

Lehrstuhl für Kommunikationsnetze  
Technische Universität München

# **Routing and Dimensioning in Satellite Networks with Dynamic Topology**

**Markus Werner**

Vollständiger Abdruck der von der Fakultät für  
Elektrotechnik und Informationstechnik der Technischen Universität München  
zur Erlangung des akademischen Grades eines

Doktor-Ingenieurs

genehmigten Dissertation.

Vorsitzender: Univ.-Prof. Dr.-Ing. J. Hagenauer

Prüfer der Dissertation: 1. Univ.-Prof. Dr.-Ing. J. Eberspächer  
2. Prof. des Univ. G. Maral, Ecole Nationale Supérieure des  
Télécommunications – Site de Toulouse / Frankreich

Die Dissertation wurde am 31.01.2002 bei der Technischen Universität München  
eingereicht und durch die Fakultät für Elektrotechnik und Informationstechnik  
am 13.06.2002 angenommen.



# Acknowledgements

This thesis originates from my work as research scientist at the Institute of Communications and Navigation of the German Aerospace Center (DLR). During this time, many people have contributed in various ways to the successful outcome of this work.

First of all, I want to thank my supervisor Prof. Dr.-Ing. Jörg Eberspächer for his constant advice, support and many helpful and stimulating discussions during the various phases of the dissertation. I am especially grateful for his confidence in my work, giving me considerable freedom for creative research.

I would also like to thank Prof. Gérard Maral who readily accepted to act as the second auditor of the thesis. In fact, he has been more than an examiner of the final thesis only; for many years before, I had the pleasure to cooperate with him in various research projects, in most fruitful co-supervision of students in a bilateral student exchange program, and for a number of joint publications. Over all these years, I have personally drawn invaluable stimulation and feedback, experience and knowledge gain from this cooperation.

A key factor for such a thesis to be successfully completed is the daily working environment. I sincerely thank all my colleagues and students of the Department Digital Networks (DN) at DLR, who have jointly contributed to a highly positive spirit over all those years. Without the friendly, creative and stimulating atmosphere in the DN research “family” this work would have been virtually impossible. Representing all these colleagues, I would like to mention the head of this group, Dr. Erich Lutz, whom I also thank for his continuous interest, stimulation and support in my particular area of research.

For some focussed parts of the research work towards this thesis, I have also had the opportunity and pleasure to cooperate with several colleagues from Prof. Eberspächer’s research team at the Institute of Communication Networks (LKN) of the Munich University of Technology. I would like to express my thanks to them for all the related stimulating and helpful discussions, and to the team as a whole for my integration in the LKN “family.”

I would also like to thank Dr. Axel Jahn and Dr. Erich Lutz for devoting valuable time and effort to proof-reading the manuscript, which contributed to an improved readability of the text in terms of both language and content.

Finally, I thank my wife Brigitte and my children Franziska, Johannes and Magdalena, for their love and support, for their patience and tolerance missing their husband and father in many hours of physical and mental absence — and for the continuing companionship in the real *promovere* in life.

*Oberpfaffenhofen, July 2002*

*Markus Werner*

*To Brigitte, Franziska, Johannes, and Magdalena*

# Contents

<b>1</b>	<b>Introduction</b>	<b>1</b>
1.1	Motivation and Background . . . . .	1
1.2	Scope and Contributions of the Thesis . . . . .	2
<b>2</b>	<b>Satellite Constellation Networks</b>	<b>7</b>
2.1	System Architecture . . . . .	8
2.1.1	Space Segment . . . . .	9
2.1.2	Ground Segment . . . . .	10
2.1.3	Air Interface . . . . .	10
2.2	Satellite Orbits . . . . .	11
2.2.1	Basic Orbit Parameters . . . . .	11
2.2.2	Useful Circular Orbits . . . . .	13
2.2.3	Satellite Ground Tracks . . . . .	14
2.3	Satellite–Earth Geometry . . . . .	15
2.3.1	Basic Geometric Relations . . . . .	15
2.3.2	Coverage Area . . . . .	16
2.4	Satellite Constellations and Intersatellite Link (ISL) Networks . . . . .	17
2.4.1	Basic Design Considerations . . . . .	17
2.4.2	Constellation Types . . . . .	18
2.4.2.1	Star Constellations . . . . .	19
2.4.2.2	Delta Constellations . . . . .	21
2.4.3	ISL Network Topologies . . . . .	21
2.5	Multiple Coverage, Handover and Satellite Diversity . . . . .	23
2.5.1	Multiple Satellite Coverage . . . . .	23
2.5.2	Handover . . . . .	25
2.5.3	Satellite Diversity . . . . .	27
2.6	Network Traffic Dynamics . . . . .	28
2.6.1	User Link Capacity and Traffic Dynamics . . . . .	29
2.6.2	Long-Term Variation and System Period . . . . .	31

<b>3</b>	<b>ATM-Based Satellite Networking Concept</b>	<b>34</b>
3.1	Motivation and Scenario . . . . .	34
3.2	Basic ATM Principles . . . . .	35
3.2.1	Transmission and Multiplexing Scheme . . . . .	36
3.2.2	Virtual Connections . . . . .	36
3.3	ATM in Satellite Systems . . . . .	38
3.3.1	Application and Service Scenario . . . . .	38
3.3.2	Protocol Architecture and Challenges . . . . .	39
3.4	Overall Networking Concept . . . . .	41
3.4.1	Network Segmentation . . . . .	41
3.4.2	Problem Outline . . . . .	41
3.4.3	Continuous End-to-End Networking . . . . .	42
3.4.4	UDL Segment: Air Interface Access Network . . . . .	44
3.4.5	ISL Segment: Meshed Space Backbone Network . . . . .	45
3.5	Feasibility and Implementation Aspects . . . . .	46
<b>4</b>	<b>Routing</b>	<b>49</b>
4.1	Routing in LEO/MEO Satellite Networks . . . . .	49
4.1.1	ISL Topology Dynamics . . . . .	50
4.1.2	Network Segmentation . . . . .	51
4.2	UDL Routing . . . . .	52
4.2.1	Satellite Handover . . . . .	53
4.2.1.1	Handover Procedure . . . . .	53
4.2.1.2	Handover Strategies . . . . .	54
4.2.2	Satellite Diversity . . . . .	59
4.2.2.1	Scenario and User Environment . . . . .	59
4.2.2.2	Evaluation of Satellite Diversity Gain . . . . .	61
4.3	Off-Line Dynamic ISL Routing Concept . . . . .	64
4.3.1	Discrete-Time Network Model . . . . .	64
4.3.2	Dynamic Virtual Topology Routing (DVTR) . . . . .	66
4.4	On-Line Adaptive ISL Routing . . . . .	73
4.4.1	Distributed Approach Based on Deterministic Algorithms . . . . .	75
4.4.1.1	Dynamic and Adaptive Routing Concepts . . . . .	75
4.4.1.2	Decentralized Traffic Adaptive Routing . . . . .	75
4.4.2	Isolated Approach Based on Neural Networks . . . . .	79
4.4.2.1	Motivation and Background . . . . .	79
4.4.2.2	Basics of Multilayer Perceptrons . . . . .	81
4.4.2.3	Traffic Adaptive Routing with Distributed Multilayer Perceptrons . . . . .	84

<b>5</b>	<b>ISL Network Design</b>	<b>87</b>
5.1	Introduction and Overview . . . . .	87
5.2	Topological Design . . . . .	88
5.2.1	Motivation and Satellite Constellation Options . . . . .	88
5.2.2	ISL Topology Design Procedure for Delta Constellations . . . . .	88
5.3	ISL Routing Framework . . . . .	92
5.3.1	The Combined ISL Routing/Dimensioning Problem . . . . .	92
5.3.2	Path Grouping Concept . . . . .	96
5.4	Capacity Dimensioning . . . . .	97
5.4.1	Overall Approach and Assumptions . . . . .	97
5.4.2	Isolated Single-Step Dimensioning . . . . .	99
5.4.2.1	Heuristic Approach . . . . .	99
5.4.2.2	Optimization Approach . . . . .	102
5.4.3	History-Based Multi-Step Dimensioning . . . . .	105
5.4.3.1	Heuristic Approach . . . . .	107
5.4.3.2	Optimization Approach . . . . .	107
<b>6</b>	<b>Numerical Studies</b>	<b>111</b>
6.1	Routing . . . . .	112
6.1.1	UDL Routing . . . . .	112
6.1.1.1	Scenario and Overall Simulation Approach . . . . .	113
6.1.1.2	Channel Modeling . . . . .	113
6.1.1.3	Channel Adaptive Satellite Diversity (CASD) . . . . .	116
6.1.1.4	Satellite Diversity and Handover Performance . . . . .	118
6.1.2	ISL Routing . . . . .	123
6.1.2.1	Simulation Approach and Traffic Modeling . . . . .	123
6.1.2.2	Impact of Long-Term Traffic Variation . . . . .	125
6.1.2.3	Impact of Traffic Weight in the Link Cost Metric . . . . .	126
6.1.2.4	Impact of the Type of Link Cost Metric . . . . .	131
6.2	ISL Network Dimensioning . . . . .	132
6.2.1	Scenario and Assumptions . . . . .	133

6.2.2	Isolated Single-Step Dimensioning . . . . .	134
6.2.2.1	Worst-Case Link Traffic Load . . . . .	134
6.2.2.2	Physical Link Traffic Load . . . . .	139
6.2.2.3	Path Costs . . . . .	139
6.2.3	History-Based Multi-Step Dimensioning . . . . .	139
6.2.3.1	History Modeling . . . . .	142
6.2.3.2	Worst-Case Link Traffic Load . . . . .	146
6.2.3.3	Physical Link Traffic Load . . . . .	150
6.2.3.4	Path Traffic Load . . . . .	150
<b>7</b>	<b>Conclusions</b>	<b>154</b>
<b>A</b>	<b>Reference Satellite Constellations and ISL Topologies</b>	<b>159</b>
<b>B</b>	<b>Space Geometry and Orbital Mechanics</b>	<b>166</b>
B.1	Satellite Orbits . . . . .	166
B.1.1	Elliptical and Circular Orbits . . . . .	166
B.1.2	Satellite Velocity and Orbit Period . . . . .	168
B.2	Coordinate Systems and Satellite Ground Tracks . . . . .	169
<b>C</b>	<b>Global Traffic Models</b>	<b>172</b>
C.1	Traffic Modeling Framework TMF 1 . . . . .	172
C.2	Traffic Modeling Framework TMF 2 . . . . .	177
<b>D</b>	<b>Illustrations for MLP Simulations</b>	<b>179</b>
	<b>Notation and Symbols</b>	<b>185</b>
	Constants . . . . .	185
	Mathematical Notation . . . . .	185
	Symbols . . . . .	186
	<b>Abbreviations</b>	<b>191</b>
	<b>Bibliography</b>	<b>195</b>

# Chapter 1

## Introduction

*Nothing will come of nothing.*

— WILLIAM SHAKESPEARE, *King Lear* (Act I, Scene 1)

### 1.1 Motivation and Background

The idea of satellite constellations being used to provide communication services to large portions of the earth can be traced back to Arthur C. Clarke’s famous paper “Extra-Terrestrial Relays” in *Wireless World* in 1945 [Cla45]. In this paper, he proposed a constellation of three geostationary satellites, nearly equidistantly spaced in the geostationary earth orbit (GEO), to provide full equatorial coverage of the earth. Among other potential applications, he advertised the use of such a constellation for television broadcasting committed to education and distance learning – obviously issues of timeless significance.

Small constellations of satellites in geostationary earth orbit (GEO) or highly elliptical orbits (HEO) have been proposed, implemented and operated since the early days of satellite communications. Up until very recently, however, networking in these constellations has not been a major issue and not really tackled the satellite segment as such; rather the networking functionality has remained on ground, with the satellite mainly serving as space-based retransmitter, either in “bent-pipe” fashion as frequency shifter and amplifier, or including baseband digital signal processing for signal regeneration. Uplink traffic, however, is always directly returned to the ground in these systems, may it be for unicast, multicast, or broadcast service.

Satellite communications then witnessed a real paradigm shift in the late 1980s and early 1990s. In these years, plans were revealed or seemed to be realistic for the first time in satellite history, that scheduled to launch and operate satellite constellations of several tens of satellites in low earth orbits (LEO) or medium earth orbits (MEO). The ultimate challenge and complexity was especially incorporated in proposals such as Iridium [Leo91, Gru91, HL95, PRFT99], employing intersatellite links (ISLs) to form a backbone network in space by meshing the satellites in a dynamic topology. The tutorial paper by Maral et al. [MDER91] provides an excellent overview of the state of the art around 1990, and has certainly contributed to advertising the idea of constellation networks especially in Europe, too, for the scene had been completely dominated by US industry and research until then.

In the first half of the 1990s, emerging system proposals like Iridium, Globalstar [Sch95], and ICO [Pos96] were aiming at providing mainly (mobile) telephony and other low-bit-rate services to single users. Therefore, these systems are also called narrowband or simply voice systems, or they are assembled under the label S-PCN (satellite personal communication network). According to their modest bit rate requirements, they operate the user or service links at L or S bands (around 2 GHz), as illustrated in Fig. 1.1.

The latter half of the 1990s witnessed a growing interest in broadband constellations targeting multimedia applications. Teledesic [Stu95], Celestri [Mot97] and M-Star [Mot96] are well-known representatives of the group of proposed ISL-based broadband LEO systems, SkyBridge [Sky97] stands for a system providing pure access network capabilities without ISLs. Because of the crowded situation in the low-frequency bands around 2 GHz and the large bandwidth required for broadband applications, these systems were assigned higher frequencies bands ranging from Ku to V band, cf. Fig. 1.1.

Besides the trend towards general broadband multimedia systems, the Internet boom in the late 1990s has also raised particular interest in satellite systems being more dedicated to Internet (broadband) data applications.

The most important features of satellite systems with regard to these different application areas are illustratively summarized in Fig. 1.2, including the traditional application area of satellite broadcasting. Obviously, the latter has a pronounced affinity with GEO satellite systems, whereas in the three formerly identified areas there are some compelling reasons for considering LEO or MEO systems as potential solution, including (i) considerable benefits in terms of latency and link budgets due to shorter distance and lower free space loss, (ii) true global coverage, and (iii) the potential for better reuse of limited available ground-space communication frequencies.

Coinciding with these developments of satellite communications in the last decade, terrestrial communication networks have seen the development and implementation of the *asynchronous transfer mode (ATM)* as the transmission and switching scheme for the broadband ISDN (B-ISDN), efficiently providing multimedia and multiservice communications in integrated networks. While basically operating in connection-oriented manner, ATM combines the strengths of both circuit switching and packet switching techniques, and is therefore capable of transporting all kinds of information and supporting each type of service; in particular, different levels of quality of service (QoS) guarantees are provided. Based on the connection-orientation paradigm and on a well-specified framework of services and traffic classes (including their QoS parameters) supported, reliable and powerful traffic engineering methods can be applied in network design.

## 1.2 Scope and Contributions of the Thesis

Satellite constellations have been proposed and deployed to serve as communication systems over the last decades, and *networking* of such constellations has become an increasingly important issue for their operation and their success.

This thesis focuses on routing and capacity dimensioning for LEO and MEO satellite systems employing ISLs, i.e., such system where the satellite constellation *is* the network; they are synonymously referred to as *LEO/MEO (satellite) networks* or *constellation networks*.

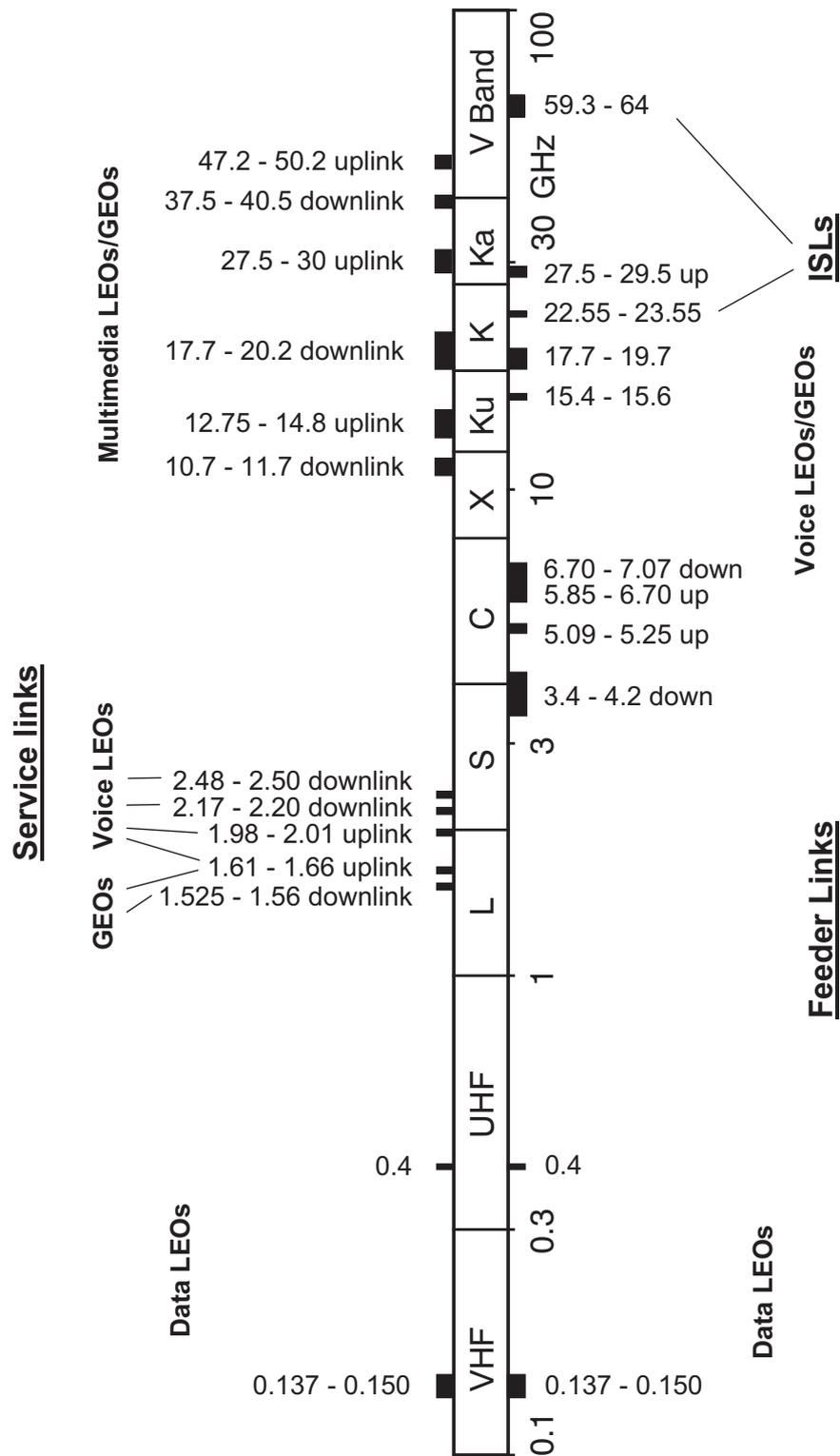
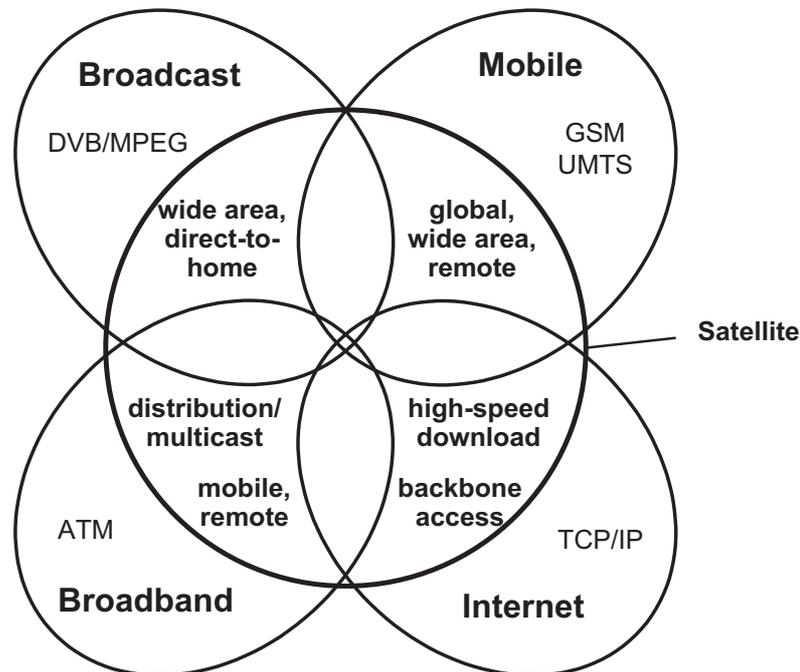


Figure 1.1: Frequency bands for mobile and broadband satellite communications.



**Figure 1.2:** Application areas of satellite communications.

A satellite constellation consists of a number of similar (or identical) satellites in complementary orbits, sharing tasks devoted to a common goal. The particular goal we are looking at in this thesis is the provision of communication services to large areas of the earth, up to global coverage. With respect to the application areas identified above, for the purpose of this thesis we do not explicitly distinguish between narrowband or broadband systems or applications, rather we concentrate on generic networking approaches and concepts that are valid in any of the possible application areas. However, we fundamentally restrict our considerations to unicast or point-to-point communications; multicast and broadcast services are not addressed.

The core contributions of this thesis are

1. an overall ATM-based networking concept,
2. an off-line dynamic ISL routing concept with on-line adaptive routing extensions, and
3. an integrated network design approach comprising (i) topology design, (ii) routing framework and (iii) capacity dimensioning / routing optimization,

all essentially based on the connection-oriented communication paradigm, for LEO/MEO satellite constellation networks with dynamic topology. Major parts of the concepts are claimed and believed to be generic so as to qualify them for a long “life-cycle” in a rapidly changing communications and networking world in general, and in an economically critical satellite environment in particular.

The research work underlying this thesis has roughly been performed in the years 1995–1997 for the first two items, and in 1998–1999 for the third one.

Up until 1995, virtually no fundamental research on routing and dimensioning for the proposed commercial ISL-based satellite networks had been published. Earlier work [Bra84, BHGV87, Cha89] had mainly concentrated on delay-tolerant data communication, distributed architectures, packet routing approaches and survivable communication concepts for ISL networks, driven more or less by a military background.

Obviously, for commercial applications with delay-sensitive voice or multimedia communications as key components, the objectives and challenges are quite different compared to (military) data communications. Thus, the announcement and ongoing discussion of the commercial ISL-based Iridium system foreseen for global voice communications was the main stimulation for our research into networking issues for ISL-based systems.

As a first step, an overall networking concept was developed, providing reliable and continuous end-to-end communication in the encountered multi-dynamic network topologies. Given the specific challenges and objectives of the considered scenario, we proposed a connection-oriented and ATM-based approach [WDV<sup>+</sup>97], mainly capitalizing on the virtual connection capabilities of ATM that allow to “hide” from the end-to-end routing schemes the relative mobility of both, (i) satellites with respect to served users on ground, and (ii) satellites with respect to their neighbors in the constellation. The proposed overall networking concept follows a strict decomposition approach, separating the sub-problems of routing in the ISL segment and routing on the up/downlinks, the UDL segment. Focusing further on the ISL segment, this work has been extended by developing a general (i.e., not bound to an ATM implementation) off-line dynamic routing concept for time-variant ISL topologies, *dynamic virtual topology routing (DVTR)* [Wer97]. DVTR essentially exploits the periodic and predictable nature of the topology by discretization into series of quasi-static topology snapshots, and providing a reliable framework for any on-line traffic-adaptive routing scheme to operate within.

It has already been recognized in literature that both mentioned concepts have been among the first to address the dynamic ISL networking and routing issues for commercial systems in a generic manner (i.e. being not subject to certain routing algorithms, signaling schemes, etc.); until today, they are referenced in several papers on ISL routing in LEO constellations.

The use of the Iridium ISL topology for related conceptual and numerical investigations has also led to identifying particular impairments for efficient routing in ISL topologies where (i) certain links have to be switched on and off, and (ii) links can not be established between satellites in certain orbits at all (as with Iridium). This reduces the overall regularity of the topology and leads inevitably to forced path switching situations, with related operational complexity and degradation of the grade of service for connection-oriented communication. Consequently, this has led us to investigate potential alternative constellations that allow to set up permanent ISL topologies [WJLB95, Wer95]. In fact, most of the broadband LEO constellations with ISLs that were proposed in the following years, 1996/97, turned out to provide such permanent ISL topologies.

Based on constellations of this type, we developed an integrated network design approach comprising (i) a systematic method for the design of permanent ISL topologies, (ii) a routing framework based on a modification of the original DVTR concept, and (iii) a capacity dimensioning procedure with routing optimization [WFWM01]. As to our knowledge, this has been the first published treatment of the combined routing/dimensioning problem for LEO/MEO ISL networks.

Around these core contributions, the remainder of the thesis is organized as follows.

Chapter 2 first presents an overview of the system architecture of satellite constellation networks, and then proceeds to cover relevant basics of satellite orbits and satellite–earth geometry. Following that, the two basic types of circular-orbit satellite constellations, *star* and *delta* constellations, are explained in detail, and implications of both approaches for possible ISL network designs are derived. The chapter further introduces handover procedures and the satellite diversity concept, both constituting the up/downlink (UDL) routing segment. The chapter concludes with the discussion of some aspects of traffic dynamics, which need to be taken into account especially for numerical performance evaluation and simulation.

The detailed description of the ATM-based overall networking concept, as mentioned before, forms the core of Chapter 3. Before, the relevant basic principles of ATM and some facets of ATM in the satellite environment are presented. The discussion of the networking concept is concluded with some feasibility and implementation considerations.

Chapter 4 covers the whole range of routing concepts, schemes and algorithms. After outlining the dynamic environment and the network segmentation into UDL and ISL segments, first the UDL segment is considered in detail. Satellite handover and satellite diversity approaches are presented and evaluated from a routing perspective; in particular, implications for ISL routing are derived. Following that, the above-mentioned off-line dynamic ISL routing concept, DVTR, is treated in depth, concluding with some performance investigations. Finally, some candidate options for on-line traffic adaptive routing in ISL networks are presented and numerically evaluated, exploring the potential of worst-case load reduction and smoothing of network traffic. Two basic approaches are considered, namely a distributed scheme based on conventional algorithms, and an isolated approach using neural networks of multilayer perceptron type.

In Chapter 5 we present the three components of the integrated network design approach – systematic ISL topology design, adapted routing framework, and capacity dimensioning – where the focus is on the mathematical formulation of the dimensioning problem and of some heuristic and optimization approaches to its solution.

Chapter 6 assembles numerical in-depth studies pertaining to the three areas UDL routing, ISL routing, and optimized capacity dimensioning. The specific models, scenarios, and simulation approaches used to derive the results are presented in detail along with the respective results. Some emphasis is laid on a step-by-step evaluation of the various capacity dimensioning approaches, from simple to sophisticated methods, in order to properly illustrate the theory and mathematics from Chapter 5.

Finally, in Chapter 7 we conclude the main part of the thesis by summarizing our work and discussing directions for future work.

Appendix A contains an overview of all systems, their constellation parameters, and ISL topologies which are addressed in this thesis. Appendix B provides supporting material on space geometry and orbital mechanics. In Appendix C, we present two global traffic modeling frameworks, which have been used for numerical investigations pertaining to various topics in this thesis. Appendix D provides some illustrations on the simulation of traffic adaptive routing with multilayer perceptrons.

Finally, a description of mathematical notation and a list of symbols, a list of abbreviations, and the bibliography assembling all literature referenced in this thesis, is given.

## Chapter 2

# Satellite Constellation Networks

*However great the initial expense, it would only be a fraction of that required for the world networks replaced, and the running costs would be incomparably less.*

— ARTHUR C. CLARKE, in *Extra-Terrestrial Relays* (1945)

*There is just one thing I can promise you about the outer-space program: your dollar will go further.*

— WERNHER VON BRAUN (1912 – 1977)

In terms of constellation size and networking complexity, satellite communications has experienced a permanent growth since the mid 1960s when commercial applications started to be offered via individual geostationary satellites. As an extension of one individual geostationary satellite, already a triplet of them can be considered as a constellation; if they are approximately equidistantly spaced around the equator, they can already provide full equatorial (i.e., excluding the polar caps) coverage, and this was in fact the original constellation proposal by Arthur C. Clarke in his visionary article [Cla45].

Small constellations of satellites in geostationary earth orbit (GEO) or highly elliptical orbits (HEO) have been proposed, implemented and operated since then. In this thesis, we will not further consider any HEO constellations, nor will we elaborate more on GEO satellites; the latter are only used as illustrative geometrical or geographical reference when presenting and discussing those constellations that have lifted networking to space, as follows.

This thesis focuses on LEO and MEO satellite constellations employing ISLs; they are synonymously referred to as *LEO/MEO (satellite) networks* or *constellation networks*. (For some particular topics not related to ISLs, however, we will also consider two example systems that don't use ISLs, namely the Globalstar LEO constellation and the MEO system ICO.)

In this chapter we present the main components, underlying concepts, characteristics and particular networking challenges of such constellation networks, focusing on such material that is relevant for the research areas covered in the following core chapters of the thesis. For a more comprehensive treatment of various aspects of satellite communications and constellations, we refer especially to the following recently published or updated textbooks:

- The standard work by Maral and Bousquet [MB98] and its predecessor editions has been read and consulted as *the* reference book on satellite communications for more than a decade now.
- Co-authored by the author of this thesis, [LWJ00] can serve as a recent resource providing a broad coverage of satellite systems for personal (narrowband) and multimedia (broadband) communications, including some core parts of the research presented herein. Some parts of this chapter have been adopted from this reference.
- For an excellent view on the topic of *mobile* satellite communication networks in general and their potential integration in a global UMTS infrastructure, one should refer to the most recently published book by Sheriff and Hu [SH01].

Before scratching the ground floor of all satellite constellations, namely orbital mechanics, to further build upon, we think it is more instructive to start into the topic with a top level view on the architecture of constellation satellite systems.

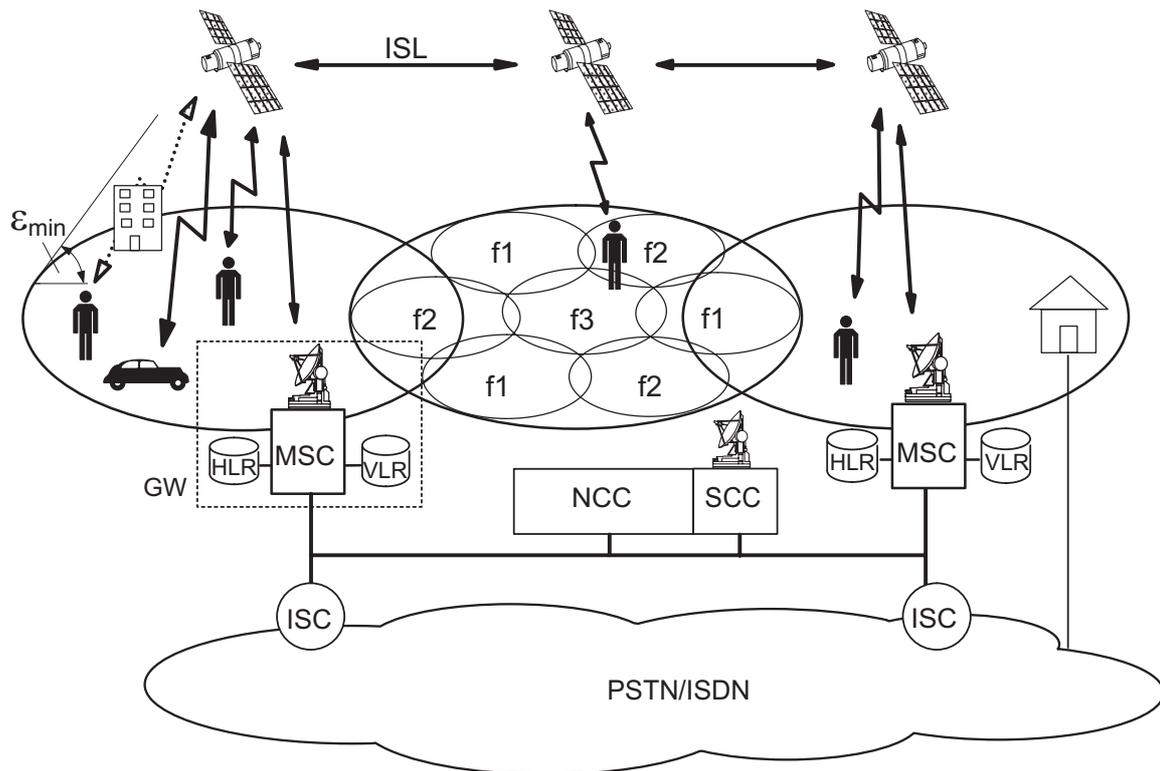
## 2.1 System Architecture

This section presents the main network elements of a satellite constellation system, using the case of (narrowband) mobile/personal communications as an example; in this scenario, such a system is often called a *satellite personal communication network* (S-PCN).

This view indeed is not a restriction of universality and justified mainly with the following arguments:

- Many distributed access points (users) to a global system, user mobility, and power-limited handheld terminals do certainly impose major challenges to a constellation network; a network architecture that can cope with these challenges will certainly do for less demanding tasks as well.
- Logically, in the decade of GSM the S-PCNs have adopted a lot of the benefits of GSM, mainly when it comes to network architecture. Being one of the most sophisticated and complete architectures communication networks have ever seen, it is today definitely the key to all directions that constellation satellite networks may take in the future; this is already happening with the path towards UMTS as the successor of GSM, where satellites again follow the terrestrial evolution.
- Any broadband satellite constellation, typically to serve fixed users at higher rates – for instance a plain backbone to interconnect LAN islands, or a sophisticated “Internet-in-the-sky” – would not bring in substantially new issues from the system and network architecture point of view, rather the additional challenges would be on the side of capacity and frequency bands used, protocols, interworking, and traffic engineering.

Figure 2.1 shows the basic system architecture of an S-PCN. By means of this illustration we will only give a brief overview of the related aspects. A variety of research issues for S-PCNs and their architecture is addressed in a special issue of the *International Journal of Satellite Communications* [Mar94]. Given the importance of GSM for satellite network architectures, we will make use of some GSM terminology without in-depth explanations; for a dedicated “networking view” on GSM concepts one may refer to [EVB01].



**Figure 2.1:** Network architecture of a satellite personal communication network (S-PCN).

### 2.1.1 Space Segment

The space segment consists of an ensemble of satellites, which may be interconnected by inter-satellite links (ISLs) allowing long-distance transmission within the space segment. Each satellite covers a circular area on the earth's surface, which increases with increasing orbit altitude and decreasing minimum elevation angle  $\epsilon_{\min}$ . The choice of orbit planes and the satellite phasing within the orbits must guarantee continuous coverage of the service area (being the full surface of the earth for global systems). These geometrical aspects will be looked at in more detail throughout the next sections.

Direct communication via satellite using a handheld terminal with low transmit power and an omnidirectional antenna requires a high antenna gain on board the satellite. Accordingly, the coverage area of an antenna beam (spot beam) is rather small, and the coverage area of the satellite is composed of several radio cells. The spot beam approach allows the reuse of frequency bands  $f_1, f_2, \dots$  in separated cells, increasing the bandwidth efficiency of the system.

From a networking viewpoint, employing ISLs in the system or not makes a major difference: with ISLs, the system is not longer a pure access network for the first or last hops, but rather provides global backbone facilities at the same time, thus making it an *autonomous system*. Evidently, satellites in an ISL-based system need to carry a much more sophisticated payload than in a non-ISL system, including ISL terminals, true on-board processing (OBP), and switching or routing functionality.

Apart from their communications task, the satellites can also perform decentralized tasks in the area of radio resource management, such as the allocation of radio channels to users.

### 2.1.2 Ground Segment

*Gateway stations (GWs)* are important elements of the ground segment. They comprise fixed earth stations, which provide communication links to the satellites and via the satellites to the mobile users within the GW service area. On the other hand, the GWs are connected via international switching centers (ISCs) to terrestrial fixed networks. Thus, GWs constitute interfaces between the satellite network and terrestrial networks.

With respect to a mobile user we distinguish two types of GWs. The *home gateway* is the GW which serves the home area of the mobile user. The *visited gateway* is a GW which serves a mobile user outside the service area of the home GW.

The GWs are equipped with databases allowing contact to be maintained with globally mobile users by means of GSM-like mobility management functions. The *home location register (HLR)* contains the data of subscribers residing within the service area of the relevant GW and being registered in that GW. If a user is currently located in a different GW service area, the HLR contains the corresponding VLR address (the visited GW). The *visitor location register (VLR)* contains information (e.g. the location area) of “visitors” currently staying in the service area of the visited GW.

The *network control center (NCC)* controls access to the satellite network and is involved in RR. Among other tasks it allocates spot beam frequencies and distributes routing tables to the satellites. Also, it performs network management tasks, such as performance management and fault management. The *satellite control center (SCC)*, via telemetry, tracking, and command (TT&C) links, keeps the satellites in their correct orbital positions.

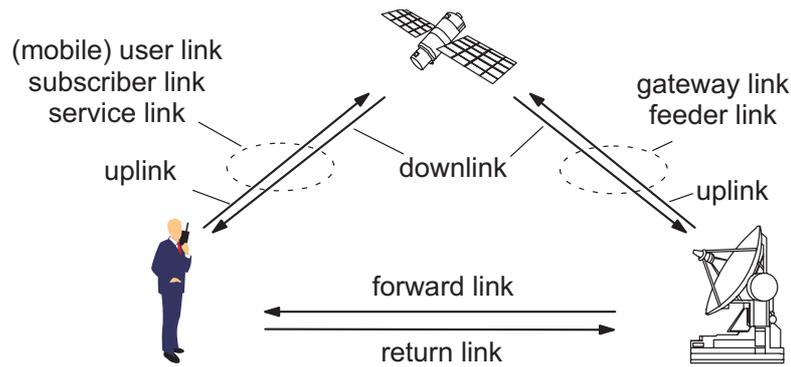
Gateways and control stations are interconnected by a fixed terrestrial network consisting of private or leased lines. Similarly, each satellite always has a connection to the NCC or to other control stations with corresponding functionalities.

### 2.1.3 Air Interface

The connection between space and ground segment is on the whole often called the *air interface*. Neglecting the ISLs for the moment as not belonging to the air interface, the radio links in satellite communications are referred to under specific terms as illustrated in Fig. 2.2.

The most critical link with regard to communication quality, and usually also the capacity bottleneck, is the link between the user and the satellite; common names for this link are *user link* (*mobile user link* in the case of mobile communication), *subscriber link*, or *service link*. The link between the satellite and a fixed earth station is called the *feeder link* (which stems from imagining that the earth station “feeds” the satellite with a broadcast signal, for example) or *gateway link* (in particular if the fixed earth station represents an interface to a terrestrial network).

Both links can be separated into an *uplink* and a *downlink*. Finally, the series of unidirectional links from the fixed earth station to the user is called the *forward link*, and the links in the opposite direction constitute the *return link*.



**Figure 2.2:** Radio link terminology for satellite communication systems.

## 2.2 Satellite Orbits

An in-depth discussion of satellite orbits can be found in the three textbooks mentioned in the introduction of this chapter. In addition, Logsdon [Log97] provides an extensive and lively treatment of orbital mechanics from a sound engineering perspective. We will here restrict ourselves to summarize the major points, often not deducing but just giving a final formula or stating a result. Some background material is also assembled in Appendix B.

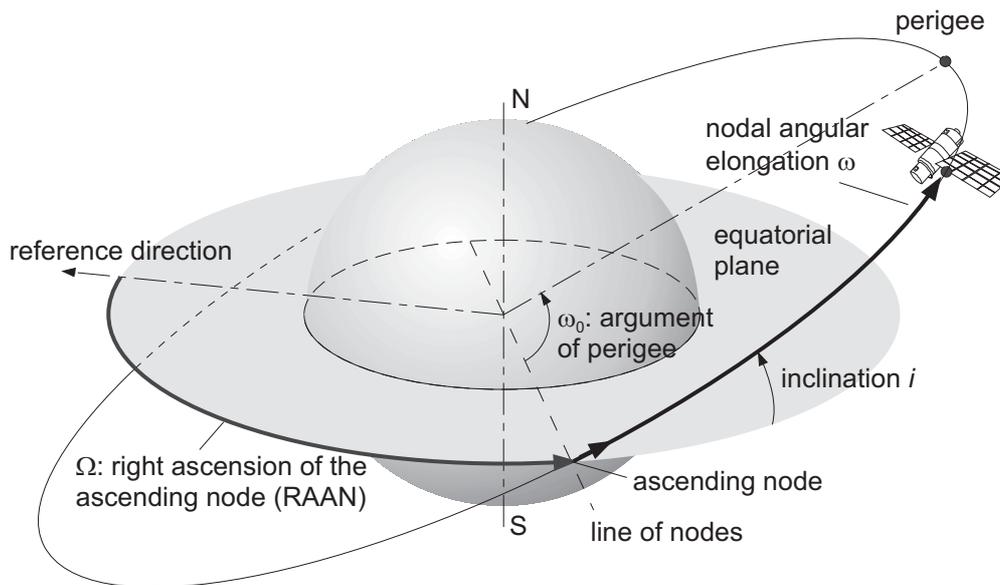
It should be noted that in the following we assume

- a two-body system consisting of earth and satellite, no gravitational influence of other bodies like sun and moon,
- the earth to be a perfect sphere, neglecting its oblateness (flattening at the poles),
- earth and satellite as point masses for gravitation consideration, and
- no other orbit perturbations through, e.g., solar radiation pressure, atmospheric drag, and distortions due to the satellite propulsion.

These simplifying assumptions are usually justified if one considers satellites, and constellations formed by them, primarily as communication elements rather than rigid bodies in space that have to be controlled and stabilized in their instantaneous location and attitude with respect to earth.

### 2.2.1 Basic Orbit Parameters

In the idealized scenario as described by the above-mentioned assumptions, a satellite orbit will remain constant in space for all times; that is, its orientation is sidereal fixed (i.e. fixed with respect to the stars) and independent of the earth's rotation. For the general case of elliptical orbits the position of a satellite in a point in space at any given time is thus completely determined by a set of six orbital parameters: three of them define the orientation of the orbital plane in space, two additional ones define the geometrical shape (parameters of the ellipse, or specifying a circle as a special case) and the last parameter describes the position of the satellite within the plane.



**Figure 2.3:** Orientation of the orbit plane in space. The reference direction points from the earth to the sun at vernal equinox.

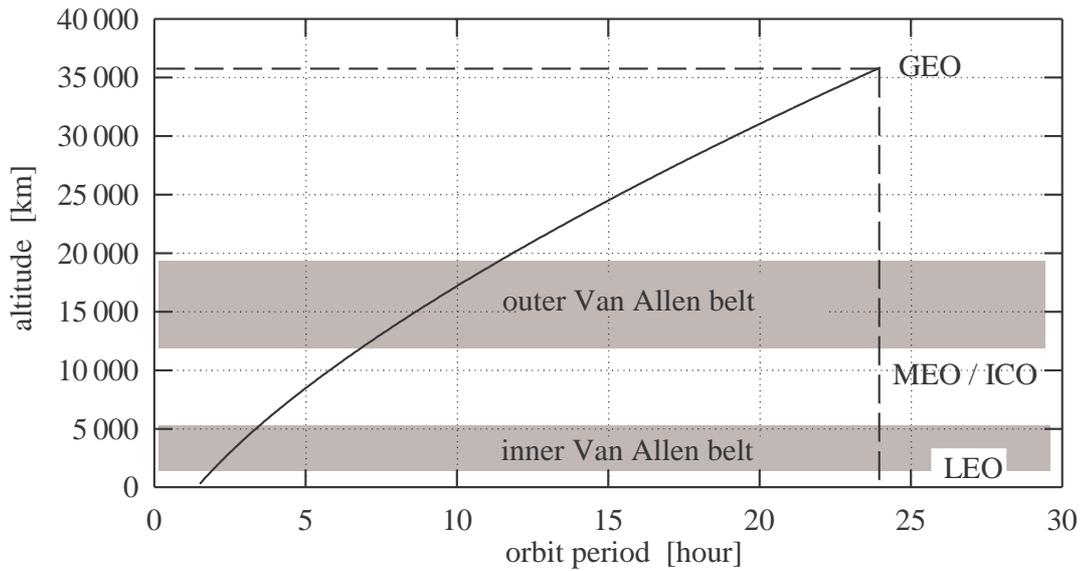
**Orientation of the Orbit Plane** Figure 2.3 shows the parameters that characterize the orbit orientation:

- The *right ascension of the ascending node (RAAN)*  $\Omega$  determines the angle between a reference direction and the line of nodes. The reference direction is given by the direction from the earth's center to the sun at vernal equinox. Equivalently, this direction corresponds to the intersection between the equatorial plane and the plane of the ecliptic. The reference direction remains fixed in space.<sup>1</sup>
- The *inclination*  $i$  defines the angle between the orbit plane and the equatorial plane. It is counted positively with respect to the ascending satellite orbit track. The line of intersection between the two planes is called the *line of nodes*. The *ascending node* is passed when the satellite enters the northern hemisphere.
- The *argument of perigee*  $\omega_0$  is the angle between the line of nodes and the semi-major axis of the ellipse. This parameter is not relevant in the special case of circular orbits.

**Geometrical Shape of the Orbit Plane** The shape of the orbit plane is determined by the *semi-major axis*  $a$  of the ellipse and the *eccentricity*  $e$ ; a detailed discussion can be found in Appendix B.1. Restricting to circular orbits now we can just shortly state here that  $e = 1$ , and  $a$  becomes the radius  $r$  of the circular orbit plane.

**Position of the Satellite in the Orbit Plane** Finally, one more parameter is needed to determine the satellite's position. This parameter is the *true anomaly*  $\theta$  which is discussed in detail in Appendix B.1. Serving as general parameter for elliptical orbits, however, it relates the satellite's

<sup>1</sup>The intersection of the planes varies somewhat due to the perturbations of the terrestrial rotation. For details see [MB98].



**Figure 2.4:** Orbit classification according to altitude.

position to the argument of perigee, which is not relevant for circular orbits, as we have just said. Therefore, the *nodal angular elongation*  $\omega$  is used instead for circular orbits; as shown in Fig. 2.3,  $\omega$  relates the instantaneous satellite position within the plane to the ascending node of the latter.

**Orbit Period** Essentially based on Newton’s law of gravity, the important *orbit period*  $T$  can be derived (see Appendix B.1) as a function of the radius  $r$  for circular orbits,

$$T = 2\pi \sqrt{\frac{r^3}{GM_E}}, \quad (2.1)$$

where  $G$  is the universal gravitational constant and  $M_E$  is the mass of the earth.

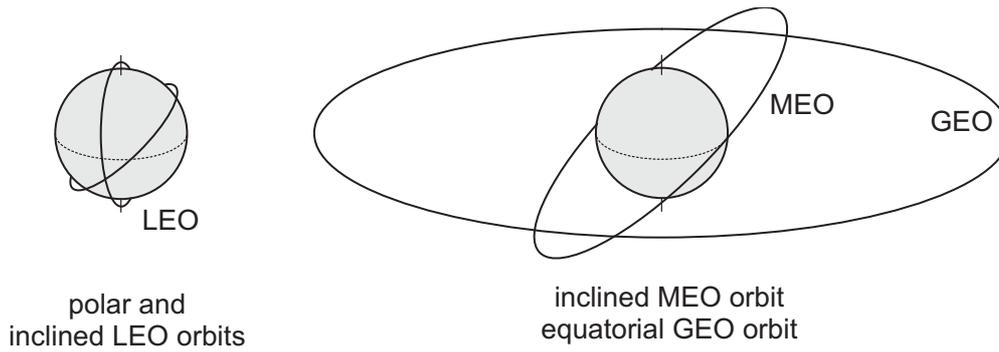
## 2.2.2 Useful Circular Orbits

For a number of reasons, including required number of satellites for earth coverage, free space loss, and propagation delay, the altitude  $h$  of the satellites over ground is a crucial parameter for any constellation. Obviously, with  $h = r - R_E$  and using Eq. (2.1), the orbit altitude is related to the orbit period  $T$  by

$$h = \sqrt[3]{GM_E \left(\frac{T}{2\pi}\right)^2} - R_E. \quad (2.2)$$

Particularly attractive orbit periods are those where  $T$  is an integer divisor of a sidereal day  $T_E$  (the time of one earth rotation, 23 h 56 min 4.1 s) since in this case the satellite positions reiterate periodically day by day: roughly speaking<sup>2</sup>, when  $T = 1, 2, 4, 6, 12, 24$  h. However, some periods may not be used since the associated altitude falls into the so-called *Van Allen belts*, which are

<sup>2</sup>The small difference between  $T_E$  and the mean solar day  $T_{Sol} = 24$ h is neglected here; see Section 2.6.2 for a discussion on this issue.



**Figure 2.5:** LEO, MEO, and GEO orbit types.

regions of the ionosphere with high ion concentration, thus reducing the satellite's lifetime. Figure 2.4 shows the relation between satellite period and altitude using Eq. (2.2). Three regions of operation can be identified:

- *low earth orbits* (LEO), with altitudes from 500–1 500 km above the earth and periods of approximately 2 h,
- *medium earth orbits* (MEO), with altitudes from 5 000–11 000 km above the earth and periods of approximately 4–6 h, and the
- *geostationary earth orbit* (GEO) at 35 786 km altitude and a period  $T = T_E$ .

The GEO orbit is a special *geosynchronous* orbit, namely the one in the equatorial plane ( $i = 0$ ), whereas LEO and MEO orbits are also called *non-geosynchronous* orbits. Finally, Fig. 2.5 illustrates typical orbit types with various inclination angles at a glance. Whereas  $i \approx 90^\circ$  for polar (or near-polar) orbits, the inclination angle of inclined orbits is usually between  $40^\circ$  and  $80^\circ$ .

### 2.2.3 Satellite Ground Tracks

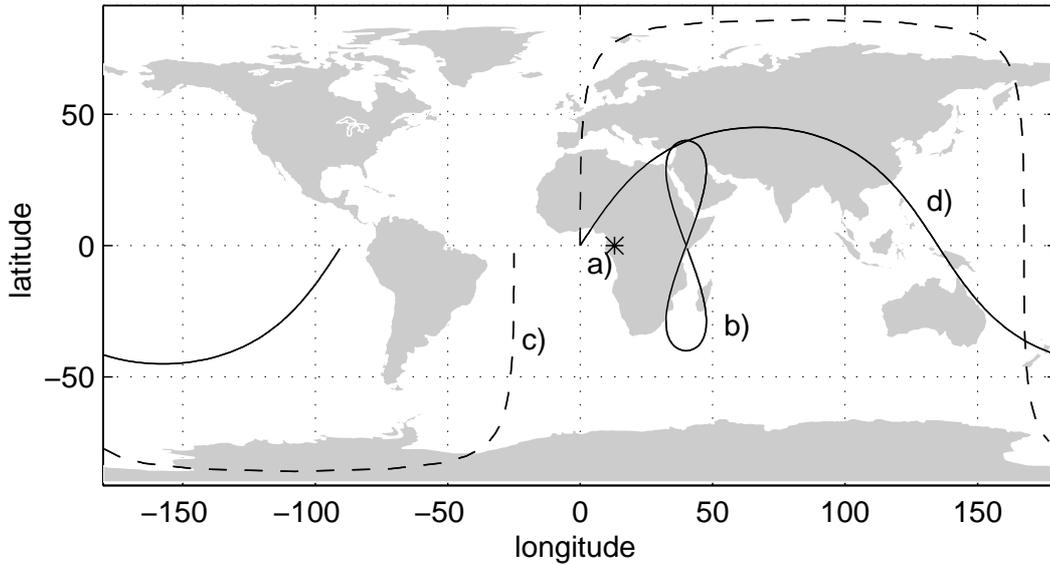
So far, the satellite position has been determined within the orbit in space, but not with respect to the rotating earth. The satellite *ground track* in terms of longitude and latitude pairs over time, can be expressed as a function of the elongation  $\omega(t)$  denoting the angle from the ascending node to the satellite (cf. Fig. 2.3), according to

$$\sin \varphi_s(t) = \sin i \sin \omega(t) \quad (2.3)$$

$$\lambda_s(t) = \lambda_0 + \arccos \frac{\cos \omega(t)}{\cos \varphi_s(t)} - \frac{2\pi}{T_E}(t - t_0), \quad (2.4)$$

where  $\lambda_0$  is the longitude of the ascending node and  $t_0$  is the time when the satellite passes the ascending node. The derivation is given in Appendix B.2.

In general, satellite tracks are not closed after a revolution due to the earth's rotation. As stated before, closed tracks can be achieved for orbit periods  $T = \frac{m}{n} \cdot T_E$ , with integers  $m, n$ . If  $m$  and  $n$  are not divisors of each other the tracks will close after  $n$  revolutions. Figure 2.6 shows typical examples of satellite ground tracks. The point of the track with highest latitude is called the vertex and corresponds to the inclination  $i$ .



**Figure 2.6:** Satellite ground tracks of different orbit types:

- a) geostationary,  $\lambda_0 = 13^\circ$ ,  $T = T_E$ , inclination  $i = 0^\circ$
- b) geosynchronous,  $\lambda_0 = 40^\circ$ ,  $T = T_E$ , inclination  $i = 40^\circ$
- c) polar LEO,  $\lambda_0 = 0^\circ$ ,  $T = 6\,000$  s, inclination  $i = 86^\circ$
- d) inclined MEO,  $\lambda_0 = 0^\circ$ ,  $T = T_E/4$ , inclination  $i = 45^\circ$

## 2.3 Satellite–Earth Geometry

### 2.3.1 Basic Geometric Relations

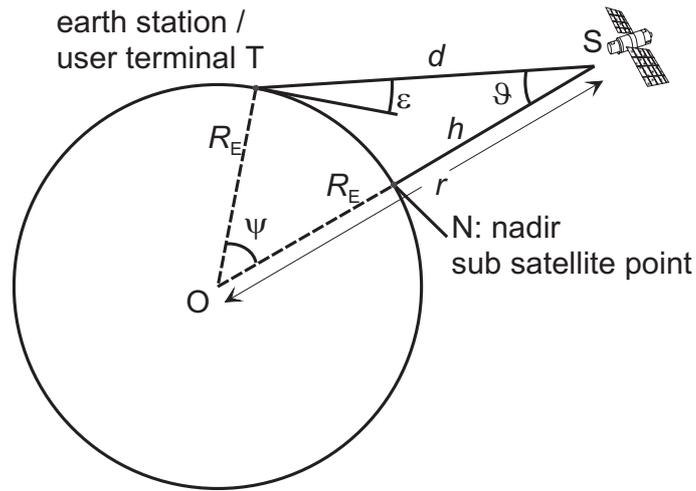
In order to appropriately consider the crucial satellite and constellation coverage of earth regions, it is first useful to look at some basic geometric relations between a satellite and a user terminal on earth, supported by Fig. 2.7. The projection of the satellite onto the earth is called the sub-satellite point (SSP) or *nadir*. Four important geometrical parameters can be identified:

- the *elevation angle*  $\varepsilon$  at which a user can see the satellite above the horizon,
- the *nadir angle*  $\vartheta$  that gives the deflection of the user from nadir as seen from the satellite,
- the *earth central angle*  $\psi$  between the sub-satellite point SSP and the user, and
- the *slant range*  $d$  denoting the distance between the user terminal and the satellite.

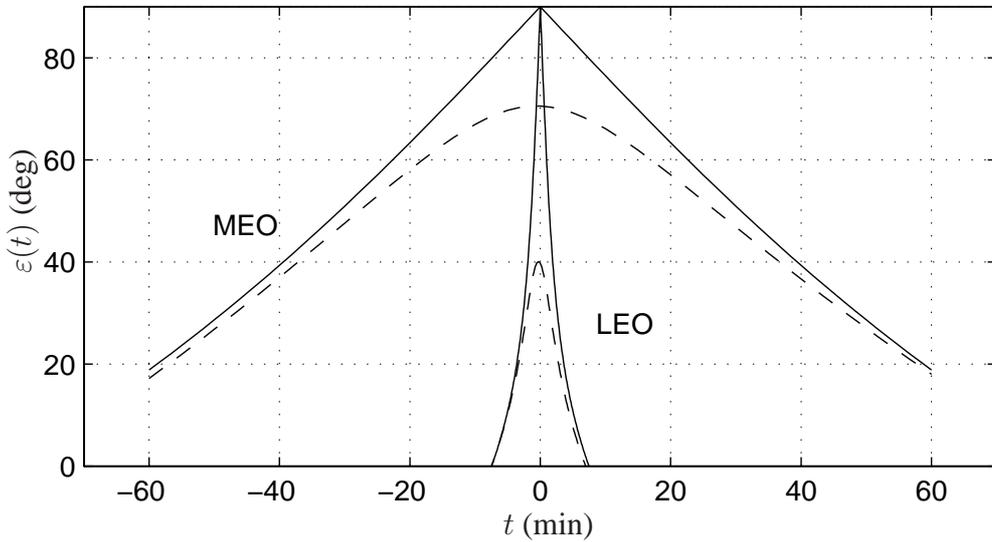
Using the sine and cosine laws for the triangles STN and OTS, the earth central angle and the nadir angle can be given as a function of the elevation angle,

$$\psi = \arccos\left(\frac{R_E}{r} \cos \varepsilon\right) - \varepsilon, \quad (2.5)$$

$$\vartheta = \arcsin\left(\frac{R_E}{r} \cos \varepsilon\right), \quad (2.6)$$



**Figure 2.7:** Geometric relations in the plane STO (satellite S, earth station T, earth center O).



**Figure 2.8:** Time dependency of the elevation angle  $\varepsilon$  for an overhead pass of a MEO and LEO satellite and a  $70^\circ$  MEO and  $40^\circ$  LEO passage.

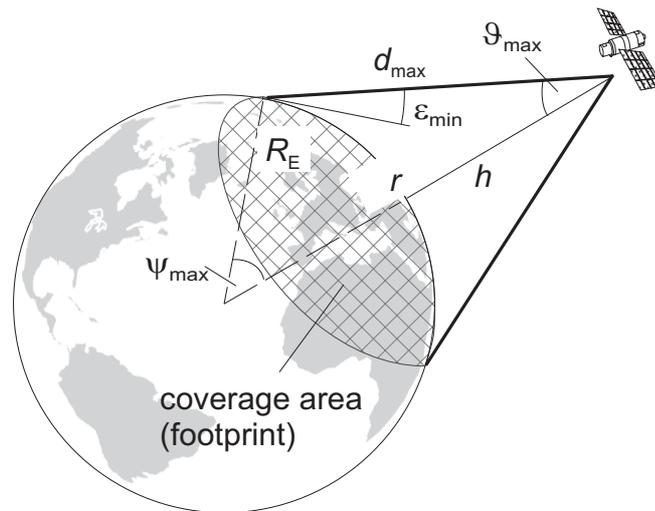
and the slant range  $d$  can be calculated from

$$d = \sqrt{R_E^2 + r^2 - 2R_E r \cos \psi}. \quad (2.7)$$

Since non-geostationary satellites move relative to an earth-fixed station, these parameters vary with time. The variations can be calculated using the Eqs. (2.3) and (2.4) for the satellite tracks. As a particularly interesting case, Fig. 2.8 illustrates the resulting time-varying elevation for some selected LEO and MEO passages.

### 2.3.2 Coverage Area

The *coverage area* or *footprint* of a satellite is defined as the area on the earth's surface where a satellite is seen with an elevation angle  $\varepsilon$  greater than a given  $\varepsilon_{\min}$ , see Fig. 2.9. The threshold  $\varepsilon_{\min}$



**Figure 2.9:** Coverage area (footprint) of a satellite.

defines the border of the (circular) coverage area. For a given orbit altitude a *coverage angle*  $\psi_{\max}$  corresponds to a certain minimum elevation  $\varepsilon_{\min}$ . The minimum elevation  $\varepsilon_{\min}$  is an important system parameter since it affects the number of satellites and orbital planes for global coverage of a system.

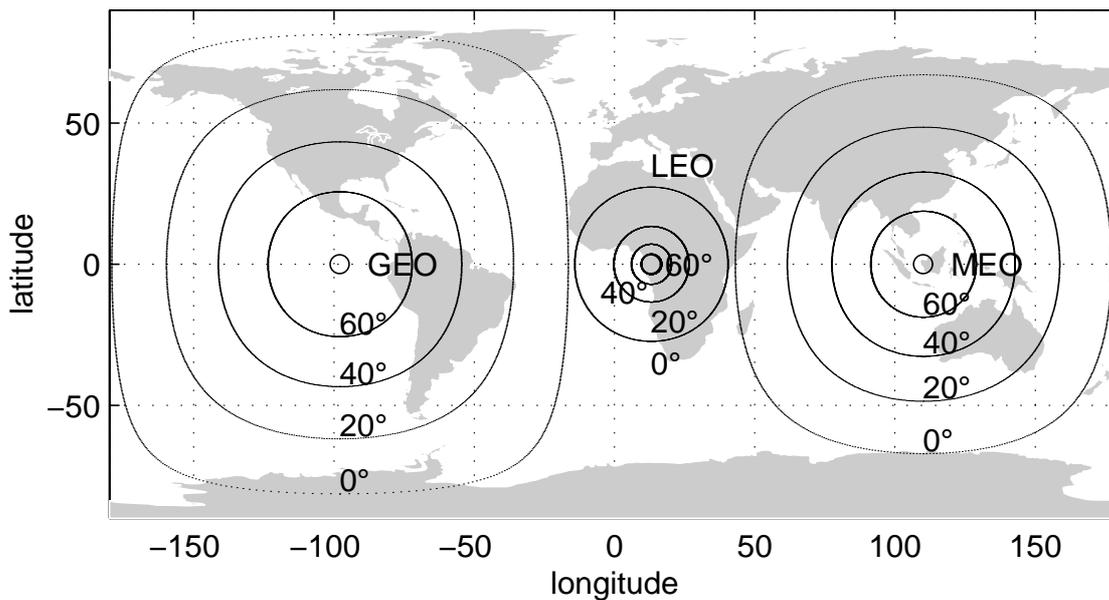
The coverage area of a satellite as defined by geometry is a spherical cap on the earth's surface. Its contour is determined by one of the following parameters which are related to each other through the Eqs. (2.5) to (2.7): the minimum elevation  $\varepsilon_{\min}$ , the maximum nadir angle  $\vartheta_{\max}$ , the maximum slant range  $d_{\max}$ , or the coverage angle  $\psi_{\max}$ . From Fig. 2.9 it can be seen that the coverage area increases, the higher the satellite altitude  $h$  and the lower the minimum elevation  $\varepsilon_{\min}$ . Examples for the coverage area of satellites in different orbit types are shown in Fig. 2.10 for several minimum elevation angles. Based on the above parameters, both the arc length and the area of the spherical cap can be calculated and used in simple estimations on lower bounds for the number of satellites and orbit planes required in a constellation for single global coverage; this has been developed in [BJLW94, WJLB95] and illustratively summarized in [LWJ00].

## 2.4 Satellite Constellations and Intersatellite Link (ISL) Networks

### 2.4.1 Basic Design Considerations

Single satellites can only provide coverage of limited areas at a given time. To extend the coverage, a satellite system may use a number of satellites. The composite of all satellites in the system is called a *constellation*. In this thesis, we consider only constellations where all orbit planes are of equal type, which is in fact the case in the majority of all commercial and study constellations that have been proposed or implemented.

In a non-geostationary satellite constellation, the *total instantaneous coverage area* consists of the union set of the coverage areas of all single satellites at a given time. Due to overlapping



**Figure 2.10:** Coverage areas of satellites in different orbit types for several minimum elevation angles.

the instantaneous constellation coverage area is in general smaller than the sum of all satellite coverage areas. Obviously, in designing a satellite constellation, a major consideration is to provide continuous coverage to a specified area on earth. Typically, this is either true global coverage (including the polar caps) or full equatorial coverage, i.e. up to a certain latitude. The *guaranteed coverage area* of a constellation is defined as the regions on earth in which at least one satellite is visible for 100% of the time. The guaranteed coverage area is a function of latitude and longitude and depends on the orbit and constellation type. Typically, satellites in geostationary provide a regional coverage that can be extended to a multiregional service by using several satellites. With GEO satellites full global coverage cannot be achieved since the polar regions cannot be reached from a GEO position. Furthermore, the elevation angle to a GEO satellite decreases at higher latitudes. Here, constellations of inclined or polar MEO or LEO satellites can provide service with reasonably high elevation angles.

*Multiple coverage* is given if satellite footprints overlap, i.e. if a user in the considered area sees more than one satellite simultaneously. This multiple visibility can be used to improve the availability and quality of the service through the concept of satellite diversity, which we will discuss later.

Finally – and particularly important from a networking viewpoint – the constellation design crucially affects the resulting dynamic ISL network topology, which will be addressed in many parts of this thesis.

## 2.4.2 Constellation Types

Over the last three to four decades, research on satellite constellation design has concentrated on providing optimal ground coverage with a minimum number of satellites in typically regular constellations. Circular orbits are particularly well suited for such an optimization, as their satellites offer a constant footprint size throughout the orbit. Scanning all relevant literature on this topic, one comes to distinguish between two basic types of constellations, namely

**Table 2.1:** Possible classifications of the two fundamental constellation types.

Criterion	Type (1)	Type (2)
Constellation pattern (polar view)	star	delta
Constellation pattern (overall view)		rosette
Inclination	(near-)polar	inclined
Equatorial angle subtended by ascending nodes	$\pi$	$2\pi$

- *Walker star* constellations, and
- *Walker delta* constellations,

named after one of the pioneers in satellite constellation design, J. G. Walker, who originally dubbed the two types *star pattern* and *delta pattern* constellations [Wal70, Wal77], or shortly, just *star* and *delta* constellations, for their view at the pole, where the orbital patterns appear either as a ‘star’ (\*) or as a ‘delta’ ( $\Delta$ ), as we will illustrate below. Before entering the presentation of both types in more detail, Table 2.1 lists other names for these constellations types that have been proposed in the literature; they will be explained in the following as well.

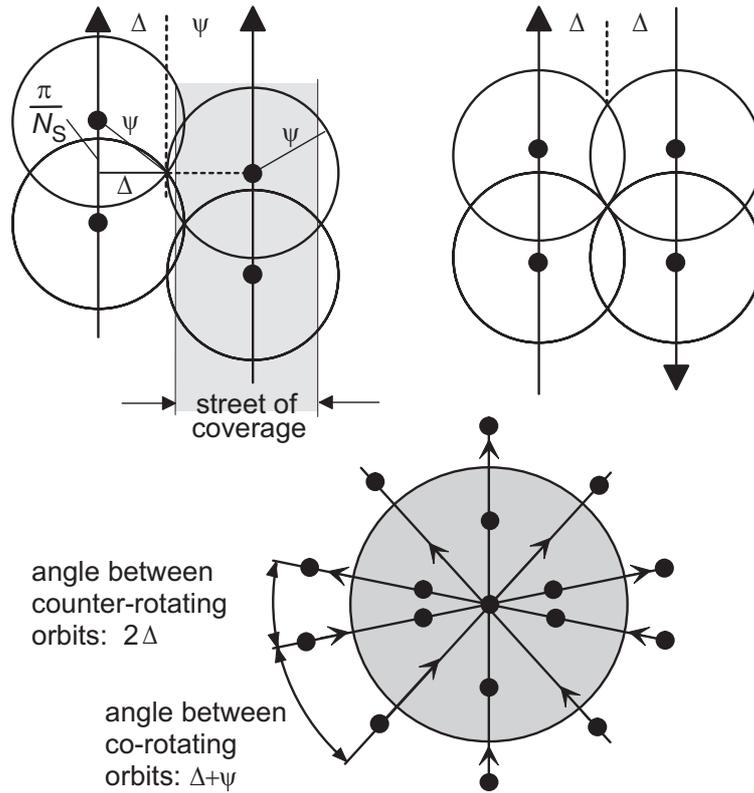
### 2.4.2.1 Star Constellations

A star constellation consists of a number of orbit planes with an inclination  $i \approx 90^\circ$ . Evidently, this inclination is the reason for classifying them also as *polar* constellations. The orbits are arranged in such a way that the total of their ascending nodes – or more precisely, the accumulated equatorial plane spacing – makes up a subtended angle of approximately  $180^\circ$  in the equatorial plane. The alternative name  $\pi$ -constellation [Gal99, Gal01] obviously stems from this feature.

As a result we have a configuration where the satellites in adjacent orbits are co-rotating; they form a *surface* of co-rotating satellites, which at a given longitude appear either *ascending* (satellites from south to north) or *descending* (satellites from north to south) [WPE01, Woo01]. At the poles, all orbit planes intersect to form a ‘star’ in a polar view, cf. Fig. 2.11. Consequently, in this configuration there are two counter-rotating orbits at the edges of the ascending and descending hemispheres; the region between them is called the *seam*.

Further to the characteristics stated above, the star constellation is characterized by the following parameters:

- $P$  is the number of orbit planes with an inclination  $i \approx 90^\circ$ , and  $N$  is the number of satellites in the constellation.
- $N_S = N/P$  denotes the number of satellites per orbit plane. The satellites are evenly spaced in the orbit plane to form a *street of coverage*, in which satellites hand over connections to following satellites (this concept was first proposed by Lüders [Lüd61] but not yet applied to constellations for global coverage at that time). The angular difference between neighbors in one plane is  $2\pi/N_S$ .



**Figure 2.11:** Illustration of a star constellation: co-rotating orbits and street of coverage, counter-rotating orbits with seam, and angles between the orbit planes.

- The satellites in adjacent orbits must be shifted relative to each other to provide coverage without gaps. The phase shift in co-rotating orbit planes is  $\pi/N_S$ .
- The spacing of the orbit planes is explained in Fig. 2.11. With  $\psi$  denoting the one-sided earth central angle of the satellite footprint (coverage angle, see also Fig. 2.7), for continuous coverage of co-rotating orbits the planes must not be spaced by an angle separation larger than  $\psi + \Delta$  where  $\Delta$  can be calculated using spherical geometry:

$$\cos \Delta = \frac{\cos \psi}{\cos(\pi/N_S)}. \quad (2.8)$$

- For continuous coverage the spacing between the two counter-rotating orbits must be smaller. From Fig. 2.11 it can be seen that the maximum spacing is  $2\Delta$ .

Hence, a condition for global coverage can be formulated:

$$\pi = (P - 1) \cdot (\psi + \Delta) + 2\Delta. \quad (2.9)$$

Note again the  $\pi$  which was also used for naming this type of constellation.

Several different optimization techniques for star constellations were derived over the years by Beste [Bes78], Rider [Rid85] and Adams and Rider [AR87], each adapted to the different underlying goals, either a particular multiple coverage or coverage above a specified latitude. The work by Adams and Rider [AR87] is also credited as the basis for the Iridium constellation, which is today probably the most well-known representative of the class of star constellations, cf. Fig. A.1 in the Appendix.

### 2.4.2.2 Delta Constellations

A delta constellation consists of  $P$  equally inclined orbit planes (where the inclination is generally less than  $90^\circ$ ) with their ascending nodes being equally spaced along the full  $360^\circ$  of the equatorial plane, thus the alternative name  $2\pi$ -constellation [Gal99, Gal01]. This leads to a situation where ascending and descending planes of satellites and their coverage continuously overlap, rather than being separated as with the Walker star. Again for the view at the pole, Walker named such constellations ‘delta’ because in the case of three orbital planes a rounded triangle, or Greek  $\Delta$ , is formed by the planes around the pole [Woo01].

The constant longitude offset  $\Delta\Omega$  between the planes is  $2\pi/P$ . On each plane a number of satellites  $N_S = N/P$  are equally distributed with an angular spacing of  $\Delta\omega = 2\pi/N_S$ . Besides the inclination angle  $i$  and the period  $T$  three other parameters are needed to describe the constellation, typically denoted by a triplet, called the

$$\textit{Walker Notation: } N/P/F$$

where  $N$  is the number of satellites in the constellation,  $P$  is the number of orbit planes, and  $F$  is the so-called *phasing factor* ( $F = 0, 1, \dots, P - 1$ ). The phasing factor  $F$  determines the angular offset  $\Delta\omega_p$  between the satellites in adjacent orbit planes,

$$\Delta\omega_p = 2\pi \cdot \frac{F}{N}, \quad (2.10)$$

so as to guarantee a completely regular appearance of the constellation geometry across the orbital sphere; in particular, the satellites on the “last” orbit plane  $P$  will find the ones on the “first” orbit plane 1 at the same offset as encountered between all pairs of planes from 1 to  $P$ .

Walker’s work has been extended and generalized by Ballard [Bal80] who concentrated on optimized multiple coverage by multiple low-inclination planes with proper phasing. He called the delta constellation an *inclined rosette* constellation for the typical patterns being similar to the petals of these flowers.

The parameters of an exemplary Walker 9/3/1 constellation are illustrated in Fig. 2.12. Except Iridium, all constellations considered in this thesis and listed in Appendix A are Walker delta constellations; the main reasons for that will be shortly mentioned in the next section and explained in more detail in Sections 4.1.1 and 5.2, as they have been one result of the research performed in the framework of this thesis.

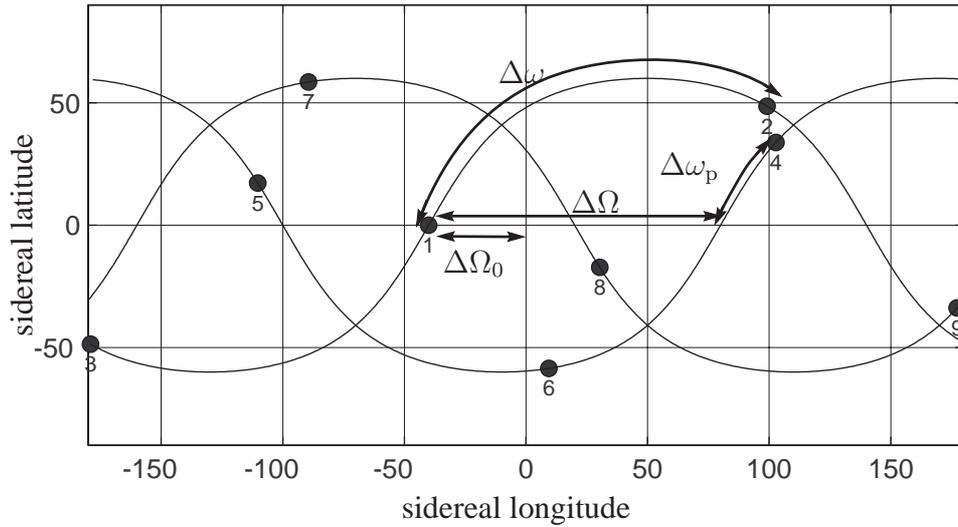
### 2.4.3 ISL Network Topologies

With each of the constellations identified before, a set of communication nodes is up in space; what remains to be done is to properly interconnect them by means of intersatellite links. Compared to terrestrial backbone networks, the *regular* distribution of nodes over the spherical surface in such constellations at a given time is an extremely nice feature, and one would certainly like to draw benefit from that as far as possible. An obvious solution would be to form a likewise *regular* network mesh composed of *intra-orbit* and *inter-orbit* ISLs.

From a topological viewpoint, one can basically distinguish between two types of ISLs:

---

<sup>3</sup>In effect, the definition of  $\Delta\Omega_0$  is related to the RAAN  $\Omega$  of the first satellite in the Walker delta constellation.



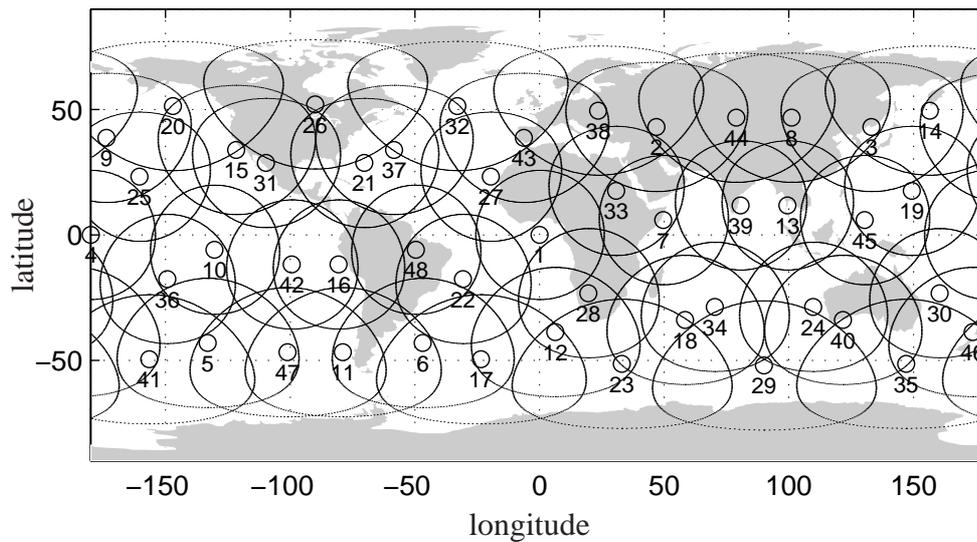
**Figure 2.12:** Walker constellation 9/3/1:  $N = 9$ ,  $P = 3$ ,  $F = 1$ ,  $\Delta\omega_p = 2\pi F/N = 40^\circ$ ,  $\Delta\omega = 2\pi P/N = 120^\circ$ ,  $\Delta\Omega = 2\pi/P = 120^\circ$ ,  $i = 60^\circ$ . The offset  $\Delta\Omega_0 = -40^\circ$  of the first orbit plane from the reference meridian defines the orientation of the Walker constellation with respect to earth for a given time.<sup>3</sup> (Note: the satellite track is displayed without earth rotation in sidereal coordinates.)

- *Intra-orbit* ISLs connect satellites following each other in the same orbit plane. Obviously, for circular orbits and controlled constant phasing of all satellites within the orbit, both the distance and the relative pointing remain fixed. Therefore, intra-orbit links will be generally permanent.
- *Inter-orbit* ISLs connect satellites in different orbit planes, and the most obvious partners are on the neighboring orbits, of course. However, in both star and delta constellations, neighboring orbits cross each other at the respective highest latitudes, and thus require either (i) tracking mechanisms that can cope with the maximum range and the change rate of pointing angles, or (ii) deactivation while partners are swapping sides (the most challenging phase) and reacquisition of the old or a new partner.

The issue of link permanence of inter-orbit ISLs is one of the key issues in ISL topology design for both star and delta constellations, and has been thoroughly considered as a part of the research work forming this thesis; consequently, it will be treated in more detail in the core chapters in the context of routing and capacity dimensioning.

However, at this point we can already state the following:

While in delta constellations the tracking requirements between two inter-orbit neighboring satellites tend to be somewhat permanent, but not too extreme over time, the case is different with star constellations: over a wide range of latitude *between* the polar regions, the distance and pointing angle variation between satellites in *co-rotating* planes is fairly low, but it is extreme over the poles and across the seam (between the counter-rotating planes); as a consequence, Iridium as a current reference system for an ISL-based star constellation switches off ISLs over polar regions and has not implemented cross-seam links at all. As shown in its ISL topology snapshot, Fig. A.7 in the Appendix, especially the latter feature breaks the ultimate regularity of the ISL network. Both the seam and switched links over the poles impose major challenges on routing in a connection-oriented scenario, and will be mainly tackled in Chapter 4. This statement should not stand without



**Figure 2.13:** Coverage snapshot of Globalstar for a minimum elevation angle of  $10^\circ$ .

mentioning the good news, too: in any case topology changes are predictable and periodic (aside from equipment failures), and the graph topology is still somewhat dense, leading to a multiplicity of similar routes to most destinations. Both of these simplifying properties should and can be exploited, as we will show. In Section 5.2 we will also show that a completely regular and permanent network topology may be built upon a delta constellation, with obvious advantages for optimized network operation in terms of traffic routing and capacity dimensioning.

## 2.5 Multiple Coverage, Handover and Satellite Diversity

### 2.5.1 Multiple Satellite Coverage

As we have previously indicated, *multiple coverage* of specified areas on earth may be one of the design goals for satellite constellations. Multiple visibility of satellites above the minimum elevation angle may be exploited to improve signal transmission quality especially in critical (mobile) user environments by the use of satellite diversity, which will be treated below. In general, if footprints overlap more than needed for guaranteeing permanent coverage without gaps, also the necessary handovers between satellites can be performed with an increased degree of freedom, as more alternative successor satellites may be available, or just because the handover time window becomes longer. A brief introduction to handover issues will also be given below.

A good example for a constellation design aiming at multiple coverage is the Globalstar system (see Appendix A). The constellation provides seamless global coverage with the exception of polar regions as shown in Fig. 2.13. The constellation was designed to provide good visibility for the temperate climate zones with main business relevance, as indicated in Fig. 2.14.

In these examples, a minimum elevation angle of  $10^\circ$  is assumed. However, as one can guess from Fig. 2.13, the multiple coverage entails in most situations at least one satellite being visible at considerably higher elevation. Figure 2.15 gives a rough impression of the time shares for the elevation of highest satellites in typical constellations, as a function of user latitude.

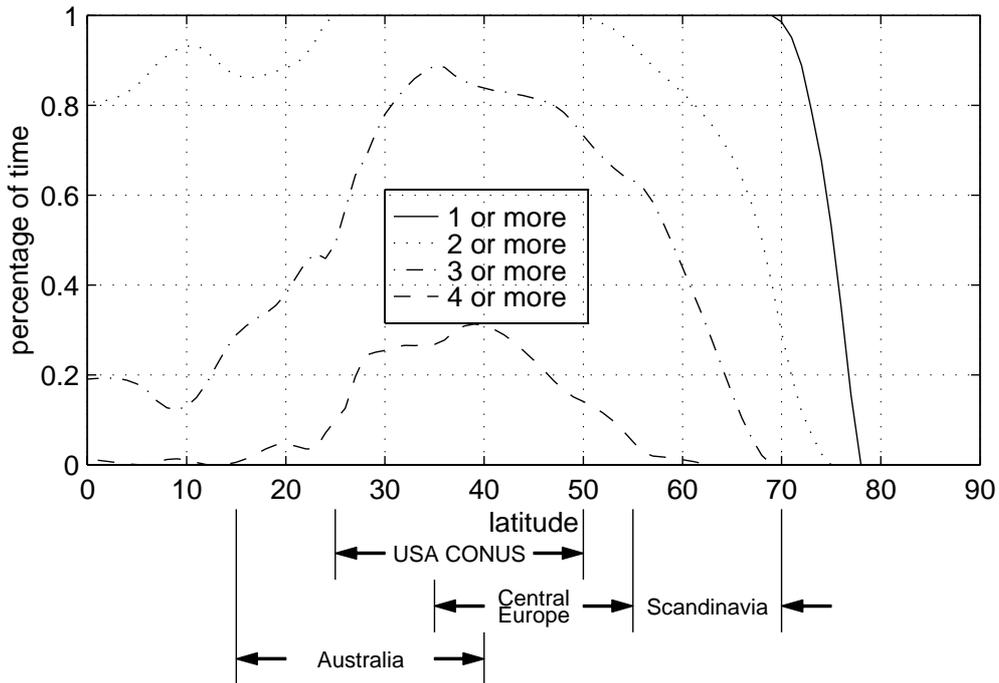


Figure 2.14: Multiple satellite coverage of Globalstar for a minimum elevation angle of 10°.

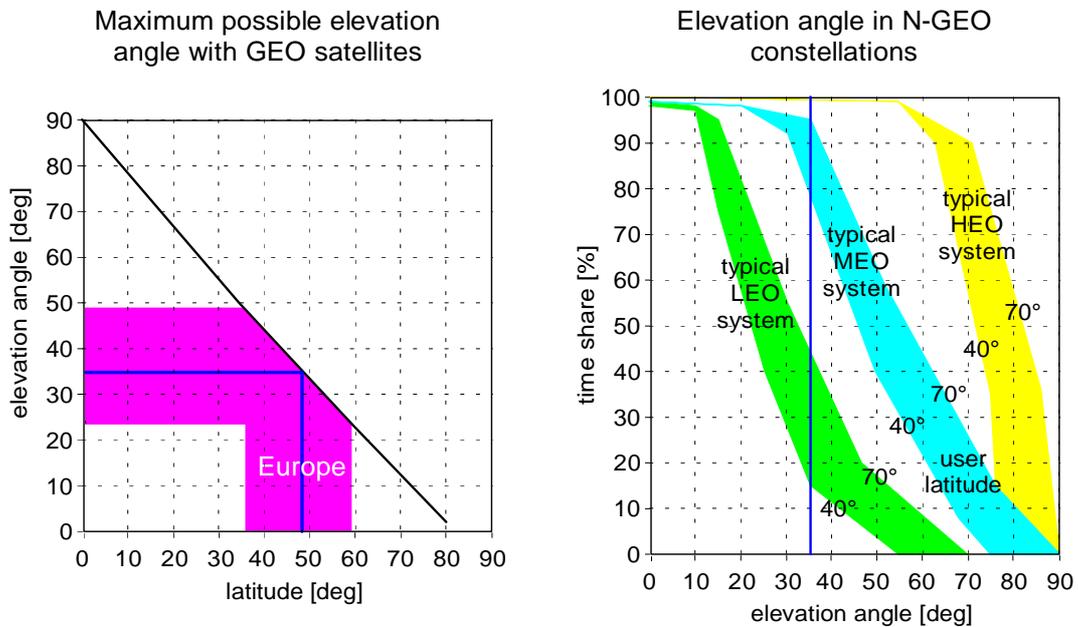
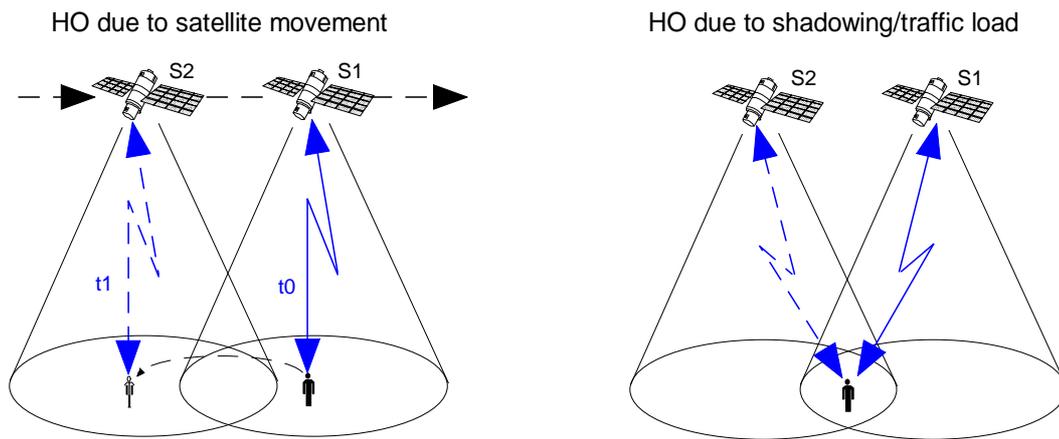


Figure 2.15: Comparison of typical elevation angles achieved with different constellations. The lines in the diagram mark the elevation angle of 34° at which a GEO satellite (positioned at the same longitude) is seen from Munich, 48° latitude.



**Figure 2.16:** Handover causes in constellation networks.

## 2.5.2 Handover

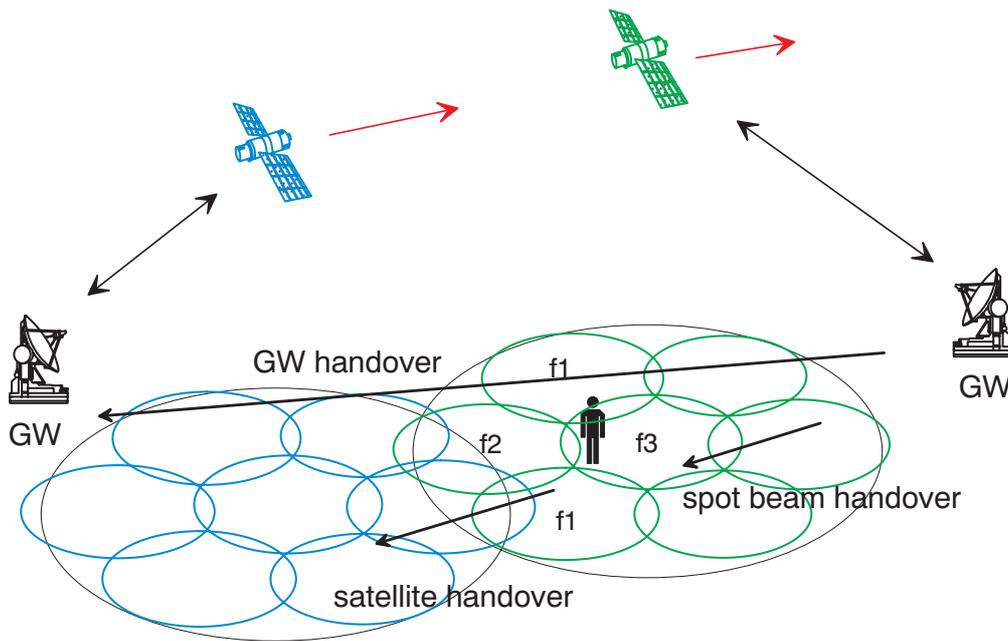
Handover (HO) actions in non-geostationary satellite constellations can be classified mainly with respect to two criteria, namely (i) the cause for the handover and (ii) the network entities involved. With respect to the former, handovers may be due to (cf. Fig.2.16)

- satellite movement, when the currently serving satellite drops too low above the horizon, which is deterministic and required handovers are thus predictable,
- user mobility leading to bad propagation conditions, for instance shadowing, which is not predictable at all, and
- requirements for or interest in peak load reduction; the instantaneous traffic load is not really deterministic, but some prediction/estimation of the short-term load variation may be possible.

With respect to the network entities between which handovers can take place, we can identify (cf. Fig.2.17)

- satellite handovers between overlapping satellites, in many cases within a “street of coverage”,
- spot beam handovers between adjacent spot beams of the same satellite, and
- gateway (GW) handovers, where due to a satellite handover the user may effectively leave the service area of the initial GW.

Obviously, a spot beam handover is not applicable to compensate for bad propagation conditions. GW handover is usually avoided, because it would have an impact on the routing within the terrestrial fixed network, and, especially, it would presuppose widely deployed intelligent network functionality. Satellite and spot beam handovers occur in a deterministic sequence mainly due to satellite movement. According to the different visibility times of LEO and MEO satellites, MEO



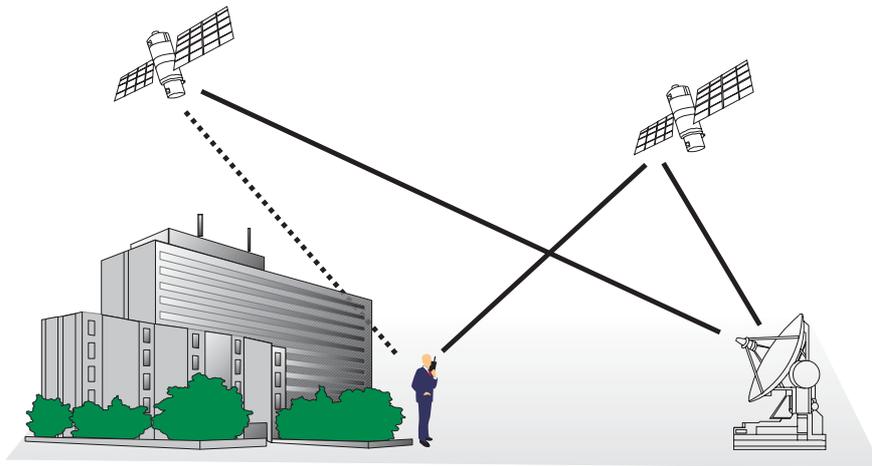
**Figure 2.17:** Handover between network entities in constellation networks.

satellite handovers may be avoided for connections with a short holding time, like typical telephone calls, whereas this is less likely in LEO systems; this issue will be looked at in more detail in Section 4.2.1 from a routing perspective.

In *satellite-fixed cell systems*, the cells move over the earth according to the satellite's motion, and handovers between spot beams of a satellite arise, occurring more frequently than satellite handovers. Especially in LEO systems with a large number of spot beams, this can lead to an enormous frequency of beam handovers, and each handover action on principle entails a certain probability of call dropping. Moreover – and this is particularly challenging for an ISL network/routing designer – in such a system handovers are completely *asynchronous*, imposing mostly unpredictable processing and rerouting burdens on the space segment as well.

This problem has been addressed by Restrepo and Maral [RM96, RM97, MRD<sup>+</sup>98] proposing and analyzing in detail the concept of *earth-fixed cell systems*, where cells are fixed geographic regions on earth. During the time interval which a beam is assigned to a cell, the beam is continuously steered in order to compensate for the satellite's motion. Within a satellite footprint, no beam handover occurs. If the satellite footprint leaves an earth-fixed cell, the respective beam is redirected to a new incoming cell. This eliminates an uncertainty as to when a handover occurs, resulting in a near-to-zero probability of handover failure. In addition, satellite handovers (and potentially resulting ISL rerouting) can be performed in synchronized manner. However, earth-fixed cell systems require a larger number of satellites compared to satellite-fixed cell systems [RM96]. Moreover, the satellites must be capable of performing beam steering and cell switching. An example of an earth-fixed cell system is the Teledesic [Stu95] system concept.

In the mobile/personal environment, the handover process should also consider the impact of the satellite channel. Optimizing the quality of service implies that the best mobile link should always be chosen in terms of received signal power or signal quality. This would produce a large number of handovers, however, corresponding to a high signaling overhead in the network. Applying a power hysteresis threshold or a time-out period to handover initiation reduces the frequency of handovers and leads to a trade-off between service quality and signaling load [CB95].



**Figure 2.18:** Principle of satellite diversity.

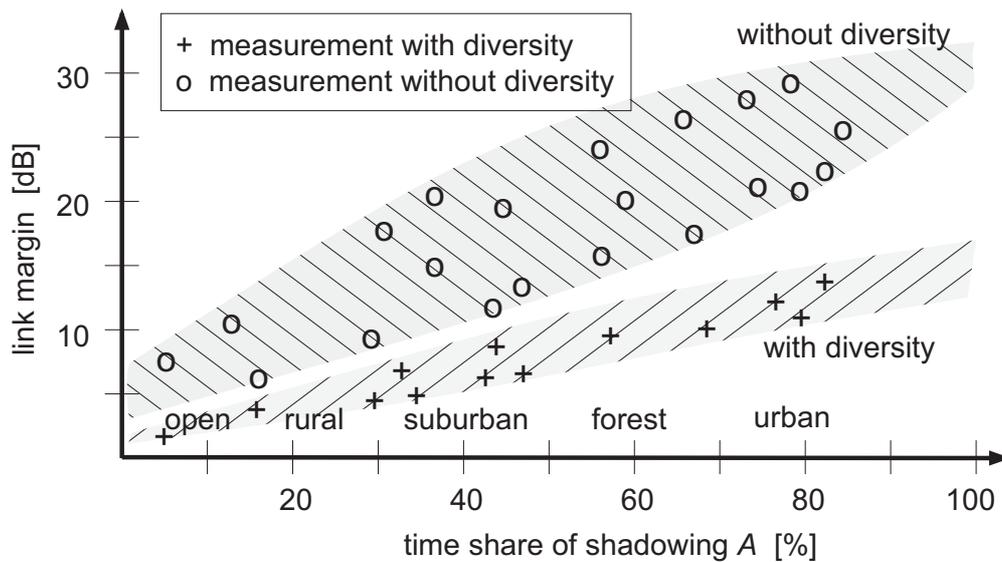
### 2.5.3 Satellite Diversity

Satellite diversity describes the concept of providing simultaneous communication with a user via two or more satellites, where the communication partner may be another user or a gateway (GW). It can be regarded as a fundamental counter-measure against the signal deterioration through shadowing and blocking in the user environment, and is thus particularly interesting for service provision to mobile users moving in or between critical environments. Unlike terrestrial cellular systems, however, which mainly rely on diversity combining of signals from several indirect (i.e., shadowed) paths between user and base station, the main potential of satellite diversity is based on increasing the possibility of having at least one satellite in view under direct line-of-sight (LOS) conditions, as shown in Fig. 2.18. Therefore, satellite diversity is also said to be a kind of *macro diversity* [Jah99] or may be classified as *artificially introduced multipath* [STTE94].

Obviously, multiple satellite coverage is a precondition for the utilization of satellite diversity, and a possible diversity gain will always depend on the extent of multiple coverage at considered points on ground. Systems such as the CDMA-based Globalstar and the TDMA-based ICO essentially provide double coverage of large (latitudinal) regions on earth, as they were designed for satellite diversity operation. In this context it is worth emphasizing that in contrast to many earlier (published) opinions on this issue, satellite diversity is *conceptually* applicable in TDMA- as in CDMA-based systems, and differences occur in terms of implementation and performance, where it is not justified to claim a general advantage of CDMA for satellite diversity if one takes into account all relevant criteria [GHJ95].

With respect to the kind of combination of two (or more) received signals, three different methods can in general be distinguished: (i) selection combining, (ii) maximal ratio combining and (iii) equal gain combining [Pro89, Hes93, Lee93]. Due to the pronounced two-state behavior of (mobile) satellite channels, as indicated above and further elaborated in Sections 4.2.2 and 6.1.1, it is for a rough estimation or simulation of satellite diversity performance acceptable to restrict to the least complex selection combining, which is in fact a *switched diversity* in that it simply switches always to the best available satellite. In this thesis, selection combining is always assumed if not otherwise stated.

The most important performance measure for satellite diversity from a system viewpoint is the improvement achieved in terms of service availability (i.e. the percentage of time when the service



**Figure 2.19:** Required link margins at L band for different shadowing with handset reception; required link availability 95%.

is available) compared to a system without diversity operation; this will be also referred to as *diversity gain* in the remainder of this thesis. From a system design perspective, one could also state that diversity allows to reduce the required link power margin for a certain nominal service availability targeted by the system.

In [Jah94] the required link margins for S-PCN systems have been determined by measurement for many environments with different shadowing. To permit an investigation of diversity performance for selection combining, two transmitters have been used with different elevation and azimuth angles. The required values for the link margins with and without diversity are shown in Fig. 2.19 versus the time-share of shadowing  $A$  for a link availability of 95%. The measurement points correspond to low and medium elevation angles  $\varepsilon = 15^\circ - 45^\circ$ . Obviously, in all environments a substantial improvement can be achieved by diversity. Even in environments with very strong shadowing (such as forest or urban areas) realistic link margins around 10 dB are sufficient for a good quality of service. For comparison, the link margin available in some LEO and MEO satellite systems for personal communications is as follows:

- Iridium: 16 dB, no satellite diversity;
- Globalstar: 3 dB + power control + satellite diversity;
- ICO: 10 dB + satellite diversity.

## 2.6 Network Traffic Dynamics

Concerning traffic and capacity considerations, the focus of this thesis is clearly on the ISL network *internal* traffic flows and mechanisms to route and optimize them. The serving source/destination satellites are then consequently regarded as collective traffic sources and drains, not considering

single end user connections anymore. Nevertheless we derive our figures for source traffic demand keeping in mind that single calls with their specific parameters form the “atomic” traffic entity for our ISL study as well.

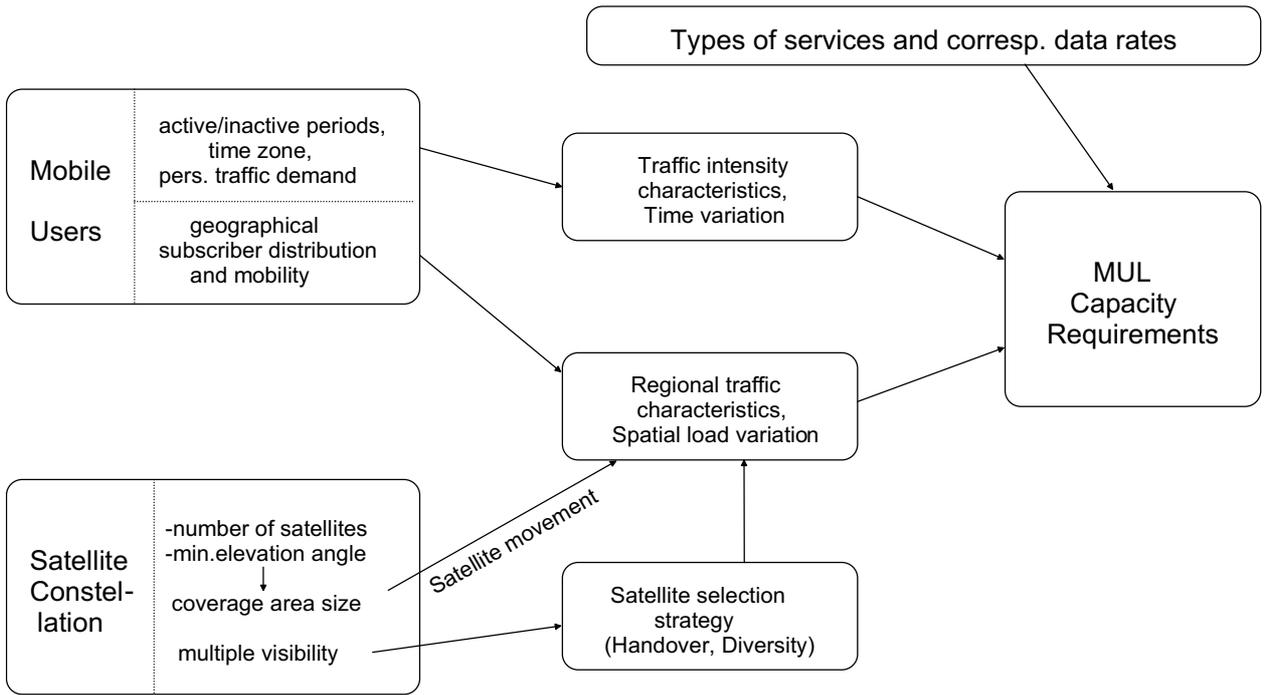
In this section, we want to present some characteristic features and relationships concerning traffic in constellation networks, mainly addressing the issue of traffic dynamics. Without restriction of universality, we will do this under simplifying assumptions for a single-service (mobile) telephony scenario; in fact some of the following material summarizes early research work from the days when the satellite community was still in the learning process to understand fundamental effects in operating these new and highly complex global satellite networks. Repeating these earlier exercises, it is from a tutorial viewpoint certainly instructive rather than out of date, as the basic purpose has been insight, not numbers (cf. the quote of Richard W. Hamming on page 111). Concluding this, one can certainly assume that basic characteristics of network traffic dynamics as presented in the following should be valid for any multiservice broadband scenario as well.

### 2.6.1 User Link Capacity and Traffic Dynamics

Whatever constellation network and whatever service and global user scenario we assume, the user uplink is the key to network traffic flows. Earlier work [LR94, BW94a, Wer95] has qualitatively and quantitatively investigated capacity requirements and traffic variations as seen from the orbiting satellite in its footprint. The major impact factors driving user link capacity requirements, and their relationship, are illustrated in Fig. 2.20. One of the key lessons learnt has been that the time variation of user link traffic is clearly dominated by the spatial load variation which an orbiting satellite, by collecting this spatial load in its footprint, effectively translates into time variation – if one takes the perspective of the space network and considers this satellite just as ingress node. This component of traffic variation is not found in any fixed terrestrial network, where traffic variation only depends on time zones, user activity variation and statistics of user group behavior. Conceptually, but quantitatively much less pronounced, a comparable effect can be observed in terrestrial cellular mobile networks when spatial load variation results from user mobility and translates into time-varying traffic from the base station perspective, for instance.

After these conceptual and qualitative considerations, it is now interesting to give some quantitative impressions. As a first useful approach we present a simple method for a rough estimation of the required capacity in the (mobile) user link of the considered global systems, if nothing but the total number of (targeted) subscribers to this system/service is known (in fact, this has been the case with MSS systems over several years in the 1990s):

- The satellite system shall serve a total number  $N_u$  of users (subscribers) with an average activity of  $A_u$  in Erl. Then, the required total system capacity in terms of duplex channels can be estimated (neglecting blocking) as  $C_{\text{sys}} = N_u A_u$ .
- The subscribers are assumed to be uniformly distributed on the continents, covering 26% of the earth’s surface.
- The global satellite constellation shall comprise  $N$  satellites, which are assumed to be uniformly distributed over the orbital sphere.  $N$  may be given from a system design or may be estimated as a function of orbit altitude and minimum elevation, as it has been proposed



**Figure 2.20:** Impact factors for MUL capacity requirements.

at the end of Section 2.3.2. Then, the capacity required for the mobile user link of a single satellite can be estimated as

$$C_{\text{sat}} = \frac{1}{0.26} \cdot \frac{C_{\text{sys}}}{N} = \frac{1}{0.26} \cdot \frac{N_u A_u}{N} . \quad (2.11)$$

Taking Iridium as an example and assuming 1 million subscribers with a 5 mErl busy hour activity results in 292 channels per satellite.

Following this simple estimation approach, which is certainly useful when one has no powerful simulation tool at hand, and which can of course be refined having more precise data about user distribution and their activity, we now present two representative numerical results which have been gained from extensive system simulations.

Equation (2.11) does not take into account user accumulation in certain regions, nor does it capture the potential influences of footprint overlaps, satellite handover scheme or the usage of more than one average channel per user to exploit satellite diversity. A more realistic capacity estimation can be achieved by assuming a refined user distribution, and by simulating the satellite constellation with sufficient resolution, taking into account satellite elevation, diversity and/or handovers.

Figures 2.21 (a) and (b) show the simulation result for an Iridium and a LEONET satellite, respectively, exhibiting high traffic peaks when the satellite crosses areas of high user density (for constellation parameters refer to Appendix A). Note especially that the global user distribution is based on the *6-regions-model* from traffic modeling framework TMF 1 presented in Appendix C.1. Assuming dynamic channel allocation to spot beams and thus taking into account blocking for the whole satellite according to the Erlang-B formula

$$E(A, C) = B = \frac{A^C / C!}{\sum_{k=0}^C A^k / k!} , \quad (2.12)$$

with  $A$  being the peak loads extracted from the simulations, the required up/downlink channels for a satellite  $C_{\text{sat}}$  can be determined.

In the considered examples, for a target blocking probability of 5% we end up with 508 channels for Iridium and 1290 channels for LEONET. The difference is of course mainly due to the larger coverage area of the MEO system. Note that the Iridium simulation result can not directly be compared with the figure from the earlier estimation mainly due to the uniform land mass user distribution assumed there.

Finally, assuming that all satellites in the constellation are identical, one can directly calculate the total “installed” system capacity (i.e. the sum of all channels offered to the ground) as

$$\begin{aligned} C_{\text{sys}} &= N C_{\text{sat}} & (2.13) \\ &= 66 \cdot 508 = 33\,528 \quad \text{for Iridium, and} \\ &= 15 \cdot 1290 = 19\,350 \quad \text{for LEONET,} \end{aligned}$$

which confirms the impression from the load curves that the average utilization of the MEO system is higher.

## 2.6.2 Long-Term Variation and System Period

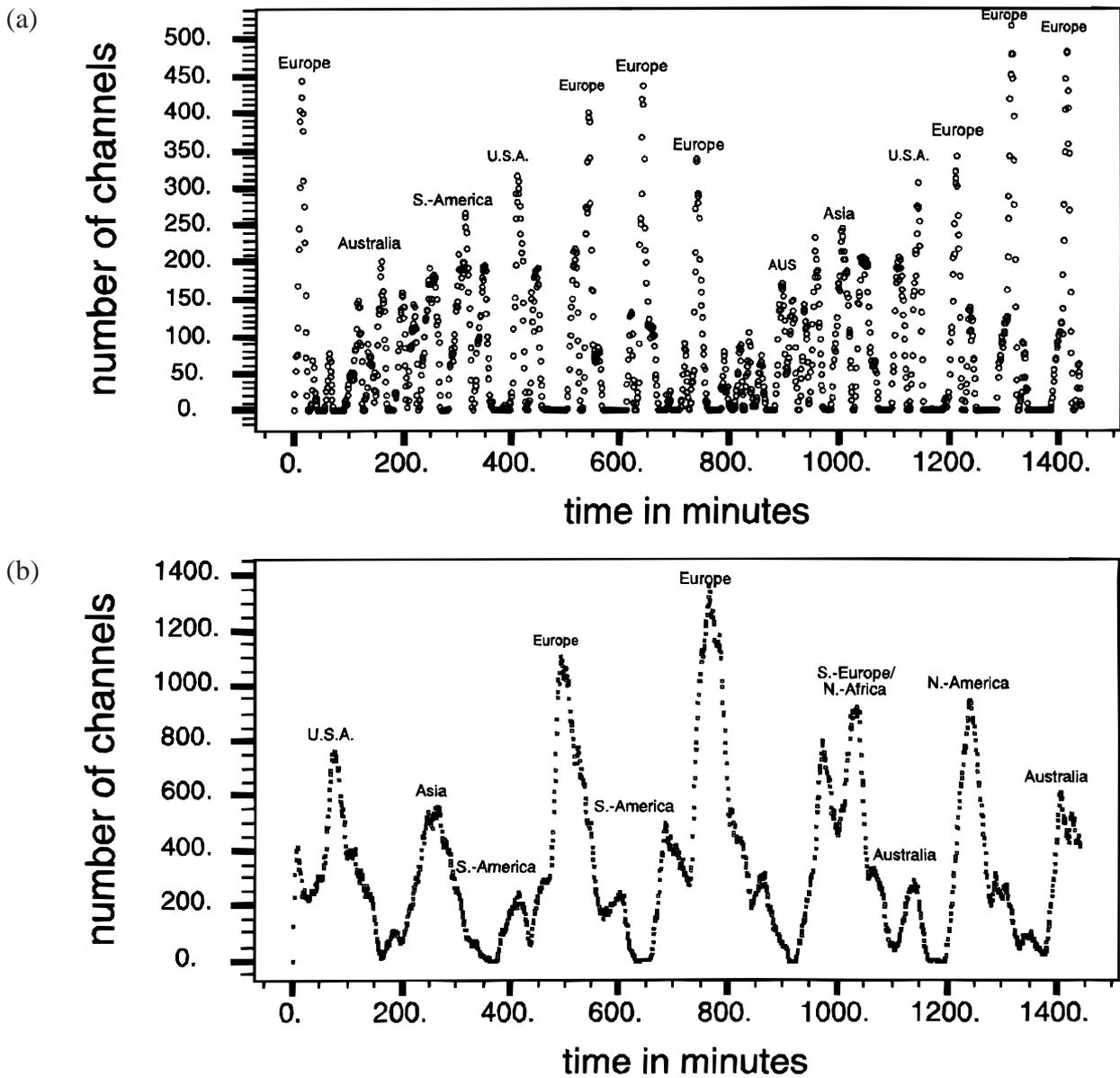
We have shown that the relative movement of satellites and user location areas on earth leads to a complex combined spatial and time variant traffic pattern collected and delivered by every single satellite in its footprint. The long-distance share of this traffic is routed through the space segment. Characteristics of source/destination traffic variation are then transferred into the ISL traffic mix in a smoothed form, because every single intersatellite link also carries some share of transit traffic.

When studying network traffic in ISL-based constellations, as soon as some “realistic” traffic input from ground is considered (and not directly a simplified ISL network traffic model, or an end-to-end demand matrix that is just assumed for a numerical exercise), one has to take care of the fundamental fact that the dynamic network topology pattern repeats itself with the constellation period  $T$ , whereas the period for repeating ground traffic patterns is usually the solar day  $T_{\text{Sol}}$ ; and, moreover, repetitions of the constellation are fixedly oriented in space whereas the “orientation” of earth traffic patterns could be said to slowly rotate in space with the earth’s rotation around the sun. In order to make sure that all relevant cases of constellation/traffic pattern pairs are taken into account, thus especially ensuring that the usually targeted worst case is included, one must specify an appropriate minimum investigation period, especially for simulations.

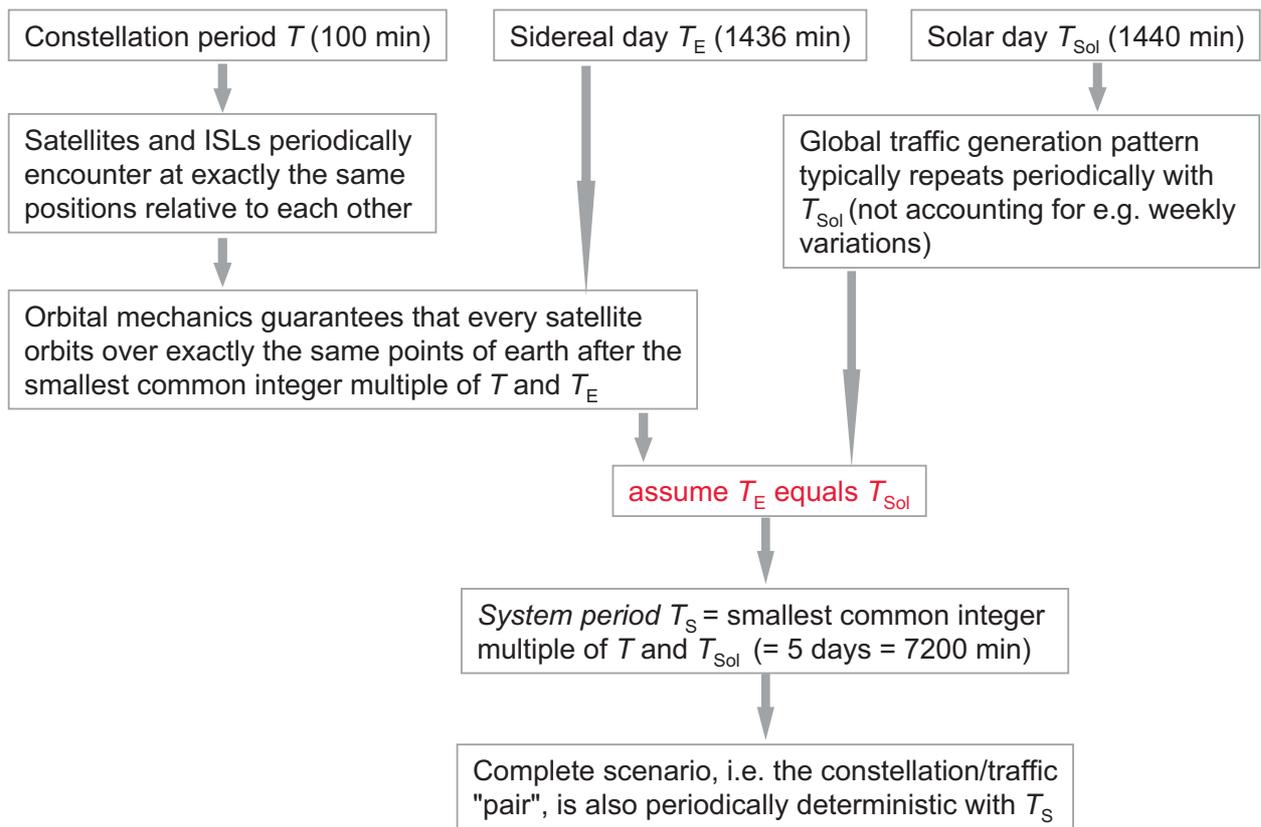
As a general solution, and as a compromise between realistic simulation modeling and acceptable simulation effort, we propose to define a *system period*  $T_{\text{S}}$ :

$$T_{\text{S}} = \min \{ \text{lcm}(T, T_{\text{E}}), \text{lcm}(T, T_{\text{Sol}}) \}, \quad (2.14)$$

where the values should be given as rounded integers (e.g., in minutes). The term *system* captures the combined constellation/traffic scenario, as illustrated in Fig. 2.22. This flowchart brings in another, less formal interpretation of the system period: for the sake of simplicity we usually neglect the small difference between sidereal and solar day for system-level traffic considerations and somewhat fit the calculation of the system period to our needs (i.e. keep it as small as possible),



**Figure 2.21:** Time-varying capacity requirements in the up/downlink of an (a) Iridium, (b) LEONET satellite; 1 million users according to TMF 1 (App. C.1), 5 mErI constant user activity, no daily traffic variation, no satellite diversity, handover always to the highest-elevation satellite.



**Figure 2.22:** The notion of a *system period*, using the Iridium constellation as an example.

knowing that in reality we will have a small “offset” between (solar) daily traffic pattern and constellation repetition pattern. Given that, the system period is well suited as *upper limit* for any ISL routing and dimensioning investigation based on exact traffic variation models. However it can even be shown that one solar day is enough in many cases, especially for LEO constellations which repeat often during one day, and taking into account the general coarseness of input traffic data. In Section 6.1.2 we will provide a kind of proof that this is well justified.

# Chapter 3

## ATM-Based Satellite Networking Concept

*Le Roi est mort! Vive le Roi!*

— JULIA PARDOE, *Life of Louis XIV.*, vol.3 (1886)

*ATM – or IP with MPLS?*

*Use a path – or a route via switched path express?*

*Based on a sound concept a simple thing:*

*The king is dead, long live the king!*

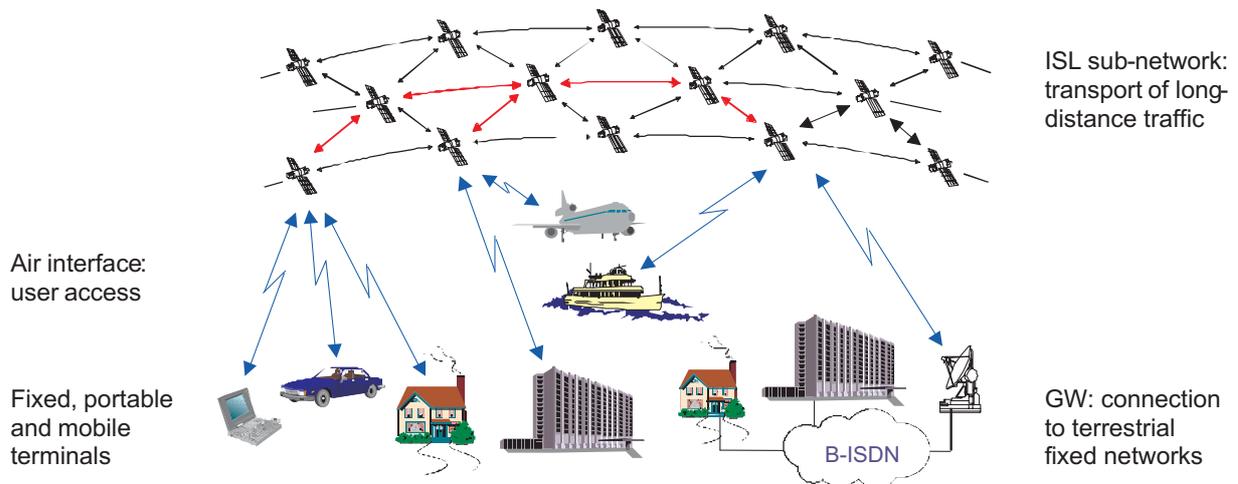
— MARKUS WERNER (2000)

### 3.1 Motivation and Scenario

The *asynchronous transfer mode (ATM)* has been proposed and widely accepted as the underlying transmission and switching scheme for the broadband ISDN (B-ISDN) in order to efficiently provide multimedia and multiservice communications in integrated networks. While basically operating in connection-oriented manner, ATM combines the strengths of both circuit switching and packet switching techniques, and is therefore capable of transporting all kinds of information and supporting each type of service; in particular, different levels of quality of service (QoS) guarantees are provided. Based on the connection-orientation paradigm and on a well-specified framework of services and traffic classes (including their QoS parameters) supported, reliable and powerful traffic engineering methods can be applied in network design.

We have chosen to consider ATM as the basis for our satellite constellation networking concept mainly for three reasons:

- The clear trend for future satellite communications is towards multimedia applications and multiservice networks.
- We think that a connection-oriented operation scheme is best capable of exploiting three main properties of dynamic ISL topologies, namely predictability, periodicity, and regularity, while inherently guaranteeing the order of arriving information entities over significant periods of end-to-end connections.



**Figure 3.1:** System concept of an ISL-based LEO/MEO satellite network for global communications and its connection to terrestrial fixed networks.

- In the design of broadband multiservice satellite networks with QoS guarantees and service differentiation, traffic engineering is a crucial component.

Having said this, we present an illustrative overview of the network scenario we consider for ATM-based constellation networks in Fig. 3.1.

The global network is essentially based on a satellite constellation which uses an ISL subnetwork for the transmission of long-distance traffic. The ISL segment comprises full network layer functionality, i.e. each satellite can be considered as an orbiting ATM switch or cross-connect from the pure networking perspective.

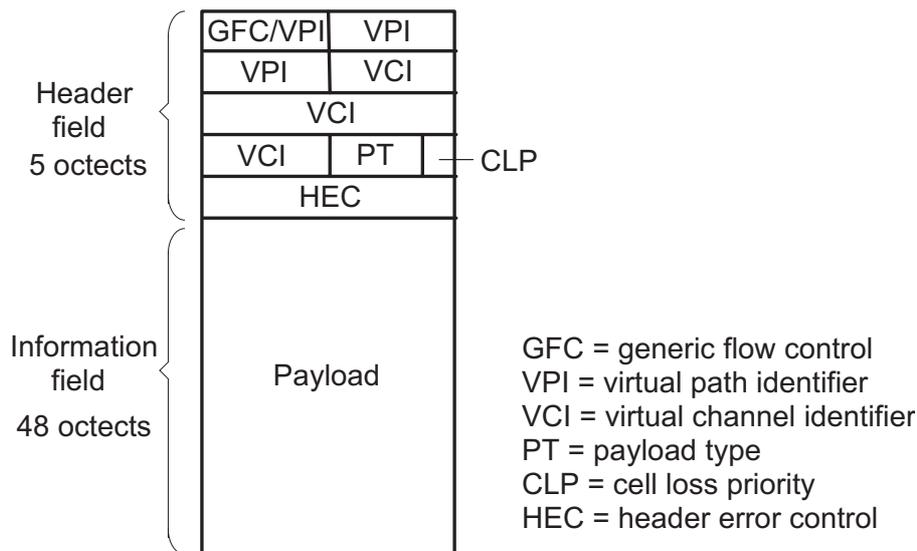
The end users can access the system over the air interface. The potential variety of user terminals may be classified into fixed, portable and mobile terminals. They can be used by individuals (e.g., handheld, laptop) or by user groups (shared access to a terminal on-board a train, for instance).

The satellite network is connected via fixed earth stations (gateways) to terrestrial networks. Note that the B-ISDN is shown as an example here; other terrestrial networks to interwork with include PSTN, ISDN, GSM, Internet, etc.

## 3.2 Basic ATM Principles

Besides the original ITU and ATM Forum documents, ATM is described in detail in many textbooks, for instance by De Prycker [dP93], Tanenbaum [Tan96], Stallings [Sta98], and McDysan and Spohn [MS98]. The book by Schwartz [Sch96] may be consulted for an overview with some focus on quantitative studies of traffic characterization, admission control, switching, and congestion control.

In the following we briefly present those basic principles of ATM which are useful or required to understand the ATM-based networking concept we develop at the core of this chapter.



**Figure 3.2:** Structure of an ATM cell (UNI/NNI).

### 3.2.1 Transmission and Multiplexing Scheme

ATM is based on an asynchronous time multiplexing technique, where all information is transported in fixed length packets called cells. Each ATM cell consists of 53 octets or bytes, cf. Fig. 3.2: the first 5 bytes constitute the cell header and the remaining 48 bytes the information field; the latter contain the user data (payload), whereas the cell header contains all information that is needed to control the flow of the cell through the ATM network.

ATM allows multiple logical connections to be multiplexed over a single physical interface. All cells belonging to one connection follow the same route through the network (connection-oriented transfer). The cells arrive at the destination in the same order as generated by the source. Moreover, they exhibit low transfer latency and small delay jitter. Established connections are virtual in the sense that no fixed line is set up between the end nodes, but cells are logically assigned to a connection. All cells belonging to one connection can be recognized by virtual connection identifiers assigned at call setup time. The use of multiple identifiers for one “bundled” connection allows the support of multimedia services.

The traffic sources can generate cells asynchronously and at different rates. In a multiplexing device (e.g. a switch), after buffering, the cells from different sources and services are multiplexed in a statistical manner, according to their arrival times and service classes. Hence, there is no fixed time relation between the input and output cells of the multiplexer. Statistical multiplexing allows the network resources to be flexibly and efficiently shared among all connections, favoring time-varying bit rates and multimedia services.

### 3.2.2 Virtual Connections

At connection setup an end-to-end route through the network (i.e. a series of links and nodes) is determined, which is used by all data cells belonging to that connection. With this procedure, a unidirectional *virtual channel (VC)* is assigned to the virtual connection in each link between two network nodes lying on the route; the series of VCs constituting the end-to-end route is called a

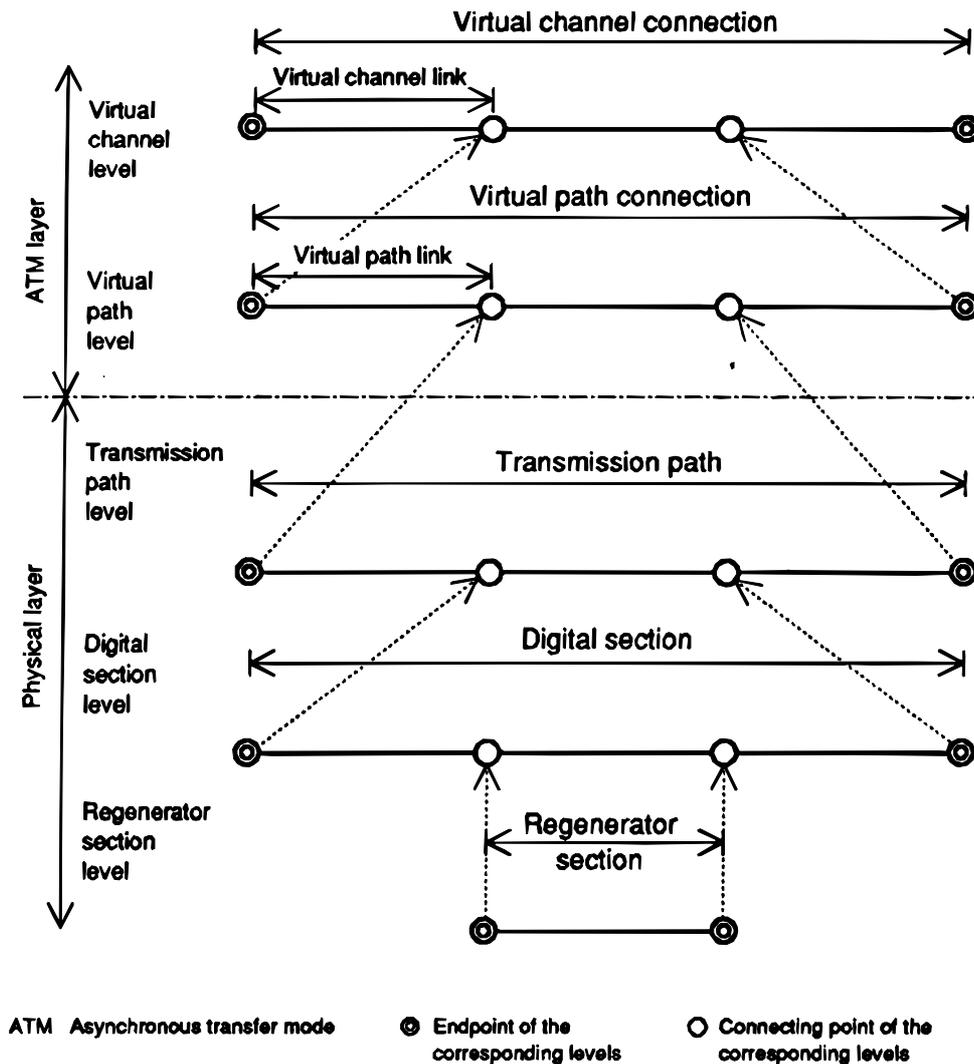
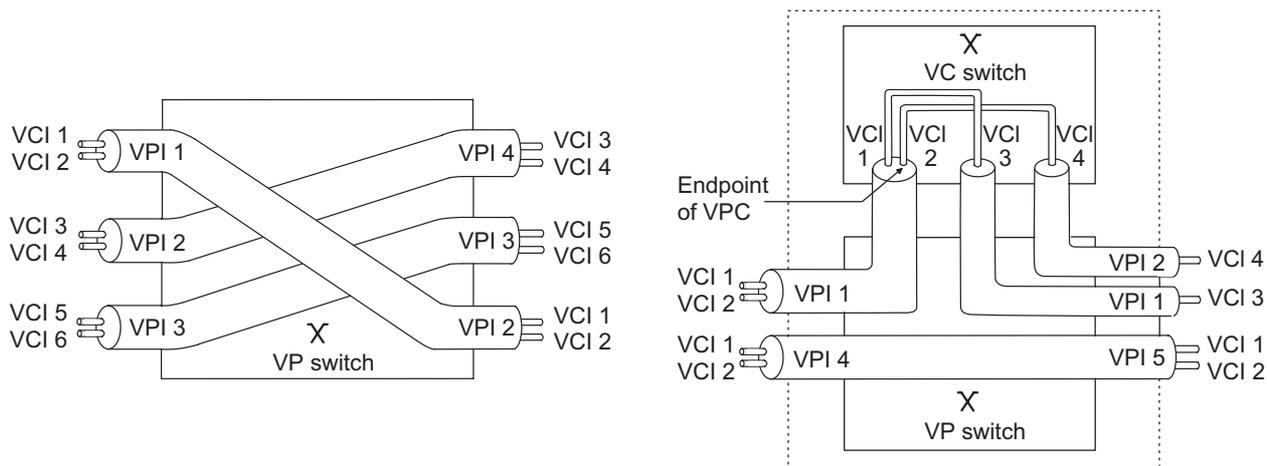


Figure 3.3: Hierarchy of virtual connections and physical path sections [HH91].

*virtual channel connection (VCC)*. Each VC of the VCC and each cell using the VC are uniquely labeled by a *virtual channel identifier (VCI)* which usually changes from link to link; thus, a VCC is determined by a number of VCIs belonging to the links on the route and a routing table in each node on the route, translating the VCIs between the incoming and outgoing links. Accordingly, each link contains a number of VCs, mapped to different connections through their VCIs.

When a group of VCCs use the same route along a part of their end-to-end routes, the respective VCs in a commonly used link can be combined into a single *virtual path (VP)* identified by a *virtual path identifier (VPI)*. The common part of the route is characterized by a series of VPs and VPIs constituting a *virtual path connection (VPC)*, cf. Fig. 3.3. Thus, each link of a connection is characterized by a VCI and optionally by a VPI. This concept of aggregating parallel VCs into a single VP substantially reduces the complexity of switching in the network nodes and eases the setup of a new VCC. While each virtual end-to-end connection is identified by a temporary VCC, VPCs may be defined between network nodes in a more static manner.

Translation and switching of VCs and VPs is done in VC switches and VP switches (also called cross-connects), see Fig. 3.4. Note that VP switches only swap VPIs but don't touch the VCI, whereas VP/VC switches change both.



**Figure 3.4:** VP switching in a cross-connect and VP/VC switching in a switch.

### 3.3 ATM in Satellite Systems

This section presents a possible application scenario and the protocol architecture for satellite ATM systems. We also summarize the particular challenges for satellite ATM in various fields. Hereby we follow largely the presentation in our earlier publication [LWJ00].

#### 3.3.1 Application and Service Scenario

Figure 3.5 shows a schematic application scenario one could think of for ATM-based satellite constellation networks.

The user applications may be based on TCP/IP, ISDN, native ATM, or MPEG protocols and formats, and via protocol conversion (ATM adaptation layer, AAL) generate ATM traffic corresponding to the categories of constant bit rate (CBR), variable bit rate (VBR), available bit rate (ABR), or unspecified bit rate (UBR). In the ATM layer, these traffic streams are multiplexed into a single stream of ATM cells. The transmission of the multiplexed ATM traffic via the satellite air interface requires a special modem at the terminal and on board the satellite, implementing a satellite-specific physical layer (S-PHY), a medium access control layer (S-MAC), and a data link control layer (S-DLC). On board the satellites, ATM switching is used to route the ATM cells into the appropriate ISL or downlink. The interworking with fixed terrestrial networks is provided by gateway stations.

It remains to be stated that a satellite ATM system in general may serve various user and application scenarios: it can either be conceived as an mobile/personal S-PCN targeting primarily single users or as a broadband data network for high data rate transfer between fixed earth stations, or another type of ATM-based network. This should actually be kept in mind when we will present our networking concept in Section 3.4. The concept has been developed and published some years back (1995/1996) at a time when satellite networks were used to primarily address the mobile/personal segment, and it has therefore been conceived as concept for ATM-based S-PCN; this terminology is retained in the presentation within this thesis, somewhat reflecting the origins. Of course, the networking concept is basically valid for any other satellite ATM scenario as well, without explicit further mentioning.

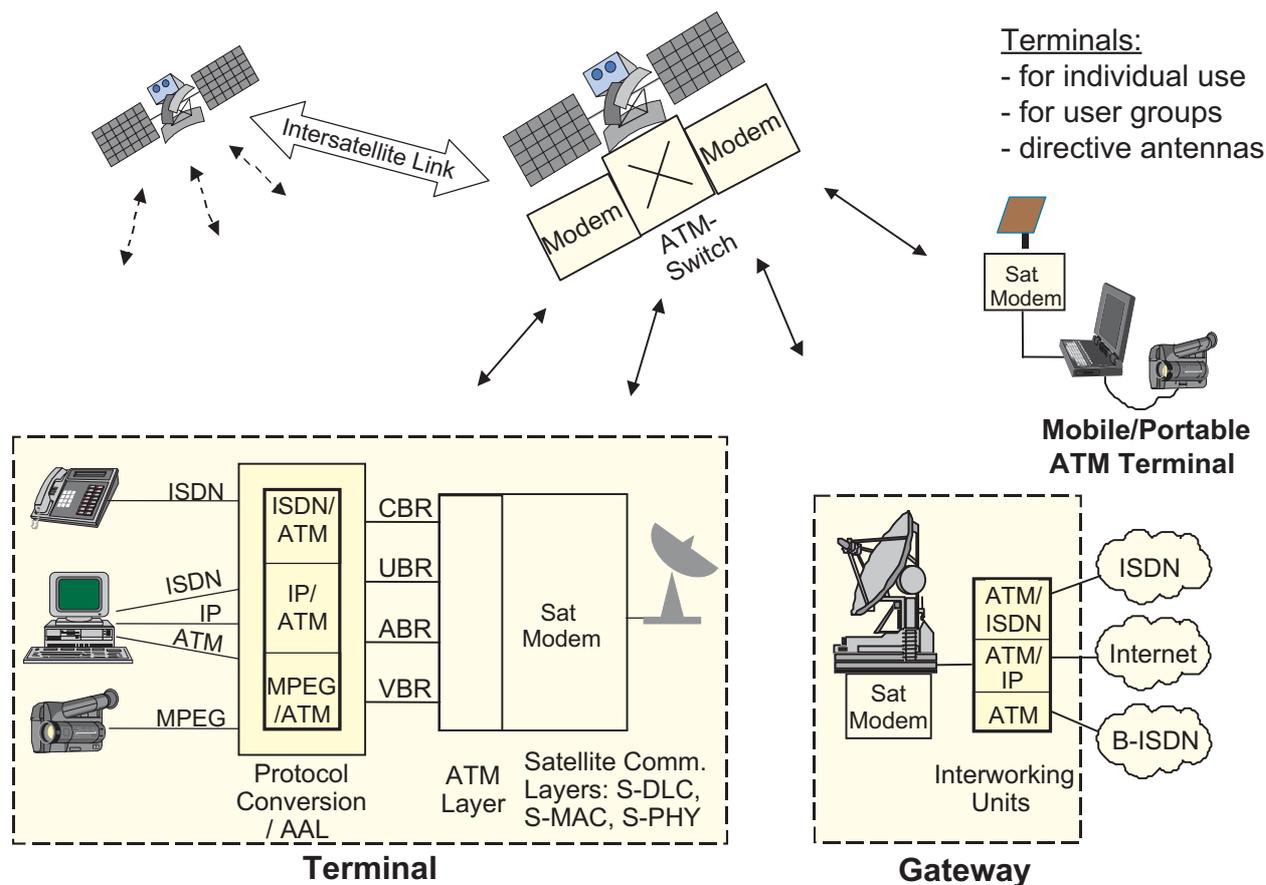


Figure 3.5: Schematic application scenario for satellite ATM.

### 3.3.2 Protocol Architecture and Challenges

A multimedia satellite system might be considered an ATM-based meshed sky network [TIA98] with dynamic network topology, including end user access, ISLs, satellite-gateway connections, and network interconnection via satellites.

In this configuration, several interconnected satellites form an in-orbit ATM network. This ATM satellite network performs ATM switching, traffic and congestion control, and QoS management equivalent to terrestrial ATM networks. The in-orbit ATM switches use NNI signaling for the ISL communication, and UNI or NNI signaling for communication with the ground stations.

The protocol reference model for the fully meshed ATM-Sat network is shown in Fig. 3.6, indicating the satellite-specific protocol layers S-PHY, S-MAC, and S-DLC that have been inserted below the ATM layer. The resource and mobility management functions can be implemented at the gateway earth station or on board the satellites and can be invoked by ATM UNI or NNI signaling via If.a or If.c. Alternatively, internal signaling channels between terminals, gateways, satellites, and an NCC (network control center) can be used (thin lines in the figure).

As indicated by the need for satellite-specific protocol layers, matching ATM and satellites brings along some particular implications and challenges, the most important of which are briefly illustrated in Fig. 3.7. Except the dynamic network topology and ISL routing schemes, we will not further consider these issues herein. The interested reader is referred to [LWJ00] for more details.

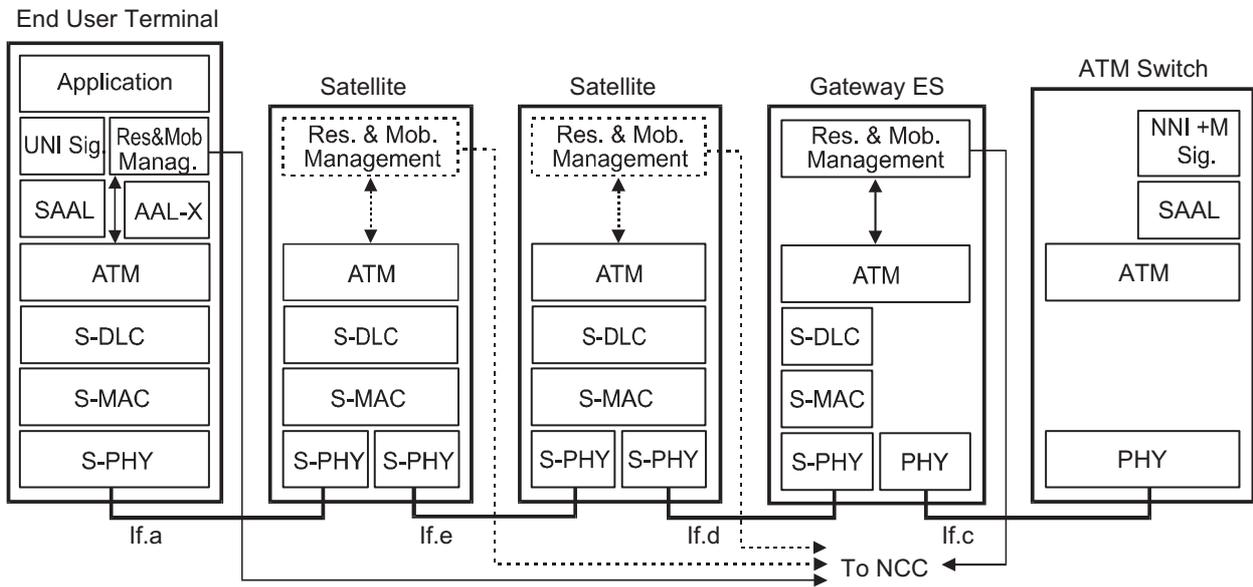


Figure 3.6: Protocol architecture of an ISL-based satellite ATM network.

**Fixed ATM networks:**

- large available bandwidth
- transmission with low bit error rate ( $<10^{-9}$ )

**Wireless ATM:**

- wireless network access
- mobile users

**Satellite communication:**

- wireless network access
- mobile users
- limited bandwidth
- long propagation delay
- signal shadowing and fading, bit errors
- dynamic network topology

**Development of satellite-specific**

- transmission schemes
- error control protocols
- multiple access protocols
- resource management protocols
- ISL routing schemes

Figure 3.7: Specific constraints and requirements for ATM-based satellite networks.

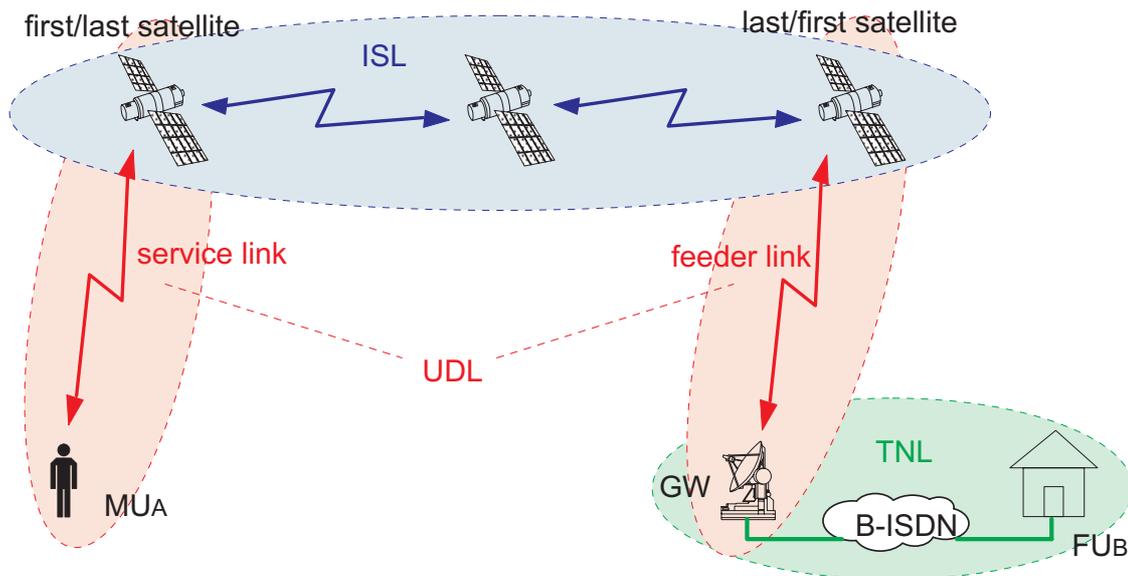


Figure 3.8: Segments of a typical end-to-end connection in ISL-based LEO/MEO satellite systems.

## 3.4 Overall Networking Concept

Forming the core of this chapter, we now present the overall networking concept for ATM-based satellite constellation systems. The concept was originally developed and published in [Wer96] and [WDV<sup>+</sup>97].

### 3.4.1 Network Segmentation

The end-to-end connection scenario of ISL-based LEO satellite systems comprises three segments, according to Fig. 3.8, which illustrates the connection between a mobile user  $MU_A$  and a remote fixed partner  $FU_B$ : the (i) *intersatellite link (ISL)* segment comprises the radio links between pairs of satellites, essentially forming a dynamically meshed subnetwork in space. The (ii) *up/downlink (UDL)* segment incorporates the *service link (SL)* between mobile users and satellites as well as the *feeder link (FL)* between satellites and fixed earth stations (*gateways GW*). The gateway stations act as interface between the satellite system and the (iii) *terrestrial network link (TNL)* segment. In the following, the satellites serving the SL and FL subsegments will be referenced to as *first* and *last satellites*, or in pairs as *terminating satellites* of a given connection. Considering bidirectional communication, clearly every first satellite in an ISL chain always acts simultaneously as last satellite and vice versa.

### 3.4.2 Problem Outline

The application of ATM concepts to satellite systems, especially to LEO/MEO constellation networks with inherent backbone capabilities, aims at the exploitation of the powerful features of the ATM layer transport functions on VC and VP level.

The connection-oriented operation mode of virtual connections – established either as VCC or as VPC – ensures that cell sequence is preserved during their lifetime. Classically, one dedicated

VCC between two ATM end users is established for the holding time of a call and then used by all related ATM cells. Redundant (standby) paths/VCCs may be provided to allow hitless path switching either on-demand (predictable) or as reaction to path failures (unpredictable). In both cases, proper alignment of the ATM cell streams during switch-over is essential for nondisruptive operation of the user connection. In [EEFK95] and [Edm96], an alignment strategy for terrestrial ATM networks is proposed that requires duplication of the cell streams during the alignment phase and thereby guarantees nearly optimal and error-free operation.

In the considered LEO scenario we face the challenge of a multiply dynamic network topology, extending on both UDL and ISL segments. Inherently, this feature in general requires switching between subsequent different paths during the lifetime of a user connection. In the following, this is referenced to as virtual connection *handover* (HO). Obviously, such connection handovers may introduce severe delay jitter resulting from the difference in transmission delay on old and new path.

Fortunately, the LEO/MEO “physical” topology dynamics – as experienced by two fictitious end user locations in the footprints of two specific satellites – is periodically deterministic with the orbit period  $T$  of the satellite constellation. For a specific constellation, this allows to set up a unique time-dependent virtual topology providing for continuous operation of end-to-end connections.

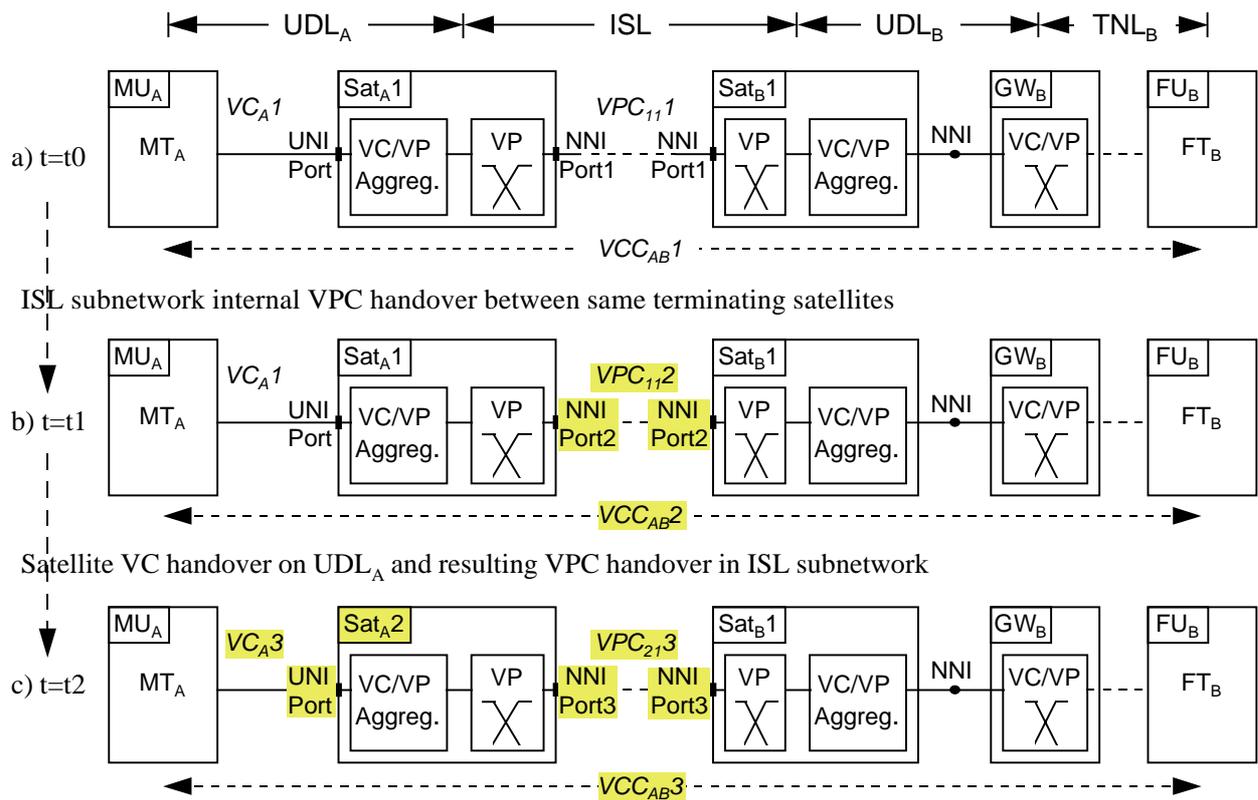
### 3.4.3 Continuous End-to-End Networking

The overall networking concept description serves as basis for a sound problem decomposition and the in-depth discussion of the ISL subnetwork alone. The systematic approach towards an ATM-based networking concept for a satellite personal communication network (S-PCN), or for potential broadband equivalents, considers ISL-based LEO satellite systems for bidirectional point-to-point communications between pairs of end users located on earth. To simplify the concept description, let us further assume “pure” VC connections between the two partners, i. e. no VCC aggregation into VPCs is performed on an end-to-end level (as would be for example appropriate in the case of multiservice communication). However, this assumption is definitely not imperative for the concept. A mobile S-PCN user  $MU_A$  may communicate with a mobile partner  $MU_B$  or with a fixed partner in the terrestrial B-ISDN,  $FU_B$ , and the partner may be within the service area of the same satellite or remote. In the following, the most demanding case of a remote fixed partner is discussed in detail, because this highlights the use of intersatellite links and the satellite/terrestrial interface at the same time.

Along with Fig. 3.9, the operation of the connection is sufficiently explained by a closer examination of the connection setup (a) and two different handover situations (b),(c).

(a) *Virtual connection setup:*

A  $VCC_{AB1}$  is established at  $t = t_0$  with a  $VC_{A1}$  entering the serving first satellite  $Sat_{A1}$  via an UNI port. After MAC/ATM protocol conversion, the aggregation into the applicable VP and the VP switching onto the outgoing VP on NNI Port P1 are performed. A strictly VPC-based virtual topology in the ISL subnetwork now provides a certain  $VPC_{111}$  between first satellite  $Sat_{A1}$  and last satellite  $Sat_{B1}$ . The  $VCC_{AB1}$  is completed through a VC/VP switch at  $GW_B$ , which is the dedicated satellite/terrestrial interface instance for  $FT_B$ , and the terrestrial ATM tail in the TNL subnetwork.



**Figure 3.9:** Call setup and continuous operation over two different virtual connection handover situations. Relevant changes due to the respective handover are emphasized in gray. MT, FT = mobile/fixed terminals.

The following procedures guarantee time continuity of the virtual connection over two different handover situations.

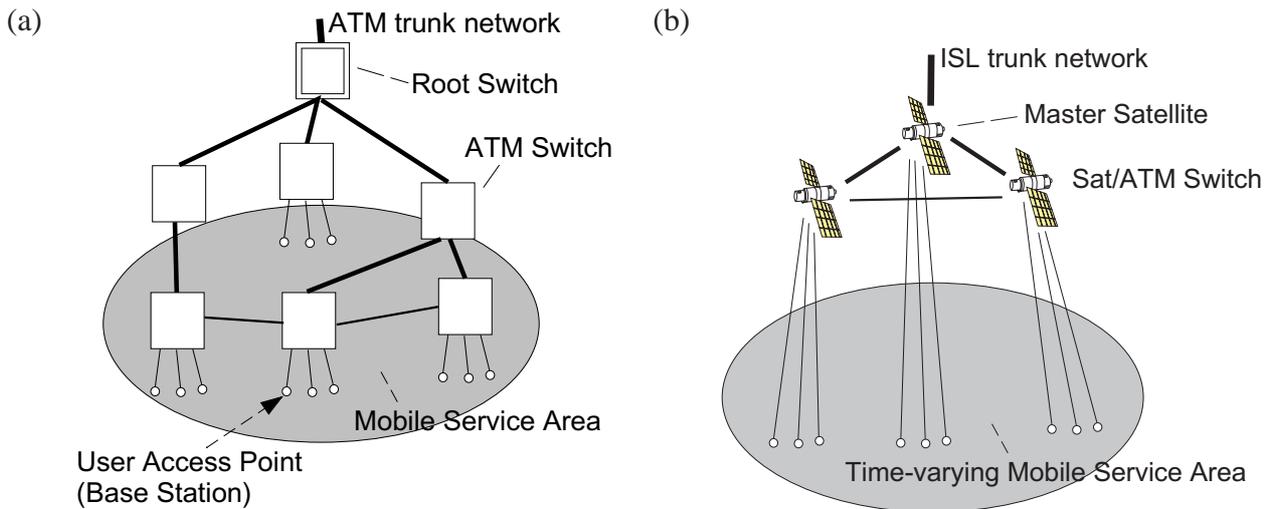
(b) *ISL subnetwork internal VPC handover:*

At  $t = t_1$  a handover from  $VPC_{11,1}$  towards  $VPC_{11,2}$  – using different paths between the *same* terminating satellites – becomes necessary due to some changes in the ISL topology. Working around the instantaneous path delay difference between the old and new VPC requires hitless path switching with corresponding hardware complexity (cell stream alignment buffers etc.). Following original ATM terminology, the overall  $VCC_{AB2}$  is obviously different from earlier  $VCC_{AB1}$ , nevertheless continuing the same call.

(c) *VC handover on UDL/SL:*

At  $t = t_2$  a satellite handover on  $UDL_A$  is performed. A twofold consequence becomes obvious: first, a handover towards a new  $VC_{A3}$  in the incoming satellite’s mobile user link has to be performed, and secondly, within the ISL subnetwork a new  $VPC_{21,3}$  will succeed  $VPC_{11,2}$ , both actions contributing to the new  $VCC_{AB3}$ . A satellite handover for  $GW_B$  works likewise, affecting a possibly large number of calls (transported through  $UDL_B$ ) simultaneously.

The networking tasks in UDL, ISL, and TNL parts of the system can be considered separately. The TNL part is easily separable by assuming that the gateway either acts as central VC/VP switch between the S-PCN internal part and the external terrestrial network or that it is implemented as



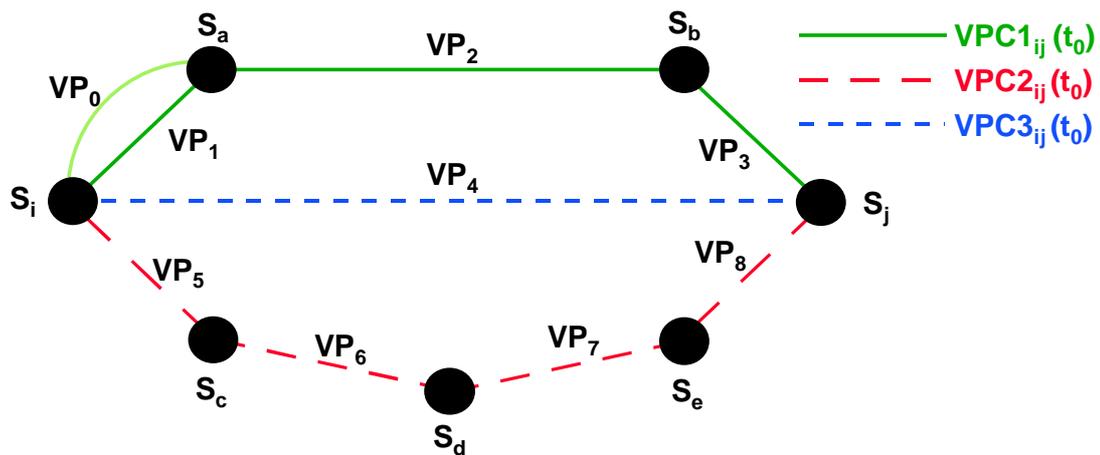
**Figure 3.10:** Virtual connection tree (VCT) concept applied to handover/rerouting in (a) terrestrial cellular and (b) multibeam satellite ATM networks.

intermediate node between two different ATM networks. The networking of the TNL part itself turns out to be the “classical” terrestrial B-ISDN one and is not further considered herein.

The periodical topological changes within the ISL subnetwork are completely deterministic, since they are only dependent on the respective satellite positions. Therefore it is possible to set up off-line, i.e. prior to the operational phase of the system, a dynamic virtual cross-connect network incorporating all satellites as pure VP switches, thereby providing (a set of) VPCs between any pair of terminating satellites at any time. This VPC topology is clearly decoupled from the VC-based UDL part by the VC/VP (de)aggregation interface in the terminating satellites and consequently does not affect UDL operation even during a VPC handover of type (b). Likewise, satellite handover on either of the UDL parts is handled on the VC level by appropriate signaling between involved mobile user(s) and gateway station; thus it works transparently over any given dynamic VPC topology and does not affect the setup of the latter.

### 3.4.4 UDL Segment: Air Interface Access Network

For the UDL routing part, an extension of the *virtual connection tree (VCT)* concept – originally proposed by Acampora and Naghshineh [AN94] for terrestrial cellular ATM networks – is considered. In the original concept one root switch manages a fixed tree serving a certain mobile service area (MSA), Fig. 3.10(a). The VCT takes care of connection setup and VCI reservation and ensures fast and transparent handovers of roaming users in the MSA. To apply the VCT architecture and methodology to the LEO satellite environment, a dynamic satellite cluster headed by a *master satellite* is defined, Fig. 3.10(b). This satellite cluster builds a time-dependent VCT with spotbeams, ensuring that fast and transparent spotbeam and satellite handovers can be performed [Lut95]. The leaves of this tree may for instance contain the spotbeams of all satellites that are above a minimum elevation angle for a certain user. In this way, a handover of virtual connections can be easily accomplished without dedicated signaling effort. For each user (group), or for each area covered by a VCT, the according VCT must be updated according to the satellite movement.



**Figure 3.11:** Set of 3 link-disjoint VPCs between first satellite  $S_i$  and last satellite  $S_j$ .  $VP_0$  indicates that in general more than one VP is defined on a physical ISL, and it represents an element of any VPC (not depicted) between  $S_k$  and  $S_l$ , where  $(k, l) \neq (i, j)$ .

### 3.4.5 ISL Segment: Meshed Space Backbone Network

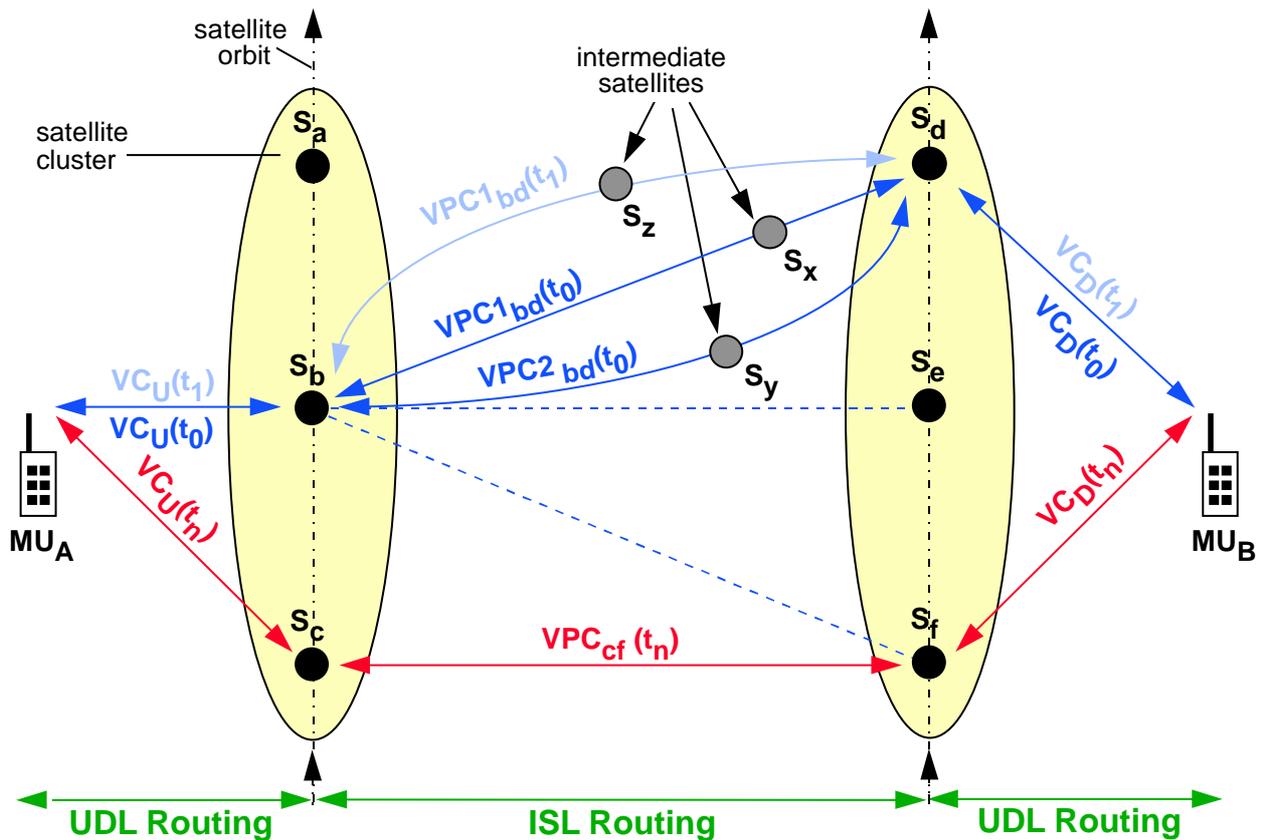
Because of the time-varying network topology, ISL routing is especially complex:

- A route between a pair of serving satellites may not be optimum or even possible during the whole connection. Rather, route changes may be necessary, resulting in problems connected with flow control and critical delay variations during path switching.
- Due to time-limited satellite visibility, the first and last satellites may change during a connection (satellite handover). This requires additional route changes within the ISL network.

The multiple ISL topology dynamics is due to the fact that pairs of satellites in adjacent orbits have a varying distance over time or can even lose sight of each other. Topological or connectivity changes may demand rerouting of calls during their lifetime, i.e. on-line switching between subsequent different paths. This highly challenging task will be tackled in depth in Section 4.3, where we develop an off-line dynamic ISL routing concept which is essentially based on the predefinition of paths between end nodes of a connection, and inherently takes care of the time continuity by controlled path (VPC) handover.

A possible route between a first and a last satellite can be modeled as a VPC. Since most satellite pairs are not directly connected by an ISL, intermediate satellites will often be needed to constitute a complete VPC route, as illustrated in Fig. 3.11. Every section of this route, i.e. every ISL between adjacent satellites, in general contains a number of VPs (resulting from the considered and all other “crossing” VPCs), all sections using the complete set of VP identifiers (VPs) independently. The routing function is incorporated in the respective dynamic VPI translation tables. By assigning VPs to the ISLs and setting up the corresponding VPI translation tables in the satellite switches, a *virtual topology* is defined on top of the physical one.

The basic idea behind this approach is that end-to-end VCCs sharing the same first and last satellites at arbitrary time can be aggregated into one common VPC across the ISL subnetwork. Every transit satellite provides – on the basis of locally available switching tables – pure VP switching



**Figure 3.12:** Virtual-connection-based routing concept in satellite systems with ISLs. The continuous operation of an existing end user connection is illustrated for the case of VPC handover between the same first/last satellite pair from steps  $t_0$  to  $t_1$ .

functionality between every pair of ISL ports, and thus the whole space segment becomes a pure fast operating cross-connect network. Avoiding any switching on VC level turns out to be especially favorable in the case of many simultaneous low bit rate connections on a trunk line; this situation is very likely to be the dominating one in systems providing primarily voice services.

Figure 3.12 illustrates the overall concept comprising the integration of UDL and ISL segment. It specifically addresses the problem to guarantee a continuous connection in the case of an ISL subnetwork internal path handover. It has been shown [Wer97] that the operation of both basic types of handover – UDL satellite handover and ISL VPC/path handover – can be managed with the proposed concept but it requires sophisticated optimization algorithms and techniques to keep the possible impairments to a minimum.

### 3.5 Feasibility and Implementation Aspects

The proposed ATM networking concept has not been implemented so far in hardware or software; however, we want to conclude the presentation of the concept with pointing at some issues relevant for a potential implementation.

**Conceptual Limitations** With respect to the conceptual limitations, the major concern must be about the maximum number of VPs required on the worst-case link, taking into account all ISLs over the whole constellation period  $T$ . The ATM cell header contains 12 bits in the VPI field (at NNI), thus allowing a maximum of  $2^{12} = 4096$  VPs on a single ISL per step. The number of instantaneously required VPs per link is determined by the number of VPCs using this link. General figures for that cannot be calculated, since they depend not only the network topology but also on the routing algorithm.

However, assuming for simplicity that only one VPC is established per terminating satellite satellite pair, we find that the number of simultaneous VPCs in the *complete network* equals the number of terminating satellite pairs,  $N(N-1)/2 = 2145$  for Iridium,  $N = 66$  (2556 for M-Star,  $N = 72$ ). If we further assume that due to the distributed connectivity of the considered ISL topologies any routing strategy should “distribute” the paths quite evenly in the network consisting of 106 (144) instantaneous links in total, and that a path consists of roughly 5 links in the average, we can estimate that the corresponding number of VPs per link is in this optimal case  $2145 \cdot 5/106 \approx 100$  ( $2556 \cdot 5/14 \approx 90$ ). In the real scenario, i.e. with non-ideal distribution of the routes, we can of course expect a worst case value which is maybe a factor 2–3 higher, but then still far below the allowed maximum.

In fact, counting VPs per link in computer simulations for shortest path routing yields a maximum of roughly 400 (230) VPs. This perfectly confirms the above estimation. The available number of VPIs is roughly by a factor of 10 higher. Note in this context, however, that short-term “overlapping” of successive VPCs during controlled VPC handover (like in the alignment concept according to [EEFK95]) will also increase the worst case VPI requirements.

**Implementation Aspects** When it comes to failure of nodes or links in the ISL network, the issue of providing backup paths becomes relevant. Considering only the ATM layer in this concept, one can in principle think of three different backup methods [Edm96]:

1. *rerouting*, where backup paths are (regularly) calculated/updated by a central management entity during system operation,
2. *protection switching*, where backup paths have been prepared and corresponding capacities have been reserved in advance by a central management entity, and the activation in case of failure is activated during system operation in decentralized manner, and
3. *selfhealing*, where backup paths have been prepared in advance *without* capacity reservation by a central management entity, and the activation in case of failure is activated during system operation in decentralized manner.

Inherently in our strictly VPC-based concept, backup VPCs should be prepared off-line in advance, because regularly updating paths for rerouting is deemed too demanding in terms of signaling, and on-demand rerouting in the global network too critical in terms of reaction times. The remaining question, if capacities should be reserved in advance or not, cannot be answered in general without knowing details about the particular constellation and its connectivity, the implemented ISL technology (RF or optical links), and the typical capacities that would be required and are dependent on the global service and traffic scenario. However, we can state in general that with (i) higher connectivity, (ii) optical ISLs and (iii) less demanding backup capacities protection switching (with reservation) tends to become more attractive and realistic than selfhealing.

Signaling for any purpose is of course a major issue during system operation, and there has been some debate about ATM's signaling complexity in recent years. The detailed discussion of such issues is beyond the scope of this thesis, but we can state that signaling effort and overhead should certainly be kept low especially in a global network. However, we claim that the proposed concept is so fundamental that adaptations to other than ATM *implementations* should be possible without major changes. In particular, we have in mind multiprotocol label switching (MPLS) as evolving technology which is commonly said to provide similar functionality as ATM at lower cost, including in particular signaling complexity.

# Chapter 4

## Routing

*I will find a way. Or I will make one.*

— SENECA (about 4 B.C. – 65)

For every connection between two end users, or every information entity (for instance, a packet) transferred between them, a route through the communication network must be chosen. Routing is a central task in large networks and influences network performance as well as quality of service (e.g. message delay).

Although the classical separation of communication networks into the circuit-switched telephone world and the packet-switched computer world (Internet) is increasingly being superseded by integration or convergence, it is especially in routing issues important and helpful to keep the respective connection-oriented and connectionless heritage in mind.

A good reference for virtually all aspects of routing in circuit-switched networks are the books by Ash [Ash97] and Girard [Gir90]; Ross [Ros95] provides a loss network modeling approach to dynamic routing in ATM networks, fusing (i) dynamic routing as known from single-service (telephone) networks and (ii) the particular ATM environment with statistical multiplexing of heterogeneous service classes. Routing in packet-switched networks is extensively discussed in the textbooks by Tanenbaum [Tan96], Schwartz [Sch87], and Bertsekas and Gallager [BG87a].

The following discussion of routing in LEO/MEO satellite networks is largely based on [Wer96], [Wer97], [WM97] and [WMMH98] concerning the ISL routing part, and on [BW94b, EVW96] and [WBL95a, WBL95b] for the UDL routing part. Connection-oriented (CO) network operation is assumed throughout; consequently, one possible routing implementation *could* be ATM-based, as outlined in Chapter 3. Besides CO operation mode, we furthermore assume (i) point-to-point communication, so there is no broadcast or multicast traffic, and (ii), for bidirectional connections, the utilization of the same physical routes in both directions.

### 4.1 Routing in LEO/MEO Satellite Networks

In satellite systems, the routing task has especially become important with the advent of LEO (and potentially MEO) satellite constellations employing ISLs, or in other words, *real* space-based

communication *networks*. Consequently, in the routing context only ISL-based systems will be considered. The movement of satellites in combination with inhomogeneous user distribution and time-varying user activity leads to a particularly time-varying source/destination traffic load on the satellites. The variance of this traffic is transferred into the ISL sub-network in a smoothed form, because the intersatellite links carry also a lot of transit traffic.

For the first generation of such systems, namely narrowband systems providing mobile telephony as key service, the operation mode will naturally be connection-oriented. Despite the fact that ATM is the designated basis for integrated-services broadband systems, we will make use of ATM terminology also in the context of narrowband (single-service) satellite systems for two good reasons: (i) the powerful concepts of ATM networking, which can be used in generalized manner for all connection-oriented communications, and (ii) the obvious trend towards service integration also in satellite communications.

### 4.1.1 ISL Topology Dynamics

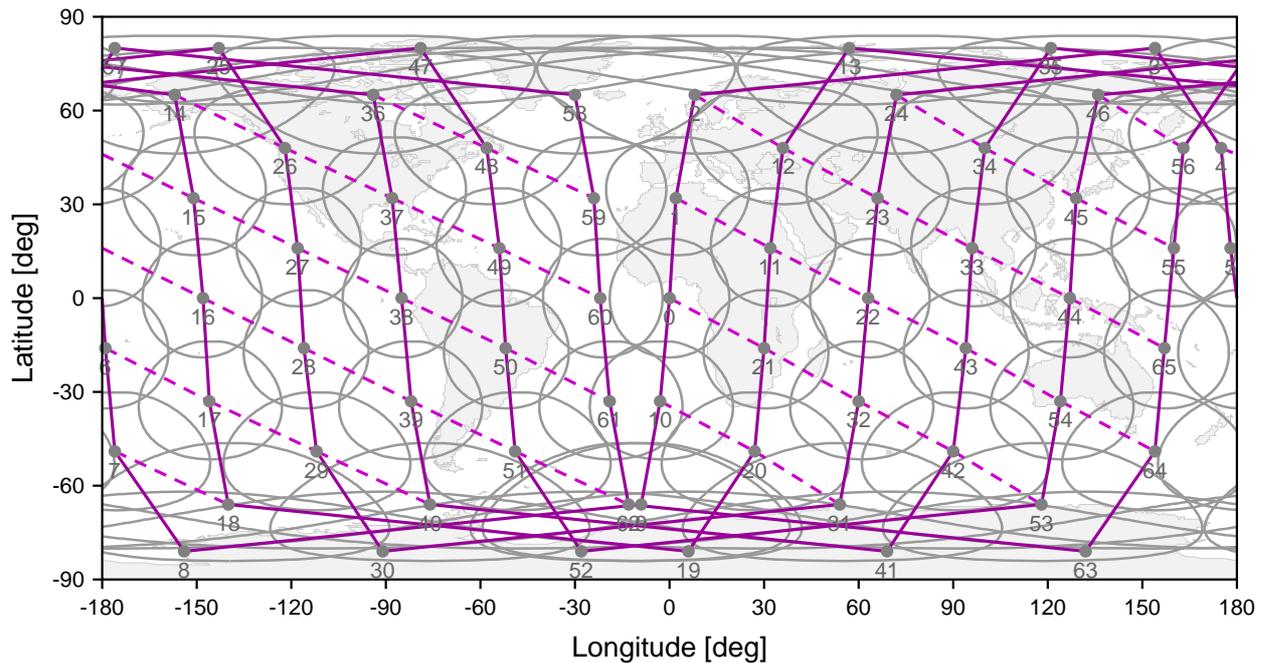
Throughout this chapter, Iridium will be used as an example, as it is the first operational LEO satellite communication system to employ ISLs, and allows some major challenges to be highlighted for connection-oriented routing.<sup>1</sup> Recalling Section 2.4, polar or near-polar constellations exhibit a *seam* between two counter-rotating orbits, effectively dividing the constellation into two hemispheres of co-rotating orbits.

At a given instant, the ISL topology of the system comprises a unique set of *intra-orbit* and *inter-orbit* ISLs as illustrated in Fig. 4.1. Intra-orbit ISLs have a fixed distance and antenna pointing, whereas inter-orbit ones are subject to continuous variations of both [WJLB95, Wer95]. For counter-rotating orbits, moreover, a frequent switching of ISLs would be necessary, a fact that has led to the avoidance of links across the seam for Iridium, see Fig. 4.1. Besides the *continuous* distance changes on co-rotating inter-orbit ISLs there is also a *discrete-time* contribution to the ISL topology dynamics: inter-orbit ISLs are deactivated in polar regions, meaning on/off switching of links. As a result, tailor-made routing strategies are required to handle the rerouting of connections.

ISLs are deactivated (reactivated) whenever one or both satellites are above (below) a given latitude threshold. Precise technical details on this issue are hardly to find in the respective literature or in system filings. One hint is given in a Motorola patent by Rahnema [Rah95] claiming that an *Iridium-like* constellation can keep inter-orbit links active up to a latitude of 68°. For our investigations, we have assumed that a reasonable solution would be to switch ISLs in such a manner that the total number of inter-orbit ISLs in the network remains constant, thus achieving a constant degree of network connectivity. A procedure to achieve this is to exactly synchronize the deactivation of “polar-bound” ISLs and the reactivation of the corresponding “equator-bound” ones. For instance, looking at the right hemisphere of ascending (north-bound) satellites in Fig. 4.1, exactly when ISL 24–34 will be deactivated soon after the displayed snapshot, ISL 31–41 should be activated and “replace” the former as fourth inter-orbit ISL connecting the two ascending co-rotating orbit planes. Having always four inter-orbit ISLs being activated at a time between any pair of co-rotating orbit planes in either ascending or descending direction (or eight considering the whole orbit planes), and assuming latitude symmetry for the activation and deactivation thresholds,

---

<sup>1</sup>It should be explicitly noted in this context that the routing concept presented here is indeed developed on top of Iridium-like dynamic ISL constellations, but it is *not* the Iridium proprietary concept.



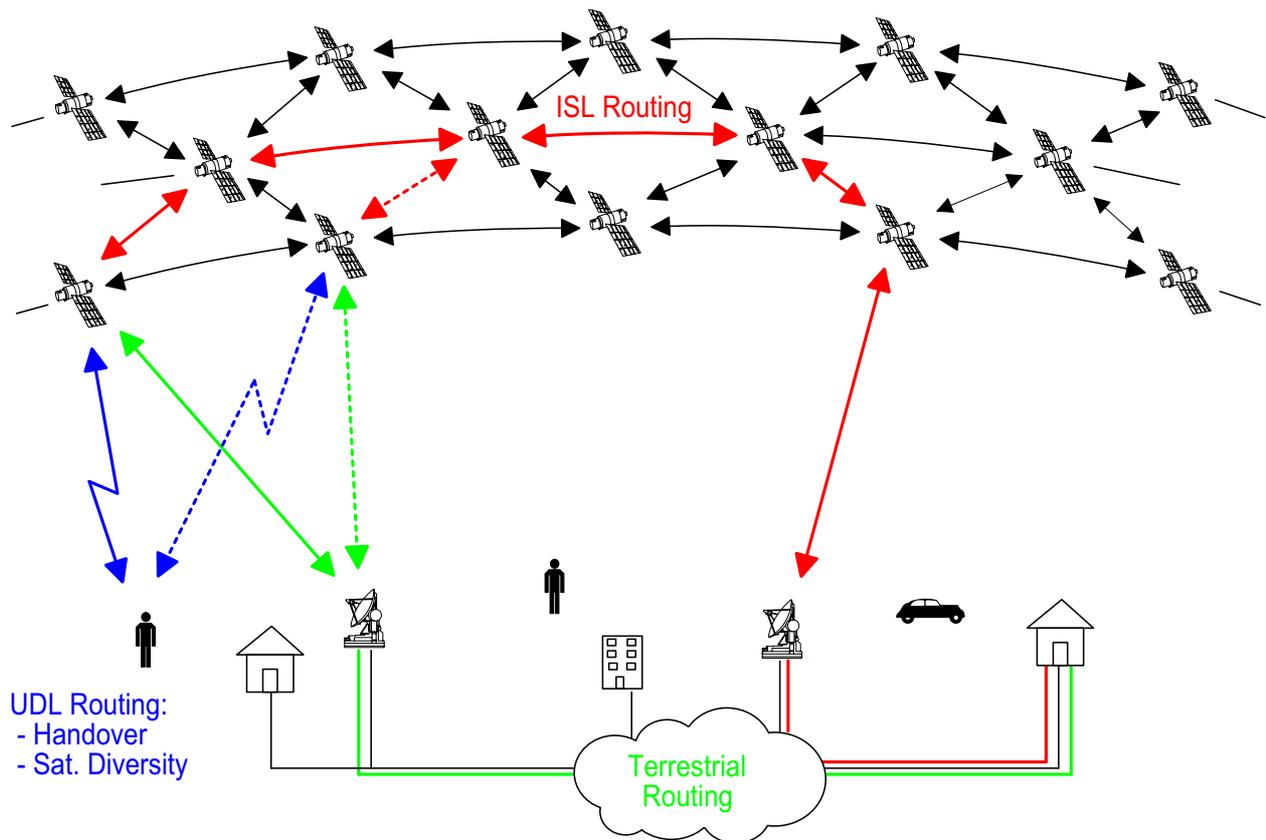
**Figure 4.1:** Snapshot of Iridium footprints and ISL topology; solid lines denote the intra-orbit and dashed lines denote the inter-orbit ISLs. The seam encounters at a longitude of roughly  $-10^\circ$  resp.  $170^\circ$ .

one effectively ends up with a latitude threshold of approximately  $65^\circ$  for the arithmetic average of the latitudes of the two satellites using the respective ISL. With this approach, we have a total of 106 active (bidirectional) ISLs at any instant in our reference Iridium topology, including 66 intra-orbit and 40 inter-orbit ones.

## 4.1.2 Network Segmentation

Although the focus of the thesis is clearly on networking issues for the space-based ISL backbone, it is especially in the routing context necessary to consider it as *one segment* of a global network providing end-to-end communications. As illustrated in Fig. 4.2, communication between a mobile or fixed user of the satellite system and a counterpart in a terrestrial network of any type (PSTN, ISDN, B-ISDN, Internet, GSM, etc.) usually involves three routing segments, namely

- *UDL routing* which essentially covers the selection of satellites serving users and gateway stations on the air interface up- and downlinks, taking into account the relative motion of satellites and ground locations;
- *ISL routing* dealing with the setup of paths between any pair of satellites terminating the space-segment part of the connection, including potential rerouting due to dynamics of the ISL topology itself or due to handover between serving satellites;
- *terrestrial routing* providing paths between the interfacing gateway station (GW) and the end user in the terrestrial network.



**Figure 4.2:** Routing segments for end-to-end communication in ISL-based LEO/MEO satellite networks.

The terrestrial routing can be assumed to be independent from satellite (UDL and ISL) routing, if the GW provides the proper routing protocol interface on the ISO/OSI network layer, and is then the standard routing for the respective type of network; terrestrial network routing will not be further considered in this thesis.

As indicated in Fig. 4.2, under UDL routing we subsume particularly the “routing features” of the more general *satellite handover* and *satellite diversity* concepts or procedures, which have already been introduced in Section 2.5. UDL and ISL routing are not completely decoupled as, for instance, satellite handover actions on the UDL part inevitably provoke some rerouting on the ISL part as well. However, mechanisms employed for (re)routing in the ISL subnetwork can be designed such that they are independent of the cause requiring the (re)routing action; this will be elaborated in detail in Section 4.3. It is nevertheless important to address some relevant UDL routing issues to a certain level of detail before, mainly in order to get a clear impression of the possible qualitative and quantitative impacts on ISL (re)routing.

## 4.2 UDL Routing

As indicated in the previous section, we are not going to elaborate on the manifold aspects of satellite handover and satellite diversity in detail; rather it is sufficient in the context of this thesis to look at some particularly interesting routing aspects of both, mainly from the following two perspectives:

1. Routing options on the UDL can be used to effectively improve the overall availability and QoS of end-to-end connections, which is *per se* an important issue knowing that the UDL is the capacity bottleneck and by far the dominating segment in terms of call dropping probability, especially in a mobile communications scenario or in general for unfavorable user environments.
2. The concrete impact of UDL routing on the operation of ISL routing should be identified more precisely, first of all in qualitative manner, and then as far as possible also quantified.

Having outlined the relationship between UDL and ISL routing so far, we will now first “return” to studying non-ISL constellations. More precisely, we consider narrowband LEO/MEO systems mainly aiming at mobile communications, for the most satellite handover/diversity research over the past years has been initiated by major representatives of them, and as they set the highest demand for both handover and diversity operation due to the critical mobile channel characteristics.

### 4.2.1 Satellite Handover

Performing satellite handovers raises mainly two interesting issues from the routing perspective, also having the combined UDL/ISL routing problem in mind:

1. The exact procedure of performing the handover, which is mainly reflected in the related signaling. This gives insight not only in potential QoS degradation due to short interruptions on the UDL alone, but also sets the requirements for the related ISL rerouting operations, mainly in terms of which satellite will play the primary role and what the time limitations for a proper reaction are.
2. The typical time between handover actions, or the frequency of handovers, is an interesting performance measure. As one of the critical networking procedures in a highly mobile satellite and user environment, satellite handover does not only increase the call dropping probability due to potential failures during the handover operation. A high frequency of handovers, i.e. of UDL routing operations, also imposes severe challenges on the related ISL routing to provide continuous end-to-end connections.

In the following we will look at both issues, summarizing relevant results from earlier in-depth investigations [BW94b, WBL95a, Eis96, EVW96].

#### 4.2.1.1 Handover Procedure

Concerning the satellite handover procedure, one can basically distinguish between (i) backward handover and (ii) forward handover. Whereas the first method is very similar to handover in GSM, we found that the second method may be especially suited for satellite networks.

Figure 4.3 illustrates the different procedures. For a *backward handover*, the mobile station (MS) submits the handover request via the old satellite, and after receiving a handover command from the network it tunes to the new satellite to acquire a new traffic channel. Between the reception of

the handover command from the old satellite and the handover complete message sent to the new satellite a *handover break* occurs with duration

$$T_{\text{break,b}} \approx 3T_{\text{prop}} + T_{\text{proc,b}}, \quad (4.1)$$

where  $T_{\text{prop}}$  is the propagation delay between MS and GW, and  $T_{\text{proc,b}}$  represents the accumulated handover processing time (approx. 150 ms for GSM), including tuning and synchronization to the new satellite. The break for a backward handover may last up to 200 ms in typical LEO constellations and up to 500 ms for representative MEO constellations [Eis96, EVW96]. Moreover, if the signal power falls below the link margin during the break, the connection is dropped.

This problem can be avoided by using a *forward handover* procedure. Here, the handover request is directly sent to the new satellite, while maintaining the old traffic channel until the new channel has been acquired. Switching to the new satellite, MS sends an assign complete message. Upon reception of this message the GW switches to the new satellite as well. Thus, the handover break is reduced to the period between sending and receiving/processing the assign complete message, where the processing time is considerably shorter than with backward handover:

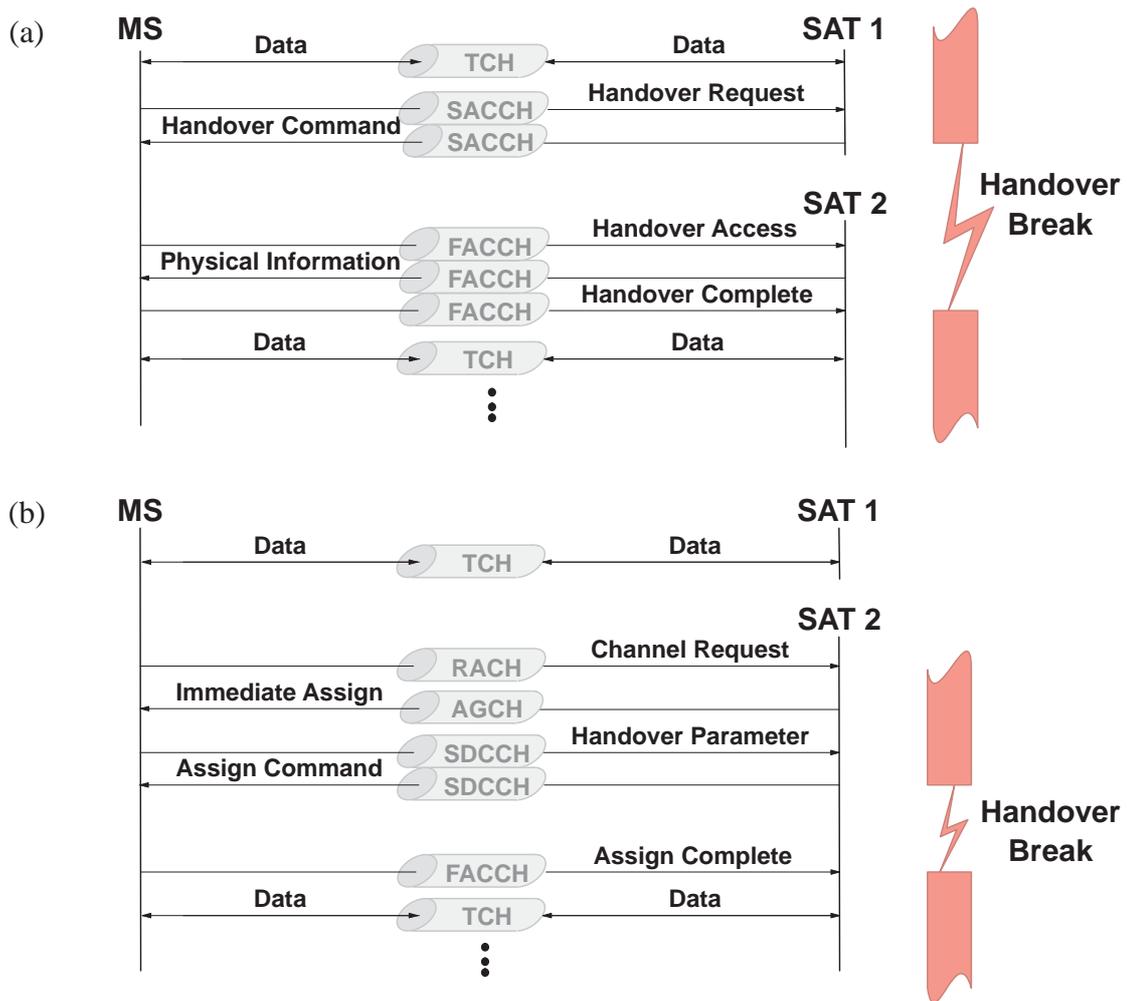
$$T_{\text{break,f}} \approx T_{\text{prop}} + T_{\text{proc,f}}. \quad (4.2)$$

Forward handover breaks in representative LEO systems typically last around 60 ms [Eis96, EVW96].

#### 4.2.1.2 Handover Strategies

We now restrict ourselves to considering deterministic handover due to satellite movement; i.e., shadowing or other events as potential stochastic handover causes are not taken into account. With this assumption, the typical time between satellite handovers crucially depends on the size of the footprints, and this, in turn, basically on the orbit altitude. Therefore we exemplarily study the Iridium and LEONET systems as representatives for the group of LEO and MEO constellations, respectively. The relevant parameters for our handover investigations are listed in Table 4.1. For an illustrative comparison of the footprint sizes one may refer to the coverage snapshots in Figs. A.7 and A.8 in Appendix A.

Clearly, freedom of choice for the selection of the instantaneously serving satellite, and thus for a potential “optimization” of inter-satellite handover, depends directly on some degree of multiple satellite visibility. If multiple satellite visibility is given, the instantaneous elevation angles, under which the alternative satellites are seen from a user position on ground, may serve as a basic criterion for the selection as the quality of the channel (and thus also the expected QoS for the user) is closely linked to the elevation angle; this will be discussed in more detail in the context of satellite diversity to follow (Section 4.2.2). Another important factor is of course the maximum period of time that a satellite is visible above the specified minimum elevation angle. These two factors are addressed in Figs. 4.4 and 4.5, respectively. Figure 4.4 shows a satellite elevation diagram for a user at 31° north latitude, using the LEONET constellation as an example. At each time, several alternative satellites are visible above the minimum elevation. The magnified part of the picture will be commented later. Figure 4.5 displays overhead passing times, i.e. the maximum possible visibility periods, in dependency of the orbit altitude and of the minimum elevation angle. From the graphs one can imagine that a typical telephone call, for instance, may be finished without the need for an in-call satellite handover in MEO systems, whereas this is much less likely for LEO constellations.



**Figure 4.3:** Simplified signaling flow, corresponding GSM-like signaling channels, and handover break for (a) backward handover and (b) forward handover [EVW96]

**Table 4.1:** Constellation and simulation parameters for satellite handover evaluation.

	Iridium	LEONET
Orbit classification	LEO	MEO
Orbit altitude $h$	780 km	6390 km
Orbit period $T$	100 min	3 h 59 min 21 s
Orbit inclination $i$	86.4°	54°
No. of satellites in the constellation $N$	66	15
No. of orbit planes $P$	6	3
No. of satellites per orbit plane $N_S$	11	5
Phasing factor $F$	not appl.	1
Minimum elevation angle $\varepsilon_{MT}$	8.2°	20°
Footprint diameter	4435 km	9350 km

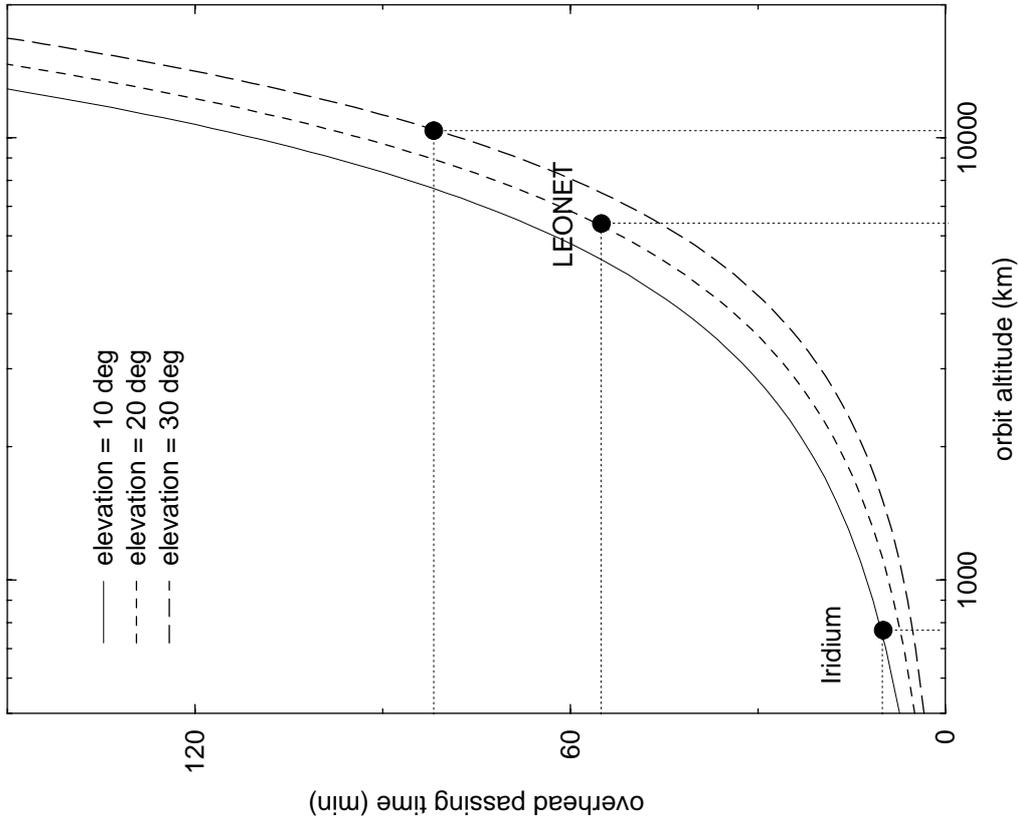


Figure 4.5: Overhead passing times for LEO/MEO satellites above a specified minimum elevation angle.

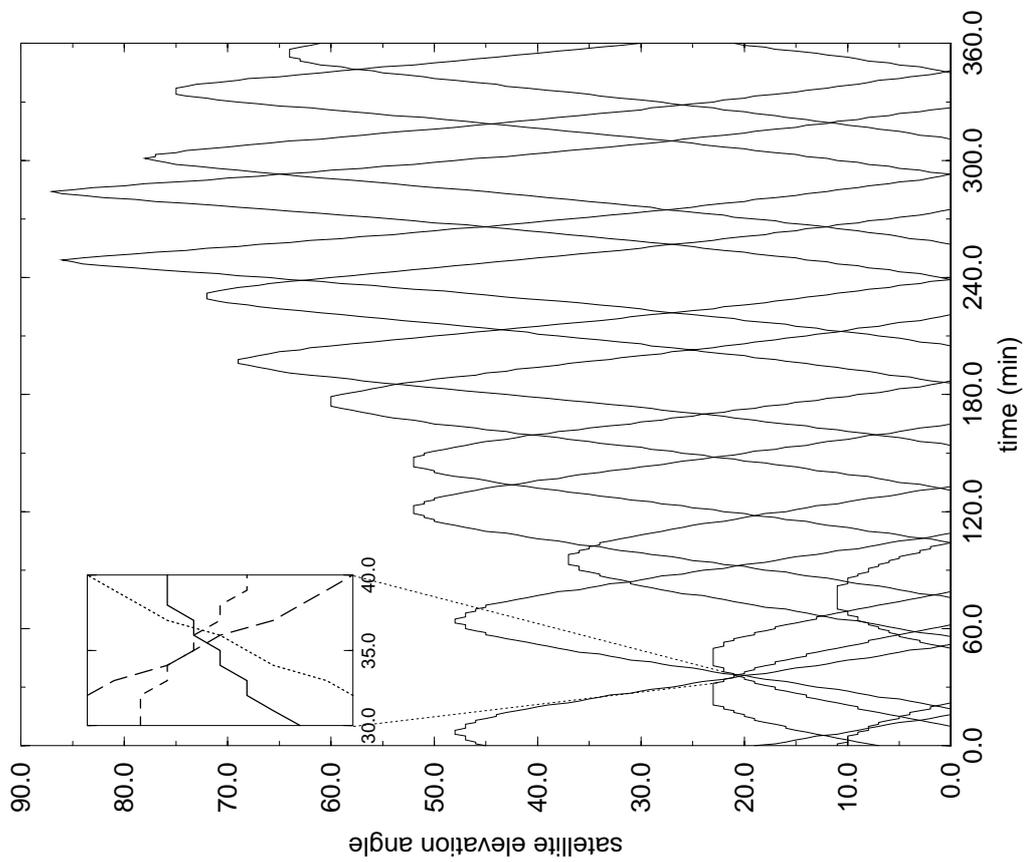


Figure 4.4: Elevation of LEONET satellites over time for a user at 31° north latitude.

Motivated by these considerations, three possible strategies for geometric satellite handover have been proposed [BW94b]:

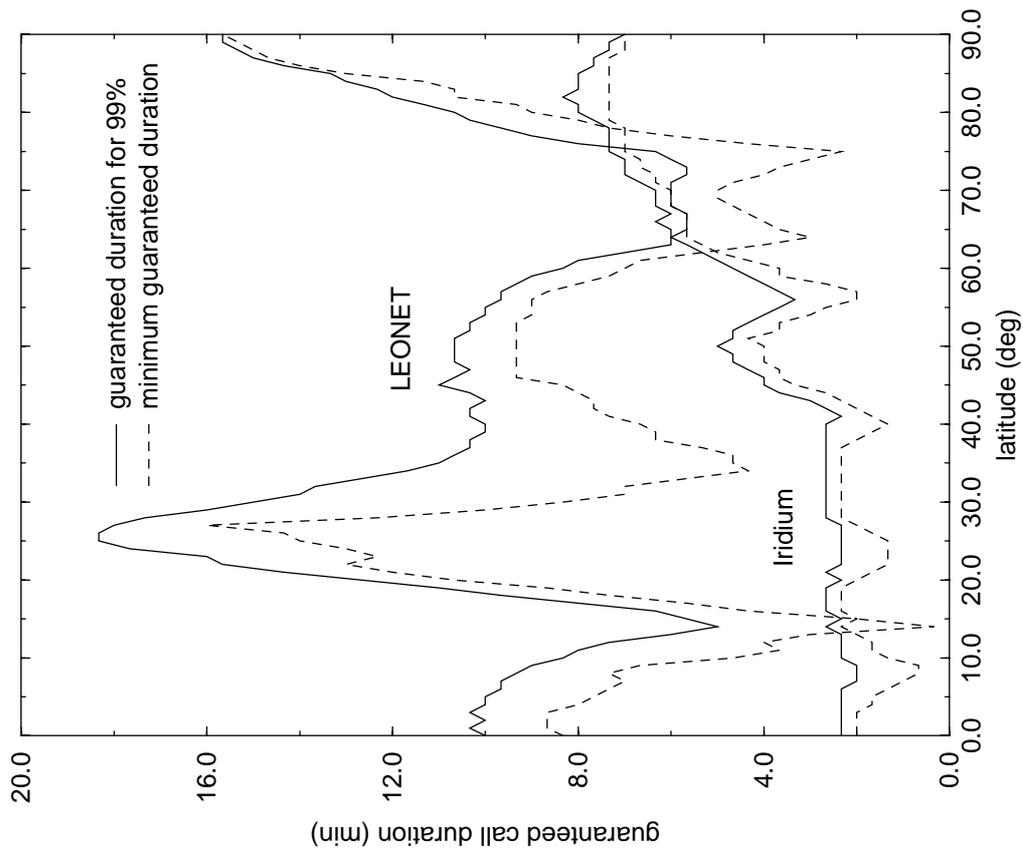
1. *Maximize the instantaneous elevation angle.* According to this strategy always the satellite with the highest instantaneous elevation angle  $\varepsilon$  is selected and handovers will be performed accordingly.
2. *Minimize the handover rate.* This must be done while maintaining the specified minimum elevation angle  $\varepsilon_{\min}$ , i.e. always that satellite is selected which will stay for the longest time above the minimum elevation angle. A satellite handover is always triggered when the currently serving satellite drops below  $\varepsilon_{\min}$ .
3. *Maximize visibility without handover.* At connection setup, always the satellite with the longest visibility time above the specified minimum elevation angle  $\varepsilon_{\min}$  is selected. With this satellite dropping below  $\varepsilon_{\min}$  the connection is terminated.

Calling option 3 a “handover strategy” sounds somewhat weird at first sight; anyway, it is a clear rule related to handover, and it draws inspiration from the observation of quite long footprint passing times encountered especially in MEO constellations, cf. Fig. 4.5.

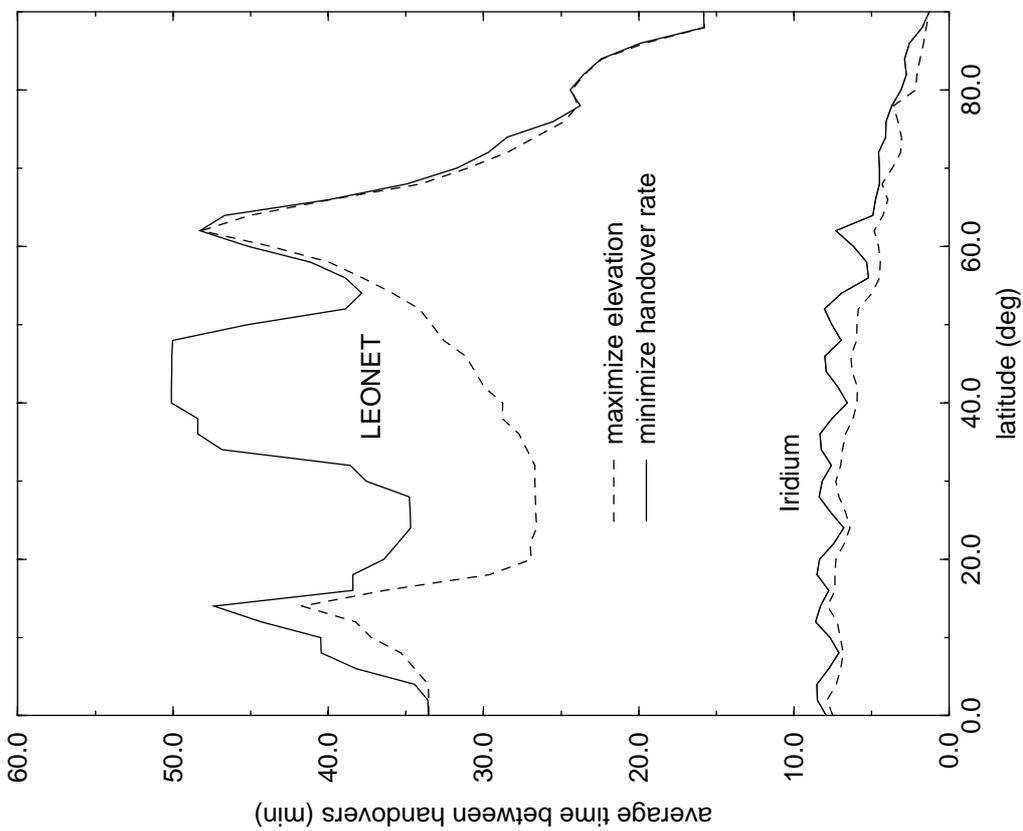
With respect to the first strategy, the magnified part of Fig. 4.4 has already indicated one potential problem: a rapid change of instantaneously highest satellites can occur at exceptional latitudes and instants of time. In such cases several handovers would be experienced in a very short time period, if no counter-measure is foreseen, as for instance an appropriate timer.

Figure 4.6 provides a performance comparison of strategies 1 and 2 with respect to handover frequency. As one expects, with strategy 2 the average interval between two successive handover actions is longer, of course at the price of a lower average elevation angle; the differences between the two strategies in terms of both elevation statistics and handover frequency are, however, not so pronounced that a general preference for one or the other strategy can be derived. These results have been independently confirmed by Krewel and Maral [KM00], focusing on the impact of handover strategies on the QoS in diversity based satellite constellations. Referring to a *nearest satellite* and a *longest remaining visibility time* procedure, they report “satellite holding times”, guaranteed for 90% of time, of 6 minutes for the former and 9 minutes for the latter, all values for Iridium.

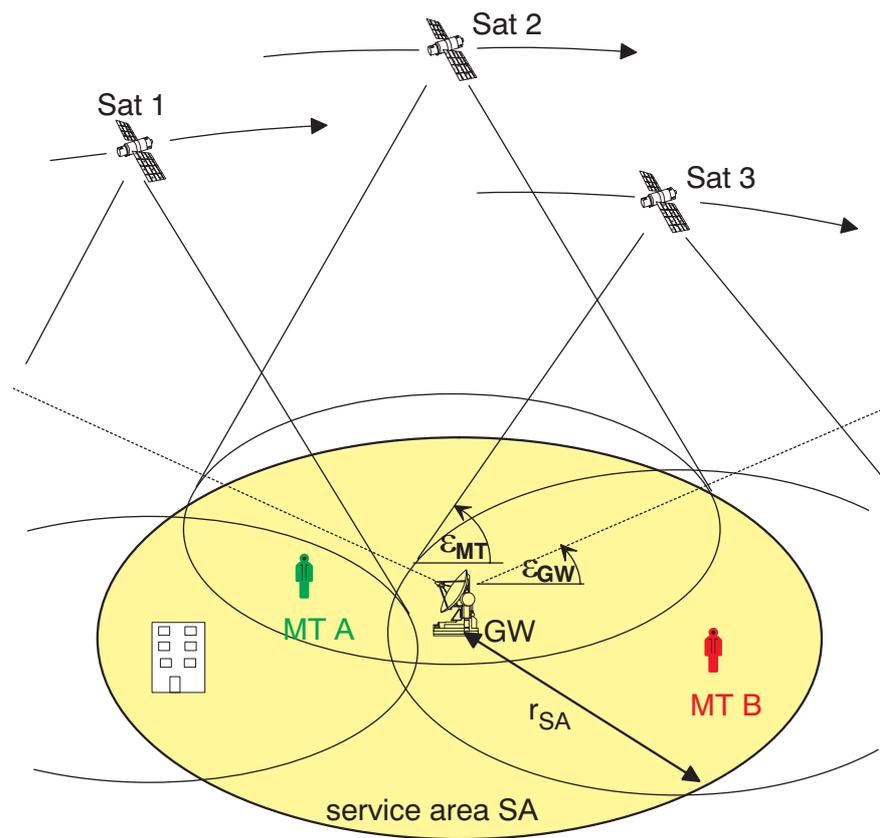
Finally, Fig. 4.6 shows the minimum guaranteed call duration (and its 99% percentile) for handover strategy 3, displaying pronounced peaks and fades over the latitude; the latter are partly due to discretization errors in the (global) simulation (deep LEONET fade at approximately  $14^\circ$ ), but in their main tendency they reflect coverage peculiarities of the respective constellation. As a conclusion from this figure, strategy 3 is only acceptable for MEO constellations assuming typical telephone call holding times, which confirms our earlier expectations.



**Figure 4.7:** Minimum guaranteed call duration and guaranteed call duration in 99% of the time without handover (handover strategy 3).



**Figure 4.6:** Average time between successive handovers, comparing the two handover strategies 1 and 2.



**Figure 4.8:** Satellite diversity scenario; for the connection between user A and GW, dual satellite diversity (via Sat 1 and Sat 2) may be exploited, whereas only one satellite (Sat 3) is jointly visible to user B and GW.

## 4.2.2 Satellite Diversity

The fundamental principle of satellite diversity and the motivation for its application in constellation networks has been outlined in Section 2.5.3. In this section we summarize earlier research work [WBL95a, WBL95b] covering a qualitative analysis and some representative numerical results for LEO/MEO satellite systems with satellite diversity implementation. The focus is on the diversity gain, i.e. improvement of service availability, that can be realized for typical user environments.

### 4.2.2.1 Scenario and User Environment

Without a restriction of universality, we exclusively consider the case of (mobile) users communicating with a gateway for the sake of a simplified scenario: one gateway GW is responsible for all communication needs of mobile users in its so-called *service area* SA with radius  $r_{SA}$ , as illustrated in Fig. 4.8. The users are represented by their mobile terminals (MT).

Using a short-hand terminology for the following discussion, we distinguish between *visibility* and *availability* of a satellite:

- A satellite is called *visible* from a position on ground if it appears under an elevation angle  $\epsilon \geq \epsilon_{\min}$  with respect to this position; thus *visibility* refers to the sheer geometrical situation and does not necessarily include clear line-of-sight.

- A satellite is called *available* for an MT or GW, if a communication link can be established, which is the case for clear line-of-sight, i.e. no shadowing. Evidently, visibility is a precondition for *availability*, and a shadowed satellite with  $\varepsilon \geq \varepsilon_{\min}$  is visible but not available.

Extending these consideration to the case of  $n$ -satellite diversity, it follows that *service availability* is given if at least one of the  $n$  satellites is visible and available to both user and GW at the same time. Usually, one can assume that a GW is positioned such that shadowing does not encounter and permanent availability is guaranteed for the whole visibility period; thus the service availability is given if only one of the user links is available.

The on/off channel behavior, which is implicitly assumed with the above definitions, is motivated by results from several measurements campaigns [LCD<sup>+</sup>91, JL94, JSBL95]. To evaluate the diversity performance of LEO/MEO satellite constellations on system level in decent time we have adopted this “simple” model, and we will shortly describe it in the context of user environment influences and permanently changing elevation angles for such systems, before entering the presentation of a few representative numerical results. Some more details of channel modeling are discussed in Section 6.1.1, where the focus is on numerical studies of satellite diversity simulations based on a more sophisticated channel model.

Satellite communications with land mobile or personal terminals suffer from strong variations of the received signal power due to signal shadowing and multipath fading.

- Shadowing of the satellite signal is caused by obstacles in the propagation path, such as buildings, bridges, trees, etc. The percentage of shadowed areas on the ground, as well as their geometric structure depends critically on the type of environment, e.g. urban, suburban, rural, etc. For low satellite elevation the shadowed areas are larger than for high elevation. Especially for streets in urban and suburban areas, the percentage of signal shadowing also depends on the azimuth angle of the satellite.
- Multipath fading occurs because the satellite signal is received not only via the direct path but also after being reflected from objects in the surroundings.

Measurements at L-band with the MARECS satellite [LCD<sup>+</sup>91] have shown that signal shadowing dominates the propagation behavior. As a first approximation, the communication channel can be considered to be interrupted (bad channel state) if there is no line-of-sight to the satellite. Vice versa, the channel can be approximated as being ideal (good channel state) if a line-of-sight connection is present.

Due to the movement of the non-geostationary satellites, the geometrical pattern of shadowed areas is changing with time. Similarly, the movement of a mobile/personal user translates the geometrical pattern of shadowed areas into a time-series of good and bad channel states. In [LCD<sup>+</sup>91] it has been shown that the time-varying channel behavior can be approximated by a two-state Gilbert-Elliott model with the mean durations in the good and bad state, respectively, depending on the type of environment, satellite elevation, and mobile user velocity. For our purpose, a separate and independent channel model was considered for each of the satellites visible to an investigated user position. Possible dependencies on the azimuth angle have been neglected. For different elevation angles, statistical values for the time-share of shadowing,  $A$ , and for the mean durations in the good and bad channel state, have been taken from the measurement results. In order to include the particular dependency on the satellite elevation, the channel parameters have been fitted over the elevation angle. As limiting cases, for 90° elevation  $A = 0$  was assumed and  $A = 1$  was used for 0° elevation.

**Table 4.2:** Constellation and simulation parameters for satellite diversity evaluation.

	Globalstar	ICO
Orbit classification	LEO	MEO
Orbit altitude $h$	1414 km	10390 km
Orbit period $T$	113.8 min	5 h 59 min 1 s
Orbit inclination $i$	52°	45°
No. of satellites in the constellation $N$	48	10
No. of orbit planes $P$	8	2
No. of satellites per orbit plane $N_S$	6	5
Phasing factor $F$	1	0
Minimum elevation angle $\varepsilon_{GW}$	5°	5°
Minimum elevation angle $\varepsilon_{MT}$	20°	20°
Footprint diameter	4760 km	10850 km
Service area radius $r_{SA}$	970 km	2600 km

#### 4.2.2.2 Evaluation of Satellite Diversity Gain

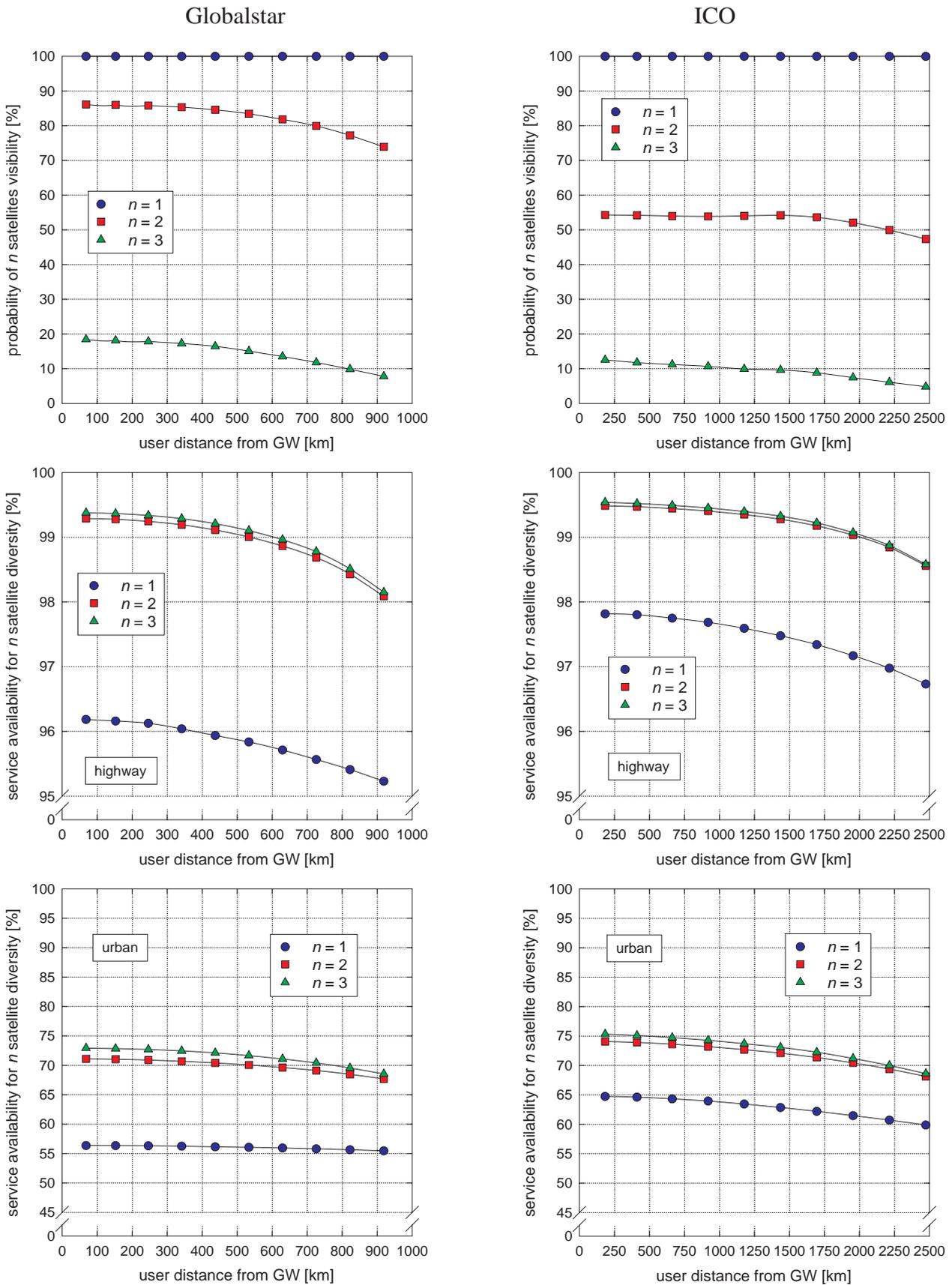
We presuppose that the implementation of satellite diversity reception is given by means of separate parallel receivers within the MT and GW, respectively. Based on that, we specify the following baseline for operating satellite diversity (cf. Fig. 4.8): A GW selects up to three satellites visible with an elevation angle  $\varepsilon \geq \varepsilon_{GW}$ . Then all MTs which are in the coverage area of at least one of these satellites, i.e.  $\varepsilon \geq \varepsilon_{MT}$ , communicate with the GW via *all* mutually visible satellites.

Extensive simulations have been performed in [WBL95a, WBL95b] for a gateway at (100°W, 40°N) and 400 user positions equally distributed in a circle around the gateway. Finally, the radius of this circular service area is chosen for each of the two considered systems, Globalstar and ICO, so as to guarantee a fair comparison between a LEO and a MEO system. Roughly speaking, the relationship between the two radii is approximately the same as the one between the respective footprint diameters, cf. Table 4.2; more details on the derivation of appropriate service radii can be found in [WBL95b].

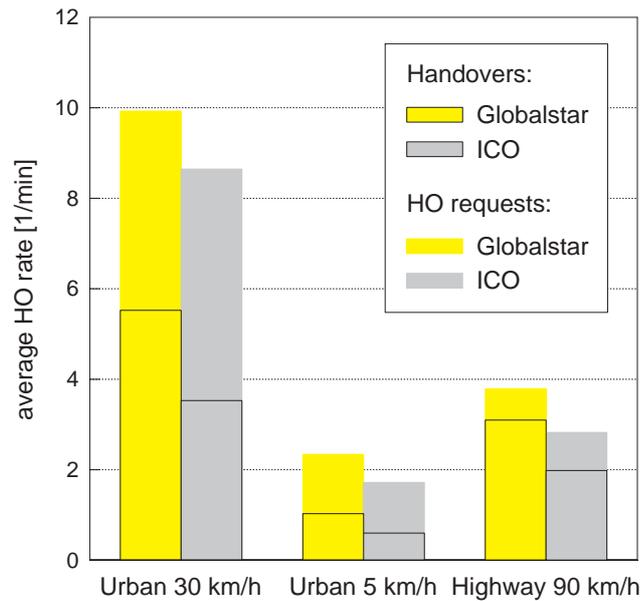
Figure 4.9 displays all relevant results at a glance, in direct comparison of the Globalstar and ICO systems. Visibility and availability curves are displayed over the distance of the user position from the GW, where the different abscissa scaling between the Globalstar and ICO diagrams – according to the above-mentioned service area radii – should be noted.

As a general result, satellite diversity yields a significant improvement of service availability. For both systems, all curves drop with increasing distance between user and GW; as one would expect, both the probability of seeing  $n$  satellites at all and also the experienced elevation angles decrease, and as a consequence of both also the service availability is reduced with distance.

The mutual dual satellite visibility turns out to be significantly higher for Globalstar than for ICO. The availability curves, on the other hand, show that the potential of diversity is nearly completely exploited in the dual satellite diversity scenario; a third satellite brings only negligible additional



**Figure 4.9:** Comparison of satellite diversity performance for Globalstar (left) and ICO (right): (i) visibility; service availability in (ii) highway environment and (iii) urban environment (from top to bottom).



**Figure 4.10:** Equivalent average handover (HO) rates for different environments.

improvement. Consequently, the diversity gain is higher for Globalstar. However, in both environments the absolute availability figures in the diversity case are nearly the same for both systems, with even slight advantages for ICO. The reason is that the first visible satellite in the ICO system obviously has a better elevation angle distribution, inherently providing higher availability in the case of no diversity. Comparing the highway environment with the city environment it can be concluded that S-PCNs can provide high service quality in highway or rural areas. In urban areas satellite diversity can substantially improve the service availability, but the maximum achieved figures are still too low from a user QoS point of view to make a stand-alone S-PCN an attractive solution; in such environments, they can only serve as complement to terrestrial systems or they must focus on certain fixed applications where terminal antennas can remain in favorable positions not affected by shadowing.

For a pure on-off channel, the concept of satellite diversity would be equivalent to a system initiating a seamless satellite handover each time the currently used satellite becomes shadowed. Comparing such a handover (HO) system with the diversity system, it is obvious that the resulting availability is the same if one assumes both perfect diversity and handover schemes. The latter means that whenever a serving satellite is no longer available, a perfect switching to an available successor is performed, if there is any; generally, such HO actions lead to a significant amount of signaling. A comparison of diversity and equivalent HO operation is therefore interesting in order to specify the signaling complexity reduction properties of a diversity system.

As a measure for the signaling effort in a comparable handover system, Fig. 4.10 shows some equivalent HO rates which a user would experience due to the particular environment and shadowing statistics. Note that all these handovers would occur in addition to the regular satellite handovers. The influence of environment and mobile user velocity on the absolute HO figures is obvious. The discrepancy between HO requests and successfully performed handovers is much higher in unfavorable urban environments, where an alternative satellite is less likely to be available.

Concluding the considerations on UDL routing, we have seen at several occasions how satellite handover and satellite diversity operation may affect the ISL routing, both qualitatively and quantitatively. Following the networking decomposition paradigm outlined earlier, we will now return to ISL routing as a separate task and mainly focus on the ISL segment in the remainder of the thesis. Nevertheless, the possible qualitative and quantitative impacts of UDL routing on ISL routing should always be kept in mind.

For aspects of integrated UDL and ISL routing the interested reader may refer to some recent research contributions of Franck [EF00, Fra01], focusing on routing algorithms, and to recent work by Wood [WPE01, Woo01], elaborating on the particular issue of extending the satellite diversity concept from the UDL segment to the ISL network, and on related constellation design and routing options.

### 4.3 Off-Line Dynamic ISL Routing Concept

Classical routing strategies have traditionally more or less focused on some kind of shortest or multiple path search in networks with a fixed topology. Dynamic routing capabilities are then only required for traffic adaptive routing or in reaction to unpredictable link or node failures.

In contrast to this, in ISL subnetworks it is generally required that the routing is dynamic already in order to cope with the permanent topological changes, enabling continuous operation in connection-oriented mode. *Dynamic Virtual Topology Routing (DVTR)* [Wer97] has been proposed for use in such environments. It builds essentially upon a very general discrete-time network model which we will present first, before returning to a more satellite-specific view again and looking at DVTR in detail.

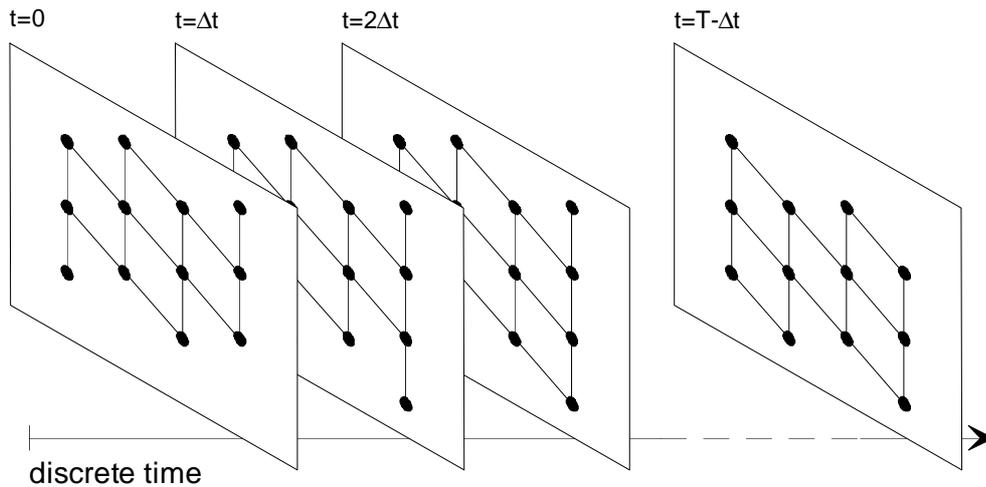
The development of the concept is *intentionally* performed using the more general and most challenging case of a constellation with switched ISLs (Iridium as example). Of course, topologies employing permanent ISLs only are covered by the concept just as a “simple” special case. Actually, we will show later in Chapter 5 how the DVTR concept can be applied within the ISL network design process for strictly permanent topologies.

#### 4.3.1 Discrete-Time Network Model

Consider an arbitrarily meshed network topology with a constant number  $N$  of network nodes (satellites in the constellation). The network topology is in the most general case subject to changes due to (i) discrete-time activation/deactivation of links, and (ii) continuous-time distance variations between nodes; depending on the actual implementation, the former may be avoided. Moreover, the complete topology dynamics is periodic with period  $T$ .

In addition, the model does not only take into account the topology, but additionally incorporates some baseline assumptions for the communication system:

- The primary services to be provided are delay-sensitive (e.g. telephony, video) and thus require clear priority for the QoS parameters delay and delay jitter; this is reflected in the link cost functions and thereby drives path search.



**Figure 4.11:** Discrete-time topology approach.

- Information flows of a single end-to-end connection always use the same path in both directions.
- The system should inherently provide a straightforward and robust scheme to cope with path failures, i.e., disjoint backup paths are desired.

Starting from these assumptions, the proposed network model is based on a discrete-time topology approach as illustrated in Fig. 4.11. The dynamic network topology is considered as a periodically repeating series of  $S$  topology snapshots separated by step width  $\Delta t = T/S$ . Each of the snapshots at  $t = s\Delta t$ ,  $s = \{0, \dots, S-1\}$ , is modeled as a graph  $G(s) = (V, E(s))$ , where  $V = \{1, \dots, N\}$  is the constant set of nodes and  $E(s)$  represents the set of undirected links  $(i, j)_s = (j, i)_s$  between neighboring nodes  $i$  and  $j$ , existing at  $t = s\Delta t$ . Associated with each link are its costs  $c_{ij}(s)$  according to an appropriate cost metric. In the following, it is assumed that the link costs are mainly determined by (i) node distance or propagation delay and (ii) an additional constant node processing/switching delay. It is stressed at this point that a simple hop count metric can of course be easily adopted as well, and it just appears as a specific case of the general approach and notation.

With  $G(s)$ , the topological situation encountered in time interval  $[s\Delta t \dots (s+1)\Delta t)$  is explicitly fixed. Considering the network characteristics and assumptions summarized above, this is appropriate

- (i) if  $\Delta t$  is adapted to the discrete-time link activation/deactivation behavior (e.g., “synchronized” with it) so as to guarantee that  $G(s)$  correctly reflects the physical topology throughout the interval, and
- (ii) if  $\Delta t$  is small enough to keep the slowly continuing distance (= cost) variations during  $[s\Delta t \dots (s+1)\Delta t)$  well below a “reasonable” limit with respect to the values at instant  $t = s\Delta t$ . For monotonous cost variation, this condition becomes

$$\frac{c_{ij}((s+1)\Delta t) - c_{ij}(s\Delta t)}{c_{ij}(s\Delta t)} \ll 1, \forall (i, j)_s \in E(s). \quad (4.3)$$

**Table 4.3:** Parameters of the discrete-time network model.

$N$	constant number of nodes (satellites)
$T$	constellation period
<i>discretization:</i>	
$S$	number of constellation/topology snapshots
$\Delta t = T/S$	discretization step width
$t = s\Delta t,$ $s = \{0, \dots, S - 1\}$	snapshot times
<i>network graph:</i>	
$G(s) = (V, E(s))$	network graph at $t = s\Delta t$ with:
$V = \{1, \dots, N\}$	constant set of nodes
$E(s) = \{(i, j)_s\}$	time-varying set of undirected links between neighboring nodes $i$ and $j$
$c_{ij}(s)$	costs associated with link $(i, j)_s$
<i>related to routing:</i>	
$W$	constant set of $N(N - 1)/2$ unordered OD pairs
$w \in W$	single OD pair
$K$	maximum number of alternative OD paths
$P_w(s)$	set of up to $K$ distinct loopless paths for $w$
$p(s) \in P_w(s)$	single path assigned to $w$
$\delta_{(i,j)_s}^{p(s)} \in \{0, 1\}$	link occupation indicator: link $(i, j)_s$ used by path $p(s)$ or not

Whereas the first condition is hard, the second one leaves some freedom for an appropriate trade-off: the choice of  $\Delta t$  may, for example, be made as compromise between achieving low distance variation within any single interval (small  $\Delta t$ ) and introducing as few as possible instantaneous distance offsets between subsequent intervals (large  $\Delta t$ ).

Throughout the period  $T$  a constant set  $W$  of  $N(N - 1)/2$  unordered origin-destination (OD) node pairs is given. A set  $P_w(s)$  of up to  $K$  distinct loopless paths is now assigned to each OD pair  $w, w \in W$ . Every path  $p(s) \in P_w(s)$  consists of a unique sequence of links  $(i, j)_s$ . A *link occupation indicator*  $\delta$  catches the relationship between a link and a path,

$$\delta_{(i,j)_s}^{p(s)} = \begin{cases} 1 & \text{if link } (i, j)_s \text{ belongs to path } p(s) \\ 0 & \text{otherwise.} \end{cases} \quad (4.4)$$

Table 4.3 lists all relevant parameters of the discrete-time network model at a glance.

### 4.3.2 Dynamic Virtual Topology Routing (DVTR)

Based on the general discrete-time network model developed in the previous section, we propose *Dynamic Virtual Topology Routing (DVTR)* for use in dynamic ISL networks, a concept that lends itself quite directly to an implementation exploiting the potential and strengths of ATM virtual connection operation – without being limited to ATM, however. Figure 4.12 illustrates the approach at

a glance. The basic idea is to exploit the fact that all topology dynamics is periodically deterministic. This allows the setting up off-line of a time-dependent virtual topology providing continuous operation of end-to-end connections. This essentially builds the framework for later on-line traffic adaptive routing strategies as indicated in the figure, i.e. traffic adaptive routing can be performed within the limits of the given virtual topology.

Assigning paths or, in ATM terminology, VPCs to OD pairs essentially determines an *instantaneous virtual topology*  $VT(s)$  upon  $G(s)$ , covered by module I-VTS(s) in Fig. 4.12. This task is performed by a  $k$ -best path search algorithm for every OD pair. The single shortest path search task can be formulated as finding the least-cost path  $p(s)$ , i. e. the path with minimum path cost

$$\min_{\forall p(s)} C_{p(s)} = \min_{\forall p(s)} \sum_{i,j} c_{ij}(s) \delta_{(i,j)_s}^{p(s)}. \quad (4.5)$$

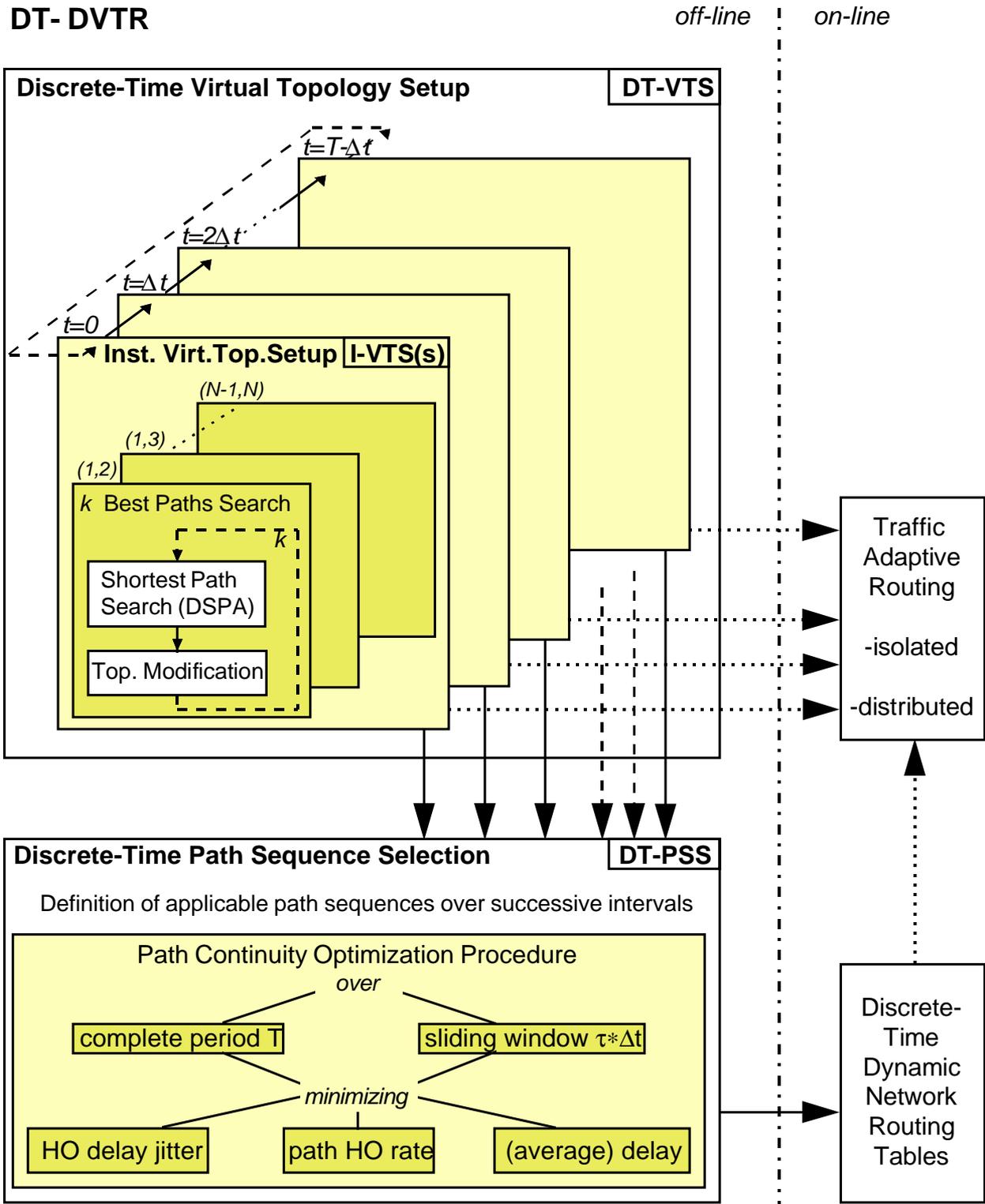
An iterative approach based on successive calls of the Dijkstra shortest-path algorithm (DSPA) [Dij59] is indicated, which allows the introduction of topology modifications in between. For instance, by “eliminating” already occupied links one can force a set of link-disjoint paths, thus providing a base for simple and robust fault recovery mechanisms and building a sound framework for shaping network traffic flow at operation time. The path cost metric can for instance consist of the sum of all single link propagation and node processing/switching delays encountered on the path.

Performing I-VTS(s) for all  $s = \{0, \dots, S-1\}$  constitutes the *discrete-time virtual topology setup DT-VTS*.

DT-DVTR is finally completed with another off-line module, namely *discrete-time path sequence selection DT-PSS*. This module solves the problem of selecting for all OD pairs sequences of VPCs over successive time intervals from the virtual topology provided by DT-VTS; here the routing becomes really dynamic and potential problems due to path switching can be avoided or minimized. Promising variants for such a *path continuity optimization procedure* are (i) minimizing HO delay jitter (i.e. instantaneous delay offsets during path handover) and (ii) minimizing path HO rate, with some restrictions also (iii) minimizing (average) delay. Concerning the duration of the optimization interval, two main approaches are incorporated:

1. The optimization is performed over the complete period  $T$ , then achieving an optimal network solution. The result is either one unique first-choice path sequence  $\{p_1(0), p_1(1), \dots, p_1(S-1)\}$  out of  $k^S$  possible ones, with  $p_1(s) \in P_w(s)$ , or a set of  $Q$  ordered (i.e. prioritized: first-choice, second-choice, etc.) sequences of such kind,  $\{p_q(0), p_q(1), \dots, p_q(S-1)\}$ ,  $q = \{1, \dots, Q\}$ .
2. The optimization is performed in a time-distributed manner within a *sliding window* of discrete-time duration  $\tau$ , i.e. extending over the interval  $[s\Delta t \dots (s+\tau)\Delta t)$  (modulo  $T$ ),  $s = \{0, \dots, S-1\}$ ,  $\tau \in \{2, \dots, S-1\}$ . The results are  $s$ -unique first-choice path sequences  $\{p_1(s), p_1(s+1), \dots, p_1(s+\tau-1)\}$  (modulo  $S$ ),  $\forall s \in \{0, \dots, S-1\}$ , respectively  $s$ -unique sets of ordered sequences,  $\{p_q(s), p_q(s+1), \dots, p_q(s+\tau-1)\}$ ,  $q = \{1, \dots, Q\}$ , for all  $s$ .

This theoretical framework provides a sound basis to extend analytical work and formulate any (minimax) target functions that are related to optimized routing in a wider sense, like, e.g., minimizing worst-case link traffic throughout the network by means of traffic adaptive routing. In the



**Figure 4.12:** Discrete-time dynamic virtual topology routing (DT-DVTR): concept and implementation.

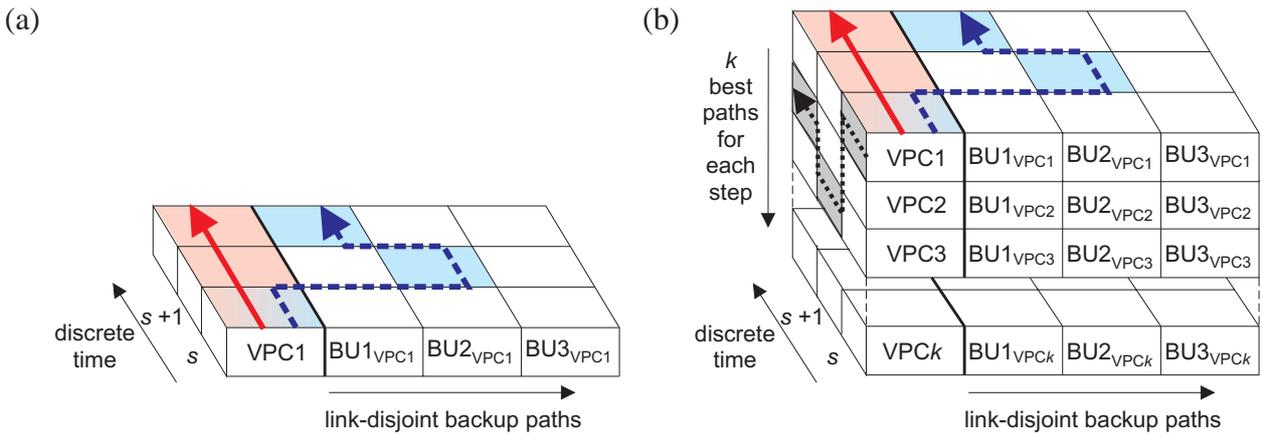
framework of ISL network dimensioning for broadband LEO systems in Section 5 we will recall and make use of this potential in the mentioned form.

After this formal description of routing concept and optimization it should be helpful to look at some more illustrative and intuitive explanations for both, (i) the setup of instantaneous virtual topologies and (ii) the path sequence selection options, under concrete assumptions.

1. *Setup of instantaneous virtual topologies.* For each of the snapshots the virtual topology is constituted by means of VPCs “directly” connecting all  $N(N - 1)/2$  pairs of satellites. In ATM terminology, a fully meshed SHVP (single hop virtual path) topology [Bau97b] is provided which contains a set of (up to)  $k$  alternative paths (VPCs) for each terminating satellite pair. Two reasonable solutions are the following:
  - (a) All paths are link-disjoint. This reduces the number of alternatives for the latter optimization just from the beginning and at the same time stocks up reasonable backup paths for the operational system. In a constellation with (up to) 4 ISLs per satellite, for instance,  $k \leq 4$  as the number of link-disjoint paths is directly limited by the number of ISL ports of the first/last satellites.
  - (b) The  $k$  “shortest” paths (with respect to a certain cost metric, but without the demanding link-disjoint restriction) are in the set. This enhances the flexibility in the latter optimization, which is specifically important when aiming at dimensioning of the operational system. Additionally, corresponding link-disjoint backup paths may be assessed for each of the  $k$  paths.

These considerations are illustrated for *one* satellite pair in Fig. 4.13. Solution (a) extends over two dimensions only and is completely defined in one “layer”, Fig. 4.13(a): for each time step  $s$ ,  $k = 4$  link-disjoint paths are in the set, and each of them can on principle be chosen as active one, leaving the other three as potential backups (BU). Solution (b) additionally comprises a third (vertical) dimension, Fig. 4.13(b): active paths can be chosen from the left “column” per time step, whereas the corresponding (link-disjoint) backup paths are already prepared as well. One should note that any link-disjoint backup (row) path can be physically identical with any of the  $k$  shortest (column) paths, as long as they are not in the same row, of course.

2. *Path sequence selection.* The second major task is now to select (for each satellite pair) appropriate *sequences* of paths from the subsequent instantaneous sets. Three possible candidates for such a selection are illustrated by the arrows in Fig. 4.13. Following the solid one simply means that always the instantaneously “best” path, e.g. the one with minimum transfer delay, is chosen. This simple approach seems to be quite promising for an ISL topology that employs only permanent ISLs, which means that the most critical path switching situations (in terms of delay jitter) are avoided straight from the beginning. The dashed arrow may reflect a path sequence that tries to minimize the worst case handover delay jitter encountered when path switching from one to another step is inevitable, like it is the case in Iridium when ISLs are switched on or off. This approach has been extensively studied and evaluated in [WDV<sup>+</sup>97]. A basic result has been that sophisticated optimization schemes can significantly reduce the handover delay jitter, but only at the expense of high complexity both before and during system operation. However, this is a general drawback of bounded ISL topologies in (near-)polar constellations. The dotted arrow finally represents a VPC



**Figure 4.13:** Illustration of instantaneous active and backup VPCs and of VPC sequence selection for one terminating satellite pair; (a)  $k$  link-disjoint paths, (b) any  $k$  shortest paths.

sequence that guarantees a constant underlying physical route over time. This is possible within a non-switched topology, if only  $k$  is chosen large enough to include this path in the set over all steps.

In the remainder of this section we will restrict ourselves to presenting some numerical results for the off-line routing approach in a system with switched ISLs (Iridium), focusing on the comparison between the system period and the sliding-window approaches with respect to the critical path switching delays.

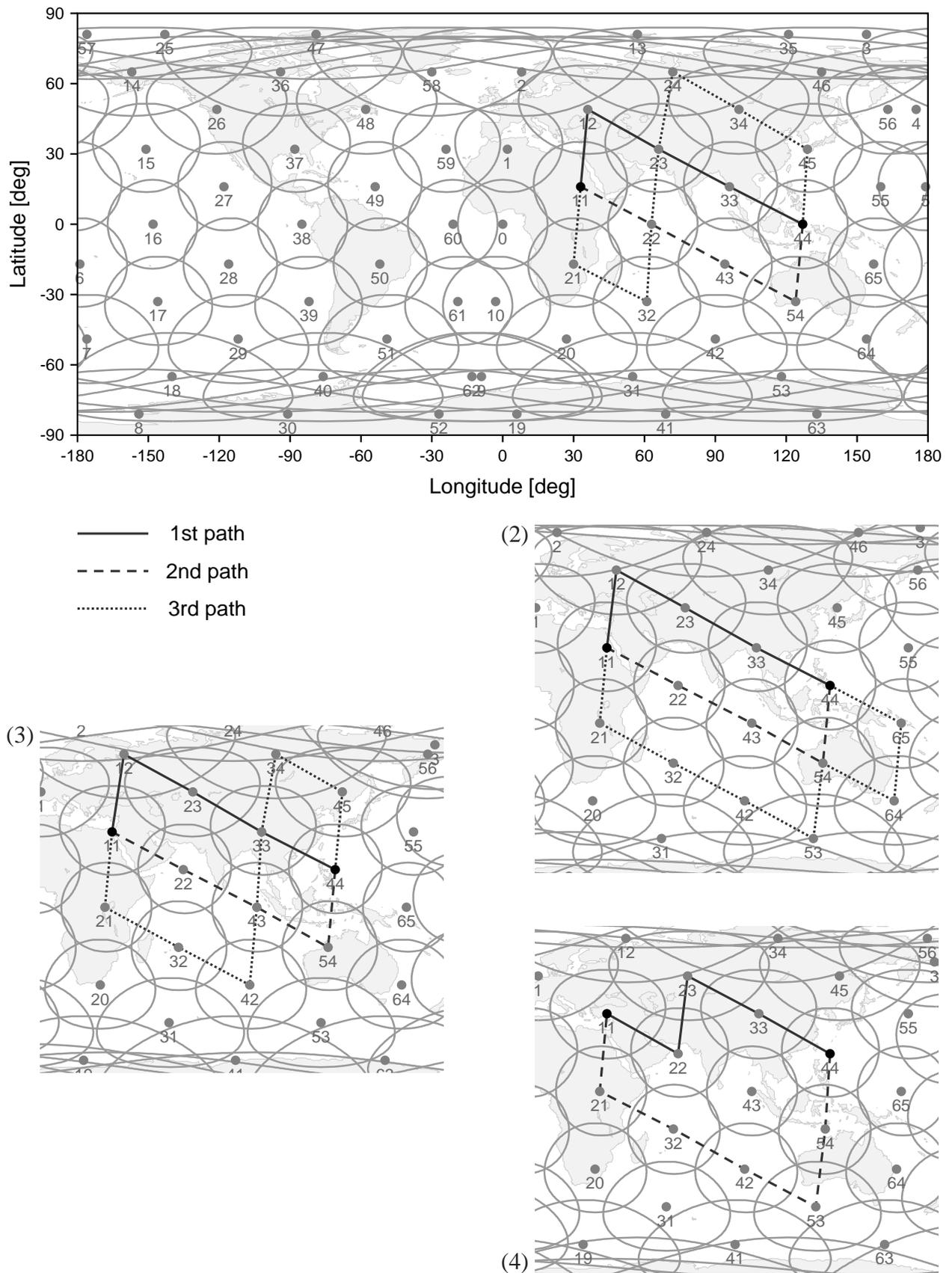
Figure 4.14 shows examples of instantaneous VPC sets over subsequent time intervals as found by an iterative Dijkstra link-disjoint SPA module of the ISLSIM<sup>2</sup> software. The effect of forced path switching becomes obvious as some ISLs are switched off when approaching the polar region. In the chosen example, the VPC sets comprise a maximum of three alternative paths although up to four would theoretically always be possible under the link-disjoint regime. The reason is an effective upper path delay bound, controlled in the I-VTS module, that does not allow paths in the set whose path delay exceeds the shortest path delay by a certain factor. Put simply, such paths are not eligible as they are “too long”.

Considering only the first-choice path sequence, both mentioned optimization methods are illustrated for the *minimize VPC HO rate* target function in Figs. 4.15 and 4.16, respectively.

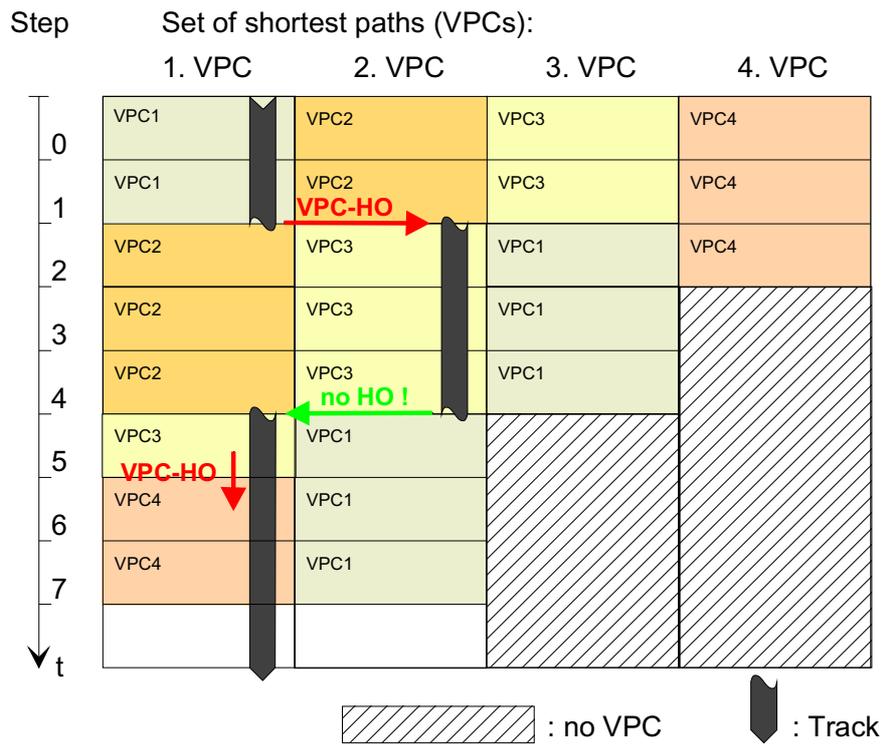
For the whole system period, one unique *track* of VPCs is defined as the first-choice VPC sequence for each terminating satellite pair in the system period approach, Fig. 4.15. The criterion for defining this track is a minimum number of VPC handovers (VPC-HO) over the whole system period. All end user connections that are served by the respective satellite pair will use the same unique VPC in a given interval. The rationale for this approach is to use a straightforward global minimization of the VPC handover rate, considering (i) that the overall signaling complexity decreases and (ii) that the critical HO delay jitter affecting single connections is reduced inherently by minimizing the probability of a HO action to encounter at all.

The main objective of the more sophisticated sliding-window approach is to extensively exploit the optimization potential with respect to single short-term connections, paying the price of course of

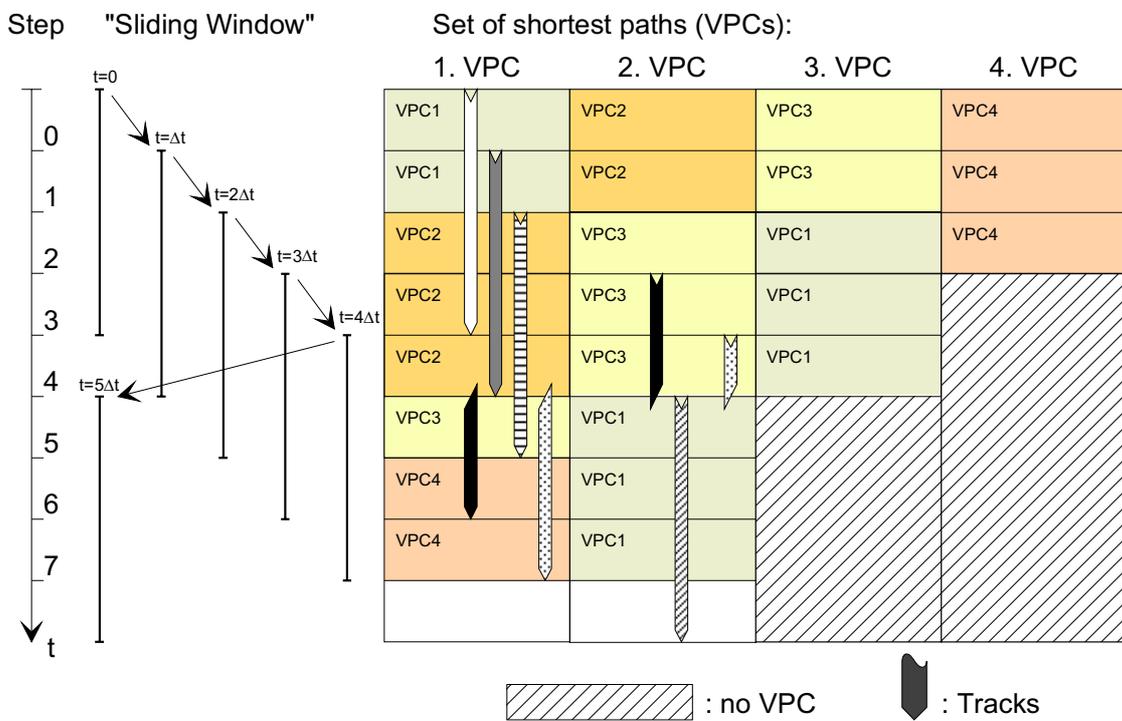
<sup>2</sup>Proprietary simulation tool of the German Aerospace Center (DLR)



**Figure 4.14:** Instantaneous sets of link-disjoint shortest-path VPCs over four subsequent time intervals ( $\Delta t = 2$  min) for satellite pair 11–44.



**Figure 4.15:** System period optimization approach for one terminating satellite pair. VPC handover situations are indicated. The same VPC numbers and grayscale levels identify one unique physical path.



**Figure 4.16:** Sliding-window optimization approach for one terminating satellite pair. The same VPC numbers and grayscale levels identify one unique physical path.

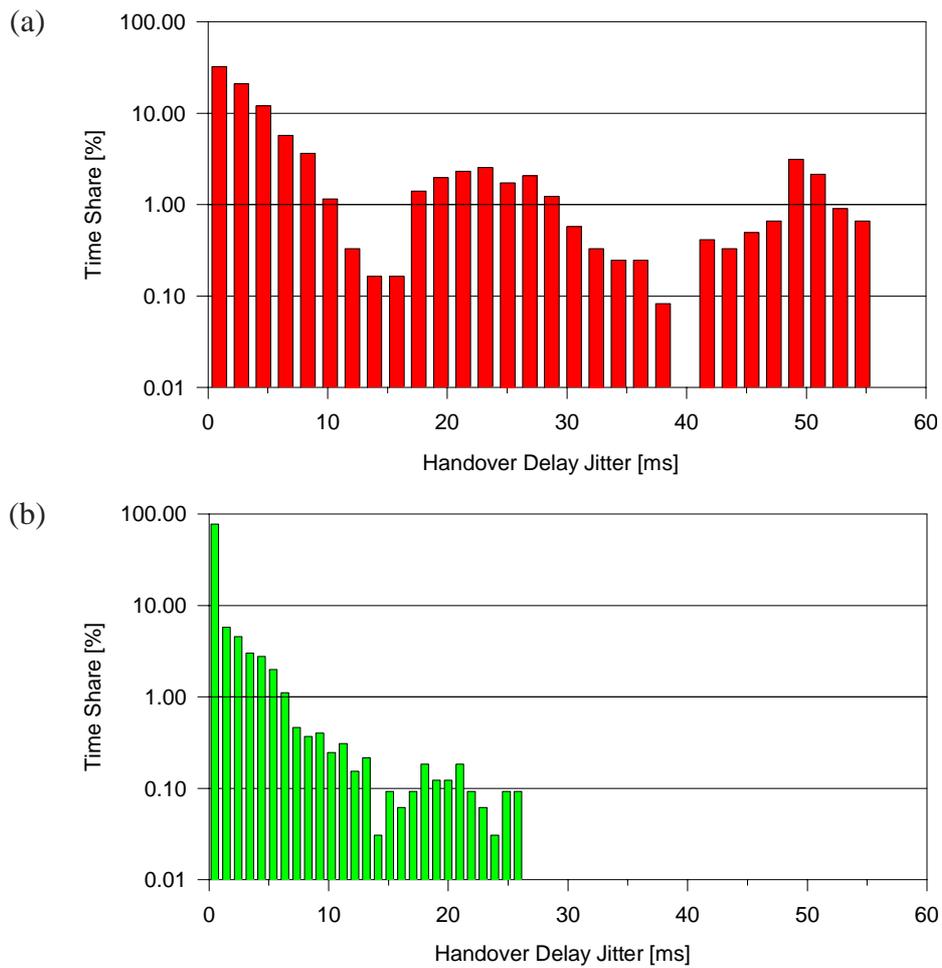
increased operational complexity. This is illustrated in Fig. 4.16, again considering one terminating satellite pair: a sliding time window (with window size typically exceeding the mean call holding time  $1/\mu$  of the target service, e.g. telephony) is used as the discrete-time optimization interval, where the evaluation of the corresponding mathematical target function is performed. The latter is in essence a weighted sum of maximum and average delay jitter. In this way, optimal tracks through the “landscape” of alternative VPCs are selected. These VPC tracks finally determine the routes, including handovers, that any connection starting in the respective time interval will use in the operational phase of the system. In other words, active connections set up in different intervals will use different VPCs (and consequently different VPs on all single links, even if physically the same!). This indeed requires an increased number of VPIs; more specifically, the *average* number of VPIs required per ISL grows proportionally with the sliding window size.

For the system period approach, further implementation details and the presentation and discussion of extensive numerical investigations can be found in the original publications [Del95] and [WDV<sup>+</sup>97]. Similar information for a variant of the system period approach and the sliding-window approach, in comparison, is given in [Ber96], [WBE97] and [Wer97].

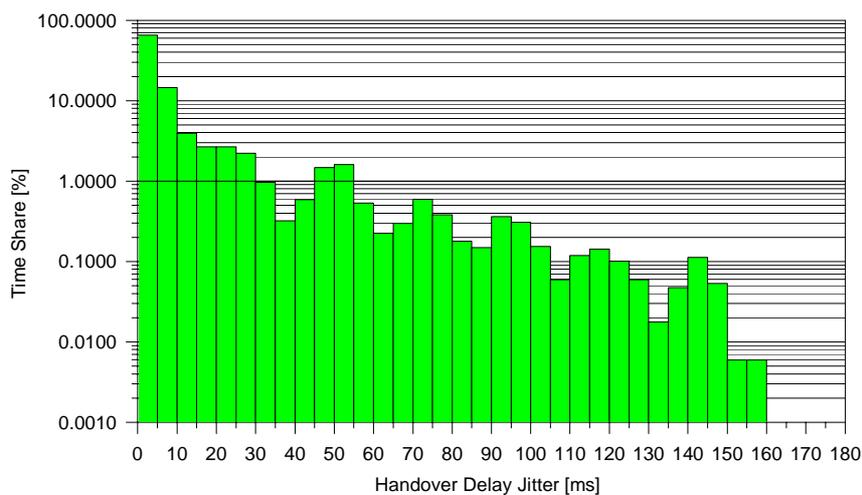
Figure 4.17 illustrates a performance comparison of the two proposed optimization approaches in terms of HO delay jitter distribution experienced by a fixed length 2 minute call in the Iridium ISL subnetwork. The simulated calls are equally distributed over time and all possible OD pairs. Considering such 2 minute fixed length calls is of course not close to reality, but it is the only way to make a fair comparison of both approaches with a discrete-time interval  $\Delta t = 1$  min. The results emphasize the superior performance of the service adaptive sliding-window approach over the simpler period-based optimization scheme. The numerical results given in Fig. 4.18 are much more speaking because they are based on a realistic service scenario. It provides the distribution of VPC-HO delay jitter for a typical telephone service, i.e. showing a negative exponentially distributed call holding time with mean  $1/\mu = 3$  min. The length of the sliding window for HO delay jitter minimization has been 5 minutes. Note that the shown distribution contains only those calls which are affected by a handover during their lifetime; with the given simulation parameters, these are in fact less than 20% of all calls. Altogether this results in a very low percentage of calls that have to cope with an HO delay jitter larger than 20 to 30 ms. However, if one wants to guarantee optimal performance in the light of ATM quality of service, the remaining delay jitter values – which are minimized within the topology-specific limitations – represent critical requirements for any hitless path switching functionality. Comparing Fig. 4.17(b) with Fig. 4.18, one can see significantly higher jitter values in the realistic traffic scenario. This is due to the fact that in the former theoretical case the optimization window equals the fixed (!) call duration, whereas in the latter case 20% of the calls extend beyond the optimized period and are then exposed to path handovers.

## 4.4 On-Line Adaptive ISL Routing

With connection-oriented communications, the task of on-line routing is to select a route at call setup, which is then only changed in case of a forced path handover. If the traffic load (on links or nodes) is taken into account as one of the decision criteria for route selection, then the on-line routing becomes traffic adaptive. Working within the presented off-line routing framework, the task of on-line routing is reduced to selecting one of the predefined alternative paths. Therefore



**Figure 4.17:** Distribution of VPC-HO related delay jitter for a 2 minute call from first satellite 0 in Iridium, averaged over all possible last satellites and one system period. *Optimization strategy:* (a) minimize number of VPC-HOs during one system period; (b) minimize HO delay jitter within a 2 minute sliding window.



**Figure 4.18:** Distribution of VPC-HO related delay jitter for a typical telephone service (negative exponentially distributed call holding time with mean  $1/\mu = 3$  min), averaged over all first/last satellite pairs and one system period. *Optimization strategy:* minimize HO delay jitter within a 5 minute sliding window (i.e. 80% of calls end within optimized track).

it becomes less complex but cannot exploit the full potential of adaptivity. Prior to a proper integration of the two concepts it is therefore reasonable to determine the maximum theoretical gain from operating an on-line adaptive routing without any restrictions, i.e., without the limitations of the off-line framework. This is the scope of this section, extracting the major results from an in-depth study [WM97]. The integrated approach, i.e. on-line routing operating within a given off-line framework, is discussed in the context of ISL network dimensioning in Section 5.4.

## 4.4.1 Distributed Approach Based on Deterministic Algorithms

### 4.4.1.1 Dynamic and Adaptive Routing Concepts

In the context of terrestrial (computer) networks, the concepts of *dynamic* and *adaptive* routing are often used synonymously since all those networks' topology is practically static (except in link or node failure situations) and the only real dynamics is introduced through traffic variations. A deterministic routing scheme is then usually also a static one based on fixed routing tables that are calculated off-line.

In a dynamic topology scenario as encountered in LEO ISL networks, both concepts have to be redefined and clearly separated. Since the topology dynamics is (periodically) deterministic, it can be appropriately modeled by a time-discrete series of quasi-static topologies with a time step  $\Delta t_s$  in between. Routing in such a network becomes inevitably dynamic as it has to perform path search for every of these topologies and ends up with a corresponding series of updated routing tables. The update period is clearly linked to the selected time step,  $\Delta t_u = \Delta t_s$ . However, in the considered on-line approach a possibly changed route for an OD pair is certainly obligatory for all new generated calls, whereas it will be used for existing calls only if the old route does no longer exist. For the purpose of numerical studies, an on-line deterministic routing scheme has been implemented which is based on a version of the Dijkstra Shortest Path Algorithm – the *Moore–Dijkstra Algorithm (MDA)* – using exclusively geometrical ISL lengths as link costs. Therefore the algorithm determines shortest routes respectively routes with minimum transfer delay. This scheme serves as reference for the performance comparison with traffic adaptive routing.

### 4.4.1.2 Decentralized Traffic Adaptive Routing

Traffic adaptive routing is in the ISL scenario mainly motivated by the goal to use the on-board ISL capacity as efficiently as possible. This can be significantly improved by smoothing extreme peaks in the ISL loads as encountered without on-line traffic adaptive routing. Strong time variations in the ISL load can become particularly critical considering the expected high data rates in broadband systems, since all satellites in the constellation have to be designed for the worst case – including ISL payload, power and antenna requirements. Using the expensive orbital resources sparingly will thus directly reduce the costs for every single satellite and the whole system. Moreover, satellites use solar power to charge on-board batteries for servicing traffic in shadowed and eclipsed portions of the earth. Peaked traffic can drain battery charges and leave the satellite effectively “dead” until recharged in the sun-illuminated portions of the orbit about the earth. Therefore, traffic demands must be constrained to optimize on-board battery life.

The study of adaptive routing schemes has shown that efficient implementations work in *distributed* or *decentralized* manner. According to [Sch87], a distributed traffic adaptive routing procedure generally consists of the following three components:

1. a measurement or monitoring process that collects the required information on the actual traffic situation in the network;
2. an update protocol that distributes (e.g. by means of flooding) this information to all nodes;
3. the shortest path computation to be performed in every single node, using the updated traffic data built into the respective link cost function.

Clearly, such a scheme is also dynamic since all topological changes have to be reported to all nodes as well during the update period, or they are built-in because they are deterministic and therefore known in advance. An important parameter for adaptive routing is the interval  $\Delta t_u$  between successive updates of the routing tables. In particular, the relationship between the update period and the typical duration of connections is essential for a reasonable trade-off between performance gain and complexity. Typically, there is some lower bound for  $\Delta t_u$  below which only little performance improvement can be achieved at the expense of unacceptably high update protocol traffic.

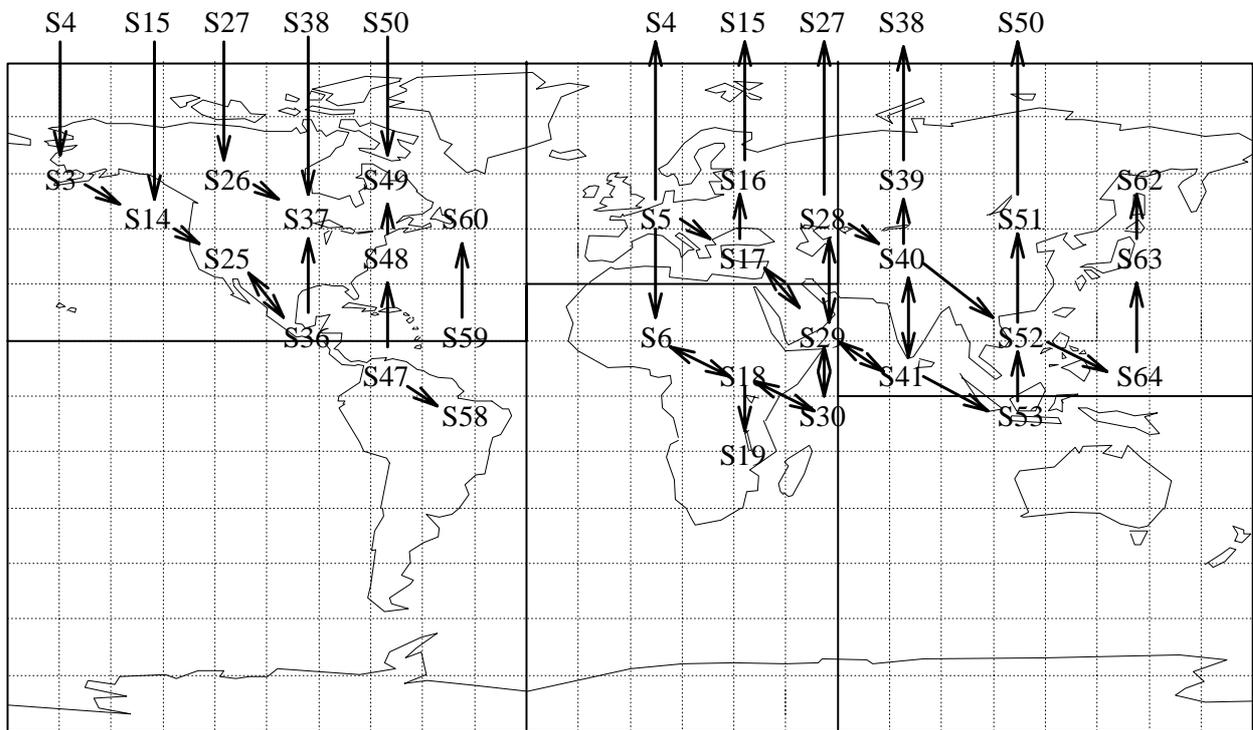
In the on-line traffic adaptive approach the *Distributed Bellman–Ford (DBF) Algorithm* [BG87b] is used for the path search. Three different path cost functions have been investigated [Gar96, WM97] in connection with this scheme: The first two are based on the sum of the costs of all participating *links*, where (i) the link cost is a function of link traffic alone, (ii) the link cost is a metric combining link traffic and link length, whereas the third one (iii) uses the sum of *node* costs (only traffic based). For the following performance comparison with MDA we restrict ourselves to (i), investigating the maximum potential with respect to worst-case link (WCL) load reduction through traffic adaptivity.

The investigations have been performed for Iridium under a global telephony scenario including a call model, an addressable market model to characterize the subscriber distribution and source traffic demand, a daily user activity model, and a global traffic flow model describing the destinations of generated calls. Details of the traffic modeling and further references can be found in [WM97]. All simulations have been performed over the complete Iridium system period  $T_S = 5$  days<sup>3</sup> in order to guarantee that all possible pairs of source traffic patterns and orbital patterns are considered. The time step  $\Delta t_s$  has been 2 minutes throughout. For DBF's update interval  $\Delta t_u$ , values ranging from 12 seconds up to 2 minutes have been evaluated in test runs; from that, an optimum value of 20 seconds has been fixed for all simulations.

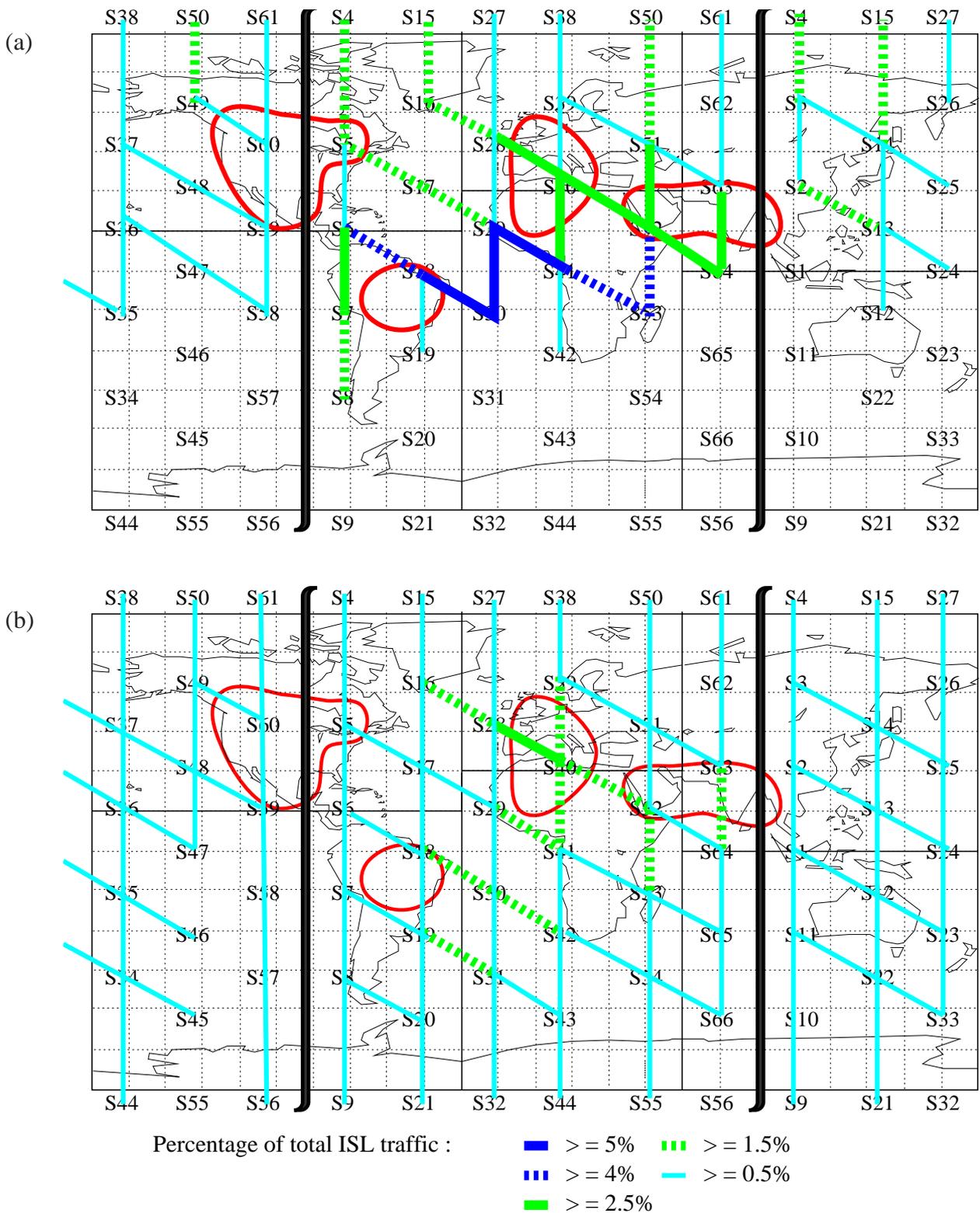
Figure 4.19 displays the pronounced asymmetry of ISL loads at the beginning of the simulation (shortest paths). This is mainly due to the combination of strong geographic source traffic variations and the seamed ISL topology. Figure 4.20(a) shows that this asymmetry is maintained with the non-adaptive MDA routing scheme, always using shortest-delay paths. Note that with 106 maintained ISLs a uniform distribution across the network would imply roughly that 1% of the total network traffic be carried on each link. This is far from being the case: the ten most loaded links together carry roughly half of the network traffic. In contrast to this, Fig. 4.20(b) proves that the traffic adaptive DBF scheme clearly achieves a more even traffic distribution in the network: except one, all links carry less than 2.5% of total network traffic each, the ten most loaded links together carry roughly 20%, and 40 links carry around 1% of the total network traffic each.

---

<sup>3</sup>Note that in this case the system period, as defined in Eq. (2.14), results as least common multiple of the orbit period  $T = 100$  min and the mean solar day  $T_{\text{Sol}} = 1440$  min.



**Figure 4.19:** Most loaded ISLs at  $t = t_0$ . Each displayed link carries more than 2% of the total network traffic. Arrows reflect the direction of *call generation* (source  $\rightarrow$  destination) rather than the amount of traffic carried in the indicated direction; carried traffic is bidirectional symmetric in all cases according to the underlying call model.



**Figure 4.20:** ISL traffic distribution with (a) the MDA and (b) the DBF scheme at  $t = 300$  min. The seam appears over the Eastern USA and India/China. High source traffic areas are highlighted.

For the following ISL load values it should be noted that all figures are normalized to the total uplink source traffic. By doing this, the influence of daily activity variations is excluded and the effects of the routing schemes can be exposed more clearly.

The highest traffic load encountered on a single link at a given instant – the worst-case link (WCL) load – is one of the most important performance parameters since it finally determines the ISL power requirements and thus the ISL link budget. Figure 4.21(a) illustrates that the DBF scheme can reduce the maximum ISL traffic by roughly 30% in the average, the instantaneous value being strongly dependent on the geographic source traffic pattern: the reduction tends to be smaller when most generated traffic is concentrated on a few adjacent zones on earth. The DBF scheme's capability to distribute traffic more evenly in the network is also stressed taking another viewpoint, as illustrated in Fig. 4.21(b). The worst-case link never carries less than 4.5% of total network traffic with MDA, but only 2–3% with DBF. On the other hand, one observes quite a constant increase in the average ISL traffic load, Fig. 4.21(c). This is due to the fact that reducing peak loads on single ISLs is often achieved by shifting parts of the traffic from the shortest paths to alternative paths with more hops. According to its link cost metric, at a given time step the DBF scheme prefers previously less loaded links, whereas the load is decreased on the critical ones. Thus the group of ISLs carrying medium traffic loads grows compared to MDA, Fig. 4.22.

## 4.4.2 Isolated Approach Based on Neural Networks

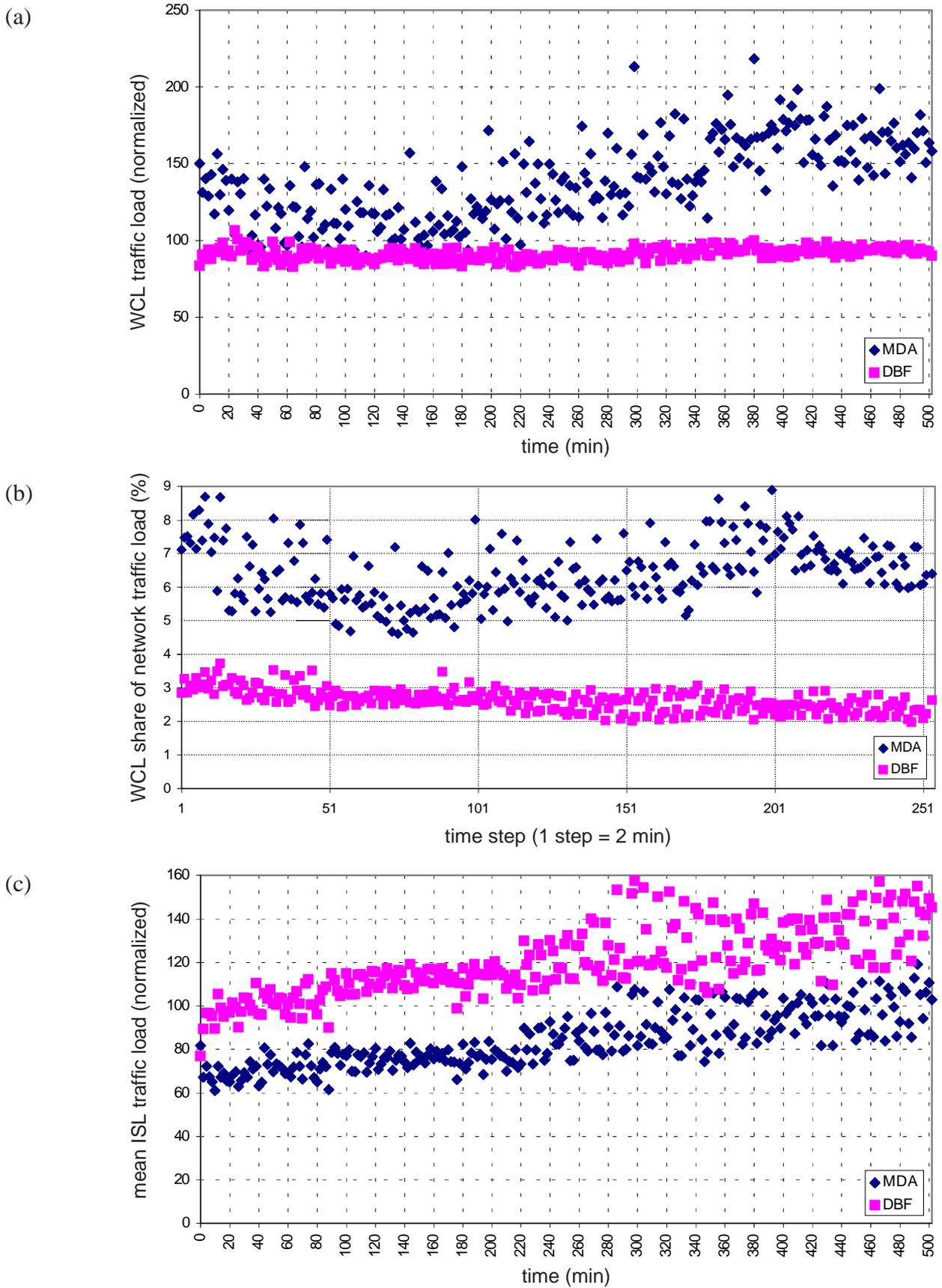
### 4.4.2.1 Motivation and Background

Neural networks (NN) or artificial neural networks (ANN) try to achieve real-time response and human-like performance using many simple processing units, so-called neurons, operating in parallel as in biological brains. The greatest potential of these networks lies in areas such as speech and image recognition, where many hypotheses are pursued in parallel, high computation rates are required, and the performance of the conventional best algorithms is far below that of humans.

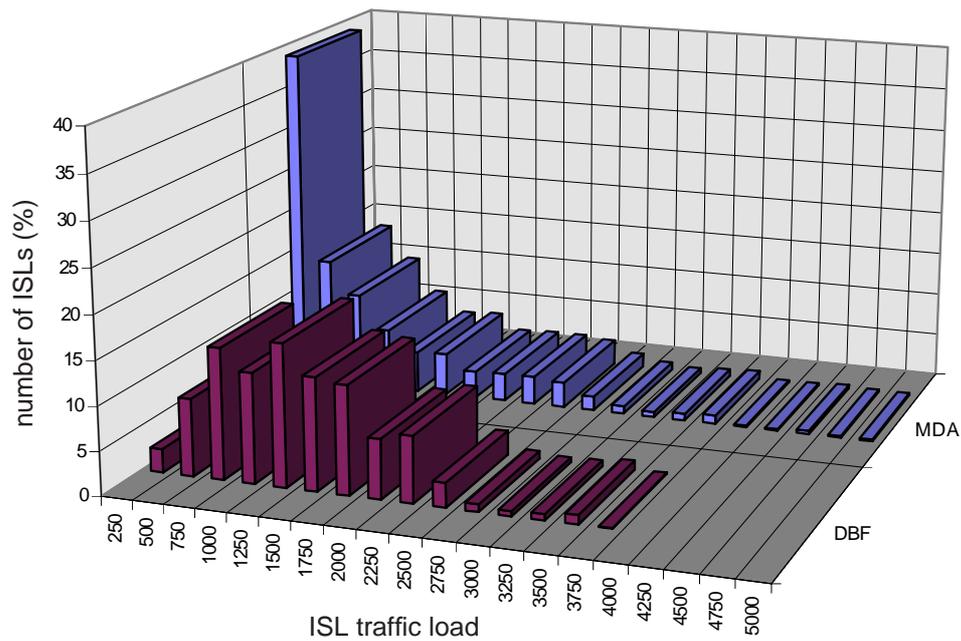
In general, an ANN takes advantage of the redundancy incorporated in its distributed processing structure, which essentially forms an internal representation of a given input vector. Based on that, it performs a functional approximation, whose accuracy only depends on the way in which this internal representation is built up [FMTG93]. The application of neural networks has been found to be particularly successful in such cases where a subject can be formulated as a pattern recognition or pattern classification problem.

Consequently, NN applications have also shown their suitability in some particular areas of telecommunication networks. Fritsch et. al. [FMTG93] provide a good summary of various NN applications in telecommunication systems. Here in general a distinction has to be made between three main classes of nets: (i) feed-forward multilayer perceptrons (MLPs) with backpropagation learning, (ii) Hopfield feedback nets, and (iii) self-organizing feature maps. The latter two have successfully been applied to the task of dynamic channel allocation (DCA) in cellular mobile networks [DFR96, SP97]. Feed-forward MLPs have been primarily proposed in the context of ATM networks for connection admission control (CAC) [TGR91, MS94, YHS97], congestion control [LD97], QoS control [Hir90] and dynamic bandwidth allocation [MCCL95].

With respect to routing, several research activities have concentrated on using self-organizing [FMTG93, Zha93] or Hopfield nets [RW88, FM91, KA93], but in many cases (partly severe)



**Figure 4.21:** Comparison of plain shortest path routing (MDA) and traffic adaptive routing (DBF): (a) WCL traffic load, (b) WCL share of total network traffic, (c) mean ISL traffic load.



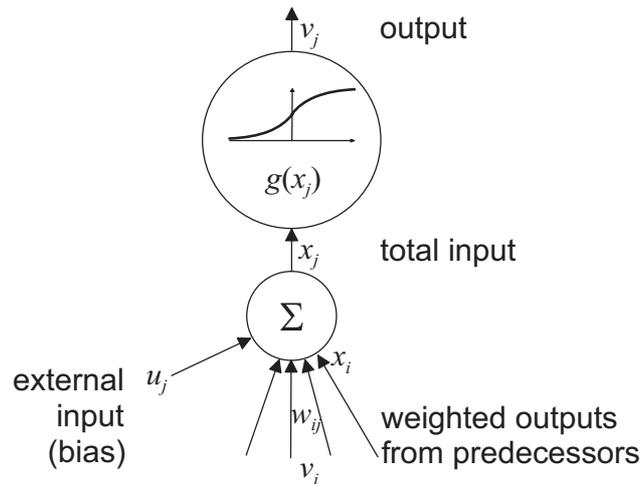
**Figure 4.22:** Comparison of plain shortest path routing (MDA) and traffic adaptive routing (DBF): ISL traffic load histogram.

problems with the so-called *penalty parameters*, one key to appropriate application of Hopfield nets, have been reported. In our original work on NN-based ISL routing [May96, WMMH98] we have also investigated a Hopfield approach but could only confirm such problems. In the following, we will therefore restrict to the case of a feed-forward MLP applied to ISL routing. Here, the basic idea has been to translate the core task of our scenario, namely the route selection or path decision itself, into a (traffic) pattern recognition problem and to let the MLP learn “optimal” solutions which are first calculated with a conventional algorithm to produce training data. A similar approach was adopted in [CSM92] for routing in a terrestrial packet switching network.

It is worth emphasizing that this approach does not appear very intuitive at first sight, because there is at least no *obvious* correlation between local traffic pattern at the source node and a reasonable traffic adaptive route selection in a wide area ISL network, where affected links are often far away from this source node region. On the other hand, if the NN should prove to be only partly successful in learning such a correlation, this would not only prove the existence of the phenomenon itself but already exploit its potential. Traffic adaptive routing would no longer induce huge amounts of signaling data to be exchanged permanently between all network nodes, which is required for any *deterministic* algorithm for distributed adaptive routing, as mentioned in the last section. Moreover, faster routing decisions and better adaptability could be achieved with the NN approach since it avoids (sometimes severe) routing information delay resulting from update intervals in deterministic schemes.

#### 4.4.2.2 Basics of Multilayer Perceptrons

A general introduction into the theory of neural networks can be found in [HKP91], and [Zel94] provides a broad overview and classification of various types of neural nets, too. For the purpose of this thesis we will now follow [FMTG93] in presenting briefly those basics of MLP-type nets



**Figure 4.23:** Schematic picture of a neuron.

which are deemed necessary for a fairly good understanding of their functionality in the context of our ISL routing problem.

The basic processing element of any type of ANN is the *neuron*. It usually performs a non-linear mapping of a set of input values  $x_i$  onto an output value  $v_j$  using a so-called *transfer function*  $g$ . Let  $w_{ij}$  be the *weight* of the link from neuron  $i$  to neuron  $j$ , representing the strength of the connection between both. Then the weight matrix  $\mathbf{W}$ ,

$$\mathbf{W} = [w_{ij}] , \quad (4.6)$$

represents the characteristic *topology* of the NN, summarizing the number and arrangement of its nodes (neurons) as well as the strength of their connections. A  $w_{ij} = 0$  means that there is no connection, whereas for a  $w_{ij} < 0$  neuron  $i$  is said to *inhibit* its successor  $j$  with the weight  $|w_{ij}|$ , and for a  $w_{ij} > 0$  neuron  $i$  is said to *excite* its successor  $j$  with the weight  $|w_{ij}|$ .

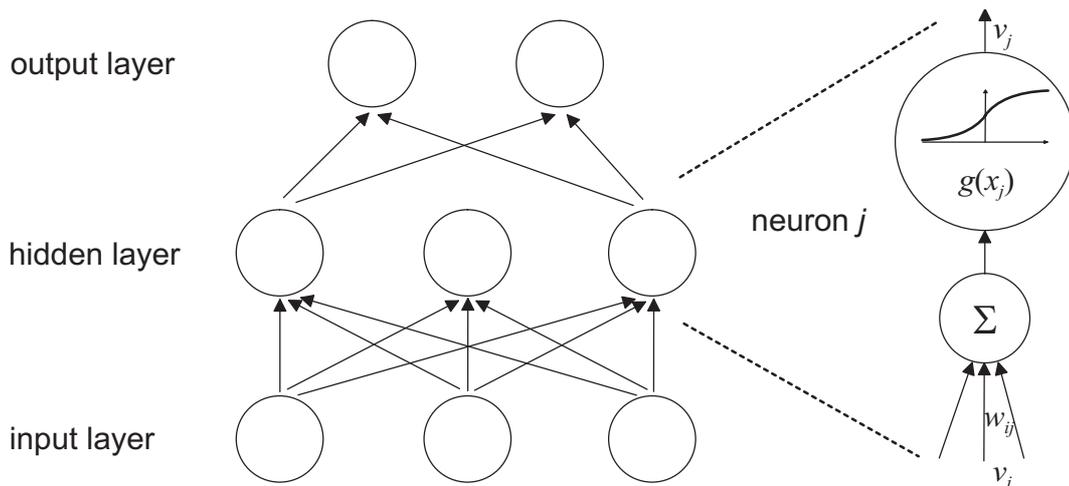
As illustrated in Fig. 4.23, neuron  $j$  first calculates the weighted sum of its inputs  $x_i = w_{ij}v_i$ , where the  $v_i$  are the outputs of its predecessor neurons, and usually adds an external input bias  $u_j$  (for numerical reasons) to calculate the total input value  $x_j$  as

$$x_j = \sum_i x_i + u_j = \sum_i w_{ij}v_i + u_j . \quad (4.7)$$

This total input is then passed to the transfer function, which is a sigmoidal function in many practical MLPs, to calculate the new output of the neuron,  $v_j$ , as

$$v_j = g(x_j) = \frac{1}{1 + e^{-x_j/x_0}} . \quad (4.8)$$

*Multilayer perceptrons* are feed-forward neural networks with error back-propagation. The basic structure of an MLP is depicted in Fig. 4.24, where an example with one hidden layer between the input and output layers is shown; this is usually called a two-layer perceptron as the input layer does not perform any computations. MLPs without hidden layer are often called *simple perceptrons*. In general an MLP may contain an arbitrary number of hidden layers. The neurons of all layers are connected in a fully meshed, feed-forward manner.



**Figure 4.24:** Basic structure of a backpropagation multilayer perceptron.

Obviously, the knowledge of an MLP is mainly coded in the weights of the links, in contrast to conventional algorithms, where it is implied in the procedure of certain calculation steps. Therefore learning or training rules are needed, which specify how the weights should be adapted, so that the network can perform a specific task and its performance can be improved.

To this end there are two operation modes of an MLP: learning and recall.

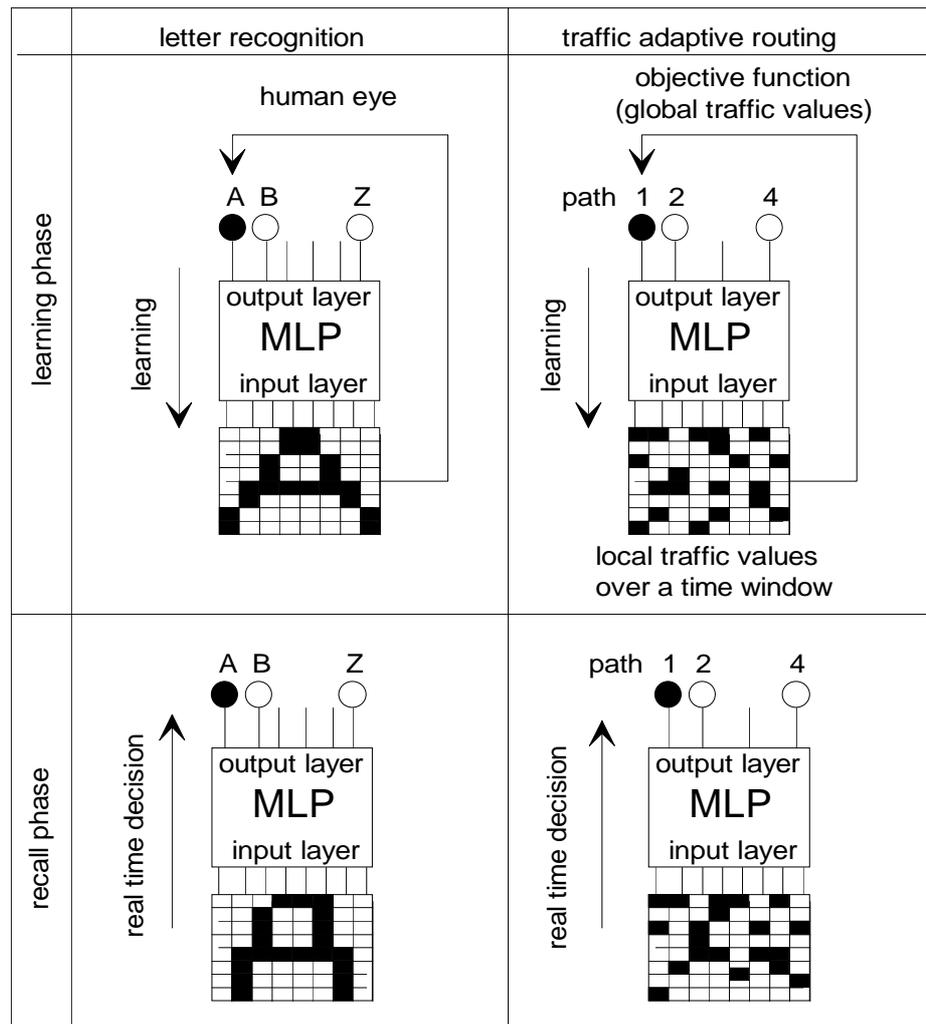
- *Learning:* During the learning phase, a number of  $N$  input/output vector pairs  $\{\mathbf{a}_n, \mathbf{b}_n\}$ ,  $n = 1, \dots, N$ , are presented to the net. It first computes its own output vector according to (4.8), and then compares this computed output vector with the target output vector. This results in an error vector  $\mathbf{e}(w_{ij})$ , which is now used to change the weight matrix  $[w_{ij}]$  according to a learning rule.

The most popular learning algorithm is the so-called *backpropagation algorithm (BPA)*. It is a gradient descent algorithm, where each  $w_{ij}$  is changed by an amount  $\Delta w_{ij}$  proportional to the gradient of  $\mathbf{e}$ ,

$$\Delta w_{ij} = -\eta \frac{\partial \mathbf{e}}{\partial w_{ij}}. \quad (4.9)$$

This algorithm assumes a differentiable transfer function  $g(x)$ , so usually a sigmoidal nonlinearity is used.  $\eta$  is the *learning factor*, which specifies the step width of the gradient descent. The learning phase ends when all input/output pairs to be learned have been presented and the total error is lower than a predefined threshold. After completion of the learning phase the information about the input/output pairs, which represents a mapping, can be regarded as being stored in the weight matrix.

- *Recall:* In the recall mode the MLP can be used as an error-correcting classifier, where the response  $\mathbf{b}_n$  is expected from the net when an input vector of type  $\mathbf{a}_n + \mathbf{r}$  is presented, where  $\mathbf{r}$  is an additive noise component. The features of computing target outputs even if incomplete or disturbed inputs (compared to the learning phase) are presented, are also called the *generalization* and *regression* capabilities of an MLP.



**Figure 4.25:** Comparison of the learning phase and the recall phase of a multilayer perceptron (MLP) type neural network used for letter recognition (left) or traffic adaptive routing (right).

#### 4.4.2.3 Traffic Adaptive Routing with Distributed Multilayer Perceptrons

The proposed NN routing concept is based on distributed multilayer perceptrons with backpropagation learning, one MLP being placed in each satellite. Using the powerful pattern recognition and regression capabilities of MLPs, the basic idea is to let them learn correlations between local traffic patterns and the network-wide traffic distribution. With this “knowledge”, they can perform a source routing in such a manner that the global traffic flows are efficiently shaped to reduce worst-case link loads and to increase average link utilization. Figure 4.25 illustrates the similarities of the onboard routing MLP with a “classical” MLP used for letter recognition; this is due to the fact that the routing decision is essentially based on (traffic) pattern recognition. However, letter recognition is strictly a classification problem, whereas the routing task implies also a kind of regression problem. It should be noted that the MLP input values in the routing case are not binary, but continuous. A significant advantage of the use of MLPs for traffic adaptive routing is that “wrong” decisions do not have such serious effects as for example in letter recognition. Non-optimal decisions up to a certain degree may rather have the advantage of avoiding oscillations in the traffic flows.

Figure 4.26 comprises the fundamental modules of the MLP routing in both, learning and recall phase. Training data for the learning phase can for instance be generated by a deterministic traffic adaptive routing scheme as presented in the previous section. In order to keep the computational complexity for training data generation low, however, for the purpose of our studies we have used a somewhat simplified two-step adaptive scheme based on centralized network knowledge and Dijkstra shortest path calculation (where “shortest” includes load information in the second of two calculation steps). The approach is more precisely explained in Section 6.1.2 in the context of numerical studies.

Generally, in the learning mode more or less sophisticated objective functions in connection with the deterministic algorithm are possible. An intuitively reasonable one (which is used in this study) is to minimize the maximal load occurring on a single link of the selected paths, because this meets the on-board power restrictions most adequately, as discussed in the previous section.

In both the learning and the recall phase, the input pattern of the MLP consists of a time series of traffic values, namely always the outgoing traffic on the four ISLs and on the downlink. The time window provides the MLP the ability to base its decisions not only on an instantaneous situation, but also to exploit the enormous “information” hidden in the gradients of the respective values.

The output pattern contains the coded information about the most appropriate path with regard to the respective objective function. This path is one of the four link disjoint paths preselected by the deterministic routing scheme. The MLP has learned to choose the “best” path in a number of given training situations/patterns. If in the recall phase a similar pattern is applied to its input layer, the MLP computes this best path in real time.

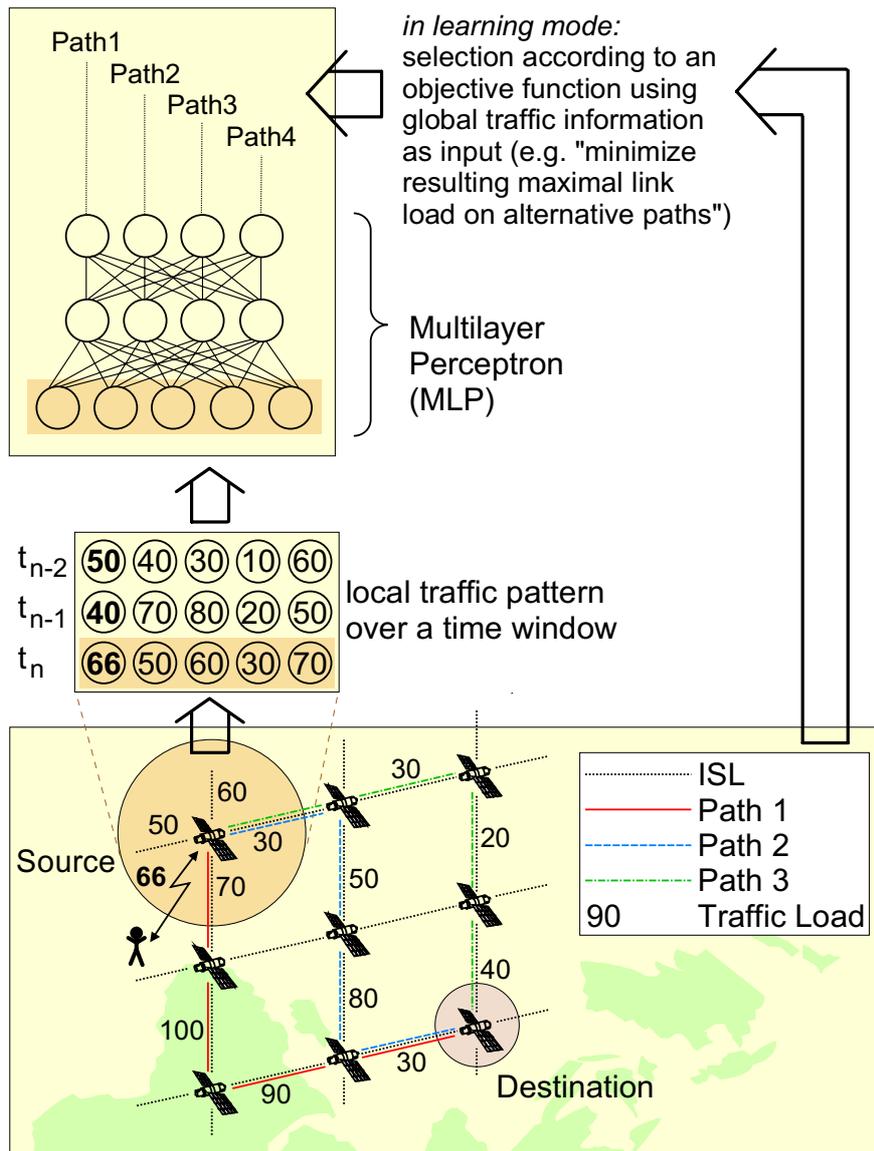
Some more illustrative details of the simulation implementation, and some graphical representations of example MLP networks and input data can be found in Appendix D.

In the following, some representative simulation results from the original extensive study [May96] are reported.

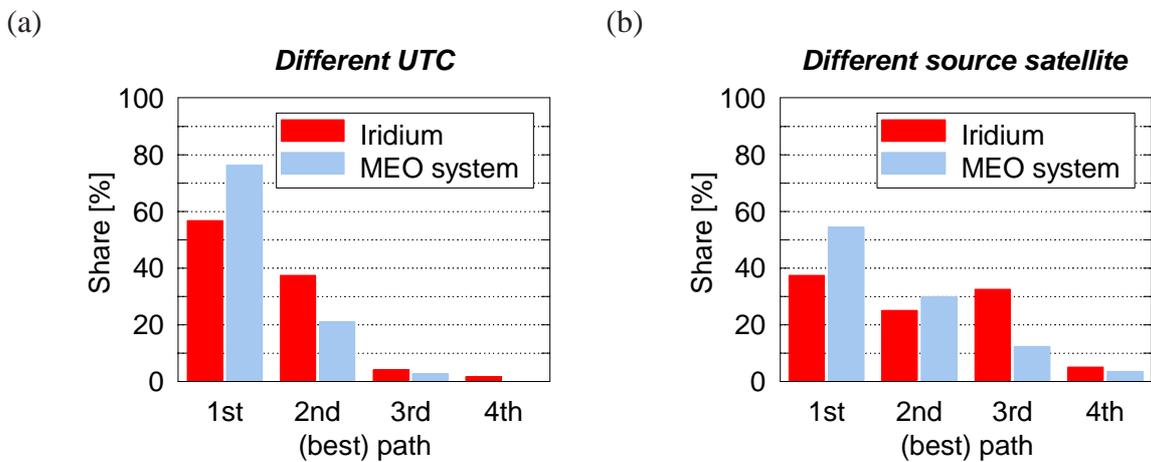
Due to the enormous processing time requirements the reported results and preliminary conclusions have been gained in a simplified simulation scenario with always only one satellite carrying the MLP on board. In all other satellites, the deterministic adaptive algorithm has been implemented. However, maximum effort has been spent on proper definition and selection of reasonable learning and evaluation runs and data sets, in order to guarantee that the results can be extrapolated to the (predicted) performance in the target scenario (all satellites carrying one MLP) at least in qualitative manner.

Figure 4.27 comprises the most important results for the evaluation of MLP routing performance. Evaluation runs have been performed for a trained MLP in two manners: (i) in the recall phase, the MLP works at a different time (UTC), and thus sees series of considerably different input traffic patterns compared to the learning phase; (ii) a similar effect is achieved by placing the MLP in another satellite (at same UTC).

The histograms prove that the MLP performs traffic adaptive routing, since it clearly tends to select those paths which are the optimal ones according to the deterministic reference scheme. The results are even better for a reference MEO system (LEONET in this case, cf. Appendix A for details) than for Iridium. One major reason for that is obviously the regular and physically permanent ISL topology (on the contrary, Iridium’s one is characterized by the seam and on/off-switching of ISLs over polar regions), which reduces the complexity of an appropriate MLP training. Future ISL topologies with such favorable features may thus show an even higher suitability for on-line NN traffic adaptive routing.



**Figure 4.26:** Traffic adaptive ISL routing concept using distributed MLPs; one MLP on board each satellite.



**Figure 4.27:** Performance of an on-board MLP router in two test scenarios: (a) different UTC, (b) different host (source) satellite.

# Chapter 5

## ISL Network Design

*Freedom and constraint are two aspects of the same necessity.*

— ANTOINE DE SAINT-EXUPÉRY, *La Citadelle* (1948)

### 5.1 Introduction and Overview

Network dimensioning is generally closely coupled with network routing in the network design or synthesis process, and it significantly influences both the installation and operating costs of a system. Whereas routing has to determine the path of connections, given the topology and the capacity of transmission and switching equipment, the task of network dimensioning is to determine these capacities assuming a particular routing method [Gir90]. Given the end-to-end traffic matrices, network dimensioning is usually treated as an (iterative) optimization problem with specific network costs as the target function and some constraints on the quality of service (delay, blocking, etc.) to be met by the network.

This chapter focuses on the combined routing and dimensioning problem for the intersatellite link segment of broadband LEO satellite systems.

The routing of complex global traffic flows over dynamic network topologies has already been addressed in Chapter 4 to a level of detail providing necessary input for the dimensioning task. For the capacity dimensioning of ISL networks one may use approaches, algorithms and tools known from terrestrial (ATM) networks to a certain extent, but has to take into account specific additional constraints like the time-variance of the topology and also potentially different target functions adapted to the LEO ISL scenario.

Given the inherent complexity through dynamics of constellation and network topology, the treatment assumes on the other hand a simplified situation on the source traffic side and thus provides a focused study of the very dimensioning *approach* and possible optimization *methods*. In order not to obscure the key issues and to keep the notation manageable, implicitly we also make the unnecessary but simplifying assumption that traffic carried in the network is of single-service type – or that multiservice traffic would not require separation in terms of routing and allocated bandwidths.

This chapter provides a comprehensive treatment of ISL network design research, mainly integrating the earlier focused publications [WFWM01], [WF99, WWFM99], and [WR99, Wer00]; it is

organized as follows. First a pragmatic method to design convenient ISL topologies for connection-oriented operation mode is presented, and a reference topology to be used in the following is derived. Then, reference is made to the earlier discussed ISL routing framework, and some specific interesting features of ISL network paths are highlighted that can be useful for tailor-made dimensioning. Based on that, the network dimensioning task itself is precisely formulated, and candidate dimensioning approaches are developed and discussed in detail.

Corresponding numerical results for a homogeneous traffic scenario are presented and discussed in Section 6.2.

## 5.2 Topological Design

### 5.2.1 Motivation and Satellite Constellation Options

Earlier studies on the routing in ISL networks of polar or near-polar *star* (pattern) [Wal70] constellations (like Iridium [HL95]) have clearly identified the seam between counter-rotating orbits and the on/off-switching of its inter-orbit ISLs as two fundamental drawbacks for the connection-oriented operation [WDV<sup>+</sup>97]. This has stimulated the investigation of moderately inclined *delta* (pattern) [Wal70] constellations employing ISLs [WJLB95, Wer95]. It was shown that such constellations generally provide the possibility to set up a number of inter-orbit ISLs that can be maintained *permanently* with acceptable pointing, acquisition and tracking (PAT) requirements; such link permanence is particularly important for optical ISLs, which seems to be the most promising candidate technology for the operation of future broadband LEO systems. Moreover, permanence of physical ISLs is in general a highly desirable feature in the light of real-time connection-oriented services, as path switching can be completely reduced to such cases where it is inevitable due to handover of the satellites serving the ground users of considered end-to-end connections. This becomes even more striking when considering, for example, the transport of jitter-sensitive ATM cell streams across these networks in space.

M-Star [Mot96] was one of the first commercial system proposals aiming at the promising combination of a delta pattern constellation and optical ISLs. A proper design of the ISL topology to be implemented is a first important step to guarantee efficient networking in the operational system. In the following we present a pragmatic approach to the ISL topology design in delta pattern constellations, deriving a reference topology for M-Star as an example. Figure 5.1 shows the M-Star constellation, and Table 5.1 lists the basic constellation parameters; a complete summary in direct comparison with other reference constellations is given in Table A.1 in the appendix.

### 5.2.2 ISL Topology Design Procedure for Delta Constellations

A closer look at the planar projection of the constellation, in Fig. 5.2, facilitates the first step in the ISL topology design, which is to identify potential links to be reasonably implemented in the network. Due to the perfect symmetry of the constellation it is sufficient to consider generic types of ISLs between satellite 0 and its eastward neighbors at  $t = 0$  as an example. These are then applicable to all other satellite pairs correspondingly. Of course, the implementation of the convenient intra-orbit ISLs 0–1 and 0–5 with both constant link length and fixed pointing angles

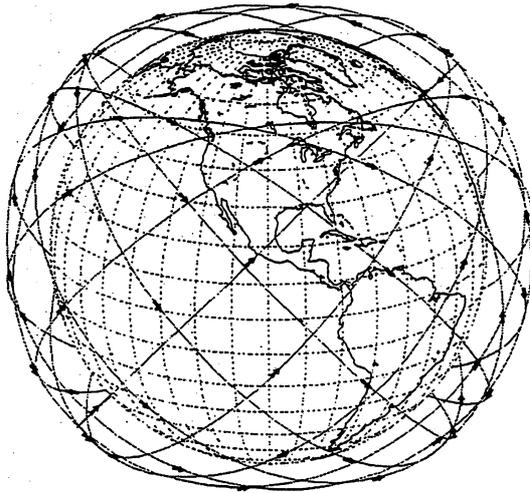


Figure 5.1: M-Star constellation [Mot96].

Table 5.1: M-Star constellation parameters.

Constellation pattern	delta
Orbit classification	LEO
Orbit altitude $h$	1350 km
Orbit period $T$	112 min 41 s
Orbit inclination $i$	$47^\circ$
No. of satellites $N$	72
No. of orbit planes $P$	12
No. of satellites per orbit plane $N_S$	6
Phasing factor $F$	5
Plane spacing at equator $\Delta\Omega$	$30^\circ$
Plane phasing angle $\Delta\omega_p$	$25^\circ$

is obvious. Then one would intuitively envisage links toward the next neighbors on the adjacent orbital plane, the one ahead (0–6) and the one behind (0–11) in terms of phasing, as illustrated in Fig. 5.3. Finally, the same procedure could be applied to partners on the next, second plane; here, the pairs 0–12 and 0–17 are selected because they are the ones with minimum phase difference.<sup>1</sup> So far, this procedure is generally applicable to delta constellations *before* considering any system specific constellation parameters.

The feasibility of these envisaged links has still to be proven taking into account the geometric and technological constraints for a specific constellation. The diagrams in Fig. 5.4 display important geometric data for this purpose. The vertical and horizontal deviation of the pointing angle from flight direction are intuitively called elevation and azimuth in this context. One can notice that both links toward the adjacent orbit show relatively little variation in length and pointing angle (both elevation and azimuth), with smaller length for ISL 0–11 and smaller length variation for ISL 0–6. This means that establishing these ISLs in permanent mode does not introduce severe problems. ISL 0–17 is less attractive but still possible, whereas earth shadowing prevents the implementation of ISL 0–12 according to simple geometric considerations as illustrated in Fig. 5.5. With given mean earth radius  $R_E$ , orbit altitude  $h$  and minimum allowed distance  $d_{\min}$  between an ISL and the earth’s surface (at any time), a maximum possible elevation angle  $\varepsilon_{\max}$  (vertical deviation of “view vector” from tangential plane),

$$\varepsilon_{\max} = \arccos \frac{R_E + d_{\min}}{R_E + h}, \quad (5.1)$$

or a maximum possible link length  $l_{\max}$ , respectively,

$$l_{\max} = 2\sqrt{(R_E + h)^2 - (R_E + d_{\min})^2}, \quad (5.2)$$

can be calculated as effective upper bounds. Considering an acceptable margin for laser link point-ahead or extra attenuation through the atmosphere, one finds that guaranteeing a minimum distance  $d_{\min}$  of 150 km leads for the M-Star orbit altitude  $h = 1350$  km to a maximum allowed length  $l_{\max}$

<sup>1</sup>From the snapshot in Fig. 5.3 the pair 0–16 also seems to be an attractive candidate; however, such “next, second partners” (in terms of phasing) have not been further considered since their attractiveness in general strongly depends on the specific combination of inclination angle and phasing factor values.

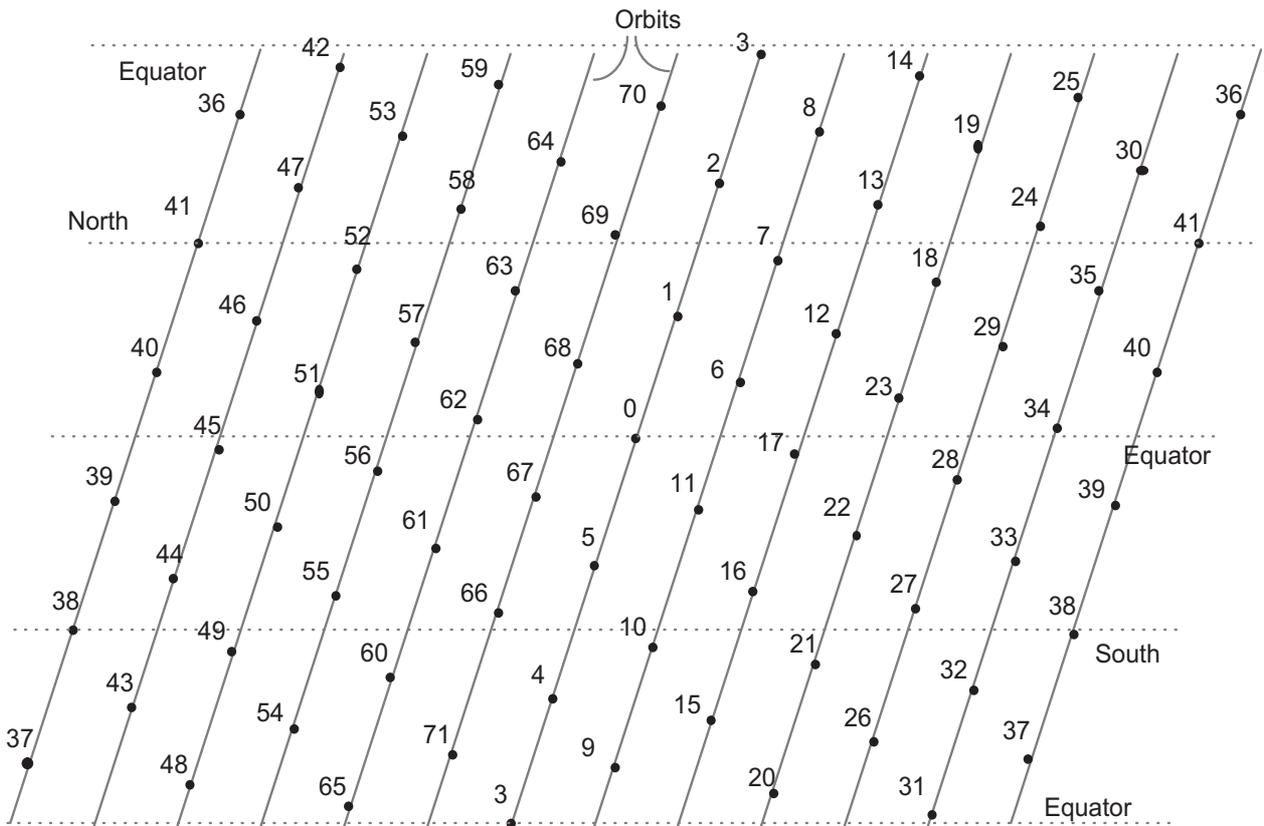


Figure 5.2: Schematic view of the M-Star constellation at  $t = 0$ .

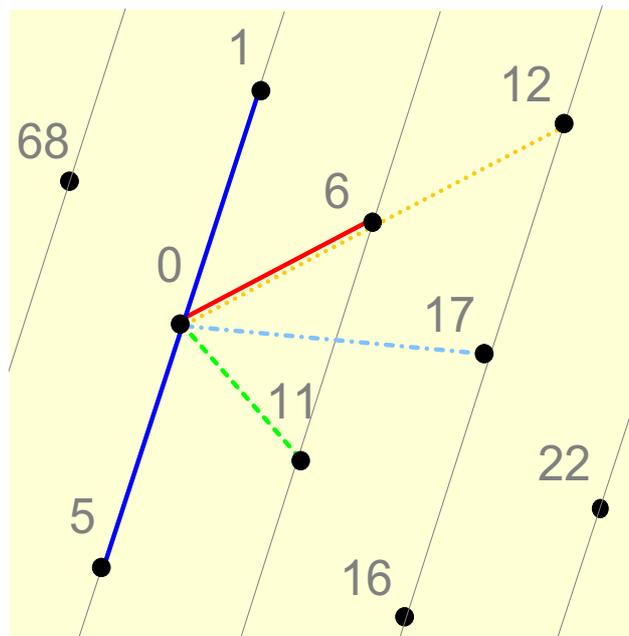
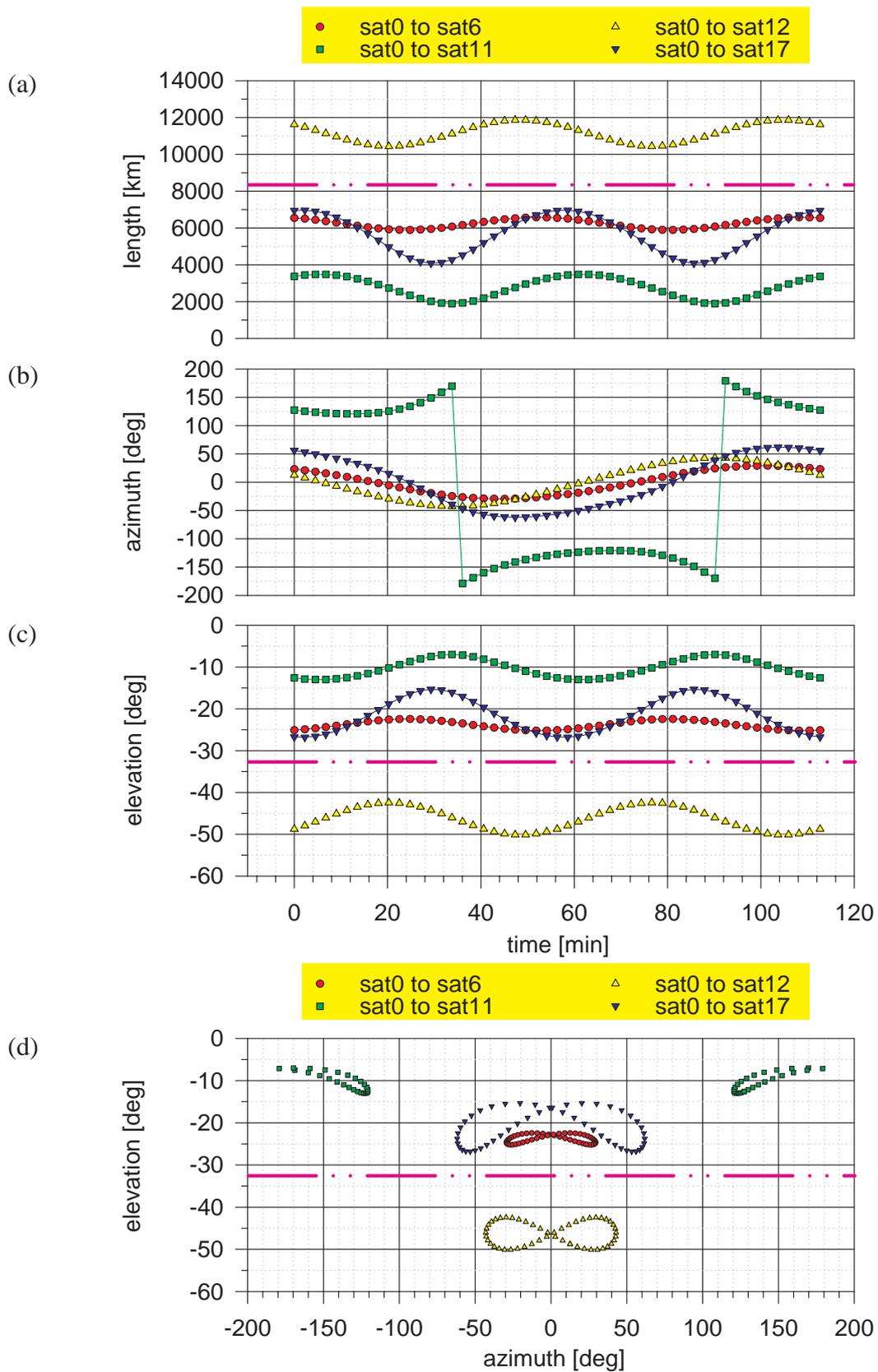
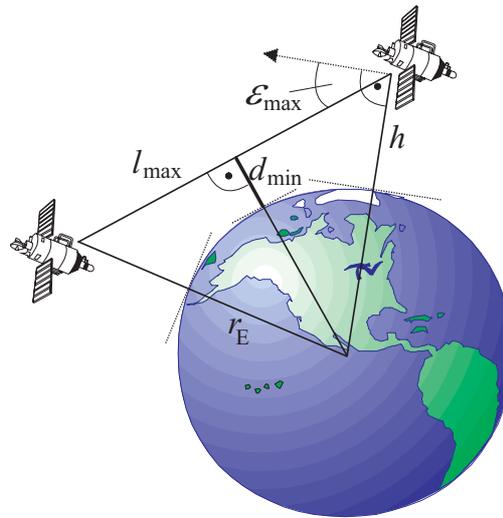


Figure 5.3: Potential generic ISLs in a regular delta constellation. The example for satellite 0 can be extended to all other satellites with their respective partners.



**Figure 5.4:** Geometric feasibility of ISLs in M-Star: time variation of (a) ISL length, (b) horizontal pointing (azimuth) and (c) vertical pointing (elevation); (d) pointing diagram. The dash-dot line represents the bound with respect to earth shadowing.



**Figure 5.5:** Geometric considerations concerning the limitation for maximum ISL length, or maximum ISL elevation, respectively, due to earth shadowing.

of roughly 8270 km for an ISL – or, equivalently, to a maximum allowed elevation  $\epsilon_{\max}$  of about  $32^\circ$  between the two connected satellites. The dash-dot lines in the diagrams indicate these values. ISL 0–12 is clearly above these limits at all instants of time.

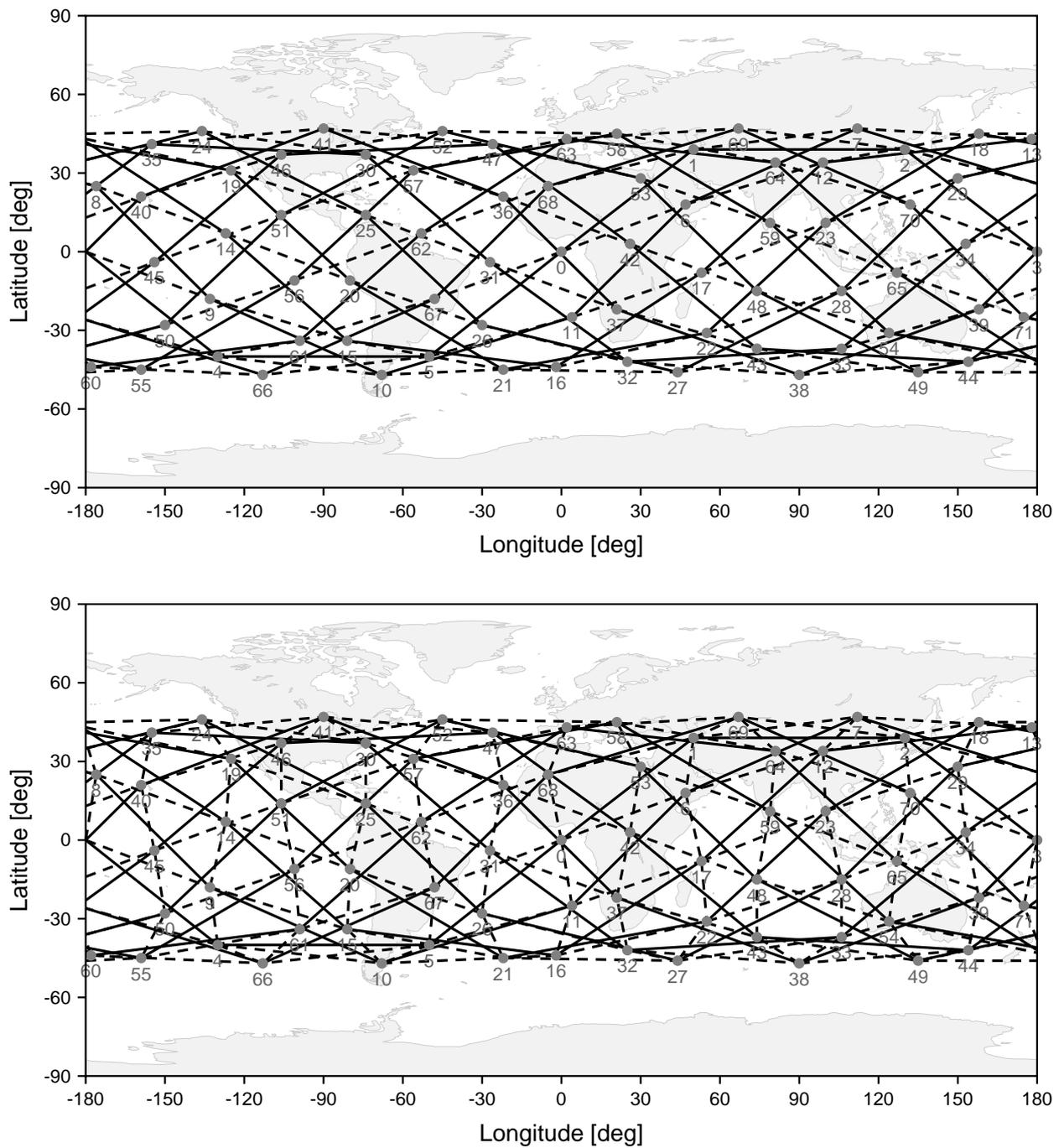
Finally, it should be noted that according to Fig. 5.4(d) the more critical azimuth swivel requirements are well below  $\pm 90^\circ$  for all potential ISLs, so that it should be possible to maintain laser links while slewing through the pointing “wings” [DKFW97].

From the viewpoint of both routing flexibility and potential for optimized dimensioning, a higher degree of meshing in the topology is of course desirable, but it should be traded off versus implementation complexity. For the following, a reference topology T1 is selected, with four bidirectional ISLs per satellite, such as the intra-orbit ones 0–1 and 0–5, and the inter-orbit ones 0–6 and 0–67. In contrast to this, a topology T2 uses six bidirectional ISLs per satellite, such as intra-orbit 0–1, 0–5, and inter-orbit 0–6, 0–11, 0–67, and 0–68. Figure 5.6 displays snapshots of T1 and T2, respectively.

## 5.3 ISL Routing Framework

### 5.3.1 The Combined ISL Routing/Dimensioning Problem

Extensive research on dynamics of traffic flows in LEO ISL networks [Gar96, May96, Kro96, WM97, WKM97, WMMH98] (with representative results presented in this thesis in Sections 4.4 and 6.1.2) has shown the gain to be expected from traffic adaptive routing, or – in other words – from shifting a part of the connections to longer paths in order to balance the flows across the network. However, deterministic on-line routing algorithms [Gar96, Kro96, WM97, WKM97] on a per-call basis or on-line neural network approaches [May96, WMMH98] with arbitrary time resolution have been considered given a classical single-service Erlang/telephone scenario. The approaches generally reveal high complexity, potential danger of uncontrolled oscillations and/or uncontrolled path switching. Moreover, the adaptation to the multiservice broadband scenario is expected to further increase their complexity significantly.



**Figure 5.6:** Schematic M-Star ISL topologies T1 (top) and T2 (bottom): solid lines, intra-orbit ISLs; dashed lines, inter-orbit ISLs.<sup>2</sup>

<sup>2</sup>Note that for illustrative reasons the ISL topology is on purpose only *schematically* overlaid the latitude/longitude map; i.e., the satellites appear at the correct latitude/longitude values of their sub-satellite points, whereas the ISLs would appear as curves rather than straight lines in a correct unprojected map representation.

On the other hand, dimensioning is a task to be performed off-line prior to system operation, and due to the proven importance of the routing for the resulting network traffic flows it must incorporate the (finally on-line) routing decisions in an off-line model. This generally requires that the on-line routing works in a deterministic manner for given topology/traffic situations, so that it is predictable in turn. Given the size and high dynamics of the ISL network, the number of path options should be reasonably restricted; in other words, only a sub-optimal solution with respect to the dimensioning target is possible. This does not only release the dimensioning procedure from excessive modeling and computing requirements, but also keeps the operational complexity for the on-line routing itself limited.

In Section 4.3, *Dynamic Virtual Topology Routing (DVTR)* has been presented as a general concept to provide connection-oriented communication in deterministic dynamic topology environments. Some emphasis has there been laid on counteracting problems caused by switched physical ISLs. In the following, we will recall those “ingredients” of the routing framework that are correlated with the dimensioning task, keeping in mind that the network design considered in the present study is limited to ISL topologies employing permanent, non-switched links only. Figure 5.7 illustrates the central position of the off-line routing framework embedded between topology design and capacity dimensioning.

At the core of the routing framework, the path search is performed by a  $K$ -shortest path algorithm (KSPA) [Shi79] for every OD pair. The single shortest path search task can be formulated as finding the least-cost path  $p(s)$ , i.e. the path with minimum path cost

$$\min_{\forall p(s)} C_{p(s)} = \min_{\forall p(s)} \sum_{i,j} c_{ij}(s) \delta_{(i,j)_s}^{p(s)}. \quad (5.3)$$

Performing this path search for all  $s = \{0, \dots, S - 1\}$  establishes a discrete-time dynamic virtual ISL topology. It consists of an ordered set of  $K$  alternative paths for any OD pair at any step  $s$ . From these  $K$  alternatives, the  $k \in \{1, \dots, K\}$  best ones can be effectively used by the routing, where  $k$  could be simply identical for all OD pairs. This may, however, lead to conflicts with the permanent VPC topology paradigm, as for a fixed  $k$  together with a strict ordering within the  $K$ -path sets, it can happen that a path does not belong to the reduced  $k$ -path set in subsequent steps. Consider  $k = 1$  and a satellite pair separated by one intra- and one inter-orbit hop as an illustrative example. Obviously, there are always two alternative shortest paths of similar length. Due to the variation of the inter-plane ISL distance, one or the other path will be the first one at a given step. If the alternative paths per step are strictly ordered according to their length, and always only the first one is chosen ( $k = 1$ ), this effectively means switching VPCs, although it is not at all mandatory given the permanence of the physical ISL topology. Fortunately, the typical path (delay) characteristics of the considered ISL topologies offer a nice opportunity to cope with this problem without discarding the cost-oriented ordering of sets. This will be shown in the following subsection.

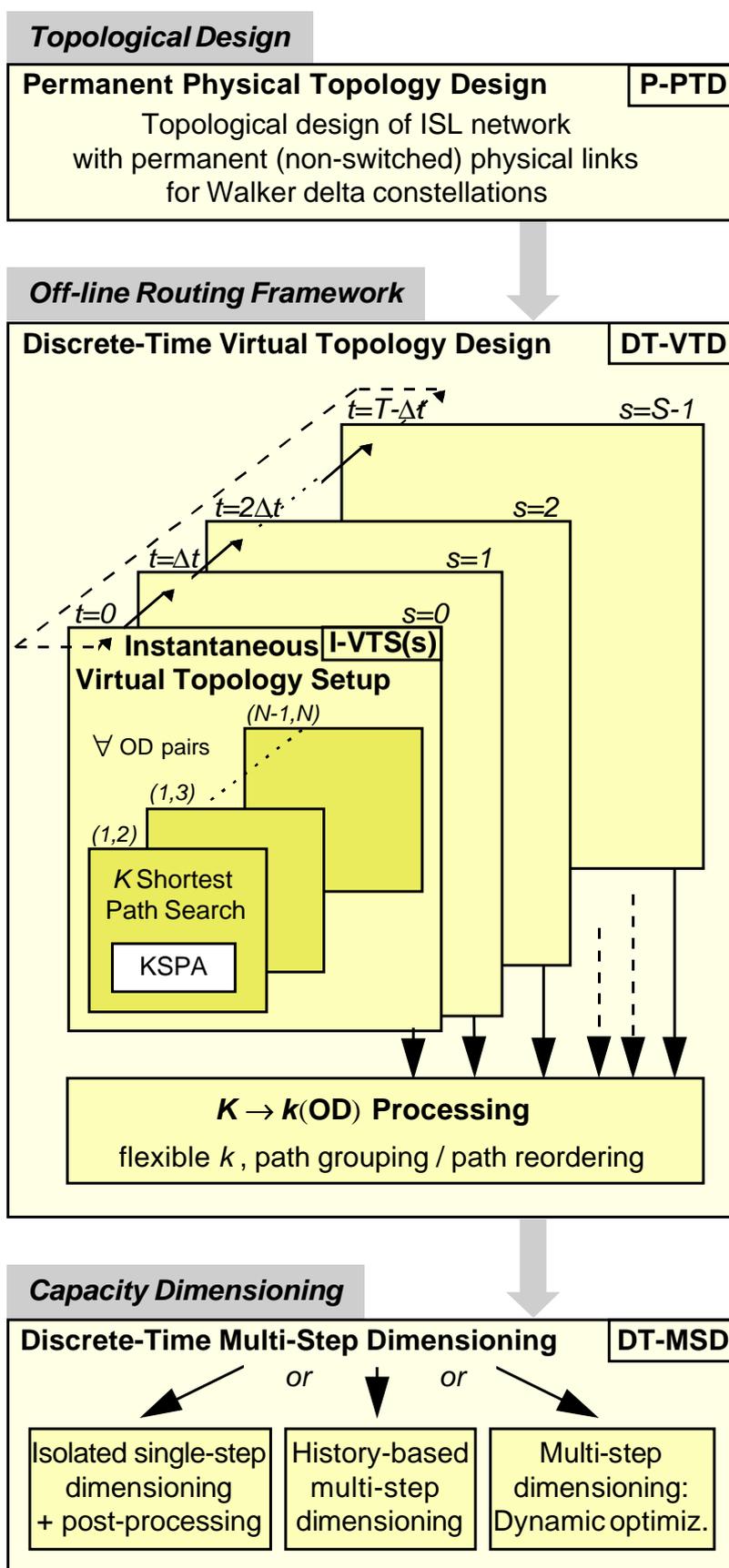
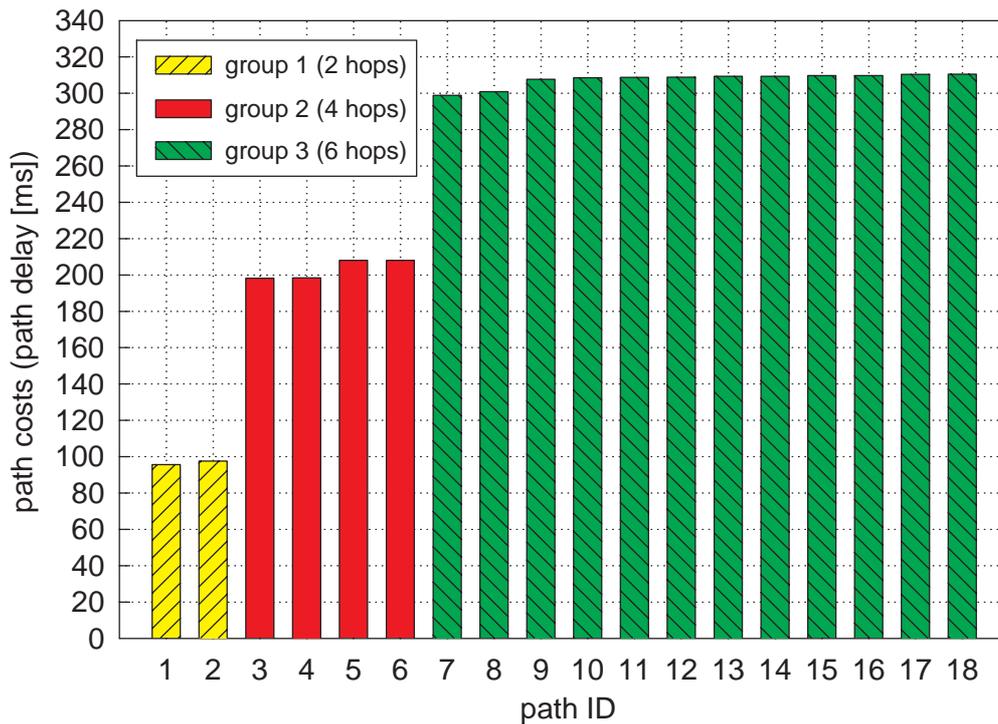


Figure 5.7: Off-line routing framework embedded in the network design process.



**Figure 5.8:** Numerical example for clear path group separation by cost ranges (OD pair sat0-sat7 at step 0). The cost values cover accumulated link propagation delays on the path and additional processing/switching delays of 20 ms per node.

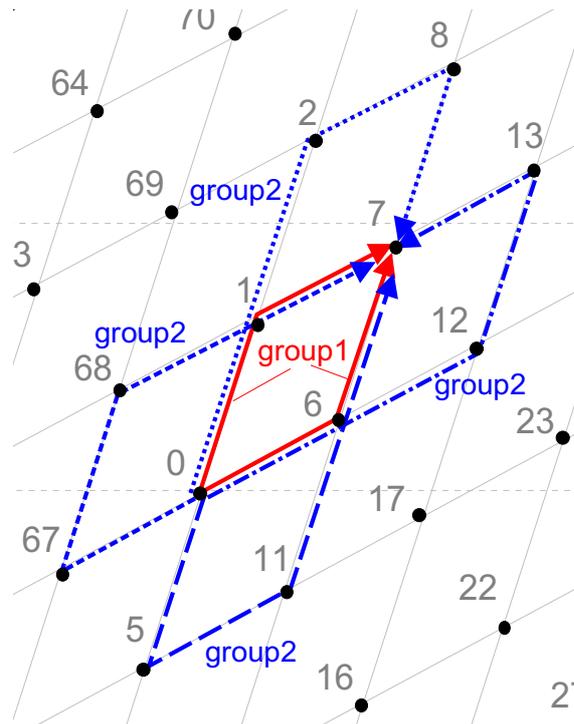
### 5.3.2 Path Grouping Concept

Studying the cost values of all ordered paths in a  $K$ -set for various OD pairs and various steps, one observes that the paths can be easily grouped according to cost ranges, as illustrated for an example OD pair of topology T1 in Fig. 5.8. Typically, one cost range corresponds to a certain number of ISL hops forming the respective paths; the paths belonging to the first two groups of the considered example are displayed in Fig. 5.9, and the relationship between hop count and path group becomes obvious.

It is now of great importance to realize that this clear path grouping extends over all steps, with characteristic group sizes and cost ranges for the respective OD pairs. In other words, the cost ranges of once identified path groups at a certain step never overlap in any of the other steps of the topology period  $T$  as well, and therefore it is guaranteed that always the same paths belong to a specific group. The ordering of paths resulting from KSPA can only vary *within* a group. The reason for this clear path grouping lies in the relatively small distance variation of the selected T1 inter-orbit ISLs over time, as displayed in Fig. 5.4(a) before.

The principle of the path grouping concept can be summarized as follows. It systematically forms groups of OD paths and then selects a specific  $k^* = k(\text{OD})$  for each OD pair in such a manner that always *complete* path groups are contained in the  $k^*$ -path set; with this, the same paths per OD pair are available for routing over all steps, and path/VPC switching can be completely avoided.

Some additional numerical treatment of the path grouping concept will be provided in Section 5.4.2 together with its application in the capacity dimensioning process.



**Figure 5.9:** Illustration of hop-based path grouping: the first two path groups for OD pair 0–7 at step 0.

## 5.4 Capacity Dimensioning

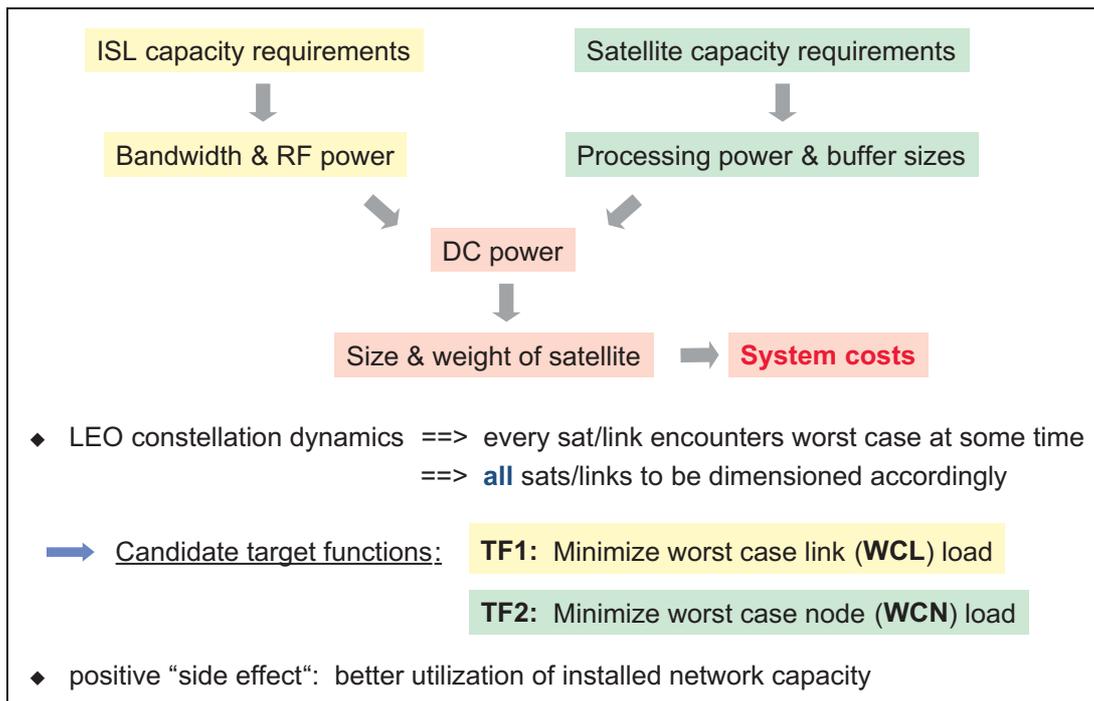
### 5.4.1 Overall Approach and Assumptions

1. *Target function.* The “classical” design process for connection-oriented networks typically aims at minimizing some total or cumulative network costs (often: implementation costs). Considering ATM with the relevant VPC (virtual path connection) concept as an example, the key optimization areas are (i) the definition of the VPC topology, (ii) the VPC capacity assignment, and (iii) the routing rules for OD traffic.

With respect to an appropriate target function for the ISL network dimensioning, there is an important difference compared to the terrestrial case, which is due to the dynamic topology encountered. With the complete satellite constellation periodically orbiting around the earth and thus the source traffic demand, each satellite and each ISL will face roughly the same worst-case traffic load at some time and hence have to be dimensioned with respect to this value; as a result, all satellites in the constellation, including especially the ISL equipment, will be identical. Capacity requirements on an ISL translate into bit rate and thus RF power requirements, whereas capacity requirements for a satellite node mainly drive processing power and buffer sizes. Altogether, this translates into DC power requirements as well as into size and weight of single on-board components and finally of the whole satellite, the latter being a major cost factor of the satellite constellation.

Based on these considerations it is straightforward to formulate the two most appropriate target functions for the ISL network dimensioning:

**TF1** Minimize the worst-case link (WCL) capacity, which is the maximum capacity required on any link at any time.



**Figure 5.10:** Selection of an appropriate target function for ISL network dimensioning.

**TF2** *Minimize the worst-case node (WCN) capacity* which is the maximum capacity required in any node (= satellite) at any time.

These considerations are illustratively summarized in Fig. 5.10.

Of course, some combined metric of TF1 and TF2 is possible as well. In this thesis, we consider the WCL target exemplary. Keeping in mind this minimax type target function as a key feature, the ISL network dimensioning process comprises the following steps and additional assumptions.

2. *Time-discrete approach.* In general, the combined routing/dimensioning task is performed in a time-discrete manner for a series of topology and traffic demand snapshots or steps. This corresponds to what is known as *multi-hour network design* from terrestrial ATM network planning (cf. e.g. [Bau97a, Bau97b]). The additional complexity usually coming in with a multi-hour scenario is in our case immediately relieved by the minimax type target function: we can break down the whole ISL network dimensioning into a number of independent dimensioning tasks, each minimizing the WCL capacity *per step*, since the overall minimized WCL capacity (our final target value) can be simply filtered out as the maximum of all the minimized per-step WCL capacities. The major remaining question is then if an interdependency between subsequent steps is assumed, for instance resulting from QoS driven constraints for in-call rerouting.
3. *Permanent virtual topology.* Overlaid on top of the time-variant physical ISL network is a VPC topology which is permanent over time; this has been discussed in detail in Section 4.3. In contrast to most approaches for terrestrial network design, the VPC topology in the considered ISL case is not subject to optimization itself; rather it is the direct result of assessing appropriate OD path sets based on higher-priority criteria (delay, delay jitter, permanence) aiming at a good performance of connection-oriented operation.

4. *Routing and VPC capacities.* The routing/distribution of given demand pair traffic on the available VPCs is either performed heuristically, according to fixed rules, or treated as an optimization problem which is formulated and solved using linear programming (LP) techniques. Assuming that a limited set of alternative VPCs may be used for splitting the traffic between a specific pair of end nodes, the main optimization parameters are then the splitting factors. Instead of Erlang traffic, we directly operate with given *OD demand pair capacities* assuming that these values have been previously calculated from the Erlang traffic<sup>3</sup> – independently for each demand pair according to the *virtual trunking concept* [DH98, Sie95, Sie96]. The VPC capacities result directly from the splitting of the demand pair capacity on all available VPCs of the OD pair.
5. *Link capacities.* According to the virtual trunking concept, the total capacity of a single link at a given step is determined by summing up all VPC capacities crossing it.

The core of the presented approach, namely the routing/splitting of OD capacities, is looked at in more detail in the following two subsections. There, addressing the target of minimizing the WCL load, three different approaches are considered. Their context and the indicative labeling is displayed in Fig. 5.11. While *equal sharing (ES)* is a heuristic design rule used as a simple reference, both *full optimization (FO)* and *bounded optimization (BO)* transform the WCL minimax problem into a smooth linear minimization problem and solve it using linear programming (LP) techniques. Applying the standard procedures directly to our discretized multi-step scenario will usually consider the complete problem as a series of independent minimization tasks, and select the maximum WCL load from all steps afterwards; this is referred to as *isolated (single-)step* dimensioning in the following. As it will be shown, however, properly taking into account correlation between subsequent steps may be an issue in this context; we address this with a method called *history-based (multi-step)* dimensioning. “True” multi-step dimensioning would require dynamic optimization methods and thus lead to a complexity that seems to be not justified in this context.

## 5.4.2 Isolated Single-Step Dimensioning

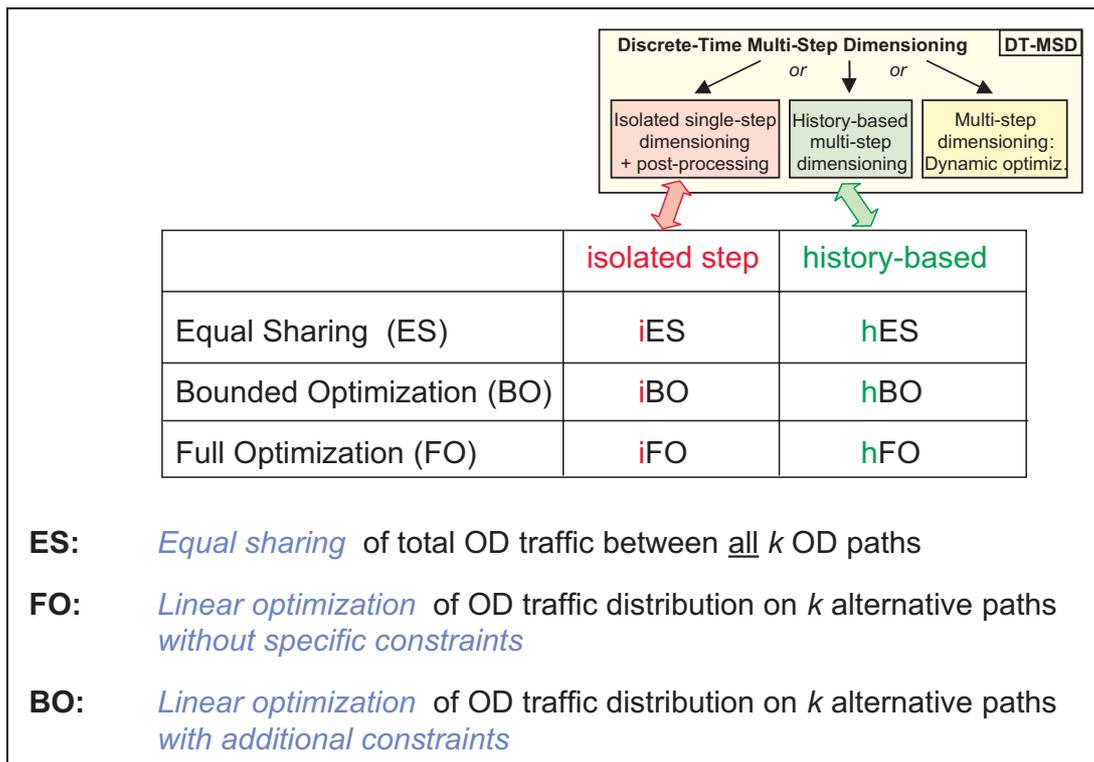
In this first approach based on [Wau98, WWFM99, WWFM01], no correlation between subsequent discretization steps is assumed. The limitations (or modeling deficiencies) of this method will become obvious with the discussion of the new history-based approach in the next subsection. On the other hand, if the “dimensioning error” due to this independence assumption should stay within certain limits, the isolated step dimensioning could be considered as an attractive low-complexity method for a first and quick step in a usually iterative network dimensioning process.

### 5.4.2.1 Heuristic Approach

A simple but pragmatic approach is based on observations made in earlier research [Bur97] that investigated some intuitive rules for traffic routing/distribution in an ISL network. For a simple reference dimensioning we use *equal sharing (ES)*; that is, each OD traffic will be equally distributed over the  $k$  best paths,  $k$  being fixed for all OD pairs in the topology and over time. Intuitively,

---

<sup>3</sup>This calculation could for instance be based on the Erlang-B blocking formula or simply assume a proportional relationship between Erlang traffic and capacity, being linked by the average bit rate of the considered services.

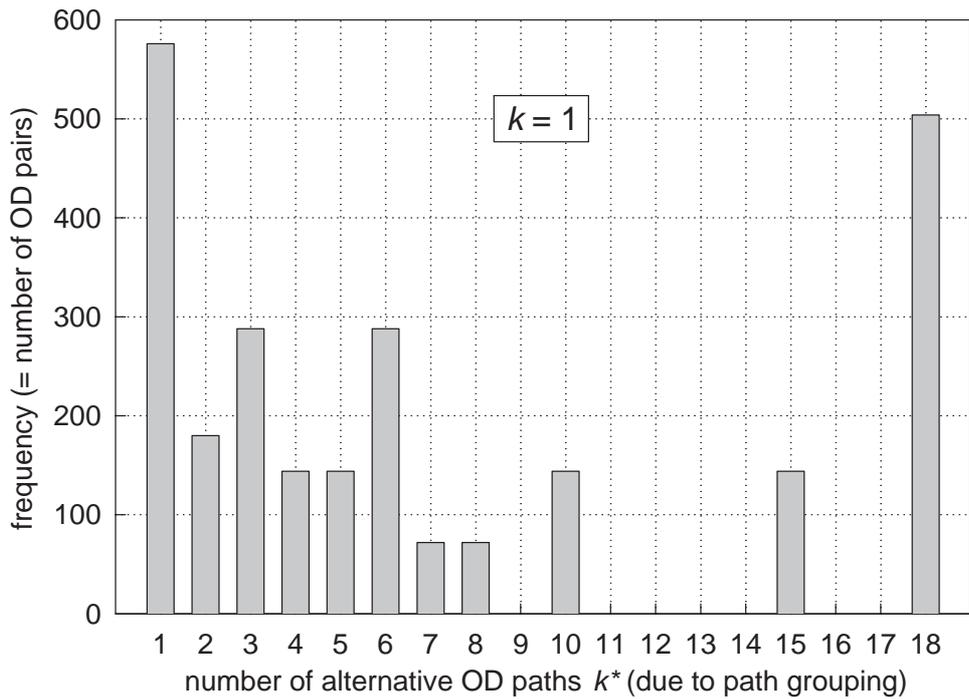


**Figure 5.11:** Overview and labeling of ISL network dimensioning approaches.

one can expect that the ES approach will lead to a decrease of WCL traffic load with increasing  $k$ , just by “somehow” smoothing link load peaks mainly through dividing peak path loads by  $k$ . However, such an effect can not be guaranteed in general, and a potential gain cannot be forecast as a systematic relation between the ES routing rule and the WCL target value does not exist.

Besides the pure ES approach, we have also investigated the combination with path grouping – *equal sharing using path grouping (ES/G)*. In this case, from the network-wide fixed  $k$  an applicable  $k^* = k(\text{OD})$  is determined for each OD pair, being simply the smallest integer value larger than or equal to  $k$  that completes a path group. Using the example from Fig. 5.8 again,  $k = 4$  translates into  $k^* = 6$ , completing path group 2 of the considered OD pair. The positive effect we expect from introducing path grouping knowledge – besides the already mentioned permanent availability of all paths in the  $k^*$ -sets over time – is that the optimization potential may be significantly increased with the number of used paths, whereas the additional paths do not introduce much higher costs than encountered on the last path in the  $k$ -set, which is already used without path grouping.

The increased optimization potential for a given  $k$  – being the lower bound for all  $k^*$  correlated with the OD pairs of the network – can be graphically reflected by a  $k^*$ -histogram as in Fig. 5.12. One may even want to catch the potential of path grouping in one single figure: an obvious and simple possibility is then to use the *average* number of applicable path alternatives per OD pair,  $\overline{k^*}$ , considering all (evenly weighted!) OD pairs in the constellation. However, one has to be quite careful in directly correlating this figure quantitatively with corresponding optimization results, as it does no longer contain the “distribution” dimension available in the histogram. For instance, the example situation illustrated in Fig. 5.12, with a broad range from 1-path to 18-path OD pairs, is simply captured in one value  $\overline{k^*} = 7.27$ . Table 5.2 exemplary lists the resulting  $\overline{k^*}$  for given  $k$  in the case of M-Star topology T1 with hop-based path grouping.



**Figure 5.12:**  $k^*$ -histogram for M-Star topology T1 and  $k = 1$ : number of OD pairs (total 2556) with  $k^*$  alternative paths in the set due to path grouping;  $k^* \leq K = 18$  (from KSPA).

**Table 5.2:** Average number of alternative OD paths,  $\overline{k^*}$ , for M-Star topology T1 with hop-based path grouping and  $K = 18$ .

$k$	1	2	3	4	5	6	7
$\overline{k^*}$	7.27	9.24	9.69	11.52	12.33	13.04	14.76

This treatment of ES/G has so far been quite general – which is, however, necessary from a dimensioning viewpoint in order to form a valid basis that covers all potential practical routing implementations. One obvious candidate for the on-line routing is the following. Using a simple hop-count metric each source node will have a specific number  $k$  of alternative  $h$ -hop routes to a given destination. If  $h$  is the minimum hop count, then  $k$  is the smallest possible  $k^*$  as introduced before. Selecting randomly (or in a round-robin fashion) one of these  $k$  routes for each new OD connection will obviously lead to a balanced load sharing between these routes. This is a pretty simple isolated source routing approach that does not need any *regular* routing update information to be exchanged in the network; the number of (shortest)  $h$ -hop routes and the corresponding routes themselves are deterministic for each OD pair for the whole lifetime, so this information can be stored in each source node. It is beyond the scope of this thesis to go into further details of the routing implementation, but it is obvious that it will need some distributed component as well in order to cope with node or link failures; distributing such failure messages would implicitly inform each source node which of its alternative routes to any destination needs to be canceled from the list of currently available ones, and, possibly, which (stored) backup route could be used instead.

### 5.4.2.2 Optimization Approach

In contrast to the heuristic approach, one may consider a dedicated optimization of a given target function, which is in our case the minimization of the overall WCL capacity. As mentioned before, the major part of this optimization consists of  $S$  dimensioning subtasks, namely minimizing the WCL capacity for all steps  $s$  independently, and the overall WCL capacity is then the maximum of all minimized WCL capacities per step. Using the formulation of the network model and dynamic routing concept presented in Section 4.3 (see Table 4.3 on page 66 for a quick brushing up), we consider one of these subtasks for a given step in the following, without explicitly indexing with  $s$  in order to enhance the readability of the notation.

We consider only permanent physical links in the ISL topology, which means that we have a fixed set of links  $l$  over all steps,  $l \in \{1 \dots L\}$ ; for topology T1,  $L = 2N = 144$ ; for T2,  $L = 3N = 216$ . The offered capacity  $n_w$  per OD demand pair  $w$  is distributed among the  $k$  shortest paths  $p$  selected from the ordered set of existing paths,  $P_w$ , so that each path carries a certain share  $n_p$  of the total demand pair capacity,

$$n_w = \sum_{p \in P_w} n_p. \quad (5.4)$$

The required capacity  $n_l$  of a link  $l$  is obtained as the sum of the capacities  $n_p$  of all paths containing this link,

$$n_l = \sum_{w \in W} \sum_{p \in P_w} \delta_l^p n_p, \quad (5.5)$$

where  $W$  denotes the set of all OD pairs and  $\delta_l^p \in \{1, 0\}$  indicates if path  $p$  uses link  $l$  or not.

Our objective to minimize the maximum required capacity on a single physical link can be expressed as a minimax optimization problem:

$$\max_l \{n_l\} = \max_l \left\{ \sum_{w \in W} \sum_{p \in P_w} \delta_l^p n_p \right\} \rightarrow \min, \quad (5.6)$$

subject to the linear constraint Eq. (5.4) for all demand pairs  $w$ .

In order to apply standard LP optimization techniques, the original minimax problem can be transformed into a smooth linear minimization problem by introducing a new scalar optimization variable  $n_{\max}$ , which is an upper bound on all  $n_l$ , and formulating the corresponding link capacity bounds as inequality constraints. Together with the set of equality constraints resulting from Eq. (5.4), the complete formulation of the LP optimization problem becomes

**iFO**

Minimize the WCL capacity,

$$\min_{\mathbf{n}_p} n_{\max}(\mathbf{n}_p), \quad (5.7)$$

subject to (i) the link capacity bounds

$$n_{\max} \geq n_l = \sum_{w \in W} \sum_{p \in P_w} \delta_l^p n_p \quad \forall l \quad (5.8)$$

and (ii) the OD/path capacity requirements

$$n_w = \sum_{p \in P_w} n_p \quad \forall w. \quad (5.9)$$

The optimization variables assembled in the vector  $\mathbf{n}_p$  are the shares of total OD capacity carried by each path belonging to the OD pair, or equivalently, the *splitting factors* that determine the OD capacity split into its correlated paths. Assuming for instance a fixed number of  $k = 3$  alternative paths for all  $N(N - 1)/2$  OD pairs in our M-Star reference topology, we end up with a number of  $kN(N - 1)/2 = 7668$  optimization variables. In comparison, the number of equality constraints according to Eq. (5.9) equals the number of OD pairs,  $N(N - 1)/2 = 2556$ , and Eq. (5.8) adds  $L = 2N = 144$  upper bounds for the link capacities. Obviously, a larger  $k$  or  $k^*$ , respectively, increases the optimization potential (and the computational complexity) with the number of optimization variables, whereas the number of constraints remains fixed for a given constellation ( $N$ ) and ISL topology ( $L = L(N)$ ).

So far, we have not introduced any specific constraint on the share of the total traffic that one path is allowed to carry. As a consequence, a single path may convey the complete offered OD traffic alone, whereas other paths may remain empty. In the following, we refer to this approach as *full optimization (FO)* which is consequently reflected in the label iFO for the isolated step version.

Although iFO certainly leads to the maximum possible WCL load reduction per step, there are some reasons – for example, consequences for operation in failure situations, potentially high load variations on single links from step to step, etc. – to introduce an additional linear constraint in the form of an upper bound  $\alpha$  for the normalized share of the OD traffic one path is allowed to carry, thus following an approach which is referred to as *bounded optimization (BO)* – with parameter  $\alpha$  – in the remainder of the thesis:

**iBO**

Eqs. (5.7)-(5.9) from **iFO** and the additional upper path capacity bounds

$$0 \leq n_p \leq \alpha n_w, \quad \alpha \in [1/k \dots 1] \quad \forall p, \forall w. \quad (5.10)$$

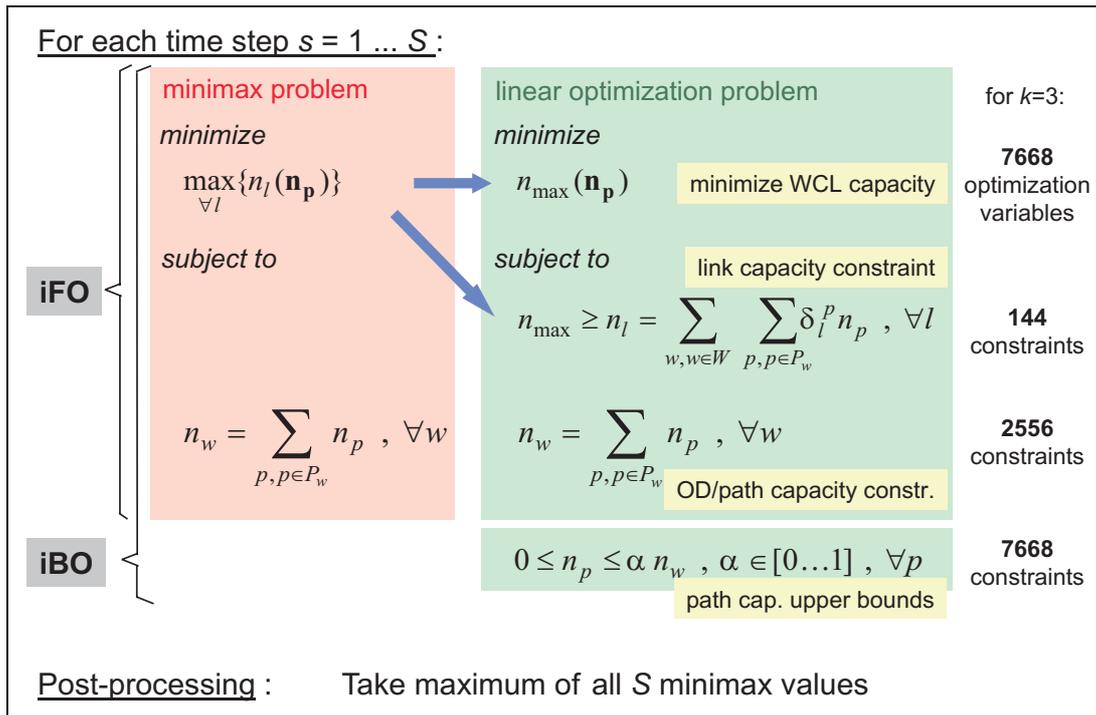
Since an additional constraint is formulated for each  $n_p$ , the number of path capacity bounds is identical to the number of optimization variables, and is thus 7668 for our numerical example. Note that for the minimum possible  $\alpha$ ,  $\alpha = 1/k$ , the BO approach becomes identical with the ES one as the OD traffic is effectively “forced” to split into equal shares on the  $k$  available paths by reducing the (remaining) optimization freedom to zero.

Figure 5.13 comprises at a glance the rationale of the dimensioning approaches described so far, highlighting the transformation of the original minimax optimization problem into a smooth linear minimization problem.

Both, iFO and iBO can additionally be combined with the path grouping concept; these approaches are identified by the corresponding acronyms iFO/G and iBO/G, respectively. Consequently, in the following also the heuristic equal sharing approaches introduced before are in the isolated step case more precisely labeled iES and iES/G, respectively.

It is again emphasized that all the above considerations on optimization are valid for a single step in the dynamic scenario, and are consequently applied to each single step independently. An advanced but highly complex approach would be to include the time dependence of subsequent steps in the optimization; this would lead to a dynamic optimization problem [Pap96], which is clearly beyond the scope of a first dimensioning approach as adopted in this work.

Like for the iES/G approach, it should be helpful to indicate how the optimized dimensioning corresponds to a practical routing implementation. Let us consider iFO/G or iBO/G as example,

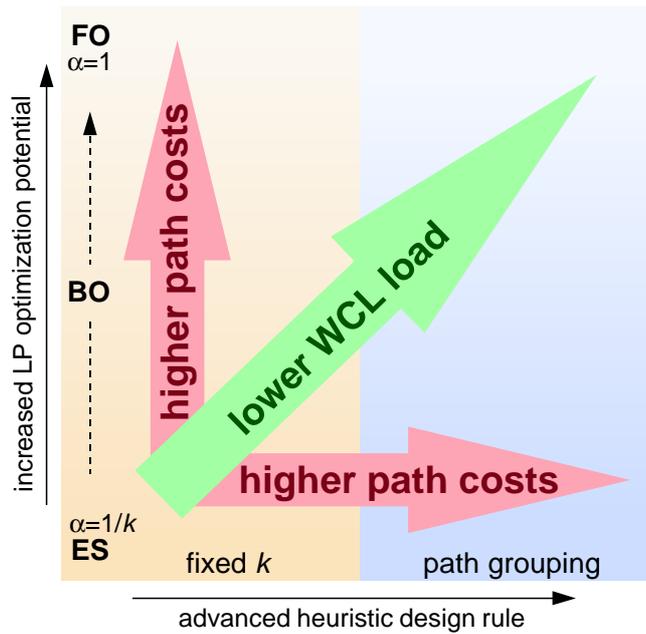


**Figure 5.13:** Overview of the isolated single-step dimensioning approaches based on the transformation of the minimax optimization problem into a smooth linear minimization problem.

and let us again assume a simple hop-count metric to identify the first group of  $k^*$  (shortest)  $h$ -hop routes to all destinations, the information being stored in all source nodes. The outcome of the optimization process in this case is a time series of  $k^*$ -tuples of load sharing coefficients defining the shares of “incoming” OD traffic to be routed on each of the  $k^*$  alternative paths for the respective time interval. A possible candidate for the corresponding on-line routing is fixed alternate routing, where upon a call request the  $k^*$  alternative routes are checked in a round-robin fashion for available capacity, and the call uses the first non-blocking route. It was shown in [Sie95] that fixed alternate routing automatically leads to load sharing values close to the coefficients resulting from the dimensioning procedure assuming virtual trunking. In operation, of course, the load sharing values will slightly vary around the corresponding dimensioning coefficients, just as the real OD traffic values vary around the “nominal” OD traffic matrices which have been input to the dimensioning optimization. In other words, the presented on-line routing implementation will react adaptively to deviations of the OD traffic flows from the nominal case and thus achieve lower blocking values than with fixed nominal coefficients.

In conclusion: for a given nominal traffic scenario, any LP optimization will on principle come up with better results than the equal sharing approach; to which extent this theoretical gain can be actually achieved in operation simply depends on how far the “real-life” traffic situations vary from the nominal case, which is essentially a prediction.

Before looking at numerical results it is worthwhile to formulate the expectations in the qualitative performance of the two major dimensioning philosophies – heuristic design rules and (LP) optimization – and their possible combination. As illustrated in Fig. 5.14, one will certainly expect that the value of the target parameter *WCL load* decreases with growing potential of both approaches, and there is some indication that a solution combining the most advanced forms of both would perform best. On the other hand, the *path costs* as major trade-off parameter should increase in



**Figure 5.14:** Expected behavior of the target parameter *WCL load* and the major trade-off parameter *path costs* in the dimensioning “square” formed by the “orthogonal” approaches (x) heuristic design rules and (y) LP optimization.

both cases, whereas it seems a bit speculative at this point to predict that the path costs would also clearly tend towards highest values in the combined approach.

### 5.4.3 History-Based Multi-Step Dimensioning

Following [Rév99, WR99, Wer00], this approach forms a special implementation of the *discrete-time multi-step dimensioning* task displayed as bottom block in Fig. 5.7. It stands between the simple isolated single-step approach and a so-called “true” multi-step dimensioning which would call for a highly complex dynamic optimization [Pap96, GMW81] solution. In contrast to the latter, the history-based method achieves the “multi-step capability” at an acceptable price in terms of complexity, by linking an arbitrary number of subsequent single steps in a particular non-trivial way.

The isolated step approach suffers from a major modeling deficiency: by assuming *uncorrelated* demand pair capacities in subsequent steps it does effectively neglect the “history” of single calls making up these instantaneous cumulative capacities. Specifically, in this approach it is implicitly assumed that calls are freely reroutable in each step according to the isolated optimization result. Consequently, the isolated step optimization results are usually “too good” as soon as any QoS requirements (which are usually defined from the end user connection perspective) have to be fulfilled, as this in turn reduces the optimization freedom. Two specific call situations must be taken into account:

1. A call remains within the same OD satellite pair in the ISL subnetwork from one step to the next. In this case the call should stick to the OD path chosen in order to avoid uncontrolled path delay offset, increased call dropping probability and unnecessary handover signaling.

2. A call has to be switched to a new OD pair from one step to the next just because of a handover at the source and/or destination satellite. Being a “new” call for the new OD pair, the call could be freely routed (according to isolated optimization), but the new route selection could as well be deterministic, for instance according to the QoS requirement to minimize delay offsets during path switching. However, in the course of this thesis we do not consider any QoS restriction for this rerouting case, postulating that all potential criteria reflecting such issues have already been taken into account during the setup of the off-line routing framework.

Based on these considerations, the demand pair  $w$  capacity is modeled as generally consisting of three components,

$$n_w(s) = n_w^{\text{new}}(s) + n_w^{\text{rem}}(s) + n_w^{\text{rer}}(s), \quad (5.11)$$

where  $n_w^{\text{new}}(s)$  is the new generated traffic (in step  $s$ ), and

$$n_w^{\text{rem}}(s) = f(n_w(s - \tau)) \text{ , } \tau = 1, \dots, \hat{\tau}_{\text{old}} \quad (5.12)$$

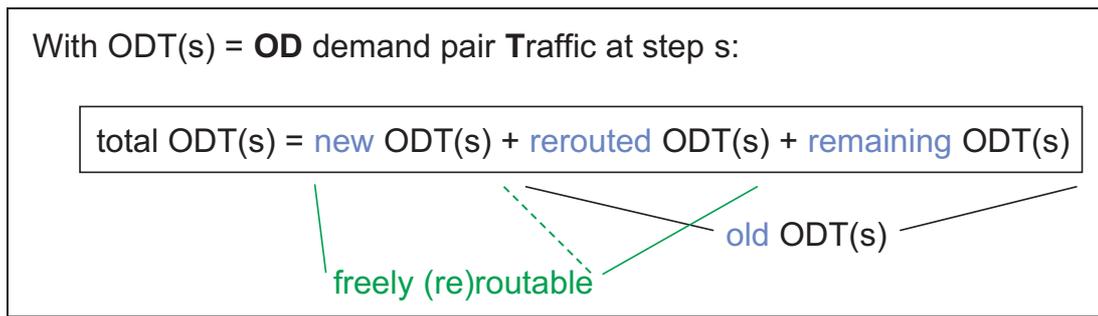
is that part of all earlier generated OD  $w$  traffic (i.e. calls of age  $\tau = 1, \dots, \hat{\tau}_{\text{old}}$ ) that has not yet expired and that remains with the considered OD pair  $w$ .  $\hat{\tau}_{\text{old}}$  is an upper age limit for calls to be considered.

Finally,  $n_w^{\text{rer}}(s)$  identifies any earlier generated traffic that has not yet expired and that is rerouted *towards* the considered OD pair  $w$  (wherever from). Obviously, in a real scenario the incoming and outgoing rerouted calls of an OD pair will be a direct consequence of the satellite handover procedures at source and/or destination satellite, given the constellation dynamics and the user distribution, and thus be figures to be given as input to the dimensioning for an operational system. In particular, incoming and outgoing rerouted traffic for an OD pair need *not* to be of the same amount. As the intention here is the study of the dimensioning approaches themselves rather than the dimensioning of concrete systems under particular traffic scenarios, we restrict ourselves to a completely homogeneous OD traffic scenario in the following. In this case, the incoming equals the outgoing rerouted traffic in the steady state, and is therefore like the remaining traffic a function of the traffic components from earlier steps:

$$n_w^{\text{rer}}(s) = f(n_w(s - \tau)) \text{ , } \tau = 1, \dots, \hat{\tau}_{\text{old}}. \quad (5.13)$$

The choice of a reasonable upper age limit  $\hat{\tau}_{\text{old}}$  to be considered depends on the discretization step width  $\Delta t$ , the call holding time characteristics and the satellite handover model, as will be shown along with the presentation of the numerical studies in Section 6.2.3.

For all calls from step  $s - 1$  it is decided according to the call model if they are still active in step  $s$  or not. Without losing the split into different ages, all still active calls in  $s$  are either classified “rerouted” or “remaining” according to the rerouting model. Given both, a concrete call model and a rerouting model, in the steady state a characteristic stationary sharing between the three classes is achieved. Whereas the call model alone determines the ratio between new and old traffic, the rerouting model together with the call model subdivides the old traffic into rerouted and remaining traffic, Fig.5.15. Some more illustrating details are given in the context of the numerical studies in Section 6.2.3.



**Figure 5.15:** Instantaneous OD demand pair traffic shares.

### 5.4.3.1 Heuristic Approach

The hES approach is identical with the iES one, except that remaining traffic is kept on the path once chosen, even if this path may no longer be in the set of currently available ones in step  $s$ . This means that the equal sharing rule is only applied to the new and rerouted traffic components in the respective step, using the current alternative paths.

### 5.4.3.2 Optimization Approach

As the path capacities  $n_p$  are proportionally coupled with the OD capacities  $n_w$ , Eq. (5.9), the classification of OD capacity shares – into new, remaining, and rerouted ones, Eq. (5.11) – applies to the path capacities accordingly,

$$n_p(s) = n_p^{\text{new}}(s) + n_p^{\text{rem}}(s) + n_p^{\text{rer}}(s). \quad (5.14)$$

The optimization potential is now reduced since remaining traffic cannot be freely routed but has to stick to the path once chosen. This can be formulated as an additional lower bound for the cumulative path capacities (our optimization variables):

**hFO**

Eqs. (5.7)-(5.9) from **iFO** (page 102) and the additional lower path capacity bounds

$$n_p^{\text{rem}}(s) \leq n_p(s) \quad \forall p, \forall w. \quad (5.15)$$

In other words, the components of the traffic which can be freely routed in the considered step to achieve the minimax goal are restricted to

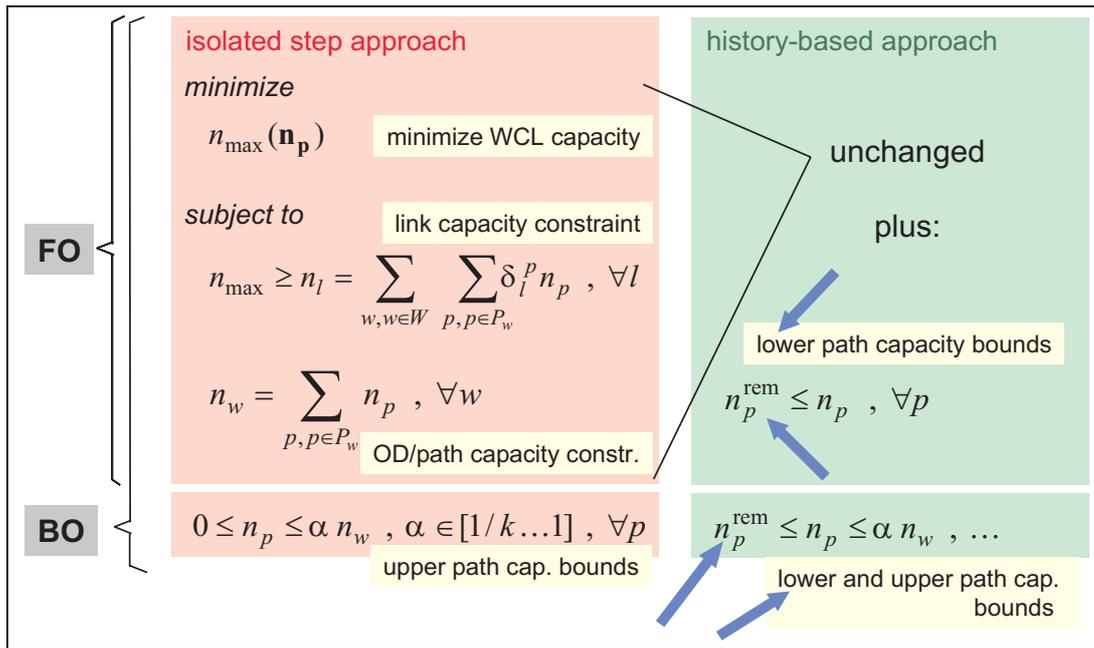
$$n_p^{\text{opt}}(s) = n_p(s) - n_p^{\text{rem}}(s) = n_p^{\text{new}}(s) + n_p^{\text{rer}}(s). \quad (5.16)$$

Finally, in straightforward extension from the iBO approach, the history-based optimization with specific upper path capacity bounds is formulated as

**hBO**

Eqs. (5.7)-(5.9) from **iFO** (page 102) and the additional lower and upper path capacity bounds

$$n_p^{\text{rem}}(s) \leq n_p(s) \leq \alpha n_w(s), \quad \alpha \in [1/k \dots 1] \quad \forall p, \forall w. \quad (5.17)$$



**Figure 5.16:** From isolated step optimization to history-based optimization.

From (5.17) one can also notice that there is an implicit impact of the traffic model characteristics on the lower bound for feasible values of  $\alpha$ . Whereas theoretically  $\alpha$  may be as low as  $1/k$  (like in the isolated step approach), it can be practically necessary to increase this lower bound for a particular network and traffic scenario, sticking to a network-wide and step-independent  $\alpha$ : this is the case if in the course of the development of effective network traffic flows there is at least one situation  $(p, w, s)$  where  $n_p^{\text{rem}}(s)/n_w(s) > 1/k$ . Intuitively, this can for instance happen when for some reason the remaining traffic on a particular path is high whereas it is low or zero on all other paths of the same OD pair, and at the same time the OD pair receives low or zero new and incoming rerouted traffic. However, for the homogeneous network traffic load and for the call and rerouting models considered in the course of this thesis,  $\alpha = 1/k$  is always a feasible value.

Figure 5.16 summarizes at a glance the relevant changes in extending the isolated single-step approach towards the history-based multi-step optimization.

Concluding this section on ISL network capacity dimensioning, Tables 5.3 and 5.4 summarize all presented approaches for isolated step and history-based dimensioning, respectively. Comparing the notation in the two tables, it should be noted that indexing with the time step  $s$  is generally omitted for a better readability, but deliberately used for the capacities  $n_w, n_l$  and  $n_p$  (as well as their components) in the history-based approach (Table 5.4) *only*, in order to highlight the crucial difference compared to the isolated step approach: in the history-based approach these values at step  $s$  include a new time or step “dimension”, namely their own history, whereas they are completely independent instantaneous values in the isolated step approach. Referring to these tables should be particularly helpful for the interpretation of the graphs that are presented in Section 6.2.

**Table 5.3:** Overview of the proposed isolated step dimensioning approaches.

	isolated step dimensioning	
	without path grouping	with path Grouping
	<p><math>k \in \{1 \dots K\}</math> alternative paths are available; <math>k</math> is identical for all OD pairs <math>w</math> and fixed for all time steps</p> <p><math>\implies P_w</math> always contains the <math>k</math> shortest paths <math>p(w)</math></p>	<p><math>k^* \geq k</math> alternative paths are available; <math>k^*</math> depends on the OD pair <math>w</math> and is therefore denoted <math>k_w^*</math>; all <math>k_w^*</math> remain fixed for all time steps</p> <p><math>\implies P_w</math> always contains the <math>k_w^*</math> shortest paths <math>p(w)</math></p>
<b>Equal Sharing</b>	<p><b>iES</b></p> <p>Calculate the WCL capacity <math>n_{\max}(\mathbf{n}_p)</math> as</p> $n_{\max} = \max_l \{n_l\} = \max_l \left\{ \sum_{w \in W} \sum_{p \in P_w} \delta_l^p n_p \right\},$ <p>where  <math>n_p = \frac{1}{k} n_w \quad \forall p \in P_w, \forall w</math>  and  <math>P_w = \{p_1(w), p_2(w), \dots, p_k(w)\}.</math></p>	<p><b>iES/G</b></p> <p>where  <math>n_p = \frac{1}{k_w^*} n_w \quad \forall p \in P_w, \forall w</math>  and  <math>P_w = \{p_1(w), p_2(w), \dots, p_{k_w^*}(w)\}.</math></p>
<b>Full LP Optimization</b>	<p><b>iFO</b></p> <p>Minimize the WCL capacity,</p> $\min_{\mathbf{n}_p} n_{\max}(\mathbf{n}_p),$ <p>subject to (i) the link capacity bounds</p> $n_{\max} \geq n_l = \sum_{w \in W} \sum_{p \in P_w} \delta_l^p n_p \quad \forall l$ <p>and (ii) the OD/path capacity requirements</p> $n_w = \sum_{p \in P_w} n_p \quad \forall w,$ <p>where  <math>P_w = \{p_1(w), p_2(w), \dots, p_k(w)\}.</math></p>	<p><b>iFO/G</b></p> <p>where  <math>P_w = \{p_1(w), p_2(w), \dots, p_{k_w^*}(w)\}.</math></p>
<b>Bounded LP Optimization</b>	<p><b>iBO</b> = iFO plus the additional upper path capacity bounds</p> $0 \leq n_p \leq \alpha n_w \quad \forall p, \forall w,$ <p>where <math>\alpha \in [1/k \dots 1]</math> and  <math>p \in \{p_1(w), p_2(w), \dots, p_k(w)\}.</math></p>	<p><b>iBO/G</b> = iFO/G plus the additional upper path capacity bounds</p> $0 \leq n_p \leq \alpha n_w \quad \forall p, \forall w,$ <p>where <math>\alpha \in [1/k_w^* \dots 1]</math> and  <math>p \in \{p_1(w), p_2(w), \dots, p_{k_w^*}(w)\}.</math></p>

**Table 5.4:** Overview of the proposed history-based dimensioning approaches (light gray entries are identical with the corresponding ones in the isolated step approach).

	<b>history-based dimensioning</b>	
	without path grouping	with path Grouping
	$k \in \{1 \dots K\}$ alternative paths are available; $k$ is identical for all OD pairs $w$ and fixed for all time steps $\implies P_w$ always contains the $k$ shortest paths $p(w)$	$k^* \geq k$ alternative paths are available; $k^*$ depends on the OD pair $w$ and is therefore denoted $k_w^*$ ; all $k_w^*$ remain fixed for all time steps $\implies P_w$ always contains the $k_w^*$ shortest paths $p(w)$
<b>Equal Sharing</b>	<b>hES</b> Calculate the WCL capacity $n_{\max}(\mathbf{n}_p)$ as $n_{\max} = \max_l \{n_l(s)\} = \max_l \left\{ \sum_{w \in W} \sum_{p \in P_w} \delta_l^p n_p(s) \right\},$ where $n_p^{\text{opt}}(s) = \frac{1}{k} (n_w^{\text{new}}(s) + n_w^{\text{rer}}(s))$ $\forall p \in P_w, \forall w$ and $P_w = \{p_1(w), p_2(w), \dots, p_k(w)\}.$	<b>hES/G</b> Calculate the WCL capacity $n_{\max}(\mathbf{n}_p)$ as $n_{\max} = \max_l \{n_l(s)\} = \max_l \left\{ \sum_{w \in W} \sum_{p \in P_w} \delta_l^p n_p(s) \right\},$ where $n_p^{\text{opt}}(s) = \frac{1}{k_w^*} (n_w^{\text{new}}(s) + n_w^{\text{rer}}(s))$ $\forall p \in P_w, \forall w$ and $P_w = \{p_1(w), p_2(w), \dots, p_{k_w^*}(w)\}.$
<b>Full LP Optimization</b>	<b>hFO</b> Minimize the WCL capacity, $\min_{\mathbf{n}_p} n_{\max}(\mathbf{n}_p),$ subject to (i) the link capacity bounds $n_{\max} \geq n_l(s) = \sum_{w \in W} \sum_{p \in P_w} \delta_l^p n_p(s) \quad \forall l,$ (ii) the lower path capacity bounds $n_p^{\text{rem}}(s) \leq n_p(s) \quad \forall p \in P_w, \forall w$ and (iii) the OD/path capacity requirements $n_w(s) = \sum_{p \in P_w} n_p(s) \quad \forall w,$ where $P_w = \{p_1(w), p_2(w), \dots, p_k(w)\}.$	<b>hFO/G</b> Minimize the WCL capacity, $\min_{\mathbf{n}_p} n_{\max}(\mathbf{n}_p),$ subject to (i) the link capacity bounds $n_{\max} \geq n_l(s) = \sum_{w \in W} \sum_{p \in P_w} \delta_l^p n_p(s) \quad \forall l,$ (ii) the lower path capacity bounds $n_p^{\text{rem}}(s) \leq n_p(s) \quad \forall p \in P_w, \forall w$ and (iii) the OD/path capacity requirements $n_w(s) = \sum_{p \in P_w} n_p(s) \quad \forall w,$ where $P_w = \{p_1(w), p_2(w), \dots, p_{k_w^*}(w)\}.$
<b>Bounded LP Optimization</b>	<b>hBO</b> = hFO plus the additional upper path capacity bounds $n_p(s) \leq \alpha n_w(s) \quad \forall p, \forall w,$ where $\alpha \in [1/k \dots 1]$ and $p \in \{p_1(w), p_2(w), \dots, p_k(w)\}.$	<b>hBO/G</b> = hFO/G plus the additional upper path capacity bounds $n_p(s) \leq \alpha n_w(s) \quad \forall p, \forall w,$ where $\alpha \in [1/k_w^* \dots 1]$ and $p \in \{p_1(w), p_2(w), \dots, p_{k_w^*}(w)\}.$

# Chapter 6

## Numerical Studies

*The purpose of computing is insight, not numbers.*

— RICHARD W. HAMMING (1962)

*The purpose of computing is not really in sight.*

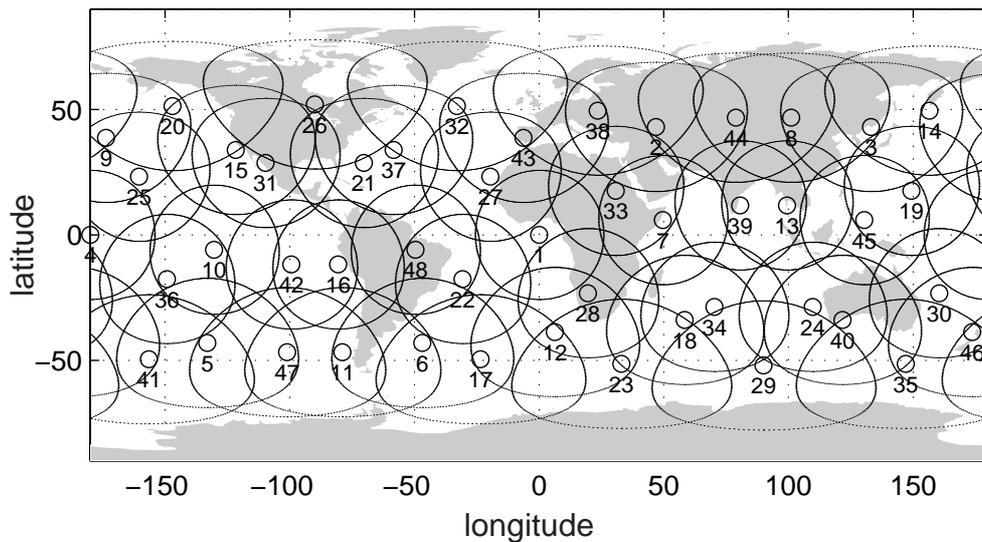
— RICHARD W. HAMMING (1973)

The numerical results assembled and discussed here refer to the different areas of research covered by Chapters 4 and 5 before. Table 6.1 provides an overview of the various numerical studies contained in this chapter, and links back to the corresponding sections where underlying theory, concepts and approaches have been treated in detail. The table also mentions the scenario and/or traffic models considered, as well as the software tools used in the respective numerical studies; the modeling assumptions will be outlined in the context of the particular investigations.

LEOSIM, ISLSIM, and ISLDIM are proprietary software packages of DLR, which have been developed to perform most of the numerical investigations underlying this thesis. All are discrete-event tools programmed in the C language. As the names indicate, LEOSIM covers all basic and general aspects of LEO (and MEO) constellations not related to ISLs, ranging from earth/space geometry and constellation modeling over channel, user and traffic models to handover and diversity schemes. ISLSIM is completely dedicated to all issues of traffic routing in the ISL subnetwork, and ISLDIM provides the framework for the ISL capacity dimensioning.

**Table 6.1:** Overview of numerical studies.

Results Section	Topic	Concepts Section	Scenario&traffic model / Constellation(s)	Software tool
6.1.1	UDL routing (handover and diversity)	4.2	Iridium, LEONET (HO); Globalstar, ICO (sat. div.)	LEOSIM
6.1.2	Adaptive ISL routing (deterministic and NN)	4.4	TMF 1, TMF 2 (App. C)	ISLSIM; SNNS
6.2	ISL dimensioning	5.3&5.4	Iridium, LEONET network uniform traffic / M-Star (Celestri)	ISLDIM/MINOS



**Figure 6.1:** Coverage snapshot of Globalstar for a minimum elevation angle of  $10^\circ$ .

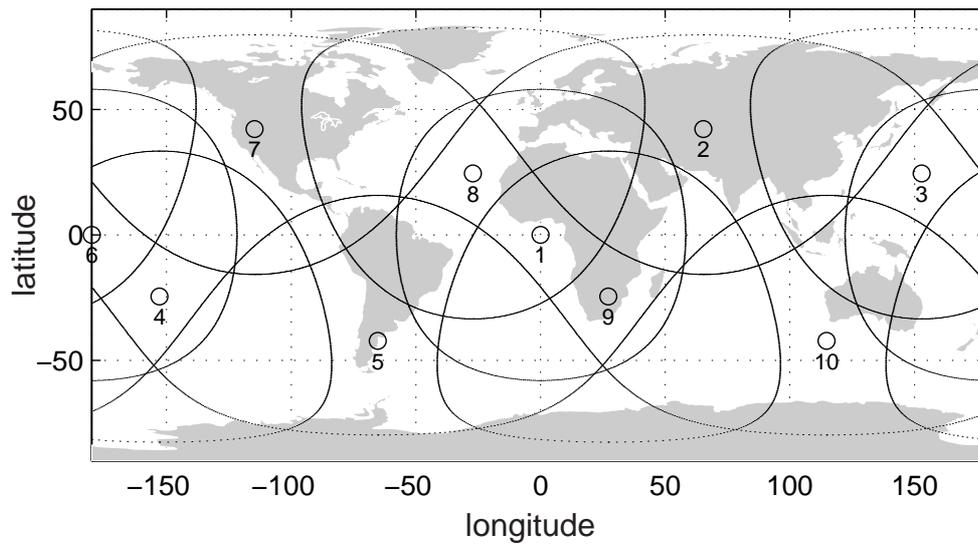
Besides this proprietary software, we have mainly used the following two tools:

- SNNS [Zel95] has been used for the isolated adaptive routing based on neural networks; it is shortly presented in an own Appendix D together with illustrated MLP implementations of ISL networks.
- MINOS [MS93] is a software tool programmed in the Fortran language, designed to solve large-scale linear and nonlinear optimization problems. The objective function and the constraints may be linear or nonlinear or a mixture of both. Stable numerical methods are employed throughout. Particularly interesting features for our work include (i) very efficient handling of linear constraints and bounds on the variables, (ii) automatic scaling of the constraints, and (iii) a standardized format for input and output data files. Finally, MINOS is callable as a Fortran 77 subroutine, which permitted a handy integration into the ISLDIM framework.

## 6.1 Routing

### 6.1.1 UDL Routing

Here we present and discuss additional numerical results pertaining to the UDL routing concepts from Section 4.2 [WBL95b, BWL96, BW97]. The focus is here on the modeling and simulation of a sophisticated satellite diversity approach, and some related handover issues are tackled as well. As in Section 4.2.2, Globalstar and ICO are studied as example systems; their coverage snapshots are displayed in Figs. 6.1 and 6.2, and a complete view on the constellations as well as the complete listing of constellation parameters can be found in Appendix A in Figs.A.2 and A.5 and Table A.1.



**Figure 6.2:** Coverage snapshot of Globalstar for a minimum elevation angle of  $10^\circ$ .

### 6.1.1.1 Scenario and Overall Simulation Approach

The basic scenario is like in the methodology Section 4.2.2, with the exception that we now restrict to *dual* satellite diversity; this is in fact a direct consequence of the fundamental results achieved in these earlier investigations based on “simple” models and approaches. Based on this scenario, the overall simulation approach is as illustrated in the flowchart in Fig. 6.3. Extensive simulations have been performed in for a gateway at ( $100^\circ\text{W}$ ,  $40^\circ\text{N}$ ) and various numbers of user positions (UP) equally distributed in a circle around the gateway forming its service area with radius  $r_{\text{SA}}$ . Each considered constellation are is simulated over a whole day (1440 min) with a time step of 1 minute. Once the satellite selection has been performed by the GW for the actual instant of time, all UPs can be investigated in terms of visibility. Then the channel model is simulated for each triple consisting of the UP and the respective diversity satellite pair, where the correlation between the two channels is taken into account (see next section). The resolution for the simulation of the channel model is kept fine enough to guarantee minimum quantization errors. The diversity evaluation concludes with the calculation of mean values for statistics.

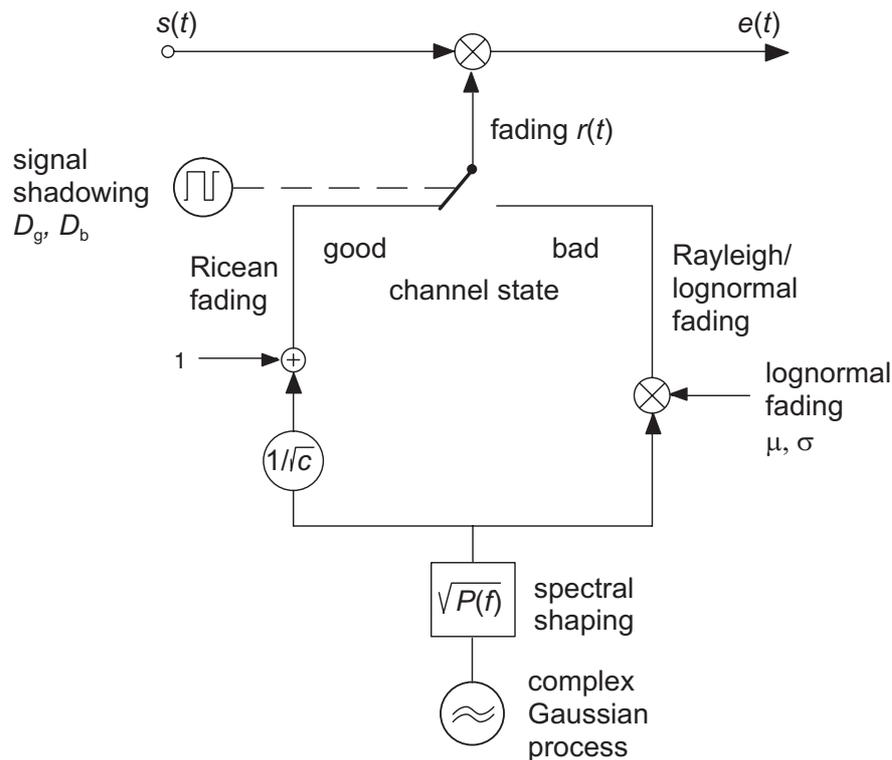
### 6.1.1.2 Channel Modeling

The channel model adopted in this simulation is essentially based on earlier channel modeling work, cf. [LCD<sup>+</sup>91], [Lut96], [BWL96]. For the sake of a better understanding of the diversity scheme that is proposed later, we recall the most important characteristics here.

The model takes into account the following three main characteristics of the non-geostationary MSS channel(s):

1. The channel depends on the user environment and the user velocity (like the geostationary land mobile satellite (LMS) channel).
2. The channel depends on the time-varying elevation angle (which is in fact quasi-static in the GEO case).





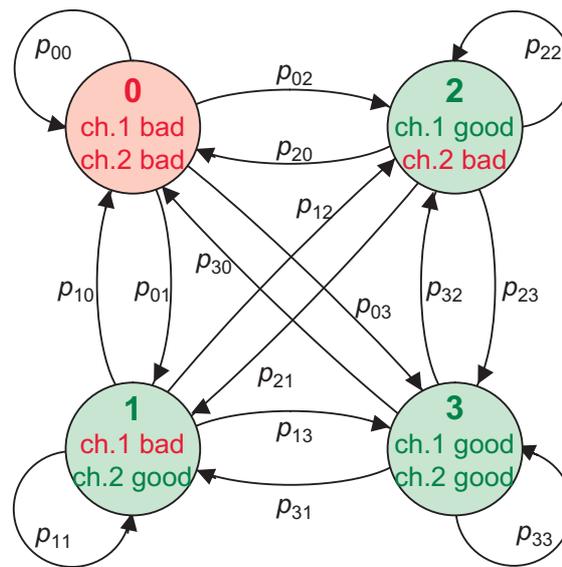
**Figure 6.4:** Narrowband model of the land mobile satellite channel [LCD<sup>+</sup>91].

3. Channels from one user to different satellites are correlated, mainly with respect to their azimuth angle difference.

**Elevation-dependent Single MSS Channel Model** The general Rice–Rayleigh/lognormal channel model used for a single land mobile satellite channel has been introduced and discussed in [LCD<sup>+</sup>91]. Fig. 6.4 illustrates the basic idea: the fading process is switched between Ricean fading, representing unshadowed areas with high received signal power (good channel state) and Rayleigh/lognormal fading, representing areas with low received signal power (bad channel state). The two global channel states are driven by a two-state Markov chain.

The original model parameters have been derived for a European geostationary satellite, evaluating measurement data for different environments. However, the general model approach can easily be applied to non-geostationary land mobile satellite systems as well by assessing elevation-dependent parameter sets for the typical LEO/MEO scenario. Aiming at this goal, several dedicated measurement campaigns have been conducted for different elevation angles and different environments [JL94, Jah95, JSBL95]. The channel measurements indicate a strong dependence on the elevation angle, especially in urban and suburban environments. In our simulation, all parameters of the model are fed with data derived from the mentioned measurement campaigns.

**Correlation of Two MSS Channels** Exploiting multiple satellite visibility on earth, the service availability may substantially be improved. Of course, gain in service availability can only be achieved if the considered satellite channels behave *different*. Therefore, any dependency between the channels influences the benefit of satellite diversity.



**Figure 6.5:** Four-state Markov model for the shadowing of two correlated land mobile satellite channels [Lut96].

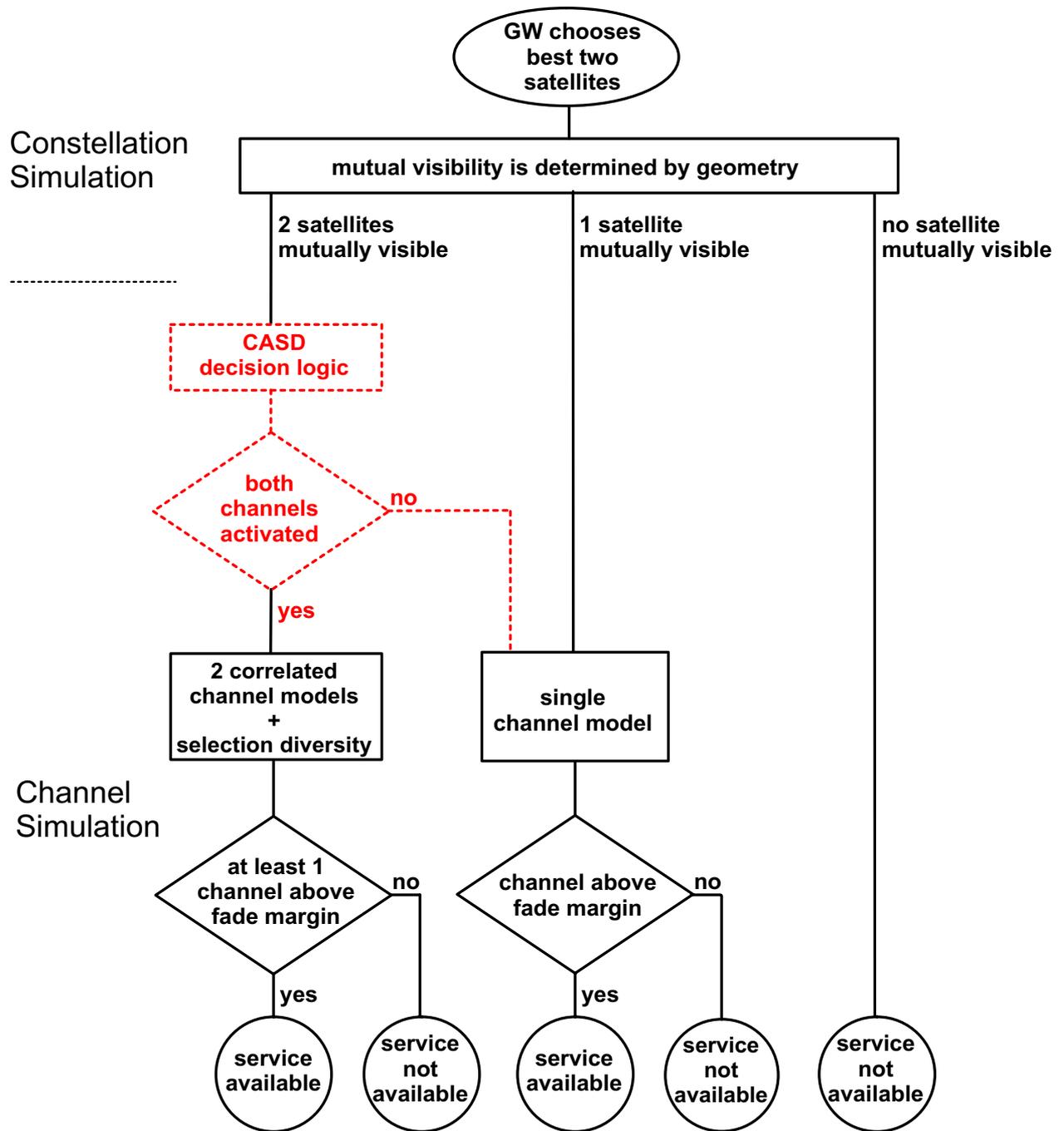
In [Lut96] a concept for modeling the shadowing behavior of two statistically dependent satellite channels was developed, which is based on a four-state Markov chain, Fig. 6.5. In this context, the mathematical notation and analytical treatment of dual satellite diversity has been discussed introducing a *correlation coefficient*, which depends on the elevation angles and on the azimuth separation of the two satellites. As an overall result, it was shown that a negative correlation between the channels increases service availability, and vice versa. For our simulation again, the numerical implementation of the correlation model and its associated parameters is validated by results from measurement campaigns.

### 6.1.1.3 Channel Adaptive Satellite Diversity (CASD)

In the methodology Section 4.2.2 a simple implementation of satellite diversity has been adopted, where both diversity channels are *always* used when the geometrical visibility is given, i.e. they both consume link resources permanently. Due to this *full-time* occupancy we refer to this scheme as *full diversity* in the following. Considering for instance the case of both channels being in good channel state and a pure selection diversity being implemented, this scheme does obviously not use the expensive user link bandwidth resources very sparingly. On the other hand, one gains significant improvement in terms of service availability. The essentials of the full diversity scheme are given by the solid line components of the flowchart in Fig. 6.6.

A less bandwidth consuming alternative to the full diversity operation is *pure satellite handover*. The major trade-off is then between bandwidth consumption and signaling complexity. Considering service availability, ideal hard handover shows the same performance as full diversity. For realistic implementations of both, however, service impairments are slightly better reduced through diversity. This is mainly due to non-ideal hard handover because of the delay for signaling message exchange.

The driving force behind the development of a new and more sophisticated scheme is the prospect of combining the advantages of full diversity and pure handover without including their respective drawbacks, i.e., essentially: minimize the bandwidth consumption while keeping the service



**Figure 6.6:** Approach for satellite diversity simulation and performance evaluation in terms of service availability and required link capacity. Replacing the direct flow line by the dashed line components makes up the adaptation of full diversity toward CASD.

availability to a maximum (as achieved with full diversity). Considering for instance a situation where both diversity satellites are typically unshadowed during long time periods (as in highway environment) one can easily conclude that a straightforward approach to reach this goal is to release one of the channels since it does not affect service availability. It is however also obvious that the bandwidth saving potential of this scheme is closely related to the user environment, with some inherent reduction for critical environments such as urban. It is therefore highly important to elaborate a sophisticated implementation of this principle in order to meet the characteristics of the respective environments. Already before looking on any concrete solution, one can generally formulate the demand for an adaptive solution. From here, it is only a last consistent step to use the channel conditions themselves for deriving the decision to activate or release a channel, because the channel behavior also reflects the user environment.

In [BW97] we have proposed such a *channel adaptive satellite diversity (CASD)* scheme as a consistent adaptation of the full diversity solution. The dashed line elements in Fig. 6.6 illustrate the required adaptation of the full diversity scheme (solid lines) on a general level: whenever two satellites are geometrically visible, an additional logic is introduced that decides if only one or both of the corresponding channels are activated. The decision is derived from the channel conditions and results in activation and release procedures of the channels. The behavior of this logic is therefore very similar to a handover scheme, where an old channel is forced to be released or dropped, and a new channel has to be set up. From this viewpoint, the CASD scheme may also be seen as a hybrid solution between full diversity and pure handover.

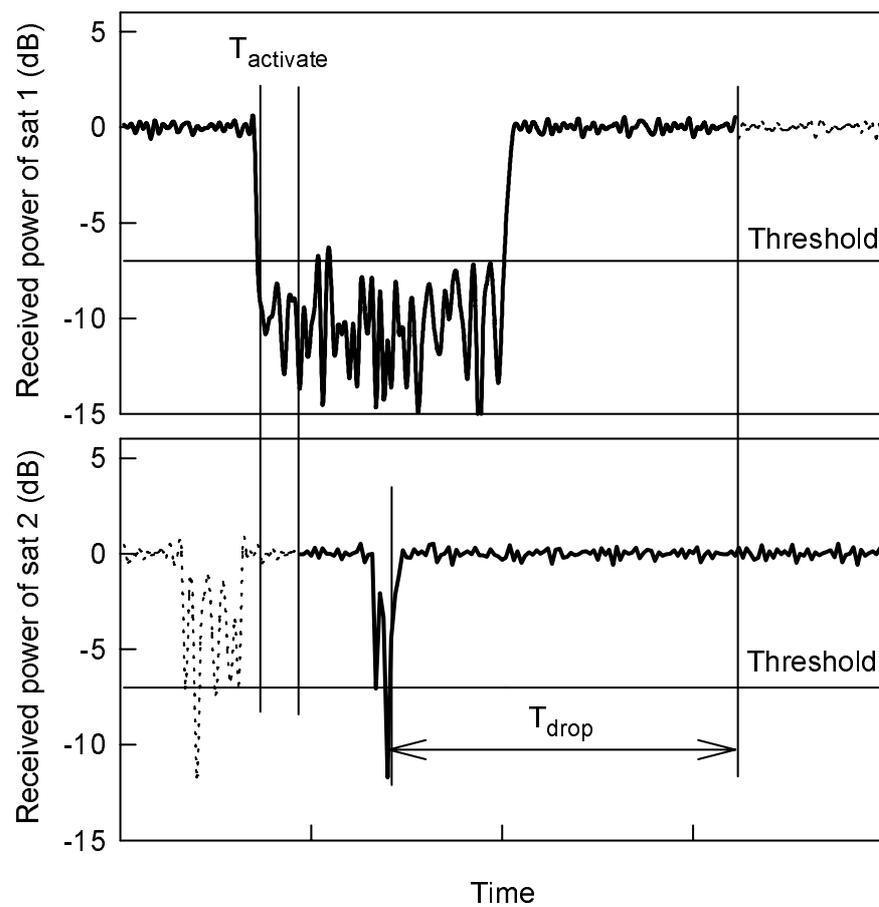
Fig. 6.7 provides insight into the details of a possible realization of CASD. The process is driven by two timers  $T_{\text{activate}}$  and  $T_{\text{drop}}$  which are implemented in the receiving branches of the user terminals. The starting point of the depicted time interval is given with the situation that due to the CASD scheme channel 2 is inactive (i.e. no link capacity is used). As soon as the received power level in channel 1 drops below a given threshold (bad channel state), the timer  $T_{\text{activate}}$  controls the setup of channel 2 via the partner diversity satellite. If channel 1 stays below the threshold during the complete timer period, channel 2 is activated. In an operational system, of course, the complete channel setup signaling has to be performed within this period. Thus the signaling protocol together with the round trip delay effectively set a lower bound to an applicable  $T_{\text{activate}}$ . The next section will provide some respective numerical considerations. Once both channels are in use, a  $T_{\text{drop}}$  timer is initialized for both fading processes whenever the power threshold is exceeded. If one of the channels experiences a timeout while staying above the threshold for the whole timer period, the *partner* channel (if there is a second active diversity channel) is released. This can be modeled to happen immediately after  $t = T_{\text{drop}}$  by assuming that the required release message(s) are sent just before the timeout.

The dashed parts of the fading processes in both channels illustrate the link capacity saving compared to the full diversity scheme.

#### 6.1.1.4 Satellite Diversity and Handover Performance

The performance of the CASD scheme shall be mainly compared to its “parents” (i) full diversity scheme and (ii) pure (non-ideal) handover operation. The three major performance quantities are

- service availability,



**Figure 6.7:** CASD operation: timers  $T_{\text{activate}}$  and  $T_{\text{drop}}$  used in the fading processes of the two satellite channels; solid lines indicate an active channel, whereas in dashed line periods link capacity is saved.

- link capacity (resp. bandwidth) requirements, which correspond to the average number of used satellites, and
- channel setup signaling requirements.

In a first step it is worthwhile investigating the influence of the timers on these quantities in order to fix reasonable values. Fig. 6.8 provides corresponding simulation results for two representative environments and user velocities, showing the influence of both timers at the same time. Given an environment/velocity pair, the task of finding optimum timer values could of course consist of solving a multidimensional optimization problem subject to a certain target function in analytical form. But then the problem is basically to formulate *one* such target function since there is obviously no unique “optimum” taking into account the three given quantities; rather there is some inherent fuzziness.

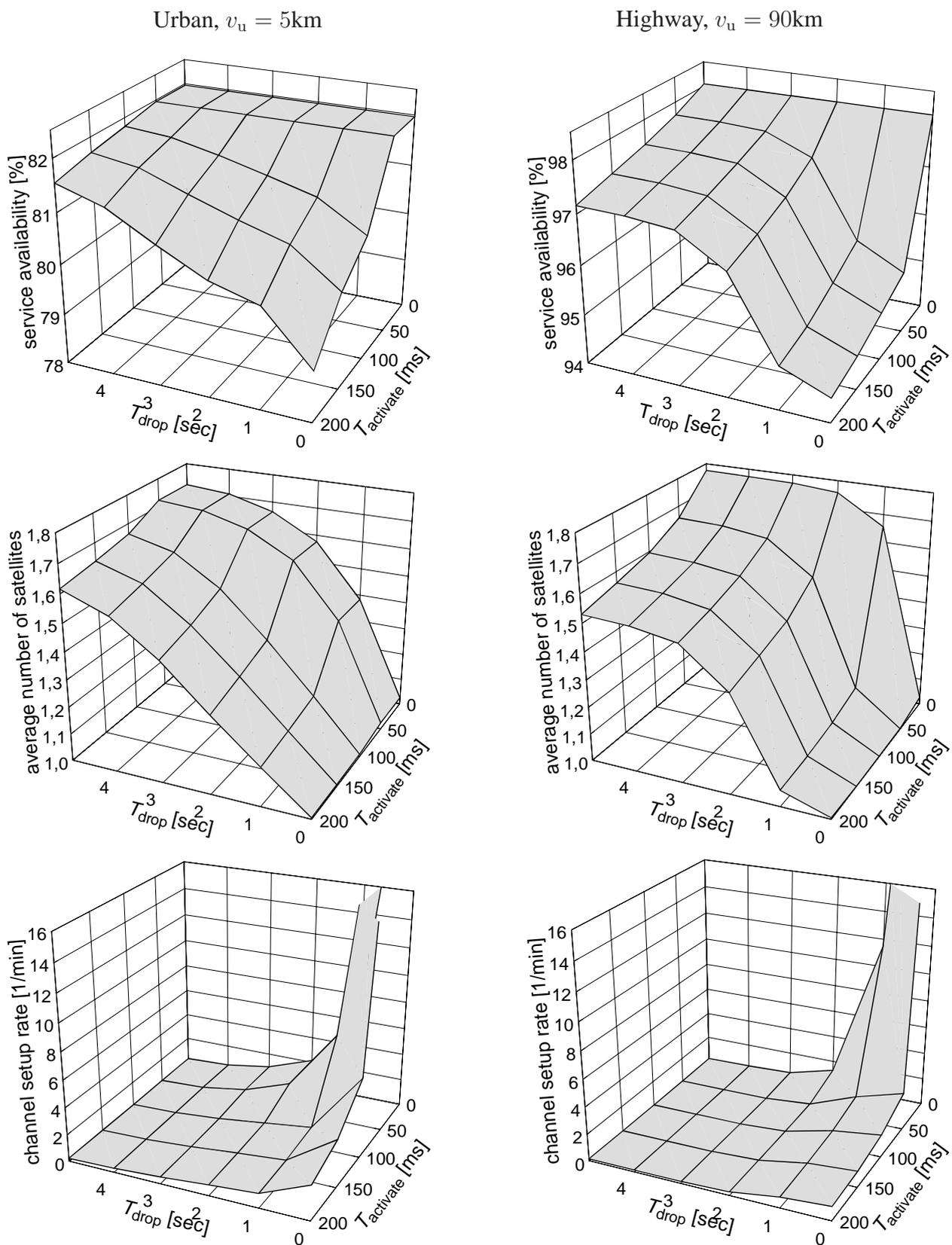
A more pragmatic approach is to consider in a first step natural lower bounds for the timer values and then to reduce possible choices step by step by taking into account the target performance quantities in prioritized manner.

First of all, a lower bound for  $T_{\text{activate}}$  and  $T_{\text{drop}}$  is given by the duration of signaling message exchange for channel setup and channel release, respectively. Here a simplified version of the formerly proposed forward satellite handover signaling protocol [Eis96, EVW96] can be applied, then basically requiring one message being sent from the mobile terminal to the GW and vice versa. Neglecting processing delay, the accumulated propagation delay alone yields a maximum (for the worst case, i.e. minimum elevation angle) of roughly 40 ms for LEO and 180 ms for MEO constellations.

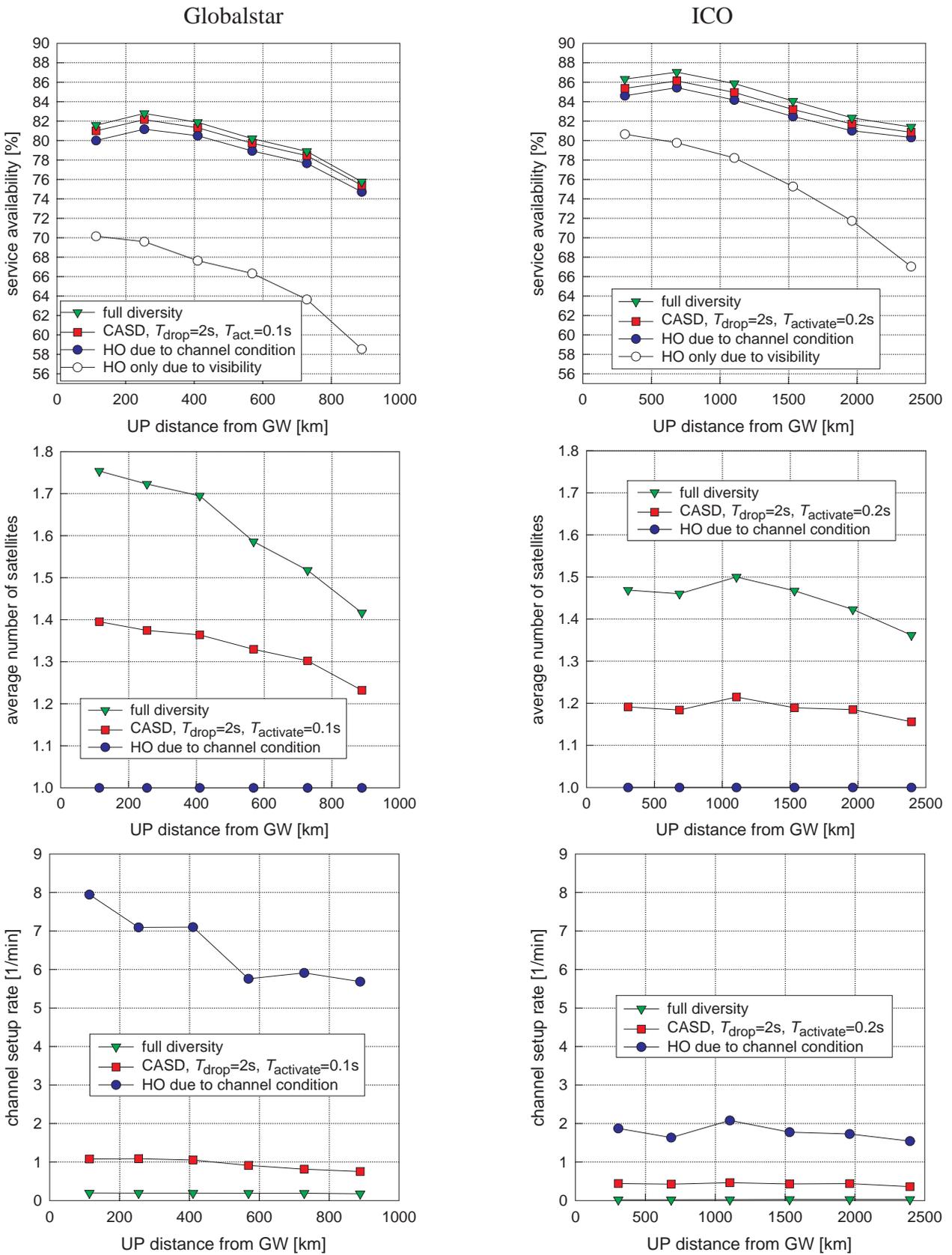
Given this lower bounds for the timer values, one may next allow a maximum of 1% decrease in service availability to be introduced by CASD compared to full diversity. (Note that the ideal full diversity case is represented in the figures by  $T_{\text{activate}} = 0$  and  $T_{\text{drop}} = \infty$ .) Applying this to the availability surface regions, Fig. 6.8 top, the applicable timer pairs are immediately reduced. In the next step, it seems reasonable to give complete priority to the link capacity requirements (middle) over the signaling requirements (bottom), because bandwidth saving is what the whole CASD is primarily for. Minimizing the link capacity requirements within the allowed timer pair region finally yields an “optimum pair”. Despite some differences with respect to the environments and user velocities one finds that  $T_{\text{activate}} = 100$  ms and  $T_{\text{drop}} = 2$  s is a generally attractive choice for LEO systems. Specifically, these values also yield a low channel setup rate although this criterion has not really been used for the selection.

With these timer settings extensive simulations have been performed to numerically evaluate the performance of CASD in comparison with full diversity and pure handover. For the remainder of this section we restrict ourselves to the presentation and discussion of simulation results for urban environment and a user velocity of 5 km/h.

Figure 6.9 displays the Globalstar and ICO results for all three mentioned performance parameters; the average number of used satellites reflects the bandwidth consumption, and the channel setup rate serves as indicator for signaling requirements. All curves are recorded versus the distance between user and gateway in order to catch the general performance decrease when approaching the edge of the service area. Note that “handover due to channel condition” (i.e. pure non-ideal handover) is given by  $T_{\text{activate}} = 100$  ms (200 ms) and  $T_{\text{drop}} = 0$ .



**Figure 6.8:** Influence of the timers  $T_{activate}$  and  $T_{drop}$  on (i) service availability, (ii) satellite link capacity and (iii) channel setup signaling (from top to bottom); Globalstar, two typical environments and user velocities, UP distance from GW = 400 km.



**Figure 6.9:** Performance comparison of different diversity / handover schemes for Globalstar (left) and ICO (right): (i) service availability, (ii) satellite link capacity and (iii) channel setup signaling (from top to bottom); urban environment and user velocity  $v_u = 5$  km in all cases.

Qualitatively, the performance of CASD – compared with full diversity and pure handover – is the same with respect to all three criteria for both systems. The figures prove that – as intended – CASD is capable of significantly reducing full diversity capacity and pure handover signaling requirements at the same time. With suitable timer settings, the respective drawbacks of both parent schemes can be largely avoided. At the same time, the decrease of service availability is virtually negligible. Other timer settings would shift the respective CASD curves either closer to the full diversity or to the pure handover ones. This characteristic can be utilized in a real implementation by introducing another level of adaptivity: the timer values need not be fixed but may adapt to different environments and/or traffic load situations dynamically during the operational phase of the system.

## 6.1.2 ISL Routing

### 6.1.2.1 Simulation Approach and Traffic Modeling

Here we present and discuss some focused numerical results pertaining to the on-line adaptive routing approaches from Section 4.4. For some basic sensitivity studies concerning (i) long-term traffic variations and (ii) the influence of relative traffic weighting in the link cost metric, a somewhat simplified two-step implementation of a deterministic adaptive routing scheme has been used, as shown in Fig. 6.10. This simplification is well justified for the considered sensitivity studies on the one hand; on the other hand, it significantly reduces the computational effort. As can be seen from the flowchart, in a first non-adaptive phase pure shortest path routing is performed using geometrical distance (or propagation delay) between adjacent satellites as the sole component of the link cost metric. Once the carried traffics in the network have been calculated according to the SPA routing decisions, *these* traffic values are in the second (adaptive) phase used as additional component in the new link cost metric besides the distance/delay. This simplified approach has also been used to generate the training data for the neural network routing from Section 4.4.2.

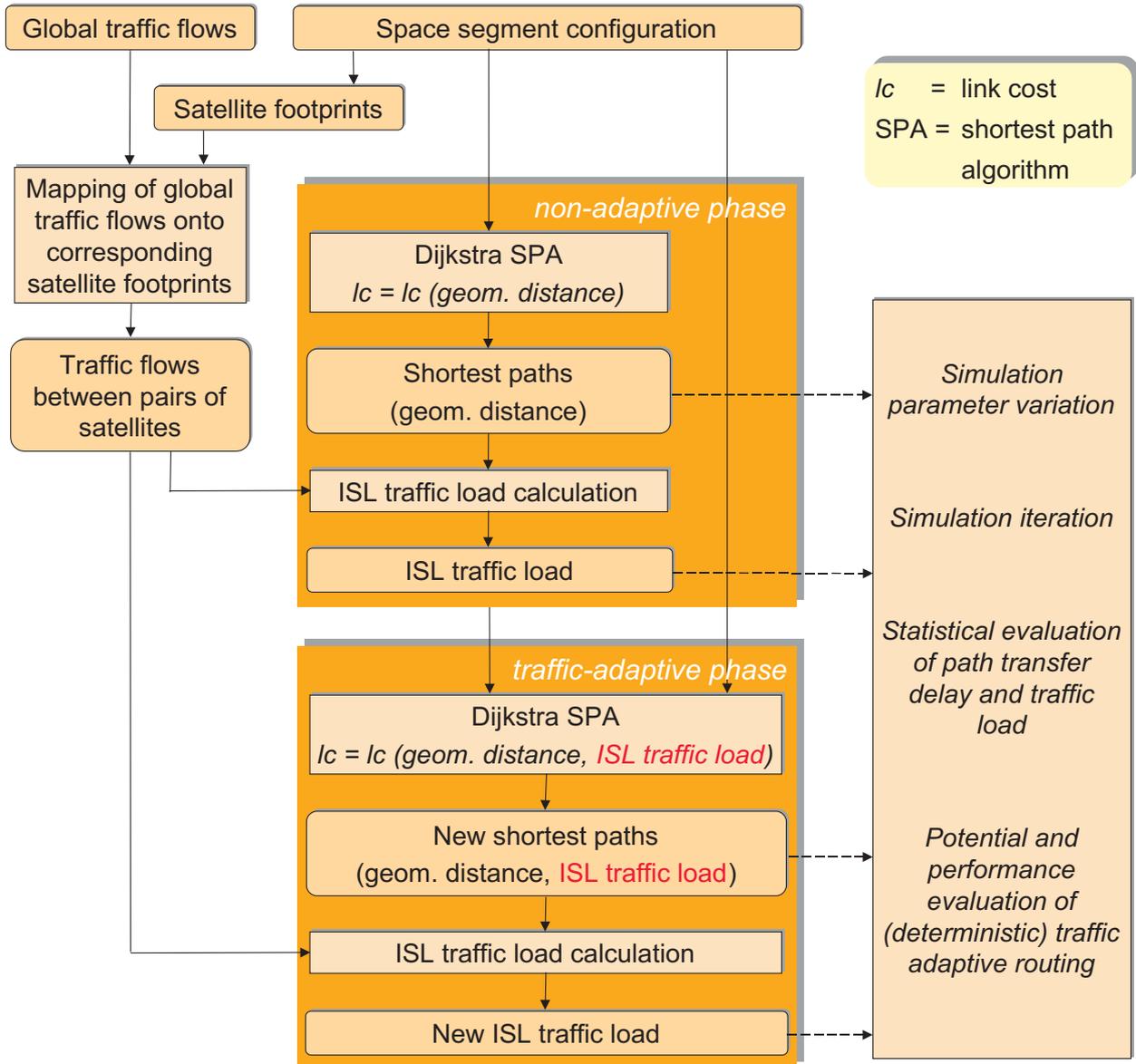
As a key input to the two routing phases, appropriate data for offered network traffic must be provided. Natural steps of network traffic modeling applied to a simulation approach for ISL routing studies, and of the corresponding sub-models to be used, are illustrated in Fig. 6.11.

The data for these models used in our numerical studies are according to

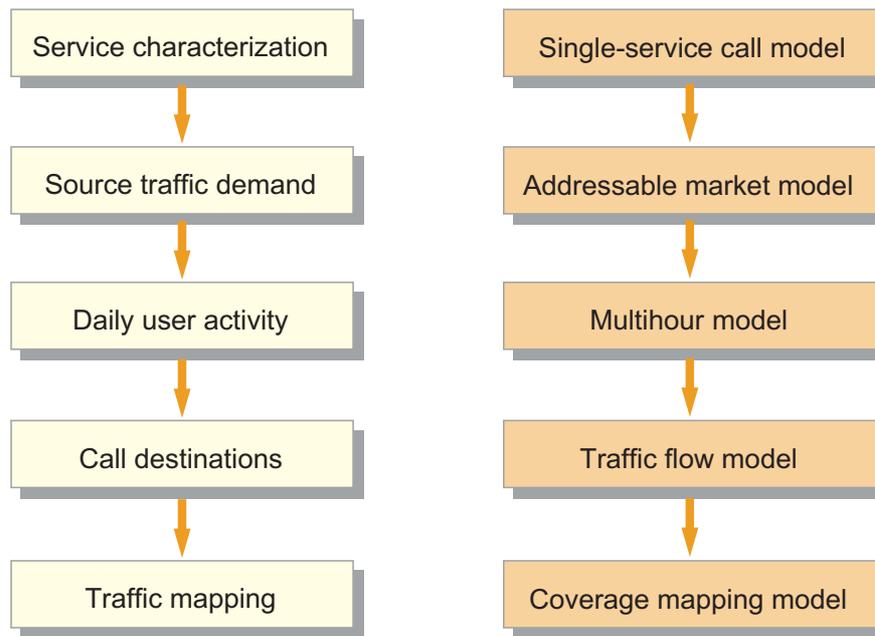
- the traffic modeling framework TMF 1 (see Appendix C.1) for the following two subsections, investigating the impact of (i) long-term traffic variations and (ii) relative traffic weighting in the link cost metric, and
- the traffic modeling framework TMF 2 (see Appendix C.2) for the studies on the impact of different link cost metrics (third subsection).

Additionally, in order to calculate the required ISL capacity  $C_{\text{ISL}}$  (i.e. the number of full-duplex channels for our traffic/service model), the Erlang-B formula (2.12) has been applied to the worst-case link load,  $A_{\text{WCL}}$  for a target blocking value of  $B = 5\%$ .

Iridium and LEONET have been studied as example systems, with relevant constellation details and ISL topology snapshots provided in Appendix A.



**Figure 6.10:** Simulation approach for a simplified two-step implementation of a deterministic traffic adaptive ISL routing scheme.



**Figure 6.11:** Required steps and corresponding sub-models for network traffic modeling used in ISL routing simulations.

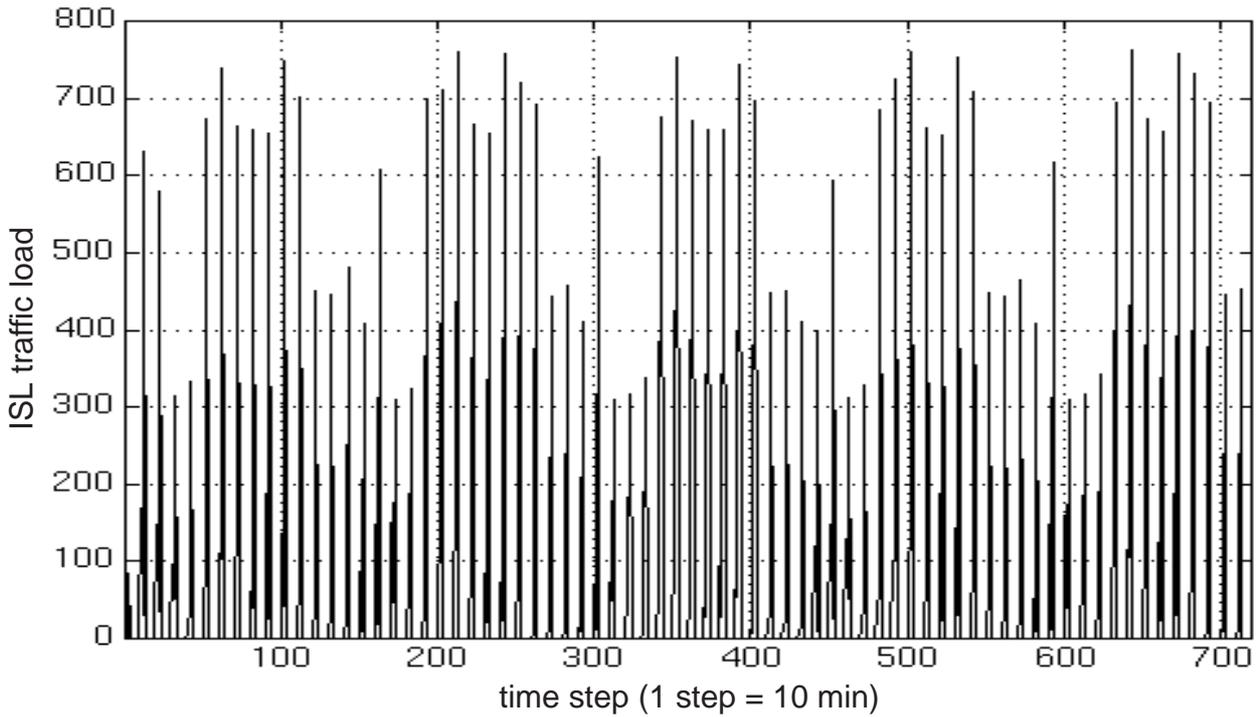
### 6.1.2.2 Impact of Long-Term Traffic Variation

In Section 2.6.2 we have introduced the important notion of a *system period*  $T_S$  (cf. Eq. 2.14) reflecting the situation that the earth rotates underneath the orbiting satellites and thus it is important to identify the when repetitions of the orbital patterns meet repetitions of earth traffic patterns for the first time. We have also stated in this context that the system period is the *maximum* period of time which needs to be taken into account and have consequently performed all numerical investigations in this thesis under this paradigm. Even more, in many cases the underlying simulations have only been performed over one (solar) day, driven by the fact that the particular topic or target of the research in those cases (for instance, assessing the worst-case link load) did not require or only indicate that longer observation periods would bring in substantial qualitative or quantitative differences. In order to support this hypothesis with some data material, Fig. 6.12 displays a representative result of earlier extensive ISL load simulations over a system period and longer [Kro96], which have been initially performed to justify the restriction to the system period or one day in later research work on the topic.

Figure 6.12 displays the traffic load of an Iridium inter-orbit ISL over the system period  $T_S = 5$  days. The system period for Iridium has been calculated according to Eq. (2.14) as follows:

$$\begin{aligned}
 T_S &= \min \{ \text{lcm}(T, T_E), \text{lcm}(T, T_{\text{Sol}}) \} \\
 &= \min \{ \text{lcm}(100 \text{ min}, 1436 \text{ min}), \text{lcm}(100 \text{ min}, 1440 \text{ min}) \} \\
 &= \min \{ 35900 \text{ min}, 7200 \text{ min} \} = 7200 \text{ min}.
 \end{aligned} \tag{6.1}$$

The particular link chosen is the ISL 32–42 (cf. Fig. A.7) between the third and fourth orbit, i.e. in the “middle” of the seamed network hemisphere, because the load variations on the central inter-orbit ISLs of Iridium are also quite similar to those encountered in delta constellations without a seam; compared to such links, the ISL load variation of the seam satellites is much more affected



**Figure 6.12:** Traffic load on an inter-orbit ISL in the Iridium network over one system period  $T_S = 5$  days.

by the existence of the seam [Kro96, WKM97] and thus not representative for a broader range of constellations. Moreover, shortest path routing has *intentionally* been adopted for these long-term simulations, as it transfers the peculiarities of both (i) the original traffic pattern variation and (ii) the influence of the ISL topology more pronounced on the resulting ISL loads, whereas traffic adaptive routing would rather conceal the effects.

In the load curve one can easily identify five sections of very similar variation patterns, being identical with a period of one day each. Roughly speaking, the ISL traffic load patterns repeat after one day, no matter what the exact relationship between a day and the constellation period is. Of course, a closer look reveals slight differences due to the orbital offset of the constellation pattern after exactly 24 hours, but the dominating characteristic is in the end a slow, permanent long-term shift of the ISL load patterns with respect to fixed 24 hour intervals. One could imagine a fixed 24 hour time window displaying *subsequent 24 hour samples* of the load curve in it: this would show (roughly) the same pattern slowly passing by again and again; obviously, extracting peak or mean values, for instance, does then no longer require to consider longer periods than any 24 hour section of the load curve.

### 6.1.2.3 Impact of Traffic Weight in the Link Cost Metric

A traffic adaptive routing algorithm is usually based on a link cost metric  $c_{ij}$  where a link traffic value  $a_{ij}$  from monitoring or measurement of the carried load is combined with the current link delay  $d_{ij}$ , for example in linear form using appropriate weighting factors:

$$c_{ij} = DWF \cdot d_{ij} + TWF \cdot a_{ij}. \quad (6.2)$$

In general,  $d_{ij}$  may consist of propagation delay *and* processing delay; in the following we only assume propagation delay, which allows to directly interpret it as geometric distance as well. The

relative impact of the two parameters on the calculated link cost is regulated with the *delay weight factor*  $DWF$  and the *traffic weight factor*  $TWF$ . Here we do intentionally not use a short-hand mathematical parameter notation for the two factors in order to highlight the important fact that they must both, besides their value, contain an “adaptive unit normalization component” which keeps the resulting link cost dimensionless and permits to provide the traffic and delay input values with respect to a particular unit value (e.g. in seconds or milliseconds).

In our first considerations to follow, the value of  $DWF$  is set to one, so that the relative weighting of delay and traffic is controlled by the value of  $TWF$  alone.

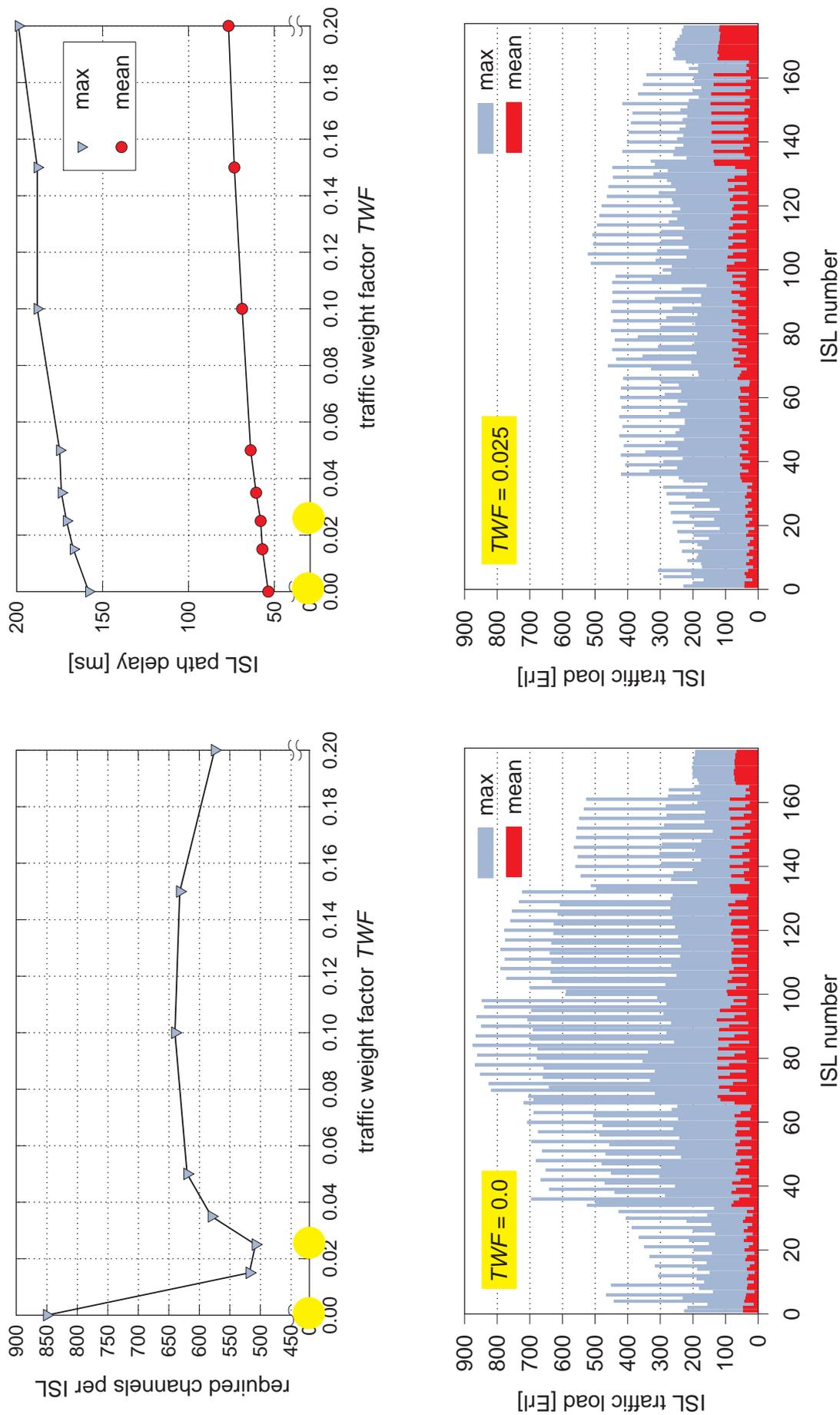
Figures 6.13 (Iridium) and 6.14 (LEONET) show at a glance the relevant simulation results for a study on the impact of the traffic weight factor  $TWF$  on (i) required ISL capacity, (ii) path delay and (iii) load distribution in the network. Presupposing that a reduction of required ISL capacity (number of channels) is the primary objective of traffic adaptive routing, the top left diagrams clearly suggest a particular  $TWF$  value as the preferred choice in both cases, namely 0.025 for Iridium and 0.3 for LEONET. While the two absolute values are of little importance (due to the contained normalization function), their relative difference of about a factor of 10 can be explained from the fact that a similar order of magnitude lies between the (average) link lengths of ISLs in Iridium and LEONET, and link distance is the parameter against which  $TWF$  has to weight the link traffic values.

The achieved capacity savings for the optimum  $TWF$ , as compared to the non-adaptive case ( $TWF = 0$ ), are impressive: roughly 40% for Iridium and 25% for LEONET; on the other hand, the increase of maximum and mean path delays (top right diagrams) is not at all critical, and particularly low (<10%) for Iridium.

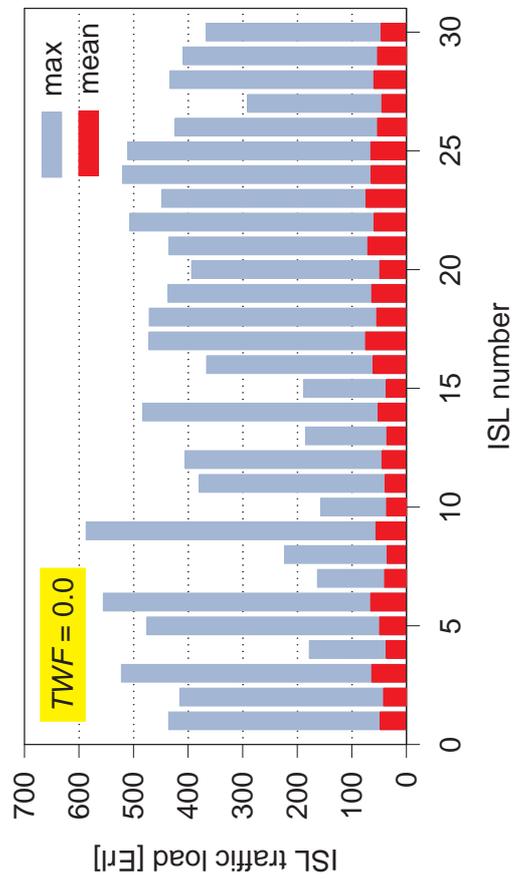
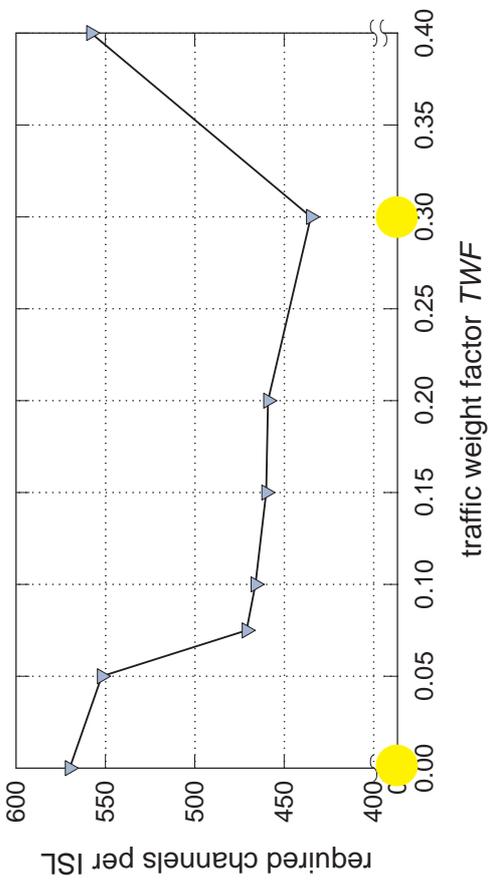
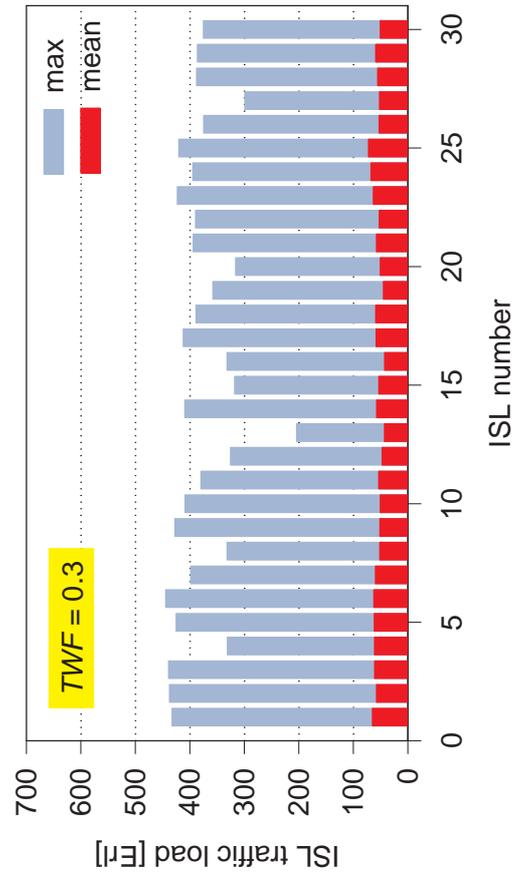
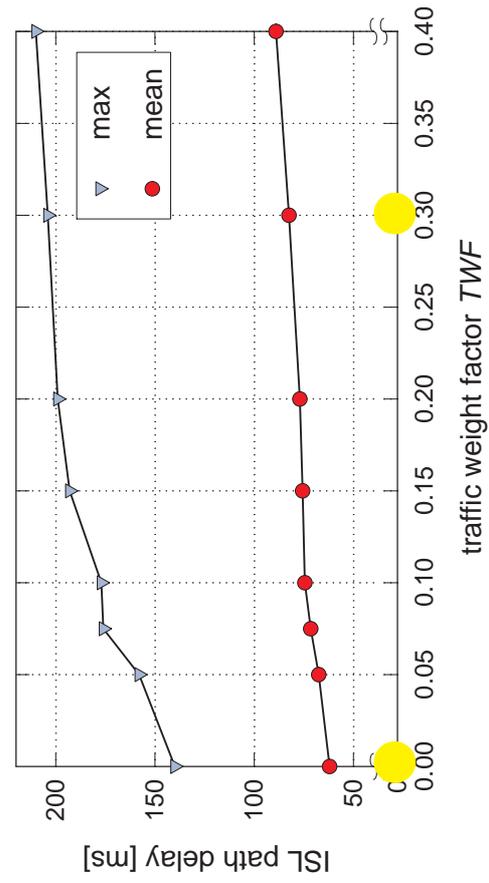
Comparing the non-adaptive with the optimum adaptive case, the bottom diagrams show the respective ISL load distributions in the network. They clearly spot the fundamental feature of traffic adaptive routing: traffic flows become much more evenly distributed across the network and peak loads are reduced; as a direct consequence, the average utilization of installed link capacities can be significantly increased. The worst-case ISL loads extracted from the bottom diagrams are linked with the already mentioned capacity (number of required channels) in the top left diagram via the Erlang-B function assuming 5% blocking.

In the case of LEONET, the concrete distribution of loads over ISL numbers (15 bidirectional intra-orbit and 15 bidirectional inter-orbit ISLs) is somewhat accidental, mainly depending on the arbitrarily chosen “start position” of constellation pattern with respect to earth, and thus, input traffic pattern. The Iridium network load pattern, on the contrary, is more meaningful, especially if one focuses on the maximum values. Full insight, however, demands knowledge of the specific ISL numbering sequence before. Starting with the first satellite in the first orbit (with the seam to the left), the respective *fore* and *right* bidirectional links are put in a sequence; this makes one intra-orbit ISL and *two* inter-orbit ISLs per satellite, one in the ascending and another one (to a different partner satellite!) in the descending orbital phase. This elementary set of three links first repeats eleven times for the satellites in the first orbit, yielding the first 33 bins in the diagram, followed by the satellites and ISLs in the second orbit, and so forth; following this rule, the last orbit with the seam to the right induces eleven intra-orbit ISLs only. In total, we have  $5 \cdot 11 \cdot (1 + 2) + 11 = 176$  separate links, 106 of which being active at a time (see the discussion in Section 4.1.1 on this issue). The six orbits can be easily identified in the diagram, as the satellites and links in center orbits clearly carry more accumulated traffic than those on the seam orbits, which is a logical

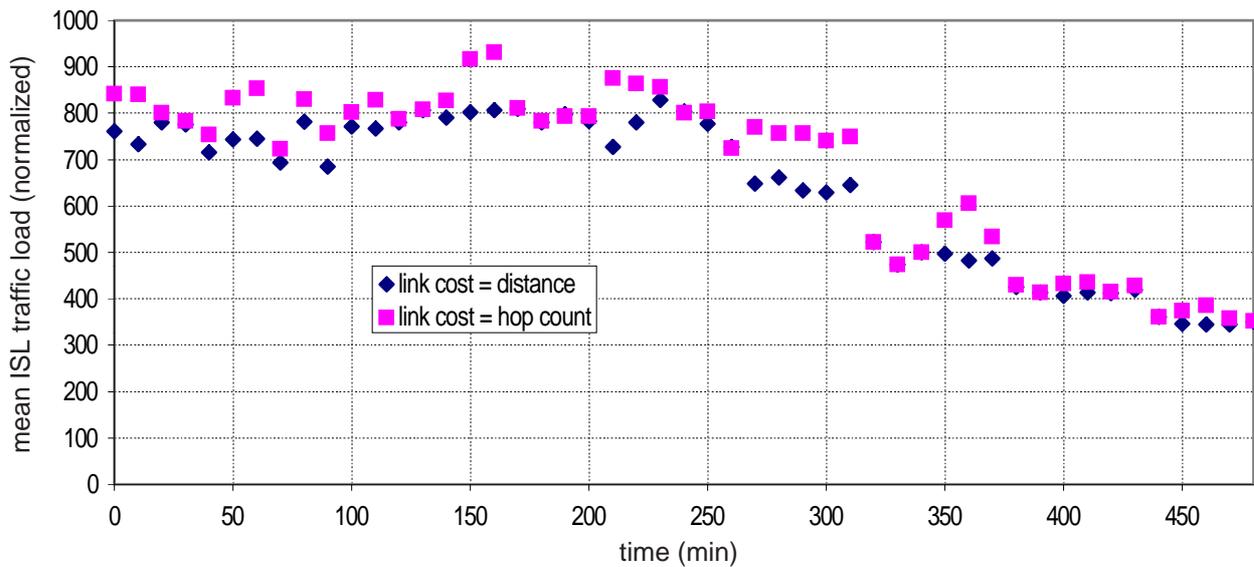
consequence of traffic concentration in the seamed network. Evidently, some reduction of peak loads is in this scenario achieved with traffic adaptive routing through shifting parts of the peak traffic flows more to the edge orbits, as can be concluded from the mean load values.



**Figure 6.13:** Study on traffic adaptive routing in the Iridium ISL network: impact of the traffic weight factor  $TWF$  on (i) required ISL capacity, (ii) path delay and (iii) load distribution in the network.



**Figure 6.14:** Study on traffic adaptive routing in the LEONET ISL network: impact of the traffic weight factor  $TWF$  on (i) required ISL capacity, (ii) path delay and (iii) load distribution in the network.



**Figure 6.15:** Comparison of alternative link cost metrics for plain shortest path routing (MDA).

#### 6.1.2.4 Impact of the Type of Link Cost Metric

We now consider the plain shortest path routing scheme MDA and the traffic adaptive routing scheme DBF as introduced in Section 4.4.1, using Iridium as an example. The underlying traffic modeling framework for these investigations has been TMF 2, with the data and geographic illustrations given in Appendix C.2.

A first comparison focuses on two possible link cost metrics for the MDA scheme, namely (i) a distance-based and (ii) a simple hop count metric. Figure 6.15 displays the resulting mean ISL traffic load in the network over time for both metrics. Obviously, the values are in many situations lower for the distance-based approach, and always as low as the load values achieved with the hop count metric. The reason for this is the fact that the inter-orbital ISL length roughly doubles between the instant when the ISLs are switched on at the edge of the polar regions and the instant when they are crossing the equator. Then, given that we often have pronounced east-west traffic flows between regions near the equator (cf. Fig. 4.20), the distance-based algorithm would often tend to route this traffic over mid-latitude inter-orbit ISLs due to their relative shortness, whereas the hop-based algorithm would select the “direct” (less hops) route along the equator.

Extending the investigation to the traffic adaptive DBF scheme as well, a new type of cost metric needs first to be mentioned. So far in this section, the *state* or traffic situation of the ISL network has been captured in *link* traffic values (and this is the case in all other chapters of this thesis), forming the basis for a *link-state* adaptive routing algorithm. There should however also be some interest in looking at a *node-state* adaptive routing approach, particularly in constellation networks where a satellite (node) has in some cases to process considerable traffic of all adjacent links. Moreover, representing the overall network state in  $N$  node states rather than in  $L$  link states – with  $L$  being always (and for a high connectivity of the ISL mesh considerably) larger than  $N$  – clearly reduces both (i) the amount of signaling for state information exchange and (ii) the processing effort for calculating route updates.

Besides two versions of link-state DBF routing as explained below, we have therefore also imple-

mented and studied a node-state DBF algorithm, with the node cost being

$$c_n = TWF \cdot a_n, \quad (6.3)$$

where the node traffic  $a_n$  consists of all accumulated adjacent (e.g. outgoing) link traffics. Evidently,  $TWF$  is here not required for relative weighting as there is no other cost metric component besides carried traffic, but only retained in its function of achieving a dimensionless cost value, depending on how *traffic* is described; this could be for instance in number of active calls or Erlang (dimensionless), but also based on monitored or measured queuing delay in on-board buffers as a handy solution in concrete simulations.

The link-state DBF scheme has been applied in two different versions, namely

- *pure traffic cost* DBF, i.e. according to (6.2) with  $DWF = 0$ ,

$$c_{ij} = TWF \cdot a_{ij}, \quad (6.4)$$

with  $TWF$  only used to keep the link cost dimensionless, and

- *combined log cost* DBF, where in some simulation test runs the linear combination of delay/distance and traffic according to (6.2) was found to be not appropriate due to the *specific* range of traffic values as measured in *this* implementation, and the link cost was instead calculated according to the logarithmic relationship

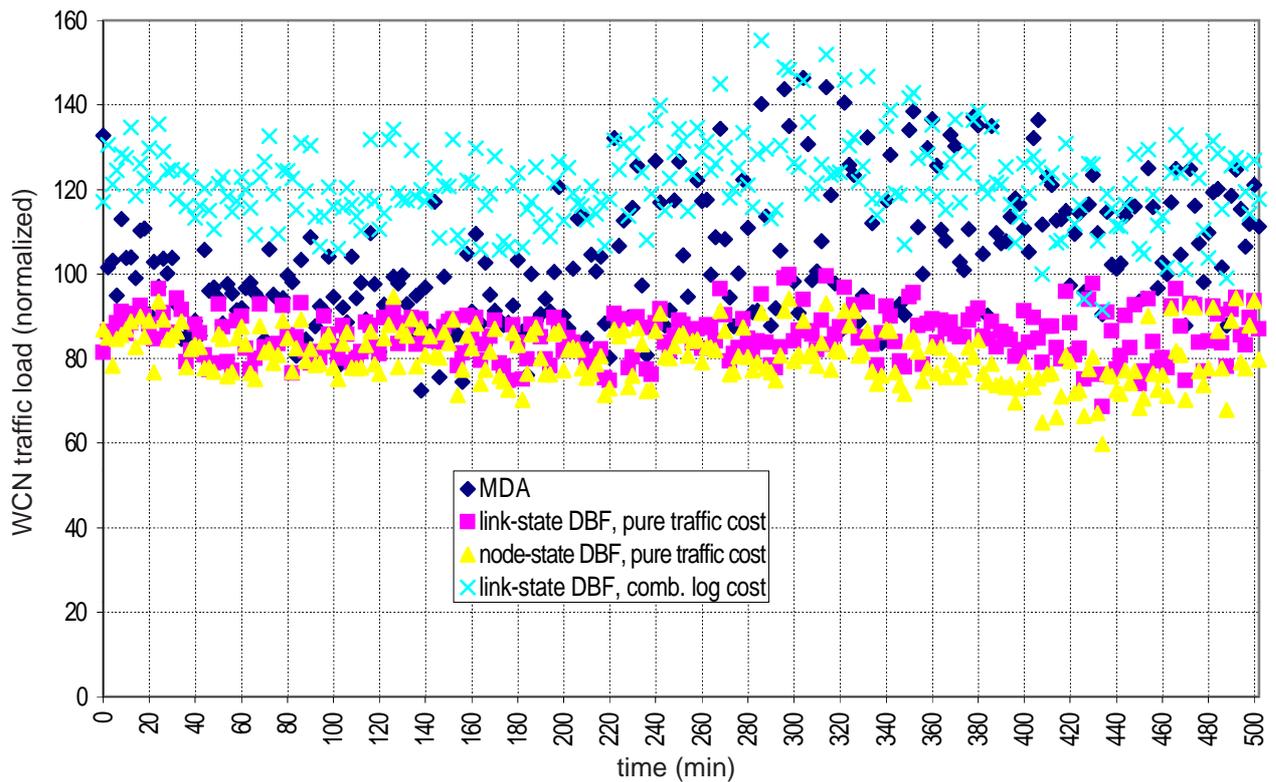
$$c_{ij} = (1 + \log(TWF \cdot a_{ij})) \cdot DWF \cdot d_{ij}, \quad (6.5)$$

where both  $TWF$  and  $DWF$  are again only used to keep the link cost dimensionless.

Figure 6.16 displays the resulting worst-case node (WCN) traffic loads over time, in direct comparison of the three adaptive DBF schemes and the reference distance-based MDA scheme. Both adaptive DBF schemes with pure traffic cost metric (link- and node-state) clearly outperform the non-adaptive MDA scheme by roughly 14 to 18% in the average. However, this WCN load reduction is clearly lower than the WCL load reduction achieved with the same schemes, which is in the order of 30% compared to MDA (cf. Fig. 4.21(a) on page 80). As one would expect, the node-state DBF version performs slightly better than the link-state one with respect to satellite (node) traffic, because it is just “made for” this target value. On the other hand, the average 4% gain in maximum satellite load has to be paid for with an increase of roughly the same percentage in maximum ISL load, where the link-state algorithm naturally performs better. The link-state DBF scheme with a combined (logarithmic) traffic/distance cost metric (“combined log cost”) exhibits an overall poor performance with respect to the target WCN load. The concrete amount of degradation compared to the MDA and the other two DBF schemes, however, is too closely linked to the particular kind of parameter combination to permit a far-reaching quantitative interpretation.

## 6.2 ISL Network Dimensioning

Here we present and discuss detailed numerical results pertaining to the network design, network dimensioning and optimization approaches from Chapter 5. Tables 5.3 (page 109) and 5.4 (page 110) provide a condensed overview and comparison of the various dimensioning procedures. The M-Star constellation (see Appendix A for details) and its ISL topologies T1 and T2, which have been designed in Section 5.2, are used as reference examples throughout this section.



**Figure 6.16:** Comparison of alternative cost metrics for traffic adaptive routing (DBF) with respect to plain shortest path routing (MDA).

### 6.2.1 Scenario and Assumptions

All numerical studies presented in this section are based on a normalized homogeneous *network* traffic scenario, i.e. each OD demand pair generates traffic of one bandwidth unit over all steps; note that this does not equal homogeneous traffic as input from the ground, mainly due to the specific multiple coverage characteristics of the considered constellations. In the following, the terms traffic (load), capacity, and bandwidth are used synonymously. Restricting to uniform network traffic allows the isolation of topological impacts on the dimensioning process and on the dimensioning results. A heterogeneous traffic scenario is likely to mask all or the most of such topological influences – at least it seems impossible to really separate the effects from the numerical results later.

Besides that, it is stressed that the intention here is to implement and perform the dimensioning *method* and show the *fundamental* optimization potential rather than to evaluate the numerical performance in realistic traffic scenarios. The presented method is of course applicable to any traffic scenario, and one should particularly expect that for less homogeneous network traffic the optimization gain will be even more distinguished, at least for the earlier identified target functions TF1 and TF2 (page 97 in Section 5.4). In this context it is especially important to note that the dimensioning method itself is clearly decoupled from multiple coverage or satellite diversity effects. The influence of both has certainly to be considered to derive effective network traffic flows correctly from given user distributions in a realistic scenario; however, from the dimensioning viewpoint, the *result* of such effects is already captured in the corresponding demand pair capacities  $n_w$  and therefore implicitly included in standard input data to the dimensioning task.

Extensive dimensioning runs have been performed for both topologies T1 and T2, where  $K = 18$  (for T1) and  $K = 10$  (T2) best paths have been in the complete set resulting from KSPA. Representative results are presented and discussed using T1 as example, after a basic comparison between T1 and T2.

In all cases,  $S = 50$  discrete steps have been considered, corresponding to a time step size  $\Delta t = T/S = 112.7/50 \approx 2.25$  min for the M-Star constellation. This is a reasonable choice, trading off accuracy and computer time.

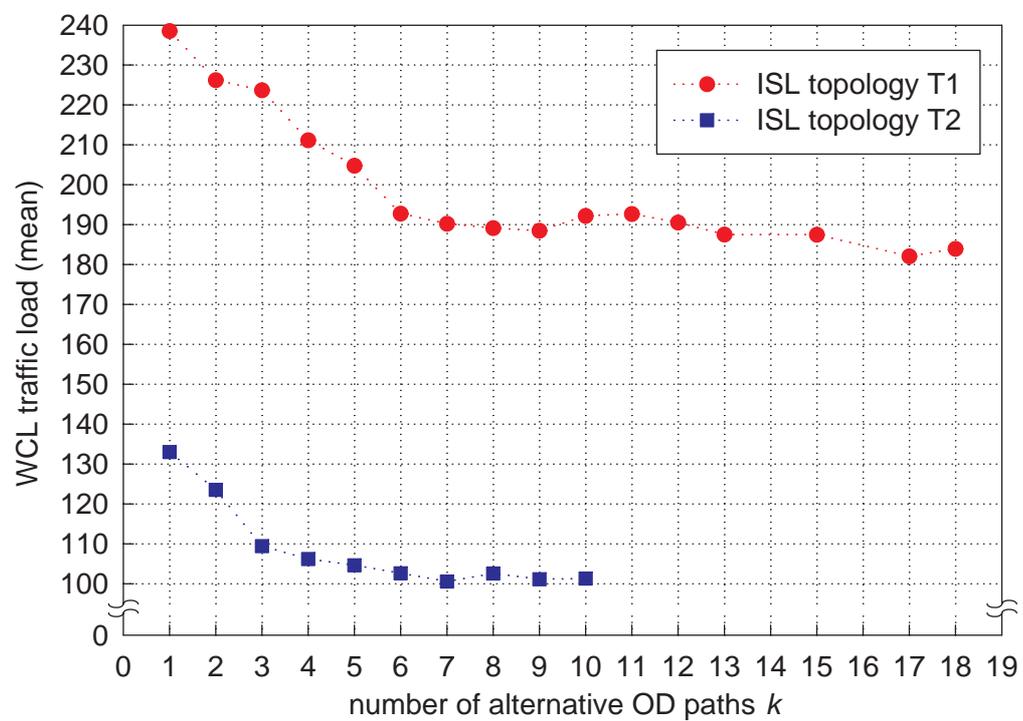
## 6.2.2 Isolated Single-Step Dimensioning

Throughout this section the indicative letter **i** in the short label names for the **i**solated approaches is intentionally omitted for two reasons: first, given that any possible confusion is excluded, for better readability; second, and more important, because *qualitatively* most results of the isolated step approaches hold for the corresponding history-based ones as well.

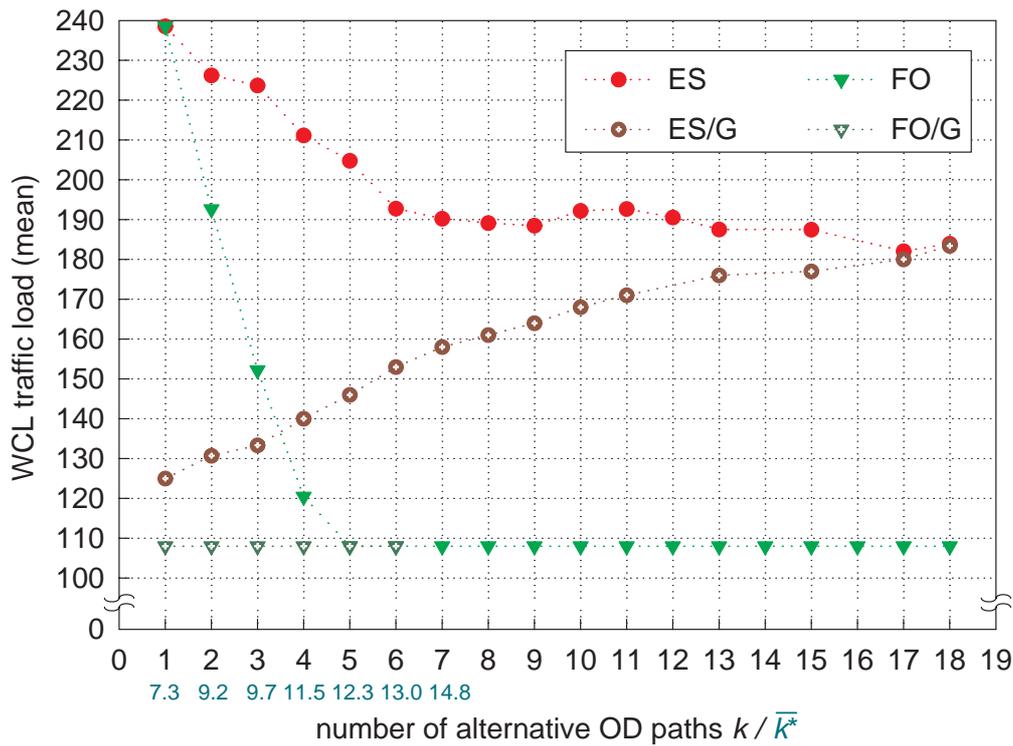
### 6.2.2.1 Worst-Case Link Traffic Load

First of all it is interesting to study the performance of the different dimensioning approaches in terms of the traffic load on the most loaded physical link (the WCL), since this has been our primary objective value. The dimensioning has always been performed over all steps, and one important observation is that the variation of the WCL load over time is very small. Already if all OD traffic is routed over the respective shortest path ( $k = 1$ ), i.e. without any traffic flow optimization, the WCL load variation over the steps is of less than 10%. And any approach that effectively reduces the WCL value (per step) leads to a significant further reduction of its variation over time as well: the lower the instantaneous WCL target value, the lower also its variation. Based on this observation we can restrict ourselves to presenting and discussing only mean values (calculated from 50 single-step values) in the following.

Figure 6.17 shows the WCL load results for the simple ES approach, comparing T1 and T2. As expected, the values for T2 are significantly lower due to the availability of many more links to distribute the traffic on. With growing  $k$  the WCL target value decreases, but in both topologies “saturation” of the optimization is more or less achieved for  $k = 6$ . The WCL load reduction of ES is then roughly 25% with respect to the case of only routing over the shortest path ( $k = 1$ ).



**Figure 6.17:** Reduction of mean WCL traffic load through ES with growing  $k$ : performance comparison between topologies T1 and T2.



**Figure 6.18:** Mean WCL traffic load versus  $k$ : performance comparison between ES, ES/G, FO and FO/G.

Focusing on topology T1, Fig. 6.18 displays at a glance the performance results for the “extreme” approaches ES, ES/G, FO and FO/G. The first impressive comparison is between the pure ES and FO curves, where FO achieves an *additional* improvement of 45% compared to ES in the saturation zone. With  $k = 5$  it is already possible to use the full optimization potential within T1. When combining the two methods with path grouping, one faces an interesting difference: Whereas FO/G achieves the lowest possible WCL value independently of  $k$ , ES/G shows significant improvements with respect to pure ES for small  $k$ , but then the WCL load monotonously grows with increasing  $k$ , which is a result of the fact that ES forces the equal distribution on all available paths, no matter how they are selected. Obviously, from the WCL perspective alone, the only reasonable implementation of ES/G is then for  $k = 1$ .

Back to the approaches without path grouping, Fig. 6.19 displays the distinct influence of the upper bound for path traffic,  $\alpha$ , in the BO approach. Figure 6.20 illustrates how the BO method performs compared to the extreme cases ES (BO with  $\alpha = 1/k$ ) and FO (BO with  $\alpha = 1$ ). As expected, the additional constraint reduces the optimization potential with respect to the WCL target value. However, with moderate values of  $\alpha$  the results come close to the FO ones in the saturation; for  $\alpha = 0.6$ , the WCL load is less than 10% higher compared to FO.

Concluding the observations of the WCL load target, Fig. 6.21 shows the excellent performance of BO when path grouping comes in again: BO/G with a moderate  $\alpha = 0.4$  achieves nearly the same WCL load as FO/G. Combining as many advantages as possible, BO/G with  $k = 5$  and  $\alpha = 0.4$  is consequently a top candidate for our ISL dimensioning problem.

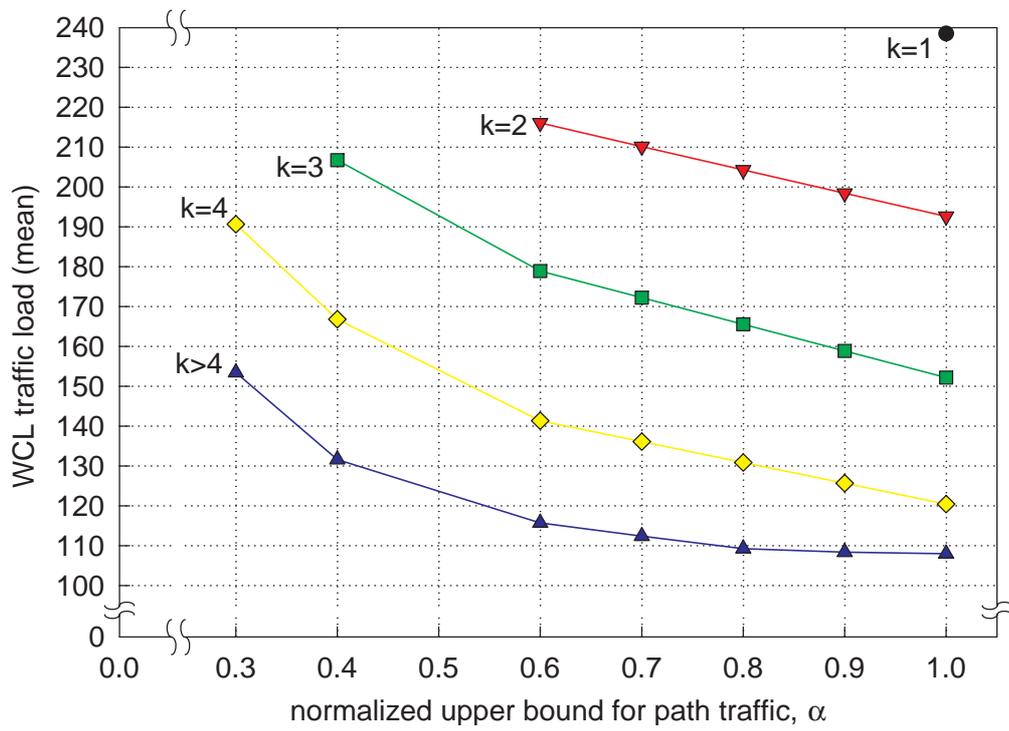


Figure 6.19: Influence of the upper bound for path traffic,  $\alpha$ , in the BO approach.

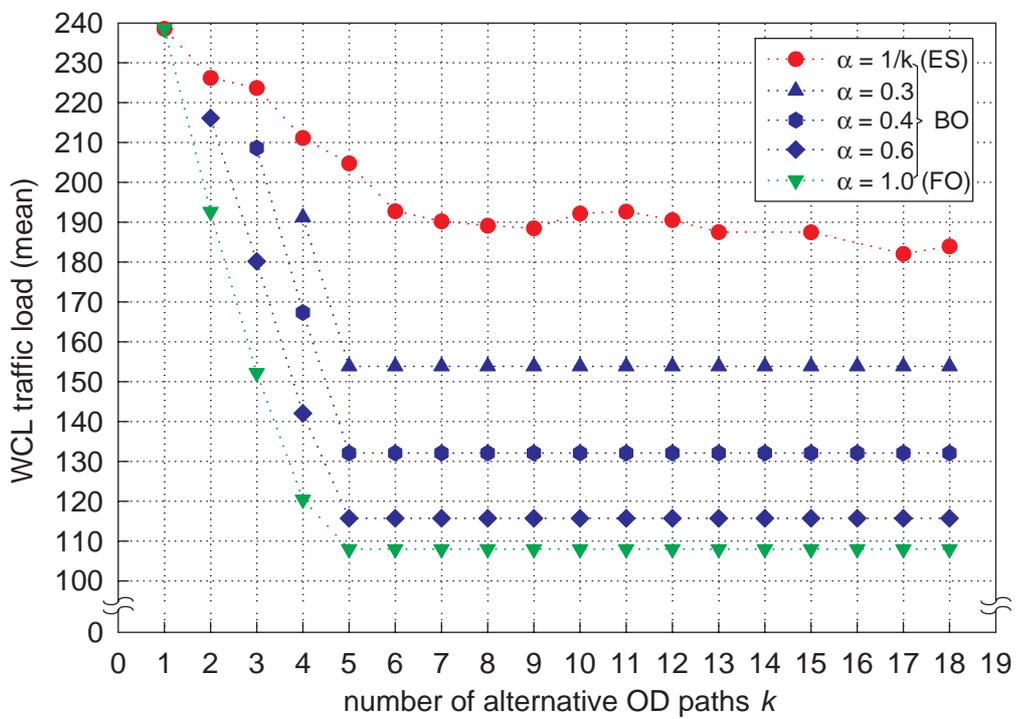


Figure 6.20: Mean WCL traffic load versus  $k$ : performance comparison for different values of  $\alpha$  in the range between the two extreme cases, ES and FO.

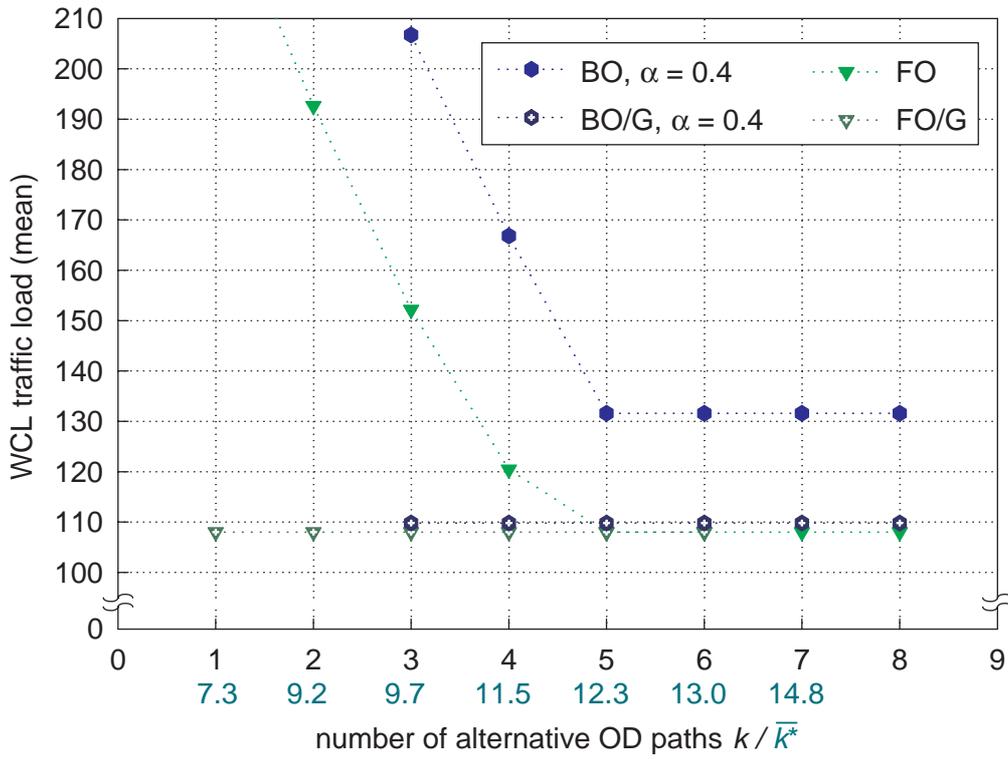


Figure 6.21: Mean WCL traffic load versus  $k$ : performance comparison between FO/G and BO/G.

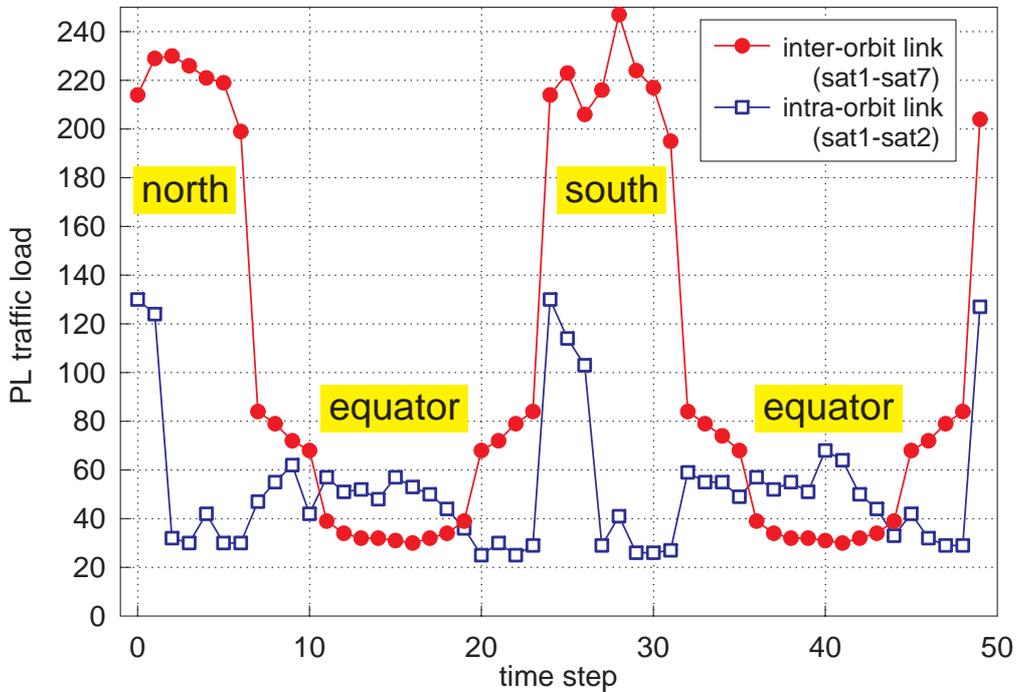


Figure 6.22: Traffic load on selected PLs over one constellation period.

### 6.2.2.2 Physical Link Traffic Load

In order to understand *how* the various traffic distribution and optimization approaches operate within the topology, it is helpful to study the load variation on fixed physical links (PLs) over the steps. Figure 6.22 displays the load over time for two selected PLs, one intra- and one inter-orbit link. In both cases we see pronounced load peaks when the respective link is over mid latitudes (i.e. the highest latitudes that the satellites reach with  $47^\circ$  inclination), whereas it is low near the equator. It is also obvious that the WCL is always an inter-orbit ISL. Concluding from both observations, WCL load reduction should typically work by shifting traffic away from those inter-orbit ISLs that are in the critical region at a certain step. Figure 6.23 confirms this expectation and illustrates the superior performance of the FO approach with respect to this shifting of traffic. For  $k = 5$ , FO leads to a constant PL load, and one can already expect from observation of this example link that this is the final limit for any optimization working on this topology – a conclusion which is in line with the shape of the FO curve in Fig. 6.20.

Another impressive confirmation of these considerations is given by the PL load distributions over all PLs of a certain step, as displayed in Fig. 6.24 for step 0 (the shape of the curves being nearly identical for other steps). It is obvious how traffic from higher loaded links is “shifted” to lower loaded ones, and a complete balance is achieved for the inter-orbit ISLs when FO with  $k \geq 5$  is used. In reality, of course, certain traffic is not just shifted from one link to another but OD traffic is shifted to other paths – sometimes to paths with more hops – resulting in an increase of *average* PL load in the network with increasing  $k$ . This is one of the prices to be paid for WCL load minimization, and it is quantified in Fig. 6.25. Again, the superior performance of FO becomes obvious; whereas ES with a reasonably high  $k = 10$  leads to an increase of more than 20%, the effect is nearly negligible for the FO counterpart.

Figure 6.26 summarizes pictorially some of the above-mentioned phenomena.

### 6.2.2.3 Path Costs

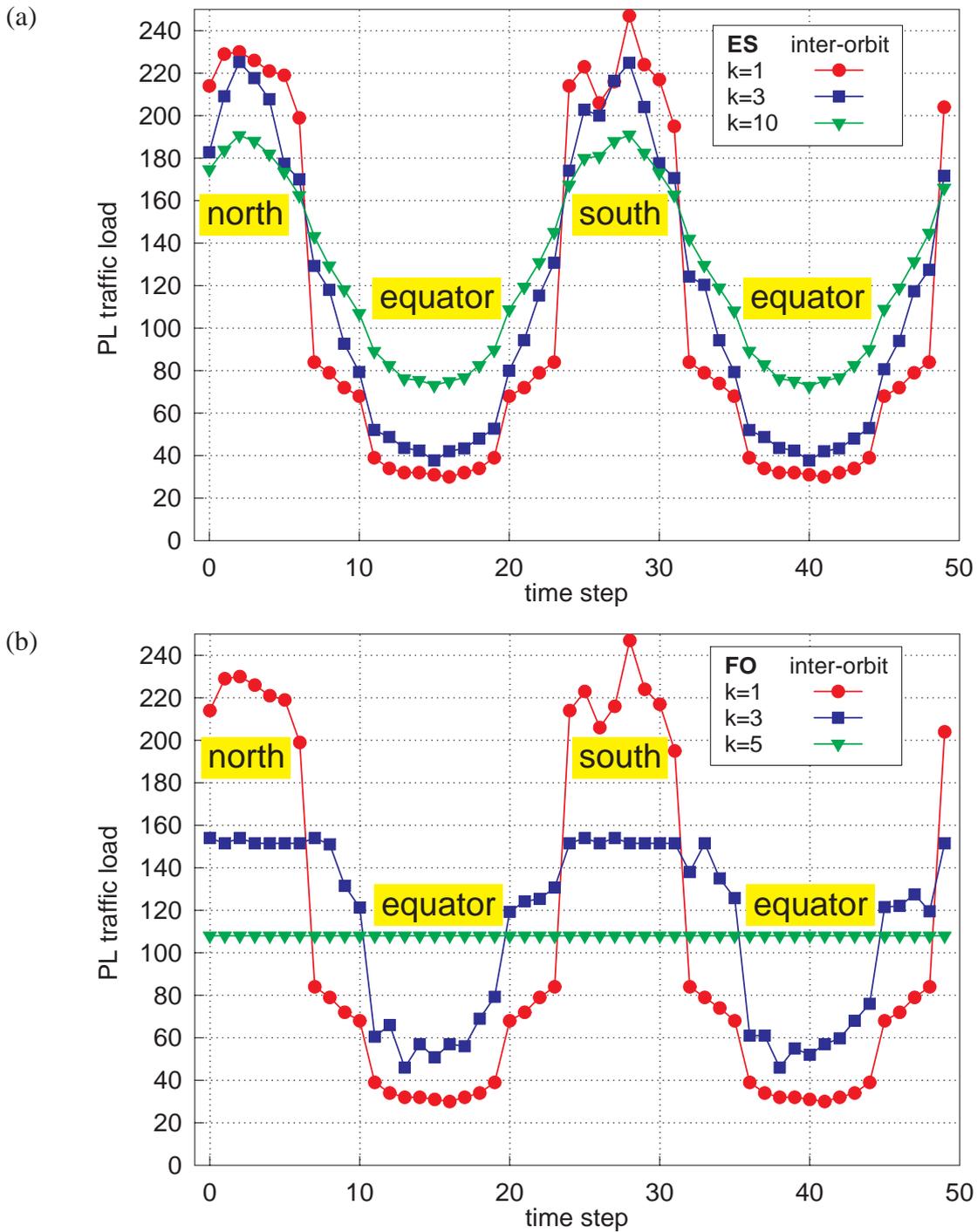
The most important trade-off, considering also the user or QoS perspective, is certainly between the WCL load minimization target (which is mainly driven by a network performance and system cost viewpoint) and the correlated degradation in terms of path delay encountered. The average costs in terms of both number of hops and path delay are summarized in Fig. 6.27 for selected approaches. One basically observes that with increasing optimization freedom (from ES over BO to FO) not only the WCL load target value, but also the average path delay can be reduced for a given  $k$ . Moreover, looking on the two most promising candidates identified above, namely FO/G and BO/G with  $\alpha = 0.4$ , both for  $k = 5$ , the increase of path delay is fairly low around 4 to 7% with respect to the reference case of routing each OD traffic completely over the shortest path alone.

## 6.2.3 History-Based Multi-Step Dimensioning

As the history-based approach is essentially a direct outgrowth of the simpler isolated step one, driven by some basic deficiencies identified in the latter, emphasis is in the following laid on direct

---

<sup>1</sup>Note that the earth map is only drawn for reasons of “orientation”, but the land masses are not correlated with the source traffic assumptions underlying the numerical results as represented by the grayscale levels in (b)–(e).



**Figure 6.23:** Equalization of traffic load on a fixed inter-orbit PL through (a) the ES and (b) the FO approaches with varying  $k$ .

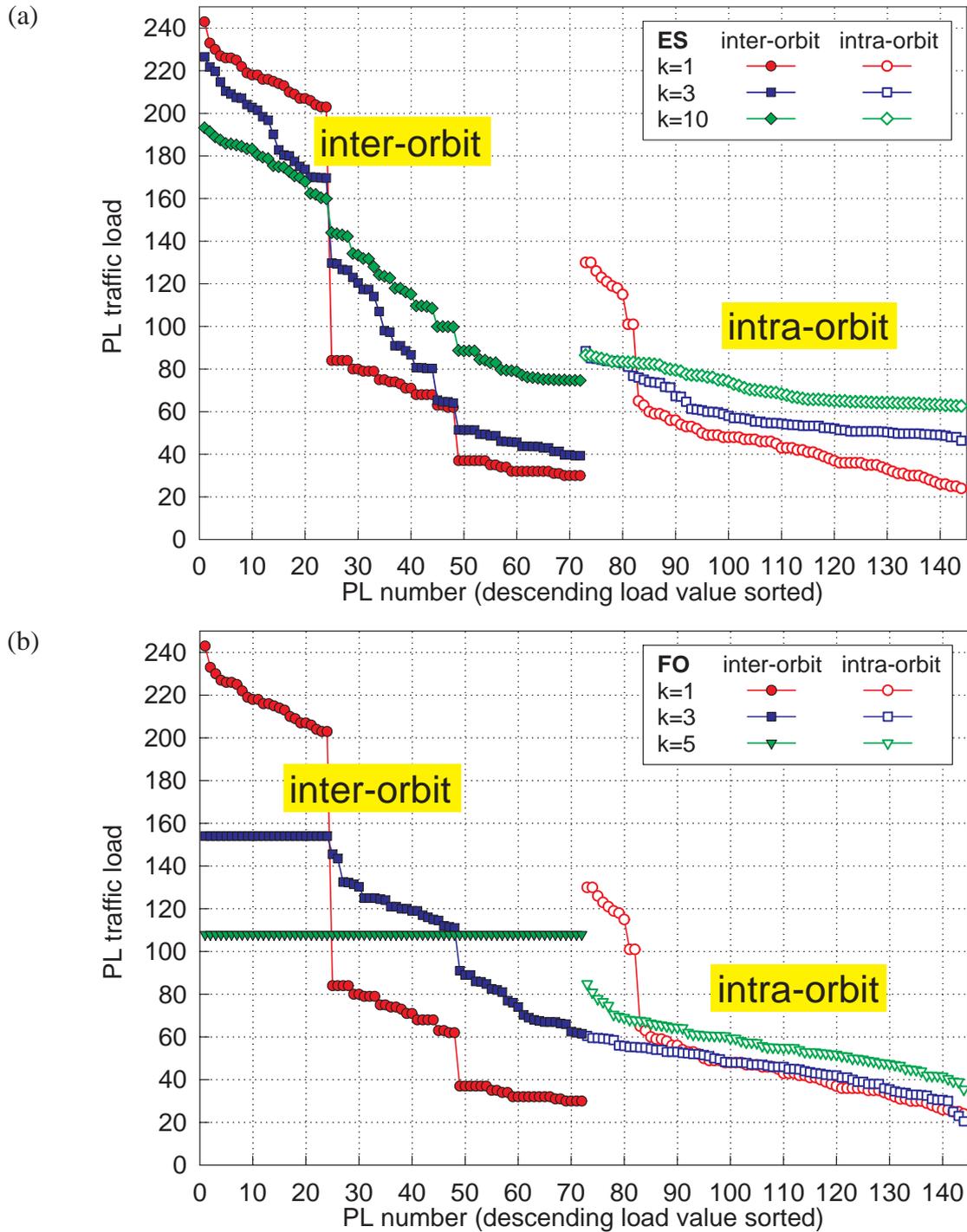
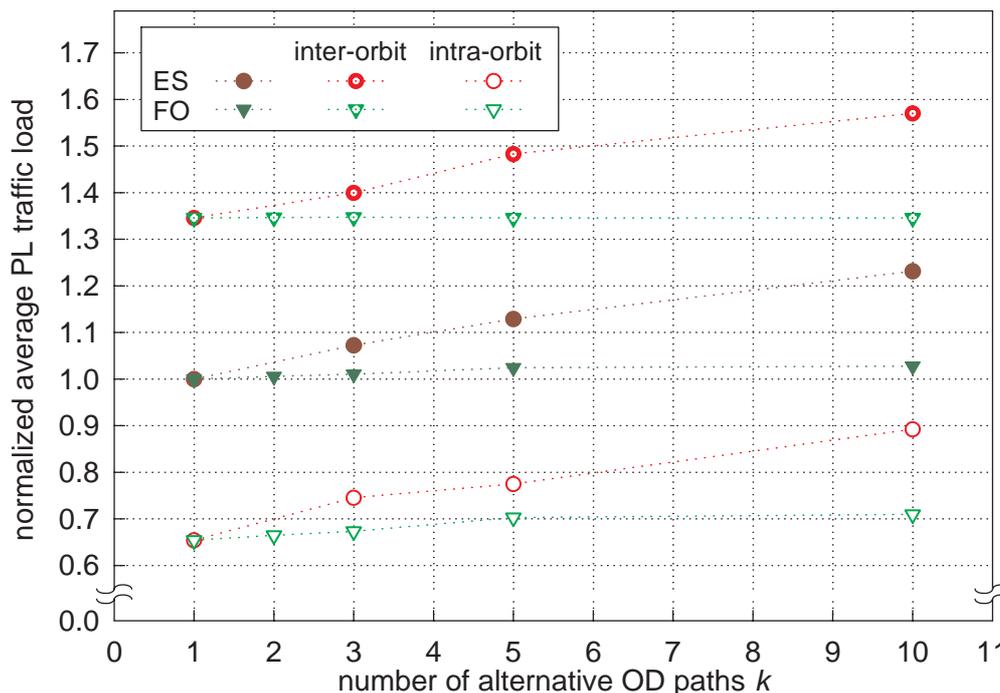


Figure 6.24: PL traffic load distribution in the network at step 0 for (a) ES and (b) FO with selected  $k$ .



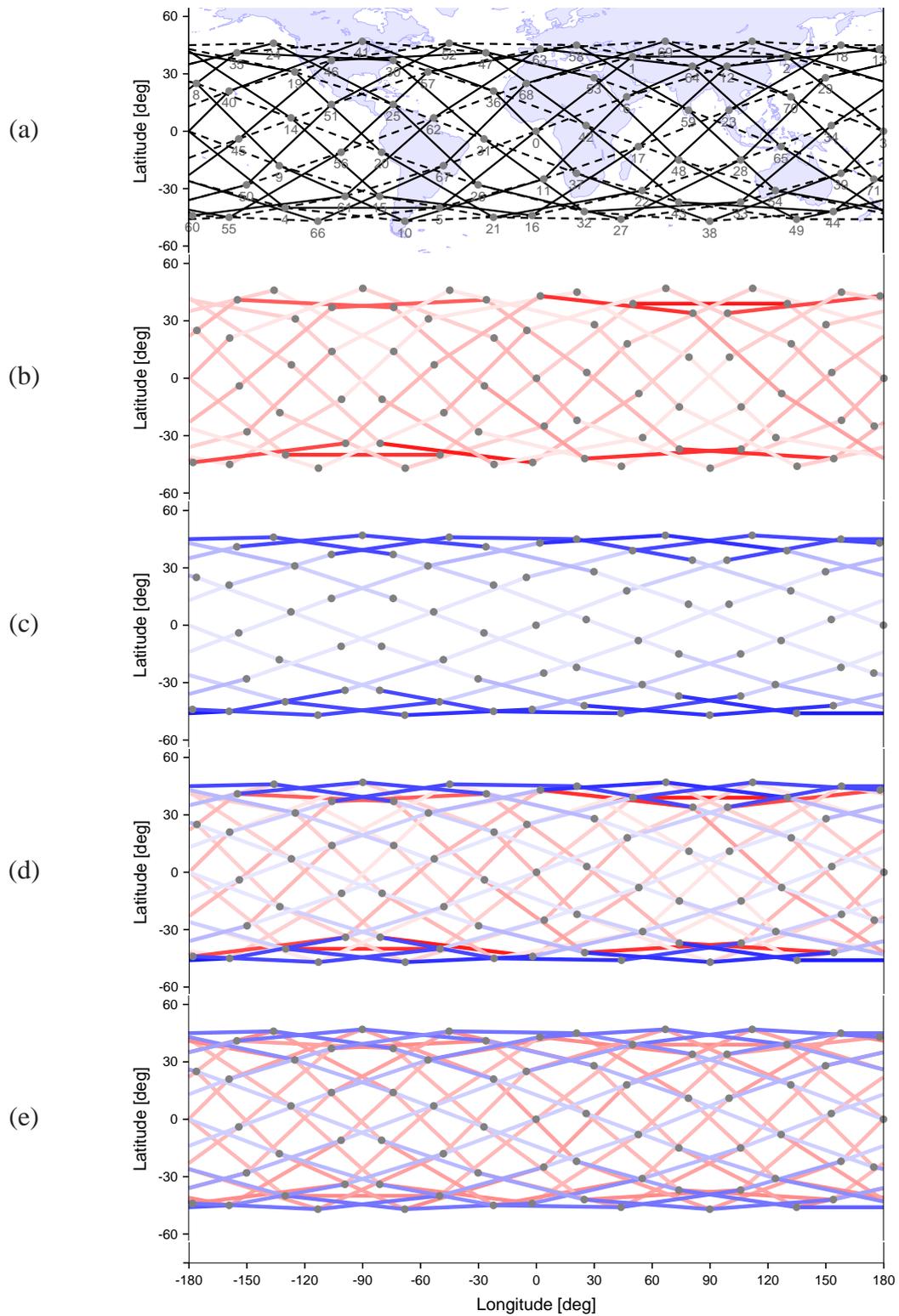
**Figure 6.25:** Normalized average PL traffic load versus  $k$ , and corresponding splits into averages for inter- and intra-orbit links, respectively.

comparison between both. We now return to the complete short labels for all approaches, including the indicative letters **i** and **h**, except when both corresponding procedures are meant at the same time (for instance, when stating a characteristic of ES in general, including both iES and hES implementations).

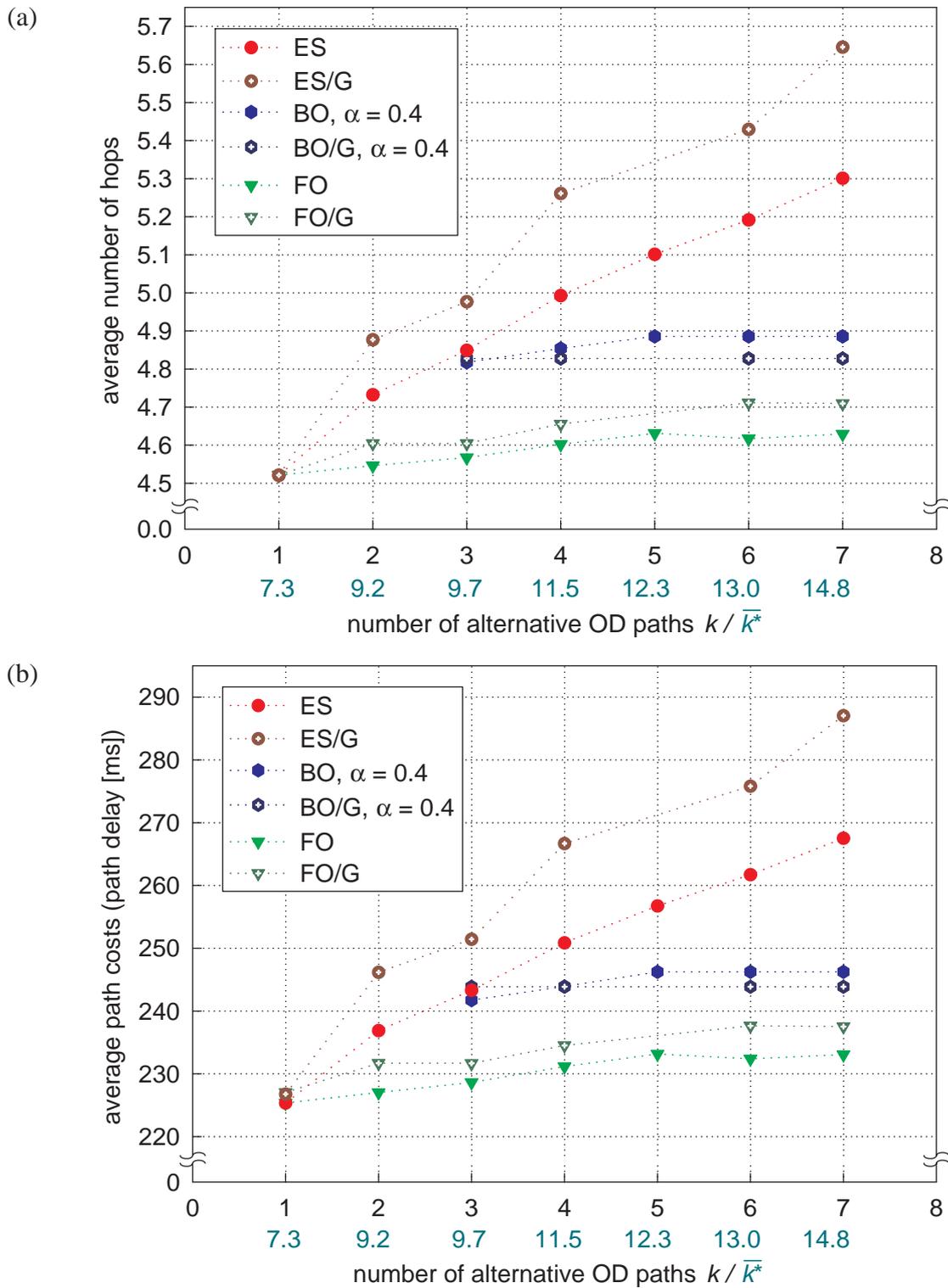
### 6.2.3.1 History Modeling

Generally, the assumptions outlined in Section 6.2.1 hold as for the isolated step approach. We stress again that for numerical studies we assume a network uniform *and single-service* traffic scenario in order (i) to isolate and understand the dynamic topology’s impact on dimensioning results and (ii) to allow a focused study of some general optimization issues, with a minimum disturbance through complicated dynamic and inhomogeneous traffic patterns. In addition to the modeling and simulation assumptions outlined so far, we present the two components for the particular “modeling of history” in the following.

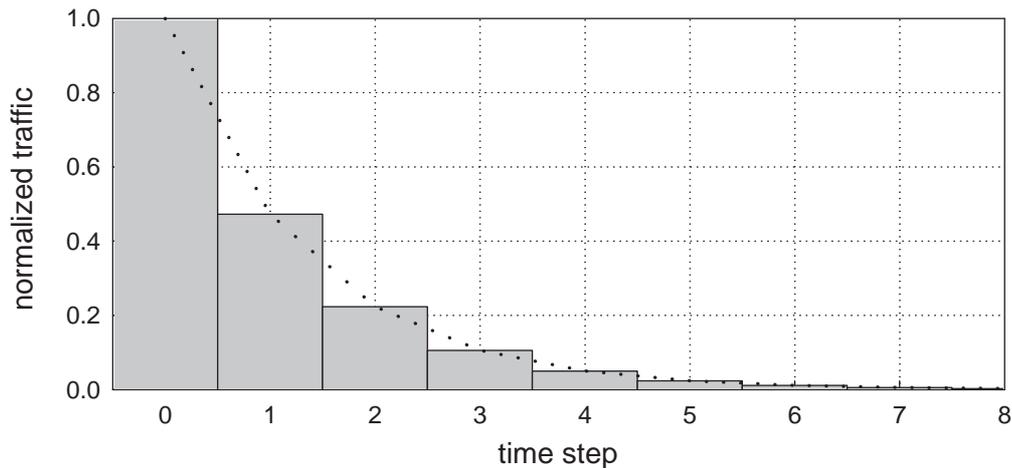
**Call Model** Single-service symmetric calls with Poisson distributed arrival and negative exponentially distributed call holding times are assumed. The mean call holding time (MCHT), being an important parameter for the history-based optimization, is set to 3 minutes except in those investigations where it has been explicitly varied to study its impact on the results. Figure 6.28 displays the normalized shares of traffic remaining active after  $\tau$  steps for MCHT = 3 min, discretizing the negative exponential curve into steps with duration  $\Delta t$ . The numerical value of the arrival rate is not interesting in our assumed homogeneous and normalized traffic scenario. This is due to the fact that the assumed homogeneity applies to OD pair traffics and *not* to the global distribution



**Figure 6.26:** Study of PL traffic distribution in the (a) M-Star topology T1 at step 0 (higher traffic load is illustrated in darker grayscale levels)<sup>1</sup>: (b)/(c) comparison of intra-/inter-orbit ISL traffic without optimization; (d)/(e) comparison of network traffic without/with optimization (ES/G with  $k = 3$ ).



**Figure 6.27:** Cost comparison in terms of (a) hops and (b) path delay for different (WCL) optimization approaches versus  $k$ .



**Figure 6.28:** Discretized remaining traffic for negative exponentially distributed call holding time with mean value MCHT = 3 min.

of user positions on ground. In other words, the (i) arrival rates together with the (ii) number of users in the respective source/destination footprints (uniform distribution *within* both) and the (iii) coverage (including time-varying multiple coverage situations) are “forced” to jointly guarantee a normalized steady-state traffic equivalent to one bandwidth unit respecting the MCHT value; concrete values of these three components are however not relevant for the optimization.

**Rerouting Model** The assumption of a uniform distribution of active users within the footprint translates into a steady state behavior where the number of incoming handover calls are equal to the number of outgoing ones for each satellite serving the ground. For a simple quantitative estimation, let the footprint be modeled by a square traveling along an *equivalent street of coverage*, equivalent in the sense that we apply the original *street of coverage* concept – as proposed by Lüders [Lüd61] and elaborated by Rider [Rid85] and Adams and Rider [AR87] for polar orbit constellations – to inclined delta constellations. Then the probability that a call of age  $\tau$  (steps) is handed over from/to a certain satellite, equaling the share of the area left/entered in one step relative to the footprint area, can be approximately calculated as

$$p_{\text{HO}}(\tau) = \tau \cdot \frac{N_S}{S}, \quad \tau = 1, \dots, \left\lfloor \frac{S}{N_S} \right\rfloor, \quad (6.6)$$

where  $N_S$  is the number of satellites per orbit plane and  $S$  is the number of discretization steps for the constellation period. After  $\lfloor S/N_S \rfloor$  steps the satellite covers a completely new area on ground and has thus already handed over all calls exceeding this age.

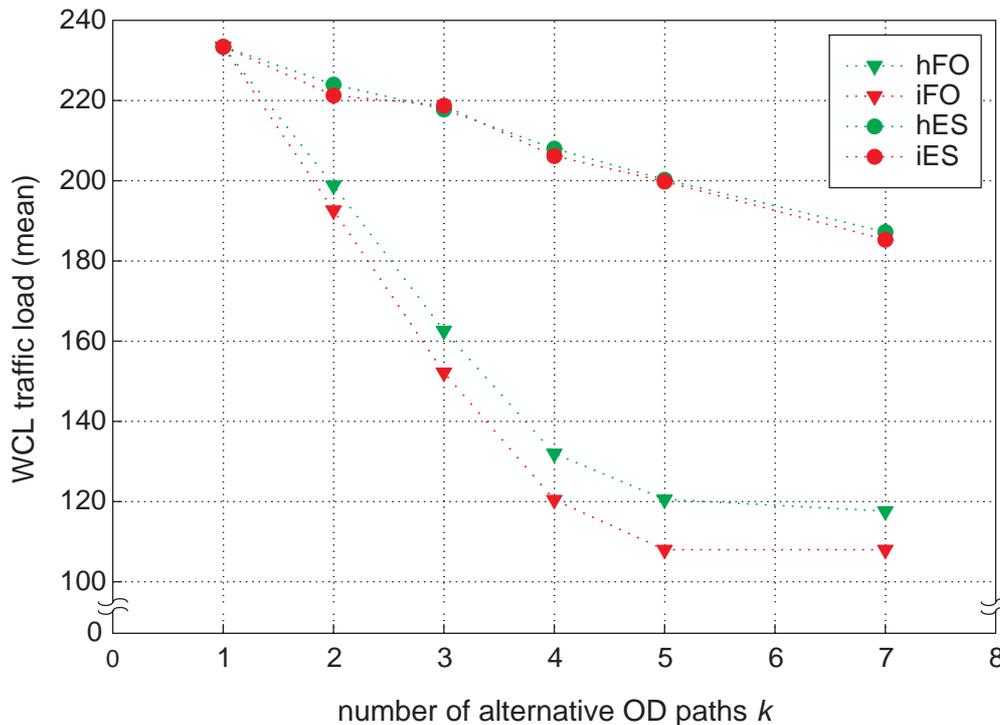
Assuming furthermore independence of the handover situations encountered at the two satellites of an OD pair, the probability that an OD call of age  $\tau$  is rerouted, becomes

$$p_{\text{rer}}(\tau) = 1 - (1 - p_{\text{HO}}(\tau))^2. \quad (6.7)$$

Finally, assuming that all calls of an OD pair at step  $s$  are uncorrelated, the value  $p_{\text{rer}}(\tau)$  can also be interpreted as the relative share of rerouted OD traffic of age  $\tau$ . Table 6.2 summarizes the

**Table 6.2:** Rerouted traffic shares as a function of the age of calls.

$\tau$	1	2	3	4	5	6	7	8
$p_{\text{HO}}(\tau)$	0.12	0.24	0.36	0.48	0.60	0.72	0.84	0.96
$p_{\text{rer}}(\tau)$	0.23	0.42	0.59	0.73	0.84	0.92	0.97	1.0

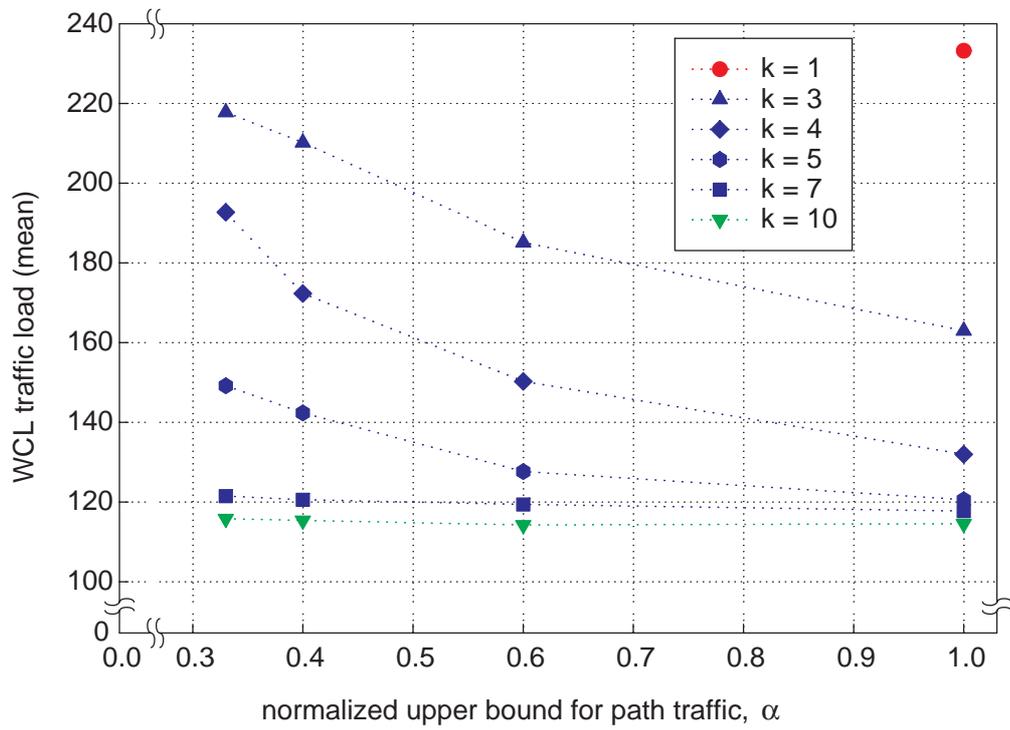
**Figure 6.29:** Mean WCL traffic load versus  $k$ : performance comparison between isolated step and history-based approaches.

resulting numerical values for the considered study case. In the simulations, the age of all active calls (or capacity shares) is tracked, in order to calculate in each step  $s$  the correct cumulative share of rerouted traffic according to the call and rerouting models. The same applies of course to the cumulative remaining traffic. As implicitly formulated already in (6.6), in our numerical example  $\hat{\tau}_{\text{old}} = \lfloor S/N_S \rfloor = 8$  would be the appropriate choice respecting all relevant remaining and rerouted traffic contributions as defined in (5.12) and (5.13), respectively.

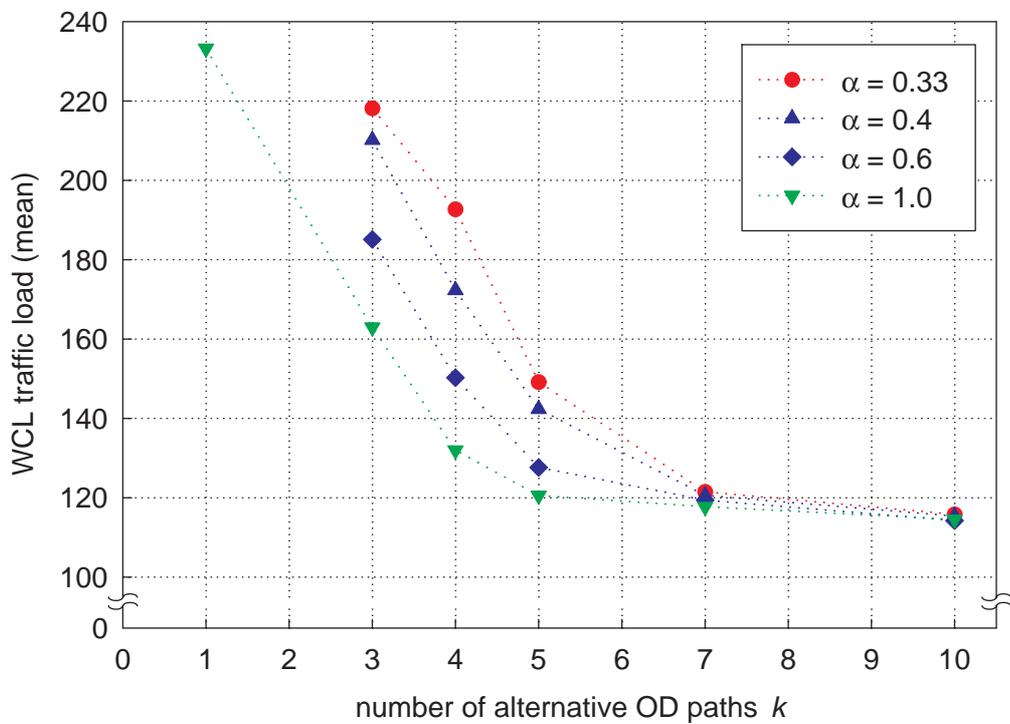
### 6.2.3.2 Worst-Case Link Traffic Load

First of all it is interesting to compare the performance of the different dimensioning approaches in terms of the traffic load on the most loaded physical link (the WCL), since this has been our primary objective value.

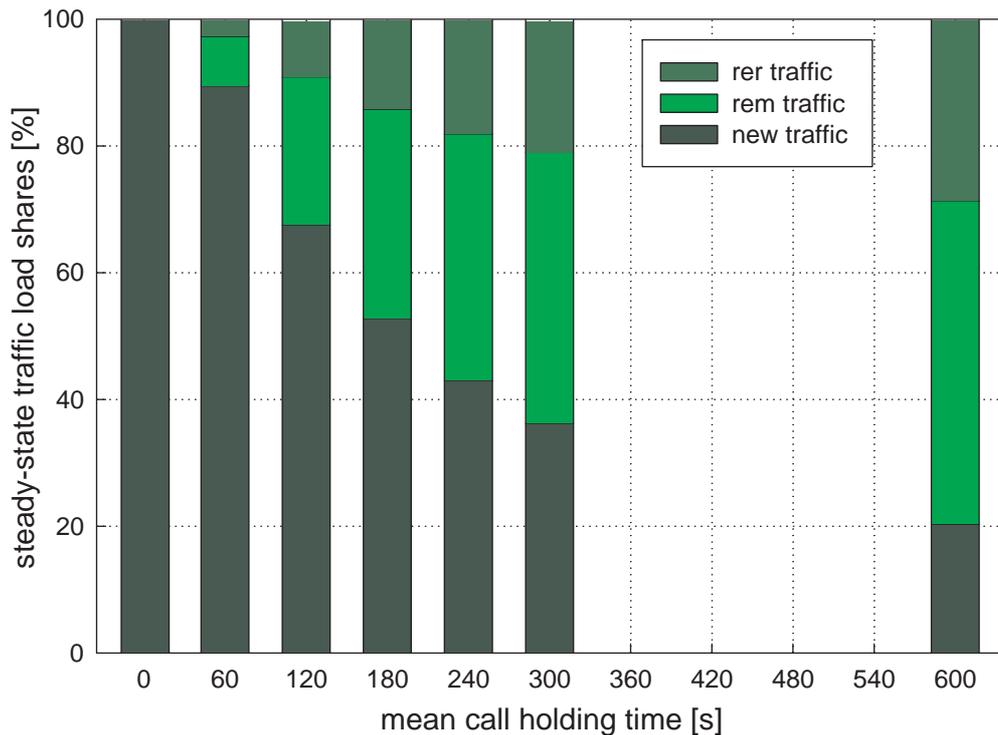
Figure 6.29 displays at a glance the performance results for the “extreme” approaches iES, hES, iFO and hFO. The first impressive comparison is between the respective ES and FO curves, where FO achieves an *additional* improvement of approximately 45% compared to ES in the saturation zone, starting roughly with  $k = 5$ . For the interesting FO case, one can clearly see that with the



**Figure 6.30:** Influence of the upper bound for path traffic,  $\alpha$ , in the hBO approach.



**Figure 6.31:** Mean WCL traffic load versus  $k$ : performance comparison for different values of  $\alpha$  in the hBO approach.



**Figure 6.32:** Sharing between the traffic components in the steady state (rem = remaining, rer = rerouted).

history-based approach the optimization gain is roughly 10% smaller than with the isolated step approach, which gives a first quantitative impression in which order of magnitude the optimization results for the simple isolated step approach have been wrong, namely “too good”.

Figure 6.30 shows the distinct influence of the upper bound for path traffic,  $\alpha$ , in the hBO approach, and Fig. 6.31 illustrates how the hBO method performs compared to the extreme cases hES (hBO with  $\alpha = 1/k$ ) and hFO (hBO with  $\alpha = 1$ ). Similarly as in the isolated step case, the additional constraint reduces the optimization potential with respect to the WCL target value, but again, with moderate values of  $\alpha$  the results come pretty close to the FO ones in the saturation.

Finally, the impact of the history, established with the call and rerouting models presented at the beginning of Section 6.2.3, is of particular interest. Figure 6.32 illustrates the characteristic stationary sharing between the three traffic components – new, remaining and rerouted – in the steady state (achieved roughly after 8 steps). Recalling that only new and incoming rerouted traffic can be freely routed in each step whereas the remaining traffic must stick to the path once chosen, one gets a fairly good impression how far the earlier isolated step optimization potential (all traffic is freely reroutable) is now reduced as a function of MCHT in the history-based approach. For instance, in case of our reference MCHT = 3 min, the “blocked” share of total traffic with respect to routing optimization (i.e., the remaining traffic share) amounts to some 30% whereas it exceeds 50% for mean call holding times over 10 minutes. Figure 6.33 provides the resulting performance in terms of WCL load reduction for different MCHT values over time. As expected, both the overall WCL load and also the WCL load variation over time clearly increase with larger MCHT values.

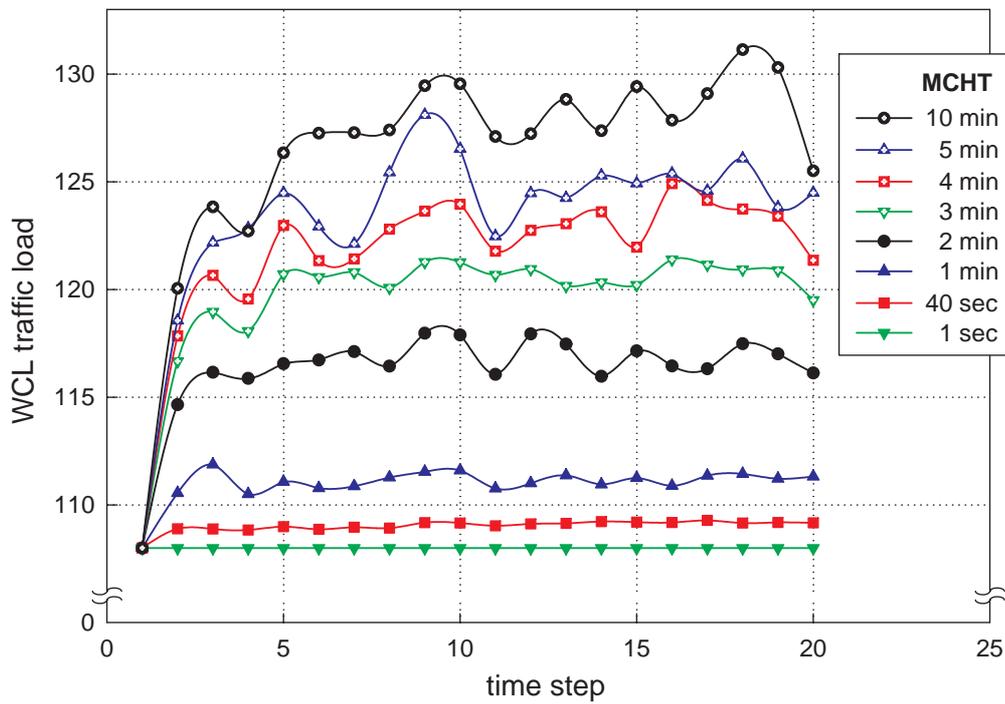


Figure 6.33: Evolution of the WCL value for different MCHT values.

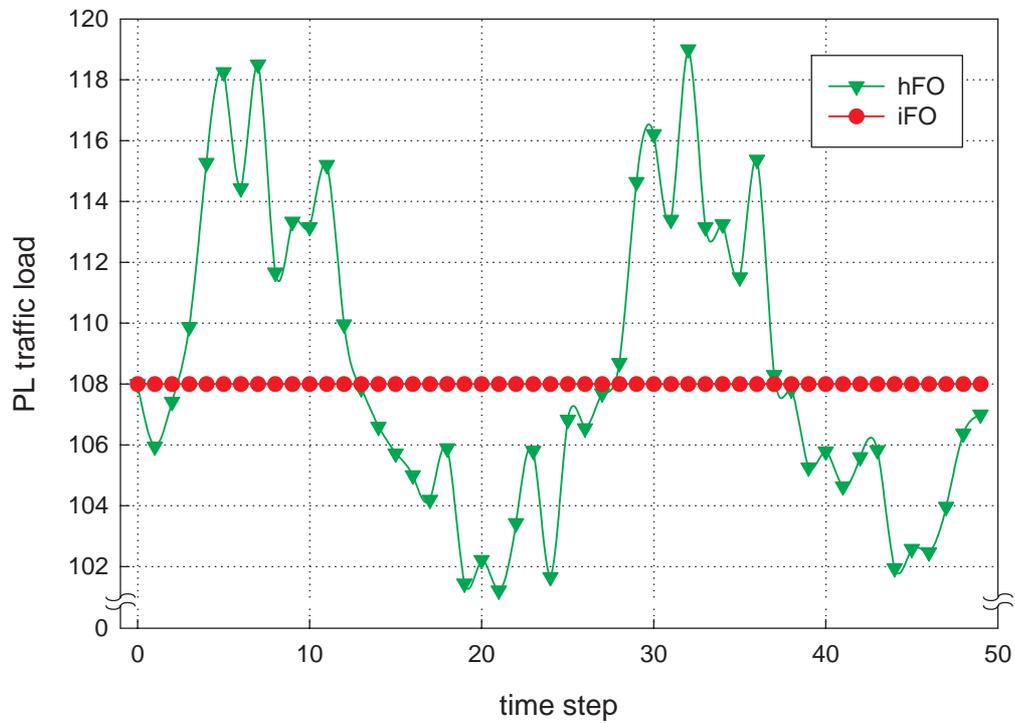


Figure 6.34: Different traffic load shaping on an inter-orbit PL for iFO and hFO with  $k = 5$ .

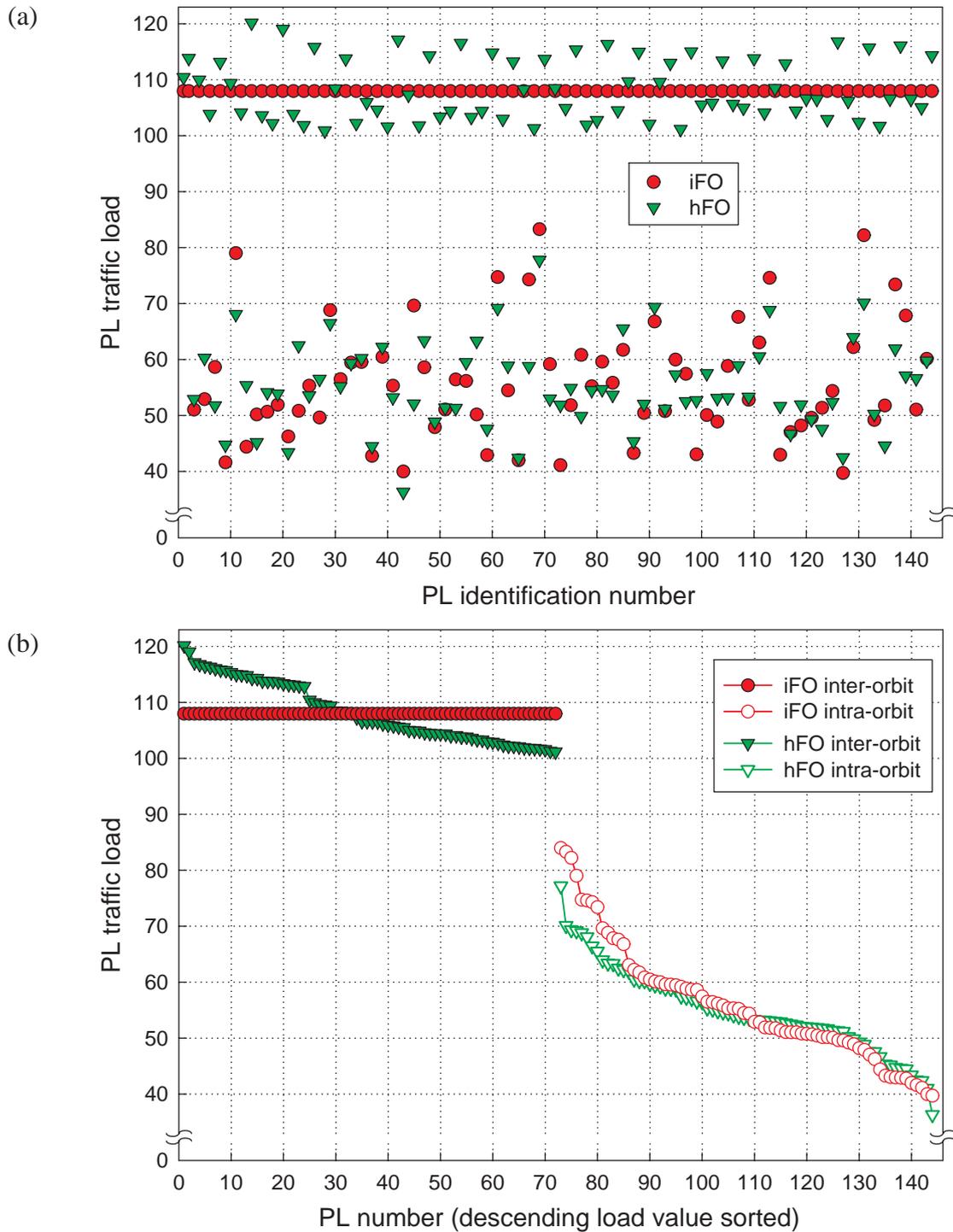
### 6.2.3.3 Physical Link Traffic Load

As it has been discussed in Section 6.2.2 (more precisely, in the context of Fig. 6.22) before, the inter-orbit ISLs tend to be generally higher loaded (in both average and peak values) than the intra-orbit counterparts, and are therefore more important with respect to WCL considerations. It has also been shown (Fig. 6.23) that WCL load reduction typically works by shifting traffic away from those inter-plane ISLs that are in the critical mid-latitude regions at a certain step. Figure 6.34 provides a performance comparison of iFO and hFO for this effective network traffic “shaping”, exemplarily displaying the PL load variation on one inter-orbit link. It becomes obvious that respecting the traffic correlation over the steps does no longer allow to reduce the PL load to a constant value as with iFO, again proving the slight but systematic error made in this isolated step approach. This is fully confirmed by the PL load distributions over all PLs of a certain step in the steady state, as displayed in Fig. 6.35.

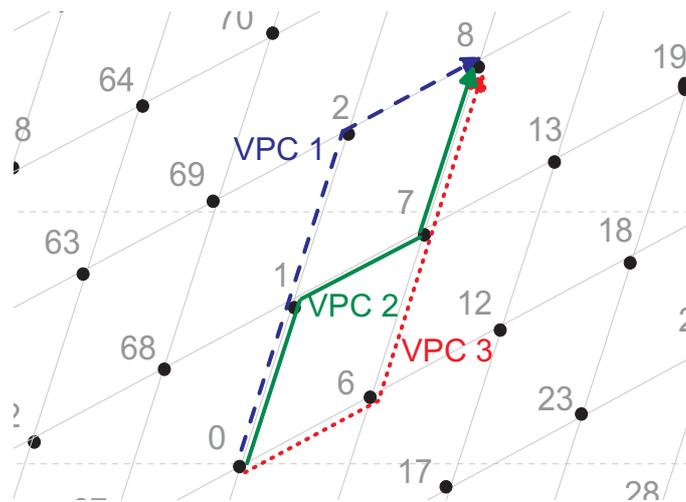
### 6.2.3.4 Path Traffic Load

In the context of isolated step dimensioning, we have not presented numerical results for the VPC (i.e. path) traffic load over time, because due to the very nature of the approach any continuity of path traffic has been neglected anyway. With complete isolated optimization freedom, the time variation of traffic on a particular path somewhat reflects the combined peculiarities of (i) the optimization algorithm and (ii) the systematic presentation of the input data to the LP optimization module. Finding a global optimum in the studied uniform network traffic scenario has often shown to result in solutions where one of the available OD paths carries the whole OD traffic (for iFO, without an upper limit for the load share) or the maximum allowed share of it (for iBO), whereas the other paths carry zero or the remaining share traffic. Monitoring the load on a particular path over time therefore often shows the extreme discontinuities between two subsequent steps of carrying the maximum traffic in one and the minimum traffic in the next step, or vice versa. In between such discontinuities, the time (number of steps) which the path load remains at a certain level is arbitrarily distributed and has no relation to “reality”. This is just one key aspect where the history-based optimization should bring in a clear difference, because it explicitly relates traffic components in subsequent steps with each other. As an example we consider the OD pair 0–8 in Fig. 6.36 and  $k = 3$  for history-based optimization. Obviously, the three displayed paths (VPC 1, VPC 2, and VPC 3) make up the 3-hop shortest-path group for this OD pair over all steps of the constellation period, so that for  $k = 3$  the whole OD traffic is always exclusively shared between these three paths. Figure 6.37 illustrates the normalized traffic load shares over time, again based on our reference MCHT = 3 min. Comparing (a) hFO and (b) hBO with  $\alpha = 0.6$ , already at first sight the increased smoothing of VPC traffic over time becomes obvious – also compared to the above-mentioned behavior of iFO. Looking at the data more closely, one can observe that maximum “step-to-step jumps” of VPC traffic for hFO are limited to approximately 0.67, reflecting the 33% share of remaining traffic that is not allowed to be switched to other paths. This effect is overlaid with the forced load sharing between the VPCs due to the upper path traffic bound of (normalized) 0.6 for the hBO dimensioning.

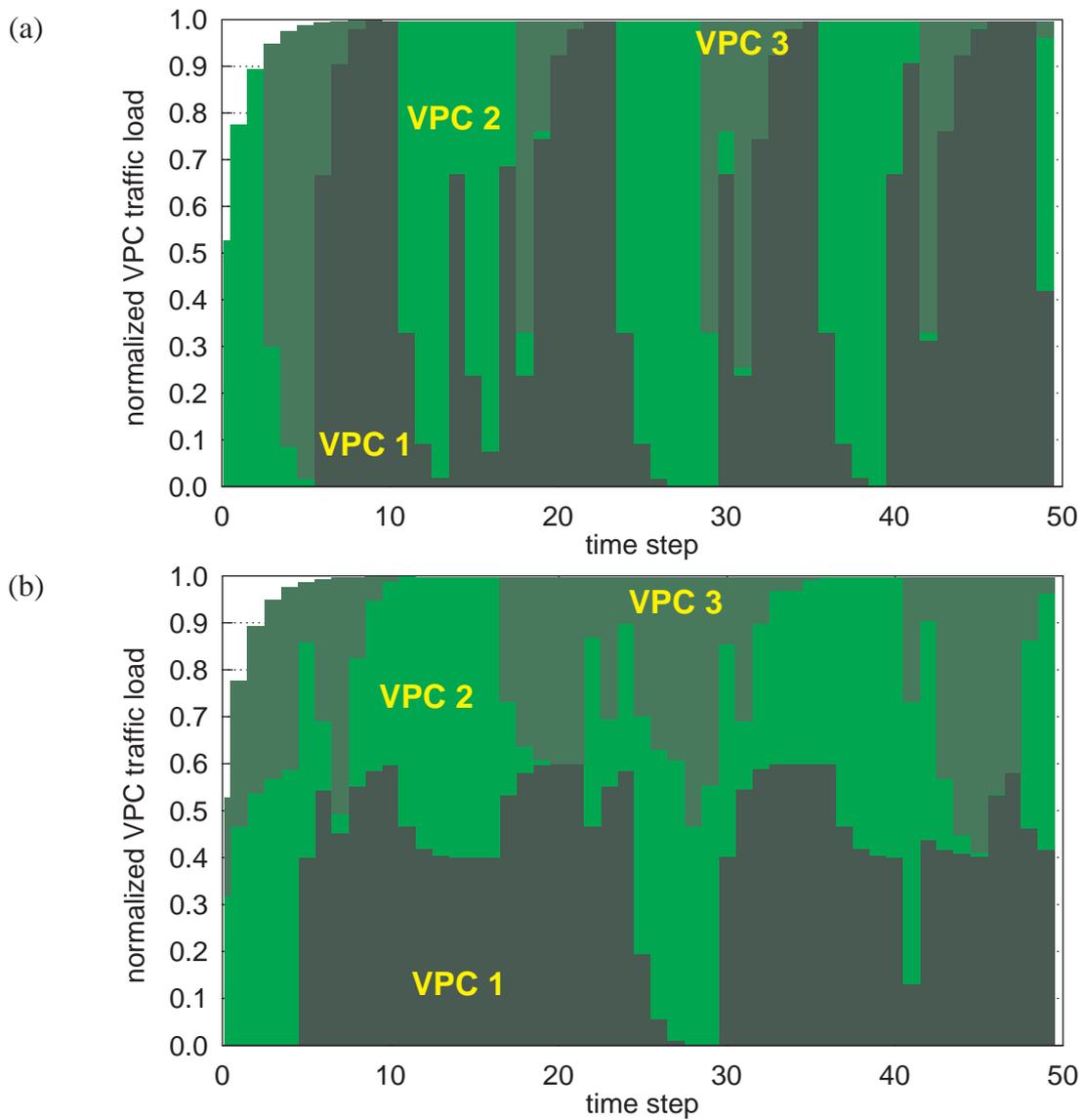
Concluding the overall numerical comparison between the isolated step and history-based optimization approaches, we have clearly shown the systematic error in the former. In the studied



**Figure 6.35:** PL traffic load distribution in the network for iFO and hFO with  $k = 5$  (a) over PL ID sequence, (b) over PLs sorted in terms of decreasing load value.



**Figure 6.36:** 3-hop shortest-path group for OD pair 0–8, consisting of VPC 1, VPC 2 and VPC 3.



**Figure 6.37:** Splitting of OD traffic on  $k = 3$  VPCs versus time for (a) hFO and (b) hBO with  $\alpha = 0.6$ .

homogeneous network traffic scenario, this error is clearly limited, so that the isolated step dimensioning could be considered as an attractive low-complexity method for a first and quick step in a usually iterative network dimensioning process. Qualitatively this should hold for inhomogeneous network traffic, too. When it comes to quantitatively “precise” dimensioning for a concrete system under real-world traffic conditions, however, one should definitely control the performance of the two approaches in the particular case, as the impact of the systematic error may become more severe and history-based optimization may become a necessity.

# Chapter 7

## Conclusions

*If you can look into the seeds of time  
And say which grain will grow and which will not,  
Speak then to me . . .*

— WILLIAM SHAKESPEARE, *Macbeth* (Act I, Scene 3)

This work has focused on routing and dimensioning in ISL-based LEO/MEO satellite networks with dynamic ISL topology, both being increasingly important issues for the operation and the success of communication satellite constellations.

The core contributions of this thesis are

1. an overall ATM-based networking concept,
2. an off-line dynamic ISL routing concept with on-line adaptive routing extensions, and
3. an integrated network design approach comprising (i) topology design, (ii) routing framework and (iii) capacity dimensioning / routing optimization,

all essentially based on the connection-oriented communication paradigm. Major parts of the concepts are claimed and believed to be generic so as to qualify them for a long “life-cycle” in a rapidly changing communications and networking world in general, and in an economically critical satellite environment in particular.

In Chapter 2, we have reviewed the two basic types of circular-orbit satellite constellations, Walker *star* and Walker *delta* constellations, elaborating on the implications of both approaches for possible ISL network designs. While in delta constellations the tracking requirements between two inter-orbit neighboring satellites tend to be somewhat permanent, but not too extreme over time, the case is different with star constellations: over a wide range of latitude *between* the polar regions, the distance and pointing angle variation between satellites in *co-rotating* planes is fairly low, but it is extreme over the poles and across the seam (between the counter-rotating planes); as a consequence, Iridium as a current reference system for an ISL-based star constellation switches off ISLs over polar regions and has not implemented cross-seam links at all. Especially the latter feature breaks the ultimate regularity of the ISL network. Both the seam and switched links over

the poles impose major challenges on routing in a connection-oriented scenario. Compared to that, a completely regular and permanent network topology may be built upon a delta constellation, with obvious advantages for optimized network operation in terms of traffic routing and capacity dimensioning. We have also shown that multiple coverage is an important goal in constellation design, as it essentially drives the potential of satellite handover and satellite diversity approaches. Finally, we have highlighted the extreme traffic dynamics in constellation networks, which essentially results from spatial traffic load variation as seen by an orbiting satellite. This effect is clearly more pronounced with LEO satellites than with MEOs as they are orbiting the earth much faster.

In Chapter 3, we have developed an overall ATM-based networking concept for LEO/MEO constellation networks. We have chosen to consider ATM as the basis for our satellite constellation networking concept mainly for three reasons:

- The clear trend for future satellite communications is towards multimedia applications and multiservice networks.
- We think that a connection-oriented operation scheme is best capable of exploiting three main properties of dynamic ISL topologies, namely predictability, periodicity, and regularity, while inherently guaranteeing the order of arriving information entities over significant periods of end-to-end connections.
- In the design of broadband multiservice satellite networks with QoS guarantees and service differentiation, traffic engineering is a crucial component.

The networking concept mainly capitalizes on the virtual connection capabilities of ATM that allow to “hide” from the end-to-end routing schemes the relative mobility of both, (i) satellites with respect to served users on ground, and (ii) satellites with respect to their neighbors in the constellation. The proposed overall networking concept follows a strict decomposition approach, separating the sub-problems of routing in the ISL segment and routing on the up/downlinks, the UDL segment.

As for the UDL segment, it could be shown that the virtual connection tree (VCT) concept known from terrestrial wireless ATM networks can be adapted to the situation in dynamic LEO/MEO networks. A dynamic satellite cluster builds a time-dependent VCT with spotbeams, ensuring that fast and transparent spotbeam and satellite handovers can be performed.

According to the overall networking concept, the ISL segment is fully meshed by VPCs. The basic idea behind this approach is that end-to-end VCCs sharing the same first and last satellites at arbitrary time can be aggregated into one common VPC across the ISL subnetwork. Every transit satellite provides – on the basis of locally available switching tables – pure VP switching functionality between every pair of ISL ports, and thus the whole space segment becomes a pure fast operating cross-connect network. Avoiding any switching on VC level turns out to be especially favorable in the case of many simultaneous low bit rate connections on a trunk line; this situation is very likely to be the dominating one in systems providing primarily voice services.

The fully meshed VPC topology appears static within the ISL segment, but cannot prevent from regular VPC handovers (path switching) due to the relative movement of the whole ISL network with respect to earth, and thus, VCC end points. Alignment of the ATM cell streams during switch-over is essential for non-disruptive operation of the user connection. We have foreseen to

implement an alignment server as proposed for similar situations in terrestrial ATM networks, and have explored the requirements for such a satellite network alignment server in terms of typical path delay offsets. The feasibility of the networking concept with respect to conceptual limitations has been shown, and protection switching has been identified as most appropriate candidate technique to cope with failure situations.

Chapter 4 has examined routing in the UDL and ISL segments of constellation networks. Satellite handover and satellite diversity approaches have been identified as the two important components of UDL routing. The higher the degree of freedom through multiple coverage in the constellation, the better can handover and diversity operation be optimized to achieve minimum rerouting requirements. For satellite handover, a GSM-like backward handover procedure has been found to produce unacceptably long handover breaks of up to 200 ms for typical LEO systems, which can be significantly reduced to about 60 ms (70% reduction) with forward handover, for the sake of a lower call dropping probability. For several investigated satellite handover strategies, we found that a *minimize handover rate* method can slightly reduce the rerouting implications for the ISL segment, compared to the *maximize elevation* reference method. Satellite diversity has been shown to yield significant improvements of service availability in many environments, and a sophisticated channel adaptive diversity scheme was able to achieve nearly the same performance as permanent dual satellite diversity while drastically reducing the additional capacity requirements.

As for the ISL segment, a generic discrete-time network model has been formulated that serves as basis for both advanced routing and dimensioning mechanisms. In ISL subnetworks it is generally required that the routing is dynamic in order to cope with the permanent topological changes, enabling continuous operation in connection-oriented mode. *Dynamic Virtual Topology Routing (DVTR)* has been proposed for use in such environments. The DVTR concept has been validated for the most challenging case of a constellation with switched ISLs (Iridium as example). Of course, topologies employing purely permanent ISLs are covered by the concept just as a “simple” special case. DT-DVTR works completely off-line, i.e. prior to the operational phase of the system. In a first step, a virtual topology is set up for all successive time intervals, providing instantaneous sets of alternative paths between all source-destination node pairs. In the second step, path sequences over a series of time intervals are chosen from that according to certain optimization procedures. Doing so in general for a set of several alternative paths and path sequences, a framework for later on-line traffic adaptive routing strategies can be build, i.e. traffic adaptive routing can be performed within the limits of the given virtual topology. DVTR lends itself quite directly to an implementation exploiting the potential and strengths of ATM virtual connection operation – without being limited to ATM, however. A VPC-based implementation of DVTR in LEO satellite ISL networks has been evaluated by means of software simulation. It could be shown that appropriate optimization strategies are specifically capable of minimizing severe delay jitter which inevitably encounters in path handover situations. The preferred option for routing optimization in the dynamic topology scenario is a sophisticated, service-adaptive sliding window scheme. The higher implementation complexity of this scheme – compared to a simple reference approach – has been shown to be tractable; on the other hand, it clearly outperforms simpler schemes in terms of minimized delay jitter, which in turn reduces complexity of the ATM cell streams alignment.

On-line routing has also been investigated. With connection-oriented communications, the task of on-line routing is to select a route at call setup, which is then only changed in case of a forced path handover. If the traffic load (on links or nodes) is taken into account as one of the decision criteria for route selection, the on-line routing becomes then traffic adaptive. Working within the presented off-line routing framework, the task of on-line routing is reduced to selecting one of

the predefined alternative paths. Therefore it becomes less complex but cannot exploit the full potential of adaptivity. Prior to a proper integration of the two concepts it has therefore been found reasonable to determine the maximum theoretical gain from operating an on-line adaptive routing without any restrictions, i.e., without the limitations of the off-line framework.

Two different adaptive routing implementations have been considered: a decentralized approach based on a conventional algorithm, and an isolated approach using neural networks. Compared to plain shortest path routing, decentralized traffic adaptive routing can significantly reduce peak loads (in the order of 30% for Iridium) and distribute much more evenly the traffic in the network. One of the drawbacks of such a distributed routing scheme is the considerable amount of signaling for exchanging state and routing update information between all satellites in the global network. This has been one of the drivers to consider an isolated approach based on a multilayer perceptron (MLP) type neural network; MLPs are generally well suited for all kinds of pattern recognition problems. The basic idea has thus been to translate the core task of our scenario, namely the route selection or path decision itself, into a (traffic) pattern recognition problem and to let the MLP learn “optimal” solutions which are first calculated with a conventional algorithm to produce training data. As an overall result, we could show that an MLP placed on board of one satellite in the constellation (all others were running conventional algorithms for the sake of computing requirements) is in principle capable of performing traffic adaptive routing, only based on information drawn from its direct environment.

In Chapter 5, a systematic approach to the network design of LEO ISL networks with dynamic topology has been presented. Moderately inclined Walker delta constellations have been identified as appropriate candidates to implement convenient ISL topologies for connection-oriented operation. In contrast to polar star constellations, they don't require on/off-switching of links at highest latitudes, and they don't exhibit a seam of counter-rotating orbits. PAT requirements to follow inter-orbit neighbors were shown to be tractable with both RF and laser technology ISL terminals, so that a completely regular mesh of permanent links can be achieved, with periodic length variation of inter-orbit ISLs being the only remaining dynamic component within the ISL network from a routing point of view. Defining a discrete-time virtual topology on top of the orbiting physical one, the combined routing and dimensioning task can be tackled with similar approaches as known from terrestrial ATM network design. Minimizing the worst-case link capacity is an appropriate target function, which can be formulated as linear minimax optimization problem. Various heuristic and optimization approaches have been numerically compared in a homogeneous traffic scenario. Optimization approaches clearly outperform heuristic methods with respect to the minimum worst-case link capacity target, while the increase of both average load per link and average path delay is limited to acceptable values.

Initially, the whole dimensioning procedure has considered isolated steps. This approach contains one systematic modeling deficiency: by assuming *independent* demand pair capacities in subsequent steps it does effectively neglect the “history” of single calls making up these instantaneous cumulative capacities. Specifically, in this approach it is implicitly assumed that calls are freely reroutable in each step according to the isolated optimization result. Consequently, the isolated step optimization results are usually “too good” as soon as any QoS requirements have to be fulfilled, as this in turn reduces the optimization freedom. Consequently, a history-based multi-step dimensioning approach has been realized and compared to the isolated step approach, numerically assessing the reduction of the optimization potential through the introduction of call and path history. As a major conclusion, the “dimensioning error” in the isolated step approach turns out to be

typically lower than 10%, which makes it an attractive low-complexity method for a first step in the network dimensioning process.

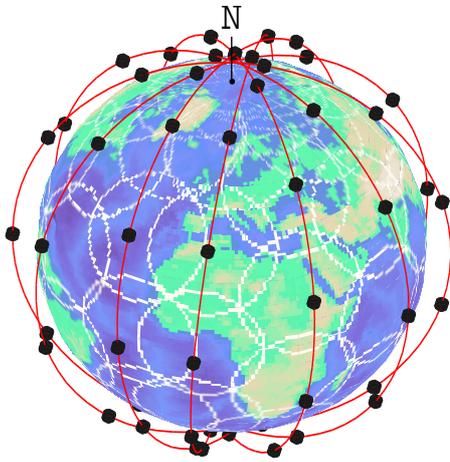
# **Appendix A**

## **Reference Satellite Constellations and ISL Topologies**

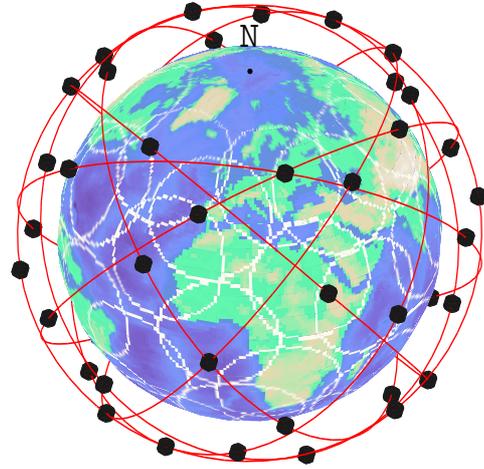
All reference constellations considered in this thesis are displayed in Figs. A.1 – A.6. They are shown (approximately) true to scale with respect to each other.

Table A.1 lists all relevant constellation parameters of these six (seven) constellations at a glance.

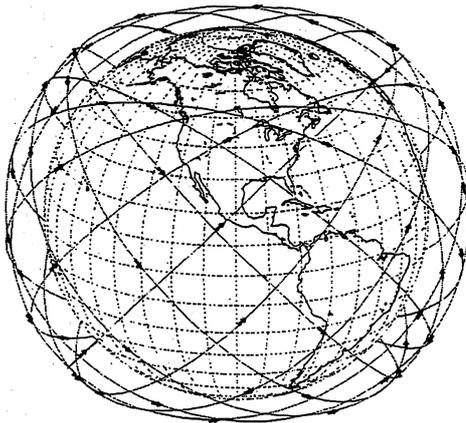
Snapshots of ISL topologies used in this thesis, or in related own research work that is herein referenced but not presented, are displayed in Figs. A.7 – A.11.



**Figure A.1:** Iridium –  $6 \times 11$  LEO star.



**Figure A.2:** Globalstar –  $8 \times 6$  LEO delta.



**Figure A.3:** M-Star –  $12 \times 6$  LEO delta.  
Source: [Mot96]



**Figure A.4:** Celestri –  $7 \times 9$  LEO delta.  
Source: [Mot97]

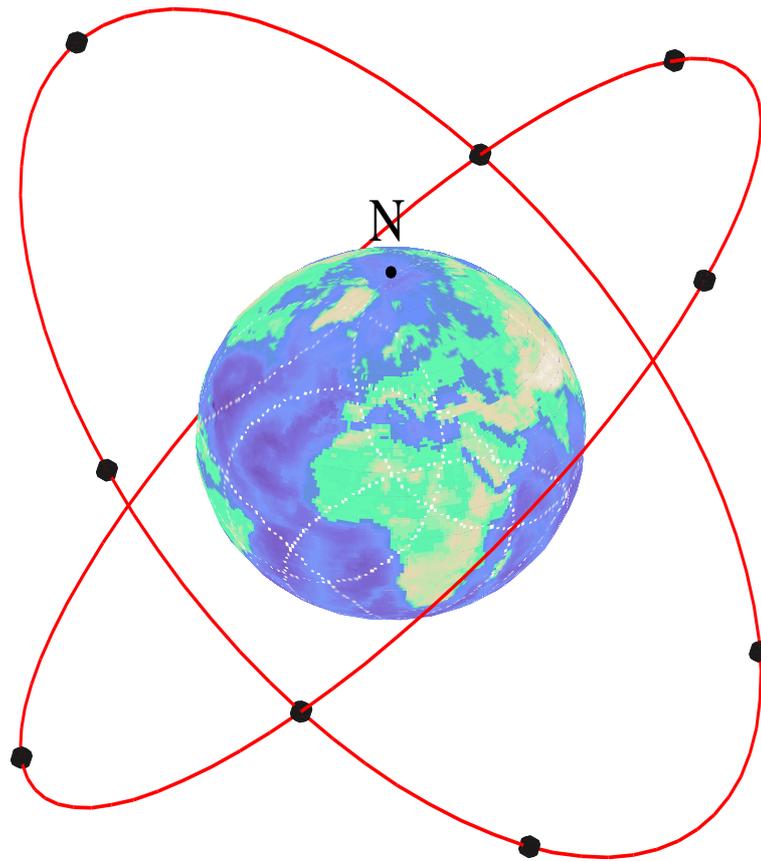


Figure A.5: ICO –  $2 \times 5$  MEO delta.

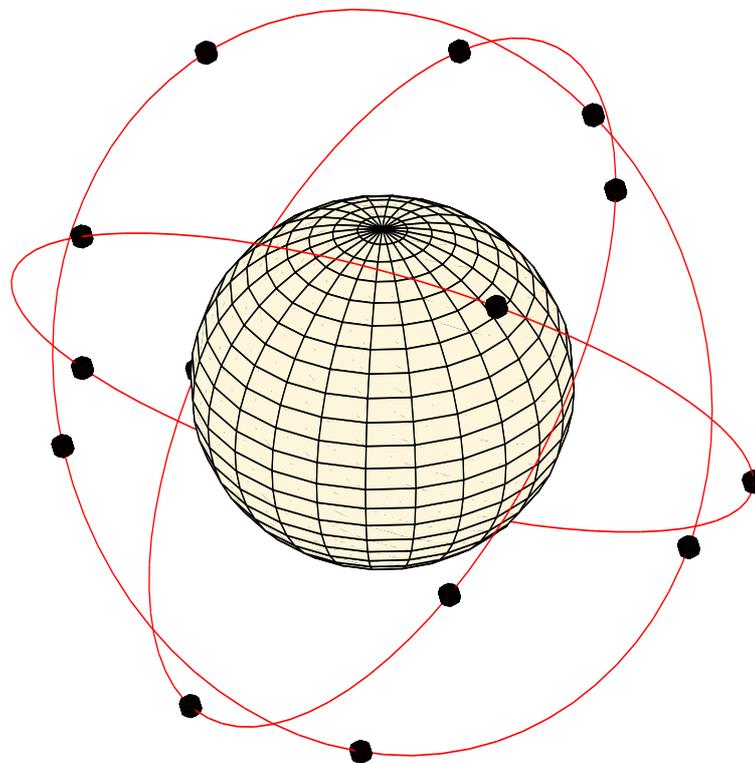
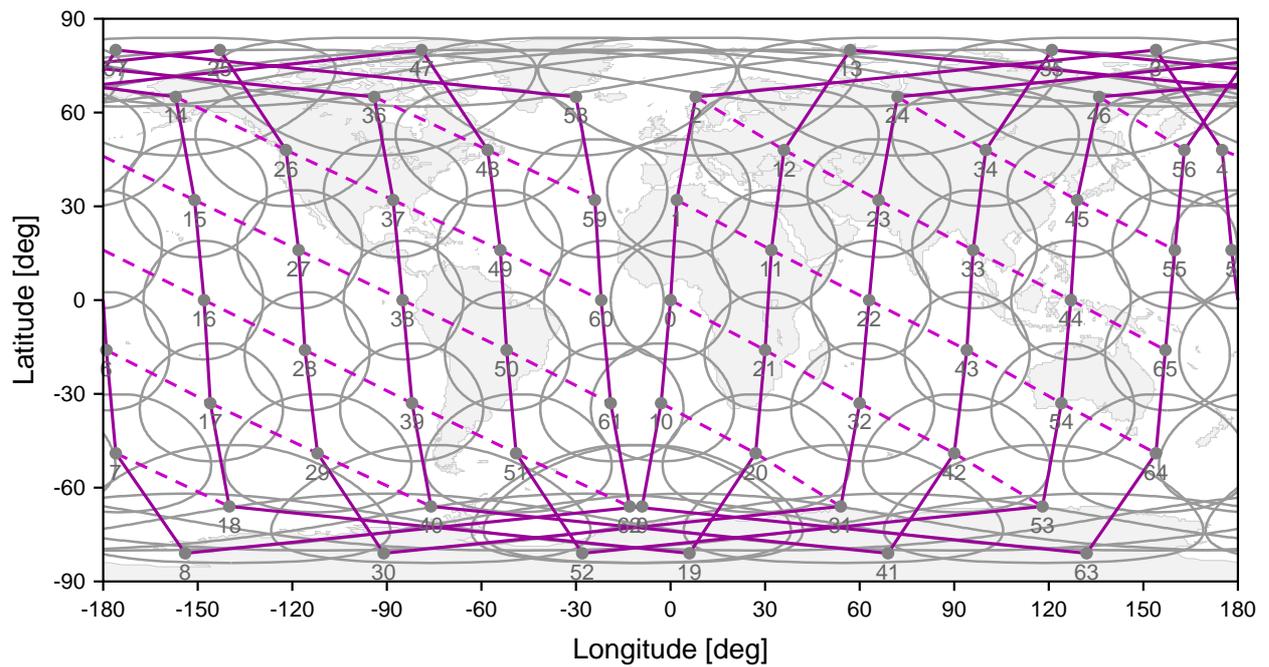


Figure A.6: LEONET –  $3 \times 5$  MEO delta.

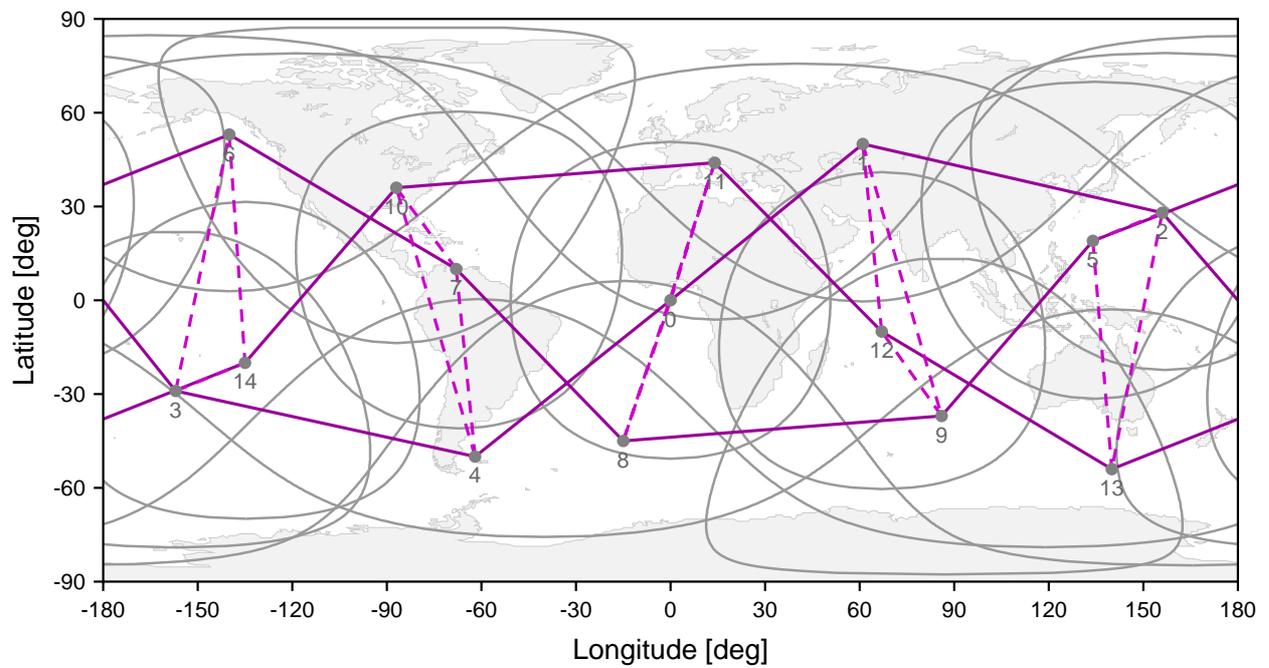
Table A.1: Constellation parameters.

	Iridium	Globalstar	ICO	M-Star	Celestri	LEONET
Constellation pattern	star	delta	delta	delta	delta	delta
Orbit classification	LEO	LEO	MEO	LEO	LEO	MEO
Orbit altitude $h$	780 km	1414 km	10390 km	1350 km	1400 km	6390 km
Orbit period $T$	100 min	113.8 min	5 h 59 min 1 s ("6 h orbit")	112 min 41 s	113.75 min	3 h 59 min 21 s ("4 h orbit")
Orbit inclination $i$	86.4°	52°	45°	47°	48°	54°
No. of satellites in the constellation $N$	66	48	10	72	63	15
No. of orbit planes $P$	6	8	2	12	7	3
No. of satellites per orbit plane $N_S$	11	6	5	6	9	5
Phasing factor $F$	NA	1	0	5	5	1
Plane spacing at equator $\Delta\Omega$ ( $/ 2\Delta$ )	31.6° / 22.0°	45°	180°	30°	51.43°	120°
Plane phasing angle $\Delta\omega_p$	16.36°	7.5°	0°	25°	28.57°	24°
No. of intra-orbit ISLs per satellite	2	none	none	2	2	2
No. of inter-orbit ISLs per satellite	0–2	none	none	2 (T1) / 4 (T2)	2	2
Type of inter-orbit ISLs	switched	NA	NA	permanent	permanent	permanent
No. of intra-orbit ISLs in the constellation	66	none	none	72	63	15
No. of inter-orbit ISLs in the constellation	40	none	none	72 / 144	63	15
No. of ISLs in the constellation	106	none	none	144 / 216	126	30
Minimum elevation angle $\varepsilon_{\min}$	8.2°	10°	10°	22°	16°	20°
$\varepsilon_{\min}$ assumed for certain simulations	8.2°	20°	20°	NA	NA	20°

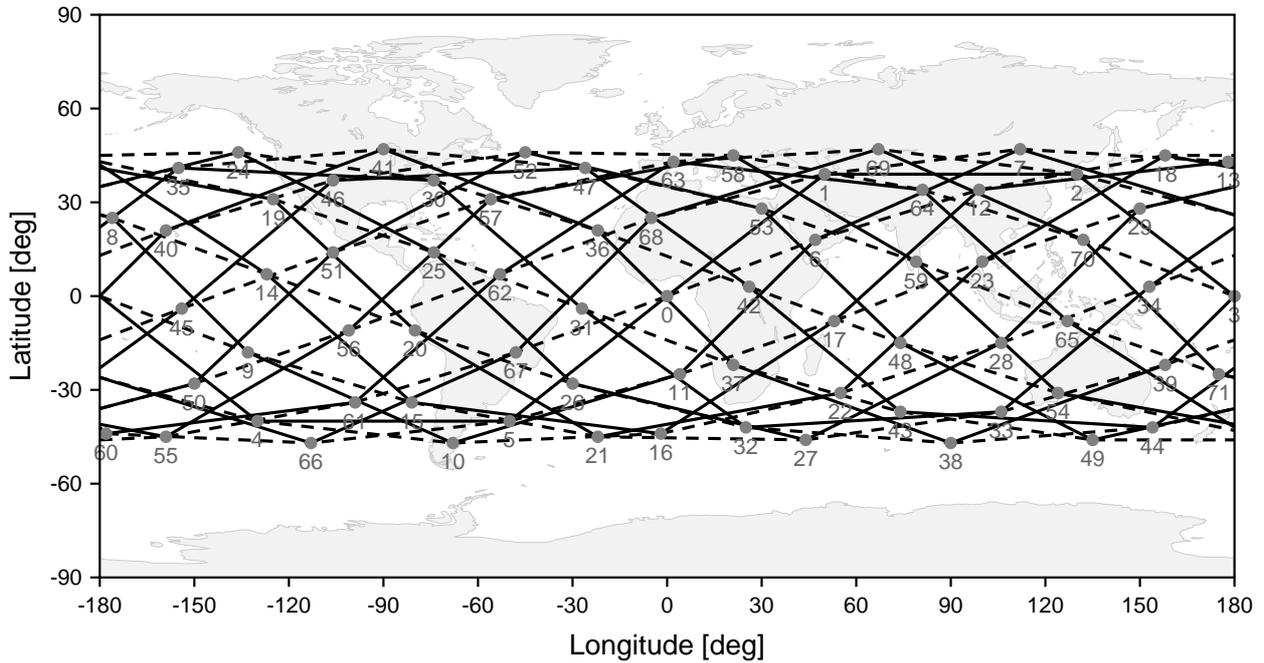
NA = not applicable



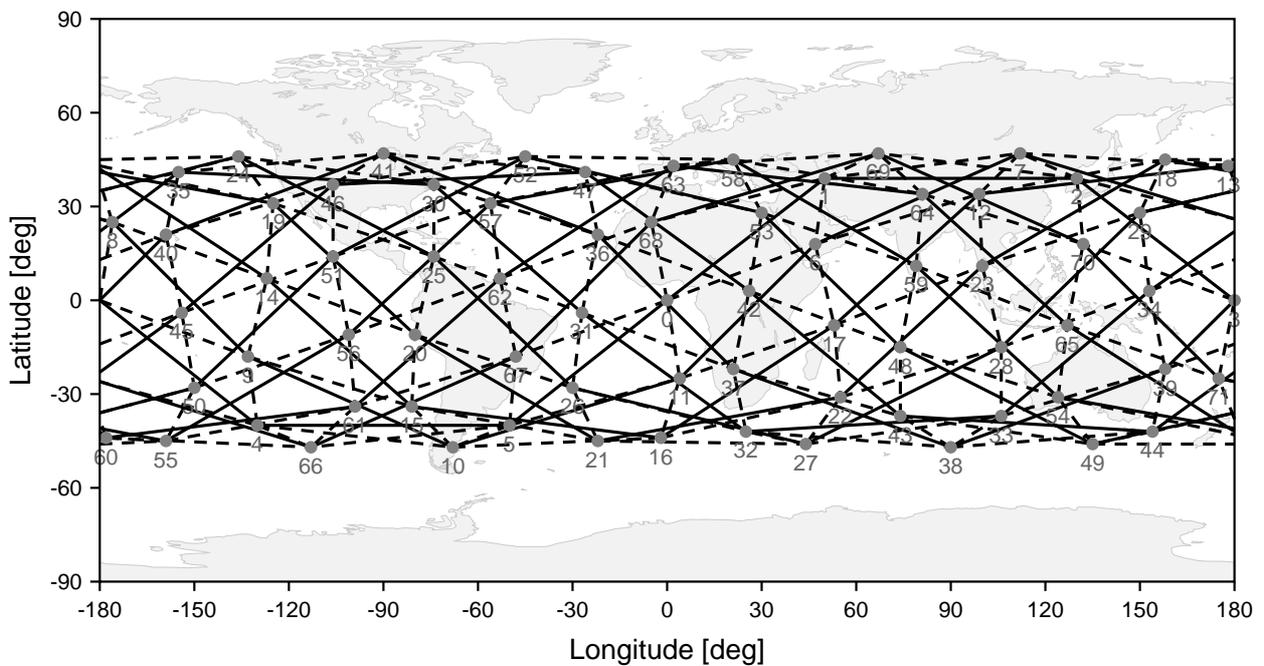
**Figure A.7:** Iridium ISL topology and constellation coverage snapshot: solid lines, intra-orbit ISLs; dashed lines, inter-orbit ISLs.



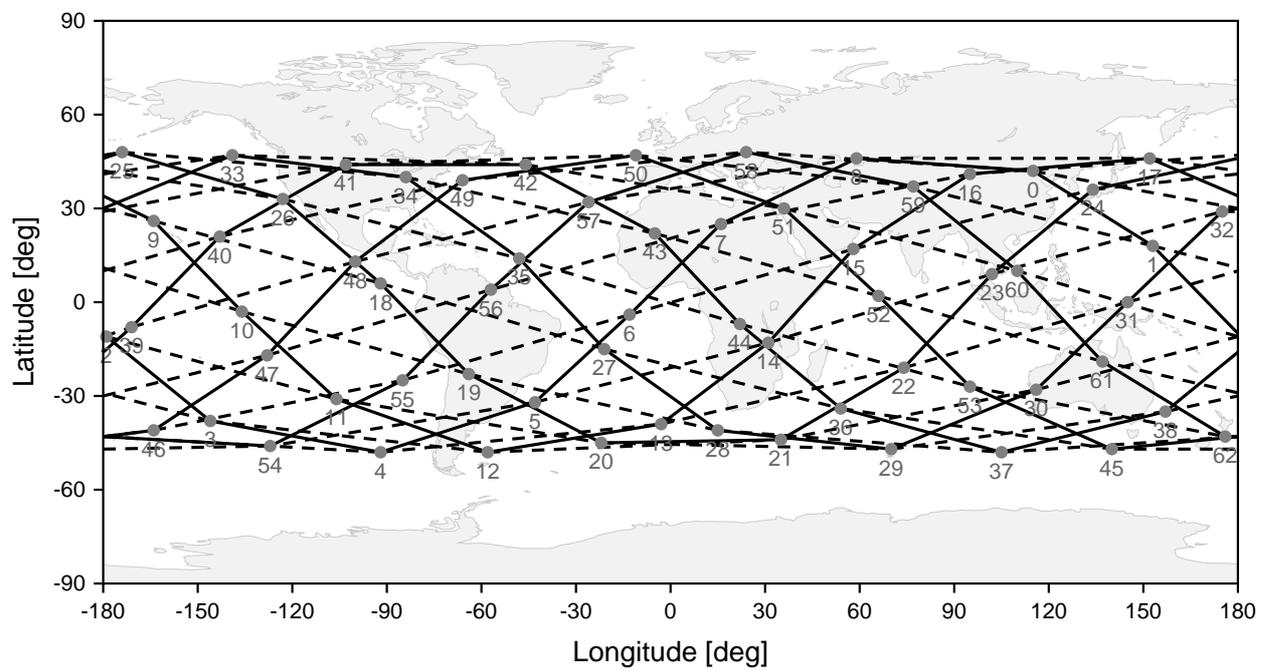
**Figure A.8:** LEONET ISL topology and constellation coverage snapshot: solid lines, intra-orbit ISLs; dashed lines, inter-orbit ISLs.



**Figure A.9:** M-Star ISL topology T1 and constellation coverage snapshot: solid lines, intra-orbit ISLs; dashed lines, inter-orbit ISLs.



**Figure A.10:** M-Star ISL topology T2 and constellation coverage snapshot: solid lines, intra-orbit ISLs; dashed lines, inter-orbit ISLs.



**Figure A.11:** Celestri ISL topology and constellation coverage snapshot: solid lines, intra-orbit ISLs; dashed lines, inter-orbit ISLs.

# Appendix B

## Space Geometry and Orbital Mechanics

*There are an infinite number of possible stable orbits, circular and elliptical, in which a rocket would remain if the initial conditions were correct.*

— ARTHUR C. CLARKE, in *Extra-Terrestrial Relays* (1945)

### B.1 Satellite Orbits

This section considers the general case of elliptical orbits, with circular orbits as a special case, and presents the derivation of satellite velocity and orbit period, introduced in Section 2.2.1.

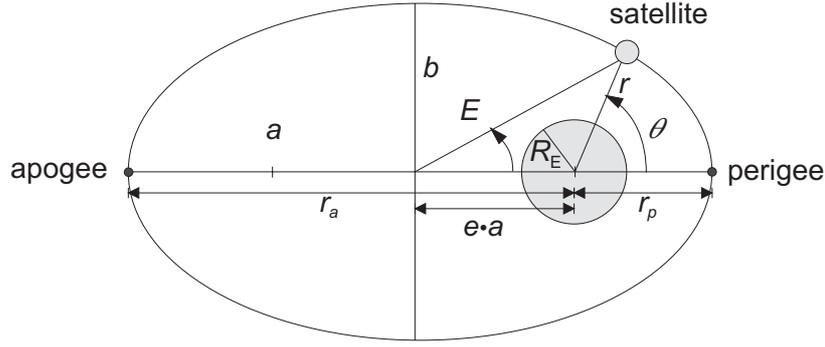
In the early 17th century Johannes Kepler discovered some important properties of planetary motion that are known since then as Kepler's laws:

- *First law (1602)*: the planets move in a plane; the orbits around the sun are ellipses with the sun at one focal point.
- *Second law (1605)*: the line between the sun and a planet sweeps out equal areas in equal intervals of time.
- *Third law (1618)*: the ratio between the square of the orbit period  $T$  and the cube of the semi-major axis  $a$  of the orbit ellipse,  $T^2/a^3$ , is the same for all planets.

These laws can be applied to any two-body system subject to gravitation, and thus also describe the motion of a satellite around the earth. Extensive treatments of orbit mechanics can be found in the textbooks [Dav85] and [MB98].

#### B.1.1 Elliptical and Circular Orbits

Figure B.1 shows the geometry of an elliptical satellite orbit according to Kepler's first law. The satellite orbit has an elliptical shape with the earth at one focal point. The ellipse is defined by



**Figure B.1:** Parameters of elliptical orbits.

two parameters: the semi-major and semi-minor axes  $a$  and  $b$ . The shape of the ellipse can also be described by the numerical eccentricity

$$e = \sqrt{1 - \frac{b^2}{a^2}} \quad \text{with} \quad 0 \leq e < 1. \quad (\text{B.1})$$

With this parameter the distance of the focal points from the ellipse center can be expressed as  $e \cdot a$ . The distance of the satellite from the earth's center is the radius  $r$ . The point of the orbit where  $r$  is smallest is called perigee with  $r = r_p$ . The point with largest  $r$  is denoted apogee with  $r = r_a$ . From Kepler's second law we can deduce that a satellite moves quickly near perigee and slowly near apogee. According to Fig. B.1 and using Eq. (B.1) we can set up the following relations:

$$\begin{aligned} a &= \frac{r_a + r_p}{2} \\ e &= \frac{r_a - r_p}{r_a + r_p} \\ r_a &= a(1 + e) \\ r_p &= a(1 - e). \end{aligned} \quad (\text{B.2})$$

The angle  $\theta$  between the perigee and the satellite as seen from the earth's center is commonly called the *true anomaly*. It can be used to determine the satellite radius  $r$  along the elliptical orbit:

$$r = \frac{a(1 - e^2)}{1 + e \cos \theta}. \quad (\text{B.3})$$

The angle between perigee and the satellite with respect to the ellipse center is denoted the *eccentric anomaly*  $E$ , which is related to  $\theta$  through

$$\cos \theta = \frac{a}{r} (\cos E - e) = \frac{\cos E - e}{1 - e \cos E}. \quad (\text{B.4})$$

The time  $t$  after perigee passing  $t_p$  can be related to the eccentric anomaly  $E$  through

$$\frac{2\pi}{T}(t - t_p) = E - e \sin E, \quad (\text{B.5})$$

where  $T$  is the orbit period of the satellite and the term  $2\pi(t - t_p)/T$  is called the *mean anomaly*. Using Eq. (B.5) and Eq. (B.4) the time can be derived as a function  $t(\theta)$ . However, since the inverse function of Eq. (B.5) cannot be solved, the time behavior of  $\theta(t)$  must be determined numerically.

The satellite altitude  $h$  above the earth's surface is

$$h = r - R_E \quad (\text{B.6})$$

with  $R_E$  being the radius of the earth. Accordingly, the orbit altitude at apogee is  $h_a = r_a - R_E$  and the altitude at perigee is  $h_p = r_p - R_E$ . Actually the earth is not an ideal sphere but exhibits some flattening at the poles. In the following, we will use  $R_E = 6\,378$  km representing the mean equatorial radius<sup>1</sup>.

**Circular Satellite Orbits** A circular satellite orbit is a special case of an elliptical orbit with zero eccentricity,  $e = 0$ . Thus,  $a = b = r = r_a = r_p$ . The earth is at the center of the circular orbit, and the satellite altitude  $h = r - R_E$  is constant. Furthermore, it follows for the time behavior of the true anomaly that

$$\theta(t) = \frac{2\pi t}{T}. \quad (\text{B.7})$$

## B.1.2 Satellite Velocity and Orbit Period

Isaac Newton extended the work of Kepler and in the year 1667 discovered the law of gravity. This law states that two bodies with masses  $m$  and  $M$  at a distance  $r$  attract each other with the gravitational force

$$F_g = G \frac{mM}{r^2}. \quad (\text{B.8})$$

Here,  $G = 6.6732 \cdot 10^{-11} \text{N m}^2/\text{kg}^2$  is the universal gravitation constant.

For a satellite orbiting around the earth the mass  $m$  represents the satellite mass and  $M = M_E = 5.9733 \cdot 10^{24}$  kg is the mass of earth. The total mechanical energy consisting of the potential energy and the kinetic energy is constant:

$$\frac{mv^2}{2} - \frac{\mu m}{r} = -\frac{\mu m}{2a}, \quad (\text{B.9})$$

where  $\mu = GM_E = 3.986 \cdot 10^{14} \text{m}^3/\text{s}^2$ . Thus, the velocity  $v$  of a satellite in an elliptic orbit may be obtained by

$$v = \sqrt{\mu \left( \frac{2}{r} - \frac{1}{a} \right)} \quad (\text{B.10})$$

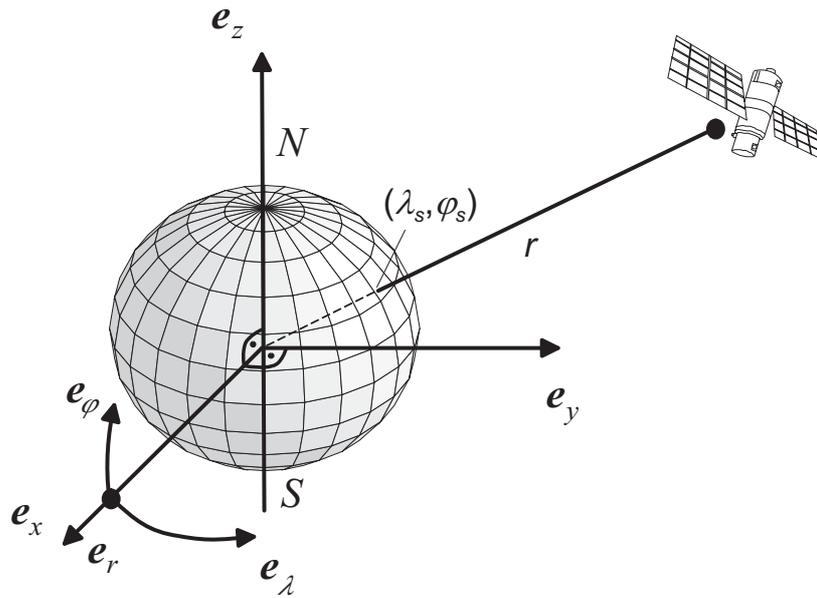
which can be simplified for circular orbits ( $r \equiv a$ ) to

$$v = \sqrt{\frac{\mu}{r}}. \quad (\text{B.11})$$

Equation (B.11) states that the velocity of satellites in circular orbits is constant, coinciding with Kepler's second law. The orbit period can now be derived as

$$T = \frac{2\pi r}{v} = 2\pi \sqrt{\frac{r^3}{\mu}} \quad (\text{B.12})$$

<sup>1</sup>The polar earth radius amounts to 6 357 km whereas the mean radius averaged over the earth's surface is 6 371 km.



**Figure B.2:** Earth coordinate systems.

which for elliptical orbits generalizes to

$$T = 2\pi \sqrt{\frac{a^3}{\mu}} \quad (\text{B.13})$$

according to Kepler's third law.

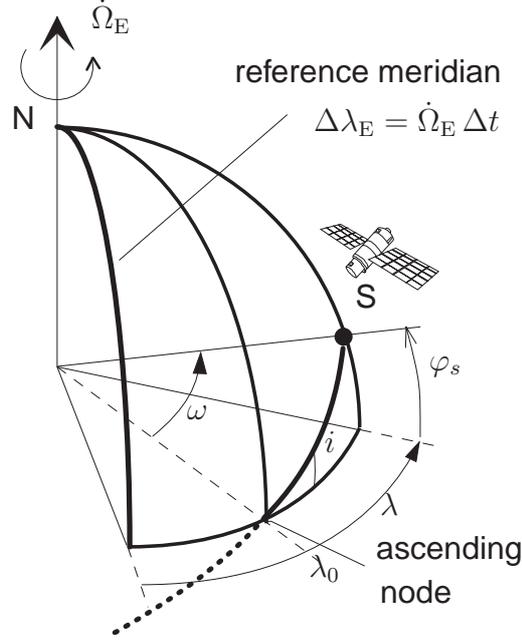
The orbit mechanics discussed so far are idealized in the sense that they assume a spherical and homogeneous earth, empty space, and the absence of any gravitational forces from sources other than the satellite and the earth. For this ideal scenario the satellite orbit will remain constant for all times.

## B.2 Coordinate Systems and Satellite Ground Tracks

In this section the time-variant coordinates of a satellite position with respect to earth are derived. To this end, it is useful to first introduce coordinate systems that allow us to relate the satellite's position to spherical or Cartesian geocentric coordinates, cf. Fig. B.2. In an inertial spherical coordinate system with earth-centered origin any point on the earth's surface is specified by two angular coordinates, the latitude  $\varphi$  and the longitude  $\lambda$ , and a distance  $r$  from the origin.<sup>2</sup> The longitude  $\lambda$  ( $-180^\circ \leq \lambda < 180^\circ$ ) is counted positively from the prime "Greenwich" meridian to the east. The latitude  $\varphi$  ( $-90^\circ \leq \varphi \leq 90^\circ$ ) determines the position on great circles with constant longitude that run through the north and south poles. The latitude is counted positively from the equator ( $\varphi = 0^\circ$ ) to the north.

Alternatively, an inertial system can be described by Cartesian coordinates. Let  $\vec{e}_x$ ,  $\vec{e}_y$ , and  $\vec{e}_z$  be the orthogonal normalized base vectors of the Cartesian coordinate system with earth-centered

<sup>2</sup>For simplicity, the distance  $r$  on the earth's surface can be assumed constant, equal to the mean equatorial earth radius  $R_E = 6378$  km.



**Figure B.3:** Ground track of the satellite on the rotating earth.

origin. The  $z$ -axis points to the geographic north pole and the  $x$ -axis to the prime meridian. The coordinate systems are inertial systems with respect to the earth's rotation, i.e. a fixed point  $P$  on earth does not change its geocentric coordinates  $(x_p, y_p, z_p)$  or  $(\lambda, \varphi, r)$  during an earth revolution.

The Cartesian and spherical coordinate systems can be transformed into each other by

$$\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} r \cos \varphi \cos \lambda \\ r \cos \varphi \sin \lambda \\ r \sin \varphi \end{pmatrix} \quad \text{or} \quad \begin{pmatrix} \lambda \\ \varphi \\ r \end{pmatrix} = \begin{pmatrix} \arctan4 y/x \\ \arctan4 z/\sqrt{x^2 + y^2} \\ \sqrt{x^2 + y^2 + z^2} \end{pmatrix}, \quad (\text{B.14})$$

with the definition of the arc tangent for four quadrants

$$\arctan4 y/x \doteq \begin{cases} \arctan y/x & \text{for } x \geq 0 \\ \arctan y/x + \pi & \text{for } x < 0, y \geq 0 \\ \arctan y/x - \pi & \text{for } x < 0, y < 0. \end{cases} \quad (\text{B.15})$$

The goal is now to calculate the satellite track  $(\lambda_s(t), \varphi_s(t))$ , i.e. the coordinates of the satellite with respect to the rotating earth. The procedure is depicted in Fig. B.3. With spherical trigonometry one can express the satellite latitude  $\varphi_s(t)$  at time  $t$  as a function of the elongation  $\omega(t)$  denoting the angle from the ascending node to the satellite,

$$\sin \varphi_s(t) = \sin i \sin \omega(t). \quad (\text{B.16})$$

The elongation  $\omega$  is related to the true anomaly  $\theta$  by

$$\theta(t) = \omega(t) - \omega_0, \quad (\text{B.17})$$

(cf. Section B.1). Furthermore, by spherical trigonometry one can find

$$\cos \varphi_s(t) \cos(\lambda(t) - \lambda_0) = \cos \omega(t). \quad (\text{B.18})$$

Here,  $\lambda(t)$  is the longitude of the satellite with respect to the non-rotating earth and  $\lambda_0$  denotes the longitude of the ascending node. Let  $t_0$  denote the time when the satellite passes the ascending node. To include the rotation of the earth with its angular frequency  $\dot{\Omega}_E$  we have to consider that at time  $t$  the reference meridian has changed its sidereal orientation by  $\Delta\lambda_E = \dot{\Omega}_E(t - t_0)$ . Therefore, we have for the geographic satellite longitude  $\lambda_s$

$$\begin{aligned}\lambda_s(t) &= \lambda(t) - \dot{\Omega}_E(t - t_0) \\ &= \lambda_0 + \arccos \frac{\cos \omega(t)}{\cos \varphi_s(t)} - \frac{2\pi}{T_E}(t - t_0) .\end{aligned}\tag{B.19}$$

# Appendix C

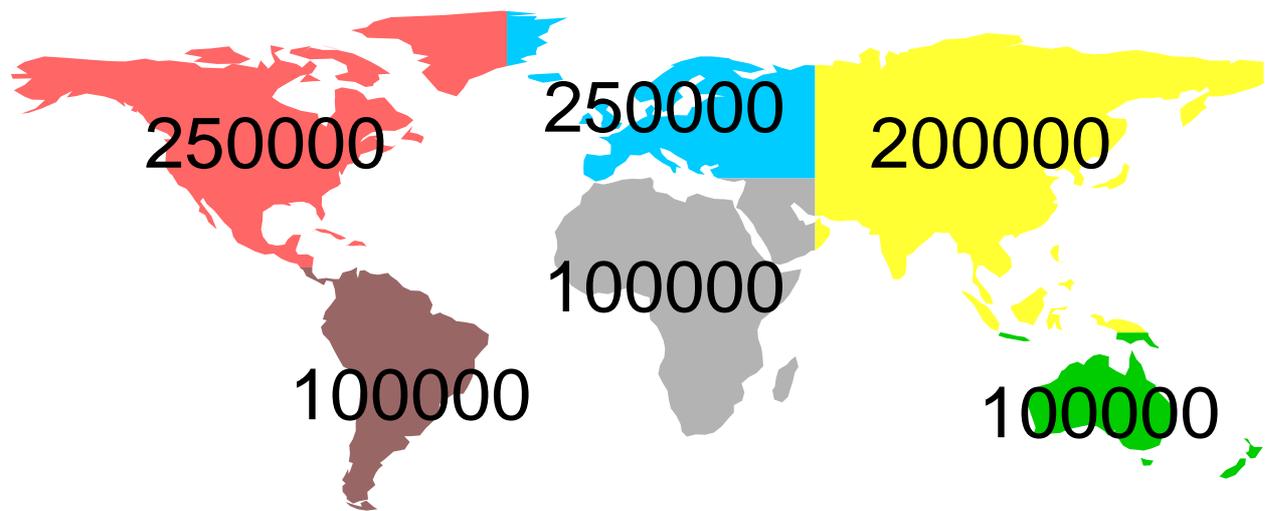
## Global Traffic Models

This appendix presents the rationale and numerical data of the two global *traffic modeling frameworks*, referred to as TMF 1 and TMF 2, which are used for numerical investigations in different areas covered by this thesis. Each traffic modeling framework consists of some model components, where some of them may be optionally used or “bypassed”. Roughly speaking, TMF 1 is the somewhat simpler model, mainly (i) in that it assumes a coarser geographic resolution (grid) for the worldwide user distribution, and (ii) providing the offered traffic input data (mainly for network routing) more directly after some off-line pre-processing, resulting in considerably lower on-line processing requirements (on call level) during the routing simulations. It should be stressed, however, that the reason for using different models (and different combinations of modeling components) is *not* due to the fact that the underlying data would be closer to reality for one or the other scenario; rather the different data sets are mainly useful for qualitative distinction of different numerical investigations, stressing certain phenomena (like daily load variation or inhomogeneity of global traffic flows, for instance) in a particular way, and for sensitivity analysis with respect to such phenomena.

### C.1 Traffic Modeling Framework TMF 1

According to [WJLB95, BJL<sup>+</sup>93], one million worldwide subscribers are assumed to be distributed on six continental regions as shown in Fig. C.1. Within each of the regions, geographic uniform distribution is assumed. This model is referred to as *6-regions model*.

Assuming a single-service call scenario (e.g., telephony) with a constant activity of 5 mErl for all one million subscribers, we have a total offered network traffic of  $A_{\text{net}} = 5000$  Erl. This target value can now be broken down into appropriate splits of traffic flows between any two of the six regions (including pairs of source/destination region identity, accounting for the important share of “local” traffic) to get the regional traffic matrix  $A_{\text{reg}}$ . Table C.1 lists the corresponding regional traffic flows in Erlang, according to [WJLB95, BJL<sup>+</sup>93]. The figures have been derived by extrapolating telephone traffic measurements in the global PSTN and making some additional assumptions; for more details on this issue one should refer to the mentioned original publications. The upper and lower triangle matrices are symmetric as we consider bidirectional symmetric services (e.g., telephony) in this modeling framework.



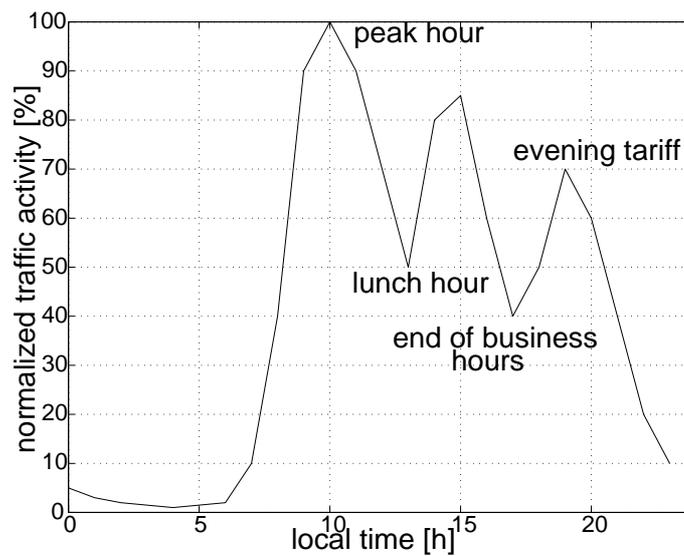
**Figure C.1:** Global user distribution over six continental regions.

**Table C.1:** Global peak traffic flows in Erlang.

	North America	Europe	Asia	South America	Africa	Australia/Oceania
North America	1030	48	48	36	24	24
Europe	48	1030	48	36	36	12
Asia	48	48	825	10	20	39
South America	36	36	10	436	10	5
Africa	24	36	20	10	436	5
Australia/Oceania	24	12	39	5	5	441

**Table C.2:** Offset of local time from UTC in hours, assumed for the six continental regions.

Region	Local time = UTC $\oplus$
North America	-7 h
Europe	+1 h
Asia	+7 h
South America	-3 h
Africa	+1 h
Australia/Oceania	+9 h



**Figure C.2:** Normalized daily activity assumed within TMF 1.

So far – by assuming a *constant and worldwide unique* activity of 5 mErl for all users – the usual daily activity variations have not been considered in the calculation of traffic flow values. Thus, the figures in Table C.1 can be regarded as *peak* values; however, they may be sufficient for several numerical studies aiming at qualitative insight into certain effects, as for instance the impact of *instantaneous* unbalanced traffic flows being injected into the ISL network.

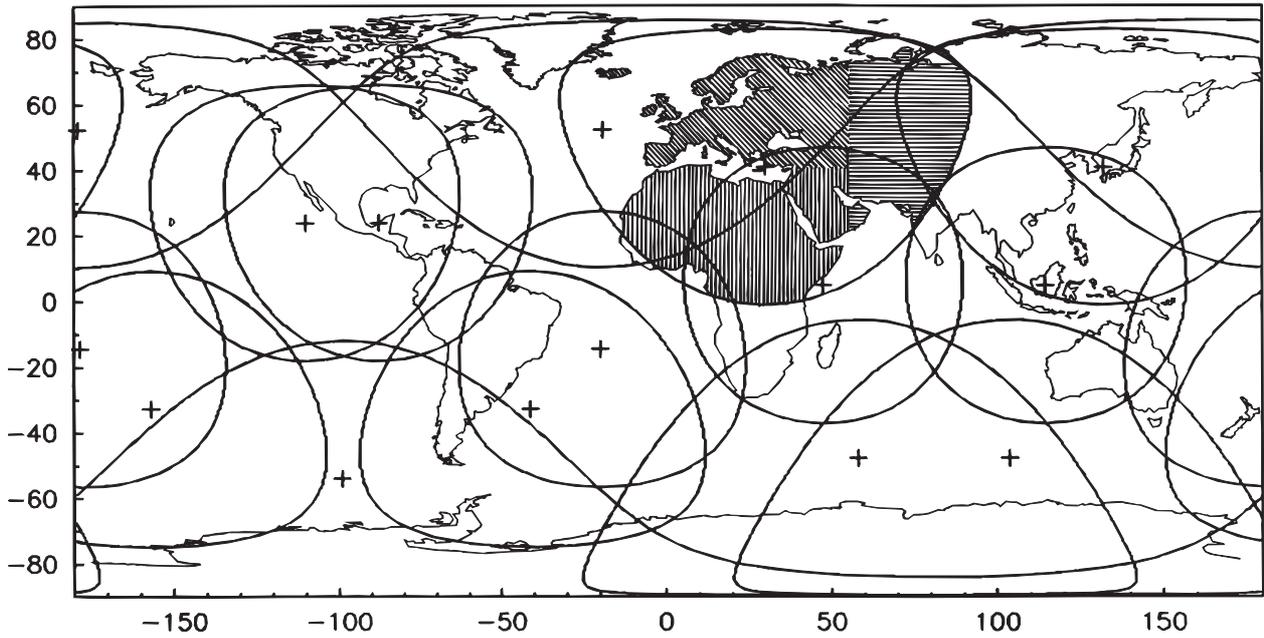
For those numerical studies where spatial and time variation of offered traffic are an issue, the modeling framework TMF 1 provides the following extension.

Using a somewhat “average” offset of the local time in the six regions from universal time UTC according to Table C.2 in combination with the normalized daily activity curve from Fig. C.2, several effective regional traffic flow matrices over a 24-hour period can be derived.<sup>1</sup> According to our particular needs in constellation network simulations, and considering the general coarseness bound up with the 6-regions model, a discretization into 4-hour intervals was found to be sufficient; this yields the six regional traffic flow matrices listed in Table C.3, each one defining a constant regional traffic situation for the respective four hours. After 24 hours, the patterns repeat as we do not assume any longer-term (e.g., weekly) activity variations.

<sup>1</sup>Note that for a flow between regions  $A$  and  $B$ ,  $A \rightarrow B$ , the local time information and the traffic scaling according to the corresponding normalized activity may be applied to source *and* destination, and several ways of “combining” both are on principle possible. However, in order to retain the symmetry of the triangle matrices, we have applied the scaling only based on the local time of the source, once for  $A \rightarrow B$  and once for  $B \rightarrow A$ , and then taken the arithmetic average of both to achieve identical flow values for  $A \leftrightarrow B$ .

**Table C.3:** Global traffic flows in Erlang with respect to UTC.

		N. America	Europe	Asia	S. America	Africa	Austr./Oc.
UTC=1.00	Local time	18.00	2.00	8.00	22.00	2.00	10.00
N. America	18.00	515	13	22	13	7	20
Europe	2.00	13	21	10	4	1	6
Asia	8.00	22	10	330	3	4	27
S. America	22.00	13	4	3	87	1	3
Africa	2.00	7	1	4	1	9	3
Austr./Oc.	10.00	20	6	27	3	3	441
UTC=5.00	Local time	22.00	6.00	12.00	2.00	6.00	14.00
N. America	22.00	206	6	22	4	3	12
Europe	6.00	6	21	18	1	1	5
Asia	12.00	22	18	577	4	7	29
S. America	2.00	4	1	4	24	0	2
Africa	6.00	3	1	7	0	9	2
Austr./Oc.	14.00	12	5	29	2	2	353
UTC=9.00	Local time	2.00	10.00	16.00	6.00	10.00	18.00
N. America	2.00	21	25	15	1	13	7
Europe	10.00	25	1030	39	19	36	9
Asia	16.00	15	39	495	3	16	22
S. America	6.00	1	19	3	9	5	1
Africa	10.00	13	36	16	5	436	4
Austr./Oc.	18.00	7	9	22	1	4	221
UTC=13.00	Local time	6.00	14.00	20.00	10.00	14.00	22.00
N. America	6.00	21	20	15	19	10	3
Europe	14.00	20	825	34	33	29	6
Asia	20.00	15	34	493	8	14	16
S. America	10.00	19	33	8	436	9	3
Africa	14.00	10	29	14	9	349	3
Austr./Oc.	22.00	3	6	16	3	3	88
UTC=17.00	Local time	10.00	18.00	24.00	14.00	18.00	2.00
N. America	10.00	1030	37	18	33	18	13
Europe	18.00	37	515	18	24	8	2
Asia	24.00	18	18	218	7	6	1
S. America	14.00	33	24	7	349	4	2
Africa	18.00	18	8	6	4	41	2
Austr./Oc.	2.00	13	2	1	2	2	9
UTC=21.00	Local time	14.00	22.00	4.00	18.00	22.00	6.00
N. America	14.00	825	21	15	24	10	13
Europe	22.00	21	206	7	13	5	2
Asia	4.00	15	7	87	4	3	1
S. America	18.00	24	13	4	218	3	1
Africa	22.00	10	5	3	3	8	1
Austr./Oc.	6.00	13	2	1	1	1	9



**Figure C.3:** Regional coverage shares and traffic mapping illustrated for a LEONET satellite.

Although not really being a traffic *modeling* component, it is in direct extension of TFM 1 useful to consider the mapping of regional traffic flows onto traffic flows between pairs of satellites in the orbiting constellation, which brings in its own dynamic component. At any considered time step, source and destination traffic from/to ground is allocated to the different satellites according to their instantaneous coverage share of any of the six land mass regions, cf. the example for a LEONET satellite in Fig. C.3. Of course, the effect of multiple coverage must be taken into account in this process so as to guarantee that the sum of traffic values on all satellites with a coverage share equals the respective regional source/destination traffic.

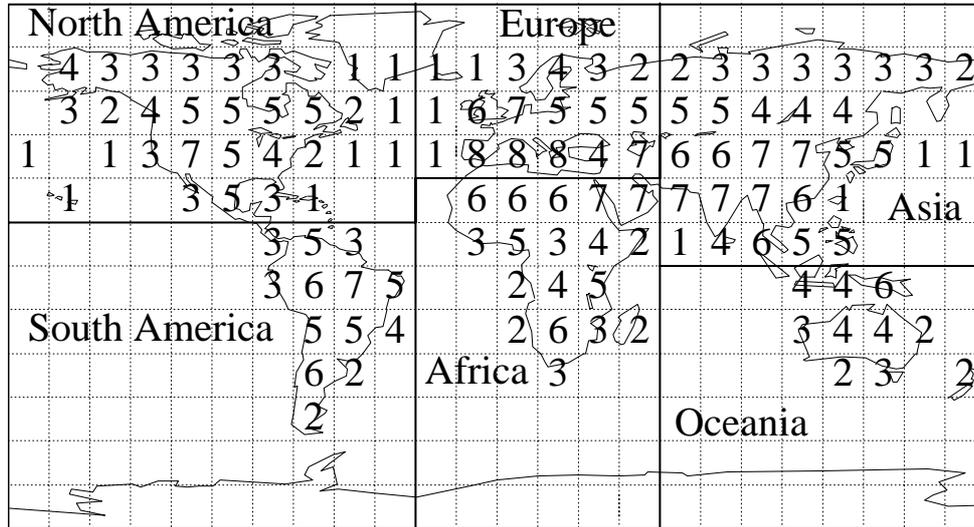
From an offered traffic viewpoint, this mapping effectively translates the regional traffic matrix  $\mathbf{A}_{\text{reg}}$  into a network matrix  $\mathbf{A}_{\text{net}} = [A_{\text{OD},ij}]$  of offered traffics  $A_{\text{OD}}$  between all pairs  $(i, j)$  of satellites, where  $i, j = 1, \dots, N$ , such that

$$\sum_i \sum_j A_{\text{OD},ij} = A_{\text{net}} . \quad (\text{C.1})$$

It should be explicitly noted that  $i = j$  is of course possible, and it covers the cases where uplink traffic is directly routed to the downlink of the same satellite, without using any ISL. In other words, the total traffic offered to the *ISL network* is in general less than  $A_{\text{net}}$ . It is important to keep this in mind whenever one looks at numerical results for ISL traffic routing or capacity dimensioning.

**Table C.4:** Traffic intensity levels and corresponding expected market (2005) in millions of addressable minutes/year [Vio95].

Intensity level	1	2	3	4	5	6	7	8
Traffic (million min/year)	1.6	6.4	16	32	95	191	239	318



**Figure C.4:** Earth zone division and regional traffic intensity levels (2005 forecast) [Vio95].<sup>2</sup>

## C.2 Traffic Modeling Framework TMF 2

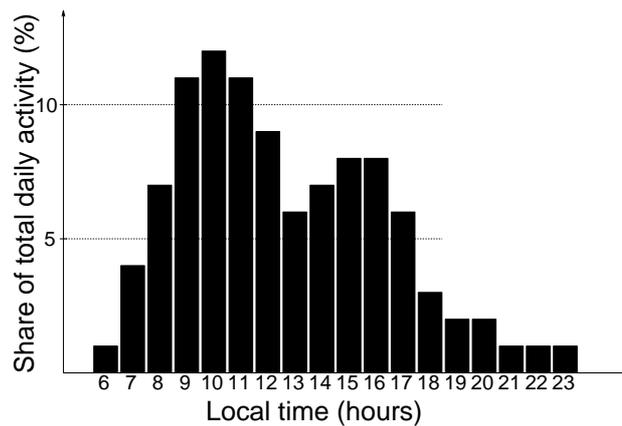
In order to fully understand the rationale of TMF 2 one should have read the previous section on TMF 1 before. Presenting TMF 2 we will mainly focus on the conceptual differences with respect to TMF 1 and provide the specific numerical data.

A single-service scenario as in TMF 1 is assumed. As a major difference to the former, a certain (peak) activity in Erlang is not explicitly set here as free parameter since it is implicitly contained in the source traffic demand figures as presented in the following *addressable market model*.

In general, source traffic demand can be expected to be largely dependent on geographical subscriber distribution and other market factors like GDP, service penetration and acceptance, etc. Using market prediction methods based on such factors, one of the most comprehensive studies that fits our needs has been published as MSc thesis at the Massachusetts Institute of Technology (MIT) [Vio95]. In essence, it provides global source traffic maps – with a resolution of 15° by 15° earth zones – forecasting the expected total MSS market in millions of addressable minutes per year beyond 2000. Figure C.4 displays the earth zones and their traffic intensity levels, the latter encoding the traffic quantities according to Table C.4. The figures show the forecast for the year 2005. This traffic map has been adopted in our simulations.

As discussed in the previous section, traffic in a global satellite communication system reveals a pronounced multihour characteristic. Consequently, the effort spent into these fine-resolution prediction figures (of peak traffic, as defined under TFM 1!), should not be made useless by remaining

<sup>2</sup>The subdivision into the six continental regions is not contained in the original work, but added for our extension of the model.



**Figure C.5:** Daily activity histogram assumed within TMF 2.

**Table C.5:** Regional traffic flow shares in %:  $x\%$  of the traffic from a particular source region flows into the respective destination region.

Source	Destination						$\Sigma$
	North America	Europe	Asia	South America	Africa	Australia/Oceania	
North America	85	4	4	3	2	2	100
Europe	4	85	4	3	3	1	100
Asia	5	5	83	1	2	4	100
South America	7	7	2	81	2	1	100
Africa	5	7	4	2	81	1	100
Australia/Oceania	5	2	7	1	1	84	100

implicitly with a globally unique busy hour assumption which introduces unacceptable inaccuracies in the basic source traffic assumptions. Therefore, daily evolution of traffic activity per user is definitely included under TFM 2. The daily activity is assumed to be the same for all users worldwide, and its discretized and normalized values are shown in Fig. C.5. Further, the local time of each traffic zone is assumed to be equal to the solar time of the respective zone's center longitude, and for a given local time the applicable activity *factor* is determined from the figure. All zones' peak traffic values are then weighted with their corresponding factor and scaled down properly so as to arrive at the effective demand (in number of call arrivals) for a given interval. The latter is equal to the interval used for time-discrete modeling of the constellation.

The last but one major step takes care of a proper assessment of the destination zones for generated calls. For this purpose, the earth is subdivided into six continental regions as displayed in Fig. C.4, each one containing a unique subset of the  $15^\circ$  by  $15^\circ$  zones. The traffic flow between these regions is assumed according to Table C.5, which is essentially a modified outgrowth of the global traffic flow matrix used under TFM 1 (cf. Table C.1). The figures reflect that most of the traffic is local per region. As far as the distribution of drain traffic *within* a given continental region is concerned, we divide it as a function of the instantaneous zone activity share.

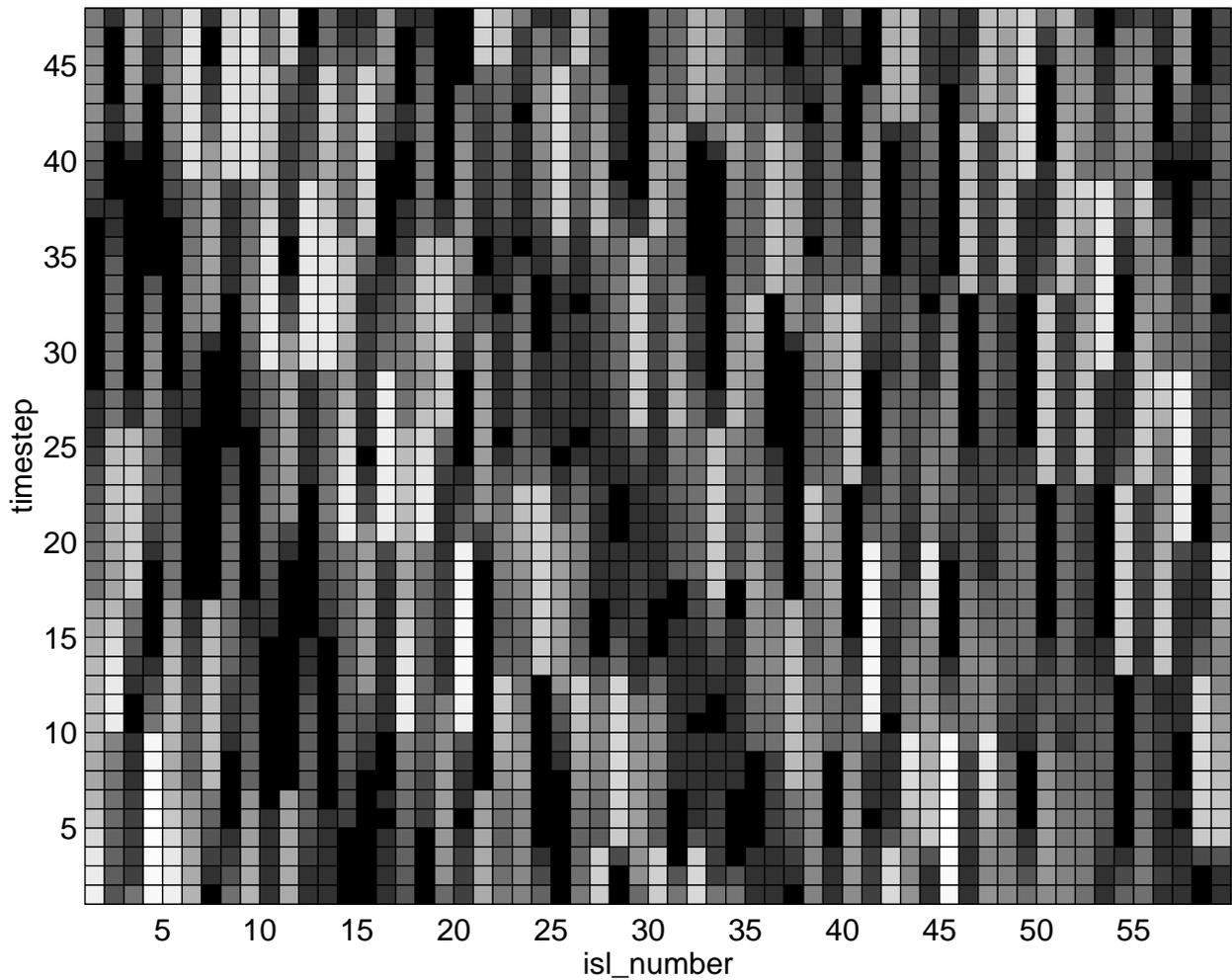
Finally, similar to the case of TFM 2, traffic generated on and delivered to earth is mapped onto the different satellites according to the fraction of the zones within their respective footprints.

# Appendix D

## Illustrations for MLP Simulations

This appendix provides some illustrative material describing the simulation of isolated traffic adaptive routing based on multilayer perceptron (MLP) type neural networks. As discussed in Section 4.4.2, the fundamental idea of the approach is to let the MLP learn a correlation between local traffic patterns (namely, the isolated sets of respective outgoing ISL and downlink loads on each satellite carrying an MLP) and global traffic flows across the ISL mesh, in order to effectively make adaptive routing decisions in the recall phase. Figure D.1 displays for example a graphical representation of the network-wide (outgoing) ISL traffic load pattern for the LEONET ISL mesh over all 50 steps of the constellation period; the load values are encoded in brightness of colors (or grayscale). The ISL numbering in the picture is such that always sets of four neighboring “columns” represent the ISLs of one satellite. A typical input pattern for one MLP is then the pixel array made of the four respective ISLs (abscissa) and the considered number of steps (ordinate) to include gradient information – here neglecting the additional downlink traffic for simplicity. In the course of discrete simulation time, this input pattern array moves or “steps” along the time axis in sliding window manner.

The investigations presented in Section 4.4.2 were performed with the *Stuttgart Neural Network Simulator (SNNS)* [Zel95]. SNNS is a simulator for neural networks developed at the Institute for Parallel and Distributed High Performance Systems at the University of Stuttgart. It consists of four main components, which are depicted in Fig. D.2: simulator kernel, graphical user interface XGUI, batch simulator version SNNSBAT, and the network compiler SNNS2C. Figure D.3 shows a typical screen setup, where the manager panel contains buttons to call all other windows of the interface and displays the status of SNNS. As shown in the bottom right window, the current NN can be displayed with its important neuron activation values encoded in symbol sizes. Such a graphical representation is shown for an MLP on board a LEONET satellite, as used in our simulations, in Fig. D.4 for two subsequent steps. Note how the input traffic pattern of 50 values – (4 ISL loads +1 downlink load) times 10 steps history gradient – moves downwards over the MLP input layer with the corresponding 50 neurons. Finally, Fig. D.5 shows with a snapshot of an Iridium MLP how the size of the NN grows with the size of the constellation and, thus, of the pattern recognition and regression problem to be solved.



**Figure D.1:** Color map representation of the ISL load pattern in the LEONET network over time (brighter color represents higher traffic load).

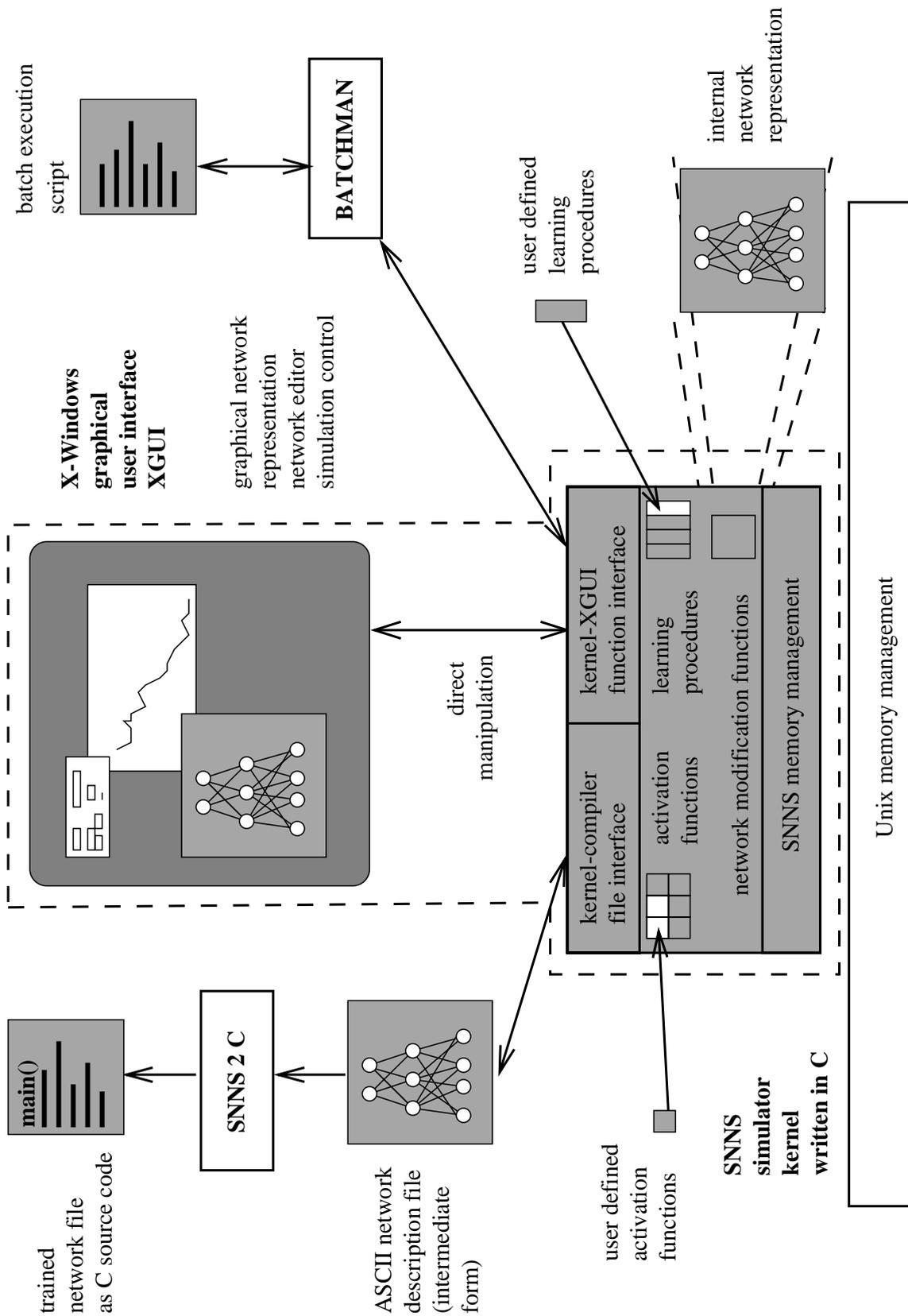


Figure D.2: SNNs components [Ze195].

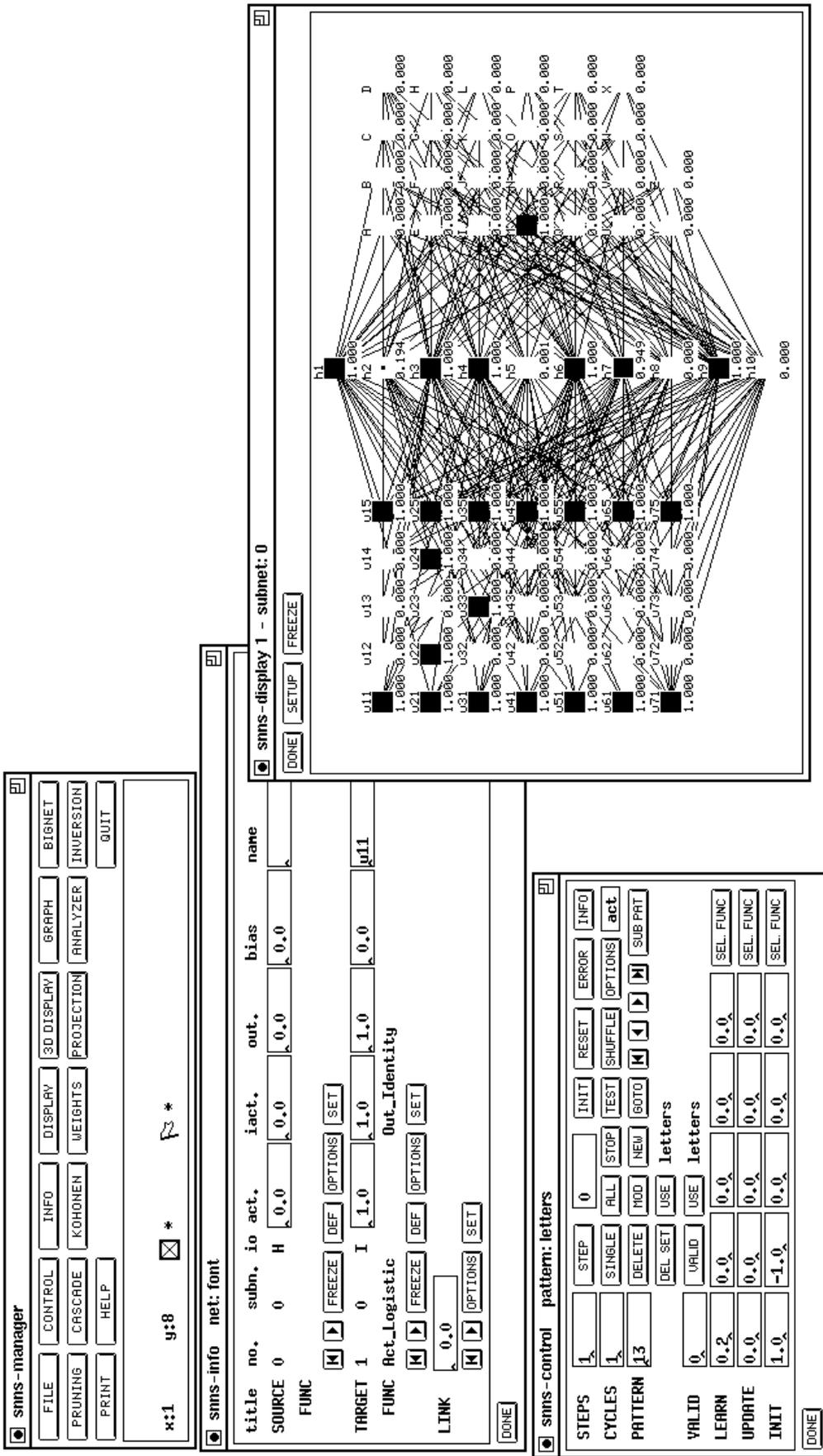
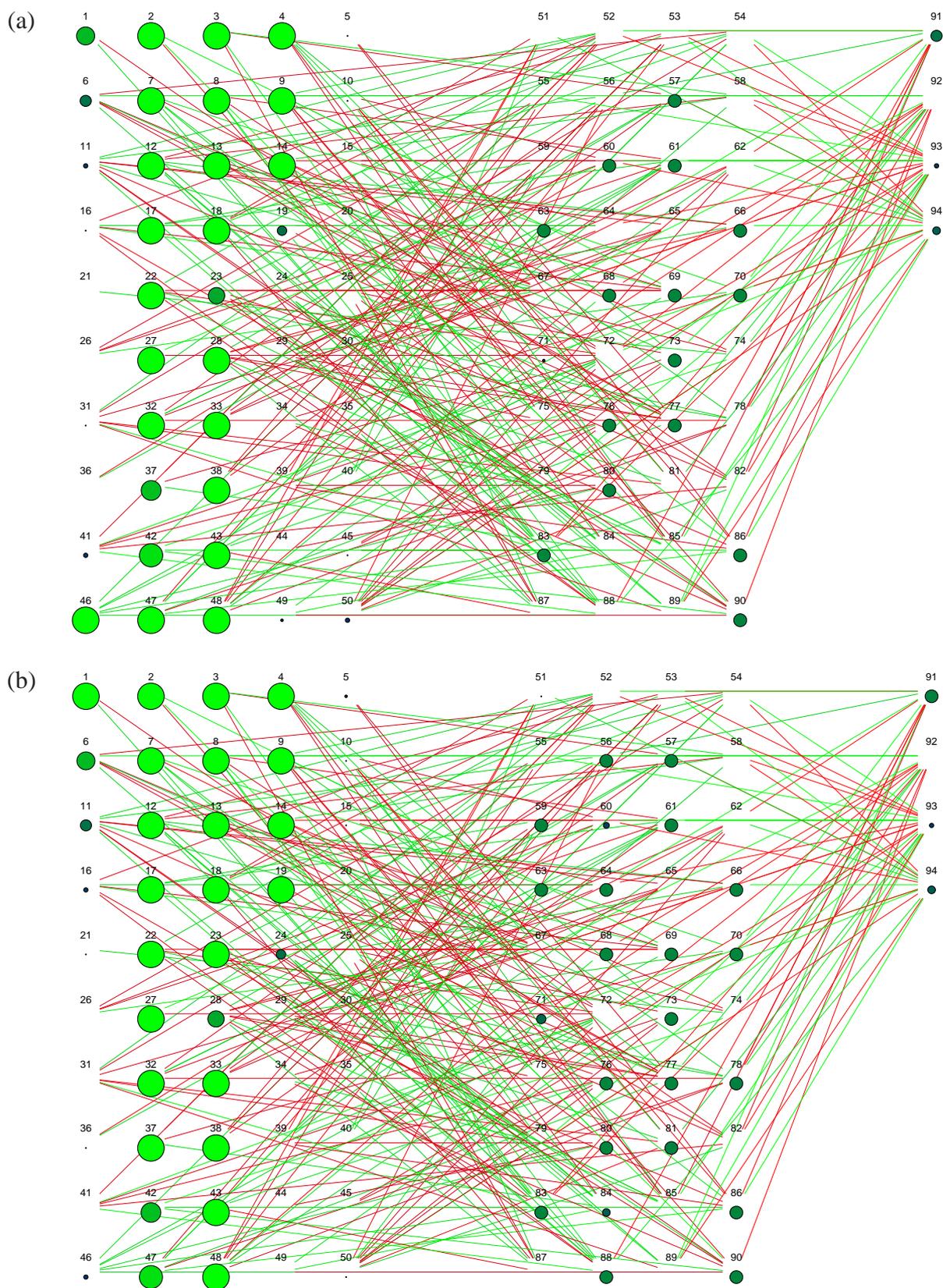
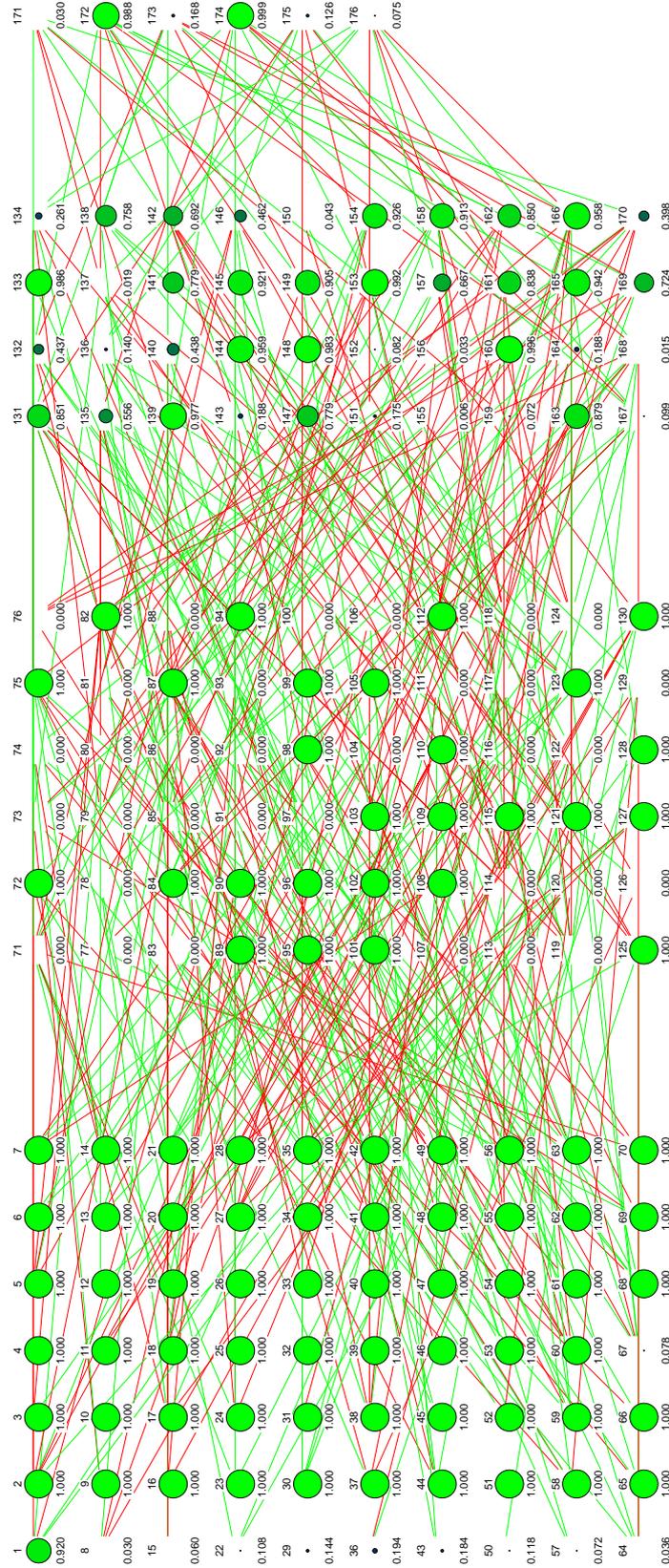


Figure D.3: SNNS manager panel, info panel, control panel and a net display [Zel95].



**Figure D.4:** MLP for the LEONET simulations (a) in time step  $s$  and (b) in time step  $s+1$ ; 50 input neurons, 40 hidden neurons and 4 output neurons.



**Figure D.5:** MLP for the Iridium simulations; 70 input neurons, 60 hidden neurons (hidden layer 1), 40 hidden neurons (hidden layer 2) and 6 output neurons.

# Notation and Symbols

*Ich weiß nicht, was soll es bedeuten . . .*

— HEINRICH HEINE (1797 – 1856)

## Constants

Constant	Value	Meaning
$c_0$	$2.998 \cdot 10^8 \text{ m/s}$	vacuum speed of light
$G$	$6.6732 \cdot 10^{-11} \text{ N m}^2/\text{kg}^2$	universal gravitational constant
$M_E$	$5.9733 \cdot 10^{24} \text{ kg}$	mass of the earth
$R_E$	6378 km	mean equatorial earth radius
$T_E$	23 h 56 min 4.1 s	sidereal day
$T_{\text{Sol}}$	24 h = 86400 s	(mean) solar day
$\mu$	$3.986 \cdot 10^{14} \text{ m}^3/\text{s}^2$	$= GM_E$

## Mathematical Notation and Operators

For this index to mathematical notation and operators, letters that are not further qualified have the following significance:

$j, k$	integer-valued arithmetic expression
$n$	nonnegative integer-valued arithmetic expression
$x, y$	real-valued arithmetic expression
$f$	real-valued function

$(B \Rightarrow E; E')$	conditional expression: denotes $E$ if $B$ is true, $E'$ if $B$ is false
$\sum_{R(k)} f(k)$	sum of all $f(k)$ such that the variable $k$ is an integer and relation $R(k)$ is true
$\min_{R(k)} \{f(k)\}$	minimum value of all $f(k)$ such that the variable $k$ is an integer and relation $R(k)$ is true
$\max_{R(k)} \{f(k)\}$	maximum value of all $f(k)$ such that the variable $k$ is an integer and relation $R(k)$ is true

$f(x) \rightarrow \min$	minimize $f(x)$
$\min_x f(x)$	minimize $f(x)$ for all $x$
$\{a \mid R(a)\}$	set of all $a$ such that the relation $R(a)$ is true
$\{a_1, \dots, a_n\}$	the set $\{a_k \mid 1 \leq k \leq n\}$
$a \in A$	$a$ is element of set $A$
$\forall a \in A$	for all $a$ in set $A$
$[a \dots b]$	closed interval: $\{x \mid a \leq x \leq b\}$
$[a \dots b)$	half-open interval: $\{x \mid a \leq x < b\}$
$ x $	absolute value of $x$ : $(x \geq 0 \Rightarrow x; -x)$
$\lfloor x \rfloor$	floor of $x$ , greatest integer function: $\max_{k \leq x} \{k\}$ ; greatest integer smaller than or equal to $x$
$\lceil x \rceil$	ceiling of $x$ , least integer function: $\min_{k \geq x} \{k\}$ ; smallest integer greater than or equal to $x$
$x \bmod y$	modulo function: $(y = 0 \Rightarrow x; x - y \lfloor x/y \rfloor)$
$j \setminus k$	$j$ divides $k$ : $k \bmod j = 0$ and $j > 0$
$\log x$	logarithm of $x$ to the base 10 (when $x > 0$ ): the $y$ such that $x = 10^y$
$\text{lcm}(j, k)$	least common multiple of $j$ and $k$ : $(j, k = 0 \Rightarrow 0; \min_{d > 0, j \setminus d, k \setminus d} \{d\})$
$\mathbf{x} = (x_k)$	real-valued vector (i.e., a linear $k$ -list of real-valued arithmetic expressions $x_k$ )
$\mathbf{X} = [x_{jk}]$	real-valued matrix (i.e., a two-dimensional $j \times k$ -array of real-valued arithmetic expressions $x_{jk}$ )

## Symbols

$\alpha$	OD traffic share factor	(dimensioning)
$\delta_l^p$	link occupation indicator: link $l$ used by path $p$ or not	(routing)
$2\Delta$	angular spacing between orbit planes at the equator for counter-rotating orbits (in star constellations)	(satellite constellation)
$\varepsilon$	elevation angle	(earth/space geometry)
$\varepsilon_{\min}$	minimum elevation angle (at edge of coverage area)	(earth/space geometry)
$\varepsilon_{\text{GW}}$	minimum elevation angle for a gateway	(satellite diversity)
$\varepsilon_{\text{MT}}$	minimum elevation angle for a mobile terminal	(satellite diversity)
$\varepsilon_{\max}$	maximum ISL elevation angle	(satellite constellation)
$\eta$	learning factor of a backpropagation algorithm	(neural network)
$\theta$	true anomaly	(earth/space geometry)
$\vartheta$	nadir angle	(earth/space geometry)
$\vartheta_{\max}$	maximum nadir angle	(earth/space geometry)
$\lambda$	longitude; call arrival rate	(earth/space geometry) (traffic model)
$\lambda_0$	longitude of the ascending node;	(earth/space geometry)
$\lambda_S(t)$	longitude of a satellite (or sub-satellite point) at time $t$	(earth/space geometry)

$\mu$	mean power-level decrease due to shadowing	(satellite diversity)
$1/\mu$	mean call holding time	(traffic model)
$\sigma$	standard deviation of power level due to shadowing	(satellite diversity)
$\tau$	sliding window duration (in discrete time steps); age of a call (in discrete time steps)	(routing) (traffic model)
$\hat{\tau}_{\text{old}}$	maximum age of a call (in discrete time steps)	(traffic model)
$\varphi_S(t)$	latitude of a satellite (or sub-satellite point) at time $t$	(earth/space geometry)
$\varphi$	latitude	(earth/space geometry)
$\psi$	(one-sided) earth central angle = angular radius of a coverage circle	(earth/space geometry)
$\psi_{\text{max}}$	coverage angle	(earth/space geometry)
$\omega$	nodal angular elongation	(earth/space geometry)
$\omega_0$	argument of perigee	(earth/space geometry)
$\Delta\omega_p$	plane phasing angle = angular phase offset between satellites in adjacent orbit planes (for delta constellations)	(satellite constellation)
$\Omega$	right ascension of the ascending node (RAAN)	(earth/space geometry)
$\Delta\Omega$	angular spacing between co-rotating orbit planes at the equator	(satellite constellation)
$\dot{\Omega}_E$	angular frequency of the earth rotation	(earth/space geometry)
$a$	carried traffic	(traffic model)
$a_{ij}$	traffic carried on link $(i, j)$	(routing)
$a_n$	(cumulative ISL) traffic in node $n$	(routing)
$\mathbf{a}$	input vector of a multilayer perceptron	(neural network)
$A$	time-share of shadowing; offered traffic	(satellite diversity) (traffic model)
$A_{\text{net}}$	total offered network traffic	(network traffic model)
$A_{\text{OD}}$	offered traffic for an OD pair (of satellites)	(network traffic model)
$A_{\text{WCL}}$	offered traffic on the worst-case link (ISL)	(network traffic model)
$\mathbf{A}_{\text{net}}$	network traffic matrix	(network traffic model)
$\mathbf{A}_{\text{reg}}$	regional traffic matrix	(network traffic model)
$\mathbf{b}$	output vector of a multilayer perceptron	(neural network)
$B$	(Erlang) blocking probability	(traffic model)
$c$	direct-to-multipath signal power ratio (Rice factor)	(satellite diversity)
$c_0$	vacuum speed of light	
$c_{ij}$	costs associated with link $(i, j)$	(network model; routing)
$c_n$	costs associated with node $n$	(network model; routing)
$C$	capacity, number of channels	(traffic model)
$C_{\text{ISL}}$	ISL capacity (number of channels)	(network traffic model)
$C_p$	costs associated with path $p$	(network model; routing)
$d$	slant range from earth terminal to satellite	(earth/space geometry)
$d_{ij}$	delay associated with link $(i, j)$	(routing)

$d_{\min}$	minimum distance between an ISL and the earth's surface	(earth/space geometry)
$d_{\max}$	maximum slant range	(earth/space geometry)
$D_b$	mean duration of bad channel state	(satellite diversity)
$D_g$	mean duration of good channel state	(satellite diversity)
$e$	eccentricity of an orbit	(earth/space geometry)
$\mathbf{e}$	error vector of a multilayer perceptron	(neural network)
$E$	set of undirected links (edges); eccentric anomaly	(network model) (earth/space geometry)
$E(A, C)$	Erlang-B (blocking) function	(traffic model)
$E(s)$	time-varying set of undirected links (edges)	(network model)
$f$	frequency	
$F$	phasing factor of a delta constellation (Walker notation)	(satellite constellation)
$F_g$	gravitational force	(earth/space geometry)
$g(x)$	transfer function of a neuron	(neural network)
$G$	universal gravitational constant	(earth/space geometry)
$G(V, E)$	network graph: graph with set of nodes $V$ and set of undirected links $E$	(network model)
$G(s)$	$= G(V, E(s))$ , network graph at step $s$	(network model)
$h$	orbit altitude; hop count = number of hops on a path	(satellite constellation) (routing)
$i$	inclination angle of an orbit plane; node index	(satellite constellation) (network model)
$(i, j)$	undirected link between nodes $i$ and $j$	(network model)
$(i, j)_s$	undirected link between nodes $i$ and $j$ at step $s$	(network model)
$j$	node (index)	(network model)
$k$	number of alternative (shortest) paths	(routing)
$k^*$	effective number of group-completing paths	(routing)
$k_w^*$	number of group-completing paths for OD pair $w$	(routing)
$\overline{k^*}$	average number of group-completing OD paths in an ISL network	(routing)
$K$	maximum number of alternative (shortest) paths	(routing)
$l$	link (index)	(network model; routing)
$l_{\max}$	maximum ISL length	(satellite constellation)
$L$	number of intersatellite links (ISLs) in a constellation	(satellite constellation)
$m$	satellite mass (earth/space geometry)	
$M, M_E$	mass of the earth	(earth/satellite geometry)
$n$	number of potential diversity satellites	(satellite diversity)
$n_l$	link capacity	(dimensioning)
$n_{\max}$	maximum link capacity	(dimensioning)
$n_p$	path capacity	(dimensioning)
$\mathbf{n}_p$	vector of path capacities	(dimensioning)
$n_p^{\text{new}}(s)$	new path traffic/capacity at step $s$	(dimensioning)

$n_p^{\text{opt}}(s)$	path traffic/capacity component subject to routing optimization at step $s$	(dimensioning)
$n_p^{\text{rem}}(s)$	remaining path traffic/capacity at step $s$	(dimensioning)
$n_p^{\text{rer}}(s)$	(incoming) rerouted path traffic/capacity at step $s$	(dimensioning)
$n_w$	OD demand pair capacity	(dimensioning)
$n_w^{\text{new}}(s)$	new OD pair traffic/capacity at step $s$	(dimensioning)
$n_w^{\text{rem}}(s)$	remaining OD pair traffic/capacity at step $s$	(dimensioning)
$n_w^{\text{rer}}(s)$	(incoming) rerouted OD pair traffic/capacity at step $s$	(dimensioning)
$N$	number of satellites in a constellation;	(satellite constellation)
	number of nodes in a network	(network model)
$N_S$	number of satellites per orbit plane	(satellite constellation)
$p$	path (index)	(network model; routing)
$p(s)$	path existing at step $s$ (for an OD pair $w$ )	(network model; routing)
$p_{\text{HO}}$	handover probability	(satellite constellation)
$p_n(w)$	$n$ -th path in the path set for OD pair $w$	(routing)
$p_{\text{rer}}$	rerouting probability	(satellite constellation)
$P$	number of orbit planes in a constellations	(satellite constellation)
$P_w$	set of distinct loopless paths for OD pair $w$	(network model)
$P_w(s)$	time-varying set of distinct loopless paths for OD pair $w$	(network model)
$Q$	number of ordered path sequences for an OD pair	(network model; routing)
$r$	orbit radius; distance of a satellite from the earth's center	(earth/space geometry)
$r_a$	apogee distance	(earth/space geometry)
$r_p$	perigee distance	(earth/space geometry)
$r_{\text{SA}}$	radius of a gateway's service area	(satellite diversity)
$\mathbf{r}$	noise component of an input vector of a multilayer perceptron	(neural network)
$R_E$	mean equatorial earth radius	(earth/space geometry)
$s$	time step (index)	(network model)
$S$	number of constellation/topology snapshots;	(satellite constellation)
	number of discrete time steps per constellation period $T$	(network model)
$t$	time	
$t_0$	reference or initial time	
$\Delta t$	time difference;	
	(discretization) step width	(network model)
$\Delta t_s$	time step	(routing)
$\Delta t_u$	routing update period	(routing)
$T$	orbit period; constellation period	(satellite constellation)
	dynamic topology period	(network model; routing)
	optimization period	(routing; dimensioning)
$T_{\text{activate}}$	activation timer for a diversity satellite channel	(satellite diversity)
$T_{\text{break,b}}$	handover break for backward handover	(satellite handover)
$T_{\text{break,f}}$	handover break for forward handover	(satellite handover)
$T_{\text{drop}}$	deactivation timer for a diversity satellite channel	(satellite diversity)

$T_E$	sidereal day	(earth/space geometry)
$T_{\text{proc}}$	accumulated processing delay for a satellite handover	(satellite handover)
$T_{\text{prop}}$	propagation delay between user and gateway (via a satellite)	(satellite handover)
$T_S$	system period	(satellite constellation)
$u_j$	external input bias of neuron $j$	(neural network)
$v$	satellite velocity	(earth/space geometry)
$v_i$	output value of predecessor neuron $i$	(neural network)
$v_j$	output value of neuron $j$	(neural network)
$v_u$	(mobile) user velocity	(satellite diversity)
$V$	set of nodes (vertices)	(network model)
$w_{ij}$	weight between neurons $i$ and $j$	(neural network)
$w$	OD pair (index)	(network model)
$W$	set of unordered OD pairs	(network model)
$\mathbf{W}$	weight matrix of a neural network	(neural network)
$x_i$	input value of a neuron from predecessor neuron $i$	(neural network)
$x_j$	total input value of neuron $j$	(neural network)

# Abbreviations

*... this is the short and the long of it.*

— WILLIAM SHAKESPEARE, *The Merry Wives of Windsor* (Act II, Scene 2)

AAL	ATM adaptation layer
ABR	available bit rate
ANN	artificial neural network
ATM	asynchronous transfer mode
B-ISDN	Broadband Integrated Services Digital Network
BO	bounded optimization
BO/G	BO with path grouping
BPA	backpropagation algorithm
BU	backup (path)
CAC	connection admission control
CASD	channel adaptive satellite diversity
CBR	constant bit rate
CDMA	code division multiple access
CL	connectionless
CO	connection-oriented
DBF	distributed Bellman–Ford (algorithm)
DCA	dynamic channel allocation (or assignment)
DLC	data link control
DLR	<i>Deutsches Zentrum für Luft- und Raumfahrt</i> , German Aerospace Center
DSPA	Dijkstra shortest-path algorithm
DT-MSD	discrete-time multi-step dimensioning
DT-PSS	discrete-time path sequence selection
DT-VTD	discrete-time virtual topology design
DVTR	dynamic virtual topology routing
DWF	delay weight factor
EHF	extremely high frequency
EQX	equator crossing
ES	equal sharing
ES/G	ES with path grouping

FL	feeder link
FO	full optimization
FO/G	FO with path grouping
FT	fixed (user) terminal
FU	fixed user
GEO	geostationary earth orbit
GSM	Global System for Mobile Communication
GW, GTW	gateway (= fixed earth station)
HEO	highly elliptical orbit
hES	history-based ES
hES/G	history-based ES with path grouping
hFO	history-based FO
hFO/G	history-based FO/G
HLR	home location register
HO	handover
ICO	intermediate circular orbit (synonym for MEO); also name of a MEO satellite system
ID	identifier
iES	isolated step ES
iES/G	isolated step ES/G
iFO	isolated step FO
iFO/G	isolated step FO/G
IP	internet protocol
ISDN	Integrated Services Digital Network
ISL	intersatellite link
ISLDIM	ISL Network Dimensioning Tool (proprietary software tool of DLR)
ISLSIM	ISL Network Simulator (proprietary software tool of DLR)
ISC	international switching center
ISO	International Organization of Standardization
I-VTS	instantaneous virtual topology setup
KSPA	$k$ -shortest-path algorithm
LEO	low earth orbit
LEOSIM	LEO System Simulator (proprietary software tool of DLR)
LOS	line-of-sight
LP	linear programming
MAC	medium access control
MCHT	mean call holding time
MDA	Moore–Dijkstra algorithm
MEO	medium earth orbit
MHVP	multi hop virtual path
MINOS	Modular In-core Nonlinear Optimization System (commercial software tool)
MLP	multilayer perceptron
MPEG	moving pictures expert group

MPLS	multiprotocol label switching
MS	mobile station
MSA	mobile service area
MSS	mobile satellite service
MT	mobile (user) terminal
MU	mobile user
MUL	mobile user link
NCC	network control center
NN	neural network
NNI	network–network interface
OD	origin–destination
OSI	Open System Interconnection
PAT	pointing, acquisition and tracking
PCN	personal communication network
PG	path grouping
PHY	physical layer
PL	physical link
P-PTD	permanent physical topology design
PSTN	public switched telephone network
QoS	quality of service
RAAN	right ascension of the ascending node
RF	radio frequency
SA	service area
SCC	satellite control center
SHVP	single hop virtual path
SL	service link
SNNS	Stuttgart Neural Network Simulator (public domain software tool)
SPA	shortest-path algorithm
S-PCN	satellite personal communication network
SSP	sub-satellite point
T1, T2	(ISL) topology 1, (ISL) topology 2
TCP/IP	transport control protocol / internet protocol
TDMA	time division multiple access
TF	target function
TMF	(global) traffic modeling framework
TNL	terrestrial network link
TT&C	telemetry, tracking, and command
TWF	traffic weight factor
UBR	unspecified bit rate
UDL	up/downlink
UMTS	Universal Mobile Telecommunication System
UNI	user–network interface
UP	user position

UTC	universal time coordinated
VBR	variable bit rate
VC	virtual channel
VCC	virtual channel connection
VCI	virtual channel identifier
VCT	virtual connection tree
VLR	visitor location register
VP	virtual path
VPC	virtual path connection
VPI	virtual path identifier
WCL	worst-case link
WCN	worst-case node

# Bibliography

*Knowledge is of two kinds. We know a subject ourselves,  
or we know where we can find information upon it.*

— SAMUEL JOHNSON (1775)

*You will, I am sure, agree with me . . . that if page 534 finds us only in the  
second chapter, the length of the first one must have been really intolerable.*

— SHERLOCK HOLMES, in *The Valley of Fear* (1888)

- [AN94] A. S. Acampora and M. Naghshineh, “An architecture and methodology for mobile-executed handoff in cellular ATM networks,” *IEEE Journal on Selected Areas in Communications*, vol. 12, pp. 1365–1375, Oct. 1994.
- [AR87] W. S. Adams and L. Rider, “Circular polar constellations providing continuous single or multiple coverage above a specified latitude,” *Journal of the Astronautical Sciences*, vol. 35, pp. 155–192, Apr.–June 1987.
- [Ash97] G. R. Ash, *Dynamic Routing in Telecommunications Networks*. New York: McGraw-Hill, 1st ed., 1997.
- [Bal80] A. H. Ballard, “Rosette constellations of earth satellites,” *IEEE Transactions on Aerospace and Electronic Systems*, vol. 16, pp. 656–673, Sept. 1980.
- [Bau97a] T. Bauschert, “Multihour design of multi-hop virtual path based wide-area ATM networks,” in *Teletraffic Contributions for the Information Age, Proceedings 15th International Teletraffic Congress (ITC-15)*, V. Ramaswami and P. E. Wirth (eds.), Washington, DC, USA, pp. 1019–1029, June 1997. Elsevier.
- [Bau97b] T. Bauschert, *Optimale Dimensionierung von ATM-Weitverkehrsnetzen mit mehrstufiger Durchschaltung*. PhD thesis, Munich Technical University, Institute of Communication Networks, Munich, Germany, 1997. Herbert Utz Verlag, Munich, Germany, 1997. GERMAN LANGUAGE.
- [Ber96] G. Berndl, “Ein Routingkonzept zur störungsfreien ATM-Pfadumschaltung in LEO-Satellitennetzen mit Intersatellitenverbindungen (Engl.: A routing concept for hitless ATM virtual path switching in LEO satellite networks employing intersatellite links),” Master’s thesis, Munich Technical University, Institute of Communication Networks, Munich, Germany, June 1996. GERMAN LANGUAGE.

- [Bes78] D. C. Beste, "Design of satellite constellations for optimal continuous coverage," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 14, pp. 466–473, May 1978.
- [BG87a] D. Bertsekas and R. G. Gallager, *Data Networks*. Englewood Cliffs, NJ: Prentice-Hall, 1987.
- [BG87b] D. Bertsekas and R. G. Gallager, *Data Networks*, ch. 5.2. Englewood Cliffs, NJ: Prentice-Hall, 1987.
- [BHG87] R. Binder, S. D. Huffman, I. Gurantz, and P. A. Vena, "Crosslink architectures for a multiple satellite system," *Proceedings of the IEEE*, vol. 75, pp. 74–82, Jan. 1987.
- [BJL<sup>+</sup>93] A. Böttcher, A. Jahn, E. Lutz, B. Schmidt, and M. Werner, "Networking requirements for user oriented LEO satellite systems." ESA Study No. 9732/91/NL/RE, Final Report, ESA/ESTEC, Noordwijk, The Netherlands, Feb. 1993.
- [BJLW94] A. Böttcher, A. Jahn, E. Lutz, and M. Werner, "Analysis of basic system parameters of communication networks based on low earth orbit satellites," *International Journal of Satellite Communications*, vol. 12, pp. 85–93, Jan./Feb. 1994.
- [Bra84] K. Brayer, "Packet switching for mobile earth stations via low-orbit satellite network," *Proceedings of the IEEE*, vol. 72, pp. 1627–1636, Nov. 1984.
- [Bur97] K. Burchard, "Application of virtual path concepts to the broadband LEO satellite system M-Star," Master's thesis, Munich Technical University, Institute of Communication Networks, Munich, Germany, July 1997.
- [BW94a] A. Böttcher and M. Werner, "Personal satellite communications: Traffic and capacity considerations for the mobile user link," in *Proceedings IEEE GLOBECOM '94*, San Francisco, California, USA, pp. 760–764, Nov./Dec. 1994.
- [BW94b] A. Böttcher and M. Werner, "Strategies for handover control in low earth orbit satellite systems," in *Proceedings 44th IEEE Vehicular Technology Conference (VTC '94)*, Stockholm, Sweden, pp. 1616–1620, June 1994.
- [BW97] H. Bischl and M. Werner, "Channel adaptive satellite diversity for non-geostationary mobile satellite systems," in *Proceedings 5th International Mobile Satellite Conference (IMSC '97)*, Pasadena, California, USA, pp. 25–31, June 1997.
- [BWL96] H. Bischl, M. Werner, and E. Lutz, "Elevation-dependent channel model and satellite diversity for NGSO S-PCNs," in *Proceedings 46th IEEE Vehicular Technology Conference (VTC '96)*, Atlanta, Georgia, USA, pp. 1038–1042, Apr./May 1996.
- [CB95] P. Carter and M. A. Beach, "Evaluation of handover mechanisms in shadowed low earth orbit land mobile satellite systems," *International Journal of Satellite Communications*, vol. 13, pp. 177–190, 1995.
- [Cha89] D. Chakraborty, "Survivable communication concept via multiple low earth-orbiting satellites," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 25, pp. 879–889, Nov. 1989.

- [Cla45] A. C. Clarke, "Extra-terrestrial relays," *Wireless World*, vol. 51, pp. 305–308, Oct. 1945.
- [CSM92] S. Cavalieri, A. Stefano, and O. Mirabella, "A neural-network-based approach for routing in a packet switching network," in *Proceedings International Joint Conference on Neural Networks*, pp. 913–918, 1992.
- [Dav85] M. R. Davidoff, *The Satellite Experimenter's Handbook*. 225 Main Street, Newington, CT 06111: The American Radio Relay League, 1985.
- [Del95] C. Delucchi, "Routing strategies in LEO/MEO satellite networks with intersatellite links," Master's thesis, Munich Technical University, Institute of Communication Networks, Munich, Germany, Aug. 1995.
- [DFR96] E. Del Re, R. Fantacci, and L. Ronga, "A dynamic channel allocation technique based on hopfield neural networks," *IEEE Transactions on Vehicular Technology*, vol. 45, pp. 26–32, Feb. 1996.
- [DH98] B. Doshi and P. Harshavardhana, "Broadband network infrastructure of the future: Roles of network design tools in technology deployment strategies," *IEEE Communications Magazine*, vol. 36, pp. 60–71, May 1998.
- [Dij59] E. W. Dijkstra, "A note on two problems in connection with graphs," *Numerische Mathematik*, vol. 1, pp. 269–271, 1959.
- [DKFW97] T. Dreischer, H. Kellermeier, E. Fischer, and B. Wandernoth, "Advanced miniature optical terminal family for inter-satellite links in space communication networks," in *Proceedings 4th European Conference on Satellite Communications (ECSC 4)*, Rome, Italy, pp. 73–78, Nov. 1997.
- [dP93] M. de Prycker, *Asynchronous Transfer Mode: Solution for Broadband ISDN*. Series in Computer Communications and Networking, New York, London: Ellis Horwood Limited, 2nd ed., 1993.
- [Edm96] B. Edmaier, *Pfad-Ersatzschaltverfahren mit verteilter Steuerung für ATM-Netze*. PhD thesis, Munich Technical University, Institute of Communication Networks, Munich, Germany, 1996. Herbert Utz Verlag, Munich, Germany, 1996. GERMAN LANGUAGE.
- [EEFK95] B. Edmaier, J. Eberspächer, W. Fischer, and A. Klug, "Alignment server for hitless path-switching in ATM networks," in *Proceedings XV International Switching Symposium (ISS '95)*, Berlin, Germany, pp. 403–407, Apr. 1995.
- [EF00] I. El Khamlichi and L. Franck, "Study of two policies for implementing routing algorithms in satellite constellations," in *Proceedings 18th AIAA International Communications Satellite Systems Conference (ICSSC '00)*, Oakland, CA, USA, pp. 579–589, Apr. 2000.
- [Eis96] M. Eisenschmid, "Handover in LEO/MEO-Satellitennetzen (Engl.: Handover in LEO/MEO satellite networks)," Master's thesis, Munich Technical University, Institute of Communication Networks, Munich, Germany, Mar. 1996. GERMAN LANGUAGE.

- [EVB01] J. Eberspächer, H.-J. Vögel, and C. Bettstetter, *GSM – Switching, Services and Protocols*. Chichester: John Wiley & Sons, 2nd ed., 2001.
- [EVW96] M. Eisenschmid, H.-J. Vögel, and M. Werner, “Handover signalling in LEO/MEO satellite systems,” in *Proceedings 2nd International Conference on Personal, Mobile and Spread Spectrum Communications (ICPMSC '96)*, Hong Kong, pp. 117–120, Dec. 1996.
- [FM91] T. Fritsch and W. Mandel, “Communication network routing using neural nets - numerical aspects and alternative approaches.” Research Report No. 30, University of Würzburg, Institute of Computer Science, July 1991.
- [FMTG93] T. Fritsch, M. Mittler, and P. Tran-Gia, “Artificial neural net applications in telecommunication systems,” *Neural Computing and Applications*, pp. 124–146, Apr. 1993.
- [Fra01] L. Franck, *Routing Algorithms for Inter-Satellite Link Networks*. PhD thesis, Ecole Nationale Supérieure des Télécommunications, Toulouse, France, 2001.
- [Gal99] J. Galtier, “Geographical reservation for guaranteed handover and routing in low earth orbit constellations,” in *Proceedings Workshop de Comunicacao Sem Fio (WCSF '99)*, Belo Horizonte, Brazil, July 1999.
- [Gal01] J. Galtier, “Geographical reservation for guaranteed handover and routing in low earth orbit constellations,” *Telecommunication Systems*, 2001.
- [Gar96] B. Garnier, “Network concepts for ATM traffic flows in low earth orbit satellite constellations with intersatellite links (ISL),” Master’s thesis, ENST Toulouse / DLR Oberpfaffenhofen, Toulouse, France / Wessling, Germany, Sept. 1996.
- [GHJ95] T. Goerke, N. Hart, and A. Jahn, “A discussion on mobile satellite systems and the myths of CDMA and diversity revealed,” in *Proceedings 4th International Mobile Satellite Conference (IMSC '95)*, Ottawa, Canada, pp. 469–475, June 1995.
- [Gir90] A. Girard, *Routing and Dimensioning in Circuit-Switched Networks*. Reading, MA: Addison-Wesley, 1990.
- [GMW81] P. E. Gill, W. Murray, and M. H. Wright, *Practical Optimization*. London: Academic Press Ltd., 1981.
- [Gru91] J. Grubb, “Iridium overview,” *IEEE Communications Magazine*, vol. 29, Nov. 1991.
- [Hes93] G. C. Hess, *Land-Mobile Radio System Engineering*. Norwood, MA: Artech House Inc., 1993.
- [HH91] R. Händel and M. Huber, *Integrated Broadband Networks: An Introduction to ATM-Based Networks*. London: Addison-Wesley, 1991.
- [Hir90] A. Hiramatsu, “ATM communications network control by neural networks,” *IEEE Transactions on Neural Networks*, vol. 1, pp. 122–130, Mar. 1990.
- [HKP91] J. Hertz, A. Krogh, and R. G. Palmer, *Introduction to the Theory of Neural Computation*. Reading, MA: Addison-Wesley, 1991.

- [HL95] J. Hutcheson and M. Laurin, "Network flexibility of the IRIDIUM global mobile satellite system," in *Proceedings 4th International Mobile Satellite Conference (IMSC '95)*, Ottawa, Canada, pp. 503–507, June 1995.
- [Jah94] A. Jahn, "Measurement programme for generic satellite channels." Final Report, Inmarsat Purchase Order P004001, Inmarsat, London, UK, Nov. 1994.
- [Jah95] A. Jahn, "Propagation data and channel model for LMS systems." Final Report, ESA Purchase Order 141742, ESA/ESTEC, Noordwijk, The Netherlands, Jan. 1995.
- [Jah99] A. Jahn, *Ressourcenverwaltung in Kommunikationsnetzen mit niedrigfliegenden Satelliten*. PhD thesis, FernUniversität Hagen, Hagen, Germany, 1999. Herbert Utz Verlag, Munich, Germany, 2000. GERMAN LANGUAGE.
- [JL94] A. Jahn and E. Lutz, "DLR channel measurement programme for low earth orbit satellite systems," in *Proceedings 3rd International Conference on Universal Personal Communications (ICUPC '94)*, San Diego, CA, USA, pp. 423–429, Sept./Oct. 1994.
- [JSBL95] A. Jahn, M. Sforza, S. Buonomo, and E. Lutz, "Narrow and wideband channel characterization for land mobile satellite systems: Experimental results at L-band," in *Proceedings 4th International Mobile Satellite Conference (IMSC '95)*, Ottawa, Canada, pp. 115–121, June 1995.
- [KA93] F. Kamoun and M. Ali, "Neural networks for shortest path computation and routing in computer networks," *IEEE Transactions on Neural Networks*, vol. 4, pp. 941–953, Nov./Dec. 1993.
- [KM00] W. Krewel and G. Maral, "Analysis of the impact of handover strategies on the QoS of satellite diversity based communications systems," in *Proceedings 18th AIAA International Communications Satellite Systems Conference (ICSSC '00)*, Oakland, CA, USA, pp. 393–403, Apr. 2000.
- [Kro96] O. Kroner, "Traffic load analysis and adaptive routing in the Iridium intersatellite link network," Master's thesis, Munich Technical University, Institute of Communication Networks / ENST Toulouse, Munich, Germany / Toulouse, France, Dec. 1996.
- [LCD<sup>+</sup>91] E. Lutz, D. Cygan, M. Dippold, F. Dolainsky, and W. Papke, "The land mobile satellite communication channel – recording, statistics, and channel model," *IEEE Transactions on Vehicular Technology*, vol. 40, pp. 375–386, May 1991.
- [LD97] Y.-C. Liu and C. Douligieris, "Rate regulation with feedback controller in ATM networks – a neural network approach," *IEEE Journal on Selected Areas in Communications*, vol. 15, pp. 200–208, Feb. 1997.
- [Lee93] W. C. Y. Lee, *Mobile Communications Design Fundamentals*. New York: John Wiley & Sons, 2nd ed., 1993.
- [Leo91] R. J. Leopold, "Low-earth orbit global cellular communications network," in *Proceedings IEEE International Conference on Communications (ICC '91)*, pp. 1108–1111, June 1991.

- [Log97] T. Logsdon, *Orbital Mechanics: Theory and Applications*. New York: John Wiley & Sons, 1997.
- [LR94] E. H. Lipper and M. P. Rumsewicz, "Teletraffic considerations for widespread deployment of PCS," *IEEE Network Magazine*, vol. 8, pp. 40–49, Sept. 1994.
- [Lüd61] R. D. Lüders, "Satellite networks for continuous zonal coverage," *American Rocket Society Journal*, vol. 31, pp. 179–184, Feb. 1961.
- [Lut95] G. Luton, "An ATM based concept for handover operation in LEO/MEO satellite systems," Master's thesis, ENST Toulouse / DLR Oberpfaffenhofen, Toulouse, France / Wessling, Germany, Aug. 1995.
- [Lut96] E. Lutz, "A markov model for correlated land mobile satellite channels," *International Journal of Satellite Communications*, vol. 14, pp. 333–339, July/Aug. 1996.
- [LWJ00] E. Lutz, M. Werner, and A. Jahn, *Satellite Systems for Personal and Broadband Communications*. Berlin, Heidelberg, New York: Springer, 1st ed., 2000.
- [Mar94] G. Maral (Ed.), "Personal communications via satellite," *International Journal of Satellite Communications*, vol. 12, pp. 1–134, Jan./Feb. 1994. Special Issue.
- [May96] C. Mayer, "Neural networks for routing in LEO/MEO satellite systems with inter-satellite links," Master's thesis, Munich Technical University, Institute of Communication Networks, Munich, Germany, Sept. 1996.
- [MB98] G. Maral and M. Bousquet, *Satellite Communications Systems*. Series in Communication and Distributed Systems, Chichester: John Wiley & Sons, 3rd ed., 1998.
- [MCCL95] W. M. Moh, M.-J. Chen, N.-M. Chu, and C.-D. Liao, "Traffic prediction and dynamic bandwidth allocation over ATM: A neural network approach," *Computer Communications*, vol. 18, pp. 563–571, Aug. 1995.
- [MDER91] G. Maral, J.-J. De Ridder, B. G. Evans, and M. Richharia, "Low earth orbit satellite systems for communications," *International Journal of Satellite Communications*, vol. 9, pp. 209–225, Sept./Oct. 1991.
- [Mot96] Motorola Satellite Systems, Inc., "Application for authority to construct, launch and operate the M-Star system." FCC filing, Washington, D.C., USA, Sept. 1996.
- [Mot97] Motorola Global Communications, Inc., "Application for authority to construct, launch and operate the Celestri multimedia LEO system." FCC filing, Washington, D.C., USA, June 1997.
- [MRD<sup>+</sup>98] G. Maral, J. Restrepo, E. Del Re, R. Fantacci, and G. Giambene, "Performance analysis for a guaranteed handover service in an LEO constellation with a "satellite-fixed cell" system," *IEEE Transactions on Vehicular Technology*, vol. 47, pp. 1200–1214, Nov. 1998.
- [MS93] B. A. Murtagh and M. A. Saunders, "MINOS 5.4 user's guide," Tech. Rep. SOL 83-20R, Stanford University, Stanford, CA, USA, Mar. 1993.

- [MS94] R. J. T. Morris and B. Samadi, "Neural network control of communications systems," *IEEE Transactions on Neural Networks*, vol. 5, pp. 639–650, July 1994.
- [MS98] D. McDysan and D. Spohn, *ATM Theory and Applications*. New York: McGraw-Hill, 1998.
- [Pap96] M. Papageorgiou, *Optimierung: statische, dynamische, stochastische Verfahren für die Anwendung*. München, Wien: Oldenbourg, 2nd ed., 1996. GERMAN LANGUAGE.
- [Pos96] P. Poskett, "Satellite system architectures," in *Mobile and Personal Satellite Communications 2, Proceedings 2nd European Workshop on Mobile/Personal Satcoms (EMPS '96)*, F. Vatalaro and F. Ananasso (eds.), Rome, Italy, pp. 485–500, Oct. 1996. Springer.
- [PRFT99] S. R. Pratt, R. A. Raines, C. E. Fossa, and M. A. Temple, "An operational and performance overview of the Iridium low earth orbit satellite system," *IEEE Communications Surveys*, vol. 1, no. 3, pp. 2–10, 1999.
- [Pro89] J. G. Proakis, *Digital Communications*. New York: McGraw-Hill, 1989.
- [Rah95] M. Rahnema, "Method and apparatus for adaptive directed route randomization and distribution in a richly connected communication network." US patent No. 5430729, 1995.
- [Rév99] P. Révillon, "Optimization issues in capacity dimensioning of LEO intersatellite link networks," Master's thesis, SUPAERO Toulouse / DLR Oberpfaffenhofen, Toulouse, France / Wessling, Germany, Aug. 1999.
- [Rid85] L. Rider, "Optimized polar orbit constellations for redundant earth coverage," *Journal of the Astronautical Sciences*, vol. 33, pp. 147–161, Apr.–June 1985.
- [RM96] J. Restrepo and G. Maral, "Constellation sizing for non-geo 'earth-fixed cell' satellite systems," in *Proceedings 16th AIAA International Communications Satellite Systems Conference (ICSSC '96)*, Washington, DC, USA, pp. 768–778, Feb. 1996.
- [RM97] J. Restrepo and G. Maral, "Guaranteed handover (HO) service in a non-geo constellation with "satellite-fixed cell" (SFC) systems," in *Proceedings 5th International Mobile Satellite Conference (IMSC '97)*, Pasadena, California, USA, pp. 19–24, June 1997.
- [Ros95] K. W. Ross, *Multiservice Loss Models for Broadband Telecommunication Networks*. Series in Telecommunication Networks & Computer Systems, Berlin, Heidelberg, New York: Springer, 1995.
- [RW88] H. Rauch and T. Winarske, "Neural networks for routing communication traffic," *IEEE Control System Magazine*, pp. 26–31, Apr. 1988.
- [Sch87] M. Schwartz, *Telecommunication Networks: Protocols, Modeling and Analysis*. Series in Electrical and Computer Engineering, Reading, MA: Addison-Wesley, 1987.

- [Sch95] J. Schindall, "Concept and implementation of the Globalstar mobile satellite system," in *Proceedings 4th International Mobile Satellite Conference (IMSC '95)*, Ottawa, Canada, pp. A11–A16, June 1995.
- [Sch96] M. Schwartz, *Broadband Integrated Networks*. Upper Saddle River, NJ: Prentice-Hall, 1996.
- [SH01] R. E. Sheriff and Y. F. Hu, *Mobile Satellite Communication Networks*. Chichester: John Wiley & Sons, 2001.
- [Shi79] D. R. Shier, "On algorithms for finding the k shortest paths in a network," *Networks*, vol. 9, pp. 195–214, 1979.
- [Sie95] R. Siebenhaar, "Multiservice call blocking approximations for virtual path based ATM networks with CBR and VBR traffic," in *Proceedings IEEE INFOCOM '95*, Boston, Massachusetts, USA, pp. 321–329, Apr. 1995.
- [Sie96] R. Siebenhaar, *Verkehrslenkung und Kapazitätsanpassung in ATM-Netzen mit virtuellen Pfaden*. PhD thesis, Munich Technical University, Institute of Communication Networks, Munich, Germany, 1996. Herbert Utz Verlag, Munich, Germany, 1996. GERMAN LANGUAGE.
- [Sky97] SkyBridge L.L.C., "Application for authority to launch and operate the SkyBridge system." FCC filing, Washington, D.C., USA, Feb. 1997.
- [SP97] K. Smith and M. Palaniswami, "Static and dynamic channel assignment using neural networks," *IEEE Journal on Selected Areas in Communications*, vol. 15, pp. 238–249, Feb. 1997.
- [Sta98] W. Stallings, *High-Speed Networks: TCP/IP and ATM Design Principles*. Upper Saddle River, NJ: Prentice-Hall, 1998.
- [STTE94] A. Sammut, P. Taaghoul, R. Tafazolli, and B. G. Evans, "Multipath by satellite diversity for CDMA based S-PCN," in *Integrated Space/Terrestrial Mobile Networks, Temporary Documents of European COST 227 Action, 227TD(94)036*, Aveiro, Portugal, Sept. 1994. COST 227 Management Committee.
- [Stu95] M. A. Sturza, "Architecture of the TELEDESIC satellite system," in *Proceedings 4th International Mobile Satellite Conference (IMSC '95)*, Ottawa, Canada, pp. 212–218, June 1995.
- [Tan96] A. S. Tanenbaum, *Computer Networks*. London: Prentice-Hall International, 3rd ed., 1996.
- [TGR91] P. Tran-Gia and O. Rose, "Structure and performance of neural nets in broadband system admission control." Research Report No. 37, University of Würzburg, Institute of Computer Science, Dec. 1991.
- [TIA98] TIA/EIA, "Satellite ATM networks: architecture and guidelines," *TIA/EIA Telecommunications Systems Bulletin (TSB)*, 1998.

- [Vio95] M. D. Violet, "The development and application of a cost per minute metric for the evaluation of mobile satellite systems in a limited-growth voice communications market," Master's thesis, Massachusetts Institute of Technology, Cambridge, MA, USA, Sept. 1995.
- [Wal70] J. G. Walker, "Circular orbit patterns providing continuous whole earth coverage," Tech. Rep. 70211 (UDC 629.195:521.6), Royal Aircraft Establishment, UK, Nov. 1970.
- [Wal77] J. G. Walker, "Continuous whole-earth coverage by circular-orbit satellite patterns," Tech. Rep. 77044 (UDC 629.195:527), Royal Aircraft Establishment, UK, Mar. 1977.
- [Wau98] F. Wauquiez, "Capacity dimensioning of intersatellite link networks in broadband LEO satellite systems," Master's thesis, SUPAERO Toulouse / DLR Oberpfaffenhofen, Toulouse, France / Wessling, Germany, Aug. 1998.
- [WBE97] M. Werner, G. Berndl, and B. Edmaier, "Performance of optimized routing in LEO intersatellite link networks," in *Proceedings 47th IEEE Vehicular Technology Conference (VTC '97)*, Phoenix, Arizona, USA, pp. 246–250, May 1997.
- [WBL95a] M. Werner, H. Bischl, and E. Lutz, "Handover and satellite diversity in personal satellite communications systems," in *Proceedings European Personal and Mobile Communications Conference (EPMCC '95)*, Bologna, Italy, pp. 140–145, Nov. 1995.
- [WBL95b] M. Werner, H. Bischl, and E. Lutz, "Mobile user environment and satellite diversity for NGSO S-PCNs," in *Proceedings 4th International Mobile Satellite Conference (IMSC '95)*, Ottawa, Canada, pp. 476–481, June 1995.
- [WDV<sup>+</sup>97] M. Werner, C. Delucchi, H.-J. Vögel, G. Maral, and J.-J. De Ridder, "ATM-based routing in LEO/MEO satellite networks with intersatellite links," *IEEE Journal on Selected Areas in Communications*, vol. 15, pp. 69–82, Jan. 1997.
- [Wer95] M. Werner, "Analysis of system connectivity and traffic capacity requirements for LEO/MEO S-PCNs," in *Mobile and Personal Communications, Proceedings 2nd Joint COST 227/231 Workshop*, E. Del Re (ed.), Florence, Italy, pp. 183–204, Apr. 1995. Elsevier.
- [Wer96] M. Werner, "ATM concepts for satellite personal communication networks," in *ATM, Networks and LANs, Proceedings European Conference on Networks & Optical Communications (NOC '96), Part II*, D. W. Faulkner and A. L. Harmer (eds.), Heidelberg, Germany, pp. 247–254, June 1996. IOS Press. Invited Keynote Paper.
- [Wer97] M. Werner, "A dynamic routing concept for ATM-based satellite personal communication networks," *IEEE Journal on Selected Areas in Communications*, vol. 15, pp. 1636–1648, Oct. 1997.
- [Wer00] M. Werner, "Multi-step capacity dimensioning of dynamic intersatellite link topologies in low earth orbit satellite systems," in *ITC Specialist Seminar on Mobile Systems and Mobility, Proceedings 12th ITC Specialist Seminar (ITC-SS-12)*, P. J. Emstad (ed.), Lillehammer, Norway, pp. 225–236, Mar. 2000. Norwegian University of Science and Technology, Trondheim, Norway.

- [WF99] M. Werner and J. Frings, "Network design of the intersatellite link segment in broadband LEO satellite systems," in *Multimedia Satellite Networks*, hot topic session at the *16th International Teletraffic Congress (ITC-16)*, Edinburgh, UK, 12 pages, June 1999.
- [WFWM01] M. Werner, J. Frings, F. Wauquiez, and G. Maral, "Topological design, routing and capacity dimensioning for ISL networks in broadband LEO satellite systems," *International Journal of Satellite Communications*, vol. 19, pp. 499–527, Nov./Dec. 2001.
- [WJLB95] M. Werner, A. Jahn, E. Lutz, and A. Böttcher, "Analysis of system parameters for LEO/ICO-satellite communication networks," *IEEE Journal on Selected Areas in Communications*, vol. 13, pp. 371–381, Feb. 1995.
- [WKM97] M. Werner, O. Kroner, and G. Maral, "Analysis of intersatellite links load in a near polar LEO satellite constellation," in *Proceedings 5th International Mobile Satellite Conference (IMSC '97)*, Pasadena, California, USA, pp. 289–294, June 1997.
- [WM97] M. Werner and G. Maral, "Traffic flows and dynamic routing in LEO intersatellite link networks," in *Proceedings 5th International Mobile Satellite Conference (IMSC '97)*, Pasadena, California, USA, pp. 283–288, June 1997.
- [WMMH98] M. Werner, C. Mayer, G. Maral, and M. Holzbock, "A neural network approach to distributed adaptive routing of LEO intersatellite link traffic," in *Proceedings 48th IEEE Vehicular Technology Conference (VTC '98)*, Ottawa, Ontario, Canada, pp. 1498–1502, May 1998.
- [Woo01] L. Wood, *Internetworking with Satellite Constellations*. PhD thesis, University of Surrey, Centre for Communication Systems Research, Guildford, UK, June 2001.
- [WPE01] L. Wood, G. Pavlou, and B. G. Evans, "Managing diversity with handover to provide classes of service in satellite constellation networks," in *Proceedings 19th AIAA International Communications Satellite Systems Conference (ICSSC '01)*, Toulouse, France, 10 pages (paper No. 194), Apr. 2001.
- [WR99] M. Werner and P. Révillon, "Optimization issues in capacity dimensioning of LEO intersatellite link networks," in *Proceedings 5th European Conference on Satellite Communications (ECSC 5)*, Toulouse, France, 8 pages (CD-ROM), Nov. 1999.
- [WWFM99] M. Werner, F. Wauquiez, J. Frings, and G. Maral, "Capacity dimensioning of ISL networks in broadband LEO satellite systems," in *Proceedings 6th International Mobile Satellite Conference (IMSC '99)*, Ottawa, Ontario, Canada, pp. 334–341, June 1999.
- [YHS97] S. A. Youssef, I. W. Habib, and T. N. Saadawi, "A neurocomputing controller for bandwidth allocation in ATM networks," *IEEE Journal on Selected Areas in Communications*, vol. 15, pp. 191–199, Feb. 1997.
- [Zel94] A. Zell, *Simulation Neuronaler Netze*. Bonn, Paris, Reading (MA): Addison-Wesley, 1994. GERMAN LANGUAGE.
- [Zel95] A. Zell, "SNNS V4.0." User Manual, June 1995.

- [Zha93] C. Zhang, "Optimal traffic routing using self-organization principle," in *Proceedings of the International Workshop on Applications of Neural Networks to Telecommunications*, J. Alspector (ed.), pp. 225–231, 1993. Lawrence Earlbaum Associates.