Resource and Network Management Framework for Large-Scale Satellite Communication Systems

February 2020

Yuma Abe

A Thesis for the Degree of Ph.D. in Engineering

Resource and Network Management Framework for Large-Scale Satellite Communication Systems

February 2020

Graduate School of Science and Technology
Keio University

Yuma Abe

Thesis Abstract
Resource and Network Management Framework for Large-Scale Satellite Communication Systems

Satellite communication (SATCOM) systems have been required in a wide range of applications, such as mobile communications for aircraft and ships and emergency communications in a disaster. Thus, the demand for SATCOM systems has increased in recent years. Furthermore, several companies are planning to launch a large number of communication satellites. In such large-scale SATCOM systems, time-variability within system components, such as the amount of communication requests from users, the number of available satellites, and their resources, must be efficiently managed to maximize performance. Therefore, it is necessary to optimize the resource allocation and network structure of SATCOM systems to continuously meet the time-varying requests. In this thesis, the author proposes improved resource and network management frameworks for large-scale SATCOM systems.

In Chapter 1, the author summarizes the applications of SATCOM systems and recent research trends, and also describes the contributions of this thesis. The purpose of this thesis is to propose resource and network management frameworks that reduce the operations cost of large-scale SATCOM systems, including a large number of system components, while continuously accommodating the time-varying communication requests from users.

In Chapter 2, previous work related to this thesis is introduced. Basic methodologies of SATCOM systems, model predictive control, and sparse optimization are utilized in the following chapters.

In Chapter 3, the author proposes a frequency allocation method featuring a function that can flexibly change the frequency allocation of satellite beams. The proposed method is based on model predictive control and sparse optimization, thus enabling the reduction of bandwidth loss and the number of control actions. In numerical examples, the proposed method was applied to time-varying requests of aircraft and the effectiveness of the proposed method was verified.

In Chapter 4, the author proposes a network design method for SATCOM systems with a large number of communication satellites, user terminals, and gateway stations. The proposed method manages a large number of heterogeneous system components in a unified manner and enables various kinds of operations by setting a cost function according to the management strategy. In the numerical examples, the author investigated the basic performance of the proposed method and analyzed the relationship between the increase in the number of satellites and the performance to verify that an efficient network could be obtained when using the proposed method.

In Chapter 5, the author proposes a chance-constrained model predictive control approach for managing stochastic and time-varying communication requests. The proposed method is an extension of the proposed frameworks in Chapters 4 and 5 and enables the system to cope with stochastic requests by adding a stochastic constraint about the bandwidth loss. In numerical examples, the author verified that an efficient resource allocation and network structure was obtained according to a user-specified performance even when the communication requests changed stochastically.

In Chapter 6, this thesis is concluded and future work in the field is discussed.

This thesis is dedicated to my parents, Yoshinori and Taiko.

Acknowledgements

This thesis summarizes my research contributions, which have been carried out in the doctoral course of the Graduate School of Science and Technology, Keio University.

First, I would like to express the deepest appreciation to my supervisor, Prof. Shuichi Adachi, who has continuously supported me since my undergraduate and master's courses. He has given me continuous guidance and always motivated me throughout my doctoral studies. In addition, he has provided me with a lot of advice to improve my presentation and writing skills.

I would like to thank Prof. Takanori Uchiyama, Prof. Hiromitsu Omori, and Prof. Yukitoshi Sanada for serving as my doctoral thesis committee. They have given me many constructive comments and suggestions to improve my thesis.

I would like to thank members of the National Institute of Information and Communications Technology (NICT) for their kind help and support. In particular, Dr. Hiroyuki Tsuji and Dr. Amane Miura have always supported my research and given me helpful guidance. Dr. Morio Toyoshima, Director of Space Communications Laboratory, Dr. Kiyoshi Hamaguchi, Director General of Wireless Networks Research Center, and Dr. Naoto Kadowaki, Vice President of NICT supported me to start the doctoral course and have always encouraged me.

Researchers and engineers from various other institutions have also helped me. First, I would like to thank Dr. Masaki Ogura from Osaka University for fruitful discussions. Without his advice and help, my research would not have been possible. His feedback on academic writing was also very helpful to me. I would also like to thank Dr. Takahiro Kawaguchi from the Tokyo Institute of Technology and Dr. Kazumune Hashimoto from Osaka University for the discussion of the problem formulation in Chapter 3. I give additional thanks to Ms. Ayako Matsumoto from ANA Holdings Inc. for her valuable guidance on the analysis of aircraft data in Chapter 3. Finally, I would like to thank Mr. Hideaki Nakao from

the University of Michigan for his productive comments on the problem formulation for the network optimization problem in Chapter 4.

I would like to give a special thanks to all my friends, who have supported me and encouraged me in my research. Moreover, I would like to thank the members of Adachi Lab for their direct and indirect support during my doctoral course.

Last but not least, my deep and sincere gratitude to my parents Yoshinori and Taiko for always supporting and encouraging me. Without your help, my research would not have proceeded. *Thank you as always!*

February 2020

Yuma Abe

Contents

List of Figures

List of Tables

List of Symbols

c	The speed of light: $c = 299,792,458$ m/s
k_{B}	Boltzmann constant: $k_{\mathrm{B}} = 1.380649 \times 10^{-23}$ J/K
$\mathcal{N}(\mu, \sigma^2)$	A set of random variables with mean μ and variance σ^2
$\mathbb{R}$	A set of real numbers
$\mathbb{R}^n$	A set of n-dimensional real vectors
$\mathbb{R}^{n \times m}$	A set of $n \times m$ real matrices
$\{0,1\}^{n \times m}$	A set of $n \times m$ binary matrices
$\mathrm{diag}(x_1, \ldots, x_n)$	A square matrix whose entries along the diagonal are $x_1, \ldots, x_n$
$\mathbf{0}_n$	n-dimensional vectors whose elements are all 0
$\mathbf{1}_n$	n-dimensional vectors whose elements are all 1
$\boldsymbol{O}_n$	$n \times n$ zero matrix
$\boldsymbol{I}_n$	$n \times n$ identity matrix
$\max\{\boldsymbol{x}, \boldsymbol{y}\}$	A function $\mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}^n$: An elementwise max function retrieving the vector with the largest values among pairs of elements for vectors $\boldsymbol{x} \in \mathbb{R}^n$ and $\boldsymbol{y} \in \mathbb{R}^n$
$\|\boldsymbol{x}\|_2$	ℓ_2-norm of a vector $\boldsymbol{x} \in \mathbb{R}^n$, defined as $\|\boldsymbol{x}\|_2 = \sqrt{\sum_{i=1}^{n} x_i^2}$
$\|\boldsymbol{x}\|_1$	ℓ_1-norm of a vector $\boldsymbol{x} \in \mathbb{R}^n$, defined as $\|\boldsymbol{x}\|_1 = \sum_{i=1}^{n} \lvert x_i \rvert$
$\|\boldsymbol{x}\|_0$	ℓ_0-norm of a vector $\boldsymbol{x} \in \mathbb{R}^n$, defined as $\|\boldsymbol{x}\|_0 = \sum_{i=1}^{n} \delta(x_i)$, where $\delta(x_i) = \begin{cases} 1, & x_i \neq 0, \\ 0, & x_i = 0, \end{cases}$
$\bigcap$	A logical conjunction
$\mathrm{erf}(x)$	The error function, defined as $\mathrm{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-t^2} \, \mathrm{d}t$

List of Abbreviations

3GPP	Third Generation Partnership Project
5G	Fifth-generation mobile communication system
5GPPP	5G Infrastructure Public Private Partnership
A/D	Analog-to-digital
AESA	Active electronically scanned array
ARIMA	Auto-regressive integrated moving-average
ARTEMIS	Advanced Relay and Technology Mission
ARTES	Advanced Research in Telecommunications Systems
ASV	Autonomous surface vehicle
AUV	Autonomous underwater vehicle
bps	bit per second
BSS	Broadcasting satellite service
CARATS	Collaborative Actions for Renovation of Air Traffic Systems
CDF	Cumulative distribution function
D/A	Digital-to-analog
DBF	Digital beamformer
DEMUX	Demultiplexer
DNC	Down-converter
EDRS	European Data Relay Satellite System
EIRP	Equivalent isotropically radiated power
eMBB	enhanced Mobile Broadband
ESA	European Space Agency
ETS	Engineering Test Satellite
EU	European Union
FL	Feeder link
FSS	Fixed satellite service
GEO	Geostationary earth orbit
gNB	Next-generation Node B
GW	Gateway station
HAPS	High-altitude pseudo-satellite
HF	High frequency
HTS	High-throughput satellite
IEEE	Institute of Electrical and Electronic Engineers
IF	Intermediate frequency

IoT	Internet of Things
IP	Internet Protocol
ISL	Inter-satellite link
ITU-R	International Telecommunication Union Radiocommunication Sector
LASSO	Least absolute shrinkage and selection operator
LEO	Low earth orbit
LHCP	Left-hand circularly polarized
LNA	Low noise amplifier
M2M	Machine-to-machine
MEO	Medium earth orbit
mMTC	massive Machine-Type Communications
MPC	Model predictive control
MSS	Mobile satellite service
MUX	Multiplexer
NCA	Number of control actions
NFV	Network functions virtualization
NGEO	Non-Geostationary earth orbit
OGS	Optical ground station
OICETS	Optical Inter-Orbit Communications Engineering Test Satellite
QKD	Quantum key distribution
QoS	Quality of service
ReLU	Rectified linear unit
RF	Radio frequency
RHCP	Right-hand circularly polarized
RMP	Rank minimization problem
RMSE	Root-mean-square error
RNMC	Resource and network management center
RNSS	Radio navigation satellite service
SAT	Communications satellite
SATCOM	Satellite communications
SDN	Software-defined network
SOC	Satellite operations center
TWTA	Traveling wave tube amplifier
UAV	Unmanned aerial vehicle
UHF	Ultra-high frequency
UL	User link
UPC	Up-converter
URLLC	Ultra-Reliable and Low Latency Communications
UT	User terminal
VHF	Very-high frequency

Chapter 1

Introduction

1.1 Introduction of Satellite Communication Systems

Satellite communications (SATCOM) have been widely used all over the world because of multiple desirable features such as wide-area coverage, broad spectrum, and independence from the terrestrial infrastructure [1]. As a result of these characteristics, there has been a gradual increase in the demand for SATCOM in recent years. The total capacity demand for high-throughput SATCOM systems is expected to increase further at an average rate of approximately 32% from 2014 through 2023 [2]. It is estimated that SATCOM systems will be required to achieve Tbps-class throughput by the early 2020s in order to meet the demand for their services [3].

Figure 1.1 shows a typical SATCOM system, which is focused on in this thesis. To describe the SATCOM system, the following terms are used.

- **Communications satellites (SATs)** are used to relay signals from/to ground stations (also called earth stations) via wireless satellite links. Satellites have a receive/transmit antenna and a transponder that functions as both a transmitter and responder. After satellites receive the signal from ground stations via the receive antenna, the signal is processed by the transponder and amplified. Then, the signal is radiated and transmitted to other ground stations by the transmit antenna.

  To communicate between satellites and the ground stations, radio waves in radio frequency (RF) bands such as the C-band (approximately 4–8 GHz), the Ku-band (approximately 12–18 GHz), and the Ka-band (approximately 26–40 GHz) are commonly

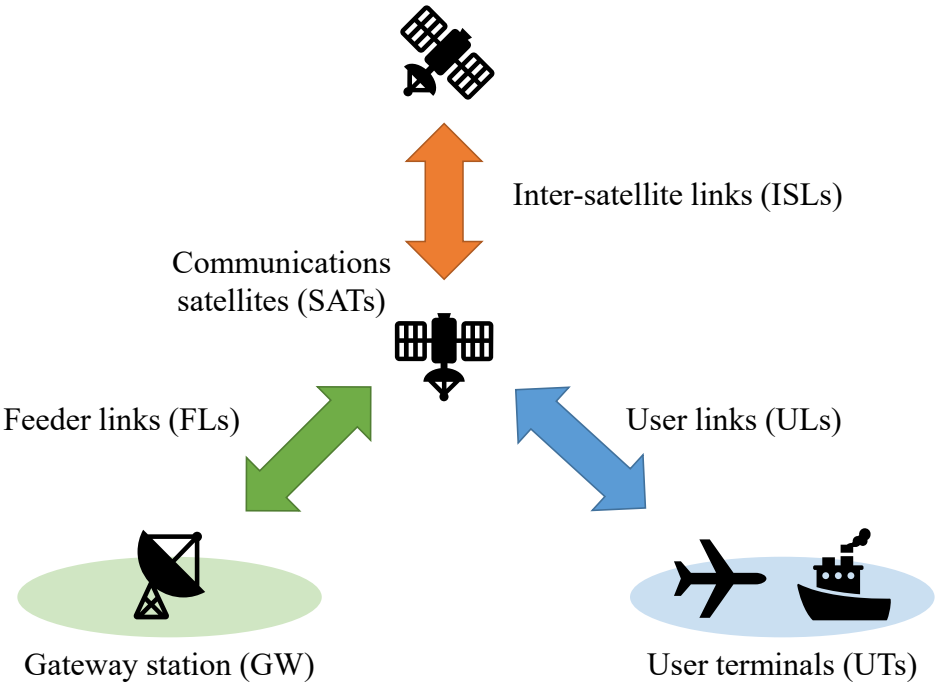


Fig. 1.1: A typical SATCOM system and SATCOM links.

used [4, 5]. A satellite radiates the radio waves, called *satellite beams*, and illuminates particular areas. Frequency resources such as bandwidth are allocated to each beam, and ground stations that exist in the satellite beam can communicate with the satellite by utilizing the allocated resources. The radio waves are attenuated due to factors such as the distance between satellites and the ground and weather conditions such as rainfall. Thus, compensation methods are applied to reduce the influence of attenuation (e.g., adaptive coding and modulation) [6].

Satellite orbits related to the topic of this thesis are roughly divided into the following two types:

- **Geostationary earth orbit (GEO)** is an orbit at an altitude of approximately 35,786 km above the earth's equator. Satellites placed at a GEO appear to be stationary from the earth.

- **Non-geostationary earth orbit (NGEO)** is a general term describing orbits at an altitude of lower than that of a GEO. Low earth orbits (LEOs, approximately 2,000 km or less) and medium earth orbits (MEOs, approximately from 2,000 km to 35,786 km) are

examples of NGEOs. Satellites placed at NGEOs revolve around the earth periodically on prescribed orbits.

Ground stations related to this thesis are roughly divided into the following two types:

- **User terminals (UTs)** are ground stations such as aircraft and ships that require SATCOM links. UTs with a small aperture antenna are called very small aperture terminals (VSATs).
- **Gateway stations (GWs)** are ground stations connected to the ground network such as the Internet. GWs typically have a larger antenna and a larger capacity than that of UTs, and the data from/to UTs are aggregated in the GWs via satellites.

There are several types of SATCOM links composed of satellites, UTs, and GWs. This thesis mainly focuses on the following three:

- **User links (ULs)** are bidirectional communication links between satellites and UTs. In the user links, UTs receive/transmit signals from/to GWs via satellites.
- **Feeder links (FLs)** are bidirectional communication links between satellites and GWs. In the feeder links, GWs receive/transmit signals from/to UTs via satellites.
- **Inter-satellite links (ISLs)** are bidirectional communication links between satellites [7, 8]. By routing communication signals via inter-satellite links, various kinds of operations can be realized [9, 10] (e.g., a data relay system among multiple satellites). For example, the communication time of NGEO satellites with a certain ground station is limited because NGEO satellites revolve around the earth. Thus, by routing the signals from an NGEO satellite to a GEO satellite via an inter-satellite link, NGEO satellites can communicate with the ground station indirectly.

A SATCOM network composed of user links, feeder links, and inter-satellite links works as a data relay system between UTs and GWs. UTs request SATCOM links, and satellites provide the links for the UTs. After the SATCOM network is constructed, communication signals between the UTs and GWs are relayed by the satellites.

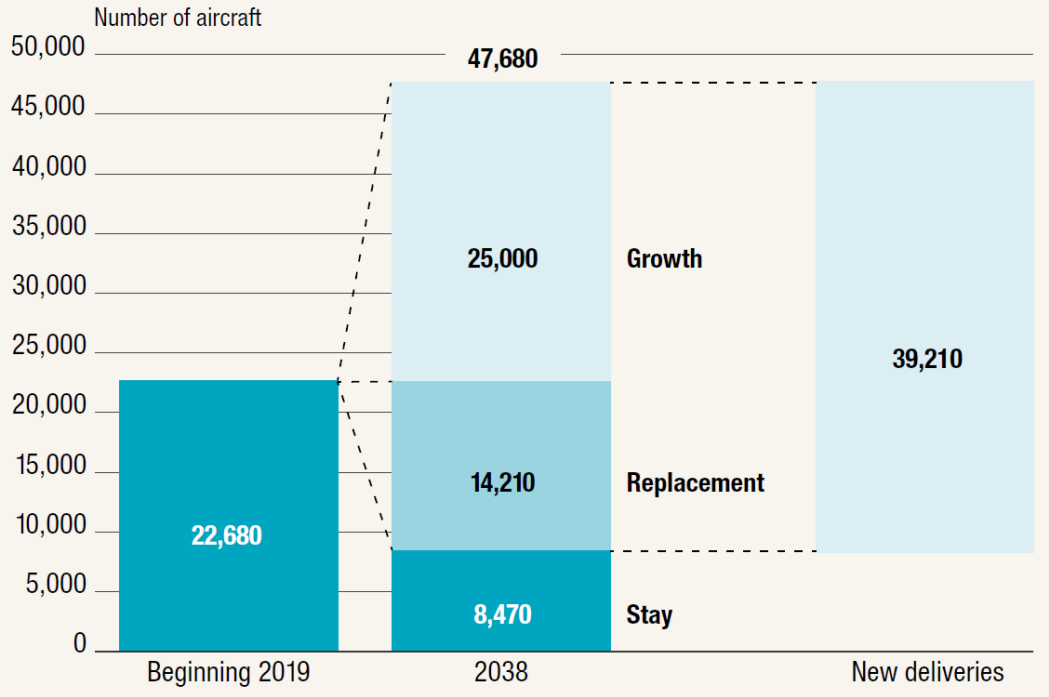


Fig. 1.2: The number of commercial aircraft from 2019 to 2038. The number is expected to have a twofold increase over the next 20 years. Reproduced from [11].

1.2 Applications of Satellite Communications

In this section, multiple applications of SATCOM systems are introduced.

1.2.1 Aeronautical Communications

The use of SATCOM systems enables in-flight connectivity, such as Wi-Fi within aircraft cabins for passenger use. The number of commercial aircraft is expected to have a twofold increase over the next 20 years as shown in Fig. 1.2 [11]. Moreover, current estimates suggest that over 50% of aircraft will be equipped with in-flight Wi-Fi by 2025 [12]. The notable increase in this specific application of SATCOM systems highlights the importance of developing the underlying technology [13, 14]. By utilizing SATCOM systems to cover wide areas in air, a large number of aircraft including drones, unmanned aerial vehicle (UAVs), and high-altitude pseudo-satellites (HAPSs) can be connected [15].

To provide seamless high-speed communications for fast-moving aircraft, effective handover schemes are required [16, 17]. The aircraft have to frequently change a connecting satellite or a connecting satellite beam in a single flight because aircraft typically move

across a wide area where that one satellite or one satellite beam cannot cover. Thus, it is necessary to perform multiple satellite handovers or multiple beam handovers. Additionally, the handover schemes should be performed with low latency so that communications for passengers are not disconnected due to the handovers.

1.2.2 Maritime Communications

Maritime and aeronautical communications are two of the most important applications of SATCOM. In particular, maritime regions have been interpreted as one of the *digital divide* regions because ships have lacked broadband communication services for a long time.

In recent years, marine resource exploration has been actively conducted with unmanned and autonomous vessels such as autonomous surface vehicles (ASVs) with autonomous underwater vehicles (AUVs) [18]. ASVs are equipped with VSATs to utilize the SATCOM links that enable efficient exploration. By using the SATCOM links, the ASVs can be remotely controlled by a nearby mother ship or the main office on the ground. Furthermore, exploration data obtained by the AUVs are sent to the ASVs, and then the data are sent to the mother ship or the main office via the SATCOM links. In this way, the SATCOM increases the efficiency of the exploration. Other use cases and research challenges for the connectivity of autonomous vessels were addressed in [19].

The authors of [20] proposed an architecture with a mesh network formed by multiple ships and satellites to realize a high-speed and low-cost system for maritime communication. Furthermore, to support effective communication in maritime regions, the authors of [21] investigated the propagation environment and proposed channel models for maritime communications. The latter study modeled unique features of the maritime propagation environment, including both the air-to-sea and near-sea-surface channel links.

1.2.3 Emergency Communications

In emergency situations that disrupt terrestrial networks, such as those caused by earthquakes, SATCOM systems are indispensable [22–24]. SATCOM can provide communication links to regions where the terrestrial network is disabled by utilizing mobile stations with VSATs, which can form a SATCOM network. The SATCOM network contributes to

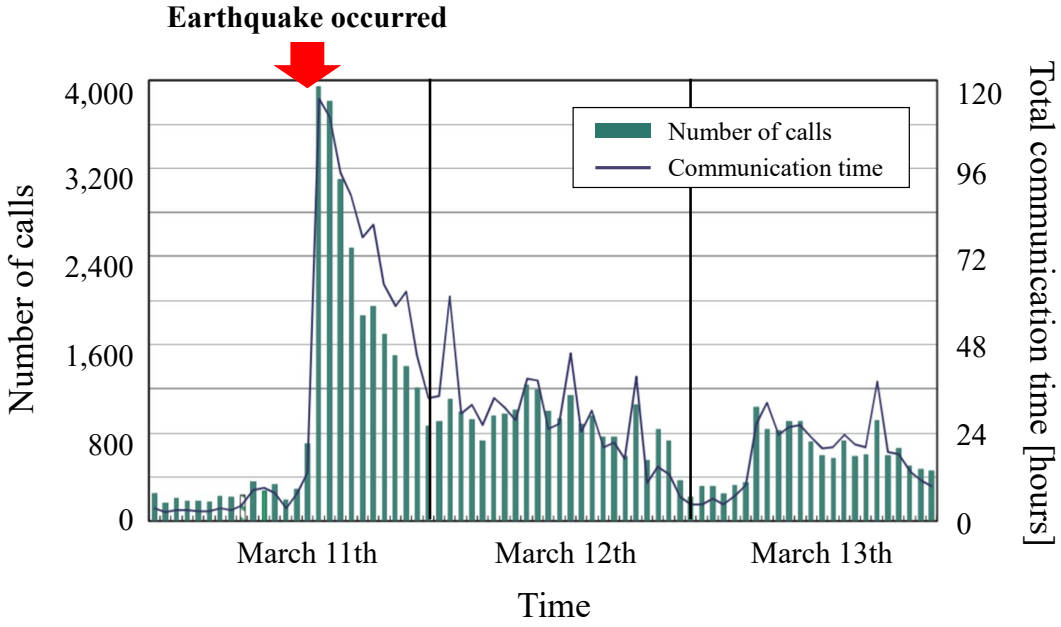


Fig. 1.3: Time dependence of the communication request at the Great East Japan Earthquake (March 11th-13th, 2011). This figure shows that the communication traffic increased rapidly after the earthquake. Adapted from [26].

understanding the emergency situation by sending messages, images, and videos of the disaster area through the satellite links. The government can utilize the information to develop a plan for rapid recovery of the disaster and can provide government communication services (e.g., police corps, fire departments, and ambulance services) [25]. Furthermore, the mobile stations can also work as cellular backhaul stations for the mobile communication, which the people who suffered from the disaster can use for personal communications.

In Japan, a large number of earthquakes occur every year. These earthquakes drive intense spikes in activity in Japan's communication networks. For example, the communication requests increased approximately tenfold after the Great East Japan Earthquake, which occurred on March 11th, 2011. This is shown in Fig. 1.3 [26]. This demonstrates that SATCOM systems with a larger capacity and a function to cope with the time-varying communication requests are required to effectively respond to emergency situations.

1.2.4 IoT/M2M Communications

The Internet of Things (IoT) and the machine-to-machine (M2M) communications are key technologies in cyber-physical systems, including the fifth-generation mobile communication system (5G) [27]. These terms refer to a large number of devices with sensors and

actuators that will be connected to the Internet. According to [28], "global M2M Internet Protocol (IP) traffic will grow more than sevenfold ... from 3.7 exabyte per month in 2017 (3 percent of global IP traffic) to more than 25 exabyte by 2022." The devices from the IoT/M2M communications are distributed over a wide area, and some of them could exist in an area without a terrestrial network. For this reason, SATCOM systems also play an important role in IoT/M2M communications [29, 30]. By utilizing SATCOM links, data obtained by the sensors can be collected for a wide area and users can access the data via the Internet. Some systems were proposed to provide IoT communication services based on the use of multiple LEO satellites [31].

1.3 Recent Trends in Satellite Communication Research

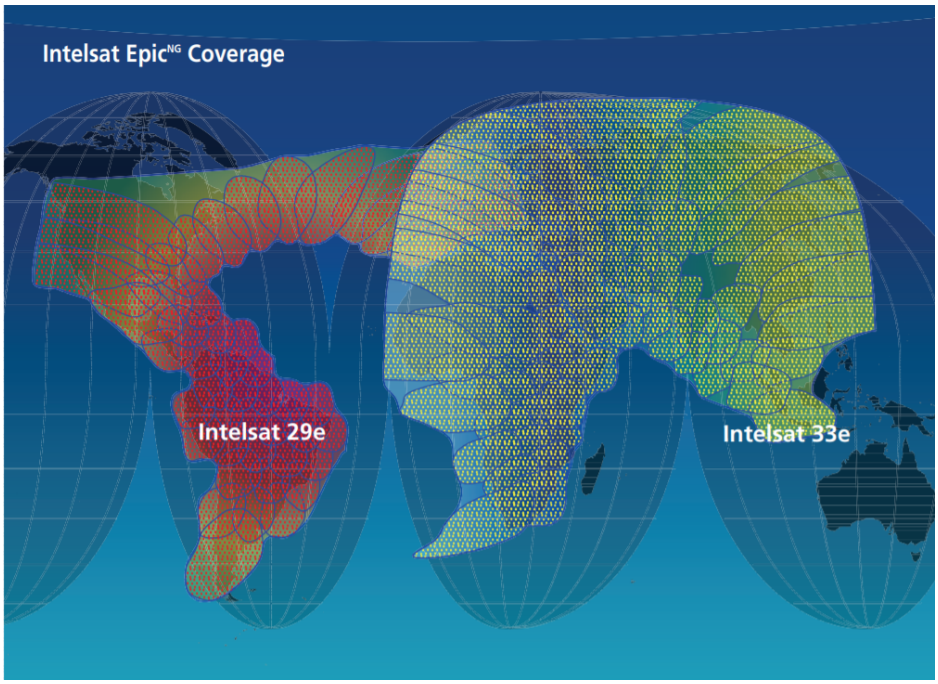
To meet the increasing SATCOM demand for the multiple applications stated in the previous section, larger communication capacity and more efficient frequency utilization are required. This is because available frequency resources, such as the bandwidth of the SATCOM systems, are limited.

In the following subsections, recent SATCOM research trends that aim to increase the communication capacity and improve the frequency utilization efficiency are introduced.

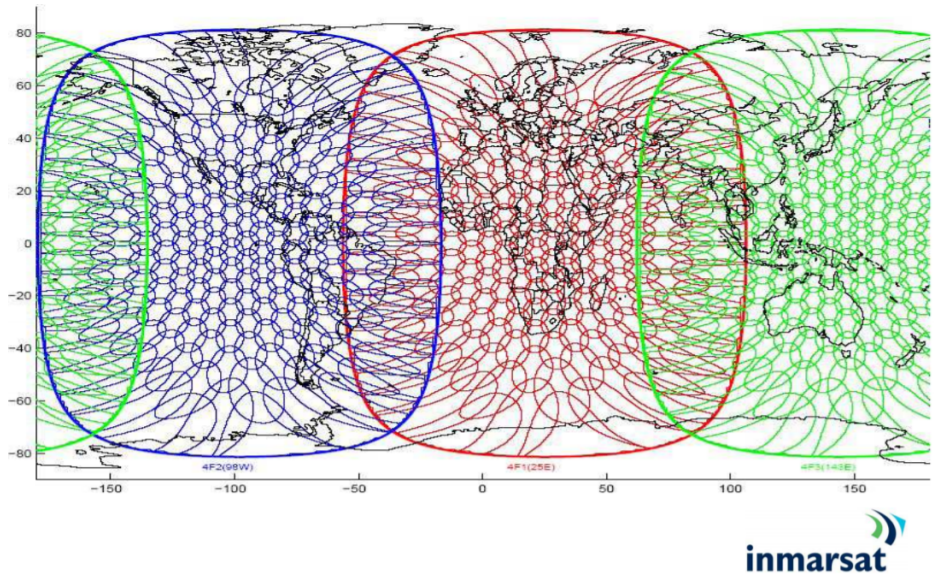
1.3.1 High-Throughput Satellites

Research and development on high-throughput satellites (HTSs) have been actively conducted in recent years [32–34]. The main objective of the HTSs is to reduce cost per bit by increasing the capacity. HTSs usually use the Ka-band, which has a wider bandwidth and can improve the frequency utilization efficiency by using multiple spot beams. To utilize as much of the bandwidth as possible, the frequency band is reused spatially among multiple beams to reduce interference. This is accomplished by separating places that use the same frequency band [35].

Figure 1.4 shows examples of multibeam coverage of Intelsat Epic (29e and 33e) in the Ku-band [36] and Inmarsat I-4 satellites in the Ka-band [37]. These satellites can provide communication services globally because they cover a wide area by using multiple beams.



(a)



(b)

Fig. 1.4: Examples of multibeam coverage. (a) Intelsat 29e and 33e coverage and (b) Inmarsat I-4 coverage. Reproduced from [36] and [37], respectively.

1.3.2 Flexibility

Current SATCOM systems allocate a fixed bandwidth to each beam. Consequently, the maximum communication capacity is also fixed. Therefore, the systems cannot cope with communication requests when the requests vary in time and exceed the maximum capacity. For example, as shown in Fig. 1.3, the communication requests vary in time and rapidly increase after natural disasters such as earthquakes. Furthermore, there is a large difference in the requests between areas that require communication resources according to local variables such as population and the number of mobile terminals. These situations demand introducing flexibility functions in next-generation SATCOM systems.

A frequency flexibility function can flexibly change the frequency bandwidth allocation for each beam. This flexibility can be realized by installing a digital channelizer in satellites as a communication payload, thus enabling the handling of time-varying communication requests by sharing the bandwidth among adjacent beams [38–40].

Beamforming technology is also effective in coping with time-varying and spatially distributed communication requests [41]. This function, which is called area flexibility, is realized by installing a digital beamformer (DBF) in satellites. By using the area flexibility function, the satellites can flexibly change areas and sizes of spot beams. Thus, satellites can target areas with increasing requests and track mobile terminals (such as aircraft and ships) to provide continuous communication services. Reconfigurable antennas such as active electronically scanned array (AESA) antennas have the area flexibility function to steer satellite beams and to achieve efficient handovers [42, 43].

Furthermore, beam hopping technology is gathering attention as one of the flexibility functions [44, 45]. In each time slot of the satellites, several beams out of all beams are selected and the data are transmitted/received by using only the selected beams. This enables a reduction in the power consumption of the satellites.

By managing and controlling the flexibility functions installed in satellites, SATCOM systems enable the utilization of limited satellite resources that can effectively meet a wide variety of communication requests.

1.3.3 Satellite Constellation

Recently, several companies have initiated the construction of global SATCOM networks by launching hundreds of thousands of communication satellites into space [46, 47]. These SATCOM systems are generally called satellite constellation systems. Examples of the satellite constellation systems [48] include:

- Telesat's Ka-band constellation, which includes at least 117 satellites distributed in two sets of orbits.
- OneWeb's Ku- and Ka-band constellation, which is comprised of 720 satellites in 18 circular orbital planes at an altitude of 1,200 km.
- SpaceX's Ku- and Ka-band constellation, which will be made up of 4,425 and more satellites that will be distributed across several sets of orbits.

These constellation systems cover the entire earth by using a large number of NGEO satellites.

The constellation networks will involve collaboration between GEO and NGEO satellites [49]. By cooperating and coordinating multiple types of satellites, SATCOM systems can achieve various communication objectives [4, 50]. To construct an efficient constellation system, spectrum sharing between the GEO and NGEO satellites is required [51, 52].

1.3.4 Small Satellites and CubeSats

Small satellites generally refer to satellites that weigh under 500 kg, while medium satellites weigh between 500–1,000 kg and large satellites weigh over 1,000 kg [53]. Small satellites have been developed vigorously all over the world in recent years [54, 55]. In particular, CubeSats are considered a class of smaller satellites with a standard size called "1U", which represents a volume of 10 cm $\times$ 10 cm $\times$ 10 cm. Based on the 1U dimensions, the sizes of CubeSats can be extended to larger sizes (e.g., 2U, 3U, 6U, and 12U) [56]. The small satellites and CubeSats are developed and launched significantly cheaper than conventional large satellites. Thus, private companies and universities are driven to develop their own satellites [57].

Figure 1.5 shows the number of nanosatellite launches from 1998 to 2023 (those in 2020–2023 are predicted). In this context, nanosatellites are defined as satellites with masses

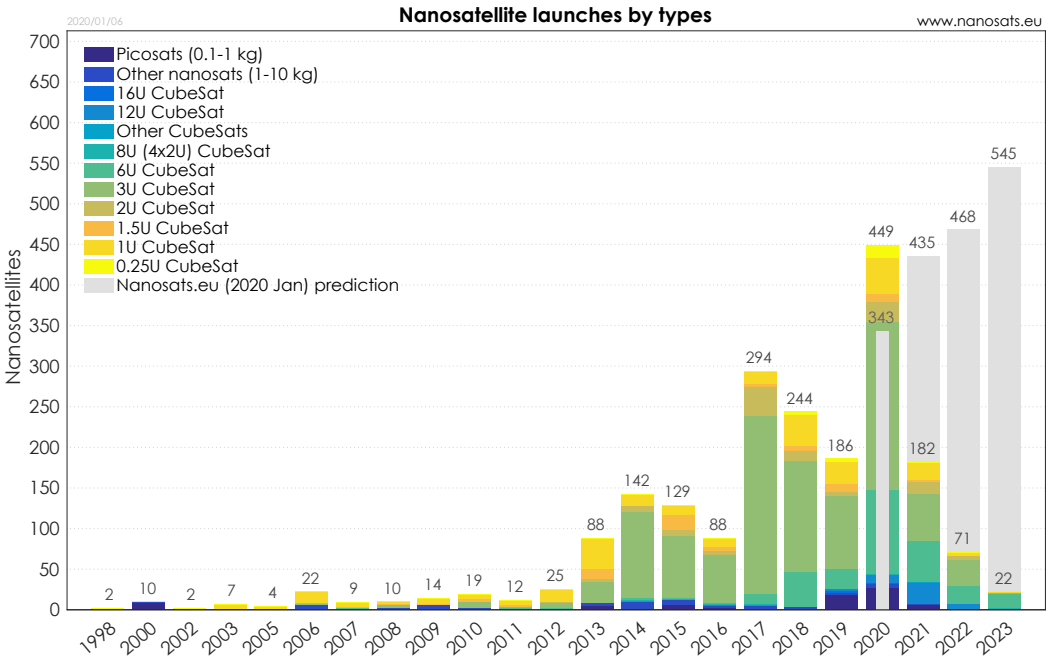


Fig. 1.5: Nanosatellite launches by type. Reproduced from [53].

from 1 kg to 10 kg (CubeSats are included in the nanosatellites). The first nanosatellite was launched in 1998 and the number of launches rapidly increased around 2013–2014. It is predicted that 545 nanosatellites will be launched in 2023 alone.

Compared to the large satellites, the power budget of the small satellites and CubeSats is limited. However, there is a possibility to replace large satellites with a large number of small satellites and CubeSats. Precision formation flying [58, 59] with a large number of satellites enables similar missions to those conducted by large satellites, but at a lower cost.

1.3.5 Space Optical Communications

Optical communications in space [7, 60] are desirable because SATCOM systems use limited frequency domains in the RF-bands, which are conventionally used. This limitation is a result of the recent rapid increase in the frequency of requests from communication and earth observation satellites. Optical communications have multiple advantages over the RF-bands, such as larger bandwidth, a higher data rate, and lower power consumption. They can also make use of the license-free spectrum. Despite these favorable characteristics, there are some obstacles to efficient optical communications in space. The beamwidth of the optical wave, which depends on wavelength, is narrower than the RF beamwidth. Thus, the required

tracking accuracy is very high for both satellites and optical ground stations (OGSs). For example, an LEO satellite requires μrad-class tracking accuracy. Furthermore, the optical wave can be blocked by clouds and almost completely attenuated; thus, site diversity technology must be applied by using a distribution of multiple OGSs [61]. SATCOM systems with site diversity can select an optimal OGS to use by monitoring a weather condition.

Unlike SATCOM links between satellites and the ground, inter-satellite links are not affected by the weather conditions; accordingly, inter-satellite links will be conducted by optical communications. By introducing inter-satellite links, even satellites that are not directly visible from any OGS can establish communication links indirectly from anywhere. Equipped with the inter-satellite links, satellite constellations work more efficiently because the number of OGSs visible to each satellite is no longer limited. In response to these clear advantages, satellite data relay systems using optical links are currently under development.

The space-ground optical bidirectional link was first successfully demonstrated by the Engineering Test Satellite (ETS-VI), which was launched to a GEO altitude above Japan in 1996. Furthermore, the Optical Inter-Orbit Communications Engineering Test Satellite (OICETS) succeeded in establishing an optical link between its LEO and the ground in 2006. The inter-satellite link was first accomplished by the Advanced Relay and Technology Mission (ARTEMIS) satellite (in a GEO) and OICETS (in a LEO) in 2005 [7]. After these demonstrations, optical communication experiments were conducted on the European Data Relay Satellite System (EDRS), which successfully demonstrated 1.8-Gbps optical communication links between Alphasat (in a GEO) and Sentinel-1 (in a LEO) in 2014 [62].

1.3.6 Satellite-Terrestrial Integration

The integration of the SATCOM system and the terrestrial mobile communication system has been considered for a significant amount of time. Due to the fact that some frequency bands are allocated to both systems, the most important component of the integration is spectrum sharing technology that shares the same frequency band between the SATCOM and terrestrial systems [63, 64].

In recent years, the integration of SATCOM systems and 5G have been making progress [65–67]. This endeavor aims to not only increase the total communication capacity of the whole system, but also accommodate various types of communication requests aris-

ing from emerging information technologies such as IoT and M2M communications. Three advantages of 5G networks are: enhanced Mobile Broadband (eMBB), massive Machine-Type Communications (mMTC), and Ultra-Reliable and Low Latency Communications (URLLC). The underlying technologies needed to implement 5G, as well as the satellite-5G integrated system are the network virtualization technologies such as network slicing, software-defined networking (SDN), and network functions virtualization (NFV). Owing to these technologies, 5G networks can provide a flexible IoT/M2M communication platform and can cope with a wide variety of use cases according to the quality of service (QoS) of each use case. In the satellite-5G integrated system, satellites will be added to the 5G network and each satellite will work as a 5G node, which is called next-generation NodeB (gNB) in the 5G literature. The satellite-5G integration enables the SATCOM links to utilize the 5G network features, which include a large capacity and a large number of connections. The 5G network also supports various types of use cases including the IoT/M2M communications. However, compared to the terrestrial 5G network, there is a much higher latency in SATCOM links because of the long propagation distances (especially for GEO satellites). Therefore, the integrated system may be limited to cases that do not require very low latency.

The satellite-5G integration projects are progressing mainly in Europe — for example, the SAT5G project [68] funded by the European Union (EU) and the SATis5 project [69] funded by the European Space Agency (ESA) Advanced Research in Telecommunications Systems (ARTES) program. These projects seek to identify use cases, applications, and technical problems of the satellite-5G integrated system. In addition, standardization of the integration process has been starting at the Third Generation Partnership Project (3GPP) [70] and the 5G Infrastructure Public Private Partnership (5GPPP) [71]. Such projects will be carried out continuously and will be applied not only to the 5G network but also to the subsequent generation of mobile communication.

1.4 Large-Scale Satellite Communication Systems: Characteristics and Difficulties

Due to the applications and recent research trends of the SATCOM systems described in the previous sections, SATCOM systems have been expanded into large-scale ones in recent

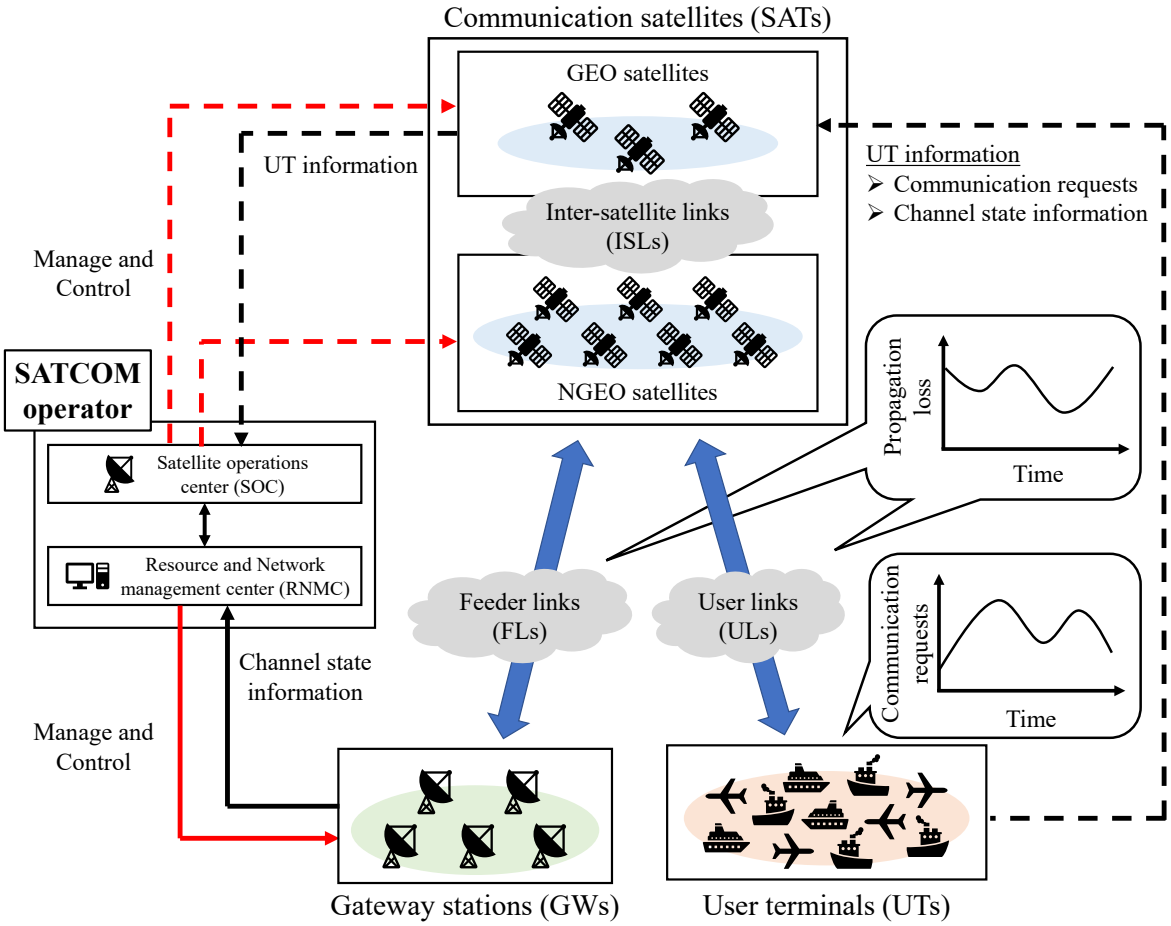


Fig. 1.6: Architecture of large-scale SATCOM systems focused upon in this thesis. Solid and dashed lines represent wired and wireless communication links, respectively. Red lines represent the actions of a SATCOM operator that works as a controller of this system by managing and controlling the satellites and GWs.

decades [48]. This is because SATCOM systems have been required to provide increasingly larger capacities of communication and to meet various novel types of communication needs arising from new technologies such as IoT and M2M [29]. Therefore, several private companies have begun constructing satellite constellation systems that consist of hundreds of thousands of communication satellites [72, 73]. Equipped with such a large number of satellites, large-scale SATCOM systems can cover land, sea, and air globally. Therefore, they can provide broadband communication links for a large number of users.

In this study, the author focuses on the large-scale SATCOM system illustrated in Fig. 1.6. This system includes multiple satellites, multiple UTs such as aircraft and ships that request communication links, and multiple GWs connected to the ground network. A SATCOM operator works as a controller of this system by managing the satellites and GWs and tries to

construct an efficient SATCOM network to provide communication links for a large number of UTs.

In current SATCOM systems, each UT typically connects to only one specific satellite determined in advance (termed a *conventional non-integrated system*), whereas in this large-scale SATCOM system, UTs have several candidate satellites to connect with (termed an *integrated system*). As stated in Section 1.1, the satellites are divided into two different types: GEO satellites and NGEO satellites. GEO satellites have a larger bandwidth and wider beam coverage than the NGEO satellites, whereas the number of satellites that can be placed at a GEO is limited. The cost of developing and launching GEO satellites is also high. On the other hand, NGEO satellites can perform lower latency communication and can be developed at a lower cost than GEO satellites. However, a large number of NGEO satellites must be launched to achieve global coverage [73]. Thus, integrating GEO and NGEO satellites produces a variety of advantages. For example, a large number of UTs can communicate with satellites at all times and wider bandwidths can be used.

To establish an efficient SATCOM network in this system, it is important for the SATCOM operator to appropriately allocate the communication resources to the user links, feeder links, and inter-satellite links. The GEO satellites can work as aggregators of UT information, such as communication requests and channel state information of the user links. By utilizing this function, the SATCOM operator aggregates the UT information via the GEO satellites and the channel state information of the feeder links from the GWs. Based on the aggregated information, the operator determines the resource allocation and network structure in a resource and network management center (RNMC). Then, from a satellite operations center (SOC), the operator sends command signals to change the state of the satellites. Finally, the satellites provide communication links to the UTs and the SATCOM network is established.

Two major difficulties are present in the management of large-scale SATCOM systems. The first one is the time-variability within the system components, such as the number of users, the amount of communication requests, the number of available satellites, and the propagation environments. In current SATCOM systems, resource allocation and network structure are static because only a few satellites and users form the system. Under time-variability conditions, the static resource allocation and network structure can result in net-

work disconnection and high latency. Furthermore, a part of a request may change stochastically due to unknown reasons, such as atmospheric conditions and natural disasters [74, 75]. Therefore, it is important for the SATCOM operator to dynamically change the resource allocation. By using the flexibility functions and the SATCOM network structure, the satellite resources can be efficiently utilized in the presence of hundreds of thousands of system components experiencing time-variability.

The second difficulty is to connect a large number of heterogeneous components in the SATCOM systems. For example, large-scale SATCOM systems contain satellites that have different sizes (e.g., large, small, or nano), orbits (e.g., GEO or NGEO), and frequencies (e.g., Ku-band, Ka-band, or optical) [72]. In the future, the network will be expanded into three dimensions; not only satellites, but also drones, UAVs, aircraft, and HAPSs will work as repeaters of signals. Furthermore, spatially distributed users send qualitatively different communication requests depending on their specific communication needs, such as mobile, emergency, and IoT/M2M communications. Therefore, the operator of the SATCOM systems has to construct a network with a large number of different types of system components.

1.5 Contribution

There is a lack of effective methodologies for the efficient management of large-scale SATCOM systems that address the aforementioned difficulties. Furthermore, SATCOM systems should be managed and controlled automatically because complicated operations are required to manage a large number of satellites and GWs and to cope with a wide variety of communication requests. Therefore, this study addresses how to establish management frameworks for large-scale SATCOM systems.

The main contributions of this thesis are summarized as follows:

- A management framework for the frequency flexibility is proposed in Chapter 3. In this framework:
  - Dynamical models of the SATCOM system utilizing the frequency flexibility are incorporated into the resource allocation problem. By solving the formulated problem, which combines model predictive control (MPC) and sparse optimization, an

optimal solution that utilizes limited bandwidth and copes with time-varying UT requests can be obtained.

  - The effectiveness of the proposed method is verified through numerical experiments. In these numerical experiments, the UT requests are time-varying and one is based on actual flight tracking data from aircraft.

- A management framework for the massive connected networks is proposed in Chapter 4. In this framework:

  - The proposed method obtains resource allocation for UTs and SATCOM network structures simultaneously, allowing the management and control of multiple GEO and NGEO satellites in a unified manner. This method also copes with time-varying UT requests by constructing the flexible SATCOM network.

  - The effectiveness of the proposed method is verified through several types of numerical experiments. These numerical experiments compare the performance of the proposed method in the integrated system with GEO and NGEO satellites to the conventional non-integrated system under time-varying UT requests. Furthermore, the effect of the number of satellites in the system is verified by comparing the performance of the integrated and non-integrated systems.

- A chance-constrained resource and network management strategy that can cope with stochastic and time-varying communication requests is proposed in Chapter 5. This management strategy can be applied to the management frameworks for both the frequency flexibility and massive connected networks and can be included in the problem formulation in the same manner.

The proposed management strategy contributes to the implementation of large-scale SATCOM systems that efficiently allocate resources and the construction of large-scale SATCOM networks connected with a large number of satellites, UTs, and GWs.

1.6 Outline of Thesis

Figure 1.7 displays a large-scale SATCOM system labeled with the chapters of this thesis that the correspond to those topics, and Figure 1.8 shows the configuration of this thesis[*1].

The remainder of this thesis is organized as follows.

Chapter 2: Related Work

In Chapter 2, related work for this thesis is described. First, an overview of SATCOM systems is given. This section includes the description of the configuration of communications satellites, frequency allocation, and link budget analysis. Then, technical methodologies of the MPC [78] and sparse optimization [79–81] are introduced. The MPC is a finite-time optimal control method, which determines control input sequences online while predicting future states and satisfying system constraints. The sparse optimization is one of the optimization methods used to obtain a solution vector with few nonzero elements. These methodologies support the management frameworks proposed in Chapters 3, 4, and 5.

Chapter 3: Management Framework for Frequency Flexibility

In Chapter 3, the author proposes a frequency resource allocation method by utilizing frequency flexibility. This allows the frequency bandwidth allocation to flexibly change for each beam in order to cope with time-varying communication requests in SATCOM systems.

In SATCOM systems, communication requests from UTs in each satellite beam vary temporally. Thus, bandwidth loss, which represents the bandwidth gap required to meet the requests, could occur unless the bandwidth allocation is adjusted according to the requests. To reduce the bandwidth loss, the author focuses on a bandwidth allocation method with frequency flexibility for HTSs. Furthermore, a large amount of power could be consumed when an HTS utilizes the frequency flexibility to cope with large time variations in the requests. Thus, frequent control actions are avoided in the bandwidth allocation when the

[*1] Part of this thesis is based on "Frequency Resource Management Based on Model Predictive Control for Satellite Communications System" Copyright © 2018 IEICE [76] and "Resource and Network Management Framework for a Large-Scale Satellite Communications System" Copyright © 2020 IEICE [77], which appeared in IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences.

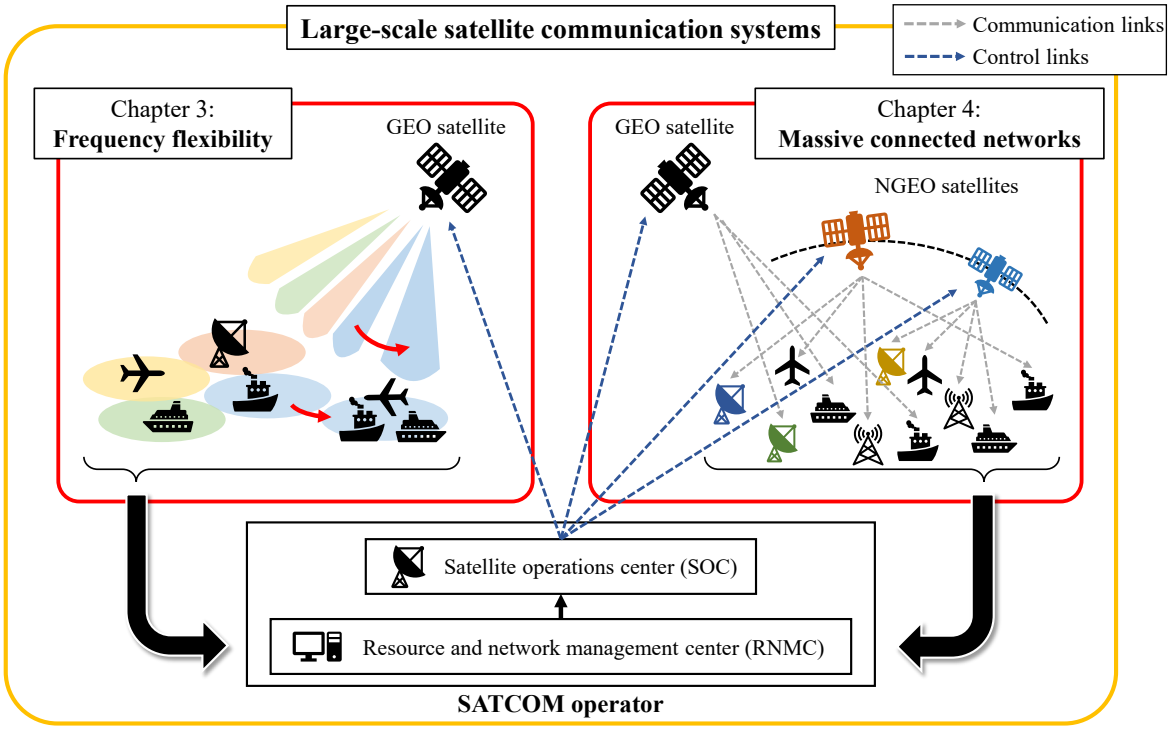


Fig. 1.7: Large-scale SATCOM systems and relationship among the chapters of this thesis.

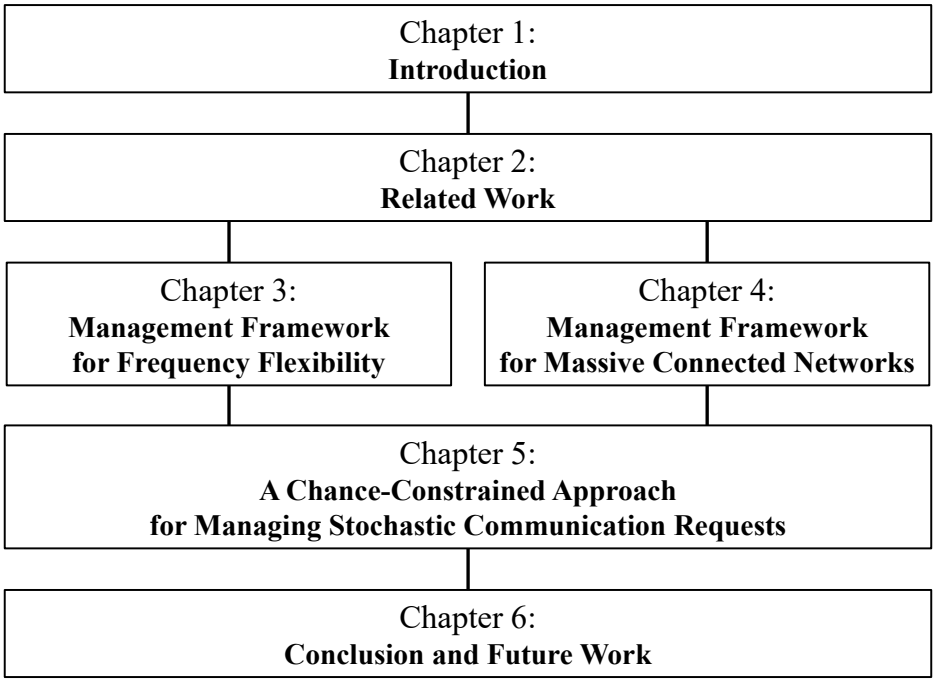


Fig. 1.8: Configuration of this thesis.

frequency flexibility is applied. To this end, the author proposes an optimal bandwidth allocation method to reduce the bandwidth loss and number of control actions (NCA) by combining MPC and sparse optimization, which are introduced in Chapter 2.

Similar studies on resource allocation are reported in [82–84], where static allocation methods are proposed. However, few studies have focused on dynamical resource allocation by using the frequency flexibility function. Moreover, unlike similar approaches, the proposed method effectively handles time-varying communication requests and prevents very frequent control actions required for changing the bandwidth allocation in the digital channelizer. The above is a flexible payload, designed to apply the frequency flexibility of HTSs while preserving their limited power resources.

The optimization in the proposed method is summarized as follows. First, the author defines dynamical models of the digital channelizer in an HTS and also defines communication requests in each satellite beam assuming that the dynamics of the requests can be obtained in advance. Then a cost function and constraints are defined, and an optimal bandwidth allocation problem is formulated. In this problem, realistic satellite constraints such as beam arrangement (avoiding beam interference) are included. The concepts of MPC and sparse optimization are mainly utilized to define the cost function in the optimization problem.

To verify the effectiveness of the proposed method, two types of numerical experiments were conducted. The first case considered a simple time variation model of the communication requests to verify the basic performance of the proposed method. The second case focused on an application of the bandwidth allocation for aircraft. The author analyzed the actual flight tracking data of aircraft and investigated the time variation of the expected communication requests in each beam. The proposed method was applied to the obtained communication requests and the effectiveness of the proposed method using the frequency flexibility was verified in the practical scenario of the aircraft requests.

The results given in this chapter are related to the following paper [76]:

- Y. Abe, H. Tsuji, A. Miura, and S. Adachi, “Frequency Resource Management Based on Model Predictive Control for Satellite Communications System,” *IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences*, vol. E101-A, no. 12, pp. 2434–2445, 2018.

The results are also related to the following peer-reviewed conference paper [85]:

- Y. Abe, H. Tsuji, A. Miura, and S. Adachi, "Frequency Resource Allocation for Satellite Communications System Based on Model Predictive Control and Its Application to Frequency Bandwidth Allocation for Aircraft," Proceedings of the 2018 IEEE Conference on Control Technology and Applications (CCTA2018), pp. 165–170, 2018.

Chapter 4: Management Framework for Massive Connected Networks

In Chapter 4, the author proposes a network optimization method to efficiently allocate the bandwidth resource to each user and to efficiently construct a SATCOM network for large-scale SATCOM systems connected with a large number of system components.

Large-scale SATCOM systems include a large number of satellites, UTs, and GWs. The time-variability within the system components, such as the amount of communication requests from users and the number of available satellites and their resources, is described in Section 1.4. Thus, the SATCOM operator should design a time-varying SATCOM network, which consists of user links, feeder links, and inter-satellite links. Furthermore, frequent satellite handovers, which lead to wasteful power consumption can be expected when designing the time-varying SATCOM network with a large number of components because UTs and GWs change satellites to allow connectivity at all times. Thus, the operator should avoid these handovers in the network design strategy.

Network design problems have been investigated in many areas [86], including communication networks [87], power networks [88], transportation networks [89], and supply chain networks [90]. Although previous literature has reported various frameworks for the operation of SATCOM systems (e.g., [9, 91, 92]), few studies have considered the time-varying UT requests in large-scale SATCOM systems.

To operate the system to continuously meet the user requests and connect the large number of system components, the author proposes a framework for establishing a flexible network in large-scale SATCOM systems. The author first describes the network structure of user links, feeder links, and inter-satellite links in terms of network connection matrices. By using these matrices, the author then describes constraints in the network design and candidate cost functions, including the sum of the bandwidth loss of the UT, the number of active satellites and GWs, and the number of satellite handovers. The SATCOM operator can define the cost function by combining the candidate functions according to their

individual management strategy. Using these constraints and cost functions, the SATCOM network design problem is formulated as a mixed integer programming problem. This formulated optimization problem is solved to simultaneously obtain the resource allocation and the network structure.

The proposed method constructs a SATCOM network that efficiently utilizes the limited resources. Furthermore, the proposed framework enables the design of the SATCOM network to allow different types of satellites to be connected in a unified manner. In this formulation, sets of connectable pair candidates in user links, feeder links, and inter-satellite links are defined. By incorporating these sets into the optimization problem, various types of satellite properties, such as orbit (e.g. GEO and NGEO), frequency, beamwidth (e.g. wide and narrow spot beam), and power, can be identified. As a result, the UTs can obtain the resources as they request, and the SATCOM operator can make the flexible SATCOM network more reliable.

In numerical experiments, the author verified that an efficient network is obtained by using the proposed method to respond to time-varying UT requests. Furthermore, the relationship between the increase in the number of satellites and the performance was also analyzed.

The results given in this chapter are related to the part of the following paper [77]:

- Y. Abe, M. Ogura, H. Tsuji, A. Miura, and S. Adachi, "Resource and Network Management Framework for a Large-Scale Satellite Communications System," *IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences*, vol. E103-A, no. 2, pp. 492–501, 2020.

Chapter 5: A Chance-Constrained Approach for Managing Stochastic Communication Requests

In Chapter 5, stochastic and time-varying communication requests from UTs are addressed. To cope with the requests, the author proposes a resource and network management method that uses a chance-constrained MPC [78, 93, 94]. The proposed method can be applied to both the proposed management frameworks for the frequency flexibility in Chapter 3 and to the massive connected networks in Chapter 4.

UTs requests in large-scale SATCOM systems behave like non-stationary processes and vary in time because of both deterministic and stochastic factors [74–76]. Future requests originating from deterministic factors can be estimated in advance. For example, the requests of the mobile communications on aircraft can be roughly estimated because the aircraft trajectories can be obtained. On the other hand, a portion of the requests change stochastically according to unknown reasons. Thus, it is difficult to predict the future behavior of the changes in requests. From the above, the author proposes to apply both predictive and stochastic control approaches for managing SATCOM systems. To this end, the robustness expressed by chance constraints is incorporated into an MPC-based optimization problem. The chance constraints require satisfying constraints on the performance for a user-specified reliability.

The proposed optimization aims to design a time-varying SATCOM network that guarantees a user-specified performance with high probability given the stochastic and time-varying requests. In this study, the requests are modeled as an auto-regressive integrated moving-average (ARIMA) model, which enables the representation of non-stationary processes [95]. This model is utilized to predict future requests. The purpose of the optimization, which is to ensure a specified performance in the future, is described as a joint chance constraint on the bandwidth loss rate. This constraint can be added to the optimization problem formulated in Chapters 3 and 4. However, the joint chance constraint is nonlinear in the design variables and the formulated problem cannot be solved efficiently. Therefore, the constraint is relaxed to a linear deterministic one under some assumptions. With the linear deterministic constraint, the original problem is reformulated as a relaxed deterministic problem that can be solved more efficiently. By solving the formulated optimization problem, the SATCOM operator obtains the resource allocation and network structure to accommodate the stochastic and time-varying communication requests.

In numerical examples, the author verified that an efficient resource allocation and network structure was obtained according to a user-specified performance even when the communication requests changed stochastically.

Chapter 6: Conclusion and Future Work

Chapter 6 summarizes the thesis and discusses future work in the field that will be required to implement large-scale SATCOM systems.

Chapter 2

Related Work

In this chapter, related work for this thesis is described. Basic methodologies introduced in this chapter are utilized to propose management frameworks in Chapters 3, 4, and 5.

2.1 Satellite Communication Systems

This section introduces the basic characteristics of the SATCOM systems that are associated with satellite hardware configuration, frequency allocation, and the link budget analysis.

2.1.1 Satellite Configuration

Communication satellites are mainly composed of a transponder and receive and transmit antennas [1]. A basic configuration of communication satellites is shown in Fig. 2.1. In this figure, a single beam per feed type is described (i.e., one feed forms a single beam). The transponder, which is short for *trans*mitter-res*ponder*, is an RF communication payload designed to amplify the received RF signal, perform frequency conversion, and transmit the amplified RF signal to the earth. The transponder constitutes the following components: a low noise amplifier (LNA), a down-converter (DNC), an analog-to-digital (A/D) converter, a demultiplexer (DEMUX), switches, a multiplexer (MUX), a digital-to-analog (D/A) converter, an up-converter (UPC), and a traveling wave tube amplifier (TWTA). For a multi-beam satellite, the transponder can switch beams to transmit the signal by implementing the multiple switches. This type of satellite includes digital payloads of the DEMUX, switches, and MUX. A set of these digital payloads is collectively called a digital channelizer, which

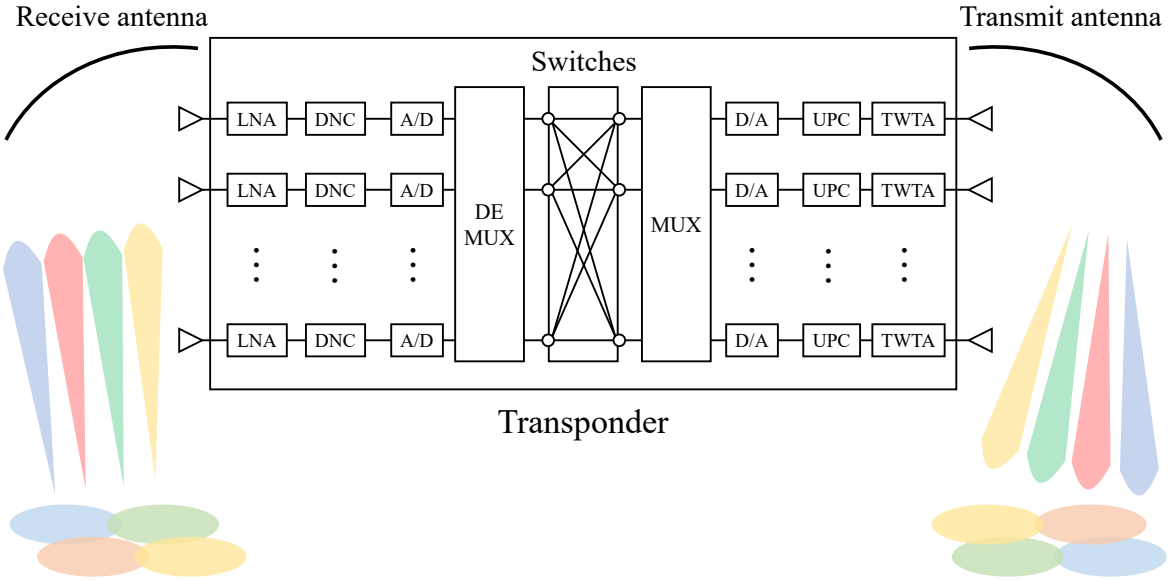


Fig. 2.1: A basic configuration of communication satellites.

is introduced in Subsection 1.3.2.

Here, a flow of signal processing in the communication satellite is described. After receiving the RF analog signal from the earth by the receiver antenna, the signal is first amplified by the LNA and downconverted by the DNC to a low frequency, called the intermediate frequency (IF), for processing the signal stably. For a satellite with digital payloads, after this conversion, the analog signal is converted to a digital one by the A/D converter. To relay the signal from one beam to another, the multiplexed signal is demultiplexed by the DEMUX. Then, beams to which each demultiplexed signal is transmitted are switched, and the signal is multiplexed again by the MUX. After converting the signal from digital to analog, the signal is upconverted from the IF-band to RF-band by the UPC. The signal is amplified by the TWTA, radiated from each feed, and transmitted to the earth by the transmit antenna.

2.1.2 Frequency Allocation

Different SATCOM systems use different frequency bands. Radio waves are defined by the International Telecommunication Union Radiocommunication Sector (ITU-R) as “electromagnetic waves of frequencies arbitrarily lower than 3,000 GHz, propagated in space without artificial guide” [96]. Table 2.1 and Fig. 2.2 show the definition of the frequency

Table 2.1: Definition of frequency bands as determined by the IEEE [5]. HF stands for "high frequency," VHF stands for "very high frequency," UHF stands for "ultra high frequency," and mm stands for "millimeter."

Band	Frequency range
HF	3–30 MHz
VHF	30–300 MHz
UHF	300–3,000 MHz
L	1–2 GHz
S	2–4 GHz
C	4–8 GHz
X	8–12 GHz
Ku	12–18 GHz
K	18–26 GHz
Ka	26–40 GHz
V	40–75 GHz
W	75–110 GHz
mm	110–300 GHz

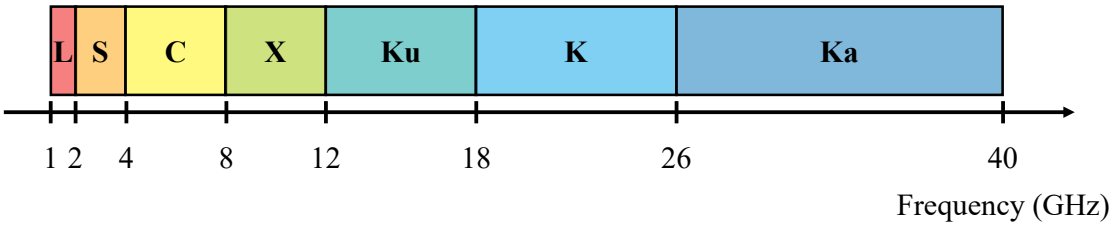


Fig. 2.2: Definition of frequency bands (L-band to Ka-band) as determined by the IEEE[5].

bands as determined by the Institute of Electrical and Electronic Engineers (IEEE) [5][*2][*3]. In general, using a high carrier frequency has a large capacity because of the large available bandwidth. However, in contrast, a high-frequency signal is strongly attenuated due to spatial propagation obstacles such as rainfall. Thus, an appropriate frequency band must be chosen for each SATCOM system according to its individual objective.

The frequency bands shown in Table 2.1 and Fig. 2.2 cannot be used freely. Frequency allocation is mainly determined by the ITU-R, and SATCOM operators are required to register the frequency they use in their systems. Table 2.2 shows an example of the frequency allocation that is available for SATCOM systems[96, 97]. The Ka-band, which is generally used by HTSs, is mainly allocated to the fixed satellite service (FSS) and mobile satellite service (MSS).

[*2]This paper discusses the frequency band used in radar systems, but this commonly overlaps in SATCOM systems.

[*3]In some cases, "K-band" is also interpreted as "Ka-band."

Table 2.2: Frequency allocation as determined by the ITU-R[96, 97].

Radio communication service	Typical frequency band (for uplink/downlink)	Frequency band defined in Table 2.1
Fixed satellite service (FSS)	6/4 GHz	C-band
	8/7 GHz	X-band
	14/11–12 GHz	Ku-band
	30/20 GHz	Ka-band
	50/40 GHz	V-band
Mobile satellite service (MSS)	1.6/1.5 GHz	L-band
	30/20 GHz	Ka-band
Broadcasting satellite service (BSS)	2/2.2 GHz	S-band
	12 GHz	Ku-band
	2.6/2.5 GHz	S-band
Radio navigation satellite service (RNSS)	1.164–1.3 GHz (down)	L-band
	1.559–1.617 GHz (down)	L-band
	5.000–5.010 GHz (up)	C-band
	5.010–5.030 GHz (down)	C-band

2.1.3 Link Budget Analysis

One of the most important aspects to incorporate into the design of SATCOM systems is the link budget analysis. This determines satellite parameters, such as power and antenna gain, so that the communication links can be established. In this analysis, every possible factor that is related to communication links between a transmitter and a receiver should be included.

In this section, a basic link budget analysis is expressed by using an example situation in which the transmitter is a satellite and the receiver is an earth station, as shown in Fig. 2.3. For simplicity, this example focuses only on a downlink, though the link budget analysis generally focuses on both the uplink and downlink. Appropriate units must be used to calculate each parameter. In the following analysis, each system parameter has a subscript of "dB" when the unit of the parameter is decibels.

The power received P_r by the earth station equals the transmitted power from the satellite plus all gains minus all losses as expressed by

$$P_{r,\text{dB}} = P_{t,\text{dB}} - L_{t,\text{dB}} + G_{t,\text{dB}} - L_{f,\text{dB}} + G_{r,\text{dB}} - L_{r,\text{dB}} - L_{s,\text{dB}}, \tag{2.1}$$

where P_t is the transmit power, L_t is the transmit feeder loss, G_t is the transmit antenna gain, L_f is the free-space loss, G_r is the receive antenna gain, L_r is the receive feeder loss,

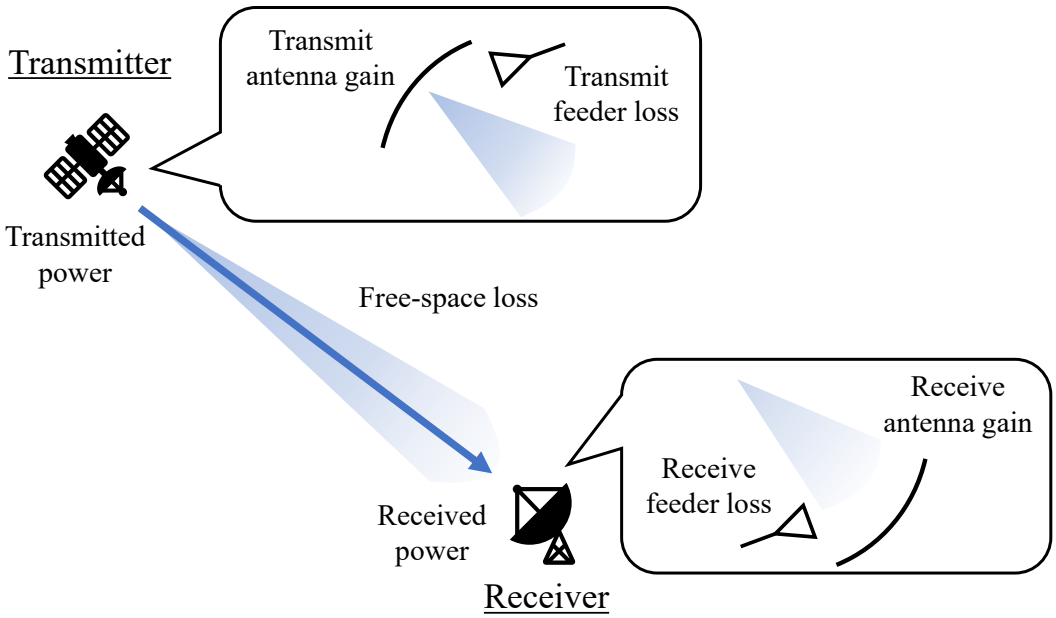


Fig. 2.3: An example of factors that are involved in the calculation of the link budget.

and L_s is other losses that should be included such as atmospheric loss, rain attenuation, and pointing loss. The sum of the first three factors in the transmitter,

$$\text{EIRP}_{\text{dB}} = P_{t,\text{dB}} - L_{t,\text{dB}} + G_{t,\text{dB}}, \tag{2.2}$$

is called the equivalent isotropically radiated power (EIRP), which indicates "how the satellite is performing compared to an isotropic source with 1 W of RF drive power (i.e., 0 dBW)" [1]. The free space loss L_f caused by the spatial propagation of radio waves attenuates the power of the signal by a factor of

$$L_f = \left(\frac{c}{4\pi df}\right)^2, \tag{2.3}$$

where d is the distance between the transmitter and receiver, f is the carrier frequency, and c is the speed of light ($c = 299,792,458$ m/s). This expression means that the amount of the free-space loss is inversely proportional to the square of both the distance and the carrier frequency.

To measure the quality of communication links, the carrier-to-noise ratio C/N_r, which represents the ratio of the received carrier power P_r to the system noise power, needs to be calculated. In the link budget analysis, the system noise of interest is mainly thermal noise, which acts as a disturbance that influences the received signal and deteriorates the quality of

the signal in the receiver. Thus, the received system noise power N_r is represented by

$$N_r = k_B T_n B_c, \tag{2.4}$$

where k_B is the Boltzmann constant ($k_{\text{B}} = 1.380649 \times 10^{-23}$ J/K), T_n is the system noise temperature, and B_c is the carrier bandwidth.

The link margin is defined as the difference between the minimum value and the threshold value of the ratio of the energy per bit to the noise power density E_b/N_0. It is known that "E_b/N_0 is a good measure to compare digital systems using different modulation or encoding schemes and data rates" [98]. The relationship between the carrier-to-noise ratio C/N_r and the minimum value of E_b/N_0 is represented by

$$C/N_r = E_b/N_0 \cdot R_b/B_c, \tag{2.5}$$

where R_b is the information bit rate. The link margin is calculated by subtracting the required $(E_b/N_0)_{\text{req.,dB}}$, which represents the threshold that ensures the specified link performance, from the obtained $(E_b/N_0)_{\text{dB}}$ as

$$\text{Link margin} = (E_b/N_0)_{\text{dB}} - (E_b/N_0)_{\text{req.,dB}}. \tag{2.6}$$

This margin represents how much additional losses can be allowed in this system to exhibit the specified link quality.

Table 2.3 represents an example of link budget analysis. Assume that $f = 14$ GHz, $d = 35,786$ km, $T_n = 140$ K, $B_c = 27$ MHz, $R_b = 30$ Mbps, $(E_b/N_0)_{\text{req.,dB}} = 8.0$ dB, and L_s is ignored. In this example, the link margin is a positive value (3.1 dB). Thus, this system may provide the specified quality of a service link. However, if there is heavy rain or another multipath fading, the margin could be negative. To avoid this, the SATCOM system should be designed so that the system has enough link margin to exhibit the required performance.

Remark 2.1. If both the uplink and downlink are considered, the total carrier-to-noise ratio $(C/N_r)_{\text{total}}$ can be calculated by

$$(C/N_r)^{-1}_{\text{total}} = (C/N_r)^{-1}_{\text{up}} + (C/N_r)^{-1}_{\text{down}}, \tag{2.7}$$

where $(C/N_r)^{-1}_{\text{up}}$ and $(C/N_r)^{-1}_{\text{down}}$ represent the carrier-to-noise ratio of the uplink and down-

Table 2.3: Example of link budget analysis based on [1]. All parameters are expressed in decibels.

Parameter	Value	Units
EIRP: Eq. (2.2) (A=A1+A2+A3)	53.0	dBW
Transmitted power P_t (250 W) (A1)	24.0	dBW
Transmit waveguide loss L_t (A2)	-1.0	dB
Transmit antenna gain G_t (A3)	30.0	dBi
Free-space loss L_f: Eq. (2.3) (B)	-206.4	dB
Receive antenna gain G_r (C)	32.7	dBi
Receive waveguide loss L_r (D)	-0.5	dB
Received power P_r: Eq. (2.1) (A+B+C+D)	-121.2	dBW
Received system noise power N_r: Eq. (2.4)	-132.8	dBW
$C/N_r (= P_r/N_r)$	11.6	dB
E_b/N_0: Eq. (2.5) (E)	11.1	dB
$(E_b/N_0)_{\text{req.}}$ (F)	8.0	dB
Link margin: Eq. (2.6) (E-F)	3.1	dB

link, respectively. Furthermore, gains of the satellite payloads, such as the transponder introduced in Subsection 2.1.1, should be included in the analysis.

2.2 Model Predictive Control

In this section, a standard model predictive control (MPC) formulation is described.

MPC is a finite-time optimal control method, which determines control input sequences online while predicting future states and satisfying system constraints [78]. This framework has been applied to various fields such as process control, as system hardware constraints can be included in the MPC.

Figure 2.4 shows a schematic diagram of MPC. The control objective is to find an optimal input sequence to make the output follow the reference. Finite horizons T_p and T_u ($\leq T_p$) are called the prediction and control horizons, respectively. A future value of the output is predicted based on the system model and an optimal input sequence is determined using the MPC scheme. Here, the control input is assumed to be constant after the time $k + T_u$.

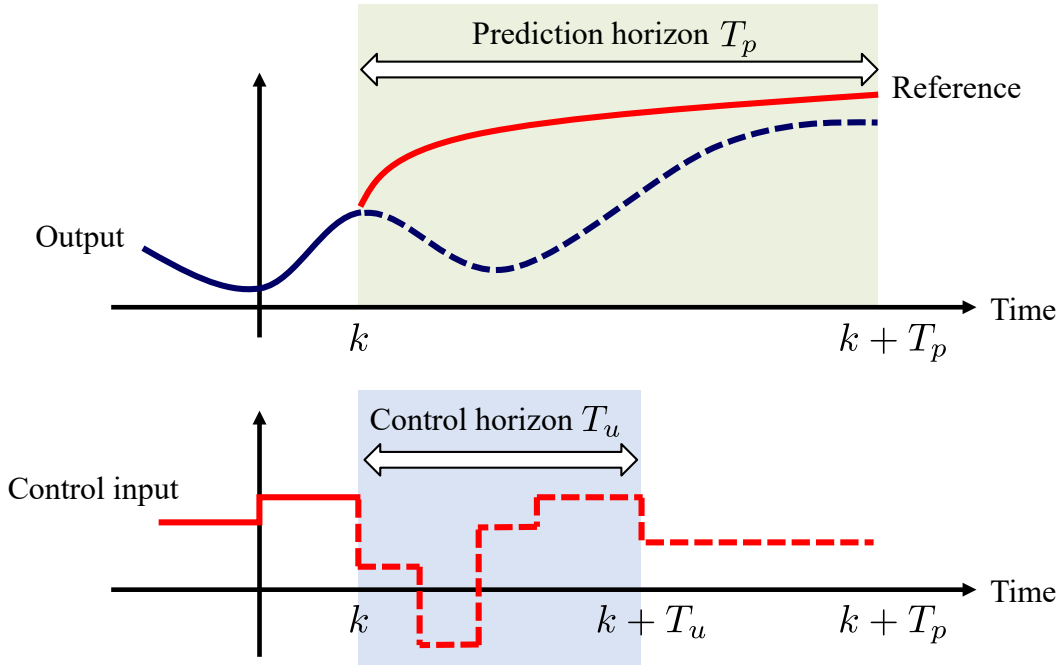


Fig. 2.4: Schematic diagram of MPC. Upper panel: output and reference. Lower panel: control input.

Consider a discrete-time linear dynamical system described by

$$\begin{cases} \boldsymbol{x}_{k+1} = \boldsymbol{A}\boldsymbol{x}_k + \boldsymbol{B}\boldsymbol{u}_k, \\ \boldsymbol{y}_k = \boldsymbol{C}\boldsymbol{x}_k, \end{cases} \tag{2.8}$$

where k is the discrete time; $\boldsymbol{x}_k \in \mathbb{R}^n$ is the state; $\boldsymbol{u}_k \in \mathbb{R}^m$ is the input; $\boldsymbol{y}_k \in \mathbb{R}^p$ is the output; and $\boldsymbol{A} \in \mathbb{R}^{n\times n}$, $\boldsymbol{B} \in \mathbb{R}^{n\times m}$, and $\boldsymbol{C} \in \mathbb{R}^{p\times n}$ are constant matrices. By using this system model, a future output sequence can be predicted.

In a standard MPC problem, a cost function is defined as

$$J_k = \sum_{\tau=1}^{T_p} \|\boldsymbol{y}_{k+\tau} - \boldsymbol{r}_{k+\tau}\|_{\boldsymbol{Q}}^2 + \sum_{\tau=1}^{T_u} \|\Delta\boldsymbol{u}_{k+\tau}\|_{\boldsymbol{R}}^2, \tag{2.9}$$

where $\Delta\boldsymbol{u}_{k+\tau}$ represents the difference between $\boldsymbol{u}_{k+\tau}$ and $\boldsymbol{u}_{k+\tau-1}$, defined as

$$\Delta\boldsymbol{u}_{k+\tau} := \boldsymbol{u}_{k+\tau} - \boldsymbol{u}_{k+\tau-1},$$

and $\boldsymbol{Q} \in \mathbb{R}^{M\times M}$ and $\boldsymbol{R} \in \mathbb{R}^{M\times M}$ are weighting matrices with the (i,i)-th elements denoted by $q^{(i,i)}$ and $r^{(i,i)}$, respectively. The notation $\|\boldsymbol{y}_{k+\tau} - \boldsymbol{r}_{k+\tau}\|_{\boldsymbol{Q}}^2$ represents the quadratic form as $(\boldsymbol{y}_{k+\tau} - \boldsymbol{r}_{k+\tau})^\top \boldsymbol{Q}(\boldsymbol{y}_{k+\tau} - \boldsymbol{r}_{k+\tau})$. The first term aims to reduce the error $\boldsymbol{y}_{k+\tau} - \boldsymbol{r}_{k+\tau}$,

and the second term aims to reduce the frequency change of the output. To formulate an optimization problem, system constraints can be included. The optimization problem is solved at every time instant, and the first optimal input is applied to the system.

MPC has been widely studied from both theoretical and applied points of view. The stability and optimality of MPC are proven in [99]. Furthermore, a wide variety of MPC applications have been developed in recent years - for example, chance-constrained MPC [100], distributed MPC [101, 102], event-triggered MPC [103, 104].

2.3 Sparse Optimization

Sparse optimization, which is also called compressive sensing, compressive sampling, or sparse sampling, is a theory designed to reconstruct an unknown sparse vector from a smaller dimension of measurements than the size of the unknown vector [79, 105]. The sparse optimization is applied to a wide variety of domains such as control systems [106], digital signal processing [81], and wireless communications [80, 107].

In this theory, it is important to select an optimized norm. For a vector $\boldsymbol{x} \in \mathbb{R}^n$, the ℓ_2, ℓ_1, and ℓ_0-norms are defined as

$$\|\boldsymbol{x}\|_2 = \sqrt{\sum_{i=1}^{n} x_i^2}, \quad \|\boldsymbol{x}\|_1 = \sum_{i=1}^{n} |x_i|, \quad \|\boldsymbol{x}\|_0 = \sum_{i=1}^{n} \delta(x_i),$$

respectively, where

$$\delta(x_i) = \begin{cases} 1, & x_i \neq 0, \\ 0, & x_i = 0. \end{cases}$$

Figure 2.5 shows schematic diagrams of these three types of norms. The ℓ_2-norm represents the sum of the squares of x_i. This norm, known as the Euclidean norm, is the most commonly used norm. The ℓ_1-norm is the sum of the absolute values of x_i and the ℓ_0-norm represents the number of nonzero elements in $\boldsymbol{x}$[*4].

Before formulating the sparse optimization, a general vector estimation problem is first introduced. Consider an equation represented by $\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x} + \boldsymbol{b}$, where $\boldsymbol{x} \in \mathbb{R}^n$ is an unknown

[*4]Strictly speaking, the ℓ_0-norm is not a norm because the definition of the ℓ_0-norm does not satisfy the norm axioms [108]: (i) $\|\boldsymbol{x}\| = 0 \Leftrightarrow \boldsymbol{x} = \boldsymbol{0}$, (ii) $\|\alpha\boldsymbol{x}\| = |\alpha|\|\boldsymbol{x}\|, \alpha \in \mathbb{R}$, (iii) $\|\boldsymbol{x} + \boldsymbol{y}\| \leq \|\boldsymbol{x}\| + \|\boldsymbol{y}\|$.

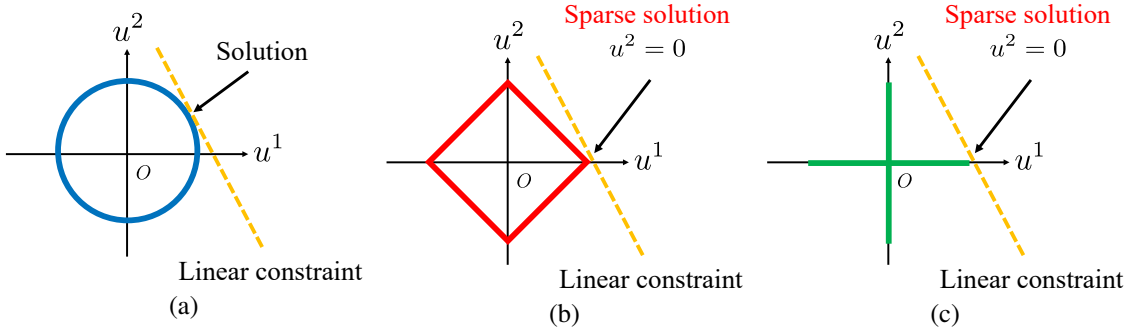


Fig. 2.5: Schematic diagrams of the (a) ℓ_2-norm, (b) ℓ_1-norm, and (c) ℓ_0-norm. The solid lines represent the contours of the norms of $\boldsymbol{u} \in \mathbb{R}^2$ and the dashed lines represent linear constraints.

vector, $\boldsymbol{y} \in \mathbb{R}^m$ is a measurement vector, $\boldsymbol{A} \in \mathbb{R}^{m \times n}$ is a measurement matrix, and $\boldsymbol{b} \in \mathbb{R}^m$ is a noise vector. When $n \leq m$, the vector estimation problem is represented by

$$\underset{\boldsymbol{x}}{\text{minimize}} \quad \|\boldsymbol{y} - \boldsymbol{A}\boldsymbol{x}\|_2^2, \tag{2.10}$$

and a least-squares solution can be obtained by

$$\hat{\boldsymbol{x}} = (\boldsymbol{A}^\top \boldsymbol{A})^{-1} \boldsymbol{A}^\top \boldsymbol{y}. \tag{2.11}$$

When $n > m$, the vector estimation problem is represented by

$$\underset{\boldsymbol{x}}{\text{minimize}} \quad \|\boldsymbol{y} - \boldsymbol{A}\boldsymbol{x}\|_2^2 + \lambda \|\boldsymbol{x}\|_2^2, \tag{2.12}$$

where λ is a regularization parameter. This problem is called regularized least squares or ridge regression. A regularized least-squares solution can be obtained by

$$\hat{\boldsymbol{x}} = (\boldsymbol{A}^\top \boldsymbol{A} + \lambda \boldsymbol{I}_n)^{-1} \boldsymbol{A}^\top \boldsymbol{y}. \tag{2.13}$$

The aim of sparse optimization is to recover the unknown vector $\boldsymbol{x} \in \mathbb{R}^n$ as a sparse vector from the measurement $\boldsymbol{y}$ when $\|\boldsymbol{x}\|_0 \ll n$ (which means $\boldsymbol{x} \in \mathbb{R}^n$ is sparse and $n > m$), where the number of unknown variables is more than that of the measurements. When the unknown vector $\boldsymbol{x}$ is known as sparse, the optimized norm of $\boldsymbol{x}$ is chosen as the ℓ_0-norm and the vector estimation problem is represented by

$$\underset{\boldsymbol{x}}{\text{minimize}} \quad \|\boldsymbol{y} - \boldsymbol{A}\boldsymbol{x}\|_2^2 + \lambda \|\boldsymbol{x}\|_0. \tag{2.14}$$

This problem is called the ℓ_0 regularization. However, as shown in Fig. 2.5(c), this norm is nonlinear and nonconvex in the vector $\boldsymbol{x}$. Therefore, an optimization problem under the ℓ_0-norm is an NP-hard problem, which is difficult to solve. To solve the problem efficiently, the ℓ_0-norm is replaced by the ℓ_1-norm, which is linear in the vector $\boldsymbol{x}$ as shown in Fig. 2.5(b). Then, the optimization problem becomes

$$\underset{\boldsymbol{x}}{\text{minimize}} \quad \|\boldsymbol{y} - \boldsymbol{A}\boldsymbol{x}\|_2^2 + \lambda\|\boldsymbol{x}\|_1. \tag{2.15}$$

This problem is called the ℓ_1 regularization, or the least absolute shrinkage and selection operator (LASSO) regression.

Next, sparse optimal control [106] is described. This control scheme tries to achieve the control objective with a sparse input sequence.

Consider a discrete-time linear system:

$$\boldsymbol{x}_{k+1} = \boldsymbol{A}\boldsymbol{x}_k + \boldsymbol{B}\boldsymbol{u}_k,$$

where $\boldsymbol{x} \in \mathbb{R}^n, \boldsymbol{u} \in \mathbb{R}^m, \boldsymbol{A} \in \mathbb{R}^{n\times n}$ and $\boldsymbol{B} \in \mathbb{R}^{n\times m}$. The control objective is to compute the control input sequence $\{\boldsymbol{u}_1, \boldsymbol{u}_2, \ldots, \boldsymbol{u}_{N-1}\}$ that drives the state $\boldsymbol{x}_k$ from a given initial state $\boldsymbol{x}_1$ to a desired value $\bar{\boldsymbol{x}}$ at a fixed final time N (that is, $\boldsymbol{x}_N = \bar{\boldsymbol{x}}$).

To find an optimal input sequence, ℓ_0- and ℓ_1-optimal control problems are defined as

$$\underset{\boldsymbol{u}}{\text{minimize}} \quad \sum_{\tau=1}^{N-1} \|\boldsymbol{u}_\tau\|_0$$

and

$$\underset{\boldsymbol{u}}{\text{minimize}} \quad \sum_{\tau=1}^{N-1} \|\boldsymbol{u}_\tau\|_1,$$

respectively. These problems are equivalent if the following condition holds. The ℓ_1-optimal control problem is considered to be normal if $(\boldsymbol{A}, \boldsymbol{B})$ is controllable, which means

$$\text{rank}\begin{bmatrix} \boldsymbol{B} & \boldsymbol{A}\boldsymbol{B} & \boldsymbol{A}^2\boldsymbol{B} & \cdots & \boldsymbol{A}^{n-1}\boldsymbol{B} \end{bmatrix} = n,$$

and $\boldsymbol{A}$ is nonsingular. These conditions are called the normality sufficient condition [109].

If it is satisfied, the following theorem holds.

Theorem 2.1 (Equivalence of ℓ_0- and ℓ_1-optimal control problems [106])**.** Let $\mathcal{U}_0^*$ and $\mathcal{U}_1^*$ be sets of the optimal solutions of the ℓ_0- and ℓ_1-optimal control problem, respectively. If the ℓ_1-optimal control problem is normal and has at least one solution, $\mathcal{U}_0^* = \mathcal{U}_1^*$ holds.

Chapter 3

Management Framework for Frequency Flexibility

This chapter is organized as follows. In Section 3.1, the system configuration and beam arrangement of a SATCOM system with frequency flexibility is detailed. An example of bandwidth allocation is also given. In Section 3.2, the proposed method for efficient bandwidth allocation is presented. In Section 3.3, the effectiveness of the proposed method is verified through two examples of numerical experiments. Subsection 3.3.1 details the verification of the proposed method based on a simple time variation model of communication requests. Subsection 3.3.2 analyzes actual flight data to derive communication demand from aircraft, and the proposed method is applied to this demand. In Section 3.4, the performance of the proposed method is discussed based on parameter tuning and computation time. Finally, this chapter is summarized in Section 3.5.

3.1 Problem Statement

In this section, the author states a problem setting for managing an HTS with the frequency flexibility. Its system configuration and beam arrangement are detailed, and then bandwidth allocation is illustrated. In this chapter, the author calls a communication satellite an HTS because it is assumed that the satellite aims to increase its capacity by using the Ka-band and that it implements multiple beams.

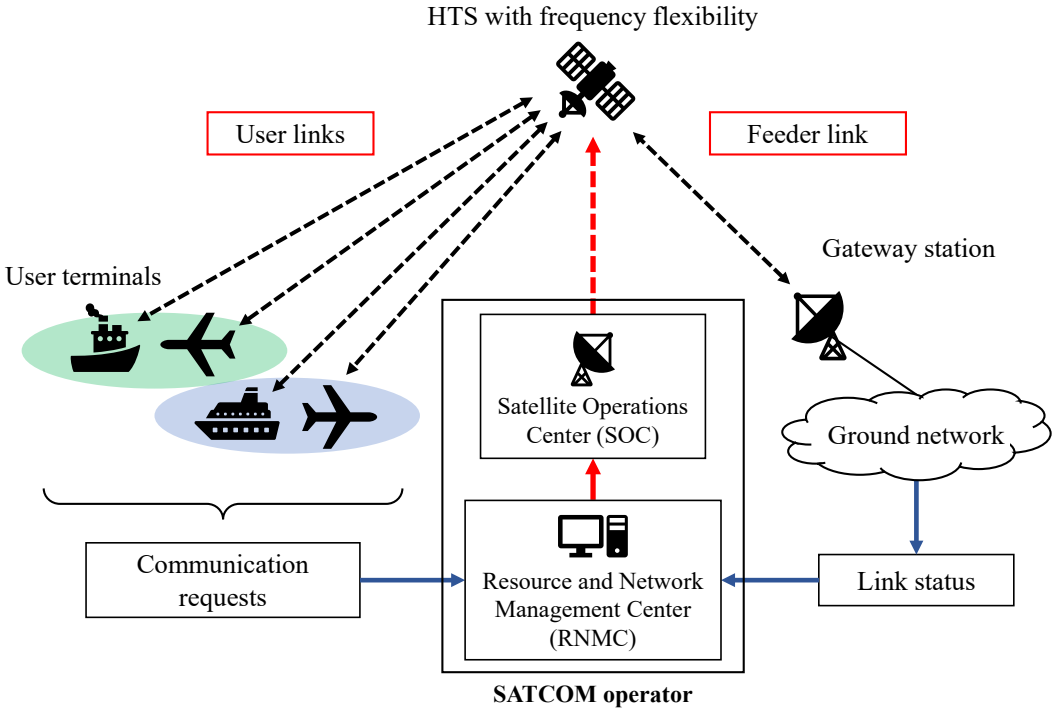


Fig. 3.1: Diagram of a SATCOM system with frequency flexibility. The RNMC and SOC serve as controllers for the HTS.

3.1.1 System Configuration

A diagram illustrating a SATCOM system is shown in Fig. 3.1. This system includes an HTS with frequency flexibility in a GEO, multiple UTs, and a GW. The SATCOM operator manages the HTS by controlling the digital channelizer in the HTS and changing the bandwidth allocation per beam. In this chapter, the author focuses on the bandwidth allocation only for the user links and assumes that the feeder link can deliver enough bandwidth to accommodate all the communication requests.

3.1.2 Satellite Beam Arrangement

Figure 3.2 shows an example of a beam arrangement for the HTS. Each beam is assumed to have a half-power beamwidth of 0.46°(i.e., the approximate beam diameter is 300 km per beam). In the HTS beam arrangement, either the frequency or polarization of adjacent beams should be different to prevent interference. To this end, the multiple beams of the HTS constitute a cluster composed of four beams with two frequency bands and two polarizations. Specifically, for pairs of beams (**A**, **B**) and (**C**, **D**), the bandwidth is allocated based on the

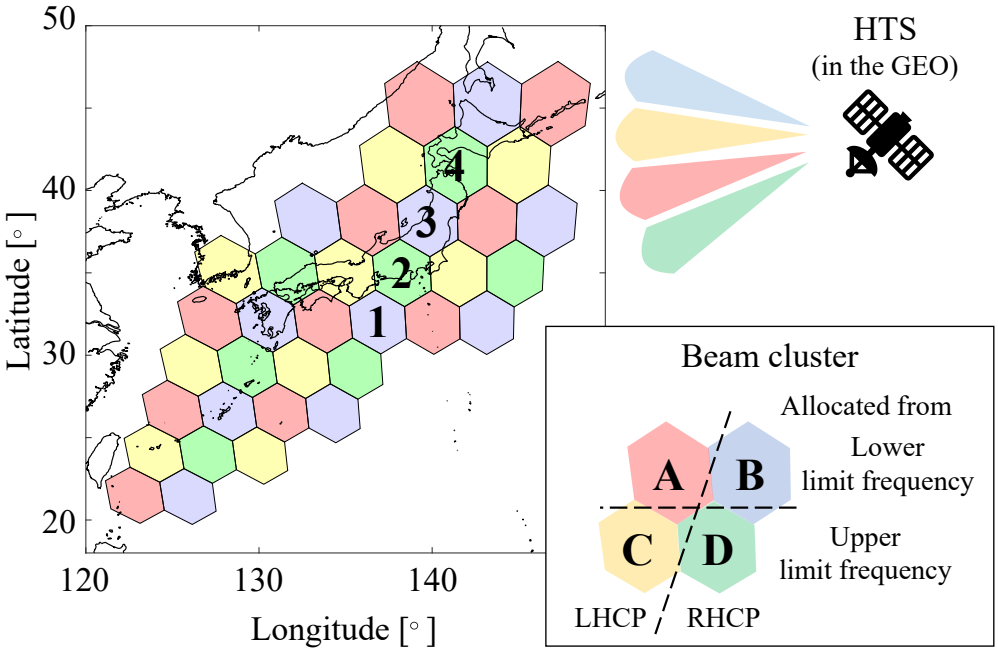


Fig. 3.2: Arrangement of the 36 beams of the HTS over Japan and the surrounding ocean.

lower frequency limit $f_{\min}$ and the upper frequency limit $f_{\max}$ assigned to the SATCOM system. Hence, the available bandwidth

$$B_{\max} := f_{\max} - f_{\min}$$

is divided into two. In addition, for the pairs of beams (**A**, **C**) and (**B**, **D**), left- and right-hand circularly polarized (LHCP and RHCP) waves are used, respectively.

The amount of interference is small at different polarizations, even if the same frequency band is used, such as for beams **A** and **B**. Therefore, this chapter focuses on the bandwidth allocation for beams exhibiting the same polarization (e.g., beams 1 to 4 in Fig. 3.2).

3.1.3 Example of Bandwidth Allocation

An example of the bandwidth allocation for the above-mentioned four beams is illustrated in Fig. 3.3. This example compares the allocation without frequency flexibility, as shown in Fig. 3.3(a), and with frequency flexibility, as shown in Fig. 3.3(b). In this example, assume that $B_{\max} = 500$ MHz and the aggregate communication requests from multiple UTs are interpreted as the requests per beam. Table 3.1 summarizes the communication requests and bandwidth allocation of each beam.

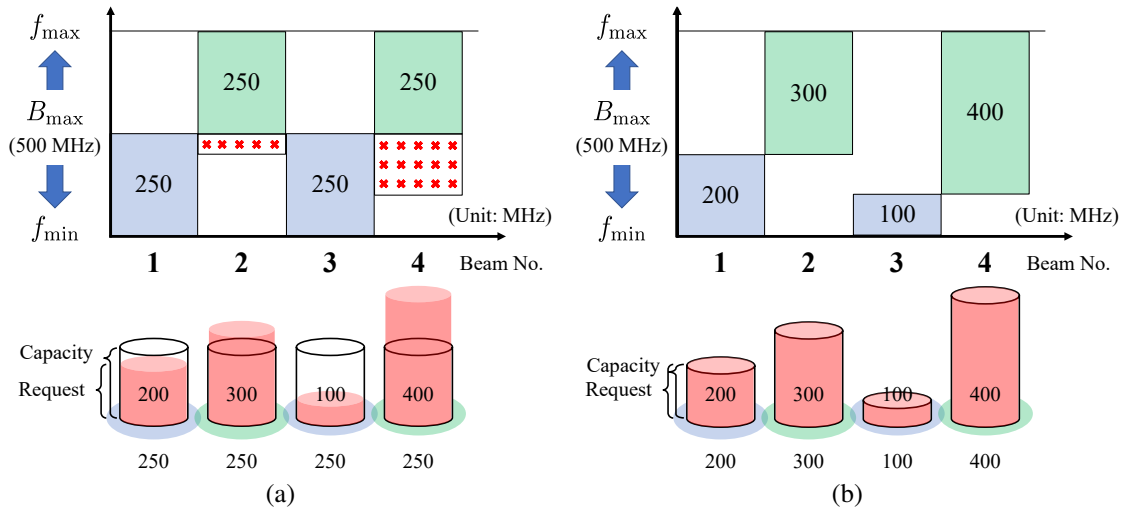


Fig. 3.3: Example of frequency allocation without and with frequency flexibility. (a) Without frequency flexibility, bandwidth loss occurs in beams 2 and 4. (b) Adjusting the allocation of each beam by exploiting the frequency flexibility prevents the bandwidth loss.

Table 3.1: Frequency allocation according to the frequency flexibility and communication requests in Fig. 3.3.

Communication requests [MHz]	Beam 1 200	Beam 2 300	Beam 3 100	Beam 4 400
Allocation without frequency flexibility [MHz]	250	250	250	250
(Difference)	(+50)	(−50)	(+150)	(−150)
Allocation with frequency flexibility [MHz]	200	300	100	400
(Difference)	(**0**)	(**0**)	(**0**)	(**0**)

In Fig. 3.3(a), as bandwidths of $B_{\max}/2 = 250$ MHz are fixed for all four beams, the requests for beams 2 and 4 exceed the allocated bandwidths and hence bandwidth loss occurs. At the same time, an excess in the bandwidth allocation occurs in beams 1 and 3. In Fig. 3.3(b), the bandwidths are adjusted by using the given frequency flexibility function. Thus, the bandwidths of beams 1 and 3 are reduced and those of beams 2 and 4 are increased to prevent both allocation excess and bandwidth loss, respectively. The amount of interference would be small if the frequency bands of the adjacent beams do not overlap.

Here, consider dynamical bandwidth allocation as a problem of efficiently allocating bandwidths for the time-varying communication requests and of preventing the frequency bands among adjacent beams from overlapping. This chapter focuses on only the allocation of the total amount of bandwidth per beam and disregards the number of UTs in each beam. Bandwidth allocation for the multiple UTs in each beam can be achieved using multiple access methods, such as frequency-division multiple access or time-division multiple

access [110, 111].

3.2 Proposed Bandwidth Allocation Method

The bandwidth allocation for time-varying communication requests is treated as an optimal control problem of the HTS. In this section, the dynamics of both the digital channelizer in the HTS and the communication requests per beam are defined. Then a cost function and constraints are defined to formulate an optimization problem based on MPC. The latter determines input sequences online while considering the finite-time horizon and satisfying the constraints. In part of the cost function, the ℓ_1-norm of control input is included based on sparse optimization to obtain a sparse input sequence, which reduces the number of control actions (NCA) of the HTS. Furthermore, it is assumed that the bandwidth is represented by a continuous value and the frequency allocation can be changed simultaneously with changes in the communication requests.

The control objective of the bandwidth allocation is to reduce the bandwidth loss and the NCA. To achieve this objective, an optimization problem to determine the most efficient bandwidth allocation is formulated in the following subsections.

3.2.1 System Description

Let M beams with the same polarization be adjacent to each other, just as beams 1–4 are adjacent to each other in Fig. 3.2. Let x_k^i and u_k^i represent the amount of the allocated bandwidth to the i-th beam at discrete time k and the control input applied to the i-th beam at time k (i.e., the variation of the allocated bandwidth: $x_{k+1}^i - x_k^i$), respectively. The dynamics of the digital channelizer in the HTS, then, is given by the following discrete-time linear system:

$$\boldsymbol{x}_{k+1} = \boldsymbol{x}_k + \boldsymbol{u}_k. \tag{3.1}$$

Here,

$$\boldsymbol{x}_k := \begin{bmatrix} x_k^1 & x_k^2 & \cdots & x_k^M \end{bmatrix}^\top \in \mathbb{R}^M,$$
$$\boldsymbol{u}_k := \begin{bmatrix} u_k^1 & u_k^2 & \cdots & u_k^M \end{bmatrix}^\top \in \mathbb{R}^M.$$

Let d_k^i be the communication request from the i-th beam at time k. Assume that the variation of communication requests is independent of the beams and of the allocated bandwidth to the beams. Then the dynamics of the communication requests is described by

$$\boldsymbol{d}_{k+1} = \boldsymbol{f}_d(\boldsymbol{d}_k), \tag{3.2}$$

where

$$\boldsymbol{d}_k := \begin{bmatrix} d_k^1 & d_k^2 & \cdots & d_k^M \end{bmatrix}^\top \in \mathbb{R}^M$$

and $\boldsymbol{f}_d : \mathbb{R}^M \to \mathbb{R}^M$ is a nonlinear function assumed to be available (i.e., the time variation of the communication requests can be predicted). The allocated bandwidth x_k^i should be larger than the requested bandwidth d_k^i so that all the UTs in each beam can communicate with the HTS.

Figure 3.4 represents the relationship between the variables x_k^i, x_{k+1}^i, u_k^i, and d_k^i and the bandwidth allocation of beams 1 and 2 in the example described in Fig. 3.3 and Subsection 3.1.3. By utilizing the frequency flexibility, the allocated bandwidth of beam 1 is reduced from $x_k^1 = 250$ to $x_{k+1}^1 = 200$ by $u_k^1 = -50$ and that of beam 2 is increased from $x_k^2 = 250$ to $x_{k+1}^2 = 300$ by $u_k^2 = 50$. To measure the performance of the bandwidth allocation, the amount of bandwidth loss at instant k in the i-th beam is defined as

$$L_k^i := \max\left(0, d_k^i - x_k^i\right), \tag{3.3}$$

which represents the amount of the additional bandwidth required to meet the request from the i-th beam. In the example in Fig. 3.4, the bandwidth loss of beams 1 and 2 is $L_k^1 = 0$ and $L_k^2 = 50$ at time k and $L_{k+1}^1 = 0$ and $L_{k+1}^2 = 0$ at time $k+1$, respectively. As this demonstrates, the frequency flexibility function minimizes the bandwidth loss by adjusting the bandwidth allocation.

Figure 3.5 shows the block diagram of the SATCOM system with the frequency flexibility. The digital channelizer of the HTS and communication requests follow the dynamics in Eqs. (3.1) and (3.2), respectively. To accommodate the requests by the HTS, the RNMC and SOC work as a controller and try to determine the control input $\boldsymbol{u}_k$.

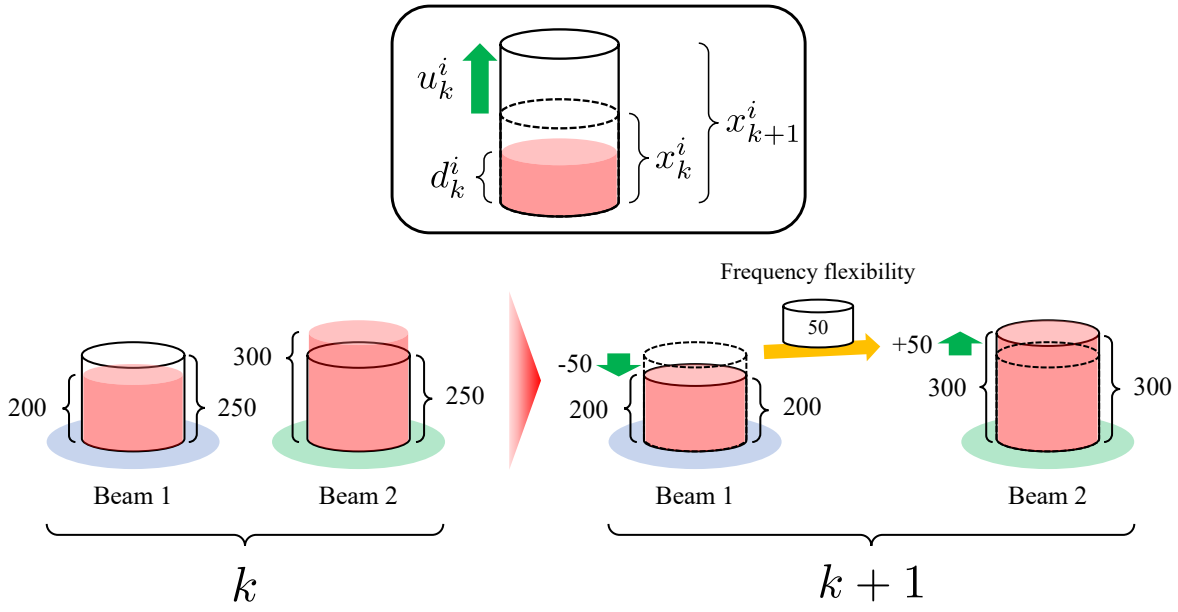


Fig. 3.4: Relationship between the variables x_k^i, x_{k+1}^i, u_k^i, and d_k^i and the bandwidth allocation of beams 1 and 2 in the example described in Fig. 3.3 and Subsection 3.1.3.

3.2.2 Cost function

In this subsection, a cost function is defined and it is minimized in the optimization problem. Some measures have been introduced so far (e.g., in [82]). In these studies, a quadratic form regarding the difference between $\boldsymbol{x}_k$ and $\boldsymbol{d}_k$, described by

$$\tilde{J}_k := (\boldsymbol{d}_k - \boldsymbol{x}_k)^\top \boldsymbol{Q}(\boldsymbol{d}_k - \boldsymbol{x}_k), \tag{3.4}$$

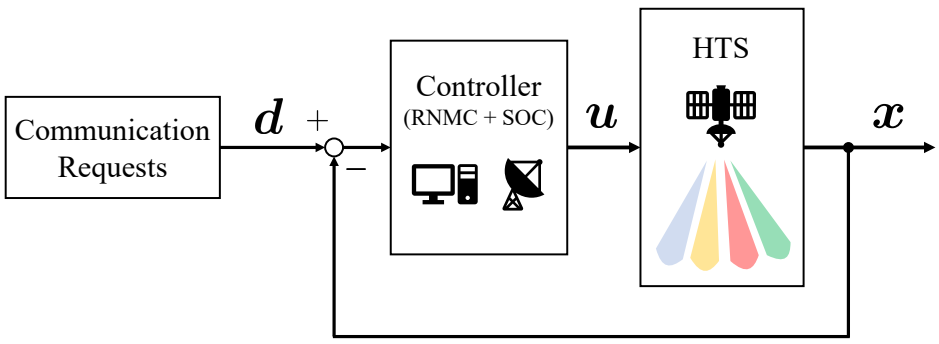


Fig. 3.5: Block diagram of SATCOM system with frequency flexibility. The communication requests of M beams vary independently. The control input $\boldsymbol{u}$ is determined by MPC and applied to the HTS.

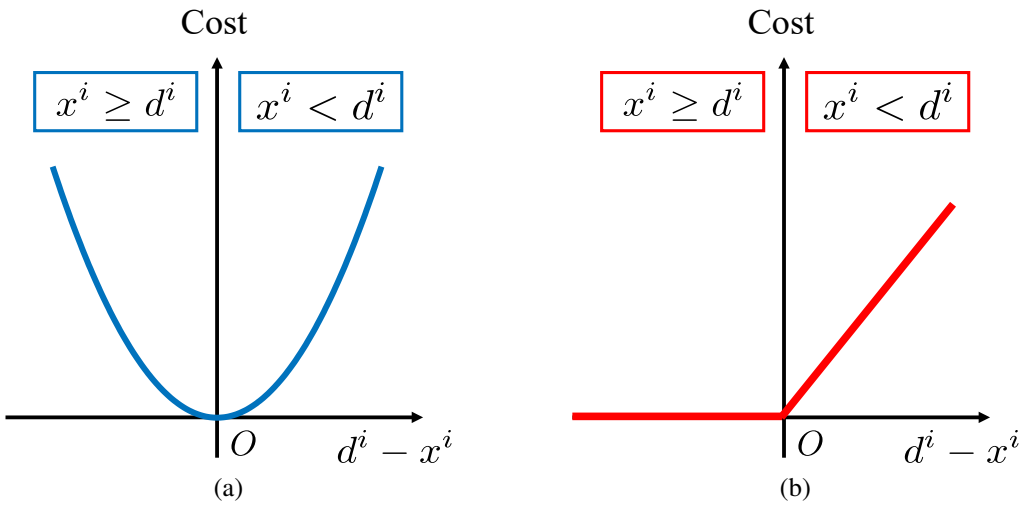


Fig. 3.6: Candidate cost functions J_k as a function of bandwidth loss. (a) Quadratic form and (b) ReLU function, in which the cost increases linearly with increasing $d^i - x^i$ for $x^i < d^i$ and is zero for $x^i \geq d^i$. (Copyright © 2018 IEICE, [76] Fig. 5)

is employed (this is illustrated in Fig. 3.6(a)). However, $\tilde{J}_k$ cannot consider predicted communication requests, nor can it distinguish the domain where bandwidth loss occurs (i.e., $x_k^i < d_k^i$) from that where excessive allocation with respect to the communication requests occurs (i.e., $x_k^i \geq d_k^i$).

For an efficient allocation that reduces the bandwidth loss and NCA, the communication requests should be predicted, and the bandwidth should be allocated so that it has a sufficient margin for the predicted demand. Thus, the cost in $x_k^i < d_k^i$ should be larger than that in $x_k^i \geq d_k^i$. To represent this effect, cost function can be defined over the finite-time horizon by using the MPC-based formulation as

$$J_k := \sum_{\tau \in \mathcal{T}_p} \mathbf{1}_M^\top \cdot \max\{\mathbf{0}_M, \boldsymbol{Q}(\boldsymbol{d}_{k+\tau} - \boldsymbol{x}_{k+\tau})\} + \sum_{\tau \in \mathcal{T}_u} \|\boldsymbol{R}\boldsymbol{u}_{k+\tau-1}\|_1, \tag{3.5}$$

where T_p and T_u $(\leq T_p)$ represent the prediction horizon and the control horizon, respectively, $\mathcal{T}_p = \{1, 2, \ldots, T_p\}$ and $\mathcal{T}_u = \{1, 2, \ldots, T_u\}$ represent sets of time indices from the initial time to the prediction horizon and control horizon, respectively, and $\boldsymbol{Q} \in \mathbb{R}^{M \times M}$ and $\boldsymbol{R} \in \mathbb{R}^{M \times M}$ are weighting matrices with the (i, i)-th elements denoted by $q^{(i,i)}$ and $r^{(i,i)}$, respectively. Larger weights $q^{(i,i)}$ are set for beams that need to be given higher priority so as to prevent a large bandwidth loss. On the other hand, larger weights $r^{(i,i)}$ are set for beams that need to reduce the NCA.

Figure 3.6(b) illustrates the i-th element of the first term of the cost in Eq. (3.5),

$$\max\left\{0, q^{(i,i)}(d^i_{k+\tau} - x^i_{k+\tau})\right\}.$$

This function is called the rectified linear unit (ReLU) function in machine learning literature [112]. The cost increases linearly with an increasing difference between x^i_k and d^i_k for $x^i_k < d^i_k$ and the cost is zero for $x^i_k \geq d^i_k$. Note that $q^{(i,i)}$ also means an increase in the slope of the cost in $x^i_k < d^i_k$.

The second term of the cost in Eq. (3.5) is the weighted ℓ_1-norm of the control input, which is given by

$$\|\boldsymbol{R}\boldsymbol{u}_{k+\tau-1}\|_1 = \sum_{i=1}^{M} \left|r^{(i,i)}u^i_{k+\tau-1}\right|.$$

The ℓ_2-norm is not used because a sparse solution is obtained by using the ℓ_1-norm, as illustrated in Fig. 3.7. A solution by using the ℓ_2-norm is not suitable for the control objective because the solution minimizes the value of each u^i_k on average and does not consider the sparsity.

To reduce the NCA, it is necessary to directly minimize the ℓ_0-norm of $\boldsymbol{u}_k$: $\|\boldsymbol{u}_{k+\tau-1}\|_0$. However, the ℓ_0-norm is nonconvex in the decision variable, and it is difficult to numerically solve the problem for a high-dimensional $\boldsymbol{u}$ (i.e., a large number of beams M). Therefore, the ℓ_0-norm is replaced by the ℓ_1-norm, which is convex in the decision variable. As shown in Section 2.3, the optimal solutions obtained by using the ℓ_0-norm and ℓ_1-norm are equivalent under some conditions [79, 106].

An efficient allocation that considers both the bandwidth loss and NCA is obtained by combining the first and second terms of the cost function in Eq. (3.5).

Remark 3.1. The second term of the cost in Eq. (3.5) is transformed as

$$\begin{aligned}
&\sum_{\tau\in\mathcal{T}_u} \|\boldsymbol{R}\boldsymbol{u}_{k+\tau-1}\|_1 \\
&= \sum_{\tau\in\mathcal{T}_u} \left(|r^{(1,1)}u^1_{k+\tau-1}| + \cdots + |r^{(M,M)}u^M_{k+\tau-1}|\right) \\
&= \left(|r^{(1,1)}u^1_k| + \cdots + |r^{(M,M)}u^M_k|\right) \\
&\qquad + \cdots + \left(|r^{(1,1)}u^1_{k+T_u-1}| + \cdots + |r^{(M,M)}u^M_{k+T_u-1}|\right)
\end{aligned}$$

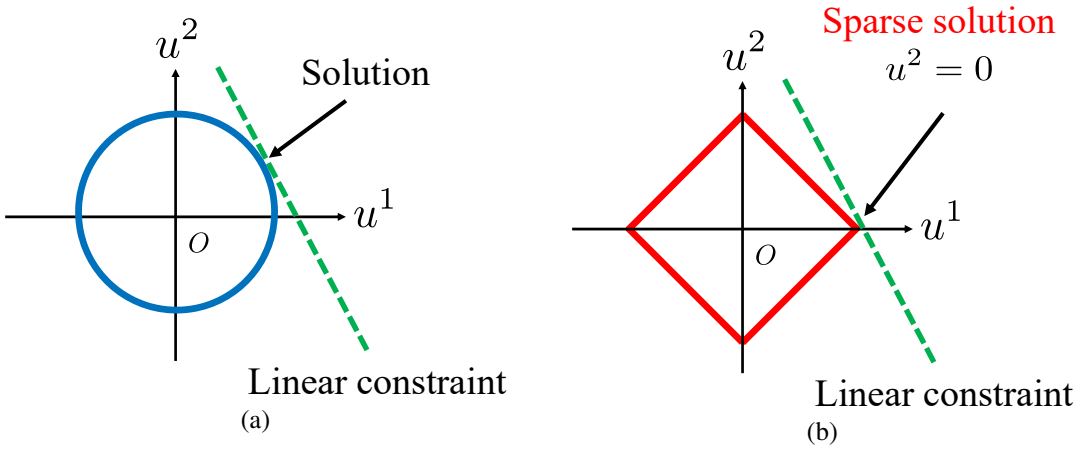


Fig. 3.7: Candidate cost functions J_k according to the NCA. (a) ℓ_2-norm and (b) ℓ_1-norm. The solid lines represent the contours of the norms of $\boldsymbol{u} \in \mathbb{R}^2$, and the dashed lines represent linear constraints. In (a), an optimal solution $u^1 \neq 0$ and $u^2 \neq 0$ is obtained, whereas in (b), a sparse solution $u^1 \neq 0$ and $u^2 = 0$ is obtained. (Copyright © 2018 IEICE, [76] Fig. 6)

$$
\begin{aligned}
&= r^{(1,1)} \left(|u_k^1| + \cdots + |u_{k+T_u-1}^1| \right) + \cdots + r^{(M,M)} \left(|u_k^M| + \cdots + |u_{k+T_u-1}^M| \right) \\
&= r^{(1,1)} \|\boldsymbol{u}_{k|k+T_u-1}^1\|_1 + \cdots + r^{(M,M)} \|\boldsymbol{u}_{k|k+T_u-1}^M\|_1,
\end{aligned} \tag{3.6}
$$

where

$$
\boldsymbol{u}_{k|k+T_u-1}^i := \begin{bmatrix} u_k^i & \cdots & u_{k+T_u-1}^i \end{bmatrix}^\top.
$$

The expression in Eq. (3.6) represents the weighted-sparse term regarding the time domain. Thus, minimizing the second term of the cost function (Eq. (3.5)) results in decreasing the NCA.

3.2.3 Constraints

To formulate the optimization problem, the following constraints for the bandwidth allocation are considered:

- Constraint on nonnegative x_k^i:

$$
x_k^i \geq 0, \quad \forall i = 1, \ldots, M. \tag{3.7}
$$

- Constraint on adjacent inter-beam interference (i.e., the sum of the bandwidth of ad-

jacent beams with the same polarization must not exceed the maximum bandwidth $B_{\max}$):

$$x_k^j + x_k^{j+1} \leq B_{\max}, \quad \forall j = 1, \ldots, M-1. \tag{3.8}$$

3.2.4 Optimization Problem

Combining the above-mentioned dynamics, cost function, and constraints, an optimization problem (based on the MPC-formulation) to find the optimal control input for efficient bandwidth allocation is formulated. Let the optical input sequence obtained by solving the problem be $\{\hat{\boldsymbol{u}}_{k+\tau-1}\}_{\tau\in\mathcal{T}_p}$.

Problem 3.1 (Proposed bandwidth allocation optimization problem)**.**
For given $\boldsymbol{x}_k$, $\boldsymbol{d}_k$, and the dynamics in Eqs. (3.1) and (3.2), solve

$$\begin{aligned} &\underset{\{\boldsymbol{u}_{k+\tau-1}\}_{\tau\in\mathcal{T}_p}}{\text{minimize}} && J_k \\ &\text{subject to} && \text{Eqs. (3.1), (3.2), (3.7), (3.8)} \end{aligned}$$

The cost function and constraints are linear in $\boldsymbol{u}_k$ because the dynamics of $\boldsymbol{x}_k$ in Eq. (3.1) are linear in $\boldsymbol{u}_k$, the dynamics of $\boldsymbol{f}_d$ are independent of $\boldsymbol{x}_k$, and $\boldsymbol{f}_d$ is available for the prediction as defined in Subsection 3.2.1. Thus, Problem 3.1 is a linear programming problem and can be solved efficiently. This problem is solved by the SATCOM operator at every instant k, and the first optimal input $\hat{\boldsymbol{u}}_k$ is applied to the HTS. Then the state of the digital channelizer in the HTS is changed according to the dynamics in Eq. (3.1).

Furthermore, the following proposition shows that the solution obtained from Problem 3.1 is equivalent to that obtained from the ℓ_0-norm optimization problem. Therefore, the solution obtained by solving Problem 3.1 using the ℓ_1-norm can reduce the NCA.

Proposition 3.1. Problem 3.1 is equivalent to the ℓ_0-norm optimization problem.

Proof. Problem 3.1 is a kind of the ℓ_1-optimal control problem. The matrices $\boldsymbol{A} = \boldsymbol{I}_n$ and $\boldsymbol{B} = \boldsymbol{I}_n$ $(n = m)$ for the system in Eq. (3.1) clearly satisfy the normality condition as described in Section 2.3. Thus, Problem 3.1 is normal.

Assume the problem has at least one solution. By applying Theorem 2.1 to Problem 3.1, sets of the optimal solutions of the ℓ_0- and ℓ_1-optimal control problem are equivalent. Therefore, Problem 3.1 is equivalent to the ℓ_0-norm optimization problem. □

Remark 3.2. Note that the first term of the cost function of Problem 3.1 includes the max function, and the second term can be rewritten by using the max function as

$$\|\boldsymbol{R}\boldsymbol{u}_{k+\tau-1}\|_1 = \mathbf{1}_M^\top \cdot \max\left\{\boldsymbol{R}\boldsymbol{u}_{k+\tau-1}, -\boldsymbol{R}\boldsymbol{u}_{k+\tau-1}\right\}.$$

It is known that introducing slack variables works well to optimize the max function [113]. Thus, to solve Problem 3.1, slack variables $t \in \mathbb{R}$ and $s \in \mathbb{R}$ are introduced. Then, by using the transformation stated in [113], Problem 3.1 is reduced to the following equivalent optimization problem.

Problem 3.2 (Proposed bandwidth allocation optimization problem equivalent to Problem 3.1)**.**

For given $\boldsymbol{x}_k$, $\boldsymbol{d}_k$, and the dynamics in Eqs. (3.1) and (3.2), solve

$$\begin{aligned}
\underset{\{\boldsymbol{u}_{k+\tau-1}\}_{\tau\in\mathcal{T}_p}}{\text{minimize}} \quad & \sum_{i=1}^{M}\left(\sum_{\tau\in\mathcal{T}_p} t^i_{k+\tau} + \sum_{\mu\in\mathcal{T}_u} s^i_{k+\mu}\right) \\
\text{subject to} \quad & \text{Eqs. (3.1), (3.2), (3.7), (3.8),} \\
& q^{(i,i)}(d^i_{k+\tau} - x^i_{k+\tau}) \le t^i_{k+\tau},\ 0 \le t^i_{k+\tau}, \\
& r^{(i,i)}u^i_{k+\mu-1} \le s^i_{k+\mu},\ \ -r^{(i,i)}u^i_{k+\mu-1} \le s^i_{k+\mu}, \\
& \forall i = 1,\dots,M,\ \ \tau\in\mathcal{T}_p,\ \ \mu\in\mathcal{T}_u.
\end{aligned}$$

3.3 Numerical Experiments

In this section, the effectiveness of the proposed method is verified through numerical experiments. The MATLAB YALMIP [114] and SeDuMi [115] packages are used for solving the optimization problem.

3.3.1 Example 1: Bandwidth Allocation for Simple Time-Varying Requests

3.3.1.1 Simulation Setup

Consider an HTS with two beams (i.e., $M = 2$). The effectiveness of the proposed method is analyzed by adopting the following dynamic model of the communication requests, described as

$$\begin{aligned} \boldsymbol{d}_{k+1} &= \boldsymbol{f}_d(\boldsymbol{d}_k) \\ \begin{bmatrix} d^1_{k+1} \\ d^2_{k+1} \end{bmatrix} &= \begin{bmatrix} \alpha^1 & 0 \\ 0 & \alpha^2 \end{bmatrix} \begin{bmatrix} d^1_k \\ d^2_k \end{bmatrix}. \end{aligned} \tag{3.9}$$

The parameters (α^1, α^2) change in time interval $k \in [1, 90]$ as

$$(\alpha^1, \alpha^2) = \begin{cases} (0.99, 1.05), & k \in [1, 24], \\ (1, 1), & k \in [25, 34], \\ (1.04, 0.98), & k \in [35, 90]. \end{cases}$$

The requests monotonically increase when $\alpha > 1$, monotonically decrease when $\alpha < 1$, and remain unchanged when $\alpha = 1$. The other simulation parameters are listed in Table 3.2 (e.g., $T_p = 5$ and $T_u = 5$).

The effectiveness of the proposed method and the effect of the weighting matrices $\boldsymbol{Q}$ and $\boldsymbol{R}$ were investigated by exploring the three conditions specified in Table 3.3, where conditions A-1 and A-2 demonstrate the effect of $\boldsymbol{Q}$ and conditions A-1 and A-3 demonstrate the effect of $\boldsymbol{R}$. Conditions A-1 and A-2 do not consider the NCA. In addition, these results were compared with those obtained from fixed and conventional allocations. Specifically, the fixed allocation does not consider the frequency flexibility, and hence the initial allocation $\boldsymbol{x}_1$ is maintained over time, whereas the conventional allocation considers the quadratic cost function $\tilde{J}_k$ given by Eq. (3.4).

Each condition is evaluated by two performance indices, namely, the root-mean-square

Table 3.2: Simulation parameters to evaluate varying communication requests.

Parameter	Value
M	2
T	90
$B_{\max}$ [MHz]	500
(T_p, T_u)	$(5, 5)$
$\boldsymbol{x}_1$ [MHz]	$[250 \;\; 250]^\top$
$\boldsymbol{d}_1$ [MHz]	$[100 \;\; 100]^\top$

Table 3.3: Weighing matrices to evaluate varying communication requests.

Condition	$\boldsymbol{Q}$	$\boldsymbol{R}$
A-1	$\mathrm{diag}(500, 500)$	$\mathrm{diag}(0, 0)$
A-2	$\mathrm{diag}(100, 500)$	$\mathrm{diag}(0, 0)$
A-3	$\mathrm{diag}(500, 500)$	$\mathrm{diag}(100, 100)$

error (RMSE) and the NCA. The RMSE of the bandwidth loss is defined as

$$\mathrm{RMSE} := \sqrt{\frac{1}{T}\sum_{k=1}^{T}(L_k)^2}, \tag{3.10}$$

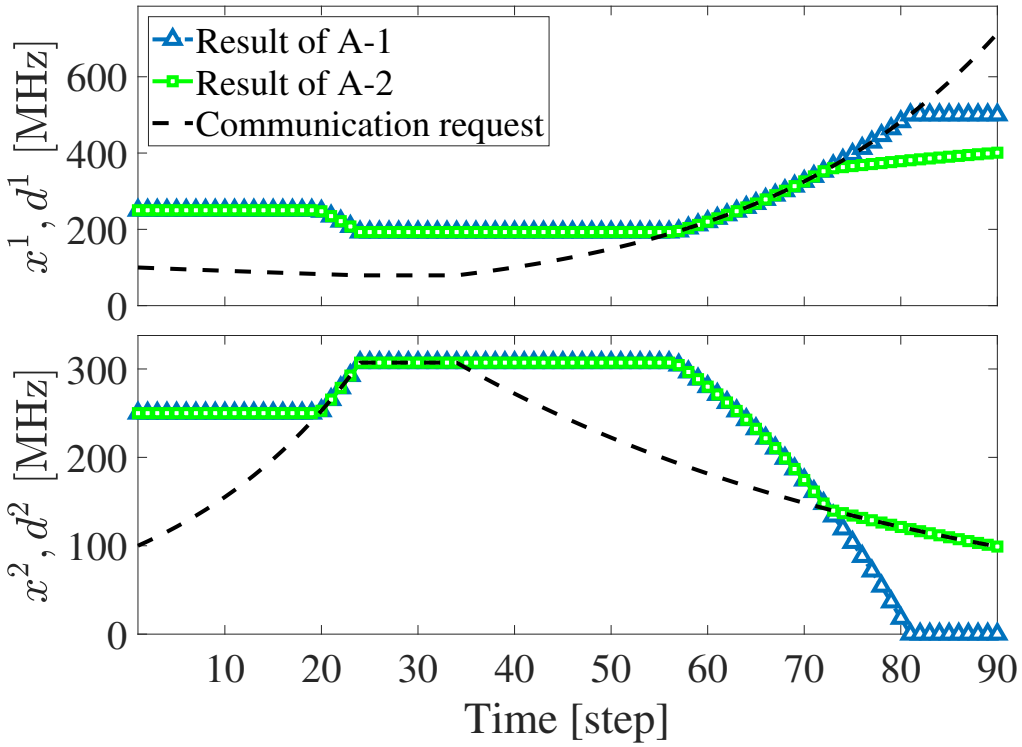
where T is the simulation time interval and L_k is defined in Eq. (3.3). The NCA of the HTS over the simulation time interval is defined as

$$\mathrm{NCA} := \sum_{k=1}^{T} \|\hat{\boldsymbol{u}}_k\|_0 \,. \tag{3.11}$$

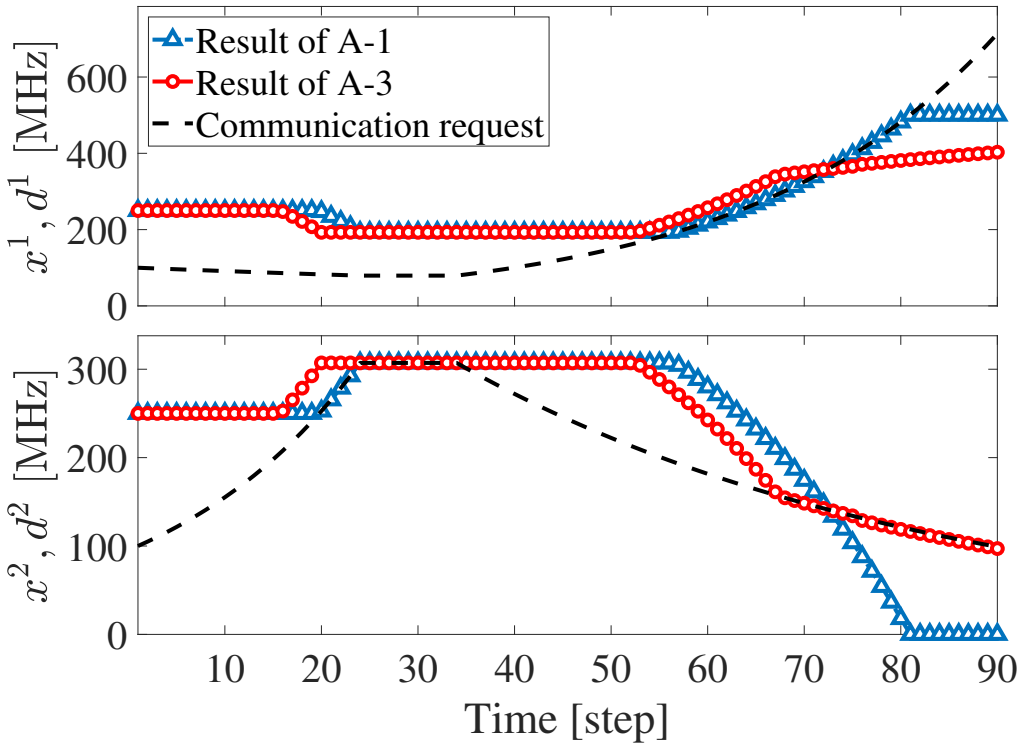
3.3.1.2 Results

Figure 3.8 shows the evolution of bandwidth allocation $\boldsymbol{x}_k$ and communication requests $\boldsymbol{d}_k$ for each condition shown in Table 3.3. The figure illustrates that bandwidth loss occurs when the black dashed line has a higher value than the colored lines. When the colored lines do not change, the bandwidth allocation remains fixed. Table 3.4 shows the results of the evaluation indices in Eqs. (3.10) and (3.11) for the compared methods.

The results in Fig. 3.8 show that the bandwidth allocation $\boldsymbol{x}_k$ increases before the onset of bandwidth loss, which does not occur until the time $k = 73$ in any condition. Because of the beam interference constraint in Eq. (3.8), the increase in the bandwidth in one beam is compensated by the decrease in the bandwidth in the other beam. Therefore, the bandwidth



(a)



(b)

Fig. 3.8: Allocated bandwidth $\boldsymbol{x}_k$ according to communication requests $\boldsymbol{d}_k$ using the proposed method with different conditions of the weighing matrices as shown in Table 3.3. (a) Comparison between conditions A-1 and A-2. (b) Comparison between conditions A-1 and A-3.

Table 3.4: Evaluation indices of the bandwidth allocation.

Condition/Method	RMSE of bandwidth loss [MHz]		
	Beam 1	Beam 2	Total
A-1	40.8	40.6	81.4
A-2	75.4	0	75.4
A-3	74.6	0.9	75.5
Fixed allocation	131.9	23.0	154.9
Conventional allocation	44.1	32.1	76.2

Condition/Method	The number of control actions		
	Beam 1	Beam 2	Total
A-1	30	30	60
A-2	39	39	78
A-3	43	43	86
Fixed allocation	0	0	0
Conventional method	74	74	148

loss for both beams does not reach zero after $k = 73$ because the communication requests exceed $B_{\max} = 500$ MHz at $k = 73$.

On the other hand, when the requests neither change nor exceed the allocated bandwidth for the prediction horizon, the allocation remains fixed, and thus no control action is performed in the HTS. Therefore, the NCAs of conditions A-1, A-2, and A-3 are fewer than that of the conventional method, which changes the allocation at every timestep.

Figure 3.8(a) compares the results of conditions A-1 and A-2, where the allocation results differ for $k > 73$. In fact, bandwidth loss does not occur in beam 2 because it has a larger weight ($q^{(2,2)} = 500$) than that of beam 1 ($q^{(1,1)} = 100$). Consequently, the RMSE of beam 2 decreases from 40.6 to 0 MHz and that of beam 1 increases from 40.8 to 75.4 MHz. Therefore, the weight $\boldsymbol{Q}$ has the effect of prioritizing the bandwidth allocation for beams with higher weights.

Figure 3.8(b) compares the results of conditions A-1 and A-3. In Condition A-3, the allocation increases by predicting the increase in the requests, thus confirming the effectiveness of the second term in the cost function, which includes the ℓ_1-norm of the input. Figure 3.9 illustrates the schematic of the optimal input sequence of conditions A-1 and A-3, $\{\hat{u}^i_{k+\tau-1}\}_{\tau\in\mathcal{T}_p}$. In condition A-1, the total amount of the input was distributed over the

prediction horizon as

$$\|\hat{\boldsymbol{u}}^i_{k|k+T_p-1}\|_1 = \sum_{\tau\in\mathcal{T}_p} \left|\hat{u}^i_{k+\tau-1}\right| = \sum_{\tau\in\mathcal{T}_p} \hat{u}^i_{k+\tau-1},$$

where

$$\hat{\boldsymbol{u}}^i_{k|k+T_p-1} := \begin{bmatrix} \hat{u}^i_k & \cdots & \hat{u}^i_{k+T_p-1} \end{bmatrix}^\top.$$

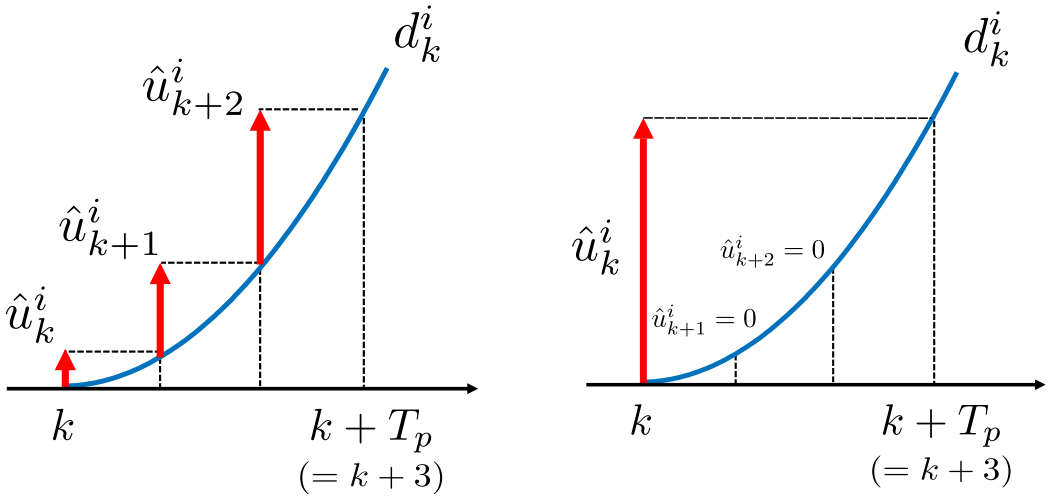
On the other hand, in condition A-3, the input sequence was applied only at the instant k. As a result, the ℓ_0-norm of the input ($\|\hat{\boldsymbol{u}}^i_{k|k+T_p-1}\|_0$) during condition A-3 is below that during condition A-1, whereas the ℓ_1-norms of the input ($\|\hat{\boldsymbol{u}}^i_{k|k+T_p-1}\|_1$) during both conditions are equal. Indeed, $\|\hat{\boldsymbol{u}}^i_{k|k+T_p-1}\|_0$ was minimized as planned in the problem formulation in Section 3.2. The bandwidth allocation increases in advance during condition A-3 because the first optimal input, $\hat{\boldsymbol{u}}_k$, which is larger than that during condition A-1, is applied. Hence, it would be possible to handle a sudden increase in requests that do not follow the model in Eq. (3.9).

The NCA during condition A-3 is slightly higher than that during condition A-1, contrary to expectations. However, the NCA is reduced if $\boldsymbol{R}$ is set to be much larger than that in condition A-3 (i.e., much larger than $\boldsymbol{R} = \mathrm{diag}(100, 100)$). Furthermore, for a sufficiently large $\boldsymbol{R}$, the allocation does not change and the result is equivalent to the case of the fixed allocation because the second term of the cost function (i.e., the ℓ_1-norm of the input) becomes dominant compared to the first term of the cost function. These results confirm that $\boldsymbol{R}$ influences bandwidth allocation in advance and reduces the NCA.

In summary, the effectiveness of the frequency flexibility and the proposed bandwidth allocation method have been verified through the simulation described in this subsection. Moreover, the proper settings of the weighing matrices $\boldsymbol{Q}$ and $\boldsymbol{R}$ allow an efficient bandwidth allocation that prioritizes either the bandwidth loss or the NCA per beam.

3.3.2 Example 2: Bandwidth Allocation for Aircraft Communication Demand

In this section, the effectiveness of the proposed method is evaluated in a practical application to the bandwidth allocation for aircraft. First, the time variation of aircraft communica-



(a) Condition A-1. (b) Condition A-3.

Fig. 3.9: Optimal input sequence $\{\hat{\boldsymbol{u}}_{k+\tau}^i\}_{\tau \in \mathcal{T}_p}$ of conditions A-1 and A-3 (i-th beam, $T_p = T_u = 3$). The ℓ_0-norm of the input during condition A-3 in (b) is lower than that during condition A-1 in (a), whereas the ℓ_1-norms of both conditions are equal. (Copyright © 2018 IEICE, [76] Fig. 8)

tion requests is analyzed by using actual flight tracking data. Then the proposed method is applied to the obtained requests.

3.3.2.1 Evolution of Aircraft Communication Demand

To understand the variation of aircraft communication requests, the author analyzed the Open Data from the Collaborative Actions for Renovation of Air Traffic Systems (CARATS) [116] provided by the Ministry of Land, Infrastructure, Transport and Tourism of Japan. These data contain actual flight tracking information including latitude, longitude, and altitude of each regularly scheduled flight around Japan and is updated every 10 seconds. The data is structured according to the following format:

Time (hh:mm:ss.s), Flight number, Latitude (°), Longitude (°), Altitude (ft), Model.

For example:

08:00:00.0, FLT0464, 35.137487, 139.147647, 23720, B763.

The CARATS Open Data for Sunday, July 19, 2015 (which includes 4,183 flights), was analyzed. Specifically, the time variation of communication requests was analyzed for beams 1 to 4 shown in Fig. 3.2.

Table 3.5: Aircraft types and required bandwidth.

Type	Required bandwidth [MHz]	Model
Large	30	**Airbus** 330-{200,300}, 340-{300, 500, 600}, 380-800 **Boeing** 747-{400, 8}, 777-{200, 200LR, 300, 300ER}, MD-11
Medium	10	**Airbus** 300-600, **Boeing** 757-200, 767-300, 787-{8, 9}
Small	5	**Airbus** 319, 320, 321, **Boeing** 737-{400, 500, 700, 800, 900}, MD-90
Regional Jet	0	**Bombardier** CRJ-{200, 700}, **Embraer** E-{170, 190}, **Sukhoi** SSJ-100
Turboprop	0	**Bombardier** DHC8-{Q100, Q200, Q300, Q400}, **Saab** SF340

Aircraft models were divided into five types, namely, large, medium, small, regional jet, and turboprop as listed in Table 3.5. In this analysis, the following assumptions are made: (i) half of the 4,183 aircraft are equipped with in-flight Wi-Fi (this assumption is based on the forecast of the in-flight Wi-Fi equipping ratio in 2025 mentioned in Subsection 1.2.1); (ii) large, medium, and small aircraft require bandwidths of 30, 10, and 5 MHz, respectively, whereas regional jets and turboprops do not require bandwidth (these values are larger than the current bandwidth allocated for in-flight Wi-Fi [14]); (iii) aircraft request bandwidth at altitudes above 10,000 ft (this assumption is reasonable according to the restrictions of the use of portable electronic devices within aircraft [117]).

Figure 3.10 shows the time evolution of the communication demand from each aircraft type for all four beams, in intervals of 10 minutes. Half of the aircraft (i.e., 2,092) were randomly selected for this estimation and each of them are assumed to have in-flight Wi-Fi. The communication demand in beam 2 is the largest because the Tokyo International Airport is included in beam 2. In all the beams, the communication demand varies temporally. For example, the request demand in beam 2 suddenly increases at around 18:00 and the demand decreases in adjacent beam 3 at around 10:00 and 19:00. Such situations enhance the capabilities of the frequency flexibility adapting to the evolution of the different communication demands among adjacent beams.

3.3.2.2 Results

To verify the effectiveness of the proposed method in the practical aircraft scenario, the bandwidth demand depicted in Fig. 3.10 is applied instead of using the dynamics $\boldsymbol{f}_d$ in Eq. (3.2). The simulation parameters are set as listed in Table 3.6 (e.g., $T_p = 5$ and $T_u = 5$), and two conditions detailed in Table 3.7 are compared. The aim of this simulation is to compare the proposed method to the fixed allocation and to evaluate the effect of the

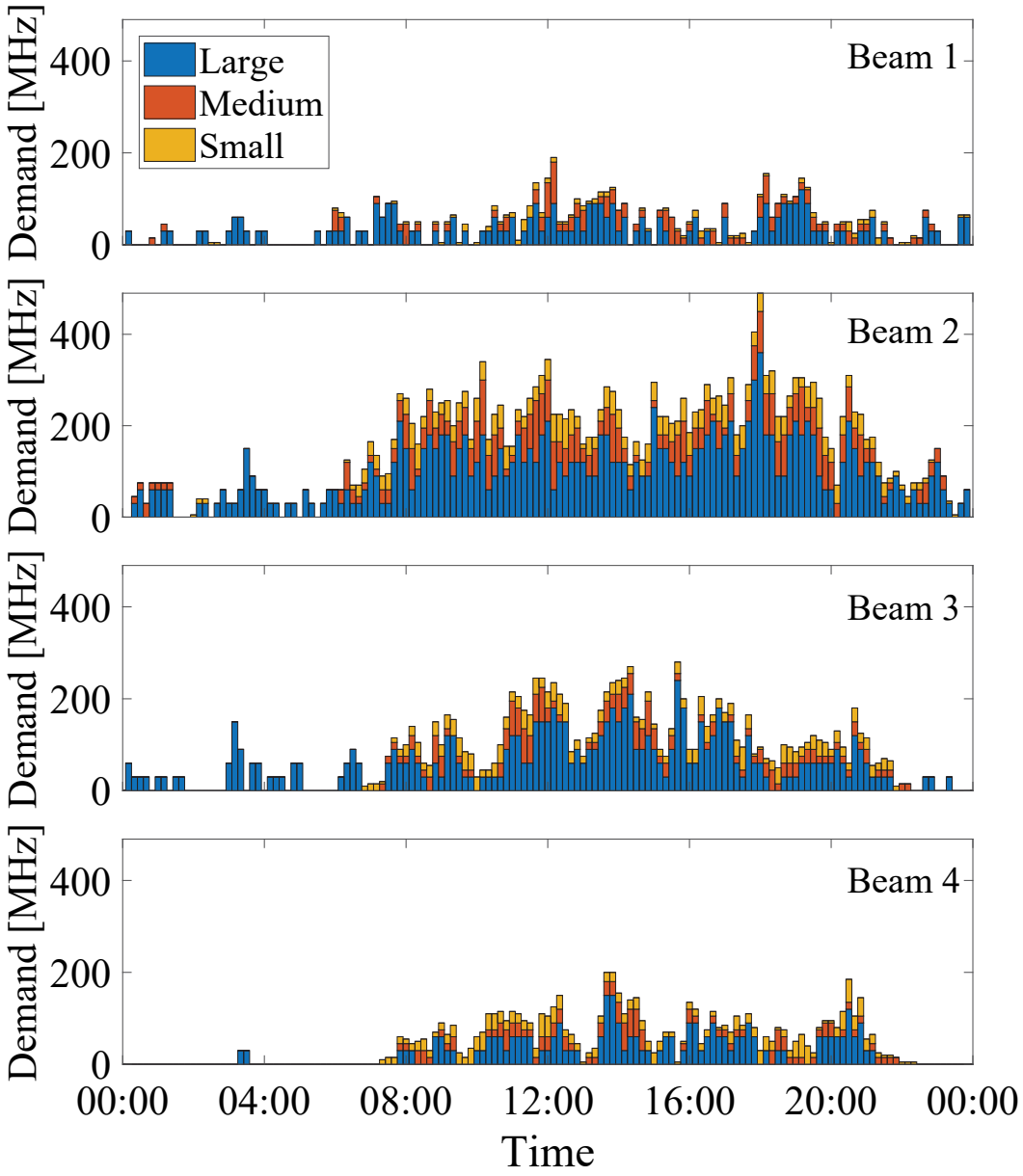


Fig. 3.10: Time variation of aircraft communication demand in the four beams according to aircraft type. (Copyright © 2018 IEICE, [76] Fig. 9)

Table 3.6: Simulation parameters for aircraft communication demand evaluation.

Parameter	Value
M	4
T	145
$B_{\max}$ [MHz]	500
(T_p, T_u)	$(5, 5)$
$\boldsymbol{x}_1$ [MHz]	$[250 \;\; 250 \;\; 250 \;\; 250]^\top$

Table 3.7: Weighting matrices for aircraft communication demand evaluation.

Condition	$\boldsymbol{Q}$	$\boldsymbol{R}$
B-1	$\mathrm{diag}(220, 520, 285, 210)$	$\mathrm{diag}(0, 0, 0, 0)$
B-2		$\mathrm{diag}(300, 300, 300, 300)$

weighting matrix $\boldsymbol{R}$ compared to the fixed weighting matrix $\boldsymbol{Q}$. The matrix $\boldsymbol{Q}$ is determined based on the largest value of the bandwidth demand of each beam, as shown in Fig. 3.10.

The results of evaluating the proposed method using the aircraft communication demand are shown in Fig. 3.11, which shows the evolution of the bandwidth allocation $\boldsymbol{x}_k$ and communication requests $\boldsymbol{d}_k$ under conditions B-1 and B-2. In addition, Table 3.8 lists the evaluation indices in Eqs. (3.10) and (3.11) for the compared methods. These results show a similar behavior to that of the outcomes from Subsection 3.3.1. In fact, the variation of the requests was predicted accurately, and the allocated bandwidth was increased when a large bandwidth loss would occur over the prediction horizon. Moreover, the bandwidth allocation remained fixed when $\boldsymbol{x}_k$ exceeded $\boldsymbol{d}_k$ over the prediction horizon. However, for beams adjacent to those with increased allocation, bandwidth loss increased due to the constraint in Eq. (3.8). Overall, bandwidth allocation during condition B-2 exhibited less variation than that during condition B-1 (i.e., the NCA during condition B-2 was below that during condition B-1). This result confirms that the weighting matrix $\boldsymbol{R}$ reduces the NCA (in this case, from 71 to 26). Nevertheless, the RMSE of beams 1 and 2 and the total RMSE were higher during condition B-2 than during condition B-1. Therefore, there exists a tradeoff between the bandwidth loss and the NCA, and the balance can be adjusted by the weighting matrices $\boldsymbol{Q}$ and $\boldsymbol{R}$.

Finally, to illustrate the performance of the proposed method in each condition, Fig. 3.12 shows the bandwidth loss of beam 2 during the interval from $k = 50$ to $k = 145$ for con-

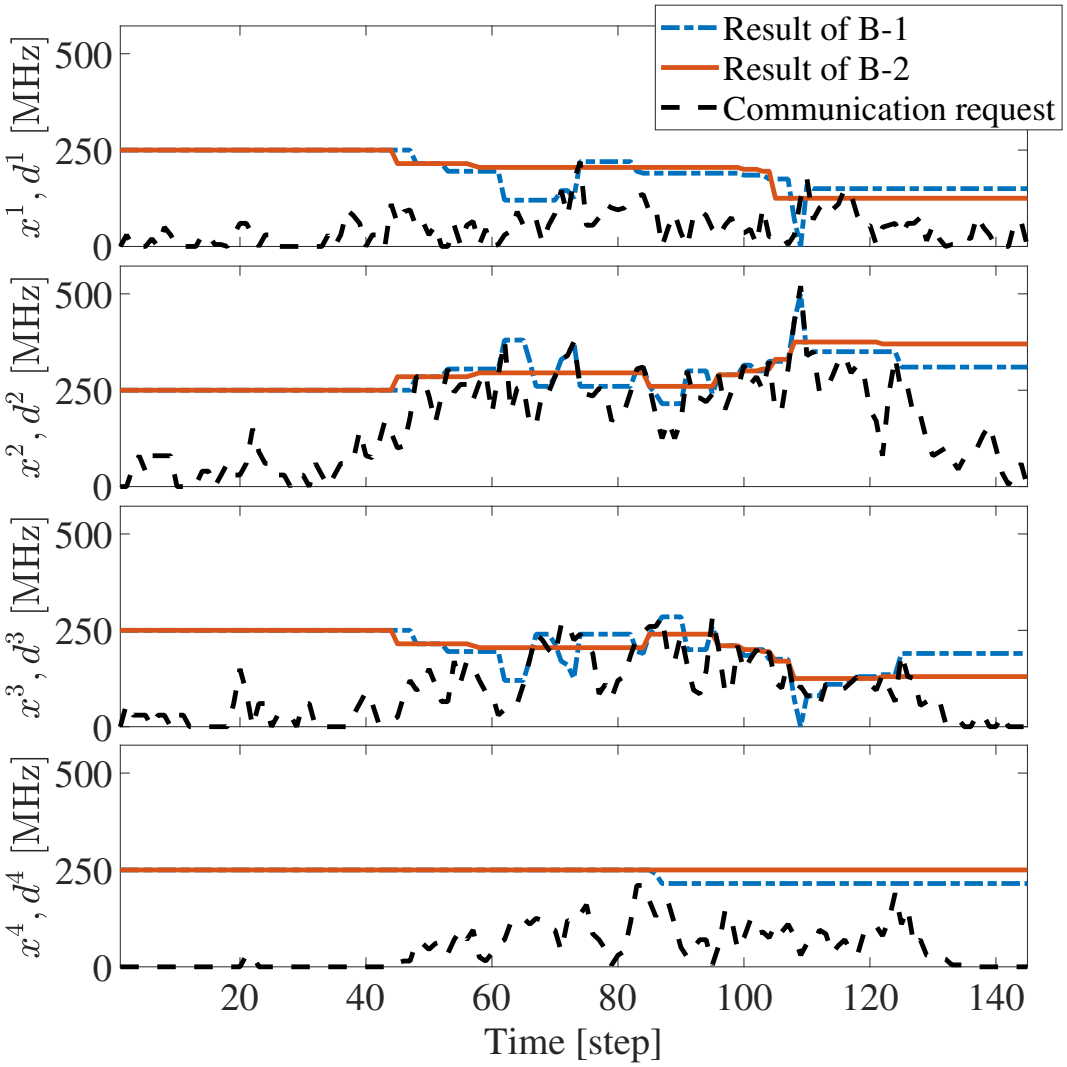


Fig. 3.11: Allocated bandwidth $\boldsymbol{x}_k$ according to demand $\boldsymbol{d}_k$ of aircraft.

ditions B-1, B-2, and the fixed allocation. This figure and Fig. 3.11 suggest that bandwidth loss tends to occur when the demand increases rapidly. Moreover, the fixed allocation shows a higher bandwidth loss than conditions B-1 and B-2 for the entire $50 < k < 145$ interval. This verifies the effectiveness of the frequency flexibility in a practical scenario considering aircraft demand.

Note that the effective use of the frequency flexibility depends on the HTS configuration, including maximum bandwidth $B_{\max}$ and beam arrangement as shown in Fig. 3.2. Furthermore, it is important to accurately predict the communication demand for each SATCOM application. The performance of the frequency flexibility is influenced by the dynamics of the requests in Eq. (3.2).

Table 3.8: Evaluation indices of the bandwidth allocation using the estimated communication demand from aircraft.

Condition/Method	RMSE of bandwidth loss [MHz] Beam No. 1	2	3	4	Total
B-1	11.1	1.7	18.2	0	30.1
B-2	4.9	17.4	12.3	0	34.6
Fixed allocation	0	40.5	4.8	0	45.3

Condition/Method	The number of control actions Beam No. 1	2	3	4	Total
B-1	14	26	29	2	71
B-2	6	10	10	0	26
Fixed allocation	0	0	0	0	0

3.4 Discussions on Performance

3.4.1 Parameter Tuning of Weighting Matrices

For the weighting matrices $\boldsymbol{Q}$ and $\boldsymbol{R}$, a larger weight $q^{(i,i)}$ makes the bandwidth loss of the i-th beam less, and a larger weight $r^{(i,i)}$ makes the input for the i-th beam sparser (i.e., the NCA is smaller), as mentioned in Subsection 3.2.2. Thus, it is necessary to properly tune the weights $q^{(i,i)}$ and $r^{(i,i)}$ based on the purpose and situation because a tradeoff exists between the bandwidth loss and the NCA. Furthermore, if the time variation of the communication requests is known or can be predicted in advance, the matrix $\boldsymbol{Q}$ should be chosen according to this prior information as determined in the experiment in Subsection 3.3.2.

The performance of specific beams or the effect of the frequency flexibility could be negated if the parameters are chosen inappropriately. For example, the following results were obtained in the experiment in Subsection 3.3.1:

- By setting $\boldsymbol{Q} = \mathrm{diag}(500, 0)$ and $\boldsymbol{R} = \mathrm{diag}(0, 0)$, which means ignoring the performance of beam 2, the bandwidth allocation of beam 2 is not increased at all even though the bandwidth loss occurs in beam 2.

- By setting $r^{(i,i)}$ over 1,250 for $i = 1, 2$ and $\boldsymbol{Q} = \mathrm{diag}(500, 500)$, the result is the same as the fixed allocation.

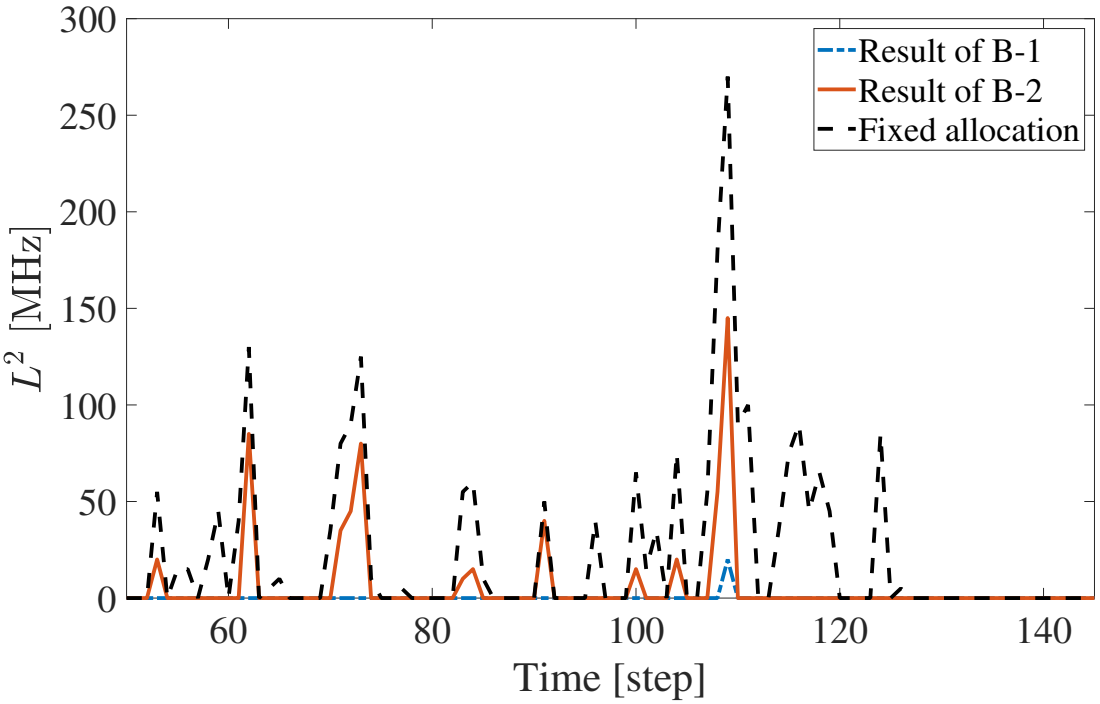


Fig. 3.12: Bandwidth loss of beam 2 for fixed allocation and the proposed method under conditions B-1 and B-2.

The author further analyzes the effect of the weighing matrix $\boldsymbol{R}$ used in the experiment described in Subsection 3.3.2. The matrix was set to $\boldsymbol{R} = r\boldsymbol{I}_4$ and r was increased from 0 to 1000 in increments of 50. The weighting matrix for the bandwidth loss was set to $\boldsymbol{Q} = \mathrm{diag}(220, 520, 285, 210)$ for all r as used in Subsection 3.3.2.

Figure 3.13 represents the relationship between the total bandwidth loss and the total NCA for each r setup. This indicates that the total NCA is decreased by setting a large value for r, although the total bandwidth loss is increased. Furthermore, when $r \geq 900$, the NCA was 0 and the result was the same as the fixed allocation.

This result can be applied to the selection of appropriate weighting matrices for the operation of the SATCOM system. For example, consider a situation where the SATCOM operator tries to reduce NCA less than 30. In this example, $r = 300$ was the best of these r settings for the operator because the total NCA was 26.

3.4.2 Computation Time

In the simulations described in Subsections 3.3.1 and 3.3.2, computation times were 19.7 s for 90 steps in total (i.e., 0.219 s/step) and 28.2 s for 145 steps in total (i.e., 0.194 s/step),

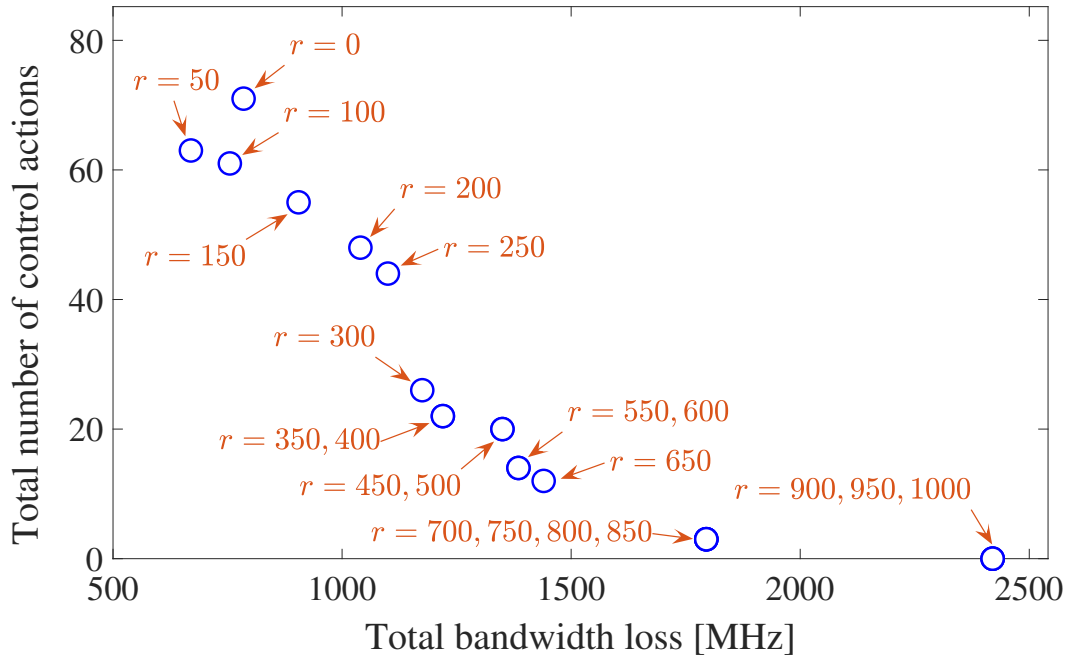


Fig. 3.13: Relationship between the total bandwidth loss and the total number of control actions for each r.

respectively. These simulations were conducted using an Intel(R) Core(TM) i7-7560U CPU at 2.4 GHz with 8 GB RAM.

The computation time for each step depends on M, T_p, and T_u (especially T_p and T_u). Thus, as the value of these parameters increases, the computation time increases correspondingly. However, the proposed bandwidth allocation problem can be solved online in the SATCOM system because the control period of the HTS will be relatively long (e.g., 10 minutes) compared to the computation time due to hardware constraints. This computation is conducted by the RNMC in the SATCOM operator as shown in Fig. 3.1.

3.5 Summary

In this chapter, the author proposed a bandwidth allocation method for SATCOM systems with the frequency flexibility based on MPC and sparse optimization. In this proposed method, the dynamics of the HTS and communication requests were incorporated into the problem formulation. Then, the effectiveness of the proposed method was verified through two numerical experiments: one that used a simple time-varying model and another that used predicted aircraft communication requests determined from actual flight tracking data.

This evaluation demonstrated that efficient bandwidth allocation can be expressed in terms of the bandwidth loss and the NCA. Furthermore, the proposed method allows adjusting control objectives and beam priorities by tuning the weighting matrices. Specifically, the bandwidth loss of a beam decreases (and the NCA increases) when the weight $q^{(i,i)}$ is set to a relatively large value, whereas the NCA decreases (and the bandwidth loss increases) when the weight $r^{(i,i)}$ is set to a relatively large value.

For efficient allocation, a dynamical model of the communication requests is assumed to be available (i.e., the time variation of the requests can be predicted). However, in general, this model cannot be fully obtained in advance because it is difficult to completely predict the future requests of a SATCOM system with a wide variety of applications. Therefore, future work will include extending the allocation method to handle the uncertainty in the time variation of communication requests. In addition, the author aims to explore a scheme to determine the bandwidth allocation based on a trend of communication requests when the model is completely unknown.

Chapter 4

Management Framework for Massive Connected Networks

This chapter is organized as follows. In Section 4.1, the system description of the large-scale SATCOM system is described. In Section 4.2, design variables, constraints, and cost functions are defined, and the optimization problem is formulated as a mixed integer programming problem to obtain the resource allocation and network structure. In Section 4.3, the effectiveness of the proposed method is verified through four examples of numerical experiments. Finally, this chapter is summarized in Section 4.4.

4.1 System Model

Figure 4.1 shows the large-scale SATCOM systems focused on in this chapter. User terminals (UTs), such as aircraft and ships, request communication links to satellites (SATs); in turn, the satellites provide the links for the UTs. Communication data from the UTs are aggregated in the satellites and sent to gateway stations (GWs), which are connected to the ground network. If some satellites cannot directly send the communication data to the GWs due to physical or bandwidth limitations, the satellites are allowed to communicate with another satellite to reroute the data. A SATCOM operator is expected to flexibly design and manage a network composed of user links, inter-satellite links, and feeder links to meet communication requests from UTs. In this chapter, a network composed of user links, inter-satellite links, and feeder links are collectively called a SATCOM network.

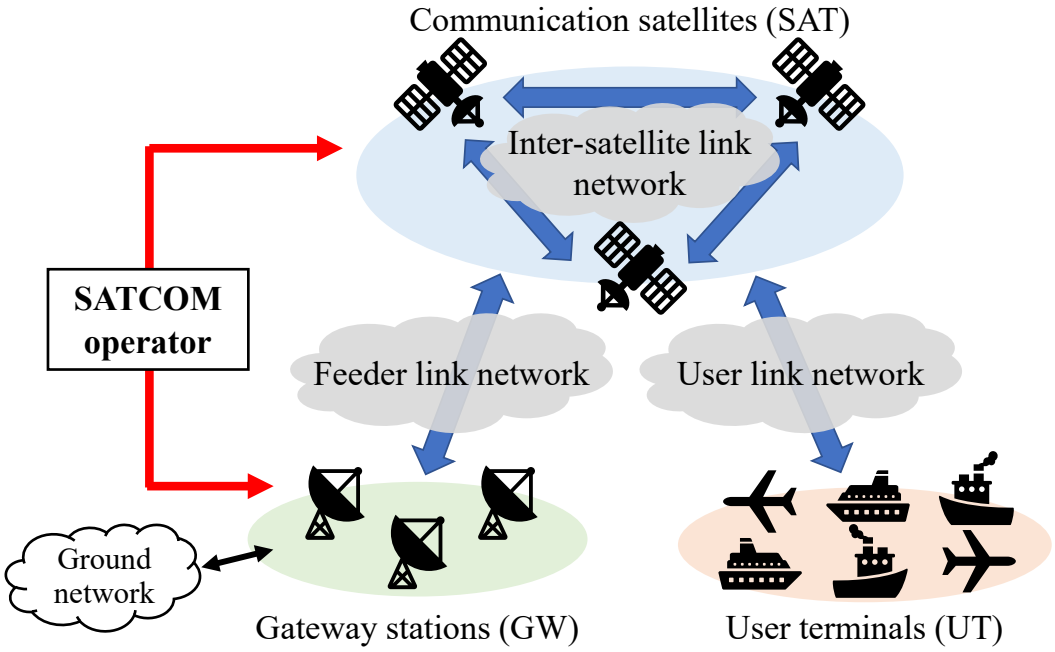


Fig. 4.1: Schematic of large-scale SATCOM systems with massive connected networks. This system includes communication satellites (SATs), user terminals (UTs), gateway stations (GWs), and a SATCOM operator that manages and controls the satellites and GWs.

Let $N_{\mathrm{U}}, N_{\mathrm{S}}$, and N_{G} be the number of UTs, satellites, and GWs in the large-scale SATCOM system, respectively. Let index sets of the UTs, satellites, and GWs be denoted as $\mathcal{N}_{\mathrm{U}} = \{1, 2, \ldots, N_{\mathrm{U}}\}$, $\mathcal{N}_{\mathrm{S}} = \{1, 2, \ldots, N_{\mathrm{S}}\}$, and $\mathcal{N}_{\mathrm{G}} = \{1, 2, \ldots, N_{\mathrm{G}}\}$, respectively. Throughout this chapter, the indices i, j, and ℓ are used to denote UTs, satellites, and GWs, respectively. The maximum bandwidth of the j-th satellite in user links and inter-satellite links is denoted by $B_{\mathrm{S}}^{j} \geq 0$ and $B_{\mathrm{ISL}}^{j} \geq 0$, respectively, and that of the ℓ-th GW is denoted by $B_{\mathrm{G}}^{\ell} \geq 0$. For each time $k \geq 0$, each UT sends a communication request to any satellite, where the amount of the communication request from the i-th UT is denoted by $d_k^i \geq 0$.

For user links (UT-SAT pairs), inter-satellite links (SAT-SAT pairs), and feeder links (SAT-GW pairs), the following connectable candidate sets are defined.

Definition 4.1. The sets of connectable pair candidates in user links, inter-satellite links, and feeder links at time k are defined as

$$\bar{\mathcal{C}}_k := \{(i, j) \in \mathcal{N}_{\mathrm{U}} \times \mathcal{N}_{\mathrm{S}} \mid i\text{-th UT and } j\text{-th satellite can be connected at time } k.\}, \quad (4.1a)$$

$$\bar{\mathcal{H}}_k := \{(j_1, j_2) \in \mathcal{N}_{\mathrm{S}} \times \mathcal{N}_{\mathrm{S}} \mid j_1\text{-th and } j_2\text{-th satellites can be connected at time } k.\}, \quad (4.1b)$$

$$\bar{\mathcal{E}}_k := \{(j,\ell) \in \mathcal{N}_{\mathrm{S}} \times \mathcal{N}_{\mathrm{G}} \mid j\text{-th satellite and } \ell\text{-th GW can be connected at time } k.\}, \tag{4.1c}$$

respectively.

Assume that sets $\bar{\mathcal{C}}_k$, $\bar{\mathcal{H}}_k$, and $\bar{\mathcal{E}}_k$ are known by the SATCOM operator in advance. These sets are determined based on the satellite position and the link budget analysis introduced in Subsection 2.1.3. The sets change over time due to time-variability in the satellite position and propagation environment. For example, because NGEO satellites revolve around the earth periodically on a prescribed orbit, the operator can determine the sets in advance. During operation of the SATCOM system, the operator can update the sets using channel state information from the UTs and GWs.

Figure 4.2(a) shows an example of the SATCOM system at time k. Suppose that $N_{\mathrm{U}} = 6, N_{\mathrm{S}} = 4$, and $N_{\mathrm{G}} = 2$ and each satellite beam is a different shape and size. In Fig. 4.2(b), the candidate sets of UT-SAT and SAT-GW pairs are illustrated. In this example, $\bar{\mathcal{C}}_k$, $\bar{\mathcal{H}}_k$, and $\bar{\mathcal{E}}_k$ are represented by

$$\bar{\mathcal{C}}_k = \{(1,1),(1,3),(2,1),(2,3),(3,1),(3,2),(3,3),(4,3),(5,3),(5,4),(6,3),(6,4)\},$$
$$\bar{\mathcal{H}}_k = \{(2,3)\},$$
$$\bar{\mathcal{E}}_k = \{(1,1),(1,2),(2,2),(3,1),(3,2),(4,2)\}.$$

The UT-SAT, SAT-SAT, and SAT-GW pairs to be connected are chosen from these candidates.

4.2 Problem Formulation

In this section, design variables and constraints are presented, and cost functions relevant to the SATCOM system are defined. After this, the optimization problem is formulated to determine the resource allocation and network structure for the large-scale SATCOM system.

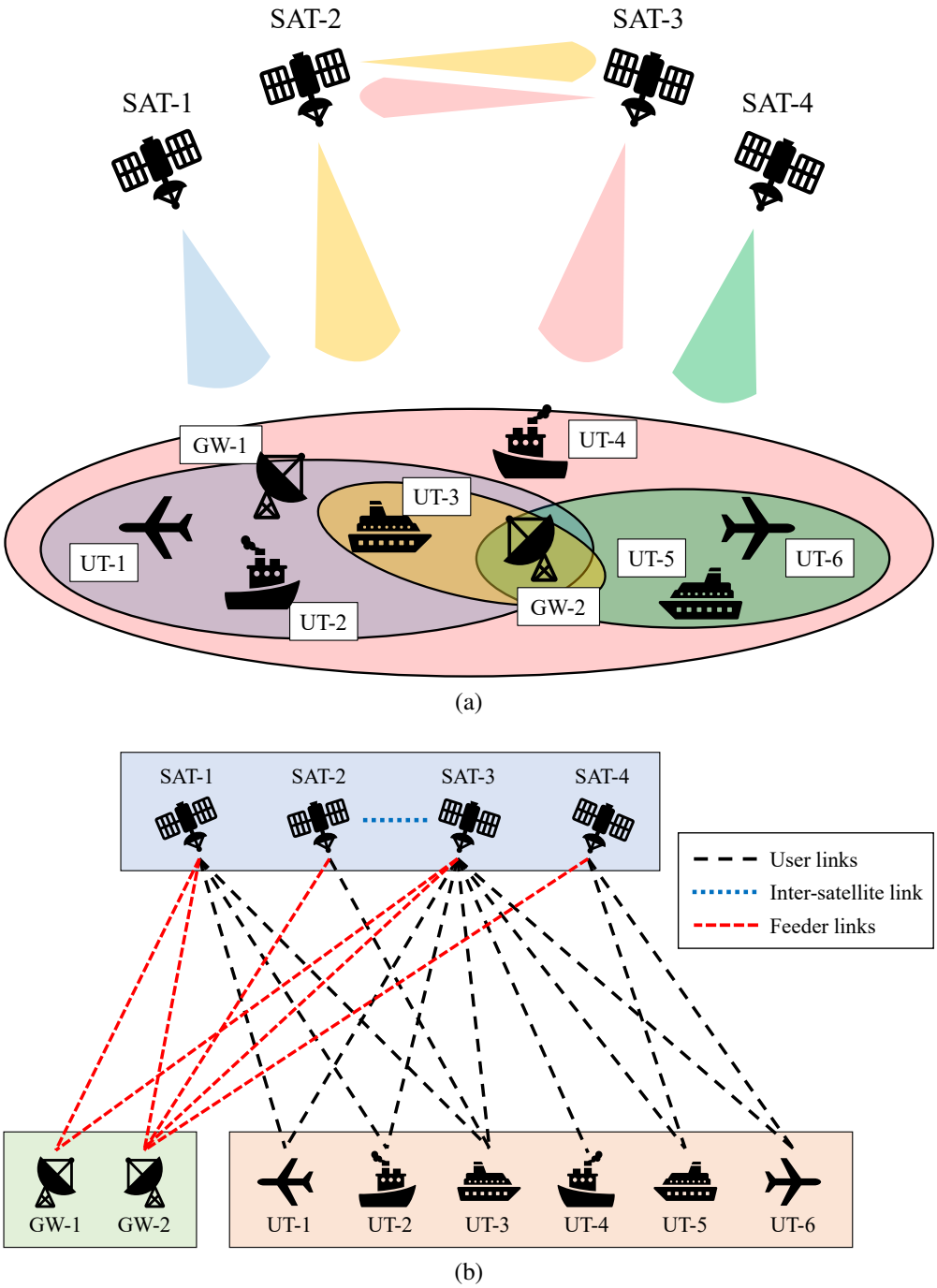


Fig. 4.2: (a) Example of the SATCOM system. (b) Connectable pair candidates in user links, inter-satellite links, and feeder links derived from (a), where the black, blue dotted, and red dashed lines represent the connectable candidate $\bar{\mathcal{C}}_k$, $\bar{\mathcal{H}}_k$ and $\bar{\mathcal{E}}_k$, respectively.

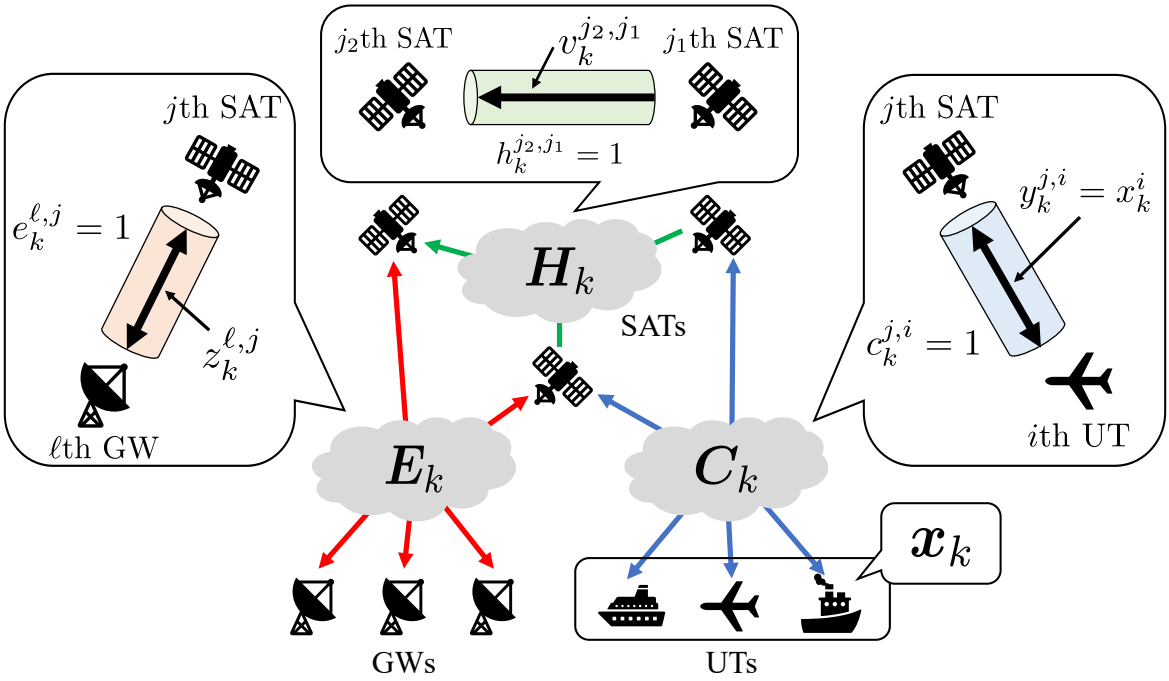


Fig. 4.3: Notation of the resource allocation and network structure in large-scale SATCOM systems.

4.2.1 Design Variables of SATCOM Systems

In this subsection, design variables of the SATCOM network structure that the SATCOM operator plans to use are defined. Notations appearing in this subsection are summarized in Fig. 4.3.

In the following problem formulation, assume that the operator of the SATCOM network can design the following two types of variables: (i) the connectivity between system components (i.e., UTs, satellites, and GWs) and (ii) the amount of data flowing over the network connections. To represent the connectivity, the following three types of network connection matrices are defined:

- The matrix of the bidirectional user link network at time k is denoted by $\boldsymbol{C}_k = [c_k^{j,i}]_{j,i} \in \{0,1\}^{N_\mathrm{S} \times N_\mathrm{U}}$. An entry of $\boldsymbol{C}_k$ is represented by

$$c_k^{j,i} := \begin{cases} 1, & \text{if the } i\text{-th UT and } j\text{-th satellite are connected,} \\ 0, & \text{otherwise.} \end{cases} \tag{4.2}$$

- The matrix of the bidirectional inter-satellite link network at time k is denoted by $\boldsymbol{H}_k =$

$[h_k^{j_2,j_1}]_{j_1,j_2} \in \{0,1\}^{N_\mathrm{S}\times N_\mathrm{S}}$. An entry of $\boldsymbol{H}_k$ is represented by

$$h_k^{j_2,j_1} := \begin{cases} 1, & \text{if there is an inter-satellite link between the } j_1\text{-th satellite} \\ & \text{and the } j_2\text{-th satellite,} \\ 0, & \text{otherwise.} \end{cases} \tag{4.3}$$

If there are no inter-satellite links, $\boldsymbol{H}_k = \boldsymbol{O}_{N_\mathrm{S}}$ holds.

Remark 4.1. The inter-satellite links are bidirectional because satellites transmit and receive not only the data, but also some notifications such as an acknowledgment based on a communications protocol. In this study, however, only the links of the data as defined in Eq. (4.3) are considered. Furthermore, the inter-satellite links are established in both links from GWs to UTs (called a forward link) and links from UTs to GWs (called a return link). Thus, the matrix $\boldsymbol{H}_k$ should be defined for both the forward and return links. In this study, the author assumes that the data flow in the same satellite path for both the forward and return links, and therefore, the matrix $\boldsymbol{H}_k$ is defined for only the one link.

- The matrix of the bidirectional feeder link network at time k is denoted by $\boldsymbol{E}_k = [e_k^{\ell,j}]_{\ell,j} \in \{0,1\}^{N_\mathrm{G}\times N_\mathrm{S}}$. An entry of $\boldsymbol{E}_k$ is represented by

$$e_k^{\ell,j} := \begin{cases} 1, & \text{if the } j\text{-th satellite and } \ell\text{-th GW are connected,} \\ 0, & \text{otherwise.} \end{cases} \tag{4.4}$$

By using these matrices, let $\mathcal{C}_k, \mathcal{H}_k$, and $\mathcal{E}_k$ denote sets of the connected pairs in each network, defined as

$$\begin{aligned} \mathcal{C}_k &:= \{(i,j) \in \bar{\mathcal{C}}_k \mid c_k^{j,i} = 1\}, \\ \mathcal{H}_k &:= \{(j_1,j_2) \in \bar{\mathcal{H}}_k \mid h_k^{j_2,j_1} = 1\}, \\ \mathcal{E}_k &:= \{(j,\ell) \in \bar{\mathcal{E}}_k \mid e_k^{\ell,j} = 1\}, \end{aligned}$$

respectively.

Assume that UTs can connect with up to one satellite through the user links and satellites can connect with up to one GW through the feeder links, while satellites are allowed to connect with multiple other satellites through the inter-satellite links. These assumptions

imply that the following constraints hold:

$$\sum_{j:(i,j)\in\mathcal{C}_k} c_k^{j,i} \in \{0,1\}, \ \ \forall i \in \mathcal{N}_{\mathrm{U}}, \tag{4.5a}$$

$$\sum_{\ell:(j,\ell)\in\mathcal{E}_k} e_k^{\ell,j} \in \{0,1\}, \ \ \forall j \in \mathcal{N}_{\mathrm{S}}. \tag{4.5b}$$

Next, the amount of data flowing through the connections are described. The amount of bandwidth that the operator plans to allocate to the i-th UT at time k is denoted by

$$x_k^i \geq 0, \ \ \forall i \in \mathcal{N}_{\mathrm{U}}. \tag{4.6}$$

For the N_{U} UTs, the allocated bandwidth can be written in vector form as

$$\boldsymbol{x}_k := \begin{bmatrix} x_k^1 & x_k^2 & \cdots & x_k^{N_{\mathrm{U}}} \end{bmatrix}^\top \in \mathbb{R}^{N_{\mathrm{U}}}.$$

Assume that, at each time $k \geq 0$, the operator can either increase or decrease the bandwidth allocated to users by adjusting an input. Therefore, the dynamics of the resource allocation is given by the following discrete-time linear system:

$$\boldsymbol{x}_{k+1} = \boldsymbol{x}_k + \boldsymbol{u}_k. \tag{4.7}$$

Here, $\boldsymbol{u}_k \in \mathbb{R}^{N_{\mathrm{U}}}$ is the control input applied at time k. The allocated bandwidth x_k^i should be larger than the requested bandwidth d_k^i because otherwise UTs cannot communicate with satellites.

The communication data from the UTs flow through the communication links by utilizing the allocated bandwidth. For the user links and the feeder links, let $y_k^{j,i}$ and $z_k^{\ell,j}$ denote the amount of the flow in each link between the i-th UT and j-th satellite and that between the j-th satellite and ℓ-th GW at time k, respectively. For the inter-satellite links, the amount of the flow from the j_1-th satellite to the j_2-th satellite is represented by $v_k^{j_2,j_1}$. In terms of these variables, the following data flow matrices can be defined: $\boldsymbol{Y}_k := [y_k^{j,i}]_{j,i} \in \mathbb{R}^{N_{\mathrm{S}}\times N_{\mathrm{U}}}$, $\boldsymbol{Z}_k := [z_k^{\ell,j}]_{\ell,j} \in \mathbb{R}^{N_{\mathrm{G}}\times N_{\mathrm{S}}}$, and $\boldsymbol{V}_k := [v_k^{j_2,j_1}]_{j_2,j_1} \in \mathbb{R}^{N_{\mathrm{S}}\times N_{\mathrm{S}}}$.

The following capacity constraints for the user links and the feeder links are required to

hold:

$$\sum_{i:(i,j)\in\mathcal{C}_k} y_k^{j,i} \leq B_{\mathrm{S}}^j, \quad \forall j \in \mathcal{N}_{\mathrm{S}}, \tag{4.8a}$$

$$\sum_{j:(j,\ell)\in\mathcal{E}_k} z_k^{\ell,j} \leq B_{\mathrm{G}}^\ell, \quad \forall \ell \in \mathcal{N}_{\mathrm{G}}. \tag{4.8b}$$

Here, B_{S}^j and B_{G}^k represent the maximum bandwidths of the j-th satellite and ℓ-th GW, respectively. Furthermore, the sum of the inflow and outflow in each inter-satellite link must satisfy

$$v_k^{j,\mathrm{in}} + v_k^{j,\mathrm{out}} \leq B_{\mathrm{ISL}}^j, \quad \forall j \in \mathcal{N}_{\mathrm{S}}, \tag{4.9}$$

where $v_k^{j,\mathrm{in}}$ and $v_k^{j,\mathrm{out}}$ represent the inflow and outflow of the j-th satellite with respect to the inter-satellite links defined by

$$v_k^{j,\mathrm{in}} := \sum_{\substack{j_1:(j_1,j)\in\mathcal{H}_k \\ j_1\neq j}} v_k^{j,j_1}, \quad v_k^{j,\mathrm{out}} := \sum_{\substack{j_2:(j,j_2)\in\mathcal{H}_k \\ j_2\neq j}} v_k^{j_2,j},$$

respectively, and B_{ISL}^j represents the maximum processing bandwidth that can be used for the inter-satellite links in the j-th satellite.

To ensure the full and efficient use of resources, the SATCOM network is designed to conserve the inflow and outflow of each link. Thus, the following conservation laws are required to hold true:

$$x_k^i = \sum_{j:(i,j)\in\mathcal{C}_k} y_k^{j,i}, \quad \forall i \in \mathcal{N}_{\mathrm{U}}, \tag{4.10a}$$

$$v_k^{j,\mathrm{in}} + \sum_{i:(i,j)\in\mathcal{C}_k} y_k^{j,i} = v_k^{j,\mathrm{out}} + \sum_{\ell:(j,\ell)\in\mathcal{E}_k} z_k^{\ell,j}, \ \forall j \in \mathcal{N}_{\mathrm{S}}. \tag{4.10b}$$

In a case where all the satellites have no payloads for inter-satellite links or the operator decides not to use the links, Eq. (4.10b) becomes

$$\sum_{i:(i,j)\in\mathcal{C}_k} y_k^{j,i} = \sum_{\ell:(j,\ell)\in\mathcal{E}_k} z_k^{\ell,j}, \ \forall j \in \mathcal{N}_{\mathrm{S}},$$

which applies only to user links and feeder links.

Schematic of the constraints defined in this section are displayed in Fig. 4.4.

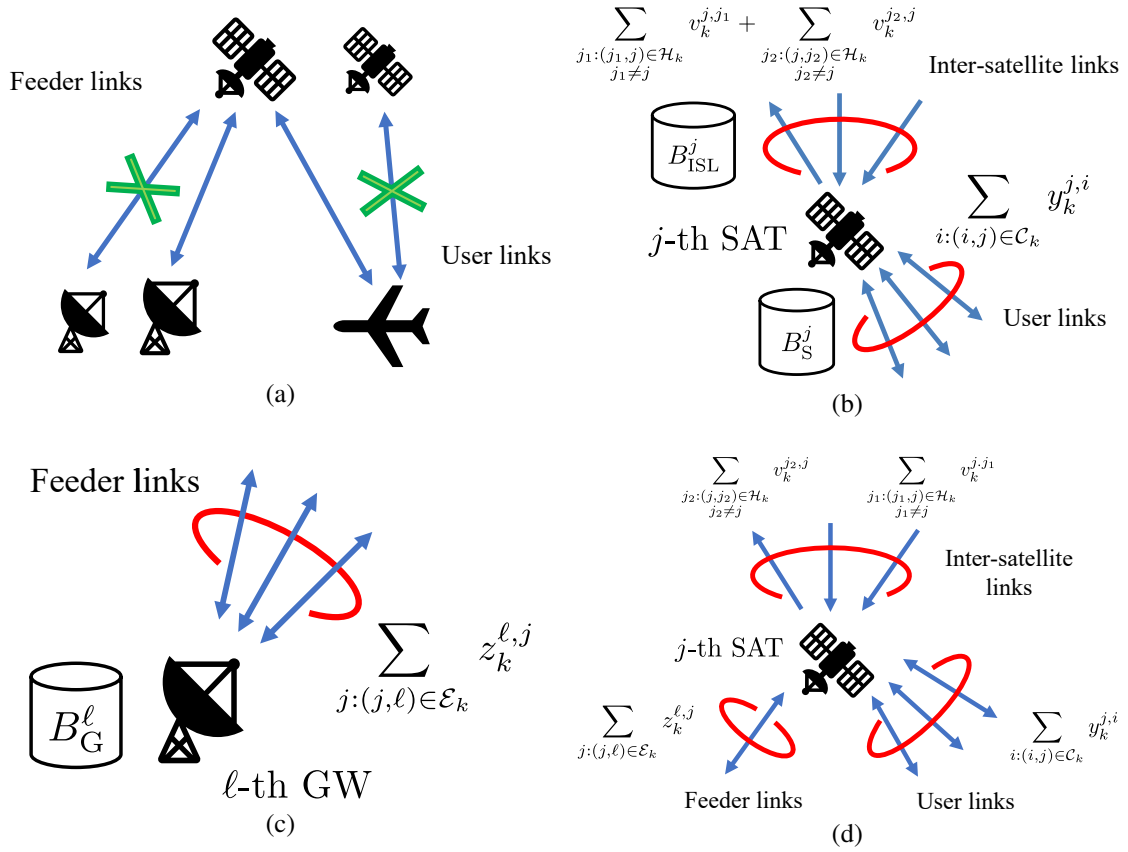


Fig. 4.4: Schematics of the constraints defined in this section. (a) Eq. (4.5), (b) Eqs. (4.8a) and (4.9), (c) Eq. (4.8b), and (d) Eq. (4.10b).

4.2.2 Cost Functions

In this subsection, the author introduces candidate factors for the cost function to ensure a cost-efficient and flexible design of the SATCOM system. These candidates also include management strategy factors related to the SATCOM operator.

1. ***Bandwidth loss***: The first factor measures the bandwidth loss of the UTs and is defined as

$$L_k := \sum_{i \in \mathcal{N}_\mathrm{U}} \max\left(0, d_k^i - x_k^i\right). \tag{4.11}$$

Minimizing this factor decreases the total bandwidth loss of the UTs. An optimization problem with the max function in Eq. (4.11) can be transformed into a linear optimization as described in Section 3.2.

2. ***Satellite and GW operations factor***: The second factor measures the number of active satellites and GWs to be used. Let the number of active satellites in user links and inter-

satellite links at time k be represented by $\hat{N}_{\mathrm{S},k}$ and the number of active GWs in feeder links at time k be represented by $\hat{N}_{\mathrm{G},k}$. To calculate the number of active satellites considering both user links and inter-satellite links, a matrix

$$\boldsymbol{F}_k := \begin{bmatrix} \boldsymbol{C}_k & \boldsymbol{H}_k \end{bmatrix} \in \{0,1\}^{N_\mathrm{S} \times (N_\mathrm{U}+N_\mathrm{S})} \tag{4.12}$$

is defined and an entry of the matrix $\boldsymbol{F}_k$ is denoted as $f_k^{j,i}$. The row sums of the matrices $\boldsymbol{F}_k$ and $\boldsymbol{E}_k$ represent the number of connected links for satellites in user links and inter-satellite links and GWs in feeder links, respectively. Therefore, the number of active satellites and GWs is represented by

$$\hat{N}_{\mathrm{S},k} = \|\boldsymbol{f}_k\|_0, \tag{4.13a}$$

$$\hat{N}_{\mathrm{G},k} = \|\boldsymbol{e}_k\|_0, \tag{4.13b}$$

respectively, where

$$\boldsymbol{f}_k := \begin{bmatrix} \sum_{j=1}^{N_\mathrm{U}+N_\mathrm{S}} f_k^{1,j} & \cdots & \sum_{j=1}^{N_\mathrm{U}+N_\mathrm{S}} f_k^{N_\mathrm{S},j} \end{bmatrix}^\top \in \mathbb{R}^{N_\mathrm{S}},$$

$$\boldsymbol{e}_k := \begin{bmatrix} \sum_{j=1}^{N_\mathrm{S}} e_k^{1,j} & \cdots & \sum_{j=1}^{N_\mathrm{S}} e_k^{N_\mathrm{S},j} \end{bmatrix}^\top \in \mathbb{R}^{N_\mathrm{G}}.$$

By minimizing the factors in Eq. (4.13), the UTs connect to as few satellites and GWs as possible. Therefore, the SATCOM operator profits because fewer active satellites and GWs reduce the operating costs of the system. On the other hand, by maximizing these factors, network disconnection risks due to satellite and GW failures and changes in the propagation environment can be reduced. This is because the accommodated data per one satellite and one GW is decreased when the number of active satellites and GWs are increased.

Remark 4.2. The number of active satellites and GWs represented by Eqs. (4.13) can be written in another form. For $\hat{N}_{\mathrm{S},k}$ and $\hat{N}_{\mathrm{G},k}$, the following lemma holds.

Lemma 4.1. The numbers of active satellites and GWs are represented by

$$\hat{N}_{\mathrm{S},k} = \mathrm{rank}(\boldsymbol{F}_k),$$

$$\hat{N}_{\mathrm{G},k} = \mathrm{rank}(\boldsymbol{E}_k),$$

respectively.

Proof. As indicated by Eqs. (4.13), the number of active satellites and GWs is represented by the number of nonzero row vectors of the matrices $\boldsymbol{F}_k$ and $\boldsymbol{E}_k$, respectively. The row vectors $\boldsymbol{C}_k$, which are part of $\boldsymbol{F}_k$, are linearly independent because the row vectors are not identical due to the constraint in Eq. (4.5). Similarly, the row vectors of $\boldsymbol{E}_k$ are also linearly independent. Because the number of linearly independent row vectors of a matrix is equivalent to the rank of the matrix, the number of active satellites and GWs is represented by the rank of $\boldsymbol{F}_k$ and $\boldsymbol{E}_k$, respectively. This completes the proof. □

Optimization of the rank functions is shown in Appendix A.

3. *Handover factor*: The third factor measures the number of satellite handovers. In practical SATCOM systems, satellite handovers should be managed efficiently because both the SATCOM operator and the UTs are responsible for the handovers. Let the numbers of satellite handovers in user links, inter-satellite links, and feeder links at each time k be denoted as $n_k^{\mathrm{UL}}, n_k^{\mathrm{ISL}}$ and n_k^{FL}, respectively. These numbers are defined for $k \geq 2$ and represented by

$$n_k^{\mathrm{UL}} = \frac{1}{2} \sum_{i \in \mathcal{N}_{\mathrm{U}}} \sum_{j \in \mathcal{N}_{\mathrm{S}}} |c_k^{j,i} - c_{k-1}^{j,i}|, \tag{4.14a}$$

$$n_k^{\mathrm{ISL}} = \frac{1}{2} \sum_{j_1, j_2 \in \mathcal{N}_{\mathrm{S}}} |h_k^{j_2,j_1} - h_{k-1}^{j_2,j_1}|, \tag{4.14b}$$

$$n_k^{\mathrm{FL}} = \frac{1}{2} \sum_{j \in \mathcal{N}_{\mathrm{S}}} \sum_{\ell \in \mathcal{N}_{\mathrm{G}}} |e_k^{\ell,j} - e_{k-1}^{\ell,j}|, \tag{4.14c}$$

respectively. The factors in Eq. (4.14) directly represent the number of pairs that change connections in the SATCOM network. By minimizing these factors, frequent changes in the network can be reduced, resulting in lower power consumption. By maximizing these factors, the network structure is changed frequently, which makes the network more secure by reducing wiretap risks.

Combining the factors in Eqs. (4.11)–(4.14), the cost function at time k is defined as

$$\begin{aligned} J_k := w_1 \sum_{i \in \mathcal{N}_{\mathrm{U}}} \max\left(0, d_k^i - x_k^i\right) + w_{2\mathrm{s}} \left\|\boldsymbol{f}_k\right\|_0 + w_{2\mathrm{g}} \left\|\boldsymbol{e}_k\right\|_0 + w_{3c} \sum_{i \in \mathcal{N}_{\mathrm{U}}} \sum_{j \in \mathcal{N}_{\mathrm{S}}} |c_k^{j,i} - c_{k-1}^{j,i}| \\ + w_{3h} \sum_{j_1, j_2 \in \mathcal{N}_{\mathrm{S}}} |h_k^{j_2,j_1} - h_{k-1}^{j_2,j_1}| + w_{3e} \sum_{j \in \mathcal{N}_{\mathrm{S}}} \sum_{\ell \in \mathcal{N}_{\mathrm{G}}} |e_k^{\ell,j} - e_{k-1}^{\ell,j}|, \end{aligned} \tag{4.15}$$

where $w_1 \geq 0, w_{2\mathrm{s}} \in \mathbb{R}, w_{2\mathrm{g}} \in \mathbb{R}, w_{3c} \in \mathbb{R}, w_{3h} \in \mathbb{R}$, and $w_{3e} \in \mathbb{R}$ are the weights. The weights $w_{2\mathrm{s}}, w_{2\mathrm{g}}, w_{3c}, w_{3h}$, and w_{3e} can be negative when the corresponding factors are maximized. The SATCOM operator designs these weights according to their individual management strategy and can add other factors to the cost function as desired. For $k = 1$, the last three terms are neglected because these are defined for $k \geq 2$, and the cost is then set as

$$J_1 := w_1 \sum_{i \in \mathcal{N}_{\mathrm{U}}} \max(0, d_k^i - x_k^i) + w_{2\mathrm{s}} \left\| \boldsymbol{f}_k \right\|_0 + w_{2\mathrm{g}} \left\| \boldsymbol{e}_k \right\|_0.$$

Assume that future UT requests and future connectable pair candidates can be predicted[*5]. When an optimization problem is formulated in an MPC manner [78], the cost function at each time is defined by $\sum_{\tau \in \mathcal{T}_p} J_{k+\tau}$, where T_p represents the prediction horizon and $\mathcal{T}_p = \{1, 2, \ldots, T_p\}$ represents a set of the finite-time horizon. Here, the constraints must be satisfied across the entire prediction horizon. Using this approach to formulate the problem, a solution to change the resource allocation and network structure in advance before the UT requests change can be obtained.

4.2.3 Optimization Problem

In the network optimization problem, the following decision variables are determined:

- Network connection matrices of the SATCOM network: $\boldsymbol{C}_k, \boldsymbol{H}_k$, and $\boldsymbol{E}_k$ (binary matrices),
- Control input $\boldsymbol{u}_k$ to determine the bandwidth allocation of UTs (a real vector),
- Resource flow matrices of the SATCOM network: $\boldsymbol{Y}_k, \boldsymbol{V}_k$, and $\boldsymbol{Z}_k$ (real matrices).

Let solutions of the optimization problem be denoted as $\{\hat{\boldsymbol{C}}_{k+\tau}, \hat{\boldsymbol{H}}_{k+\tau}, \hat{\boldsymbol{E}}_{k+\tau}, \hat{\boldsymbol{Y}}_{k+\tau}, \hat{\boldsymbol{V}}_{k+\tau}, \hat{\boldsymbol{Z}}_{k+\tau}\}_{\tau \in \mathcal{T}_p}$, an optical input sequence as $\{\hat{\boldsymbol{u}}_{k+\tau-1}\}_{\tau \in \mathcal{T}_p}$, and the resulting allocated bandwidth as $\{\hat{\boldsymbol{x}}_{k+\tau}\}_{\tau \in \mathcal{T}_p}$.

Combining the cost function and constraints defined previously, the following network optimization problem is formulated.

[*5] To obtain these future requests and connectable candidates, future trends of UT requests and the propagation environment must be predicted by analyzing past data. For example, UT requests can be effectively predicted using actual tracking data of UTs as described in Chapter 3.

Problem 4.1 (Proposed network optimization problem)**.**

For given $N_{\mathrm{U}}, N_{\mathrm{S}}, N_{\mathrm{G}}, B_{\mathrm{S}}^{j}, B_{\mathrm{ISL}}^{j}, B_{\mathrm{G}}^{\ell}$ and future values of $\{\bar{\mathcal{C}}_{k+\tau}, \bar{\mathcal{H}}_{k+\tau}, \bar{\mathcal{E}}_{k+\tau}, d_{k+\tau}^{i}\}_{\tau\in\mathcal{T}_p}$, solve

$$\begin{aligned} &\underset{\substack{\{\boldsymbol{C}_{k+\tau},\boldsymbol{H}_{k+\tau},\boldsymbol{E}_{k+\tau},\boldsymbol{Y}_{k+\tau},\\ \boldsymbol{V}_{k+\tau},\boldsymbol{Z}_{k+\tau},\boldsymbol{u}_{k+\tau-1}\}_{\tau\in\mathcal{T}_p}}}{\text{minimize}} \quad \sum_{\tau\in\mathcal{T}_p} J_{k+\tau} \\ &\text{subject to} \qquad \text{Eqs. (4.2) – (4.10)} \end{aligned}$$

Problem 4.1 is a mixed integer programming problem because $\boldsymbol{u}_k$ is a real vector and $\boldsymbol{Y}_k, \boldsymbol{V}_k$, and $\boldsymbol{Z}_k$ are real matrices, whereas $\boldsymbol{C}_k, \boldsymbol{H}_k$ and $\boldsymbol{E}_k$ are binary matrices (i.e., integer matrices). Although the mixed integer programming problem is NP-hard, this problem can be efficiently solved using the branch-and-cut algorithm, which is one of the heuristic algorithms [118].

The SATCOM operator solves Problem 4.1 at every time k and the first optimal input $\hat{\boldsymbol{u}}_k$ is utilized to determine the bandwidth allocation $\hat{\boldsymbol{x}}_{k+1}$ based on the dynamics in Eq. (4.7). Using this result, the operator manages and controls the satellites and GWs based on the most suitable resource allocation and network structure for a given objective.

Remark 4.3. The constraints in Eq. (4.5) can be reduced to the following equivalent inequality constraints:

$$\sum_{j:(i,j)\in\mathcal{C}_k} c_k^{j,i} \leq 1, \quad \forall i \in \mathcal{N}_{\mathrm{U}}, \tag{4.16a}$$

$$\sum_{\ell:(j,\ell)\in\mathcal{E}_k} e_k^{\ell,j} \leq 1, \quad \forall j \in \mathcal{N}_{\mathrm{S}}. \tag{4.16b}$$

The feasible set of constraints in Eq. (4.16) is equivalent to that of the original constraints in Eq. (4.5) under the binary variables $c_k^{j,i}$ and $e_k^{\ell,j}$. By using these constraints, Problem 4.1 can be solved more efficiently.

4.3 Numerical Experiments

In this section, four types of numerical experiments are conducted to verify the effectiveness of the proposed network design method in the integrated system, which allows UTs to have several candidates of satellites to connect to regardless of whether they are GEO or NGEO

satellites. In Subsection 4.3.1, the basic performance of the proposed method is verified for simple static UT requests. In Subsection 4.3.2, dynamic UT requests are addressed and the performance of the proposed method in the integrated system and in the conventional non-integrated system are compared. In Subsection 4.3.3, the effect of the number of satellites in the system is investigated and the performance of the integrated and non-integrated systems are compared. These three examples do not include inter-satellite links (i.e., $\boldsymbol{H}_k = \boldsymbol{O}_{N_\mathrm{S}}$). The effect of the inter-satellite links in the system is verified in Subsection 4.3.4.

In these simulations, it is assumed that satellites use a different frequency band to prevent interference between communication links. To be consistent with this assumption, the UTs and GWs are assumed to have antennas that cover all the frequency bands utilized by the satellites. Furthermore, the influence of the propagation delay due to the propagation distance is ignored.

To evaluate the performance of the method, the amount of bandwidth loss is defined as

$$L_k = \sum_{i \in \mathcal{N}_\mathrm{U}} L_k^i = \sum_{i \in \mathcal{N}_\mathrm{U}} \max\left(0, d_k^i - x_k^i\right), \tag{4.17}$$

where L_k^i represents the bandwidth loss of the i-th UT at each time k. Here, the bandwidth loss represents the total amount of the additional bandwidth required to meet UT requests.

The optimization problem is described by the MATLAB YALMIP [114] and solved by the Gurobi Optimizer [119]. This solver implements the branch-and-cut algorithm to efficiently solve the mixed integer programming problem. These experiments were conducted using an Intel(R) Xeon(R) E5-2687W CPU at 3.10 GHz with 64 GB RAM.

4.3.1 Example 1: Basic Performance Considering Static Requests

This example focuses on simple static UT requests in one step to confirm the basic performance of the proposed method. Assume that all UT-SAT and SAT-GW pairs can be connected:

$$\bar{\mathcal{C}} = \mathcal{N}_\mathrm{U} \times \mathcal{N}_\mathrm{S}, \tag{4.18a}$$

$$\bar{\mathcal{E}} = \mathcal{N}_\mathrm{S} \times \mathcal{N}_\mathrm{G}. \tag{4.18b}$$

The simulation parameters are shown in Table 4.1. Each UT request d^i is generated from

Table 4.1: Simulation parameters for static UT requests.

N_{U}	N_{S}	N_{G}	B_{S}^{j} [MHz] $(\forall j \in \mathcal{N}_{\mathrm{S}})$	B_{G}^{ℓ} [MHz] $(\forall \ell \in \mathcal{N}_{\mathrm{G}})$
10	5	3	100	100

a uniform distribution on $[30, 70]$ as follows:

$$\boldsymbol{d} = \begin{bmatrix} 58 & 36 & 59 & 50 & 34 & 52 & 40 & 61 & 63 & 38 \end{bmatrix}^{\top}.$$

The cost function is defined as the total amount of bandwidth allocated to the UTs:

$$J := \sum_{i \in \mathcal{N}_{\mathrm{U}}} \max\left(0, d^{i} - x^{i}\right),$$

where $w_1 = 1$ and the other weights are set to 0 in Eq. (4.15).

The solutions obtained by solving Problem 4.1 were as follows:

$$\hat{\boldsymbol{x}} = \begin{bmatrix} 0 & 36 & 59 & 0 & 0 & 41 & 40 & 61 & 60 & 3 \end{bmatrix}^{\top},$$

$$\hat{\boldsymbol{C}} = \begin{bmatrix} 1 & 0 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \end{bmatrix},$$

$$\hat{\boldsymbol{E}} = \begin{bmatrix} 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 \end{bmatrix},$$

$$\hat{\boldsymbol{Y}} = \begin{bmatrix} 0 & 0 & 59 & 0 & 0 & 41 & 0 & 0 & 0 & 0 \\ 0 & 36 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 60 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 61 & 0 & 3 \\ 0 & 0 & 0 & 0 & 0 & 0 & 40 & 0 & 0 & 0 \end{bmatrix},$$

$$\hat{\boldsymbol{Z}} = \begin{bmatrix} 0 & 36 & 0 & 64 & 0 \\ 0 & 0 & 60 & 0 & 40 \\ 100 & 0 & 0 & 0 & 0 \end{bmatrix}.$$

The bandwidth loss L of each UT was

$$L = \begin{bmatrix} 58 & 0 & 0 & 50 & 34 & 11 & 0 & 0 & 3 & 35 \end{bmatrix}^\top.$$

In this example, the simulation conditions were set as inequalities in the total UT requests and the total satellite and GW bandwidths

$$\sum_{i \in \mathcal{N}_\mathrm{U}} d^i \ (= 491) \leq \sum_{j \in \mathcal{N}_\mathrm{S}} B_\mathrm{S}^j \ (= 500),$$
$$\sum_{i \in \mathcal{N}_\mathrm{U}} d^i \ (= 491) \geq \sum_{\ell \in \mathcal{N}_\mathrm{G}} B_\mathrm{G}^\ell \ (= 300),$$

were held fixed. Thus, not all UT requests were accommodated and the UTs only obtained a total bandwidth of

$$\sum_{i \in \mathcal{N}_\mathrm{U}} \hat{x}^i = 300.$$

Therefore, UT-1, UT-4, and UT-5 experienced bandwidth loss, despite their connections to SAT-1. By weighting each term x^i in the cost function, this imbalance can be avoided in situations with limited bandwidth.

4.3.2 Example 2: Comparison of Integrated and Non-Integrated Systems

In this example, dynamic UT requests that show temporal variability are simulated. The simulation parameters are shown in Tables 4.2 and 4.3, and the simulation is conducted in $T = 150$ steps with $T_p = 1$.

There are a total of 65 satellites in the large-scale SATCOM system, which is composed of five GEO satellites and 60 NGEO satellites divided into four groups. Figure 4.5 shows the satellite visibility schedule. Over a continuous period of visibility, the corresponding satellites can connect to all UTs and GWs. The GEO satellites are visible throughout the entire simulation time interval, whereas the visibility periods of the NGEO satellites are limited to finite intervals and repeated periodically. In this example, the author assumes that the period of each NGEO satellite overlaps itself by two steps to enable efficient handovers. The connectable pair candidate sets $\bar{\mathcal{C}}_k$ and $\bar{\mathcal{E}}_k$ are obtained throughout the simulation time

Table 4.2: Simulation parameters of satellites.

Type of satellites	N_{S}	B_{S}^{j} [MHz] (for each satellite)	Visible period [steps]
GEO	5	500	-
NGEO group 1	24	250	5
NGEO group 2	18	250	6
NGEO group 3	12	250	10
NGEO group 4	6	250	20
	65 (in total)		

Table 4.3: Simulation parameters of UT and GW.

N_{U}	N_{G}	B_{G}^{ℓ} [MHz] $(\forall \ell \in \mathcal{N}_{\mathrm{G}})$
100	10	500

interval based on this schedule. Furthermore, the time-varying total satellite bandwidth is calculated from the bandwidth B_{S}^{j} of each satellite and this visibility schedule.

Assume that the satellites have beams large enough to connect to all UTs and GWs when these satellites are visible. The satellites have steerable antennas to connect to multiple UTs as well as fixed antennas to connect to the GWs. The UTs have steerable antennas to track the satellites and the GWs have antennas with sufficiently wide beams to track all connected satellites. Furthermore, assume that the adjacent satellites in the same NGEO group connect to the same GWs in the overlapped period .

The dynamics of the UT requests are described as

$$d_{k+1}^{i} = \gamma_{k}^{i} d_{k}^{i}, \quad \forall i \in \mathcal{N}_{\mathrm{U}}, \tag{4.19}$$

where the initial values of the UT request d_{1}^{i} are generated from a uniform distribution on $[10, 30]$, and coefficients γ_{k}^{i} are generated from a Gaussian distribution with $\mathcal{N}(1.0075, 0.05)$ in each step.

Figure 4.6 shows the total time-varying UT requests $\left(\sum_{i \in \mathcal{N}_{\mathrm{U}}} d_{k}^{i}\right)$, the total satellite bandwidth $\left(\sum_{j:(i,j) \in \bar{\mathcal{C}}_{k} \cap (j,\ell) \in \bar{\mathcal{E}}_{k}} B_{\mathrm{S}}^{j}\right)$, and the total GW bandwidth $\left(\sum_{\ell:(j,\ell) \in \bar{\mathcal{E}}_{k}} B_{\mathrm{G}}^{\ell}\right)$. According to this figure, the total UT requests exceed the total satellite and GW bandwidth at approximately $k = 85$ and $k = 120$, respectively. The total GW bandwidth is larger than the total satellite bandwidth throughout the simulation time interval.

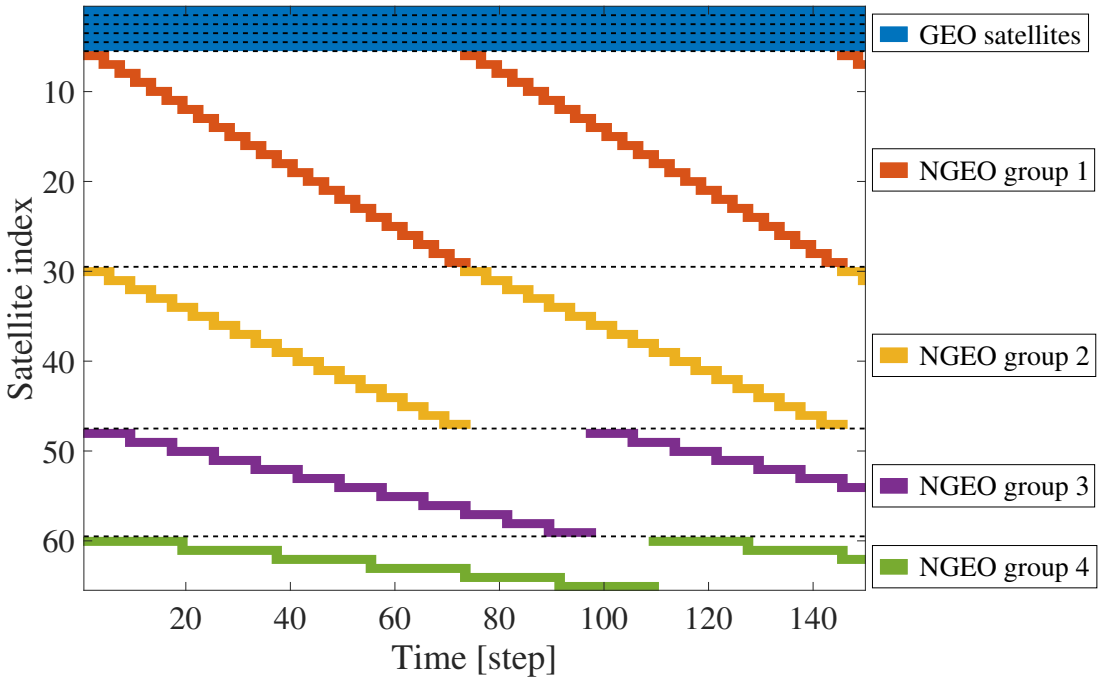


Fig. 4.5: Satellite visibility schedule. All the UTs and GWs in the system can connect to satellites that are in the period of visibility. (Copyright © 2020 IEICE, [77] Fig. 4)

Define the cost function as

$$J_k := \sum_{i \in \mathcal{N}_{\mathrm{U}}} \max\left(0, d_k^i - x_k^i\right), \tag{4.20}$$

where $w_1 = 1$ and the other weights are set to 0 in Eq. (4.15). Using this cost, the following three methods are compared: the proposed method in the integrated system that solves Problem 4.1 (Method A), a method to design a random network in the integrated system (Method B), and a method to design a random network in the conventional non-integrated system (Method C).

In Methods B and C, the connection pairs (UT and GW to GEO or NGEO satellite groups) are determined randomly. In addition, in Method B the pairs are determined at each time step and can be changed in the next step, whereas in Method C the pairs cannot be changed throughout the simulation. Due to the fixed connection pairs, the network structure $\boldsymbol{C}_k$ and $\boldsymbol{E}_k$ are determined in advance. Thus, in Methods B and C, the bandwidth allocation $\hat{\boldsymbol{x}}_k$ is obtained by solving the following optimization problem:

$$\begin{aligned} &\text{minimize} && J_k \\ &\text{subject to} && \text{Eqs. (4.6), (4.7), (4.8), (4.10).} \end{aligned}$$

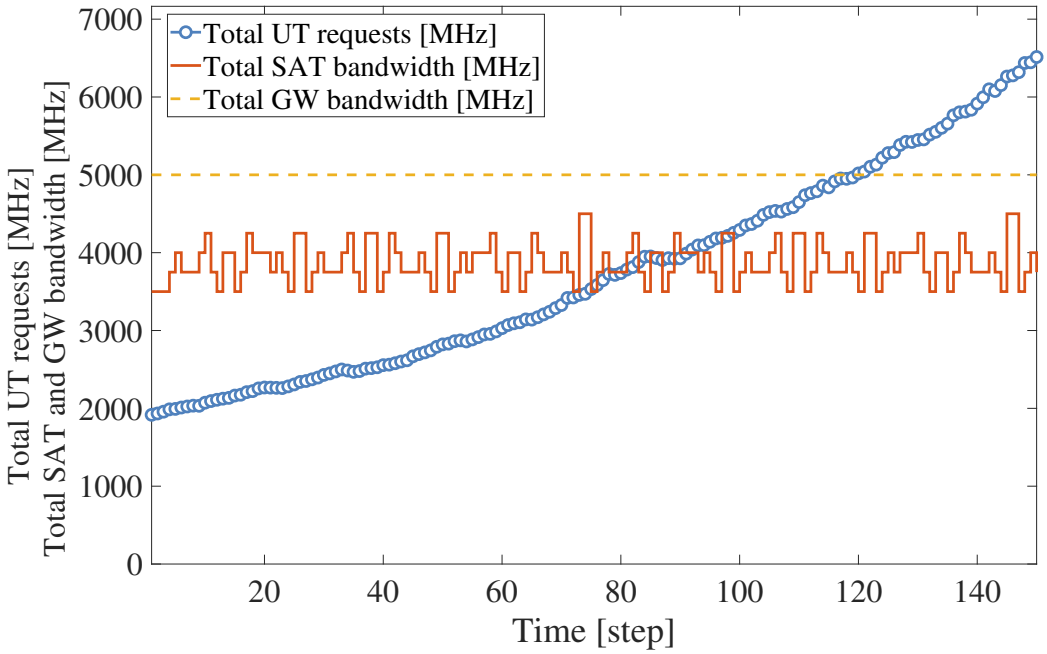


Fig. 4.6: Time-varying total UT requests (blue line), total satellite bandwidth (red line), and total GW bandwidth (dashed yellow line). (Copyright © 2020 IEICE, [77] Fig. 5)

This is a linear programming problem because $\boldsymbol{C}_k$ and $\boldsymbol{E}_k$ are given and the decision variables are $\boldsymbol{u}_k$, $\boldsymbol{Y}_k$, and $\boldsymbol{Z}_k$.

Figure 4.7 shows the results of the bandwidth loss L_k among the three methods. The simulation was conducted in 30 trials for Methods B and C and the average bandwidth loss was calculated. Method A exhibited better performance than the other methods throughout the simulation time interval. Specifically, the bandwidth loss was almost zero during the interval from $k = 1$ to approximately $k = 85$ because the total satellite bandwidth exceeded the total UT requests. Furthermore, the resulting bandwidth loss of the proposed method achieved a lower bound, which is determined by the total satellite and GW bandwidth and is represented by

$$\underline{L_k} = \max\left(0, \Delta_k^L\right), \tag{4.21}$$

where

$$\Delta_k^L = \max\left(\sum_{i\in\mathcal{N}_\mathrm{U}} d_k^i - \sum_{j:(i,j)\in\bar{\mathcal{C}}_k\cap(j,\ell)\in\bar{\mathcal{E}}_k} B_\mathrm{S}^j, \sum_{i\in\mathcal{N}_\mathrm{U}} d_k^i - \sum_{\ell:(j,\ell)\in\bar{\mathcal{E}}_k} B_\mathrm{G}^\ell\right).$$

This means the proposed method maximized the use of the available resources.

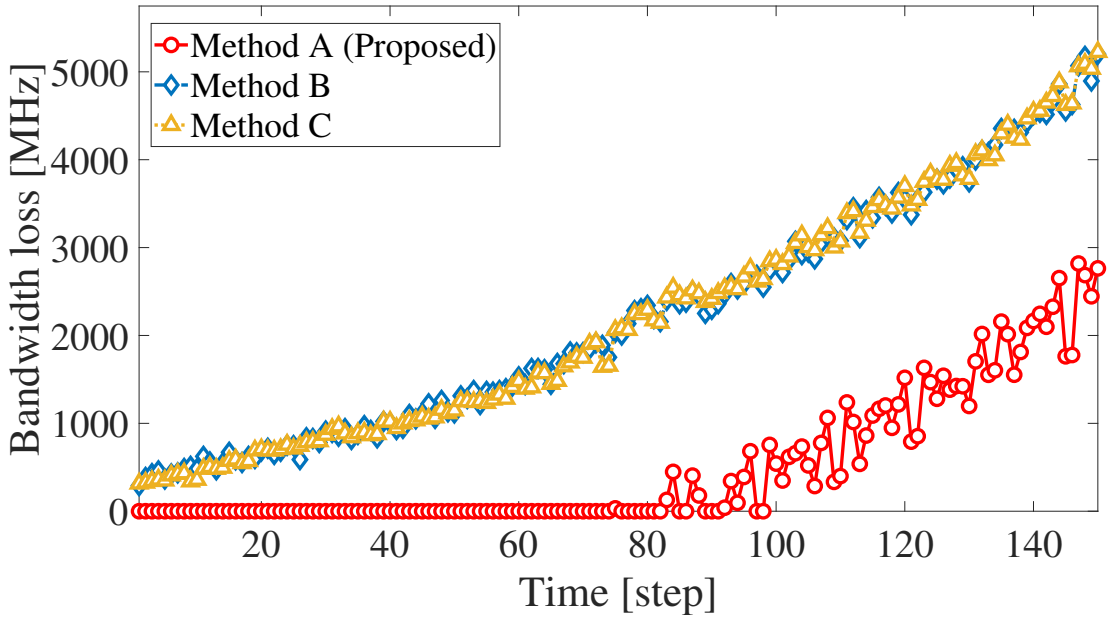


Fig. 4.7: Comparison of the bandwidth loss of the proposed method to that in integrated and conventional non-integrated systems. The red, yellow, and blue lines represent the results of Methods A, B, and C, respectively.

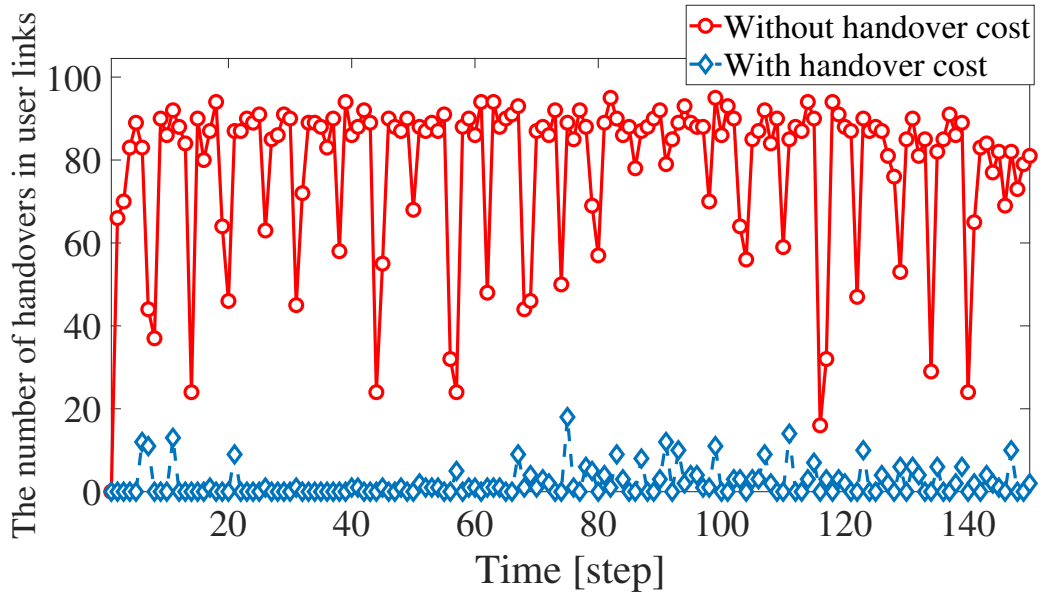
In the other two methods, the bandwidth loss was not zero at any time and the loss increased as the total UT requests increased. This was because the GEO or NGEO satellites to which the UTs can be connected were limited in number. After the time instant $k = 85$, the bandwidth loss of Method A also increased because the total UT requests exceeded the total satellite bandwidth. However, the average bandwidth loss of Method A was reduced on average by 71.2% from $k = 85$ to 150 compared to Methods B and C. Furthermore, Method B exhibited worse performance than Method A, despite both methods involving the integrated system. This indicates that the performance is reduced by determining the network randomly, even when the integrated system is considered.

Next, the handover factor in the cost function is activated, which means the cost is defined as

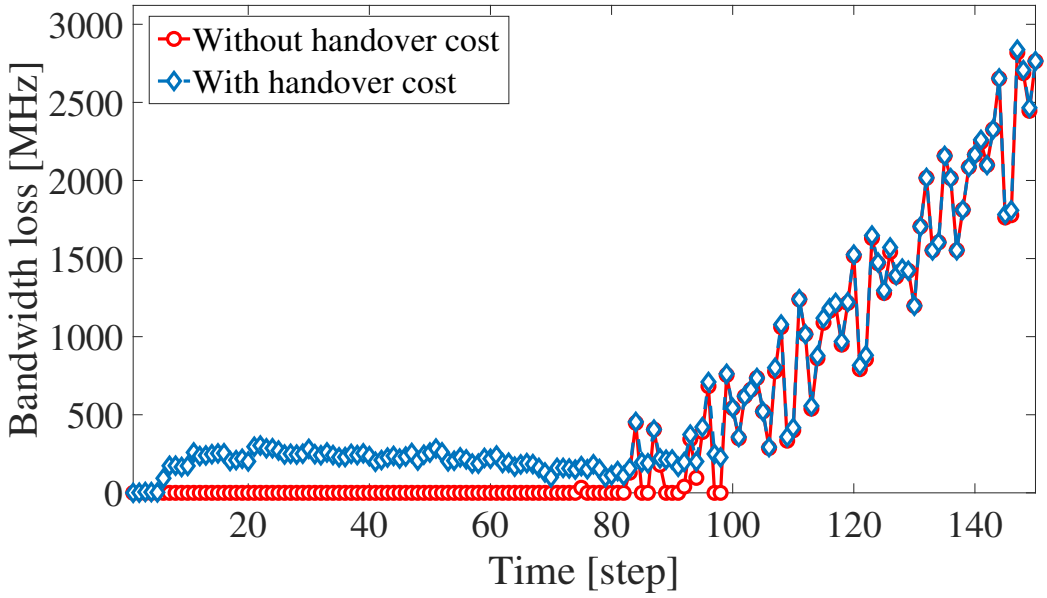
$$J_k := \sum_{i \in \mathcal{N}_\mathrm{U}} \max\left(0, d_k^i - x_k^i\right) + 20 \sum_{i \in \mathcal{N}_\mathrm{U}} \sum_{j \in \mathcal{N}_\mathrm{S}} |c_k^{j,i} - c_{k-1}^{j,i}|,$$

where $w_1 = 1, w_{3c} = 20$, and the other weights are set to 0 in Eq. (4.15). The result using this cost function is compared with the result by setting $w_{3c} = 0$, which is calculated in the aforementioned simulation.

Figure 4.8 shows the bandwidth loss and the number of handovers in user links, which



(a)



(b)

Fig. 4.8: Comparison of the performance of the proposed method with the handover factor in the cost function being deactivated ($w_{3c} = 0$, red line) and activated ($w_{3c} = 20$, blue line). (a) The number of handovers in user links and (b) bandwidth loss.

is the number of UT-SAT pairs changing the connection. In the figure, red and blue lines represent the result of the proposed method without and with the handover factor, respectively. The red line in Fig. 4.8 is the same as that in Fig. 4.7. Fig. 4.8(a) demonstrates that the number of handovers were decreased by making the handover factor active in the cost function. On the other hand, in Fig. 4.8(b), the bandwidth loss was increased compared to the result without the handover term because decreasing the number of handovers can increase the bandwidth loss by not changing the satellite connection. After $k = 99$, the results of these conditions were identical, which means both results reached the lower bound of the physical constraint of the available bandwidth represented by Eq. (4.21).

These results demonstrate the effectiveness of the proposed method in the integrated system compared to the conventional non-integrated system.

4.3.3 Example 3: Effect of the Number of Satellites

In this example, the effect of the number of satellites is investigated by a comparison of the performance of the integrated and non-integrated systems.

Assume that only GEO satellites with $B_{\mathrm{S}} = 500$ MHz exist in the system and that all pairs can be connected as defined in Eq. (4.18). In this simulation, the number of GEO satellites is increased from $N_{\mathrm{S}} = 1$ to 30. The cost function and the UT requests are set to the same as those for Example 2 (see Eq. (4.20)). The number of GWs and the GW bandwidth are set to $N_{\mathrm{G}} = 30$ and $B_{\mathrm{G}} = 500$ MHz, respectively, so that an inequality $\sum_{j\in\mathcal{N}_{\mathrm{S}}} B_{\mathrm{S}}^{j} \leq \sum_{\ell\in\mathcal{N}_{\mathrm{G}}} B_{\mathrm{G}}^{\ell}$ holds for any N_{S}. Note that the total satellite bandwidth corresponding to $N_{\mathrm{S}} = 14$ to 30 exceeds the total UT requests throughout the simulation.

Figure 4.9 shows the bandwidth loss as a function of the number of satellites for the integrated (Method A) and non-integrated (Method C) systems. For the Method C simulation, the average bandwidth loss of 50 trials was utilized for a given number of satellites. In Fig. 4.9, the vertical axis represents the time average of the bandwidth loss throughout the simulation: $(1/T)\sum_{k=1}^{T} L_k$.

This result indicates that the difference in performance depends on the number of satellites that exist in the system. For all N_{S}, the bandwidth loss of Method A was less than that of Method C and achieved zero when $N_{\mathrm{S}} \geq 14$. For a small number of satellites, the difference was small because the total satellite bandwidth was less than the bandwidth of the UT

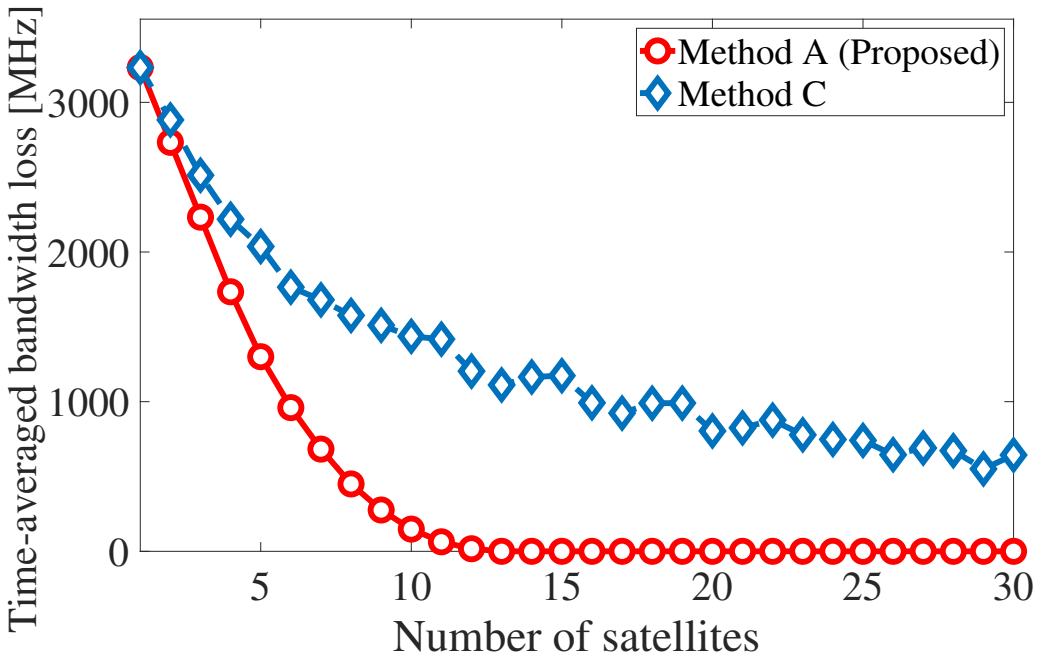


Fig. 4.9: Relationship between number of satellites and the time-averaged bandwidth loss. The red and blue lines represent the results of Methods A and C, respectively.

requests in both systems. However, the difference in bandwidth loss increased as the number of satellites increased, with an average difference from $N_{\mathrm{S}} = 14$ to 30 of approximately 835 MHz. Therefore, for a large number of satellites ($10 \geq N_{\mathrm{S}} \geq 30$), the proposed method in the integrated system exhibited better performance than the conventional non-integrated system. When there is a sufficiently large number of satellites, the bandwidth loss can be zero for Method C also, because the satellites can provide enough resources to satisfy all the UT requests with any network structure.

Remark 4.4. In the integrated system, if all UTs can obtain a bandwidth equal to the UT request at time k, the following inequalities hold:

$$\sum_{i \in \mathcal{N}_{\mathrm{U}}} d_k^i \leq \sum_{j \in \mathcal{N}_{\mathrm{S}}} B_{\mathrm{S}}^j \leq \sum_{\ell \in \mathcal{N}_{\mathrm{G}}} B_{\mathrm{G}}^\ell. \tag{4.22}$$

However, in the non-integrated system, if all UTs can obtain a bandwidth equal to the UT request, the following inequalities hold:

$$\sum_{i \in \mathcal{N}_{\mathrm{U}}^j} d_k^i \leq B_{\mathrm{S}}^j, \quad \forall j \in \mathcal{N}_{\mathrm{S}} \quad \text{and} \quad \sum_{j \in \mathcal{N}_{\mathrm{S}}} B_{\mathrm{S}}^j \leq \sum_{\ell \in \mathcal{N}_{\mathrm{G}}} B_{\mathrm{G}}^\ell. \tag{4.23}$$

Here, $\mathcal{N}_{\mathrm{U}}^j \subseteq \mathcal{N}_{\mathrm{U}}$ is defined as a set of UTs connected to the j-th satellite.

According to these relationships, the inequalities in Eq. (4.22) hold if those in Eq. (4.23) hold, which means that all UTs can always obtain a bandwidth equal to the UT request in the integrated system if this situation is achieved in the non-integrated system. This fact indicates that the performance of the integrated system is always the same or superior to that of the non-integrated system. This relationship can be proven by taking the sum of the former inequality in Eq. (4.23) using $j \in \mathcal{N}_{\mathrm{S}}$.

4.3.4 Example 4: Effect of Inter-Satellite Links

Finally, the effect of the proposed method for inter-satellite links is verified.

Assume there exist ten UTs ($N_{\mathrm{U}} = 10$), two GWs ($N_{\mathrm{G}} = 2$), and one GEO satellite in the system. Furthermore, assume there are also a sufficient number of NGEO satellites so that three NGEO satellites are always visible from the UTs throughout the simulation time interval. The maximum bandwidths of satellites for user links and inter-satellite links are shown in Table 4.4. The maximum bandwidth for the inter-satellite links is set much larger than that for the user links because the inter-satellite links will be typically conducted by optical links.

In this example, a simple data relay system, as shown in Fig. 4.10, is simulated. In this setup, a GEO satellite can aggregate signals from three NGEO satellites by using inter-satellite links. To simulate this system, sets of connectable pair candidates $\bar{\mathcal{C}}_k, \bar{\mathcal{H}}_k$, and $\bar{\mathcal{E}}_k$ are set as

$$
\begin{aligned}
\bar{\mathcal{C}} &= \{(i,2),(i,3),(i,4) \mid \forall i \in \mathcal{N}_{\mathrm{U}}\},\\
\bar{\mathcal{H}} &= \{(1,2),(1,3),(1,4)\},\\
\bar{\mathcal{E}} &= \{(1,1),(2,2),(3,2),(4,2)\}.
\end{aligned}
$$

These settings mean all UTs can connect only to the NGEO satellites and these satellites can connect to the GEO satellite via inter-satellite links. Half of the GWs have bandwidths of $B_{\mathrm{G}} = 500$ MHz for the GEO satellite and the other half have bandwidths of $B_{\mathrm{G}} = 250$ MHz for the NGEO satellites. The cost function is set to the same as that for Example 2 (see Eq. (4.20)). UT requests are generated from the same Gaussian distribution as that from Example 2 and the simulation is conducted in $T = 150$ steps.

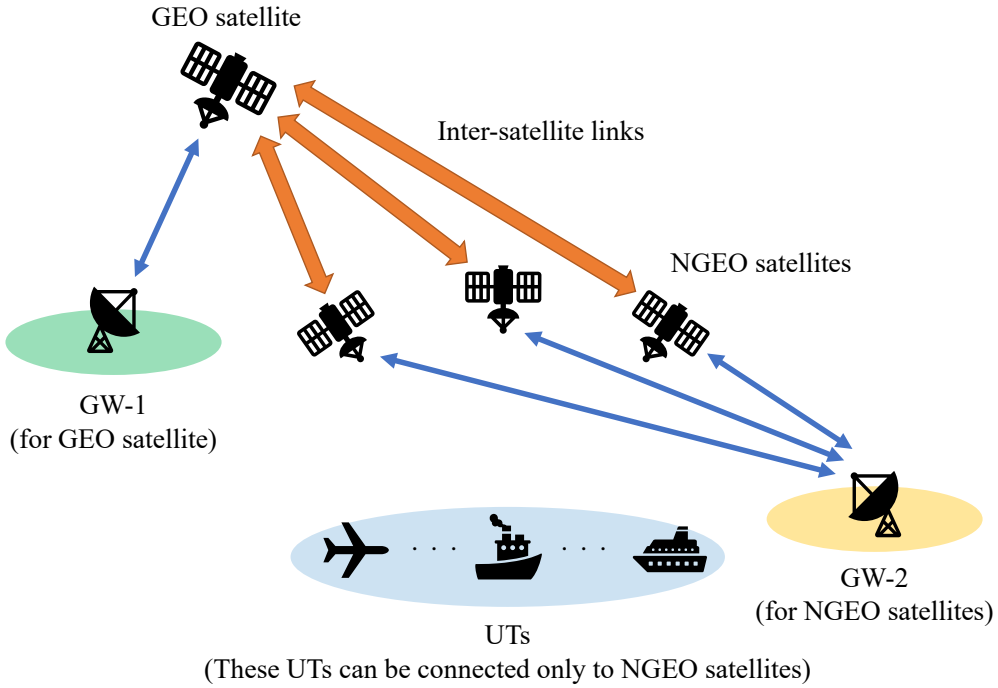


Fig. 4.10: Data relay system (including inter-satellite links) simulated in Example 4.

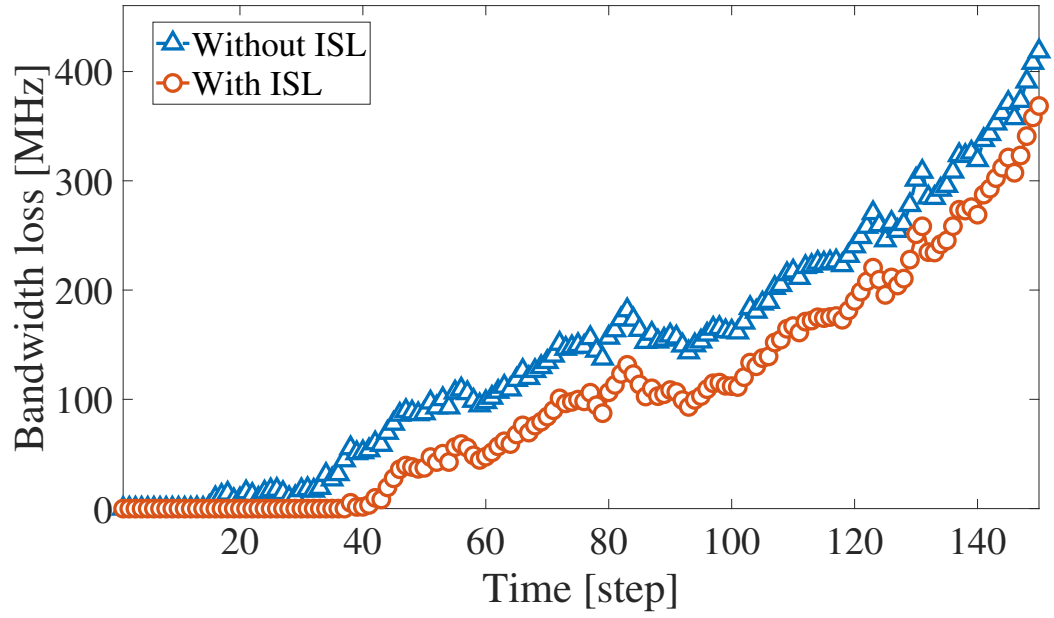
Table 4.4: Simulation parameters of satellites for inter-satellite links.

	B_{S}^{j} [MHz]	B_{ISL}^{j} [MHz]
GEO satellite	500	1000
NGEO satellite	100	500

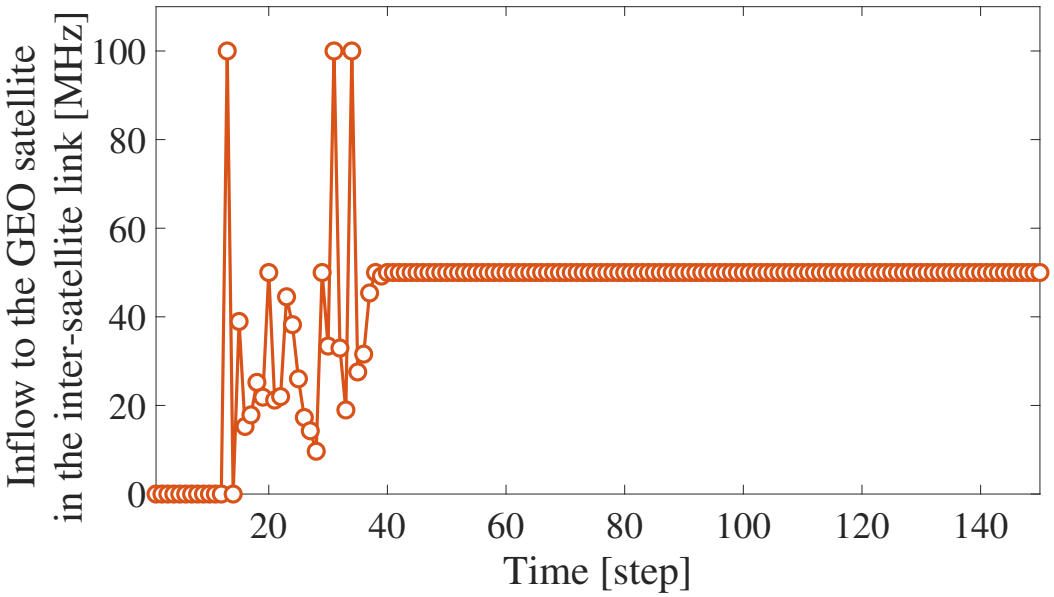
Figure 4.11 shows the effect of the inter-satellite links on the proposed strategy. In Fig. 4.11(a), the bandwidth loss without and with the inter-satellite links are presented. By introducing the inter-satellite links, the NGEO satellites were connected to the GEO satellite, which routed the data to the GW as a substitute for the NGEO satellites. As a result, the bandwidth loss was decreased because the communication data were routed to the GEO satellite as shown in Fig. 4.11(b). The difference between these results after $k = 40$ was 50 MHz, which arose from the following difference:

$$\left| B_{\mathrm{G}}^{2} - \sum_{j \in \{2,3,4\}} B_{\mathrm{S}}^{j} \right| = |250 - 300| = 50.$$

The result in this example demonstrated the effectiveness of the proposed strategy for inter-satellite links.



(a)



(b)

Fig. 4.11: (a) Comparison of the bandwidth loss results without and with inter-satellite links. (b) Inflow to the GEO satellite in the inter-satellite link.

4.4 Summary

In this chapter, the author proposed a framework for the efficient resource and network management of large-scale SATCOM systems integrating multiple satellites, UTs, and GWs. The network optimization problem was formulated for the SATCOM operator to determine efficient resource allocation and design reliable SATCOM network structure that can accommodate multiple UT requests. In this framework, the SATCOM operator can design the network according to their individual management strategy by adjusting the weights of the cost function. The numerical experiments verified the effectiveness of the proposed method in the integrated system and demonstrated that the proposed management strategy performed better than the conventional strategy. Finally, increasing the number of satellites and inter-satellite links was discovered to further improve the performance of the proposed strategy.

Chapter 5

A Chance-Constrained Approach for Managing Stochastic Communication Requests

This chapter is organized as follows. In Section 5.1, the stochastic and time-varying communication requests are modeled using an auto-regressive integrated moving-average (ARIMA) model. In Section 5.2, a chance constraint on the bandwidth loss rate is described. A joint chance constraint is utilized to ensure that the user-specified performance is achieved with high probability through the prediction horizon. In Section 5.3, the joint chance constraint is relaxed to the deterministic linear constraint, and then the optimization problem is reformulated. In Section 5.4, the effectiveness of the proposed chance-constrained management strategy is verified using numerical experiments. Finally, this chapter is summarized in Section 5.5.

The chance constraint can be applicable to the management framework for both the frequency flexibility described in Chapter 3 and the massive connected networks described in Chapter 4. In this chapter, the chance-constrained methodology is incorporated into the management framework for the massive connected networks. The methodology proposed in this chapter is also applicable to the frequency flexibility in the same manner.

5.1 Stochastic and Time-Varying Communication Requests

For each time $k \geq 0$, each UT can send a communication request to satellites. The amount of the communication request from the i-th UT is denoted by $d_k^i \geq 0$. As described in Section 1.4, the stochastic and time-varying requests of the SATCOM system behave like non-stationary stochastic processes. Thus, the dynamics of the i-th UT request is modeled using an auto-regressive integrated moving-average (ARIMA) model, which enables the representation of non-stationary stochastic processes [95]. The ARIMA model of order (n_i, r_i, m_i) is represented by

$$\phi^i(B) d_{k+1}^i = \theta^i(B) w_k^i + d_{\text{const.}}^i, \tag{5.1}$$

where $w_k^i \in \mathbb{R}$ is an i.i.d. (independent and identically distributed) Gaussian random variable with distribution $\mathcal{N}(\mu_i, \sigma_i^2)$, $d_{\text{const.}}^i \in \mathbb{R}$ is a constant value, B is the backshift operator that works as $B^n d_k = d_{k-n}$, and

$$\phi^i(B) := \left(1 - \sum_{j=1}^{n_i} \phi_j B^j\right)(1-B)^{r_i},$$
$$\theta^i(B) := 1 - \sum_{j=1}^{m_i} \theta_j B^j.$$

Here, the model in Eq. (5.1) is rewritten as

$$d_{k+1}^i = \tilde{\phi}^i(B) d_k^i + \theta^i(B) w_k^i + d_{\text{const.}}^i, \tag{5.2}$$

where $\tilde{\phi}^i(B)$ represents a coefficient function only for the past value of d_k^i.

The future value of the requests can be predicted by utilizing the dynamical model in Eq. (5.2) and calculating a linear sum of the time series of the past requests.

5.2 Network Optimization with Chance Constraint of Bandwidth Loss Rate

In this section, a chance constraint on the bandwidth loss is described and a network optimization problem with the chance constraint is formulated.

5.2.1 Chance Constraint of Bandwidth Loss Rate

In this subsection, a chance constraint on bandwidth loss rate is described. The bandwidth loss rate of the i-th UT at time k is defined as

$$b_k^i := \max\left(0, \frac{d_k^i - x_k^i}{d_k^i}\right), \tag{5.3}$$

which represents the ratio of the bandwidth loss to the bandwidth of the requests. This measure is adopted instead of the absolute bandwidth loss in order to evaluate the performance of each UT uniformly. The event that represents the bandwidth loss rate remaining below a threshold level $\beta_k^i \in [0, 1]$ is defined as

$$\mathfrak{B}_k^i := \{\omega \in \Omega \mid b_k^i \leq \beta_k^i\}, \tag{5.4}$$

where ω denotes a sample from a sample space Ω.

What is crucial for the reliable operation of the SATCOM network experiencing the stochastic and time-varying requests is to ensure that the event in Eq. (5.4) is achieved for all the UTs at all times with a certain reliability level. This required constraint is defined by

$$\Pr\left[\bigcap_{\tau \in \mathcal{T}_p} \mathfrak{B}_{i,k+\tau}\right] \geq 1 - \Delta^i, \quad \forall i \in \mathcal{N}_{\mathrm{U}}, \tag{5.5}$$

where T_p is the prediction horizon, $\mathcal{T}_p = \{1, 2, \ldots, T_p\}$ is a set of the finite-time horizon, and $\Delta^i \in (0, 1]$ is the reliability level of the i-th UT. This constraint is called a *joint chance constraint*, which is defined as the conjunction of state constraints [120]. In this constraint, setting a small Δ^i represents high reliability (e.g., setting $\Delta^i = 0.1$ requires 90% reliability of the joint probability for the i-th UT).

5.2.2 Optimization Problem with Joint Chance Constraint

Based on the formulated optimization problem in Chapter 4, a SATCOM network optimization problem based on the chance-constrained MPC approach is formulated. The cost

function in Eq. (4.15) is adjusted for this chapter as follows:

$$J_k := \begin{cases} \sum_{i\in\mathcal{N}_\mathrm{U}} \max\left(0, d_k^i - x_k^i\right), & (k=1) \\ w_c \sum_{i\in\mathcal{N}_\mathrm{U}} \sum_{j\in\mathcal{N}_\mathrm{S}} |c_k^{j,i} - c_{k-1}^{j,i}| + w_h \sum_{j_1,j_2\in\mathcal{N}_\mathrm{S}} |h_k^{j_2,j_1} - h_{k-1}^{j_2,j_1}| & \\ \qquad + w_e \sum_{j\in\mathcal{N}_\mathrm{S}} \sum_{\ell\in\mathcal{N}_\mathrm{G}} |e_k^{\ell,j} - e_{k-1}^{\ell,j}|, & (k\geq 2) \end{cases} \tag{5.6}$$

where $w_c \in \mathbb{R}, w_h \in \mathbb{R}$, and $w_e \in \mathbb{R}$ are the weights. Adopting the MPC approach, the SATCOM network can be optimized while estimating future requests of the UTs by using the ARIMA model in Eq. (5.2). To calculate the future requests of the UTs, the time series of the past requests $\{d_{k-\tau}^i\}_{0\leq\tau\leq n_i r_i-1}$, where n_i and r_i are the orders of the ARIMA model, is required.

Problem 5.1 (Proposed chance-constrained network optimization problem)**.**
For given $N_\mathrm{U}, N_\mathrm{S}, N_\mathrm{G}, B_\mathrm{S}^j, B_\mathrm{ISL}^j, B_\mathrm{G}^\ell$, the time series of past requests $\{d_{k-\tau}^i\}_{0\leq\tau\leq n_i r_i-1}$, the future values of $\{\bar{\mathcal{C}}_{k+\tau}, \bar{\mathcal{H}}_{k+\tau}, \bar{\mathcal{E}}_{k+\tau}, d_{k+\tau}^i\}_{\tau\in\mathcal{T}_p}$, and the ARIMA model in Eq. (5.2), solve

$$\begin{aligned} \underset{\substack{\{\boldsymbol{C}_{k+\tau},\boldsymbol{H}_{k+\tau},\boldsymbol{E}_{k+\tau},\boldsymbol{Y}_{k+\tau},\\ \boldsymbol{V}_{k+\tau},\boldsymbol{Z}_{k+\tau},\boldsymbol{u}_{k+\tau-1}\}_{\tau\in\mathcal{T}_p}}}{\text{minimize}} \quad & \sum_{\tau\in\mathcal{T}_p} J_{k+\tau} \\ \text{subject to} \quad & \text{Eqs. (4.2) – (4.10), (5.5)}, \ \forall\tau\in\mathcal{T}_p, \ i\in\mathcal{N}_\mathrm{U} \end{aligned}$$

Problem 5.1 is a mixed integer and nonlinear programming problem because $\boldsymbol{C}_k, \boldsymbol{H}_k$, and $\boldsymbol{E}_k$ are binary matrices and the joint chance constraint in Eq. (5.5) is a nonlinear constraint in the decision variables. As in Chapter 4, the SATCOM operator solves Problem 5.1 at every time k, the first optimal input $\hat{\boldsymbol{u}}_k$ is utilized to determine the resource allocation $\hat{\boldsymbol{x}}_{k+1}$ based on the dynamics in Eq. (4.7), and the SATCOM network is changed according to the obtained $\hat{\boldsymbol{C}}_{k+1}, \hat{\boldsymbol{H}}_{k+1}$, and $\hat{\boldsymbol{E}}_{k+1}$.

5.3 Reduction to Deterministic Optimization Problem

To solve Problem 5.1 efficiently, the joint chance constraint in Eq. (5.5) is reduced to the deterministic one through the following two stages:

- Stage 1: Relaxing the joint chance constraint to the individual one.

- Stage 2: Reducing the individual chance constraint to the deterministic one.

5.3.1 Stage 1: Relaxing Joint Chance Constraint

The following lemma holds for the joint chance constraint.

Lemma 5.1. The joint chance constraint in Eq. (5.5) holds if there exists $\delta^i_{k+\tau}$ $(\tau \in \mathcal{T}_p)$ such that

$$\Pr\left[\mathfrak{B}^i_{k+\tau}\right] \geq 1 - \delta^i_{k+\tau}, \tag{5.7}$$

$$\sum_{\tau \in \mathcal{T}_p} \delta^i_{k+\tau} \leq \Delta^i,\ \delta^i_{k+\tau} \geq 0, \tag{5.8}$$

hold for all $\tau \in \mathcal{T}_p$ and $i \in \mathcal{N}_{\mathrm{U}}$.

Proof. For an event $\mathfrak{E}_i$ and its negation $\bar{\mathfrak{E}}_i$, Boole's inequality holds [94]:

$$\Pr\left[\bigcap_i \mathfrak{E}_i\right] \geq 1 - \sum_i \Pr\left[\bar{\mathfrak{E}}_i\right].$$

Applying this inequality to the left side of the joint chance constraint in Eq. (5.5) results in

$$\Pr\left[\bigcap_{\tau \in \mathcal{T}_p} \mathfrak{B}^i_{k+\tau}\right] \geq 1 - \sum_{\tau \in \mathcal{T}_p} \Pr\left[\bar{\mathfrak{B}}^i_{k+\tau}\right],\ \forall i \in \mathcal{N}_{\mathrm{U}}. \tag{5.9}$$

Here, setting the right side of the inequality in Eq. (5.9) to $1 - \Delta^i$ or more implies the inequality in Eq. (5.5) holds. That is, the condition that

$$\sum_{\tau \in \mathcal{T}_p} \Pr\left[\mathfrak{B}^i_{k+\tau}\right] \geq \sum_{\tau \in \mathcal{T}_p} 1 - \Delta^i,\ \forall i \in \mathcal{N}_{\mathrm{U}} \tag{5.10}$$

holds. Furthermore, the inequality in Eq. (5.10) holds if the inequalities in Eqs. (5.7) and (5.8) hold. This completes the proof of Lemma 5.1. □

Lemma 5.1 means that the joint chance constraint is divided into T_p individual chance constraints. Thus, the equations in Eq. (5.8) can be regarded as the reliability level Δ^i allocated to each prediction horizon as $\sum_{\tau \in \mathcal{T}_p} \delta^i_{k+\tau}$.

5.3.2 Stage 2: Reducing Chance Constraint

In the next stage, the individual chance constraint in Eq. (5.7) is reduced to the deterministic one. It is known that the individual chance constraint is transformed to an equivalent deterministic constraint when a cumulative distribution function (CDF) of the random variable has an analytical form [120].

For simplicity of notation, the time horizon index τ is neglected in the following description. By applying the model in Eq. (5.2) and $\beta_k^i \geq 0$, the inequality $b_k^i \leq \beta_k^i$ is transformed as

$$\begin{aligned}
\max\left(0, \frac{d_k^i - x_k^i}{d^k}\right) &\leq \beta_k^i \\
\frac{d_k^i - x_k^i}{d_k^i} &\leq \beta_k^i \ \ (\because \beta_k^i \geq 0) \\
(1-\beta_k^i) d_k^i &\leq x_k^i \\
\tilde{\phi}^i(B) d_{k-1}^i + \theta^i(B) w_{k-1}^i + d_{\text{const.}}^i &\leq \frac{x_k^i}{(1-\beta_k^i)} \ \ (\because \text{Eq. (5.1)}) \\
\tilde{\phi}^i(B) d_{k-1}^i + \bar{w}_{k-1}^i + d_{\text{const.}}^i &\leq \frac{x_k^i}{(1-\beta_k^i)},
\end{aligned}$$

where

$$\bar{w}_k^i := \theta^i(B) w_k^i = \left(1 - \sum_{j=1}^{m_i} \theta_j B^j\right) w_k^i. \tag{5.11}$$

Thus, the event in Eq. (5.4) is equivalent to

$$\mathfrak{B}_k^i = \left\{ \omega \in \Omega \;\middle|\; \bar{w}_{k-1}^i \leq \frac{x_k^i}{(1-\beta_k^i)} - \tilde{\phi}^i(B) d_{k-1}^i - d_{\text{const.}}^i \right\}.$$

The inequality $\Pr\left[\mathfrak{B}_{k+\tau}^i\right] \geq 1 - \delta_{k+\tau}^i$ in Eq. (5.7) is then transformed as

$$F_{\bar{w}^i}\left(\frac{x_k^i}{(1-\beta_k^i)} - \tilde{\phi}^i(B) d_{k-1}^i - d_{\text{const.}}^i\right) \geq 1 - \delta_k^i,$$

where $F_{\bar{w}^i}$ represents a CDF of $\bar{w}_k^i$. Denoting an inverse function of $F_{\bar{w}^i}$ as $F_{\bar{w}^i}^{-1}$ results in

$$\frac{x_k^i}{(1-\beta_k^i)} - \tilde{\phi}^i(B) d_{k-1}^i - d_{\text{const.}}^i \geq F_{\bar{w}^i}^{-1}(1-\delta_k^i). \tag{5.12}$$

When the CDF $F_{\bar{w}^i}$ is obtained in an analytical form, the inequality in Eq. (5.12) becomes a

linear inequality constraint on the decision variable.

By using the definition of the random variable $\bar{w}_k^i$ in Eq. (5.11), $\bar{w}_k^i$ follows the Gaussian distribution with $\mathcal{N}(\bar{\mu}_i, \bar{\sigma}_i^2)$, where

$$\bar{\mu}_i = \left(1 - \sum_{j=1}^{m_i} \theta_j\right) \mu_i, \quad \bar{\sigma}_i^2 = \left(1 - \sum_{j=1}^{m_i} \theta_j^2\right) \sigma_i^2.$$

For the Gaussian with $\mathcal{N}(\bar{\mu}_i, \bar{\sigma}_i^2)$, the CDF is represented by

$$F_{\bar{w}^i}(x) = \frac{1}{2}\left(1 + \mathrm{erf}\left(\frac{x - \bar{\mu}_i}{\sqrt{2\bar{\sigma}_i^2}}\right)\right),$$

where $\mathrm{erf}(\cdot)$ is the error function defined by

$$\mathrm{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-t^2} \, \mathrm{d}t.$$

Therefore:

$$F_{\bar{w}^i}^{-1}(x) = \bar{\mu}_i + \sqrt{2\bar{\sigma}_i^2}\, \mathrm{erf}^{-1}(2x - 1).$$

By using this function and adding the time horizon index τ, the inequality in Eq. (5.12) is transformed to

$$\frac{x_{k+\tau}^i}{(1 - \beta_{k+\tau}^i)} - \tilde{\phi}^i(B) d_{k+\tau-1}^i - d_{\text{const.}}^i \geq \bar{\mu}_i + \sqrt{2\bar{\sigma}_i^2}\, \mathrm{erf}^{-1}(1 - 2\delta_{k+\tau}^i). \tag{5.13}$$

For the given dynamical model in Eq. (5.2) and $\bar{\mu}_i, \bar{\sigma}_i^2, \beta_{k+\tau}^i$, and $\delta_{k+\tau}^i$, the inequality in Eq. (5.13) is the linear constraint on the resource allocation $x_{k+\tau}^i$.

5.3.3 Relaxed Optimization Problem

From the previous subsections, Problem 5.1 is reduced to the following problem with the deterministic linear constraint.

Problem 5.2 (Relaxed chance-constrained network optimization problem)**.**
For given $N_{\mathrm{U}}, N_{\mathrm{S}}, N_{\mathrm{G}}, B_{\mathrm{S}}^j, B_{\mathrm{ISL}}^j, B_{\mathrm{G}}^\ell$, the time series of past requests $\{d_{k-\tau}^i\}_{0 \leq \tau \leq n_i r_i - 1}$, the

future values of $\{\bar{\mathcal{C}}_{k+\tau}, \bar{\mathcal{H}}_{k+\tau}, \bar{\mathcal{E}}_{k+\tau}, d^i_{k+\tau}\}_{\tau\in\mathcal{T}_p}$, and the ARIMA model in Eq. (5.2), solve

$$\begin{aligned} \underset{\substack{\{\boldsymbol{C}_{k+\tau},\boldsymbol{H}_{k+\tau},\boldsymbol{E}_{k+\tau},\boldsymbol{Y}_{k+\tau},\\ \boldsymbol{V}_{k+\tau},\boldsymbol{Z}_{k+\tau},\boldsymbol{u}_{k+\tau-1}\}_{\tau\in\mathcal{T}_p}}}{\text{minimize}} \quad & \sum_{\tau\in\mathcal{T}_p} J_{k+\tau} \\ \text{subject to} \quad & \text{Eqs. (4.2) – (4.10), (5.8), (5.13)}, \ \ \forall\tau\in\mathcal{T}_p,\ i\in\mathcal{N}_{\mathrm{U}} \end{aligned}$$

Problem 5.2 can be solved more efficiently than Problem 5.1 because the nonlinear constraint is not included. Although Problem 5.2 is still the mixed integer programming problem, this type of problem can be efficiently solved using the branch-and-cut algorithm. This is one of the heuristic algorithms [118], as stated in Chapter 4. If Problem 5.2 becomes infeasible at a certain time because the constraint in Eq. (5.13) is violated, this constraint is neglected and the total bandwidth loss rate ($\sum_{i\in\mathcal{N}_{\mathrm{U}}} b^i_k$) is directly minimized instead of the cost in Eq. (5.6).

Problem 5.2 is solved at every time k by the SATCOM operator to obtain the optimal resource allocation and network structure. Using this result, the operator manages and controls the satellites and GWs to meet the stochastic and time-varying UT requests.

Remark 5.1. The relaxation attained by using Lemma 5.1 produces a gap between Problem 5.1 and Problem 5.2 unless $T_p = 1$ and this gap increases with T_p. Thus, the solution obtained by solving Problem 5.2 is suboptimal because the obtained solution is actually not optimal for Problem 5.1.

5.4 Numerical Experiments

In this section, numerical experiments are conducted to verify the performance of the proposed chance-constrained network management methodology with the stochastic and time-varying communication requests. In this simulation, it is assumed that there are no inter-satellite links and satellites use different frequency bands to prevent interference between communication links.

The optimization problem is described by the MATLAB YALMIP [114] and solved by the Gurobi Optimizer [119]. This solver implements the branch-and-cut algorithm to efficiently solve the mixed integer programming problem. These experiments were conducted using an Intel(R) Xeon(R) E5-2687W CPU at 3.10 GHz with 64 GB RAM.

Table 5.1: Simulation parameters of UT and GW.

N_{U}	N_{G}	B_{G}^{ℓ} [MHz] $(\forall \ell \in \mathcal{N}_{\mathrm{G}})$
50	5	500

Table 5.2: Simulation parameters of satellites.

Type of satellites	N_{S}	B_{S}^{j} [MHz] $(\forall \ell \in \mathcal{N}_{\mathrm{S}})$	Visible period [steps]
GEO	3	500	-
NGEO group 1	24	250	5
NGEO group 2	18	250	6
(in total)	45		

5.4.1 Simulation Conditions

The simulation parameters are shown in Tables 5.1 and 5.2. The simulation was conducted in $T = 50$ steps and the prediction horizon was set to $T_p = 5$.

The dynamics of the UT requests in Eq. (5.1) was set to the ARIMA model of order $(1, 1, 2)$. For all the UTs, the coefficients ϕ_1, θ_1, and θ_2 were generated from the uniform distributions on $[0.1, 0.2], [-0.2, -0.1]$, and $[-0.5, -0.4]$, respectively. Furthermore, $d_{\text{const.}}^{i} = 0.5$ and $(\mu_i, \sigma_i^2) = (0.4, 5)$ for all $i \in \mathcal{N}_{\mathrm{U}}$.

There are a total of 45 satellites, which includes three GEO satellites and 42 NGEO satellites divided into two groups. As with the setting in Chapter 4, the satellites can connect to all UTs and GWs over a continuous period of visibility. The GEO satellites are visible throughout the entire simulation time interval, whereas the visibility periods of the NGEO satellites are limited to finite intervals and repeated periodically. The time-varying connectable pair candidate sets $\bar{\mathcal{C}}_k$ and $\bar{\mathcal{E}}_k$ and the total satellite capacity are calculated based on the visibility schedule.

Figure 5.1 shows the total time-varying UT requests and the total satellite and GW bandwidth. The figure indicates that after $k = 43$, the SATCOM system could not supply enough resources to meet all the UT requests.

The reliability level Δ^i is set as $\Delta^i = 0.5, 0.3, 0.1, 0.05$, and 0.01 for all $i \in \mathcal{N}_{\mathrm{U}}$, and results for each Δ^i are compared. As shown in Fig. 5.2, the level Δ^i is allocated to each prediction horizon as higher reliability is required in the near future. This means that δ_{k+1}^{i} is the smallest among $\tau \in \mathcal{T}_p$. This setting satisfies the inequalities in Eq. (5.8) as

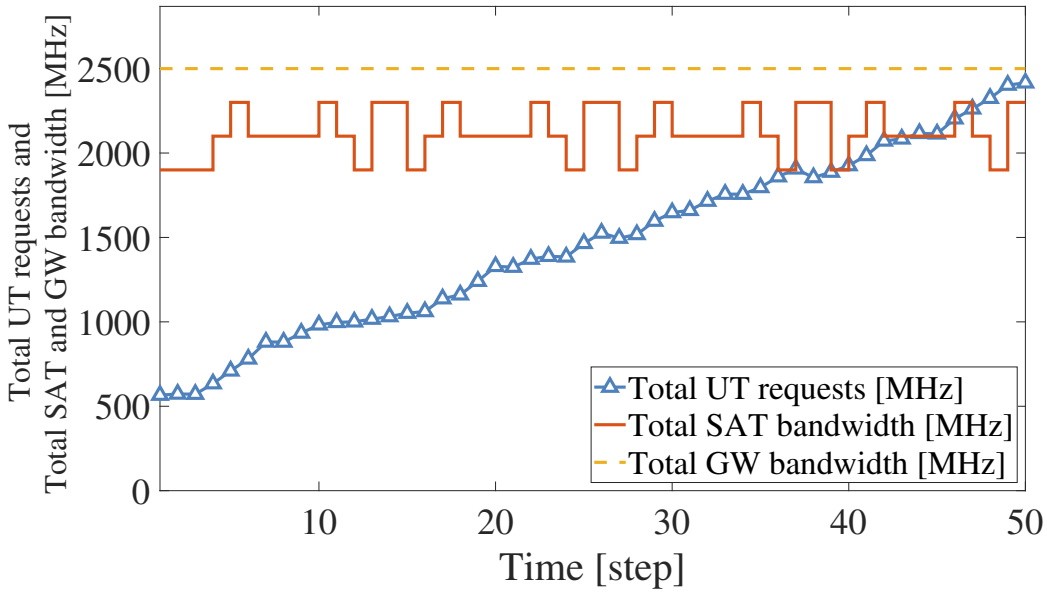


Fig. 5.1: Time-varying total UT requests (blue line), total satellite bandwidth (red line), and total GW bandwidth (dashed yellow line).

$\sum_{\tau\in\mathcal{T}_p}\delta^i_{k+\tau} = \Delta^i$ $(\forall i \in \mathcal{N}_{\mathrm{U}})$. Furthermore, the bandwidth loss rate threshold is set to $\beta^i_k = 0.1$ for all the UTs throughout the simulation time interval. In the cost function in Eq. (5.6), the weights are set as $w_c = 10, w_e = 1$, and $w_h = 0$ for $k \geq 2$.

To evaluate the performance of the proposed chance-constrained network design methodology, the number of the achieved user-specified events in Eq. (5.4) through the simulation is defined as

$$\bar{N} = \sum_{i\in\mathcal{N}_{\mathrm{U}}}\sum_{k=1}^{T} N^i_k,$$

where

$$N^i_k = \begin{cases} 1, & \text{if } \mathfrak{B}^i_k \text{ is achieved,} \\ 0, & \text{otherwise.} \end{cases}$$

An achieved rate is expressed by $\bar{N}/(N_{\mathrm{U}}T)$, which represents how many times the specified performance is achieved in all trials. Furthermore, a baseline methodology is introduced for comparison. In this baseline methodology, the chance constraint is simply replaced by the corresponding deterministic ones: $b^i_{k+\tau} \leq \beta^i_{k+\tau}, \forall i \in \mathcal{N}_{\mathrm{U}}, \tau \in \mathcal{T}_p$.

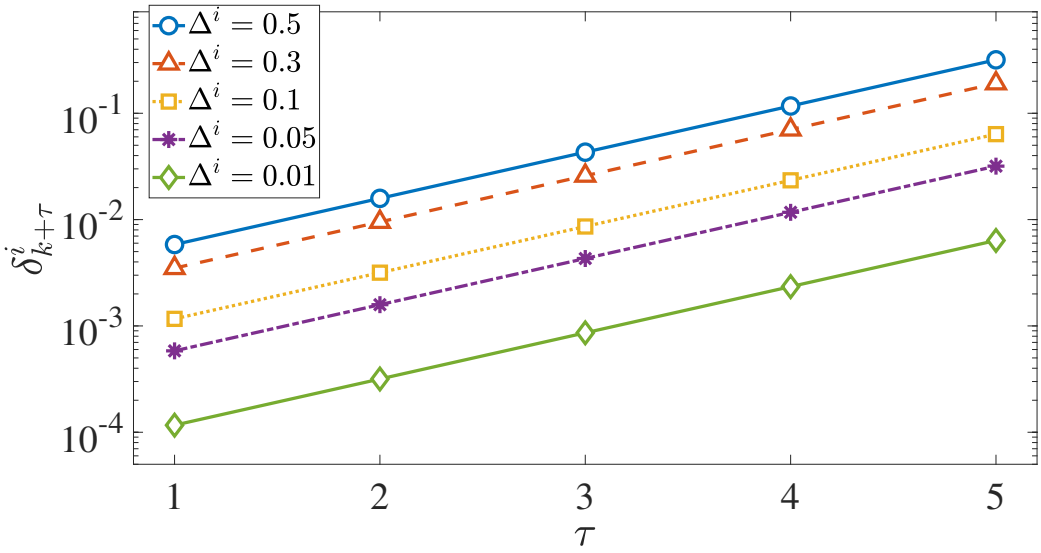


Fig. 5.2: Reliability allocation for each Δ^i when $T_p = 5$.

5.4.2 Results

First, the author compares the results of the proposed method with $\Delta^i = 0.1$ and the comparison method. In Fig. 5.3, the red and blue lines represent the resulting total bandwidth loss $\sum_{i \in \mathcal{N}_\mathrm{U}} \max(0, d^i_k - x^i_k)$ of the proposed method with $\Delta^i = 0.1$ and the comparison method, respectively. Note that the bandwidth loss was high after $k = 43$ for both methods because the SATCOM system could not supply enough resources to meet all the UT requests. The achieved rate was 69.8% for the proposed method and 36.2% for the comparison method, and thus the proposed method outperformed the comparison method.

Next, the performance of the five different Δ^i settings are compared. Table 5.3 shows a comparison of the achieved rate of the events represented by $\bar{N}/(N_\mathrm{U}T)$ for each Δ^i. These results indicate that the achieved rate was increased with smaller Δ^i, which means that the specified performance achieved higher reliability. As a reference, the achieved rate of the comparison method was 36.2%. Thus, the result of the proposed method with $\Delta^i = 0.01$ performed 36.5 percentage points higher than that of the comparison method.

These results demonstrate the effectiveness of the proposed chance-constrained method in managing the stochastic and time-varying requests in the large-scale SATCOM system.

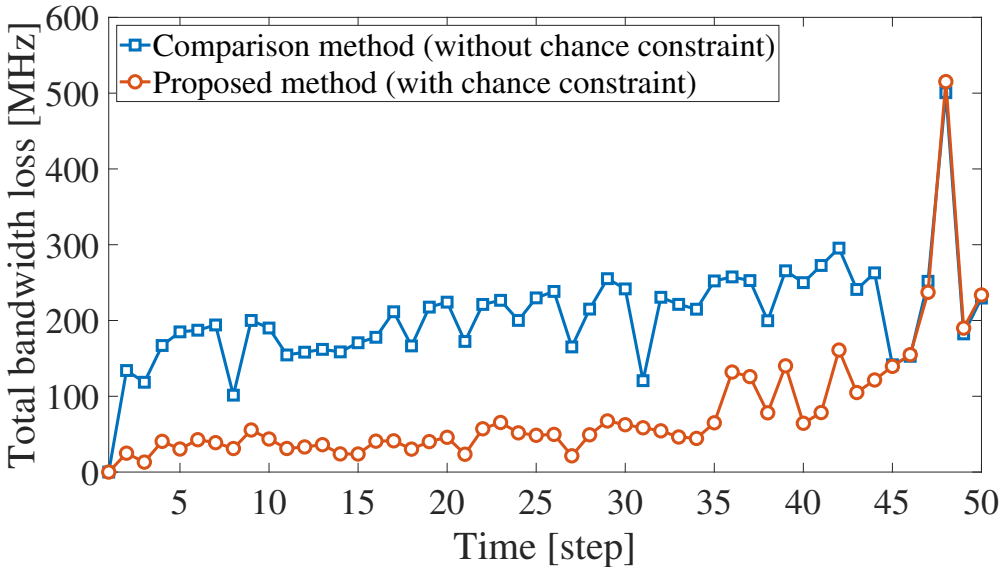


Fig. 5.3: Total bandwidth loss of the proposed method with $\Delta^i = 0.1$ (red line) and the comparison method (blue line).

Table 5.3: Achieved rate of the event with Δ^i.

$\Delta^i(\forall i \in \mathcal{N}_\mathrm{U})$	Achieved rate: $\bar{N}/(N_\mathrm{U}T)$
(Comparison method)	(36.2%)
0.5 (50% reliability)	66.1%
0.3 (70% reliability)	67.0%
0.1 (90% reliability)	69.8%
0.05 (95% reliability)	70.6%
0.01 (99% reliability)	72.7%

5.5 Summary

In this chapter, the author proposed the chance-constrained resource and network management method for large-scale SATCOM systems with stochastic and time-varying communication requests. To efficiently solve the formulated optimization problem with the joint chance constraint on the bandwidth loss rate, the constraint is relaxed to the individual chance constraint and then reduced to the deterministic constraint. From the results of the numerical simulation, the effectiveness of the proposed method to obtain an efficient SATCOM network was verified. Furthermore, the user-specified event in Eq. (5.4) was achieved with higher reliability when a smaller Δ^i is adopted.

In future work, the reliability allocation $\sum_{\tau \in \mathcal{T}_p} \delta_k^i$ in Eq. (5.8) will be optimized. Furthermore, the gap between Problem 5.1 and Problem 5.2 caused by the relaxation will be analyzed and reduced.

Chapter 6

Conclusion and Future Work

In this chapter, conclusions of this thesis and future work are provided.

6.1 Conclusion

In this thesis, the author proposed management frameworks for large-scale SATCOM systems. The proposed management strategy enables the two major difficulties of large-scale SATCOM systems to be addressed: the time-variability within the system components and connecting a large number of heterogeneous components, which are detailed in Section 1.4.

First of all, in Chapter 1, the author explained the reasons for the increased SATCOM demand in recent years by introducing the applications of SATCOM systems. Then, the recent SATCOM research trends aiming to meet the increasing SATCOM demand were summarized. The latest research aims to increase the communication capacity and improve the frequency utilization efficiency because of the current lack of frequency resources. Furthermore, the necessity of the large-scale SATCOM systems and the difficulties of realizing these systems were described. To tackle the difficulties, the author proposed the following management frameworks for SATCOM resources and networks.

In Chapter 3, the author proposed the management framework for the frequency flexibility to efficiently allocate the frequency resources to each beam for coping with time-varying communication requests from users. The dynamical models of the HTS, which increases the capacity by using multiple beams, and those of the communication requests were introduced. By incorporating the dynamical models and combining MPC and sparse optimiza-

tion, the proposed method obtains the optimal solution that utilizes the limited bandwidth efficiently and can cope with the time-varying requests. Numerical experiments showed that the proposed method reduced the bandwidth loss and NCA in multiple scenarios, such as the communication demand originating from aircraft.

In Chapter 4, the author proposed the management framework to construct an efficient SATCOM network connected with a large number of system components consisting of multiple satellites, user terminals, and gateway stations connected via user links, inter-satellite links, and feeder links. The proposed method generates efficient resource allocation for UTs and SATCOM network structure and allows the management of multiple GEO and NGEO satellites in a unified manner. In numerical experiments simulating time-varying UT requests, the proposed method in the integrated system, which allows UTs to have several candidate satellites to connect, outperformed the non-integrated system, which allows UTs to connect only to predetermined satellites. Furthermore, increasing the number of satellites and inter-satellite links increased the performance of the proposed method.

Finally, in Chapter 5, the author proposed the chance-constrained approach to managing the stochastic and time-varying communication requests. This method solves the optimization problem that includes the chance constraint satisfying the bandwidth loss rate with user-specified reliability. In the formulation of the optimization problem, the nonlinear joint chance constraint was relaxed to the linear deterministic one, which in turn allowed the network optimization problem to be solved sub-optimally. This management strategy can be applied to the management framework for both the frequency flexibility in Chapter 3 and the massive connected networks in Chapter 4.

The proposed management frameworks contribute to the efficient arrangement of large-scale SATCOM systems that optimally allocate frequency resources among SATCOM networks connected with a large number of satellites, user terminals, and gateway stations. Furthermore, implementing the proposed strategies enables communication links to meet various types of communication needs in future communication systems.

6.2 Future Work

The proposed management framework for large-scale SATCOM systems may be further extended to enable more flexible operations and to cope with various types of communication needs as described below.

Management Strategy Including Incentive Mechanism

The proposed management strategy can include an incentive mechanism. The SATCOM operator should prepare multiple operation strategies to utilize the limited resources of the large-scale SATCOM system. In some cases, the total user requests exceed the total satellite resources when the SATCOM system includes a large number of users, and each user requires a large amount of communication resources. To overcome this situation, incorporating an incentive mechanism into the SATCOM system may work to reduce the strain on the capacity of the SATCOM system.

In the literature on power systems, a smart grid incorporating an incentive mechanism, such as a dynamic pricing, is proposed [121, 122]. Dynamic pricing is a method designed to control user power consumption by controlling prices according to the supply and demand balance. Applying this to the context of SATCOM systems, the incentive mechanism can be incorporated into a management scheme for communication requests from users (this is equivalent to drawing an arrow from the SATCOM operator to the users in Fig. 1.6). In this system, the price of allocating resources should be defined and a user preference function with respect to the price should be implemented. When the total user communication requests exceed the total satellite resources, the operator increases the price of the satellite resources. Users who have high sensitivity to price increases can set the user preference function to decrease or stop communication if the price reaches a certain level. By continuously operating this mechanism in the SATCOM system, the operator can ensure that the total satellite resources exceed the total user requests. This mechanism can be analyzed according to game theory, which is the study and analysis of strategic interaction among competitive decision-makers [123].

The author will model SATCOM systems that include an incentive mechanism. The most important aspects of this system are the price determined by the SATCOM operator and

the user preference function, which defines user actions. Then, an optimization problem will be formulated to obtain a more efficient resource allocation and network structure even when the total satellite resources exceed the total number of user requests. This can be expanded to a system concept where users select a SATCOM operator from multiple operators to receive communication services based on their operation strategies.

Multi-Satellite Beamforming

In a system with a large number of satellites using the same frequency band, there is a possibility of interference caused by beam overlapping and unwanted out-of-band emission. To avoid this interference, the system should allow satellites to form flexible beams by installing digital beamformers (DBFs), as introduced in Subsection 1.3.2. By utilizing digital beamforming functions, satellites can construct multiple beam positions and beamwidths for multiple situations [41, 124]. Thus, when the beam position and frequency of GEO and NGEO satellites overlap, the system can avoid the interference by changing not only the power of the beam, but also the beam position or the beamwidth.

The author will analyze the effect of interference when there is a large number of satellites in a SATCOM system and these satellites use the same frequency band. This analysis will lead to an understanding of the amount of interference that is caused in each resource allocation and network structure. Then, an optimization problem will be formulated to determine the optimal beamforming or optimal network structure to avoid this interference.

Security

Security is an important factor in SATCOM systems because it ensures the privacy of communication links. In large-scale SATCOM systems with many satellites connected, the impact of cyber-attacks may extend to the entire system, including the terrestrial system, even though only a single satellite may be infected. Thus, space security is needed to keep the systems and signals confidential [125, 126], especially for governmental communication links.

To provide secure satellite communication links, quantum cryptography methods such as quantum key distribution (QKD) play an important role [127, 128]. The world's first

quantum satellite, named Micius, was launched in 2016 and a number of QKD-based experiments have been conducted with it [129–131]. In addition, quantum transmission can be conducted by a 50-kg-class microsatellite [132]. Secure links such as the latter two examples are required in large-scale SATCOM systems with multiple system components connected. Secure schemes should be applied to not only the physical layers, but also to others layers such as the data links or networks.

The author will define the degree of security in large-scale SATCOM systems. Then, an optimization problem will be established to obtain a SATCOM network structure that maximizes the degree of security. To include the quantum cryptography methods in this optimization, models of communication signals or networks which are consistent with these methods should be defined. This result will provide a strategy to make communication links more confidential.

Bibliography

[1] B. R. Elbert, *Introduction to Satellite Communication*. Artech House, 2008.

[2] Euroconsult, *High Throughput Satellites: On Course for New Horizons*. Euroconsult Executive Report, 2014.

[3] H. Hauschildt, C. Elia, H. L. Moeller, and J. M. P. Armengol, "HydRON: High throughput optical network," in *Proceedings of SPIE 10910, Free-Space Laser Communications XXXI*, 2019, pp. 109100K–1–9.
DOI: 10.1117/12.2511391

[4] Y. Hu and V. O. Li, "Satellite-based Internet: a tutorial," *IEEE Communications Magazine*, vol. 39, no. 3, pp. 154–162, 2001.
DOI: 10.1109/35.910603

[5] J. A. Bruder, "IEEE radar standards and the radar systems panel," *IEEE Aerospace and Electronic Systems Magazine*, vol. 28, no. 7, pp. 19–22, 2013.
DOI: 10.1109/MAES.2013.6559377

[6] H. Bischl *et al.*, "Adaptive coding and modulation for satellite broadband networks: From theory to practice," *International Journal of Satellite Communications and Networking*, vol. 28, no. 2, pp. 59–111, 2010.
DOI: 10.1002/sat.932

[7] H. Kaushal and G. Kaddoum, "Optical communication in space: Challenges and mitigation techniques," *IEEE Communications Surveys & Tutorials*, vol. 19, no. 1, pp. 57–96, 2017.
DOI: 10.1109/COMST.2016.2603518

[8] M. Gregory, F. Heine, H. Kämpfner, R. Lange, M. Lutzer, and R. Meyer, "Commercial optical inter-satellite communication at high data rates," *Optical Engineering*, vol. 51, no. 3, pp. 57–96, 2012.
DOI: 10.1117/1.OE.51.3.031202

[9] H. Nishiyama, Y. Tada, N. Kato, N. Yoshimura, M. Toyoshima, and N. Kadowaki, "Toward optimized traffic distribution for efficient network capacity utilization in two-layered satellite networks," *IEEE Transactions on Vehicular Technology*, vol. 62,

no. 3, pp. 1303–1313, 2013.
DOI: 10.1109/TVT.2012.2227861

[10] L. Bai, L. Zhu, X. Zhang, W. Zhang, and Q. Yu, "Multi-satellite relay transmission in 5G: Concepts, techniques, and challenges," *IEEE Network*, vol. 32, no. 5, pp. 38–44, 2018.
DOI: 10.1109/MNET.2018.1800038

[11] Airbus, "Global market forecast 2019-2038 *Cities, Airports & Aircraft*," 2019, (accessed January 21st, 2020).

[12] Airbus, "Global market forecast 2017-2036 *Growing Horizons*," 2017, (accessed May 23rd, 2018).

[13] E. Dinc, M. Vondra, S. Hofmann, D. Schupke, M. Prytz, S. Bovelli, M. Frodigh, J. Zander, and C. Cavdar, "In-flight broadband connectivity: Architectures and business models for high capacity air-to-ground communications," *IEEE Communications Magazine*, vol. 55, no. 9, pp. 142–149, 2017.
DOI: 10.1109/MCOM.2017.1601181

[14] M. Vondra, E. Dinc, M. Prytz, M. Frodigh, D. Schupke, M. Nilson, S. Hofmann, and C. Cavdar, "Performance study on seamless DA2GC for aircraft passengers toward 5G," *IEEE Communications Magazine*, vol. 55, no. 11, pp. 194–201, 2017.
DOI: 10.1109/MCOM.2017.1700188

[15] X. Huang, J. A. Zhang, R. P. Liu, Y. J. Guo, and L. Hanzo, "Airplane-aided integrated networking for 6G wireless: Will it work?" *IEEE Vehicular Technology Magazine*, vol. 14, no. 3, pp. 84–91, 2019.
DOI: 10.1109/MVT.2019.2921244

[16] S. Hofmann, A. E. García, D. Schupke, H. E. Gonzálezy, and F. H. Fitzekz, "Connectivity in the air: Throughput analysis of air-to-ground systems," in *Proceedings of the 2019 IEEE International Conference on Communications*, 2019.
DOI: 10.1109/ICC.2019.8761108

[17] P. K. Chowdhury, M. Atiquzzaman, and W. Ivancic, "Handover schemes in satellite networks: state-of-the-art and future research directions," *IEEE Communications Surveys & Tutorials*, vol. 8, no. 4, pp. 2–14, 2006.
DOI: 10.1109/COMST.2006.283818

[18] A. Zolich, D. Palma, K. Kansanen, K. Fjørtoft, J. Sousa, K. H. Johansson, Y. Jiang, H. Dong, and T. A. Johansen, "Survey on communication and networks for autonomous marine systems," *Journal of Intelligent & Robotic Systems*, 2018.
DOI: 10.1007/s10846-018-0833-5

[19] M. Höyhtyä, J. Huusko, M. Kiviranta, K. Solberg, and J. Rokka, "Connectivity for autonomous ships: Architecture, use cases, and research challenges," in *Proceedings of the 2017 International Conference on Information and Communication Technology Convergence*, 2017, pp. 345–350.
DOI: 10.1109/ICTC.2017.8191000

[20] M.-T. Zhou, V. D. Hoang, H. Harada, J. S. Pathmasuntharam, H. Wang, P.-Y. Kong, C.-W. Ang, Y. Ge, and S. Wen, "TRITON: high-speed maritime wireless mesh network," *IEEE Wireless Communications*, vol. 20, no. 5, pp. 134–142, 2013.
DOI: 10.1109/MWC.2013.6664484

[21] J. Wang, H. Zhou, Y. Li, Q. Sun, Y. Wu, S. Jin, T. Q. S. Quek, and C. Xu, "Wireless channel models for maritime communications," *IEEE Access*, vol. 6, pp. 68070–68088, 2018.
DOI: 10.1109/ACCESS.2018.2879902

[22] T. Pecorella, L. S. Ronga, F. Chiti, S. Jayousi, and L. Franck, "Emergency satellite communications: Research and standardization activities," *IEEE Communications Magazine*, vol. 53, no. 5, pp. 170–177, 2015.
DOI: 10.1109/MCOM.2015.7105657

[23] M. Berioli, A. Molinaro, S. Morosi, and S. Scalise, "Aerospace communications for emergency applications," *Proceedings of the IEEE*, vol. 99, no. 11, pp. 1922–1938, 2011.
DOI: 10.1109/JPROC.2011.2161737

[24] L. Franck, M. Berioli, P. Boutry, G. Harles, L. S. Ronga, R. Suffritti, and L. Thomasson, "On the role of satellite communications for emergency situations with a focus on europe," *International Journal of Satellite Communications and Networking*, vol. 29, no. 5, pp. 387–399, 2011.
DOI: 10.1002/sat.979

[25] M. Casoni, C. A. Grazia, M. Klapez, N. Patriciello, A. Amditis, and E. Sdongos, "Integration of satellite and lte for disaster recovery," *IEEE Communications Magazine*, vol. 53, no. 3, pp. 47–53, 2015.
DOI: 10.1109/MCOM.2015.7060481

[26] Local Authorities Satellite Communications Organization, "A report on the great east japan earthquake and status of the use of the regional satellite communication network," 2012 (in Japanese) (accessed January 21st, 2020). [Online]. Available: http://www.lascom.or.jp/information/report/higashinihon

[27] A. Al-Fuqaha, M. Guizani, M. Mohammadi, M. Aledhari, and M. Ayyash, "Internet of things: A survey on enabling technologies, protocols, and applications," *IEEE*

Communications Surveys & Tutorials, vol. 17, no. 4, pp. 2347–2376, 2015.
DOI: 10.1109/COMST.2015.2444095

[28] Cisco, "Cisco visual networking index: Forecast and trends, 20172022 white paper," 2019 (accessed January 21st, 2020). [Online]. Available: https://www.cisco.com/c/en/us/solutions/collateral/service-provider/visual-networking-index-vni/white-paper-c11-741490.html

[29] I. F. Akyildiz and A. Kak, "The internet of space things/cubesats: A ubiquitous cyber-physical system for the connected world," *Computer Networks*, vol. 150, pp. 134–149, 2019.
DOI: 10.1016/j.comnet.2018.12.017

[30] M. D. Sanctis, E. Cianca, G. Araniti, I. Bisio, and R. Prasad, "Satellite communications supporting internet of remote things," *IEEE Internet of Things Journal*, vol. 3, no. 1, pp. 113–123, 2016.
DOI: 10.1109/JIOT.2015.2487046

[31] Z. Qu, G. Zhang, H. Cao, and J. Xie, "LEO satellite constellation for internet of things," *IEEE Access*, vol. 5, pp. 18391–18401, 2017.
DOI: 10.1109/ACCESS.2017.2735988

[32] M. Hasan and C. Bianchi, "Ka band enabling technologies for high throughput satellite (HTS) communications," *International Journal of Satellite Communications and Networking*, vol. 34, no. 4, pp. 483–501, 2016.
DOI: 10.1002/sat.1161

[33] Y. Vasavada, R. Gopal, C. Ravishankar, G. Zakaria, and N. BenAmmar, "Architectures for next generation high throughput satellite systems," *International Journal of Satellite Communications and Networking*, vol. 34, no. 4, pp. 523–546, 2016.
DOI: 10.1002/sat.1175

[34] H. Fenech, A. Tomatis, S. Amos, V. Soumpholphakdy, and J. L. S. Meri, "Eutelsat HTS systems," *International Journal of Satellite Communications and Networking*, vol. 34, no. 4, pp. 503–521, 2016.
DOI: 10.1002/sat.1171

[35] M. Ángel Vázquez, A. Pérez-Neira, D. Christopoulos, S. Chatzinotas, B. Ottersten, P.-D. Arapoglou, A. Ginesi, and G. Tarocco, "Precoding in multibeam satellite communications: Present and future challenges," *IEEE Wireless Communications*, vol. 23, no. 6, pp. 88–95, 2016.
DOI: 10.1109/MWC.2016.1500047WC

[36] Intelsat, "The intelsat epic platform: High throughput, high performance to support next-generation requirements," (accessed January 21st,

2020). [Online]. Available: http://www.intelsat.com/wp-content/uploads/2016/03/Intelsat-Epic-Positioning-6493-wp.pdf

[37] A. Ornés, "Satellite services inmarsat," (accessed January 21st, 2020). [Online]. Available: https://www.itu.int/en/ITU-D/Emergency-Telecommunications/Documents/Bogota_2012/presentation/PresentationAlejandraOrnesEn.pdf

[38] K. Kaneko, H. Nishiyama, N. Kato, A. Miura, and M. Toyoshima, "Construction of a flexibility analysis model for flexible high-throughput satellite communication systems with a digital channelizer," *IEEE Transactions on Vehicular Technology*, vol. 67, no. 3, pp. 2097–2107, 2018.
DOI: 10.1109/TVT.2017.2736010

[39] S. Tani, K. Motoyoshi, H. Sano, A. Okamura, H. Nishiyama, and N. Kato, "Flexibility-enhanced HTS system for disaster management: Responding to communication demand explosion in a disaster," *IEEE Transactions on Emerging Topics in Computing*, Early Access.
DOI: 10.1109/TETC.2017.2688078

[40] D. Gupta, T. V. Filippov, A. F. Kirichenko, D. E. Kirichenko, I. V. Vernik, A. Sahu, S. Sarwana, P. Shevchenko, A. Talalaevskii, and O. A. Mukhanov, "Digital channelizing radio frequency receiver," *IEEE Transactions on Applied Superconductivity*, vol. 17, no. 2, pp. 430–437, 2007.
DOI: 10.1109/TASC.2007.898255

[41] Q.-Y. Yu, W.-X. Meng, M.-C. Yang, L.-M. Zheng, and Z.-Z. Zhang, "Virtual multi-beamforming for distributed satellite clusters in space information networks," *IEEE Wireless Communications*, vol. 23, no. 1, pp. 95–101, 2016.
DOI: 10.1109/MWC.2016.7422411

[42] Y. Rahmat-Samii and A. C. Densmore, "Technology trends and challenges of antennas for satellite communication systems," *IEEE Transactions on Antennas and Propagation*, vol. 63, no. 4, pp. 1191–1204, 2015.
DOI: 10.1109/TAP.2014.2366784

[43] C. G. Christodoulou, Y. Tawk, S. A. Lane, and S. R. Erwin, "Reconfigurable antennas for wireless and space applications," *Proceedings of the IEEE*, vol. 100, no. 7, pp. 2250–2261, 2012.
DOI: 10.1109/JPROC.2012.2188249

[44] S. Shi, G. Li, Z. Li, H. Zhu, and B. Gao, "Joint power and bandwidth allocation for beam-hopping user downlinks in smart gateway multibeam satellite systems," *International Journal of Distributed Sensor Networks*, vol. 13, no. 5, pp. 1–11, 2017.
DOI: 10.1177/1550147717709461

[45] I. Aykin, M. Krunz, and Y. Xiao, "Adaptive frequency-hopping schemes for cr-basedmulti-link satellite networks," *International Journal of Satellite Communications and Networking*, vol. 36, no. 4, pp. 315–331, 2018.
DOI: 10.1002/sat.1235

[46] J. Foust, "SpaceX's space-Internet woes: Despite technical glitches, the company plans to launch the first of nearly 12,000 satellites in 2019," *IEEE Spectrum*, vol. 56, no. 1, pp. 50–51, 2019.
DOI: 10.1109/MSPEC.2019.8594798

[47] M. Harris, "Tech giants race to build orbital internet," *IEEE Spectrum*, vol. 55, no. 6, pp. 10–11, 2018.
DOI: 10.1109/MSPEC.2018.8362213

[48] I. del Portillo, B. G. Cameron, and E. F. Crawley, "A technical comparison of three low earth orbit satellite constellation systems to provide global broadband," *Acta Astronautica*, vol. 159, pp. 123–135, 2019.
DOI: 10.1016/j.actaastro.2019.03.040

[49] J. N. Pelton and B. Jacque, "Distributed internet-optimized services via satellite constellations," in *Handbook of Satellite Applications*, 2nd ed., J. N. Pelton, S. Madry, and S. Camacho-Lara, Eds. Springer, 2017, pp. 249–269.
DOI: 10.1007/978-3-319-23386-4_96

[50] L. Wood, A. Clerget, I. Andrikopoulos, G. Pavlou, and W. Dabbous, "IP routing issues in satellite constellation networks," *International Journal of Satellite Communications*, vol. 19, no. 1, pp. 69–92, 2001.
DOI: 10.1002/sat.655

[51] S. K. Sharma, S. Chatzinotas, and B. Ottersten, "In-line interference mitigation techniques for spectral coexistence of GEO and NGEO satellites," *International Journal of Satellite Communications and Networking*, vol. 34, no. 1, pp. 11–39, 2016.
DOI: 10.1002/sat.1090

[52] M. Höyhtyä, A. Mämmelä, X. Chen, A. Hulkkonen, J. Janhunen, J.-C. Dunat, and J. Gardey, "Database-assisted spectrum sharing in satellite communications: A survey," *IEEE Access*, vol. 5, pp. 25322–25341, 2017.
DOI: 10.1109/ACCESS.2017.2771300

[53] Nanosats Database, (accessed January 21st, 2020). [Online]. Available: https://www.nanosats.eu/

[54] F. Davoli, C. Kourogiorgas, M. Marchese, A. Panagopoulos, and F. Patrone, "Small satellites and cubesats: Survey of structures, architectures, and protocols," *Interna-*

tional Journal of Satellite Communications and Networking, vol. 37, no. 4, pp. 343–359, 2019.
DOI: 10.1002/sat.1277

[55] H. Heidt, J. Puig-Suari, A. S. Moore, S. Nakasuka, and R. J. Twiggs, "Cubesat: A new generation of picosatellite for education and industry low-cost space experimentation," in *Proceedings of the 14th Annual AIAA/USU Conference on Small Satellites*, 2000, pp. 1–19. [Online]. Available: https://digitalcommons.usu.edu/smallsat/2000/All2000/32/

[56] NASA, "What are smallsats and cubesats?" (accessed January 21st, 2020). [Online]. Available: https://www.nasa.gov/content/what-are-smallsats-and-cubesats

[57] B. Doncaster, J. Shulman, J. Bradford, and J. Olds, "SpaceWorks' 2016 nano/microsatellite market forecast," in *Proceedings of the 30th Annual AIAA/USU Conference on Small Satellites*, 2016, pp. 1–6. [Online]. Available: https://digitalcommons.usu.edu/smallsat/2016/TS2Launch/1/

[58] D. P. Scharf, F. Y. Hadaegh, and S. R. Polen, "A survey of spacecraft formation flying guidance and control (Part I): Guidance," in *Proceedings of the 2003 American Control Conference*, 2003, pp. 1733–1739.
DOI: 10.1109/ACC.2003.1239845

[59] D. P. Scharf, F. Y. Hadaegh, and S. R. Polen, "A survey of spacecraft formation flying guidance and control (Part II): Control," in *Proceedings of the 2004 American Control Conference*, 2004, pp. 2976–2985.
DOI: 10.23919/ACC.2004.1384365

[60] M. Toyoshima, "Trends in satellite communications and the role of optical free-space communications," *Journal of Optical Networking*, vol. 4, no. 6, pp. 300–311, 2005.
DOI: 10.1364/JON.4.000300

[61] C. Fuchs and F. Moll, "Ground station network optimization for space-to-ground optical communication links," *Journal of Optical Communications and Networking*, vol. 7, no. 12, pp. 1148–1159, 2015.
DOI: 10.1364/JOCN.7.001148

[62] H. Zech, F. Heine, D. Trndle, S. Seel, M. Motzigemba, R. Meyer, and S. Philipp-May, "LCT for EDRS: LEO to GEO optical communications at 1.8 Gbps between alphasat and sentinel 1a," in *Proceedings of the SPIE Unmanned/Unattended Sensors and Sensor Networks XI; and Advanced Free-Space Optical Communication Techniques and Applications*, vol. 9647, 2015, pp. 96470J–1–8.
DOI: 10.1117/12.2196273

[63] S. Chatzinotas, B. Evans, A. Guidotti, V. Icolari, E. Lagunas, S. Maleki, S. K. Sharma, D. Tarchi, P. Thompson, and A. Vanelli-Coralli, "Cognitive approaches to enhance spectrum availability for satellite systems," *International Journal of Satellite Communications and Networking*, vol. 35, no. 5, pp. 407–442, 2017.
DOI: 10.1002/sat.1197

[64] C. Zhang, C. Jiang, L. Kuang, J. Jin, Y. He, and Z. Han, "Spatial spectrum sharing for satellite and terrestrial communication networks," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 55, no. 3, pp. 1075–1089, 2019.
DOI: 10.1109/TAES.2018.2889585

[65] L. Zhang, M. Xiao, G. Wu, M. Alam, Y.-C. Liang, and S. Li, "A survey of advanced techniques for spectrum sharing in 5G networks," *IEEE Wireless Communications*, vol. 24, no. 5, pp. 44–51, 2017.
DOI: 10.1109/MWC.2017.1700069

[66] L. Boero, R. Bruschi, F. Davoli, M. Marchese, and F. Patrone, "Satellite networking integration in the 5G ecosystem: Research trends and open challenges," *IEEE Network*, vol. 32, no. 5, pp. 9–15, 2018.
DOI: 10.1109/MNET.2018.1800052

[67] R. Gopal and N. BenAmmar, "Framework for unifying 5G and next generation satellite communications," *IEEE Network*, vol. 32, no. 5, pp. 16–24, 2018.
DOI: 10.1109/MNET.2018.1800045

[68] SaT5G, (accessed January 21st, 2020). [Online]. Available: https://www.sat5g-project.eu

[69] SATis5, (accessed January 21st, 2020). [Online]. Available: https://satis5.eurescom.eu

[70] 3GPP, (accessed January 21st, 2020). [Online]. Available: https://www.3gpp.org/

[71] 5GPPP, (accessed January 21st, 2020). [Online]. Available: https://5g-ppp.eu/

[72] J. Liu, Y. Shi, Z. M. Fadlullah, and N. Kato, "Space-air-ground integrated network: A survey," *IEEE Communications Surveys & Tutorials*, vol. 20, no. 4, pp. 2714–2741, 2018.
DOI: 10.1109/COMST.2018.2841996

[73] Y. Su, Y. Liu, Y. Zhou, J. Yuan, H. Cao, and J. Shi, "Broadband LEO satellite communications: Architectures and key technologies," *IEEE Wireless Communications*, vol. 26, no. 2, pp. 55–61, 2019.
DOI: 10.1109/MWC.2019.1800299

[74] M. Bacco, P. Cassar, M. Colucci, and A. Gotta, "Modeling reliable M2M/IoT traffic over random access satellite links in non-saturated conditions," *IEEE Journal on Selected Areas in Communications*, vol. 36, no. 5, pp. 1042–1051, 2018.
DOI: 10.1109/JSAC.2018.2832799

[75] A. Donner, C. Kissling, and R. Hermenier, "Satellite constellation networks for aeronautical communication: traffic modelling and link load analysis," *IET Communications*, vol. 4, no. 13, pp. 1594–1606, 2010.
DOI: 10.1049/iet-com.2009.0260

[76] Y. Abe, H. Tsuji, A. Miura, and S. Adachi, "Frequency resource management based on model predictive control for satellite communications system," *IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences*, vol. E101-A, no. 12, pp. 2434–2445, 2018.
DOI: 10.1587/transfun.E101.A.2434

[77] Y. Abe, M. Ogura, H. Tsuji, A. Miura, and S. Adachi, "Resource and network management framework for a large-scale satellite communications system," *IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences*, vol. E103-A, no. 2, pp. 492–501, 2020.
DOI: 10.1587/transfun.2019EAP1088

[78] J. M. Maciejowski, *Predictive Control with Constraints*. Prentice Hall, 2000.

[79] Y. C. Eldar and G. Kutyniok, *Compressed Sensing: Theory and Applications*. Cambridge University Press, 2012.

[80] K. Hayashi, M. Nagahara, and T. Tanaka, "A user's guide to compressed sensing for communications systems," *IEICE Transactions on Communications*, vol. E96-B, no. 3, pp. 685–712, 2013.
DOI: 10.1587/transcom.E96.B.685

[81] E. J. Candès and M. B. Wakin, "An introduction to compressive sampling," *IEEE Signal Processing Magazine*, vol. 25, no. 2, pp. 21–30, 2008.
DOI: 10.1109/MSP.2007.914731

[82] A. Kyrgiazos, B. G. Evans, and P. Thompson, "On the gateway diversity for high throughput broadband satellite systems," *IEEE Transactions on Wireless Communications*, vol. 13, no. 10, pp. 5411–5426, 2014.
DOI: 10.1109/TWC.2014.2339217

[83] S. Dimitrov, S. Erl, B. Barth, S. Jaeckel, A. Kyrgiazos, and B. G. Evans, "Radio resource management techniques for high throughput satellite communication systems," in *Proceedings of the 2015 European Conference on Networks and Communi-*

cations, 2015, pp. 175–179.
DOI: 10.1109/EuCNC.2015.7194063

[84] J. Lei and M. Á. Vázquez-Castro, "Multibeam satellite frequency/time duality study and capacity optimization," *Journal of Communications and Networks*, vol. 13, no. 5, pp. 472–480, 2011.
DOI: 10.1109/JCN.2011.6112304

[85] Y. Abe, H. Tsuji, A. Miura, and S. Adachi, "Frequency resource allocation for satellite communications system based on model predictive control and its application to frequency bandwidth allocation for aircraft," in *Proceedings of the 2018 IEEE Conference on Control Technology and Applications*, 2018, pp. 165–170.
DOI: 10.1109/CCTA.2018.8511589

[86] R. K. Ahuja, T. L. Magnanti, and J. B. Orlin, *Network Flows: Theory, Algorithms, and Applications*. Prentice Hall, 1993.

[87] K. Xie, J. Cao, X. Wang, and J. Wen, "Optimal resource allocation for reliable and energy efficient cooperative communications," *IEEE Transactions on Wireless Communications*, vol. 12, no. 10, pp. 4994–5007, 2013.
DOI: 10.1109/TWC.2013.081913.121709

[88] L. Bahiense, G. C. Oliveira, M. Pereira, and S. Granville, "A mixed integer disjunctive model for transmission network expansion," *IEEE Transactions on Power Systems*, vol. 16, no. 3, pp. 560–565, 2001.
DOI: 10.1109/59.932295

[89] T. G. Crainic, "Service network design in freight transportation," *European Journal of Operational Research*, vol. 122, no. 2, pp. 272–288, 2000.
DOI: 10.1016/S0377-2217(99)00233-7

[90] T. Santoso, S. Ahmed, M. Goetschalckx, and A. Shapiro, "A stochastic programming approach for supply chain network design under uncertainty," *European Journal of Operational Research*, vol. 167, no. 1, pp. 96–115, 2005.
DOI: 10.1016/j.ejor.2004.01.046

[91] Y. Kawamoto, H. Nishiyama, N. Kato, and N. Kadowaki, "A traffic distribution technique to minimize packet delivery delay in multilayered satellite networks," *IEEE Transactions on Vehicular Technology*, vol. 62, no. 7, pp. 3315–3324, 2013.
DOI: 10.1109/TVT.2013.2256812

[92] H. S. Chang, H. S. Chang, B. W. Kim, C. G. Lee, S. L. Min, Y. Choi, H. S. Yang, D. N. Kim, and C. S. Kim, "FSA-based link assignment and routing in low-earth orbit satellite networks," *IEEE Transactions on Vehicular Technology*, vol. 47, no. 3, pp.

1037–1048, 1998.
DOI: 10.1109/25.704858

[93] A. Nemirovski and A. Shapiro, "Convex approximations of chance constrained programs," *SIAM Journal on Optimization*, vol. 17, no. 4, pp. 969–996, 2006.
DOI: 10.1137/050622328

[94] M. Ono, "Joint chance-constrained model predictive control with probabilistic resolvability," in *Proceedings of the 2012 American Control Conference*, 2012, pp. 435–441.
DOI: 10.1109/ACC.2012.6315201

[95] G. E. P. Box, G. M. Jenkins, and G. M. L. Gregory C. Reinsel, *Time Series Analysis: Forecasting and Control*, 5th ed. John Wiley & Sons, 2015.

[96] ITU-R, "Radio regulations articles (Edition of 2016)," 2016.

[97] M. Bousquet, "Satellite communications and space telecommunications frequencies," in *Handbook of Satellite Applications*, J. N. Pelton, S. Madry, and S. Camacho-Lara, Eds. Springer, 2013, pp. 239–270.
DOI: 10.1007/978-1-4419-7671-0_13

[98] D. Minoli, *Satellite Systems Engineering in an IPv6 Environment*. Auerbach Publications, 2009.

[99] D. Mayne, J. B. Rawlings, C. V. Rao, and P. Scokaert, "Constrained model predictive control: Stability and optimality," *Automatica*, vol. 36, no. 6, pp. 789–814, 2000.
DOI: 10.1016/S0005-1098(99)00214-9

[100] M. Ono, L. Blackmore, and B. C. Williams, "Chance constrained finite horizon optimal control with nonconvex constraints," in *Proceedings of the 2010 American Control Conference*, 2010, pp. 1145–1152.
DOI: 10.1109/ACC.2010.5530976

[101] P. D. Christofides, R. Scattolini, D. M. de la Peña, and J. Liu, "Distributed model predictive control: A tutorial review and future research directions," *Computers and Chemical Engineering*, vol. 51, no. 5, pp. 21–41, 2013.
DOI: 10.1016/j.compchemeng.2012.05.011

[102] A. Richards and J. P. How, "Robust distributed model predictive control," *International Journal of Control*, vol. 80, no. 9, pp. 1517–1531, 2007.
DOI: 10.1080/00207170701491070

[103] W. Heemels, K. Johansson, and P. Tabuada, "An introduction to event-triggered and self-triggered control," in *Proceedings of the 51st IEEE Conference on Decision and*

Control, 2012, pp. 3270–3285.
DOI: 10.1109/CDC.2012.6425820

[104] D. V. Dimarogonas, E. Frazzoli, and K. H. Johansson, "Distributed event-triggered control for multi-agent systems," *IEEE Transactions on Automatic Control*, vol. 57, no. 5, pp. 1291–1297, 2012.
DOI: 10.1109/TAC.2011.2174666

[105] D. L. Donoho, "Compressed sensing," *IEEE Transactions on Information Theory*, vol. 52, no. 4, pp. 1289–1306, 2006.
DOI: 10.1109/TIT.2006.871582

[106] M. Nagahara, D. E. Quevedo, and D. Nešić, "Maximum hands-off control: A paradigm of control effort minimization," *IEEE Transactions on Automatic Control*, vol. 61, no. 3, pp. 735–747, 2016.
DOI: 10.1109/TAC.2015.2452831

[107] S. K. Sharma, E. Lagunas, S. Chatzinotas, and B. Ottersten, "Application of compressive sensing in cognitive radio communications: A survey," *IEEE Communications Surveys & Tutorials*, vol. 18, no. 3, pp. 1838–1860, 2016.
DOI: 10.1109/COMST.2016.2524443

[108] S. Boyd and L. Vandenberghe, *Introduction to Applied Linear Algebra: Vectors, Matrices, and Least Squares*. Cambridge University Press, 2018.

[109] M. Athans and P. L. Falb, *Optimal Control*. Dover Publications, 1966.

[110] E. Castañeda, A. Silva, A. Gameiro, and M. Kountouris, "An overview on resource allocation techniques for multi-user MIMO systems," *IEEE Communications Surveys & Tutorials*, vol. 19, no. 1, pp. 239–284, 2017.
DOI: 10.1109/COMST.2016.2618870

[111] Z. Jiang and V. C. M. Leung, "A predictive demand assignment multiple access protocol for Internet access over broadband satellite networks," *International Journal of Satellite Communications and Networking*, vol. 21, no. 4–5, pp. 451–467, 2003.
DOI: 10.1109/ICC.2002.997385

[112] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. MIT Press, 2016.

[113] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge University Press, 2004.

[114] J. Löfberg, "YALMIP: A toolbox for modeling and optimization in MATLAB," in *Proceedings of the 2004 IEEE International Symposium on Computer Aided Control Systems Design*, 2004, pp. 284–289.
DOI: 10.1109/CACSD.2004.1393890

[115] J. F. Sturm, "Using SeDuMi 1.02, a MATLAB toolbox for optimization over symmetric cones," *Optimization Methods and Software*, vol. 11, no. 1, pp. 625–653, 1999. DOI: 10.1080/10556789908805766

[116] Ministry of Land, Infrastructure, Transport and Tourism of Japan, "Collaborative actions for renovation of air traffic systems (CARATS)," (accessed January 21st, 2020). [Online]. Available: https://www.mlit.go.jp/en/koku/koku_fr13_000000.html

[117] Federal Aviation Administration of US Department of Transportation, "Use of portable electronic devices aboard aircraft," 2017, AC No: 91.21-1D. [Online]. Available: https://www.faa.gov/documentLibrary/media/Advisory_Circular/AC_91.21-1D.pdf

[118] R. A. Stubbs and S. Mehrotra, "A branch-and-cut method for 0-1 mixed convex programming," *Mathematical Programming*, vol. 86, no. 3, pp. 515–532, 1999. DOI: 10.1007/s101070050103

[119] Gurobi Optimization, LLC, "Gurobi optimizer reference manual," (accessed January 21st, 2020). [Online]. Available: http://www.gurobi.com

[120] P. Kall and S. W. Wallace, *Stochastic Programming*. Wiley, 1995.

[121] J. Ma, J. Deng, L. Song, and Z. Han, "Incentive mechanism for demand side management in smart grid using auction," *IEEE Transactions on Smart Grid*, vol. 5, no. 3, pp. 1379–1388, 2014. DOI: 10.1109/TSG.2014.2302915

[122] Y. Xing, R. Chandramouli, and C. Cordeiro, "Price dynamics in competitive agile spectrum access markets," *IEEE Journal on Selected Areas in Communications*, vol. 25, no. 3, pp. 613–621, 2007. DOI: 10.1109/JSAC.2007.070411

[123] Z. Han, D. Niyato, W. Saad, T. Başar, and A. Hjørungnes, *Game Theory in Wireless and Communication Networks: Theory, Models, and Applications*. Cambridge University Press, 2012.

[124] W. Li, X. Huang, and H. Leung, "Performance evaluation of digital beamforming strategies for satellite communications," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 40, no. 1, pp. 12–26, 2004. DOI: 10.1109/TAES.2004.1292139

[125] C. Jiang, X. Wang, J. Wang, H.-H. Chen, , and Y. Ren, "Security in space information networks," *IEEE Communications Magazine*, vol. 53, no. 8, pp. 82–88, 2015. DOI: 10.1109/MCOM.2015.7180512

[126] D. He, X. Li, S. Chan, J. Gao, and M. Guizani, "Security analysis of a space-based wireless network," *IEEE Network*, vol. 33, no. 1, pp. 36–43, 2019.
DOI: 10.1109/MNET.2018.1800194

[127] N. Hosseinidehaj, Z. Babar, R. Malaney, S. X. Ng, and L. Hanzo, "Satellite-based continuous-variable quantum communications: State-of-the-art and a predictive outlook," *IEEE Communications Surveys & Tutorials*, vol. 21, no. 1, pp. 881–919, 2019.
DOI: 10.1109/COMST.2018.2864557

[128] R. Bedington, J. M. Arrazola, and A. Ling, "Progress in satellite quantum key distribution," *npj Quantum Information*, vol. 3, no. 30, pp. 1–13, 2017.
DOI: 10.1038/s41534-017-0031-5

[129] S.-K. Liao *et al.*, "Satellite-to-ground quantum key distribution," *Nature*, vol. 549, pp. 43–47, 2017.
DOI: 10.1038/nature23655

[130] J. Yin *et al.*, "Satellite-based entanglement distribution over 1200 kilometers," *Science*, vol. 356, no. 6343, pp. 1140–1144, 2017.
DOI: 10.1126/science.aan3211

[131] J.-G. Ren *et al.*, "Ground-to-satellite quantum teleportation," *Nature*, vol. 549, pp. 70–73, 2017.
DOI: 10.1038/nature23675

[132] H. Takenaka, A. Carrasco-Casado, M. Fujiwara, M. Kitamura, M. Sasaki, and M. Toyoshima, "Satellite-to-ground quantum-limited communication using a 50-kg-class microsatellite," *Nature Photonics*, vol. 11, pp. 502–508, 2017.
DOI: 10.1038/nphoton.2017.107

[133] M. Fazel, H. Hindi, and S. Boyd, "Rank minimization and applications in system theory," in *Proceedings of the 2004 American Control Conference*, 2004, pp. 3273–3278.
DOI: 10.23919/ACC.2004.1384521

[134] J.-B. Hiriart-Urruty and H. Y. Le, "Convexifying the set of matrices of bounded rank: applications to the quasiconvexification and convexification of the rank function," *Optimization Letters*, vol. 6, no. 5, pp. 841–849, 2012.
DOI: 10.1007/s11590-011-0304-4

Appendix A

Rank Optimization

To minimize the rank function, the following rank minimization problem is solved.

Problem A.1 (Rank minimization problem)**.**

$$\begin{aligned} &\text{minimize} && \text{rank}(\boldsymbol{X}), \\ &\text{subject to} && \boldsymbol{X} \in \mathcal{P}, \end{aligned}$$

where $\boldsymbol{X} \in \mathbb{R}^{n \times m}$ is a decision variable and $\mathcal{P}$ is a convex set denoting system constraints.

The rank minimization problem is an NP-hard problem. Thus, two heuristic methods, nuclear norm heuristics and log-det heuristics, are proposed to solve the rank minimization problem [133].

Problem A.2 (Nuclear norm heuristics for the rank minimization problem)**.**

$$\begin{aligned} &\text{minimize} && \|\boldsymbol{X}\|_* \\ &\text{subject to} && \boldsymbol{X} \in \mathcal{P}, \end{aligned}$$

where $\|\boldsymbol{X}\|_*$ represents the nuclear norm of the matrix $\boldsymbol{X}$ defined as

$$\|\boldsymbol{X}\|_* = \sum_{i=1}^{\min\{n,m\}} \sigma_i(\boldsymbol{X}),$$

where $\sigma_i(\boldsymbol{X})$ represents the i-th singular value of the matrix $\boldsymbol{X}$.

The following log-det heuristics is restricted to a square matrix $\boldsymbol{X} \in \mathbb{R}^{n \times n}$.

Problem A.3 (Log-det heuristics for the rank minimization problem)**.**

$$\begin{aligned} \text{minimize} \quad & \log\det(\boldsymbol{X} + \delta \boldsymbol{I}_n) \\ \text{subject to} \quad & \boldsymbol{X} \in \mathcal{P}, \end{aligned}$$

where $\delta > 0$ is a small regularization constant.

In both problems, the rank function is replaced by the nuclear norm or the log-det function, respectively. The nuclear norm was proven to be the best approximation of the rank function among other convex functions [134]. On the other hand, to solve Problem A.3, an iterative linearization and minimization scheme should be applied because the log-det function is not a convex function [133].

List of Achievements

Peer-Reviewed Journal Articles

[1] Y. Abe, M. Ogura, H. Tsuji, A. Miura, and S. Adachi, "Resource and Network Management Framework for a Large-Scale Satellite Communications System," *IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences*, vol. E103-A, no. 2, pp. 492–501, 2020.
DOI: 10.1587/transfun.2019EAP1088

[2] Y. Abe, H. Tsuji, A. Miura, and S. Adachi, "Frequency Resource Management Based on Model Predictive Control for Satellite Communications System," *IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences*, vol. E101-A, no. 12, pp. 2434–2445, 2018.
DOI: 10.1587/transfun.E101.A.2434

Peer-Reviewed International Conferences

[1] Y. Abe, H. Tsuji, A. Miura, and S. Adachi, "Frequency Resource Allocation for Satellite Communications System Based on Model Predictive Control and Its Application to Frequency Bandwidth Allocation for Aircraft," 2018 IEEE Conference on Control Technology and Applications (CCTA), Copenhagen, Denmark (August 22nd, 2018).
DOI: 10.1109/CCTA.2018.8511589

Domestic Conferences (in Japanese)

[1] Y. Abe, M. Ogura, H. Tsuji, A. Miura, and S. Adachi, "A Study on Design Method for Satellite Communications Network via Geometric Programming," Technical Committee on Reliable Communication and Control, The Institute of Electronics, Information and Communication Engineers (IEICE), Okinawa, Japan (December 7th, 2018).

[2] Y. Abe, H. Tsuji, A. Miura, and S. Adachi, "A Study on Modeling and Control of Resource Allocation for Satellite Communications," The fifth Multi-symposium on Control Systems, The Society of Instrument and Control Engineers (SICE), Tokyo, Japan (March 9th, 2018).

[3] Y. Abe, H. Tsuji, A. Miura, and S. Adachi, "A Study on Frequency Assignment for Satellite Communications Based on Model Predictive Control," Technical Committee on Reliable Communication and Control, The Institute of Electronics, Information and Communication Engineers (IEICE), Okinawa, Japan (December 15th, 2017).

Award

[1] Y. Abe, M. Inoue, and S. Adachi, "The 2018 Best Paper Award and Takeda Award from Society of Instrument and Control Engineers (SICE)," for "Subspace Identification Method Involving a Priori Information Characterized in Frequency Domain" (September 13th, 2018).