illustrated in Figure 1 below.

Figure 1 [A sketch of Figure 7. 4] Concept of a market driven design methodology

Research targeted at the addressable cliental provides a set of telecommunication requirements, terminal characteristics and service area, which are then synthesized for a top level system architecture to estimate approximate costs, technology and schedule. Iterations are necessary before a realistic solution can be achieved. The preliminary design is evolved into a detailed design taking into consideration the state of technology, costs, schedule, etc. If the detailed design leads to unrealistic specifications, the requirements are modified until an acceptable solution is found; the detailed design specifications are used for developing each system element. The system elements are integrated and tested and the space segment launched leading to the operational phase; finally, operational experience and market trends are utilized to evolve the system.

Due to large variability in constraints and assumptions, the architecture is not bound by rigid rules, nor does a given set of requirements have a unique solution.

Influences

Communication service, the size of the market, the service area and the cost model create the basis of a preliminary design. Consider each in turn.

1. Communication service

Services range from interactive to non-interactive services with a range of permissible delay from a few tens of ms to tens or hundreds of seconds and throughputs needs of a few bps to tens of Mbps.

The orbital characteristics of a constellation can be influenced by the chosen service. For example, non-real-time applications can be served through intermittent LEO or MEO coverage; whereas a seamless coverage is desirable for delay-sensitive interactive services.

2. Market

Market size, scope and service area govern the space segment capacity, size and cost of terminals and the service cost.

Users' ability to pay and type of use – business, pleasure, distress or remote area communications – govern the service portfolio, system cost, complexity and service cost. For example, a system targeted for personal communications must be capable of low cost hand-held terminal in conjunction with an attractive tariff at the expense of a complex space segment. On the other hand, a system targeting the transport industry, where size and cost of terminals or services are less critical, can deploy a simpler space segment.

Space segment capacity, cost and anticipated revenue must be matched to break even as quickly as possible.

3. Service area

The system architecture depends on the service area, traffic distribution and the need of features such as geographic call-barring facility. These inputs determine the type of constellation, orbital inclination and eccentricity for non-geostationary constellations, spot beam distribution and identify area-dependent system features.

4. Cost model

A preliminary business plan is adequate to investigate the sensitivities of the business case to system attributes such as spacecraft cost, schedule, traffic growth, service cost, etc. Empirical cost models are available in the literature for estimating spacecraft and launch costs from system-level requirements. Established companies possess extensive database and up-to-date market information for accurate trade-off analysis.

Constraints and considerations

System entities and their relationships must be synthesized applying real-world constraints and considerations. Such constraints may result in solutions, which are, at best, sub-optimal in an academic sense but acceptable in practice.

1. Business

Financial viability is fundamental to a commercial venture. Examples of relevant business issues are as follows.

- *Financial risk*

A number of assumptions impacting on profitability would have been made initially. These would include traffic forecast, geographical distribution of traffic, service penetration rate, user requirements, user's ability to pay, impact of competition, etc. The applicability of such assumptions can deviate substantially for a variety of reasons such as delay in the introduction of the service, unexpected competition, poor service quality due to technical problems, delay in availability of user terminals, high usage costs, etc. Such risks can be detrimental in raising initial capital or after implementation.

- *Capital* for an MSS venture is raised from a number of sources and financial instruments; success in this respect depends on the state of economy, performance of the satellite industry in general and MSS in particular, the credibility of the proposers and the soundness of the business plan
- *A clear understanding of commercial issues* is central to commercial success. These include cost per call or Kbyte, operational and maintenance expenditure and obligations, segmentation of revenue between participating entities – i.e. infrastructure provider, service provider, etc.

2. Regulatory considerations

Regulatory issues deal with spectrum allocation, operating license(s) and local rules. For example, a regulatory authority may impose technical and/or financial conditions for granting a license. The authority may require an operator to prove its financial position, modify the proposed design, accept a tentative license to prove the viability of the proposed design, etc.

3. Orbital characteristics

Orbital characteristics are determined by coverage and service requirements. The impact of the space environment, transmission delay, constellation cost, size and maintenance together with network complexity and coverage area play a crucial role in determining the orbital characteristics.

4. Call routing, connectivity and mobility management

Network connectivity requires consideration on the location and management of gateways. Intersatellite links increase flexibility in siting gateways but increase space segment complexity. When an intersatellite link is used numerous routing alternatives are feasible and therefore a suitable routing scheme must be established. Terrestrial routing requires dealing with numerous authorities while adding costs to the end-users but without impact on the space segment complexity.

To establish a call, the mobile user must be located, paged and a route established; when a call is established the connection must be retained. This task is achieved by a mobility management subsystem. In case of an IP connection, the task is to maintain IP connectivity. After a call has been established, the radio connection can be broken when the visibility of the user to the satellite is lost due to the movement of the user or the in a NGSO by the satellite itself. Thus it becomes necessary to handover a call to a radio bearer which can retain the connectivity. Various types of handover can be envisaged:

The handover scheme depends on spacecraft design, constellation architecture and air-interface design.

When considering connectivity of inter-satellite links, regular handovers are necessary due to constellation dynamics. Numerous algorithms have been proposed based on a various optimization criteria. Since routing algorithms require on-board processing; a more complex algorithm imposes a greater burden on satellite processors but can be more efficient.

5. Network interface

The network must have provisions to interconnect mobile and fixed users with each other. This connectivity is achieved through appropriate inter and intra network interfaces. The interface with the terrestrial network depends on the inter-operability arrangements. For instance Inmarsat's BGAN network provides a seamless connectivity with the terrestrial 3G core network. Similarly, Iridium and Thuraya systems have provisions to interconnect with the GSM system.

6. Hardware realization

First generation LEO and MEO systems were based on a number of novel technologies, such as integrated handset design, multiple satellite launches, advanced spot beam technology, onboard processing, intersatellite links, etc. Depending on factors such as familiarity with technology and risk attitude, a number of constellation architectures have emerged – operators have usually combined mature technologies with novel ones, although the extent of combination differs widely.

5) Suggest the main elements and processes of a simulation scheme meant to derive system parameters such as system capacity, propagation delay, link utilisation statistics and the length of the PSTN lines, given the subscriber distribution, the traffic statistics, the space segment configuration and a gateway distribution.

Due to the inherent non-linearity and complexity and a need for repetitive analysis with flexibility, computer simulation is commonly used in performance evaluation of an MSS. Figure 1 below illustrates a flow chart of a simulation program for estimating capacity, utilization, propagation delay statistics and length of PSTN lines of LEO or MEO constellations. The simulation can be conducted for a given a subscriber distribution and calculations can be performed for successive instants of time until the statistics can be estimated. [To the reader: A brief description of the figure is desirable]

Figure 1 [A sketch of Figure 7.7]

Flow chart of a simulation program for estimating capacity, utilisation, propagation delay, statistics and length of PSTN lines of LEO or MEO constellations.

6) In evaluating LEO and MEO satellite systems, orbital altitude has a profound effect on a number of system parameters; list such parameters and outline with reasoning the impact of altitude on these parameters. Suggest a favourable range of altitudes to support a hand-held voice-medium bit rate communication service

System parameters affected by orbital altitude are:

- transmission delay;
- spectrum efficiency;
- spacecraft power;
- user terminal EIRP
- Constellation size

In the following explanation we will assume that the frequency band, number and roll-off of spot beams are identical at each altitude.

Transmission delay increases as satellite altitude increases due to a corresponding increase in satellite range. Thus low earth orbiting satellites are used to provide near zero delay performance.

Spatial spectrum efficiency reduces as the altitude is increased. With an increase in orbital altitude a satellite's field-of-view increases and thus the geographical area covered by each spot beam increases resulting in spreading of frequency band over a wider area, that in effect reduces the spatial spectral efficiency.

The satellite power per channel must increase as altitude increases to compensate the additional spreading loss due to an increase in Earth-satellite range.

A study directed to assess the influence of satellite altitude in range 500–7500 km on these parameters, produced numerous interesting conclusions. It was demonstrated that lower altitudes increase frequency reusability, capacity and power. However, results pertaining to estimated transmission delays were not intuitive. Ground routing exhibited a shorter delay with an increase in altitude. When using intersatellite links, however, medium altitude proved to have the shortest delay because the number of switching nodes in low Earth orbit increased in comparison increasing the switching delays. It was observed that the delay depended on type of traffic, i.e. local or long distance and that with evolution in technology, switching time would reduce. Furthermore it was noted that that frequency reuse increased as altitude is reduced; however, the capacity limit in this case may occur due to the power capacity of the spacecraft. As expected, handset power increased with orbital altitude, impacting the size, weight and the available throughput.

It was observed that the mass of the power system increases with altitude, even though capacity and eclipse period reduce; this occurs because capacity increases linearly but power increases as the square of the altitude.

Number of satellites in a constellation increase as orbital altitude is decreased due to a reduction in field of view of each satellite. Based on space segment radiation environment the favored orbital altitude of low and medium earth orbit lies respectively between 700-1500 km and 10000-12000 km. Both the orbital altitude ranges can provide hand-held service. If transmission delay is crucial then lower altitudes are preferred. However, a techno-economic analysis is necessary as the evaluation criteria can differ between operators. For example, ICO preferred MEO whereas Globalstar and Iridium preferred a LEO.

CHAPTER 8

1) Describe a generic protocol architecture which can be used in standardisation of MSS radio interface. What are the advantages in using such an approach?

A generic protocol architecture used in the European Telecommunication Standard Institute (ETSI) standards (and others) is presented in figure 1 below. It comprises a *network independent* layer that deals with radio frequency transmission comprising synchronisation within the satellite network, coding, modulation radio frequency conversion/filtering/amplification, multiple access and encapsulation of the incoming stream for transport in the satellite network and a *network dependent* layer that deals with transport protocols related to the core network and applications. The figure also lists critical parameters of the lower physical layers addressed in many standards.
Since a majority of the standards and recommendations use the same generic architecture, there an inherent similarity across these systems; the differences lie in the solutions applied in achieving their respective goals.

Figure 1 [A sketch of Figure 8.1]
A generic protocol architecture applicable to the satellite radio interface standards (DVB-1, 2012) ©ETSI

2) Describe the evolution of the GMR standard, mentioning the incremental advances in radio interface technology at each stage.

GMR radio interface specifications are prepared by ETSI Technical Committee-Satellite Earth Stations and Systems (TC-SES). The evolution of these specifications have continually tracked those of the GSM system, as illustrated in figure 1 below

Figure 1 [A sketch of Figure 8.2]
Evolution of the GMR system

GMR-1 (ETSI TS 101 376), released in 2001, supports circuit-switched voice and data up to 9.6 kbps, short messaging service (SMS), cell-broadcast and location-based services over GSM core network interface-A. GMR-1 Release 2 known as global mobile packet radio services (GMPRS) comprising three sub-releases over the period 2003-2008, incorporates general packet radio service (GPRS) capabilities to GMR-1 with progressively new terminal classes. GMR Release 2.1 (2003) incorporates packet switched GPRS-like services up to 144 kbps based on GPRS release 97 (r97) to provide mobile services up to 144 kbps over GPRS core network interface, Gb. GMPRS release 2.2 (2005) extends services to hand-held terminals at a throughput of 64 kbps in the forward direction and 28.8 kbps in the return direction. Release 2.3 (2008) extends the data rates to 444 kbps units in the forward direction and up to 400 kbps in the return direction. GMR Release 3, known as GMR-1 Third Generation (GMR-1 3G), a satellite adaptation of Enhanced Data Rate for GSM Evolution (EDGE), enhances the throughput to 512 kbps via the core network interface, Iu. EDGE, standardized by Third Generation Partnership Project (3GPP), is compliant to ITU's International Mobile Telecommunication-2000 (IMT-2000) recommendations. It is based on 3G release 6 protocols interfacing with the core network on Iu-PS interface (Interface between base station system and core network - Packet Switched) to provide voice and broadband services including IP multimedia services up to 592 kbps depending on terminal types which range from small handheld terminals to large high gain fixed or transportable terminals.

3) Compare the lower layer technologies of the three GMR releases.

The question has asked for a comparison. The text below outlines main features of the air interface of each release. The student should collate lower-layer features of each release and tabulate them in a table with comments

GMR-1

Features of lower layer include modulation/coding, power control and multiple access. Physical channels are a part of this layer. Each physical channel contains one or more logical channels. A logical channel contains data related to a specific function whereas a physical channel is responsible for physical transport.

Coding and Modulation

The applied channel coding rate applied on per call basis depends on factors such as the channel impairments, information block size and the desired protection level. It comprises an inner coding scheme in tandem to an outer code. The outer code comprises Cyclic Redundancy Check (CRC) (8 bits, 12 bits, or 16 bits parity, depending on the channel) whereas the inner code comprises convolutional coding with a typical constraint length of 5 and rates of 1/2, 1/3, 1/4, and 1/5, as required.

The power control messages use the (24, 12) systematic Golay encoder with a soft-decision Golay decoder. Reed-Solomon code is used for a number of channels. For example, the basic alerting channel (BACH) uses a systematic (15, 9) Reed-Solomon code.

Various puncture masks are used to fit the coded bits into the channel's bit capacity. Channel-dependent interleaving is used and can be applied intra-burst or inter-burst and is based on block interleaving methods with pseudorandom permutations. A scrambler adds a binary pseudo-noise sequence to randomize the incoming bit stream. Certain channels such as the traffic channel include data encryption to prevent eavesdropping.

The modulating symbol rate is fixed at 23.4 ksps. A majority of traffic and control channels use π/4-CQPSK (coherent quadrature phase shift keying) modulation shaped by a root raised cosine filter of 0.35 roll off factor; π/4-CBPSK (coherent binary phase shift keying) modulation with identical filter shape and roll off factor is also used. A keep-alive burst is transmitted during periods of speech inactivity to save the battery life, satellite power, reduce co-channel interference, add-comfort noise, and maintain the power control and timing/frequency synchronization. The bursts uses π/4-DBPSK (differential binary phase shift keying), with the same (i.e. 0.35) roll-off root-raised cosine filter. The BACH uses pulse shaped 6-PSK (Phase Shift Keying) modulation and the frequency correction channel (FCCH) burst is a real-chirp signal spanning three slots.

Power control

Power control is exercised in both the directions to minimize interference, conserve power at satellite and MES, and maintain signal quality. The closed-loop control utilizes a feedback from the remote receiver for power adjustment; in the open-loop method the transmitter adjusts power based on the behaviour of the received power, with the premise that receive and transmit channel behaviours are correlated. The open loop system has a faster response as it avoids the delay incurred in a feedback loop from the distant transmitter and hence suits events of rapid fading.

In the idle mode, the MES is required to lock to the broadcast control channel (BCCH) to extract the system information in readiness for communication.

Multiple access

A TDM/TDMA accessing scheme is used in the service link. The forward channel consists of a continuous time division multiplexed (TDM) channel and the return link consists of time division multiple access (TDMA) bursts, synchronized in time and frequency to avoid interference. Message synchronization is also necessary to ensure that the start of each message sequence is identifiable. Procedures for these types of synchronization have been specified.

Physical and logical channels

Frequencies are allocated in steps of 31.25 kHz in pairs separated by 101.5 MHz. There are 1087 pairs in the 34 MHz MSS allocation. The minimum allocation of channels per spot beam is five.
The physical channels can be used to carry traffic or to support various control functions to ensure proper functioning of the network. The features and uses of these physical and logical channels are summarized in Table 1 below

Table 1 [Excerpts of Table 8.4]
Example features and uses of various traffic and control logical channels.

GMPRS-1
The GMPRS-1 is release 2 of the GMR-1 specifications, appending to release 1, packet data service based on the GPRS. These specifications enable packet data services at data rates ranging from 64 kbps up to 444 kbps in the forward link, depending on the capability of mobile terminals.

Version 2.1.1 (2003) enables bidirectional packet data rates up to 144 kbps supporting QoS differentiation across users with dynamic link adaptation.

Version 2.2.1 (2005) specifies packet data services to handheld terminals at up to 64 kbps on the forward link and 28.8 kbps on the return link.

Version 2.3.1 (2008) supports broadband packet services at data rates up to 444 kbps on the forward link and 202 kbps on the return link on A5 size transportable terminals with a provision to support data rates up to 400 kbps on the return link through an external antenna. Version 2.3.1 utilizes advanced techniques such as low density parity code (LDPC) and 32-ampltude phase shift keying (APSK) modulation to enhance spectral efficiency and provides bi-directional streaming services.

Differences in radio transmission and reception characteristics with respect to GMR-1 release 1 specifications include:

- Support of unpaired frequencies
- Differences in test conditions due to different Energy per symbol to noise power density ratio (E_s/N_o) requirement in packet mode
- Modifications in the switching time requirements
- Addition of four power classes of user terminals (C, A and D - two types) and associated specifications such as radiation pattern, receiver G/T, frame error rate at receiver
- A few modifications applicable to all classes of user terminals – e.g. transmit requirements (burst ramp-up and ramp-down time, unwanted emissions)
- Updates to packet burst data structure and modulation scheme
- Radio link failure detection, capability to identify BCCH type [Temporary-BCCH (T-BCCH) related to a temporary spot beam or Anchored BCCH (A-BCCH)] in idle mode and idle mode loss of T-BCCH when camped on it
- Link adaptation procedures
- Some clarifications with regards to the reporting by MESs of link quality.

GMR-1 3GPP
GMR-1 3GPP air interface is an evolution of the GMPRS-1 system using new developments in the satellite physical layer air interface technology. The air interface is applicable in L and S MSS bands and connects with the core network over an interface known as I_u-PS.

Key features of the air interface include:

- Up to ~592 kbps mobile-user throughput
- Spectrally efficient multi-rate Voice over IP (VoIP) with zero byte header compression
- IP Multimedia Services
- IPv6 compatibility
- Differentiated QoS across users and applications.
- Robust waveforms to improve link closure reliability
- Dynamic link adaptation
- Multiple carrier bandwidth operation

- Multiple terminal types - Hand-held terminals, Personal Digital Assistant (PDA), vehicular, portable and fixed
- Terrestrial-Satellite handovers
- Unmodified Non-Access Stratum (NAS) protocols and core network

A variety of mobile earth stations (MES) are supported - hand-held, fixed, transportable and nomadic. In addition to provide IP data traffic at data rates commensurate with MES capability, MESs support voice at 2.45 kbps and 4 kbps using zero-byte header compression.

The L-band terminal receive and transmit frequencies are respectively 1525-1559 MHz and 1626.5-1660.5 MHz and the S-band frequency bands are respectively 2170-2200 MHz and 1980-2020 MHz, with tuning steps of 31.25 kHz in each band. Thus there are 1087 paired cannels in L-band and 960/1280 in S-band – note the asymmetric allocation of S-band. The L-band duplex channels are paired with a separation of 101.5 MHz.

The system uses frequency division multiplexed TDM/TDMA scheme forward/return link – with the same frame structure in each part. The TDMA frame consists of 24 slots; 16 TDMA frames constitute a multi-frame and 4 multi-frames make a super-frame.

Two categories of channels are used – traffic and control channels. Circuit mode traffic channels (A-mode) use convolutionally coded $\pi/4$ CQPSK to provide encoded voice and data rates up to 9.6 kbps. Packet data traffic channels (PDTCH) carry packet mode traffic in Gb or Iu mode at data rates between 8.8 kbps and 587.2 kbps using $\pi/2$ BPSK, $\pi/4$ QPSK, 16-APSK, or 32-APSK modulation, the modulation selection being dependent on the chosen data rate (Note: Iu = Interface between base station system and core network). The RF bandwidths range from 31.25 kHz to 312.5 kHz. The modulated signals are filtered by a 0.35 roll-off root raised cosine filter. In addition to traffic and control channels, a constant envelope frequency-modulated (called dual-chirp) channel is transmitted for initial time and frequency acquisition at the user terminals. Depending on the transmission bit rate a variety of coding schemes are used – convolution code, LDPC and Turbo code. Table 2 below lists examples of the forward error correction codes supported by GMR-1 3G.

Table 2 [Excerpts of Table 8.6]
Code rates used in GMR-1 3G.

The return PDTCHs are defined as PDTHC (m, n) where m denotes the bandwidth as a multiple of 31.25 kHz, and n the number of TDMA slots assigned to it. For example, PDTHC(1, 3) denotes a physical channel occupying a bandwidth of 31.25 kHz comprising 3 slots in its TDMA frame. A dedicated PDTCH channel can be assigned for carrying voice traffic. The bearer service ranges from 1.2 kbps to 592 kbps.

The transmission rate depends on a combination of the modulation scheme and the coding rate. For example, the peak rate of 590.4 kbps for PDTCH3 (10, 3) can be achieved by 16-APSK at a code rate of 2/3 or, by $\pi/4$ QPSK at a coded rate 5/6.

4) What are the similarities and differences between a terrestrial cellular system such as the GSM system and a geostationary satellite system, which can be used to assist in optimising the air interface of a satellite system, retaining features of the terrestrial system as far as possible?

The question is answered in Table 8.8 of the book:

5) Compare the lower layer features of GMR-1 release 1 and GMR-2 systems.

Lower layer attributes of GMR-1 release-1 are described in answer to question 3. These attributes of GMR-2 system are described below. The student should prepare a comparative table combining the details given in question 3 and those outlined below to complete the answer.

Frequency bands

Service link
Space-Earth: 1525.0 - 1 559.0 MHz, Right Hand Circular Polarization (RHCP)
Earth-space: 1626.5 - 1660.5 MHz, RHCP

Feeder link
Earth-space: 6425.0 - 6 725.0 MHz, Linear Horizontal Polarization
Space-earth: 3 400.0 - 3 700.0 MHz, Linear Vertical Polarization

Channel types
There are two categories of physical channels –traffic and control channel, each sub-categorized into logical channels consistent with their respective function.

There are four types of physical traffic channels to carry voice or user data - Full rate traffic channel (24 kbps), half rate traffic channel (12 kbps), quarter rate channel (6 kbps) and one eighth rate traffic channel (3 kbps).

Signalling channels are categorized as: Satellite Broadcast Control Channel, Satellite Common Control Channel, and Satellite Dedicated Control Channel.

Multiple Access
The FDM/TDMA scheme is used with 200 kHz carrier spacing in the forward link comprising eight slots and 50 kHz in the return link comprising two slots.

The forward time-slot lasts 15/26 ms (~ 0.58 ms) at a burst-rate of about 270.833 kbps. The return time-slot lasts 60/26 ms (~2.3 ms) at a burst-rate of ~67.708 (1 625/24 kbps).

The TDMA frame structure hierarchy comprises hyperframe of 3h 28 min 53 sec 760 ms duration, chosen to support GMR-2's cryptographic mechanism. Each hyperframe comprises 2048 superframes of 6.12 seconds which is further subdivided into multiframes. There are two types of multiframes, partitioned to carry specific logical channels - 120 ms multiframe of 26 frames and ~235.4 ms multiframe of 51 frames. Twenty-four multiframes of the 26-frame multiframe are used for traffic and two are used for in band signalling. The 51-frame multiframe is used for control channels.

Six types of TDMA bursts are used, depending on the functions – to carry traffic, access the network, MES time synchronization, frequency and time correction, etc.

Coding and modulation
Various block and convolution codes are used with interleaving – the selection depending on the function of the given logical channel. For example, the basic voice channel uses a rate 1/2, 64-state punctured convolutional code; the associated and dedicated signalling channels uses a rate 1/2, 64-state convolutional code along with a Fire Code. The information bits in some control channels utilize a rate 1/2, 16-state Convolutional code and a Fire Code. Yet another coding scheme utilizes rate 1/3, 64-state convolutional code. Interleaving dimensions for the communications channels vary from 3 bursts to 81 burst depending on the required protection.

The forward link uses OQPSK modulation followed by a 0.35 roll off square root raised cosine filter. The return link modulation (including the single hop MES-MES link) for the traffic and in-band signalling utilizes Gaussian Minimum Shift Keying (GMSK) - pre-coded in the same manner as specified in the standard GSM.

Other physical layer functions in GMR-2 are - management of dynamic power control mechanisms of the MES and the gateway; synchronisation of time and frequency at the MES receivers and measurement procedures for the selection and reselection of spot beam.

Bit error rates are specified for various types of channels under quantifiable measurement condition. For example, under nominal conditions – defined as a channel without interference and an input level ranging from 10 dB to 20 dB above the reference sensitivity level in an additive white noise Gaussian channel (AWGN) – the bit error rate prior to error correction is specified as < 10^{-4} (for voice) and the that of data traffic as < 10^{-7}.

6) With the help of a schematic describe the functional entities and the interaction of a typical MSS system such as GMR-1 and GMR-2.

Figure 1 below shows GMR-2 system elements. It comprises a geostationary satellite, Network Control Centre (NCC), Satellite Control Facility (SCF), Customer Management Information Centre (CMIS), Gateways, and a population of Mobile Earth Stations (MES). MES types include handsets, vehicle-mounted terminals, and fixed terminals. The Gateways have external interfaces to existing fixed telecommunications infrastructure, namely PSTN, PN and PLMN.

Figure 1 [A sketch of Figure 8.8]
GMR-2 system element

The gateways implement the radio modem functions of the terrestrial BTS, the radio resource/traffic channel management, call setup functions of (terrestrial) base station controller (BSC) and switching functions of mobile switching centre (MSC), along with maintaining databases for subscriber data. An example of traffic and signalling channel radio connectivity between gateways, network and satellite control station, and user terminals is presented in figure 2 below. Broadcast and Common control signalling channels, provided by the NCC, are used during the initial call set up. Note that single hop user-to-user traffic link through the satellite uses L-band links. Call control for user-to-user circuits are performed by the NCC, Gateways, and by switching on the satellite, to achieve single-hop connectivity.

Figure 2 [Figure 8.9 (b)]
Example of traffic and signalling channel connectivity

The mobile earth station comprise mobile equipment consisting of a mobile termination (MT) unit that may include an interface for terminal adapters (TA) and terminal equipment (TE). A subscriber identity module (SIM) provides a unique user identity and profile. The MT is the main unit of the MES responsible for all major radio and communication functions.

Functions performed by the network management centre include system resource management, congestion control and network synchronization; day-to-day operations and maintenance; inter-station signalling management; user terminal and gateway commissioning support, common channel signalling; defining and managing payload configurations.

The Gateway sub-system (GWS) comprises transceivers, controllers and other units interfaced to the MSC through a single A-interface The Gateway subsystem controller (GSC) controls the functioning of Gateway Transceiver Subsystem (GTS). Each GTS is responsible for call management in a given spot beam.

Gateway functions include management of dedicated signalling and traffic channels; connectivity to external fixed networks; mobility management; interoperability functions related with partner land mobile networks; authentication and encryption services; radio resource management, support for user terminal commissioning, etc.

The SCF is responsible for spacecraft monitoring, managing payload configuration, generation and distribution of satellite ephemeris, telemetry and commands processing for spacecraft management, real-time range and range rate measurement for satellite ephemeris generation, real-time payload switching, etc.

The customer management information system (CMIS) manages gateway configuration billing and accounting and processing of call records.

The network control centre functions include call management that includes network parameters broadcast, payload configuration through the SCF for mobile-mobile connections, traffic monitoring, radio resource management and network timing synchronization.

Network functions to support basic services operation involve: call handling, subscriber identity authentication, support of emergency calls, supplementary services support, and short messaging service support. Network support for satellite operations entails location registration to keep track of mobile locations, intra-spot beam handover, high penetration alerting, centralized call management from the NCC.

7) With the help of a generic IMT-2000 satellite network schematic identify various interfaces including those which are excluded from the IMT-2000 radio interface specifications. List ITU recommendations pertaining to the integration of satellite systems with terrestrial IMT-2000 systems.

Figure 1 below shows a generic IMT-2000 satellite network demonstrating various interfaces including those interfaces excluded from the IMT-2000 radio interface specifications since their implementation depends on operators' design and optimisation criteria. Excluded interfaces are - satellite-feeder link, satellite-satellite link, internal interface between terrestrial and satellite elements within a mobile earth station and the interface between satellite and the core network. It is recommended that the satellite component should interface with the terrestrial core network in a similar manner as the terrestrial system, incorporating satellite-specific alterations such that all the key IMT-2000 service requirements applicable to the given technical and marketing considerations can be implemented. In a dual mode, satellite-terrestrial, user terminal, it should be possible for the terminal to select the appropriate mode either automatically or under user control. The user terminal should provide bearer services in both terrestrial and satellite modes with a roaming facility and be aligned to IMT-2000 service management and provisioning.

Figure 1 [A sketch of Figure 8.10]
Generic IMT-2000 satellite network ©ITU-1, 2010

The ITU recognizes eight IMT-2000 compatible radio interfaces – termed, Satellite Radio Interface (SRI) A, B, C, D. E, F G and H.

These interfaces deal with the radio part of *the service links* only. Architectural and system description are included where appropriate due to their strong relationship with the radio interface.

8) Several types of handovers are used in MSS. Hand-over algorithms are system-specific but broadly follow a similar sequence. List various handover types and outline the algorithm used in any one of the ITU-recommended satellite radio interface, mentioning the air interface.

Typical handovers used in MSS include – beam-beam, satellite-satellite, LES-LES and frequency.
An outline of the algorithms used in Satellite air interface A are as follows.

Beam handover – MESs regularly report pilot signal power to noise ratio (C/N) of adjacent beams to the communicating LES; the LES initiates a beam hand-off procedure at a set threshold by transmitting the signals to the operating and the candidate beams and instructs MES to start demodulating both the signals. When the LES receives reports of successful reception of the candidate signal at the MES, it drops transmissions to the operating beam.

Intersatellite handover – The MES measures the pilot C/N and scrambling code of candidate beams. When a new scrambling code is detected, the measurements are reported to the LES, which may exploit satellite diversity through maximal ratio combining by transmitting the same signal through both satellites and invoking a hand-off at an appropriate time following further network protocol exchanges.

Inter-LES handover – Inter-LES handover is achieved by negotiations between the present and the new LES. The new LES starts transmitting a carrier towards the affected mobile which is simultaneously instructed by the current LES to search for the carrier of the new LES. When the MES confirms to the current LES, satisfactory reception of the new carrier, the current LES stops transmission thus handing over the communication to the new LES. *Inter-frequency hand-off* – This is a hard hand-off and is either intra-gateway or inter-gateway.

9) Compare the following characteristics of all the satellite radio interfaces recommended by the ITU:

- **Orbit**
- **Frequency band**
- **Mobile terminal types and throughput**
- **Modulation and coding scheme**
- **Multiple access scheme**
- **Diversity**
- **Power control**

This question requires a compilation of the parameters as used in each standard. Please see section 8.2.2 of the book. This work is for the student.

10) What are the constraints in incorporating mobility to the DVBS2/RCS standard? What are the techniques used in the DVBS2/RCS+M standard to circumvent these constraints?

A mobile environment entails RF propagation impairments which adversely impacts the conventional DVB-S2 (and DVB-S) forward and RCS return signal quality. Since DVBS2/RCS targets broadband and the L and S MSS frequency bands are congested, it is necessary to utilize Ku or Ka frequency bands for broadband communication where adequate spectrum is available. MSS allocations in the Ku band have a secondary status whereas 1200 MHz of primary allocation is available worldwide (20.1-21.3 GHz) in the Ka band.

Regulatory compliance necessitates adherence to transmission power spectral density limits; and interference tolerance at the receiver from the transmissions of the primary and shared satellite services. Antenna size and profile of mobile/nomadic user terminals are restricted in order to promote mobility resulting in side lobe levels higher than the conventional VSAT user terminals and consequently high off-axis transmissions levels. Thus specific provisions are necessary to ensure regulatory compliance, particularly for the Ku band.

Furthermore, considering the inherent shadowing and multipath effects on mobiles, the conventional DVB/RCS radio link design necessitates enhancements to meet the quality of service. For example, the user should not be logged off when the terminal gets shadowed nor suffer unacceptable quality during multi-path events. Moreover, the probability of causing interference to the adjacent region increases without any modification to the conventional DVB/RCS system - in the forward direction due to the requirement of high power per bit to counter the mobile environment, while in the return direction due to off-axis EIRP transmissions due to inferior side lobe discrimination. Thus, the radio link design and user terminal antenna performance exert a strong influence on DVB/RCS+M performance.

Forward link power spectral density
Studies conducted in the DVB technical forum estimate that the minimum antenna aperture for service provision in the regional and global beams respectively would be 30 cms and 50 cms with an unmodified DVB-RCS user terminal and such antenna sizes would restrict mobility. It is further estimated that a spreading of the DVB carriers by a factor of 4 provides the necessary reduction in the transmit power density in the forward direction, whereas the spreading factor should be at least 16 in the return direction to accommodate user terminal antenna diameters of 30 cms for the global beam (Morlet, et al, 2007-2008). The increased bandwidth would, therefore, have to be traded off with the user data rate.

The RCS+M standard utilizes channel spreading as a countermeasure towards interference mitigation in both forward and return directions.

Countermeasures for non-line of sight propagation impairment
In Ku/Ka band mobile environment shadowing, multipath and channel dispersion cause severe degradation to the received signals.

The situation is quite severe in a land mobile environment. For example, studies conducted in Italy for a railway environment in K_u band indicate 2-3 dB periodic attenuation due to electrical cables and 15-20 dB due to electrical trellises around the railway tracks, superimposed on environment dependent statistical fluctuations. The coding at the physical layer alone is inadequate in such a fading environment. The concept of upper-layer coding has been introduced in DVB/RCS+M as a fade countermeasure in such a way that the DVB-RCS transmission structure remains intact. Moreover, although the standard does not support ARQ schemes these can be implemented above the standardized components with a provision of buffering at both ends. A more robust solution (optionally supported) is the use of terrestrial re-transmitters which could take over in conditions of known prolonged shadowing such as inside a tunnel.

Demodulator synchronisation/resynchronization can be severely tested in presence of Doppler fluctuations; a receiver can lose lock during a shadowing or multipath event and hence requires rapid synchronization following an event. In the RCS+M system an interleaver is used to disperse bursty errors to counteract short-term blockages and multipath fades. Another countermeasure is a frame structure containing large numbers of pilot symbols. Solutions to minimize Doppler effects include Doppler pre-compensation at the transmitter and enhanced frequency acquisition algorithm and capture range.

Link Layer Forward Error Correction (LL-FEC) allows the use of the DVB-RCS FEC while applying FEC at an upper layer thus retaining the DVB-RCS physical layer. The LL-FEC is only directed to those RCSTs which have signalled ability to support the countermeasure in the common signal channel burst.
The technique can also be applied in the return direction by those terminals which opt for continuous return link carrier transmissions.

Spectrum spreading in return direction
The return link MF-TDMA transmission is spread in bandwidth for compliance to the power spectral density regulatory limits, particularly when using the secondary allocations. The provision is also available for continuous-carrier return link transmissions, although the process is different to that of the MF-TDMA mode.

MF-TDMA
Two methods for transmission of the return link bursts are available - π/2-BPSK modulation, equivalent to spreading a QPSK modulated signal by a factor 2, and burst repetition so that the effective bandwidth of the signal corresponds to that of the π/2-BPSK signal. Each burst is repeated F times which results in a symbol rate F times the unrepeated burst.

Logging in presence of large timing uncertainty
The mobiles are spread over a wide coverage area causing time dispersion in the received bursts. The burst timings of all the mobiles must be aligned for the bursts to arrive without overlaps. At the time of initial log-on the system timing is unknown to an RCST. The RCST then searches the largest continuous number of available slots and to minimize the probability of a collision, transmits at the centre of these slots if the number is uneven and in either one of the two centre slots when the number of available channels is even. Thereafter, the network advises the RCST of the reception time of the received signal in response so that the mobile corrects its timing.

Mobility management
The standard supports a number of handover methods for mobility management. A *gateway to gateway handover* occurs when a mobile crosses the area covered by a gateway and moves to an area covered by another gateway; a *satellite-satellite handover* occurs when a mobile crosses from the area covered by a gateway operating via satellite 1 to an area covered by another gateway operating through satellite 2; a *spot beam to spot beam* handover occurs when a mobile moves from one spot beam of a satellite to another spot beam of the same satellite; a *satellite to a gap-filler system* handover occurs when the mobile moves into an area where a prolonged shadowing is known to occur.

A number of interference avoidance mechanisms dealing with aeronautical, maritime and land mobiles have been standardised by ETSI in K_u and K_a bands as a part of terminals' control and monitoring functions to comply European Commission's (EC's) radio equipment and telecommunications terminal equipment (R & TTE) conformity directive. These are useful in the DVB/ (RCS+M) implementation context.

Continuous carrier operation in return link
The RCS+M standard addresses broadband services in a shared environment such as ships or aircrafts. Simulations indicate that continuous transmission instead of a demand assigned mode (MF-TDMA) could be more efficient from terminals which produce net traffic that appear as quasi-continuous. Thus, there is a provision for this class of RCST to support a continuous transmission mode. The RCST signals its ability to support this mode in the common signalling burst. An RCST that supports continuous carrier operation can operate either in a continuous carrier or an MF-TDMA signal, but not in both the modes simultaneously; however enhanced continuous carrier operation allows simultaneous transmissions of both carrier types. The return continuous carrier can be transmitted using π/2-BPSK modulation.

Carrier spreading for return link continuous carrier
Continuous carriers are spread to mitigate interference to adjacent satellite systems. The baseline transmission scheme is that of DVB-S2 so that the carrier can use constant code modulation (CCM) or adaptable code modulation (ACM) modes. The characteristics of the return carrier can be signalled by the

RCST or managed by the NCC. The corresponding forward carrier can be transmitted as a standard DVB carrier with or without spreading.

The spreading factor is signalled in the time slot composition table transmitted in the forward link. In the scrambling operation, a scrambling code redistributes the bits to spread the spectral density and ensure presence of sufficient number of transitions to facilitate synchronisation at the receiver. The scrambling code can either be signalled by the RCST or a default code can be used.

CHAPTER 9

1) What are the entities of an MSS network management system and goals in respect of each?

An MSS network comprises numerous functional units which interact with each other to deliver the desired services to the end user consistent with the requisite service quality. The network management task is to monitor the performance of each unit, apply configuration changes as and when needed, and take corrective action to recover from anomalies.

Figure 1 represents a global view of a network management system depicting the entities that constitute the operational MSS network lifecycle.

Figure 1 [A sketch of Figure 9.1
A Telecommunications Management Network structure of a MSS.

Constellation management involves monitoring and controlling the performance of the constellation in its entirety. The system monitors health of each satellite and its orbital location and adjusts operational parameters as necessary, applies contingency measures in case of a satellite failure, supports launch and introduction of new satellites.

Gateway management involves monitoring and managing the performance and functioning of each gateway sub-system. The *switching system*, resident within a gateway, performs switching in relation to core network on a call-by-call basis. The *security subsystem* ensures transmission security as well as physical security of vital elements of the network. Security measures are defined by each network operator according to its operational philosophy. Example of security features include maintaining subscriber confidentiality, subscriber or/and mobile authentication and signaling and data security.

Maintenance involves regular check-out and calibration of equipment throughout the network.

Operations refers to real-time monitoring of traffic-flow, radio-frequency integrity (i.e. EIRP and radio frequency) of each element of the network i.e. satellite, gateways and other feeder stations, mobile terminals and interference, interfacing with other operators and quality of service monitoring.

Configuration management takes care of sub-system/system reconfiguration – for example, spacecraft on-board switching or system reconfiguration in case of a satellite failure. It involves configuring a number of system parameters depending on the network requirement – constellation reconfiguration in case of a satellite failure or reconfiguration of spot beams.

Network management control system deals with issues such as radio resource management in real time or non-real time. It involves frequency planning and EIRP budgeting. Network control function also includes mobility management to ensure connectivity during a call or when the user moves across to other spot beams or satellites. It also involves network information broadcasts to mobile and gateways – such details include frequency of signaling channels, gateway identification flag, space segment configuration, etc.

Information flow is managed so that information is available to the network manager on time for timely corrective actions.

In addition there are numerous supporting tasks. The *commercial support system* deals with customer support on real-time basis, this includes support directly or indirectly through service providers, billing, provision of traffic trends for forward planning or revenue prediction, etc. Administrative and commercial management functions (which include aspects of configuration management) include subscriber management – commissioning of new subscribers, modification of existing user service profiles and billing; user equipment identity database management; user terminal status list management for checking fraudulent use; traffic statistics generation for network provisioning, planning, etc.; report generation, including incident, fault and network change reports.

Long-term planning tasks deal with strategic planning in terms of procurement of new satellites, development of new products, and spectrum forecast. Interface with external bodies would involve liaising with operators, service providers and regulatory bodies in support of issues such as interference management, spectrum negotiations, inter-system co-ordination, collaborative endeavors, publications, etc. The research and development management ensures participation in the industrial collaborative efforts such as standardization, development of new-generation products, tracking recent development to remain competitive.

The *operational* tasks involves monitoring, interpretation and control of a vast number of parameters in various domains - for example, EIRP, carrier to noise ratio, BER in a non-IP environment and packet loss and latency in an IP domain. Such data lie scattered throughout the network. Thus data collection, abstraction, analysis and presentation on powerful graphical user interface facilitate rapid and intuitive interpretation.

2) What are the parameters which must be monitored to ensure a satisfactory quality of service in an MSS network?

The network operator is able to comply with the promised quality of service by maintaining the performance of critical network parameters within specifications. These parameters include satellite EIRP variations with traffic load, frequency of each radio carrier, carrier to noise ratio, network congestion, numbers of repeat requests in an ARQ scheme, message delivery time, throughput statistics, failed call statistics, holding time trend, etc. The MSS network provider often has no control over the quality of service of another network from where a call may originate or to which it is destined. Thus, it is necessary to monitor end-to-end call quality ((periodically) and keep a vigil on user feedback to gain an insight into the network behavior.

Measuring the users' received signal quality during regions' busy hours when the network is fully loaded provides a realistic representation of the users' perception. During these periods, satellites are heavily loaded, resulting in increased inter-modulation noise and a possible degradation in satellite EIRP due to excessive loading; there is an increased probability of interference and congestion, and a reduction in users' throughput in packet switched networks. Quality of service measure includes grade-of-service, end-to-end connection time, bit error rate, packet loss and latency. The congestion observed by a user can be caused in the terrestrial network, the gateway or the space segment. The system has to be provisioned to ensure that congestion is managed at all the levels

3) Explain the process of spectrum management in an operational MSS, including the process of generating frequency plans.

Section 9.4.1 deals with the process of spectrum management in an operational MSS, including the process of generating frequency plans.

4) What are the similarities and differences in the frequency planning process of GEO and NGEO satellite systems?

Similarities
Irrespective of orbital characteristics, the frequency assignment process generates a frequency list applicable to the given coverage area respecting the constraints and complying with the desired quality/grade of service. Figure 1 below presents the frequency assignment process. Raw data is converted to the number of required RF channels mapping the requirements to a frequency plan, taking available spectrum, reuse matrix, optimization rules and constraints as the basis. The mapping function creates a frequency plan applicable at time t by taking into consideration the relative positions of satellites and spot beams. The constraints include practical considerations, such as receiver tuning range, signaling carriers and location-specific interference. The process can repeat for time t_1, if necessary, as would be the case for a non-geostationary satellite system or in preparing a time of day frequency plan for a geostationary satellite system.

Figure 1 [A sketch of Figure 9.3(b)]

Figure 2 shows the real-time frequency assignment process per call or session irrespective of orbital characteristics. On request for a channel (in circuit mode) or capacity (in an IP system), the network control station allocates a channel or capacity on basis of a pre-assigned allocation rule. When dealing with packet switched network where assignments are multiplexed in time the radio resource manager assigns slots on basis of the quality of service negotiated by the user at the time of assignment. For example, a streaming service would get an uninterrupted assignment whereas a delay tolerant service would be assigned slots when higher priority assignments have been served.

Figure 2 [A sketch of Figure 9.3(c)

Differences

For geostationary satellite systems, antenna patterns remain static on the surface of the Earth and hence interference geometry is stationary, whereas in non-geostationary satellite systems interference geometry is dynamic. Thus, frequency management is simpler for geostationary systems in comparison to non-geostationary satellite systems. In geostationary satellite systems, frequency plans are generally static with adjustments required periodically in response to traffic variations - except when a time of day frequency plan is in use. In non-geostationary satellite systems, frequency plans tend to be quasi-static implying that they remain static over parts of the orbit where the traffic and interference profiles are quasi-static.

5) Explain the methods used in short-term and medium term spectrum planning of an operational mobile satellite system.

Frequency lists used by the resource manager should match the traffic needs of the network in real time to comply with the specified quality/grade of service. In an MSS network, traffic varies in the short term over minutes or hours, with an underlying trend in the medium term, which manifests over weeks and months. These variations are modeled in different ways due to differences in their mechanisms and characteristics. The frequency plan is developed on basis of short-term trend, while medium term trend is used for establishing that spectrum needs are met in the medium term. The factors influencing the trends are summarized below.

Short-term traffic trend

It is known that the traffic load carried over mobile satellite networks varies diurnally, with a peak during the business hours. When the service area encompasses several time zones, diurnal traffic profile of each time zone is offset proportionally; hence, the peak traffic carried over the satellite migrates from the spot beam covering the easterly time zone towards the western spot beams.

A 24-hour traffic profile within a single time zone for typical weekday business would show that the traffic rises rapidly at the start of the business; there is a trough in mid-afternoon as people disperse for lunch over an hour, followed by another surge in the traffic. The traffic subsides rapidly after the end of the business day. Weekday traffic for social needs would be expected to rise gradually over the day, with an increase between 6 and 10 p.m. and then tapering off. A weekend traffic profile would show a small back-ground business traffic whereas the social traffic would be higher. A refinement to the model would be to segment traffic by market sector, user terminal and traffic category.

Temporal and location dependence would depend on the environment, geography, traffic type and the time-of-day. Business traffic for land, maritime and aeronautical services would continue to peak in business hours. Traffic for the aeronautical environment would, in addition, be influenced by aircraft route, flight time and duration; land traffic distribution would demonstrate a strong relationship to geography (cities, motorway, railway routes, etc.) and the local telecommunications infrastructure; and spatial distribution of maritime traffic would exhibit a strong correlation to sea routes. This type of segmentation can allow an effective service/time-of-day/spot beam/location dependent radio resource management.

In summary, traffic carried by a satellite is influenced by the following factors and frequency plans must be developed to comply with the trend. One approach used in practice is to provision the frequency plan so that the given grade of service is satisfied at the busiest hour of the day.

1. Spatial distribution of users and their respective time zones;
2. Day of week (working day or not);
3. Temporal distribution of traffic – short (minutes) and long (hours) term;
4. User base i.e. business or social;
5. Service type i.e. voice, data (Internet);
6. Operational environment – aeronautical, land, maritime
7. Anomalous events such as New Year's Eve, war, natural calamity

Medium-term traffic trend

For medium-term planning extending months to years, the underlying trend and external influences are of primary interest. Trend analysis through curve fitting is a simple and effective technique under stable conditions, i.e., periods without anomalous events. The accuracy can be improved by superimposing seasonal trends such as the effects of a holiday season, or a traffic surge such on the New Year day, knowledge of new commissioning, etc. Figure 1 below illustrates the trending approach using an exponential (a), linear (b) and negative exponential (c) curve fitted to a set of hypothetical data. We note by inspection that the error resulting from estimates a and b can become significant whereas those from curve c are accurate demonstrating that unless the method is applied judiciously the error in forecast can become significant. Similarly, unexpected deviation in the usage can distort the well-behaved trend causing errors in this type of forecast.

Figure 1 (A sketch of Figure 9.12)

Trend analysis using curve fitting. Various types of curves have been fitted to a hypothetical set of data and extrapolated for forecast. Note the sensitivity of the results to type of curve (a = exponential fit, b = linear fit and c = negative exponential fit; x = historic data; O = actual growth).

The limitations of this approach and ways to improve the forecast are as follows.

Historic data are not always available. Data are non-existent when a new operator introduces a service or an incumbent launches a new product. In such cases, theoretical estimates based on heuristics, logic, etc. are used.

The impact of external influences is absent; the steady-state assumption may not apply for a number of reasons - for example, introduction of new and/or improved services, unexpected increase in competition, societal change, etc., Refinements to the model can be made by including the impact of external influences.

Product life cycles follow the well-known S curve whose time-scales are influenced by a number of factors – evidently, the accuracy of future projections depends on how well a product life cycle has been modeled. For example, the trend in the decline of analogue systems was much slower than anticipated – operators of analogue systems were able to retain customers by call cost reduction, aided by the availability of low-cost handsets enabled by mature technology.

The accuracy of simple trend analysis should generally be adequate for periods of a few months, depending on the stage of the product cycle, and up to a year if external influences are factored in.

Forecasts should have the necessary granularity for each beam and beam cluster. Spot beams in regions within the same time zone would have identical time profiles but spatial distribution of users can vary; the largest time-offset in the profile will be experienced between beams which are longitudinally furthest apart.

Applicability of traffic forecast to non-geostationary satellite systems

The coverage pattern of satellites in a non-geostationary satellite system is time variant and illuminates a smaller area at a time (except satellite systems operating in high altitude elliptical orbits), implying lower traffic per satellite and a composite diurnal and spatially varying traffic profile through each satellite. The forecast approach outlined has to be applied to include spatial, temporal and orbital characteristics.

6) Outline the methods used to monitor satellite transmissions and interference management in an MSS network.

RF emissions are readily monitored on spectrum analyzers at judiciously sited monitoring stations; a computer-controlled spectrum-analyzer measurement system offers operators the capability to measure accurately numerous parameters remotely. The remote monitoring stations can be controlled centrally from the network control centre over dedicated communication links. To maintain accuracy calibration is performed regularly several times a day. A typical set of measurements, including alarms for anomalous conditions, includes:

- EIRP of a single modulated or unmodulated RF carrier, a group of carriers or a satellite transponder;
- Carrier centre frequency;
- C/No and Eb/No
- Periodic automatic monitoring of authorized carries
- Detection of unauthorized transmissions.
- Detection of interfering carrier

7) Various approaches are used in granting spectrum licenses. List the advantage and disadvantages of those approaches that are applicable to MSS.

In most cases, the MSS services extend over a number of countries and therefore operating license has to be obtained from various authorities, making it a complex, time-consuming task.

There are wide variations in the approach adopted for granting MSS spectrum license, although it is generally based on first-come, first-served basis using established procedures recommended by the ITU. With increasing commercial competition, some regulatory authorities use alternative methods within their jurisdiction to encourage efficient use of spectrum. Sometimes a charge is levied to recover administrative costs. Cost may sometimes be increased to encourage a more efficient use of spectrum with the rationale that the operator will then try to increase the returns by increasing spectrum efficiency. When spectrum shortage is acute, the licensing costs are sometimes increased to reduce the number of competitors. Authorities may invite potential operators to participate in competitive bidding and select the licensee on the basis of criteria such as technical excellence, funding arrangement, level of risk in the proposal, etc. A license may only be granted on a conditional basis, such as proof of technical concept or funding. A major consideration when developing a charging policy for MSS licenses is that if each country charges an operator a licensing fee, the charges may become prohibitive, as satellite operators usually provide services to a large number of countries. An analysis for terrestrial UMTS indicated that profitability and pay-back periods would increase and become detrimental beyond a threshold.

Possible approaches with their merits and demerits are summarized in Table 1 below.

Table 1 [Table 9.2 or nearest] Possible approaches for spectrum licensing with their merits and demerits

CHAPTER 10

1) State the reasons for the financial failure of various NGSO mobile satellite system start-ups in the first decade of the millennium. What lessons can be learnt from the downturn?

The reasons for failure of various NGSO mobile satellite system startups are attributed to:

- over-optimistic market projections;
- stupendous growth in terrestrial mobile systems bolstered with the advent of roaming which virtually eliminated the anticipated market of the travelling businessman;
- high service and user terminal costs, large size, weight with low aesthetic appeal of the user terminal in comparison to cell phones.
- huge infrastructure cost
- Cost escalation

The number of subscribers at the end of 2011 was of the order of 2.2 million compared to the projection of over 30 million. Similarly, the subscriber base for little-LEO satellite systems was estimated as 168 million units whereas subscribers of ORBCOMM (an incumbent Little-LEO operator with maximum share of market) towards the end of 2012 totaled about 689 000.

During the early 1990s, when many big-LEO systems were conceived, cellular telephone sizes were bigger, large areas of the world were not covered by cellular systems, and international roaming arrangements were in their infancy. By the time the MSS systems were introduced cellular phones were smaller, multi-functional, lighter, attractive and fashionable; the terrestrial coverage had shrunk enormously, while data rates increased, as the third generation systems rolled out. Users could travel large parts of the populated world and communicate on the same phone, due to inter-operator global roaming arrangements and standardization in the cellular industry.

Lessons to be learnt

The spectacular failures highlight the need of a sound business plan that allows adjustments to account for risks such as competition.

The business plan must include well-balanced aspects of:

- funding that includes capital, operation, maintenance identifying an appropriate source - own, stock market, bank loans, etc.);
- revenue and returns;
- most importantly in the context of the question, the risks i.e. competition, delays and long gestation period of MSS projects with consequent increase in cost, regulatory hurdles, political difficulties, fund scarcity, launch or satellite failure, slow market penetration, etc.
- Risk mitigation countermeasures.

2) With the help of a diagram discuss the role of the system entities and their interaction at the planning stage of (i) a new MSS venture and (ii) a new product of an incumbent MSS operator.

To be financially viable, satellite system targeting the mass market must be market led rather than technology driven. There are benefits in terms of schedule and cost, leading to healthier returns, by using a systematic planning approach.

Figure 1 below shows the main entities and their interaction in an exercise of this nature.

Figure 1 [A sketch of Figure 10.3]
Main entities and their interaction in system planning

The initiative of a new satellite venture or of a product within an existing venture is initiated through marketing studies and is based on telecommunication needs, preferred size and type of hardware, quality of service expectations, geographical areas, consumers' purchasing power, preferred cost of service, market size and anticipated growth, etc. User requirements have to be converted to revenue and returns through a business model which benefits by incorporating the entire life cycle of the product. A number of iterations are necessary between system designers, marketing and business strategists before a realistic set of requirements, technology and schedule materializes. During the design process, there is a need to incorporate regulatory issues, such as spectrum availability, licensing policy in the target service areas, time lag in completion of each requirement, possibility of obtaining the preferred orbital location and frequency bands, etc. The initial model – which may be based on the operator's own historic data, published reports, documents in the public domain, government business initiatives, empirical models extracted from the literature, etc. – has to be refined by interaction with manufacturers and vendors. Furthermore, the system design concepts require industrial validation by way of technology, risks and schedule. At the end of such an interactive design process, a business plan encapsulating system specifications, operational philosophy, a detailed program schedule, capital and revenue, etc., is ready.

3) What are the advantages of a systematic planning approach for the development of a commercial mobile satellite system?

Commercial satellite systems are planned with the goal of achieving the maximum return. Using a system approach, which includes the whole life of the venture, identifies areas of potential savings and risks. The analysis provides a clearer view of the commercial venture to investors and shareholders. It allows returns to be estimated and facilitates sensitivity analysis of system parameters in terms of profit or another figure of merit. An example would be the profit/unit spacecraft EIRP ($/W) or profit/unit bandwidth ($/kHz) or profit in terms of weighted combination of spectrum and satellite EIRP.

4) Discuss the concept and advantages of a lifecycle cost-benefit analysis with examples of inputs and outputs into the process.

Commercial satellite systems are planned with the goal of achieving the maximum return. Using a system approach, which includes the whole life of the venture, identifies areas of potential savings and risks. Such a lifetime cost-benefit analysis provides a clear profile to investors and shareholders

A preliminary cost analysis of the mission based on available information such as to provide a rough order of magnitude (ROM) costs, can assess mission viability and allows readjustments. Lifecycle program cost estimation requires space segment cost, ground segment capital costs, operational and maintenance cost for replenishing the space and ground segments, etc. Sensitivity analysis of system parameters is feasible in terms of profit or another figure of merit. An example would be the profit/unit spacecraft EIRP ($/W) or profit/unit bandwidth ($/kHz) or profit in terms of weighted combination of spectrum and satellite EIRP.

Figure 1 below shows a top-level model of a lifecycle cost–benefit analysis. The inputs to the model comprise:

- *mission objectives and constraints:* financial objectives, lifetime of the mission based on operational considerations, service environment (e.g. land, maritime, aeronautical), service type (e.g. throughput - low, medium or broadband), product range and their lifetime, etc., and system constraints;
- *system characteristics:* alternative network architectures and concepts, orbital and constellation characteristics, spacecraft and terminal characteristics, system drivers (e.g. user expectations of terminal size and throughput);
- *evaluation criteria:* Baseline designs, figure of merit (cost, financial returns quality of service, spectrum efficiency), risk (technology, schedule, etc.), receiver complexity, system capacity and cost; penetration rate and revenue; service cost and penetration rate, cost tradeoff between owned versus leased space segment;

- *knowledge*: cost models, spacecraft power/mass estimation model, user terminal cost model, risk strategy, growth trend of similar services, inflation trend, traffic distribution;
- *System data*: link margin, operational frequencies, available spectrum, available finance, expected return, historic data of similar products such as unit cost, call charge, etc.;
- Other considerations and inputs as listed in figure 1.

The output from the model provides the requisite information for system optimization. Example performance evaluation criteria include: profit and revenue as a function of time; satellite EIRP and bandwidth for profitability within acceptable time limits; service penetration rate to achieve profitability within the target time frame, etc.

The feedback loops offer the planner opportunity to vary inputs and adjust the performance to achieve the desired goal.

Engineering trade-off analysis is used by planners at the outset to cut costs with minimal risk and expenditure. For example, an ESA study estimated a cost reduction by a factor of two for production of 15 spacecraft when using economies of scale.

These studies benefit greatly through iterations as a better appreciation of requirements, conceptual design and cost sensitivities is gained. For example, the perception that the lowest spacecraft cost is the best option may not necessarily hold when the venture is viewed in its entirety. This study phase involves close interaction between the business and engineering teams and is vital to the eventual success of a venture.

Figure 1 [A sketch of Figure 10.4]
A top level model of a lifecycle cost-benefit analysis (adapted from Sultan and Groepper 1999)

5) Identify the factors which contribute towards revenue improvement of an MSS operator. Explain their individual and collective role.

Figure 1 below shows the factors likely to affect usage and hence revenue to the operator. To maximize revenue an optimum balance between these is necessary.

Consider *revenue* as the main driver. It depends on *system usage* which is influenced by user terminal cost, extent of users' satisfaction with the quality of service; revenue pilferage by competition, effectiveness of the infrastructure; marketing issues such as equipment availability, promotion, after-sales service; and of course, the call and service cost. Call cost depends on the *charging policy*. Charging policy is governed by a number of independent considerations, e.g., a minimum threshold for business viability, geographical dependence (i.e. user affordability), telecommunications trend (reducing call charges, migration to IP data services), competition from other operators or services, consumer feedback, investment return, operational and fixed costs, etc. The *business strategy* deals with the overall business direction which considers revenue goals, market conditions, shareholders interest, etc. The revenue details are fed into the company's *business plan* to assess the performance of each product to assist in making the strategic decisions. The revenue is monitored by strategists who take corrective action to meet the sales target in compliance with expected investment returns. For instance, if the usage is not up to the desired level, strategists may apply a corrective action depending on the circumstances, such as ensuring adequate availability of user equipment, altering call charges to a more acceptable level.

Figure 1 [A sketch of Figure 10.10] Factors likely to affect the usage and the revenue of an operator.

6) Outline the similarities and differences between the satellite service-distribution schemes used in practice.

Service distribution schemes are outlined in section 10.3. The student should analyze and collate.

7) Space segment usage and call charge impact the revenue of an MSS operator directly. The operator attempts to maximize the usage of the available space segment radio resource namely spectrum and EIRP. Estimate the revenue generating capability of a single-satellite regional system specified below operating a circuit-mode voice service at call charge ranging 0.25$ -2$ per minute. State the assumptions and suggest whether the system is spectrum or power limited.

Spacecraft EIRP = 73 dBW; available bandwidth = 6 MHz; bandwidth per channel = 5 kHz; spectrum utilization efficiency = 90%; average EIRP/channel = 25 dBW; cluster size = 7; spectrum re-use capability (reuse factor) = 5

The reuse factor F_{ru} relates the total number of spot beams (N_s) and cluster size (C_s) as follows (Equation (10.1) of book) :

$F_{ru} = N_s/C_s$

Given,

N_s = 7 x 5 or, 35

Bandwidth/cluster = 6 MHz, hence

Bandwidth/spot = (6000/7) x 0.9, or 771 kHz

Number of channels/spot = 771/5 or 154

Number of channels supported on satellite = 154 x N_s, or 5390

Total EIRP consumed = 5390 x EIRP (W)/channel

= 5390*316.227, or 62.316 dBW

<u>It is concluded that this is a bandwidth limited system</u>.

Revenue earning capability per minute ranges from

$ 0.25 x 5390 to $2 x 5390 or, $1347.5 to $10780

Minutes per year = 60 x 24 x 365 or 0.5256 Million

<u>Revenue per year,</u> assuming a fully loaded satellite throughout the year, therefore ranges between,

1347.5 x 0.5256 and 10780 x 0.5256 Million $

<u>Or, $708.246 m and $ 5665.9 m</u>

<u>For 50% average loading,</u> the revenue ranges between

<u>$354.123 m and $2832.98 m</u>

8) Outline the main features of the forecast methodology used by:
(i) The UMTS forum during the planning stage of the third generation mobile systems
(ii) The ITU to forecast spectrum requirements up to the year 2020.

Compare the main features of these two methodologies.

(i) The UMTS forecast methodology is covered in 10.6.1 under the sub-heading, 'UMTS methodology'
(ii) The ITU forecast methodology is covered in 10.6.2 under the sub-heading 'ITU traffic and Spectrum forecast methodology'

The student should select appropriate comparison criteria and tabulate.

CHAPTER 12

1) System features of a mobile satellite broadcast service are conditioned by numerous practical considerations and constraints such as small screen size, reception on a variety of receiver types, uninterrupted and ubiquitous service, etc. Based on such issues, suggest a set of service requirements of a mobile satellite broadcast system. Give an example of service features of an advanced mobile broadcast system.

The requirements of a mobile satellite broadcast service are governed and conditioned by the operating environment (i.e. fixed, nomadic, urban, rural, etc.), target receiver cost and form, service charge and the anticipated usage of the service. As such, these requirements differ from the conventional broadcasts. For example, the size of the viewing screen of a mobile receiver is considerably smaller which influences the transmission format and radio bandwidth, while the user is interested in a different type of content than a home user. Transmissions for display on small screens require a considerably lower bandwidth (e.g. 200 kHz) than those of conventional satellite broadcasts.

The broadcast system should be capable of the following service provisions:

- high-quality content delivery;
- flexible configuration of each service i.e. audio, video, ancillary and auxiliary data;
- access to contents and services, possibly controlled through conditional access, protocols or other content protection mechanisms;
- seamless access throughout the broadcast network;
- fast identification and selection of content and services;
- mechanisms to minimize power consumption and physical size of the receivers;
- stable and reliable service coverage in the targeted environments;
- interactivity, (e.g., interactive content)
- efficient and reliable delivery mechanisms;
- technical aspects that enable interoperability between broadcast and telecommunication networks (e.g. content format)
- service quality comparable to fixed reception
- countermeasures for shadowing, multipath, Doppler and vehicular motion

<u>Examples</u>

In practice, the requirements would be conditioned by commercial and technical factors. A basic system would comprise an Electronic Service Guide (ESG) to announce the portfolio of available contents with its associated location (e.g. local, wide-area) and a hand-over mechanism to ensure seamless coverage in situations when the user moves from a satellite-only coverage to a terrestrial retransmission zone.

A more expansive set would attempt to emulate features of an advanced terrestrial broadcast system such as DVB-H that has a provision to support:

- Mobile TV programme broadcasts, possibly with associated auxiliary data (e.g. links to the service provider's web pages, video clips, sound tracks, games, etc.);
- Enhanced mobile TV to provide interactivity to include features such as online TV shopping, chat, gaming and quiz plus voting, etc., where the interaction could be achieved through a mobile communication system;
- Scheduled download facility (announced in the ESG) to provide audiovisual content or executable software;
- Provision for (online) service purchase, service access and content protection, including pay-per-view purchase;
- Roaming facility that allows contents access when outside the home network;
- High quality of service to ensure compatibility with conventional terrestrial and provide error-free downloads of data files in presence of shadowing, multi-path, Doppler and interference;
- System and receiver architecture that facilitate low power consumption at the receiver;

- Interaction in a mobile environment - an obvious solution would be to incorporate the interaction through an available mobile system.

2) Compare the characteristics of alternative mobile satellite broadcast system architecture, including an example of each. Suggest a configuration suited for broadcasts to subscribers dispersed in a mix of rural and urban environment.

Please use Table 12.2 and Table 12.5 and associated description.

3) State the considerations applied in the selection of the space segment to provide a reliable broadcast service, including the countermeasures which can be used to increase the robustness of the broadcast radio link.

One of the prime considerations in selection of the space segment for a satellite mobile broadcast service is the earth-space geometry which has a profound impact on the link reliability because fading loss is inversely proportion to the elevation angle.

Therefore geostationary satellite systems are suitable for lower latitude regions where elevation angles exceed ~20^o. At mid and high latitudes, where elevation angles are low, the propagation losses become prohibitive.

Highly elliptical orbit (HEO) can provide quasi-stationary high elevation coverage at mid-high latitudes. For example, a Tundra orbit would cover Europe at elevation angles of ~55 to 90°. An added advantage of the configuration is that the satellite eclipse could occur near the perigee when satellites are not in service. At least two satellites phased 180° apart would be necessary to provide continuous coverage (at >52° elevation) in Europe. Since the HEO configuration requires handover between satellites, the system architecture gets more complex than an architecture using a geostationary orbit. The HEO architecture is used in the Sirius XM Radio Inc. which provides satellite radio services in the American continent; it was proven that improvement in link margin reduced the requirements of the number of the complementary terrestrial transmitters when compared to an equivalent geostationary satellite system.

Approaches to mitigate the propagation impairments are spatial diversity by multiple satellite visibility and time diversity by transmission repetition sufficiently distanced in time. The premise for spatial diversity is that the probability of at least one of the satellites being in a favorable direction of the user increases; and for time diversity the premise is that the probability of a fade lasting more than a specific duration reduces.

4) Discuss the OSI model as applied to a satellite mobile broadcast system. Illustrate its applicability to a practical system with the help of an example.

Please see section 12.7

5) State the salient features of each of the five ITU-R-recommended digital mobile satellite broadcast system and highlight the differences between them; suggest the system(s) which could be used to provide reliable broadcasts to a mid-latitude region such as Europe that harbors diverse languages and culture.

Please see Table 12.5 and the associated text.

6) With the help of a block schematic discuss the functioning of a mobile receiver of a hybrid satellite system that supports spatial, time and frequency diversity.

Here we outline the receiver architecture of Digital System $D_{H,}$ one of ITU-R recommended systems. Fig. 1 below presents the receiver architecture of a dual-diversity (space-time) space-terrestrial hybrid system configuration.

Figure 1 [A sketch of Figure 12.6]
A receiver architecture applicable to a satellite-terrestrial hybrid system with space and time diversity in

the space segment.

System D_H broadcasts digital audio and data for reception by inexpensive vehicular, fixed and portable receivers in 1452-1492 MHz band incorporating spatial (dual satellite) and time diversity. It utilises coherent QPSK modulation with block and convolutional error coding, and linear amplification. The audio programmes are multiplexed and transmitted as a TDM signal. The channel of user's interest is selected from the TDM data streams to recover the desired digital baseband.

The terrestrial delivery utilizes an orthogonal frequency division multiplex technique called multi-carrier modulation (MCM) to contend frequency-selective fades observed on terrestrial wide-band channels.
The satellites are fed with the same signal but transmission to one of the satellites is delayed by 4.28 seconds. The receiver utilizes early-late time diversity in conjunction with maximum likelihood combination of diversity signals.

The satellite-only receiver architecture is enhanced to take advantage of the terrestrial component and the diversity. Thus, it includes two additional receiver branches – one to receive the second satellite signal and the other for the reception of the terrestrial component. A single antenna can be shared for the reception of all the three signals. The diversity satellite signals are received by a receiver identical to the single satellite receiver with an additional provision to combine the early satellite signal to achieve time and space diversity gain. An additional branch is necessary to receive the terrestrial MCM signals with a provision to select the best option between satellite and terrestrial components. It is possible to conceive of a scheme wherein all the signals are combined synchronously.

In figure 1, all the receiver branches share the antenna system. One branch carries early broadcast channels and the other the delayed broadcast channels. The third branch receives the terrestrial time-division-multiplexed MCM transmissions derived from the terrestrial broadcast channels. Each satellite branch comprises a satellite tuner that selects the desired TDM satellite carrier, a QPSK demodulator to recover the TDM symbol stream and a TDM demultiplexer. The early and late versions of the TDM signals are fed into a Viterbi maximum likelihood FEC trellis decoder to combine the two signals. The delay of the early signal is implemented in the TDM demultiplexer. Viterbi decoder combining is accomplished by aligning the preambles of the early and late broadcast channel frames. The terrestrial branch operates simultaneously and independently of the satellite branches following the same sequence as in the satellite branch. Synchronization for combination with satellite reception is achieved by aligning the preambles with the early and late broadcast channel frames. A post-detector logic can be used to select the component which has a better signal quality to provide the desired broadcast channel.

7) Highlight the salient features of the DVB-SH system.

Please refer section 12.10

8) What are the main differences in requirements, characteristics and features between fixed and mobile satellite broadcast systems.

The requirements of a mobile satellite broadcast service are governed and constrained by the operating environment (i.e. fixed, nomadic, urban, rural, etc.), receiver cost and form, service charge and the anticipated usage of such a service. As such, these requirements, characteristics and features differ from the conventional broadcasts. The main differences are:

- size of the receiver and viewing screen is considerably smaller which influences the transmission format and radio bandwidth;
- user requirements of content are likely to differ;
- transmissions for display on small screens require a considerably lower bandwidth (e.g. 200 kHz) than the conventional satellite broadcasts (e.g. several megahertz).
- unlike the fixed broadcast satellite system transmission format for mobile broadcasts has to be designed to contend fading, multipath and Doppler

9) Discuss the feasibility of adapting a hand-held receiver for reception of K_u band direct broadcast transmissions.

This option has not been discussed in the book. It is for the student to research and suggest the feasibility. At least the following technical considerations are essential:

- *Link budget:* Is it at all possible to provide the desired signal quality on a hand-held terminal which has a low G/T (say, -24 dB/K) when the link has been optimized for a considerably higher G/T (say, 7 dB/K). Can the link ever close on a hand-held?
- *Rain fading and mobility-induced impairments countermeasures*
- *Handheld antennas at Ku band*
- Feasibility of local retransmission for hand-held reception in areas of interest
- Regulatory and commercial restrictions, including retransmission frequency band

CHAPTER 13

1) Highlight the salient features of Cospas-Sarsat system. State the limitations of the existing space segment and the measures being undertaken to mitigate these.

Please see section 13.2.1 of the book.

2) Outline the navigation principles used in systems such as GPS based on pseudo-range measurements including the error mechanisms and method to mitigate these errors.

Figure 1 below represents the concept of a two-dimensional navigation solution including range errors. When range between the observer's location and satellite and satellite's location is known it is possible to estimate observer's location by solving a set of simultaneous equations. As illustrated in the figure, initial position of the observer is estimated within a zone of uncertainty marked by hatched portions. When errors are eliminated, the solution converges to a single point.

Figure 1 [A sketch of Figure 13.6] Concept of position fix and impact of errors in two dimensions

The range is estimated as the product of the time taken for a signal to travel from the satellite to receiver and velocity of electromagnetic wave in the intervening medium. To estimate range precisely, each user clock must be synchronized to the satellite clock. To simplify receiver architecture in the GPS system, the range from a fourth satellite is estimated, which allows resolution of user clock uncertainty.

The travel time of the signal is estimated by measuring the time shift between identical codes generated at the satellite and the receiver. The code generated at the receiver is time shifted until a maximum correlation is achieved between the transmitted and receiver codes; the time-shift provides an approximate range or 'pseudo' range which includes numerous errors. The user must have knowledge of code to be able to use the system. Since the received signal is about 22 dB below the receiver noise in the (first generation) GPS system, code acquisition and de-spreading is necessary prior to clock recovery.

To fix a position on the Earth a terrestrial reference system is necessary. The Z-axis coincides with the earth rotation axis; the X-axis is associated with the mean Greenwich meridian; and the Y-axis is orthogonal to the other two axes to complete the right-handed coordinate system. A number of earth reference systems are used based on their approximation to fit the shape of the Earth.

Pseudo-range R_p is given as:

$$R_p = R_a + c\Delta t_e + c(\Delta t_r \pm \Delta t_s) \qquad \ldots \qquad 1(a)$$

Where,

R_a= Actual range

$$R_a = \sqrt{(X_s - X_r)^2 + (Y_s - Y_r)^2 + (Z_s - Z_r)^2} \qquad \ldots \qquad 1(b)$$

c = velocity of electromagnetic wave

Where subscript s and r are respectively represent the cartesian coordinates X, Y and Z of satellite and receiver respectively.

Δt_e= Propagation delay inclusive of various error components

Δt_s= Satellite clock-offset with respect to GPS time

Δt_r= Receiver clock offset with respect to GPS time

c = velocity of light

The true range is estimated by solving a set of four simultaneous equations obtained by substituting 1(b) into 1(a). The matrix is populated by pseudo-range measurements from four satellites.

The solution can be explained qualitatively by observing that the receiver lies on the intersection of three sphere of radius R_{pi} R_{pj} and R_{pk} where subscripts i, j and k represent pseudo-range distances from satellites i, j and k respectively. The fourth measurement is used to resolve clock accuracy of the receiver.

There are a number of sources of error in range estimation:

- **Satellite clock offset relative to GPS system time and ephemeris errors:** GPS system time is maintained by the GPS master control station (MCS) through a set of highly accurate cesium clocks; the clock offset of satellites is measured daily and transmitted to each satellite by the MCS for retransmission to receivers which apply the correction algorithmically; satellites themselves incorporate highly stable atomic clocks onboard. Since certain components in ephemeris errors cannot be isolated from satellite clock offset errors they are combined with the satellite clock error in the error budget.
- **User clock offset from GPS system time:** As mentioned in the preceding text, user clock offset can be removed by solving the range equation.
- **Error due to propagation delay:** An error in range estimates is caused by delay introduced by ray bending and velocity reduction while traversing the ionosphere. The delay is approximately inversely proportional to the square of the frequency and in the GPS system the correction can be derived by comparison at two frequencies, 1,227.6 and 1,575.42 MHz. In the GPS system, it is derived at the MCS and transmitted to users. Tropospheric delays are independent of frequency and can be estimated at the receiver by applying an elevation-angle-dependent correction.
- **Group delay** error is caused by processing delay on a satellite; its value is estimated in ground tests and transmitted to users with other corrections.
- **Multipath** errors are caused by signals arriving at the receiver from different paths. It is difficult to remove them as such but multipath rejecting antenna systems can mitigate its impact.
- **Receiver noise and resolution degradation** are caused by hardware and software limitations of the receiver. Receiver motion can introduce additional errors. Use of well-designed receivers and filtering algorithms such as the Kalman filter reduces the impact of the error.

3) Explain the principle of differential GPS with an example application.

It has been observed that ephemeris and measurement errors are correlated in time and spatially. Spatial correlation decreases with distance rather slowly – correlation is close even for a distance of a few tens of kilometers. Similarly temporal correlation decreases slowly with time. Propagation errors, and S/A errors are quite well correlated in a 5–10 s time span.

An accurate estimate of such errors can be obtained when the coordinates of a location are known precisely; if these corrections are transferred to users in the vicinity of the measurement site, and the user applies such corrections, the accuracy of the fix improves considerably. Application of this technique to the GPS system is known as differential GPS (DGPS), where errors in the navigation solution are derived at a reference site and transmitted over a radio link to GPS receivers present in the neighboring areas.

Measurements and simulation demonstrate that errors range from tens of centimeters when the reference site is a few kilometers, to about 5 m for a distance of 1,000 km.

Example of a DGPS-based navigation solution is the Wide Area Augmentation System (WAAS). WAAS provides GPS-based information for aeronautical en route, departure and approaches under conditions where ceiling and visibility are 200 feet (60.96 m) and ½ mile (804.67 m) respectively. The measured 95% horizontal and vertical accuracy within CONUS (Grand Forks) during the interval 1 July to 30 September, 2012 is reported as 1.46 m horizontal and 1.7 m vertical respectively

4) Describe the characteristics of the Galileo system and compare them to GPS and GLONASS systems.

Please refer section 13.3.3 of the book. At least compare ownership, purpose, services, navigation principle, constellation characteristics, signal characteristics, user community

5) What are the issues affecting mobility of VSATs. Suggest solutions to mitigate such mobility-associated problems.

<u>Issues</u>

1. VSAT systems belong to the FSS, which implies that regulatory matters pertain to fixed installations;
2. VSAT radio link performance, optimized for static links and rain fading in K_u and K_a bands, is unreliable in a mobile environment where the signals undergo shadowing and multipath impairments; VSAT antennas are fixed and relatively relaxed antenna mounting restrictions whereas MVSAT antennas require satellite tracking, compliance to stringent mounting space and aero-dynamic considerations;
3. MVSAT systems for high speed vehicles, i.e. railway and aircrafts, must include Doppler counter-measures;
4. MVSAT transmitters should be highly efficient in order to reduce terminal size and minimize power drain;
5. VSAT system coverage in C band is available globally but K_u band coverage is generally tailored for fixed land applications on spot beams and therefore there may be gaps in connectivity on inter-continental routes.
6. Due to the large size of MVSAT antennas they are better suited for shared and captive environments such as on ships, railway, aircrafts, etc.; although man-pack MVSATS are available for special applications but L band MSS user terminals remain the primary choice for global personal and portable usage.
7. MVSAT systems can provide a more efficient use of spectrum (i.e. the orbital arc) due to the higher directivity available from MVSAT antennas when compared to L-band MSS systems designed to support a mix of user terminals including hand-held and small terminals deploying low-directivity antennas.
8. Due to the size restrictions of MVSAT antennas, the side lobe levels are higher than those of fixed VSATs and thus to comply to radio regulatory requirements the power spectral density of return links should be reduced to remain within the specified mask while the forward link must be robust to reject interference from other satellite systems received at MVSATs.
9. When FSS (rather than MSS) allocations are used for mobility, regulatory clearance is arduous due to sharing constraints imposed in FSS bands. Trans-border communication is not permitted in large parts of the world. In contrast, MSS bands have primary and exclusive allocations throughout the world and the trans-border risk is significantly lower.
10. MVSAT system architecture needs to be more robust than at present with spares and back-up provisions to achieve the reliability offered by L-band systems.

Solutions

The key enhancements to VSAT systems to incorporate mobility apply to antenna and the transmission system.

Antenna systems should be efficient and compact with tracking and low side-lobes to minimize harmful interference into adjacent satellite systems. The small size of the antenna conflicts with low side lobe requirements and hence in order to be compliant to the radio regulations the emitted power spectral density is reduced which, therefore, affects the transmission design.

The antenna system requirements and capability depends on the restrictions imposed by the mobile, i.e., operating environment, mounting space, stability of the mobile and speed of travel. Mounting space restrictions in medium/large ships is relatively relaxed but the pitch and roll movement require correction; high speed of aircrafts requires tracking agility with low drag; the restricted mounting space on road vehicles restricts the size and profile of the antenna; limited clearance space of tunnels and high speed require low profile, agile antennas for railways. Phased arrays provide an agile low-profile solution; but their performance tends to degrade at low elevation angles. Hybrid solutions utilize a combination of mechanical and electronic steering

An efficient scheme to minimize interference in to adjacent satellite systems is to reduce the power spectral density (PSD) of signals by spreading the signal. Spread spectrum with CDMA is therefore preferred in many systems; low order PSK schemes, i.e., BPSK/QPSK are preferred over high order schemes where spectral spikes would exceed the power spectral density mask. In order to optimize radio links at connection level, schemes based on adaptive modulation/coding schemes offer advantage. The impact of forward link interference from adjacent satellites and fades can be negated on each link through modulation and code rate adaptation.

MVSATs operating on land suffer the heaviest shadowing and multipath degradations. Railways and road vehicles exhibit different fading characteristics. Aeronautical and maritime channels have stable links for a majority of time barring instances of shadowing caused by mobile structure and when operating at low elevation angle. However, the dispersive nature of channels may manifest under specific conditions.

The mechanisms impacting atmospheric propagation are identical in each environment but its manifestation at the mobile receiver depends on the speed and direction of vehicular movement. Rain cells move with wind and typically have a size of only a few kilometers – for example, rain cell size ranges from less than 1 km to over 20 km in a tropical region with about 70.5% of the cells of 1 km and less. Therefore rain fades get conditioned by the velocity of the mobile due to variations of the rain-affected portion of the radio path.

A prolonged blockage lasting tens of seconds or minutes such as while traversing a tunnel can break an IP connection and result in a loss of receiver synchronization. The ATC scheme as proposed for the MSS and satellite radio broadcasts is an effective solution in these disadvantaged propagation environments.

Implementation issues include the selection of an appropriate space segment and frequency band compliant to the radio and local regulations, network architecture, quality-of-service management at the physical layer and at the user level, selection of a transmission format commensurate with the mobile environment. These issues are identical to those applicable for MSS. Nevertheless the present MVSATs do not provide resilience available from established L-band MSS systems which incorporate robust ground and space segment redundancies, advanced network features such as handover, global technical and logistic support, multiple service providers and equipment vendors, connectivity with GMDSS system for search and rescue support, etc.

Frequency bands where MSS has a primary status (e.g. 29.9-30.0 GHz earth to space/20.1-20.2 space to earth) reduce the regulatory constraints in relation to mobility. For this reason next generation MSS systems can provide MVSAT services on these frequencies without the regulatory burden.

6) Compare the architecture of a terrestrial cellular system with a MSS spot beam system stating similarities and differences.

Figure 1 below portrays a generic architecture of a terrestrial mobile system.

Figure 1 [A sketch of Figure 13.12] The architecture of a terrestrial cellular system.

The main entities of the system are

- mobiles;
- a network of cells each served by a base station; notice that on the contrary all the satellite cells are served centrally or a few earth stations;
- one or more mobile switching centers (MSC), depending on the size of coverage area, interfaced with base stations at one end and the public network at the other;

The radio link between mobile units and a base station consists of two types of channels – a control channel and traffic channels. The control channel transfers system messages and traffic channels carry traffic and supervisory signals during a call. The reader should notice here the similarity with an MSS's broadcast channel. A base station is connected, usually by land lines or backhaul satellite links, to an MSC with a group of voice trunk and data links for exchange of information and to process calls. The analogy between a base station and MSS earth stations should be evident. The MSCs are connected to the public networks.

Typically, each mobile telephone is assigned a ‘home area’. The home MSC maintains a location database (Home location register or HLR) in which it keeps the most recent position of each mobile registered with the MSC. This information is used for routing calls to the appropriate section of the network. Each mobile automatically registers its location with the visiting MSC whenever the mobile migrates outside its home MSC. The visited MSC transfers the information to the mobile’s home MSC and stores it within its visitor location register (VLR). Whenever a mobile telephone is switched on, it scans all system control channels and locks to the strongest channel. This operation is continuous, ensuring that the mobile always operates with the strongest signal. The control channel is used for two types of messages: general system information, which contains network identity, available channels, area code, other facilities/requirements; and mobile control information, which consists of paging messages to notify a mobile and channel assignment messages used to set up calls. If the mobile signal quality degrades during a call, the call is handed over to the cell which can provide a better link quality. The handover involves signaling between the base stations involved and the MSC and is transparent to a user.

Similarities with satellite systems
Conceptually GSM based satellite systems have a similar architecture. Spot beams can be viewed as spot beams and earth stations as base station. The ground infrastructure in satellite systems can share the terrestrial core network. Many satellite systems have used such an arrangement for mobility and inter-network working.

Satellite systems use control and a traffic channel much as cellular systems and handovers bear similarity with terrestrial systems.

Differences
Air interface of a vast majority of satellite systems are different due to differences in operating environment. Spot beam sizes are hundreds of km wide whereas terrestrial cell size are 100 m to a few 10s of km at most. The concept of cell splitting does not apply to satellite systems due to satellite antenna technology limitations and sparse traffic density, Frequency of operation and available spectrum differ. Whereas cellular systems can operate in shadowed environment satellite systems can at best operate under lightly shadowed conditions. Mobiles used in MSS systems are larger. Throughputs of differ vastly. Handsets of cellular systems are cheaper and smarter.

The concept and complexities of space segment management is replaced by base-station and different genre of network management. The coverage area MSS are hugely larger than those of cellular systems.

The answer can be supplemented with extracts from table 8.8 (Similarities and differences between a satellite and terrestrial system used as the basis in the development of GMR-2)

CHAPTER 14

1) State the assumptions used in the recent market projections outlining the risks associated with each assumption.

In general the accuracy of a forecast is impacted by diverse external forces and market variables. Differentiators include economic conditions, government policies and rate of service penetration. Since forecast far ahead into the future of a rapidly changing mobile communication ecosystem is fraught with uncertainty due to the dynamics of the marketplace and the variable such forecasts are taken as an indicative guideline and deviations are factored as the technology and markets evolve. It is a practice to provide low, baseline and high range in forecasts.

The following assumptions were uses in at least one of the marketing studies.

- MSS growth is driven by data applications.
- Availability of new generation MSS systems facilitates development of new application and a broader user base.
- The emerging regions contribute to a healthy growth.
- Government and military remain key markets.
- VSAT competition is likely to erode high-end MSS market to an extent.
- Ka band MSS VSAT system is likely to influence the impact of FSS MVSAT.
- Prices are likely to reduce due to competition.
- Synergistic solutions of MSS with other technologies provide growth opportunity.

2) Assume that the total number of in-service units globally at the end of the decade as 5 million distributed between land mobile, maritime, handheld and aeronautical platforms respectively as 50%, 25%, 22% and 3%. Further, assume that the users are distributed uniformly and service is provided globally by a three region geostationary satellite system with a single satellite covering each region. Estimate the spectrum required in each region to provide a packet-mode service. State each assumption qualifying the basis used.

[Hint: Assumption may include: Aggregate busy hour traffic per satellite, traffic symmetry in forward and return direction, frequency band such as L band, number of spot beams per satellite and reuse factor, spectral efficiency in bits/Hz, etc.]

Platform split (million) at end of decade:

	Global	Each region
Land	2.5	0.833
Maritime	1.25	0.417
Hand-held	1.1	0.367
Aeronautical	0.15	0.05

The following steps may be used; but other solutions are possible.

1) Assume a digital system with different service classes based on platform. For example, three classes – One each for hand-held, land, and maritime/aeronautical.
2) Assume throughput capability of each class (e.g. 64 kbps for hand-held, 2 Mbps for land and 10 Mbps for maritime and aeronautical).
3) Estimate aggregate busy-hour traffic in each region (Assume that the aggregate can serve all quality grades (e.g. streaming, interactive non-interactive, etc.).
4) Split traffic between sectors.

5) Assume total number of beams per satellite as, say, 350 (Typical of modern satellites).
6) Assume a 7-cell reuse pattern (Typical value, although a value of 4 has also been used in practice).
7) Obtain traffic per beam for each class (Total traffic/350).
8) Assume spectral efficiency for each service class based on practice (e.g. 1.5 Hz/b/s for hand-held, etc.).
9) Estimate bandwidth for each service in each beam knowing the spectral density and busy-hour traffic.
10) Assume symmetrical traffic so that forward and return bandwidths are identical.
11) Obtain bandwidth per cluster ($B_{cluster}$).
12) Since the same frequencies are reused in each the 50 clusters (of (350/7), $B_{cluster}$ is the required bandwidth per region.
13) Since traffic is uniformly distributed the same bandwidth is necessary in each beam and in each region.
14) Since on-move service is necessary, the preferred frequency band in L-band.

3) Outline the concept of an iterative interference cancellation multi user detector.

Please see section 14.4.1

4) State the difference between frequency plan and cross-layer optimisation schemes? How are these two interrelated in view of the fact that each strives to enhance spectrum utilisation.

Frequency plan optimisation is a physical layer scheme where radio frequency carriers are arranged in a transponder optimally taking into considerations only physical layer issues such as intermodulation noise, interference, grade of service, etc. The scheme does not take into consideration upper layer issues.

The vigorous growth in data communication has been driven by IP enabled solutions which are based on a rigid hierarchical protocol architecture where each layer communicates with its immediate neighbours and its peer on a pre-defined fixed interface and rules, without cross-layer interaction. The rationale in the emerging discipline of cross-layer optimisation is that in wireless communications cross-layer protocol interactions can lead to a more efficient performance of the transmission protocol stack. By sharing information dynamically, each layer can readjust performance in a co-ordinated manner such that the overall resource utilisation is used more effectively. This approach can potentially enhance the efficiency of radio resource of wireless systems which are prone to propagation-induced impairments typical of mobile satellite systems.

The interrelationship exists indirectly through translation of various types of traffic to the desired throughput say at the busiest hour. The frequency plan is developed on this optimised throughput.

5) Describe the principle of cognitive radio with an example of a potential application in MSS.

The existing methods of spectrum allocation are rigid, bureaucratic and time consuming. Several parts of allocated spectrum remain under-utilized; and even if such allocations are in use, there are periods during a day that the spectrum remains idle. Thus there is considerable interest in promoting dynamic use of vacant spectrum on an ad-hoc, opportunistic basis using agile reconfigurable transceivers. The technology involves the terminals to sense status of spectrum usage and reconfigure their operating frequencies to utilize vacant parts, and for the network to support such an ad hoc dynamic access. This type of dynamic spectrum access can be implemented through cognitive radio (CR) systems.

A cognitive radio system (CRS) includes capabilities to:

- acquire knowledge of its operational radio environment, applicable geographical environments and policies, its internal state and usage
- adjust dynamically its operational parameters, including the protocols
- learn

The operational radio environment includes current spectrum usage in the system's surroundings, identity of operational radio systems at the present location, interference levels in available parts of the frequency bands, etc. Geographical information typically consists of CRS's own location; and orientation, range, coverage and radio characteristics of the radio systems of the CRS's interest. The internal state of the CRS includes usage requirements, available frequency band, frequencies and protocols in use. The policies refer to the applicable terms and conditions for usage of the available primary spectrum such as interference threshold and time restrictions. Usage considerations include users' immediate needs such as desired bandwidth, characteristic of the traffic, usage trends of the primary system.

A cognitive radio system can acquire the knowledge to perform dynamic access through various sources and means, i.e., spectrum sensing, assessment of its internal status and database, access to a central database, access to a CRS broadcast channel where available, trending of acquired data, location-fix from a satellite or terrestrial navigation system, interference measurement. The required knowledge depends on the design methodology of the CR. For example, in the *interweaved or spectrum sensing* approach the secondary user utilizes the available empty frequencies of the primary user as conceived initially; in the *underlay* approach the user utilizes the spectrum of the primary user such as not to cause harmful interference to the primary user; in the *overlay* approach the secondary user has and utilises the full knowledge of the primary users' transmission scheme such that the primary user can tolerate the interference caused by the secondary user, while the secondary user contends the interference of the primary user (e.g. by interference cancellation). In a *database driven* approach the CRS and the primary system share relevant information – for example, the secondary system can obtain the status of the primary system relevant to its current location in order to develop a suitable transmission technique or access a radio environment map containing real-time profiles of primary and secondary users.

In the learning mode, the CRS utilizes previous information to refine its performance.

Several CRS configurations are envisaged and are being standardized. A heterogeneous CRS network comprises a number of radio access networks using either the same or different accessing schemes; some of the radio access networks (RANs) would be the conventional type operating within a fixed frequency regime whereas others would be reconfigurable operating to reconfigurable terminals. In another configuration the RAN population of reconfigurable capability shares a band servicing their respective reconfigurable terminals.

Potential MSS application

The busiest of MSS bands exhibit unfilled spectral gaps lasting several hours on a typical working day often exceeding 12 hours and extending full day during the weekends. Thus there is considerable scope in utilizing off-peak hours for non-urgent communication on a CR basis either in a blind or cooperative spectrum assessment mode. It is thus feasible to utilize the spare spectrum either in an intra-system cognitive mode or inter-system cognitive arrangement through co-operation between the systems. A central shared database could facilitate in identifying the available channel for ad-hoc dynamic access.

6) Outline the principle and advantages of a software defined radio. State the advantage in using a common development platform such as software communications architecture.

Wireless technologies continue to evolve at a phenomenal rate, making hardware platforms obsolete within a very short span incurring heavy costs in the development of new platforms. Incorporating upgrades to an existing wireless technology incurs a similar development cycle. Software radio (SR) potentially provides a cost-effective, flexible and software upgradeable platform that can be reconfigured in software to suit the changes and upgrades to operating platforms. The technology can perform on all the layers of communication architecture so as to enable a number of standards to be supportable on the same hardware platform, allowing rapid changes from one communication standard or functionality to another. User terminals and infrastructure can be reconfigured in real time to select the appropriate medium at a fraction of the cost of a total refurbishment.

The SDR utilises hardware devices that support reprogrammable firmware and software technologies to implement the desired functionalities. Such devices include field programmable gate arrays (FPGA), digital signal processors (DSP), general purpose processors (GPP), programmable system on chip (SoC) and other application-specific programmable processors.

The SDR technology offers numerous benefits:

- Reconfigurable platform architecture facilitates rapid and cost-effective development of new products and service multiple markets;
- It offers remote reprogramming capability, allowing software upgrades and addition of new features while the unit is in service and remote;
- The SDR platform future-proofs network functionalities, services and products within limits of the platform technology;
- It reduces end-user costs and enhances user experience through ubiquitous communication availability through reconfigurable user terminals.
- The technology facilitates the evolution of cognitive radio architecture.

Figure 1 below shows conceptual diagram of a generic SDR transceiver. The main blocks of the unit are the front end, the baseband processing unit and the data processing. The functionality of the unit is controlled by changing parameters through the control bus.

Figure 1 [A sketch of Figure 14.18] A generic SDR transceiver

The use of a development platform like *software communications architecture* (SCA) ensures that the industrial partners and developers may develop components of the system independently and maintain interoperability and exchangeability. It provides for portability of applications software between different SCA implementations and thus reduces software development time through the ability to reuse design modules while benefitting from evolving commercial frameworks and architectures.

7) Comment on the assumptions and arguments presented in relation to the speculative vision of the future MSS network as presented in the chapter.

This question offers an opportunity to present one’s personal vision while commenting on the assumptions and arguments presented. The following aspects remain open for review and discussions:

- Proposed system architecture - latency issues and solution, ownership and management of sub-systems, standardization of interfaces, converged architecture approach, hybrid architecture including ATC, applicability and role of ad hoc networks;
- Spectrum – frequency bands: Q, V, W bands, regulatory matters and hurdles, investment, radio link reliability improvement solutions;
- Enabling technologies – on-board architecture, micro and pico spot beam, ultra-high EIRP satellites, advanced modulation and coding methods, fade countermeasures, cognitive and software defined radio
- System performance - Improvements