**Dissertation**

# Distributed Resource Optimization for NOMA Transmission in Beamforming SATCOM

**Yazhou Zhu**

der Bundeswehr
Universität München

Faculty of Electrical Engineering and Information Technology
Institute of Information Technology
Chair of Signal Processing

| | |
|---|---|
| Supervisor and first reviewer: | Professor Dr.-Ing. Andreas Knopp, MBA |
| Second reviewer: | Professor Dr.-Ing. Christian Hofmann |

January 2023

UNIVERSITÄT DER BUNDESWEHR MÜNCHEN

# Distributed Resource Optimization for NOMA Transmission in Beamforming SATCOM

**Yazhou Zhu**

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

Doktor-Ingenieurs

genehmigten Dissertation.

**Promotionsausschuss**

| | |
|---|---|
| Vorsitzender: | Universitätsprofessor Dr. rer. nat. Dr.-Ing. Stefan Schäffler |
| 1. Prüfer: | Universitätsprofessor Dr.-Ing. Andreas Knopp, MBA |
| 2. Prüfer: | Juniorprof. Dr.-Ing. Christian Hofmann |

Die Dissertation wurde am 29.09.2022 bei der Universität der Bundeswehr München eingereicht und durch die Fakultät für Elektrotechnik und Informationstechnik am 12.01.2023 angenommen. Die mündliche Prüfung fand am 17.01.2023 statt.

# Abstract

To increase the throughput and user connectivity of existing satellite communication (SATCOM) systems, this work studies the application of nonorthogonal multiple access (NOMA) transmission in beamforming (BF) based forward links, where multiple users are simultaneously served by the same beam. To effectively implement nonorthogonal SATCOM, the feeder link limitations and multibeam satellite payload constraints must be considered for BF design and power allocation (PA) optimization.

To address these challenges, distributed resource optimization strategies are investigated for BF and flexible payload power resource allocation in multigateway (multi-GW) nonorthogonal SATCOM systems. Specifically, a per-feed available power-constrained BF strategy via maximization of the worst-user signal-to-leakage-and-noise ratio (SLNR) is explored with local channel state information (CSI) for a distributed operation of GWs. As an upper-bound performance benchmark, a centralized multilayer BF strategy is processed in a central unit with full global CSI and data sharing. Moreover, a weighted sum-rate maximization-based (WSRM-based) payload power resource optimization strategy is locally applied at each GW to efficiently use payload power resources for higher performance increment, depending on the actual traffic demand. The nonconvex WSRM problem is further solved by the weighted sum-MSE minimization-based (WMMSE-based) and deep neural network-based (DNN-based) methods. Finally, an efficient user scheduling is designed to enable the operator to capture a substantial system-throughput gain. Accurate mathematical modeling for a realistic SATCOM scenario is conducted and applied to design the resource optimization strategies. The results over the realistic simulation environment show the efficiency of our strategies.

# Kurzfassung

Zur Verbesserung des Durchsatzes bestehender Satellitenkommunikationssysteme (SATCOM) untersucht diese Arbeit die Anwendung der nichtorthogonalen Mehrfachzugriffsübertragung (NOMA) in vorwärts gerichteten Verbindungen mit Beamforming (BF), bei denen mehrere Nutzer gleichzeitig durch denselben Beam versorgt werden. Um nichtorthogonale SATCOM-Systeme effektiv zu implementieren, müssen Einschränkungen der Vorwärtsverbindung als auch der Nutzlast von Multibeam-Satelliten beim Entwurf des BFs und der Optimierung der Leistungszuweisung berücksichtigt werden.

Um diese Herausforderungen zu bewältigen, werden verteilte Optimierungsstrategien für BF und flexible Nutzlastleistungszuweisungen in nichtorthogonalen SATCOM-Systemen mit mehreren Gateways (Multi-GW) untersucht. Insbesondere wird eine BF-Strategie mit beschränkter Leistung pro Einspeisung über die Maximierung des schlechtesten Nutzersignal-zu-Verlust-und-Rausch-Verhältnisses (SLNR) mit lokaler Kanalzustandsinformation (CSI) für einen verteilten Einsatz von GWs erforscht. Als oberer Leistungsmaßstab wird eine zentralisierte mehrschichtige BF-Strategie in einer Recheneinheit mit vollständiger globaler CSI und gemeinsamer Datennutzung angewandt. Darüber hinaus wird eine auf gewichteter Gesamtratenmaximierung (WSRM) basierende Optimierungsstrategie für die Nutzlastleistungsressourcen lokal in jedem GW verwendet. Diese wird für eine Leistungssteigerung in Abhängigkeit von der tatsächlichen Verkehrslast effizient genutzt. Das nicht-konvexe WSRM-Problem wird durch die WMMSE-basierte (weighted sum-MSE minimization) und DNN-basierte (deep neural network) Methode gelöst. Schließlich wird eine effiziente Nutzerplanung entwickelt, die es dem Betreiber ermöglicht, einen erheblichen Gewinn an Systemdurchsatz zu erzielen. Es wird eine genaue mathematische Modellierung für ein realistisches SATCOM-Szenario durchgeführt und zur Entwicklung der Ressourcenoptimierungsstrategien angewendet. Die Ergebnisse aus der realistischen Simulationsumgebung zeigen die Leistungsfähigkeit unserer Strategien.

# Acknowledgments

Since it is not always a wonderful journey toward a Ph.D., I really appreciate the following persons for their enthusiastic guidance, help, and company during the journey.

First, I would like to express my sincere gratitude and admiration for my supervisor Prof. Dr.-Ing. Andreas Knopp, who has provided me with precious research opportunity and responsible supervision during the past years. I greatly appreciate my supervisor, Andreas for giving me the academic freedom and the great support to do this research work. Especially, even though this work begins without specific funds, Andreas still supports my dissertation work. Many thanks to you, Andreas. I owe you so much that cannot easily be repaid.

Moreover, I gratefully thank Prof. Dr.-Ing. Christian Hofmann for giving his expert comments on this work and opening his door to my questions. Especially, his comments make the simulation results meaningful. Furthermore, I appreciate Simon Heine for helping me with the German translation of the abstract.

Last but not least, I thank my wife Yan Xiao along with my two children Muyi and Sven for their company during this journey.

Yazhou Zhu,

Munich, January, 2023

# Contents

# List of Figures

# List of Tables

# Acronyms, Operators and Symbols

**List Of Acronyms:**

| | |
|---|---|
| **B5G** | beyond-fifth-generation |
| **BF** | beamforming |
| **BpaMod** | **B**eam-level payload **p**ower resource-**a**llocation **Mod**ule |
| **C/I** | carrier-to-interference ratio |
| **CDF** | cumulative distribution function |
| **CPU** | central processing unit |
| **CSI** | channel state information |
| **DNN** | deep neural network |
| **DVB_S2X** | digital video broadcasting satellite second generation extended |
| **DVB_S2** | the second-generation standard for digital video broadcasting over satellite |
| **DVB** | digital video broadcasting |
| **ESA** | European Space Agency |
| **FOV** | field of view |
| **FSPL** | free-space path loss |
| **FWD** | forward |
| **G/T** | gain-to-noise temperatures |
| **GEO** | geostationary-earth-orbit |

| | |
|---|---|
| **GHz** | gigahertz |
| **GPU** | graphical processing unit |
| **GW** | gateway |
| **H/VHTS** | high or very high throughput satellites |
| **HPA** | high-power amplifier |
| **HTS** | high-throughput satellite |
| **IoT** | internet of things |
| **Km** | kilometer |
| **LMMSE** | linear mean square error |
| **LNA** | lower-noise amplifier |
| **LOS** | line of sight |
| **MDP** | Markov decision process |
| **MMSE** | minimum mean square error |
| **MPA** | multiport amplifier |
| **MU_MIMO** | multi-user multiple-input and multiple-output |
| **multi − GW** | multigateway |
| **NN** | neural network |
| **NOMA** | non-orthogonal multiple access |
| **OMA** | orthogonal multiple access |
| **PAPC** | per-feed available power-constraint |
| **PA** | power allocation |
| **PD** | power-domain |
| **RL** | reinforcement learning |
| **SATCOM** | satellite communication |

| | |
|---|---|
| **SDMA** | spatial division multiplexing access |
| **SDP** | semidefinite programming |
| **SDR** | semidefinite relaxation |
| **SIC** | successive interference cancellation |
| **SINR** | signal-to-interference-plus-noise ratio |
| **SLNR** | signal-to-leakage-and-noise ratio |
| **SNR** | signal-to-noise ratio |
| **SVD** | singular-value decomposition |
| **THP** | Tomlinson-Harasima precoding |
| **TWTA** | traveling wave tube amplifier |
| **UpaMod** | **U**ser-level payload **p**ower resource-**a**llocation **Mod**ule |
| **VHTS** | very high throughput satellites |
| **WMMSE** | weighted sum mean square error minimization |
| **WSRM** | weighted sum-rate maximization |
| **ZF** | zero-forcing |

**List Of Operators:**

| | |
|---|---|
| $\|\| \cdot \|\|_2$ | the Euclidean norm |
| $\|\cdot\|$ | the absolute value or the cardinality of a finite set relying on the context |
| $\cdot$ | the dot product |
| $(\cdot)^{\mathrm{H}}$ | the conjugate transpose |
| $(\cdot)^{\mathrm{T}}$ | the transpose |
| $\mathbb{E}[\cdot]$ | the statistical expectation |
| $\ln(\cdot)$ | the natural logarithm |

| | |
|---|---|
| $\mathrm{Tr}\,(\cdot)$ | the trace |
| $\odot$ | the element-wise product |
| **blkdiag**$[\cdot]$ | the block diagonal concatenation of matrix input arguments |
| $\times$ | the cross product |

**List Of Symbols:**

| | |
|---|---|
| $\mathbf{u}_j$ | the location vectors of user $j$ |
| $\mathbf{u}_{b,g}^{(i)}$ | the user $i$ associated with beam $b$ served by GW $g$ |
| $\mathbf{X}_{\mathbf{u}_j}, \mathbf{Y}_{\mathbf{u}_j}, \mathbf{Z}_{\mathbf{u}_j}$ | the satellite-centered coordinates for user $j$ |
| $\mathbf{X}_{\mathbf{v}_b}, \mathbf{Y}_{\mathbf{v}_b}, \mathbf{Z}_{\mathbf{v}_b}$ | the satellite-centered coordinates for feed $b$ |
| $(A_b^{(i)})_{\mathrm{dB}}$ | the magnitude measurements $(A_b^{(i)})$ in decibels (dB) |
| $(\varpi_{b,g}^{(i)})^*$ | the optimal $\varpi_{b,g}^{(i)}$ |
| $(f_{b,g}^{(i)})^*$ | the optimal $f_{b,g}^{(i)}$ |
| $(f_{b,g}^{(i)})_{\mathrm{MMSE}}^{\mathrm{opt}}$ | the optimal equalizer $f_{b,g}^{(i)}$ |
| $\delta^0$ | the predefined threshold used to measure the channel correlations of the users per beam |
| $\digamma, \beth, \daleth$ | the auxiliary variables defined as $\mathbf{Y}_{\mathbf{u}_j}\mathbf{Z}_{\mathbf{v}_b} - \mathbf{Z}_{\mathbf{u}_j}\mathbf{Y}_{\mathbf{v}_b}, \mathbf{Z}_{\mathbf{u}_j}\mathbf{X}_{\mathbf{v}_b} - \mathbf{X}_{\mathbf{u}_j}\mathbf{Z}_{\mathbf{v}_b}, \mathbf{X}_{\mathbf{u}_j}\mathbf{Y}_{\mathbf{v}_b} - \mathbf{Y}_{\mathbf{u}_j}\mathbf{X}_{\mathbf{v}_b}$ |
| $\epsilon_{b,g}^{(i)}$ | the mean-square estimation error of estimating the desired signal $x_{b,g}^{(i)}$ |
| $\kappa$ | the Boltzmann constant |
| $\lambda_i$ | the penalty coefficient |
| $\lambda_s$ | the longitude of the satellite |
| $\bar{\mathsf{T}}_j$ | the complex Gaussian random vector with zero mean and unit variance introduced to design the corresponding $\mathbf{w}_{b,g}$ |
| $\bar{\mathbf{P}}_j(:,i), \bar{\mathbf{D}}_j$ | the eigenvectors matrix and eigenvalues matrix of $\mathbf{\Theta}_{b,g}^*$ |

| | |
|---|---|
| $\mathsf{T}_j$ | the complex Gaussian random vector with zero mean and unit variance introduced to design the $j$-th inner BF vector $\mathbf{w}_{\mathrm{I}}^j$ |
| $\mho_j^*$ | the optimal solution $\mho_j$ |
| $\mathbf{\Psi}_l$ | the abbreviated neuron weight matrix of the $l$-th hidden layer |
| $\mathbf{\Theta}_{b,g}^*$ | the optimal $\mathbf{\Theta}_{b,g}$ |
| $\tilde{\mathbf{\Psi}}_L^{(n)}$ | the abbreviated neuron weight matrix of each sublayer $n$ belonging to the output layer |
| $\tilde{\mathbf{a}}_L^{(n)}$ | the abbreviated neuron output vector of each sublayer $n$ belonging to the output layer |
| $\tilde{\mathbf{A}}_L$ | the output matrix of the **UpaMod** |
| $\tilde{\mathbf{b}}_L^{(n)}$ | the abbreviated neuron bias vector of each sublayer $n$ belonging to the output layer |
| $\mathbf{a}_0$ | the output from its input layer |
| $\mathbf{a}_L$ | the output from its output layer |
| $\mathbf{a}_l$ | the abbreviated neuron output vector from the $l$-th hidden layer |
| $\mathbf{b}_l$ | the abbreviated neuron bias vector of the $l$-th hidden layer |
| $\mathbf{P}_j(:,i), \mathbf{D}_j$ | the $i$-th eigenvector and eigenvalues of $\mho_j^*$ |
| $\mathbf{\Psi}_{b,g}, \mathbf{\Theta}_{b,g}$ | the auxiliary matrices introduced to design the corresponding $\mathbf{w}_{b,g}$ |
| $\mathcal{C}_{b,g}^{(j)}$ | the user group $i$ within beam $b$ served by GW $g$ |
| $\mathcal{G}$ | the total set of GWs |
| $\mathcal{M}_{b,g}$ | the NOMA user candidates |
| $\mathcal{N}(\mu,\sigma)$ | the normal distribution with mean $\mu$ and variance $\sigma$ |
| $\mathcal{N}_g$, $N_g$ | the subset of feeds and the total number of feeds used by GW $g \in \mathcal{G}$ |

| | |
|---|---|
| $\mathcal{S}_{b,g}$ | the group of selected users in beam $b \in \mathcal{N}_g$ |
| $\mathcal{T}$ | the training set |
| $\text{rank}(\mathbf{A})$ | the rank value of a matrix $\mathbf{A}$ |
| $\text{SIC}_{b,g}^{(i)\to(k)}$ | the variable for rate conditions of successful SIC decoding |
| $\rho$ | the power allocation coefficient for the strong user presented in Section 1.3.1 |
| $\theta\left(\mathbf{u}_j, \mathbf{v}_b\right)$ | the angle between location vectors $\mathbf{u}_j$ and $\mathbf{v}_b$ |
| $\theta_{\text{3dB}}$ | the 3dB angle |
| $\theta_{b,n}^{(i)}$ | the off-axis angle of user $i$ in beam $b$, with respect to the axis of maximum power radiated by feed $n$ |
| $\theta$ | the satellite elevation angle |
| $\tilde{x}_{b,g}^{(i)}$ | the estimated data symbol after conducting the equalization at user $i$ in beam $b$ served by GW $g$ |
| $\Upsilon_{b,g}$ | the group of users belonging to beam $b$ served by GW $g$ |
| $\varepsilon_1, \varepsilon_2, \varepsilon_3$ | the given solution accuracy of solving the corresponding optimization problem for the centralized BF, distributed BF and mathematical optimization-based payload PA strategies |
| $\varphi_b, \lambda_b$ | the latitude and longitude of feed $b$ |
| $\varphi_j, \lambda_j$ | the latitude and longitude of user $j$ |
| $\varpi_{b,g}^{(i)}$ | the optimization variables for the mathematical optimization-based payload PA strategy |
| $\varsigma_{b,g}^{(i)}$ | the user-priority factor for user $i$ in beam $b$ served by GW $g$ |
| $\vartheta_p^{\text{Circle}}, \phi_p^{\text{Circle}}$ | the true view angles captured for a reference point $p$ of the circle after the rotation |
| $\vartheta_b, \phi_b$ | the true view angles of the satellite antenna feed $b$ |
| $\vartheta_{p'}, \phi_{p'}$ | the service zone contour of a spot beam $b$ |

| | |
|---|---|
| $\wp_m$ | the $m$-th column of the $N_g \times N_g$ identity matrix |
| $B$ | the carrier bandwidth |
| $c_0$ | the speed of light |
| $c_{b,g}^{(j)}$ | the mean channel gains of the users assigned to user group $\mathcal{C}_{b,g}^{(j)}$ |
| $d$ | the slant range between the satellite and on-ground terminal |
| $f$ | the downlink operating frequency |
| $f_{\text{Linear}}(\cdot)$ | the linear transfer function |
| $f_{\text{Relu}}(\cdot)$ | the rectified linear unit (ReLU) function |
| $f_{\text{Softmax}}(x_i)$ | the softmax transfer function |
| $f_{\tanh}(\cdot)$ | the Tanh function |
| $G(u)$ | the Bessel antenna pattern model |
| $G(\mathbf{u}_j, \mathbf{v}_b)$ | the beam gain from feed $b$ to user $\mathbf{u}_j$ |
| $G^{\max}$ | the satellite transmit antenna gain at the beam center |
| $G_{b,n}^{(i)}$ | the radiation pattern from the $n$-th antenna toward user $i$ in beam $b$ |
| $h^{(\text{s})}, h^{(\text{w})}$ | the channel coefficient of the strong user and the weak user presented in Section 1.3.1 |
| $h_0$ | the satellite altitude |
| $h_{b,n}^{(i)}$ | the channel coefficient from the satellite antenna $n$ to the $i$-th user in beam $b$ |
| $I_{\text{Kmeans}}$ | the number of iterations needed in the K-means algorithm |
| $I_{\text{rand}}, \dot{I}_{\text{rand}}$ | the amount of randomization for the centralized BF and distributed BF strategies |
| $J_1, J_3$ | the order one and three Bessel functions of the first kind |
| $K$ | the total number of beams |

| | |
|---|---|
| $m_{\mathrm{SDP}}, n_{\mathrm{SDP}}, \varepsilon$ | the dimensions of a semidefinite matrix variable, the number of linear constraints and the given solution accuracy |
| $N$ | the total number of feeds |
| $N_u$ | the total number of beams selected within each beam |
| $p_{b,g}^{(i)}$ | the transmit power assigned to user $i$ in beam $b$ served by GW $g$ |
| $r_b^{'}$ | the distance between the satellite and the spot beam center $b$ on the Earth |
| $r_e$ | the radius of the Earth |
| $r_o$ | the satellite altitude above the equator |
| $R_{\mathrm{NOMA}}^{(\mathrm{s})}, R_{\mathrm{NOMA}}^{(\mathrm{w})}$ | the data rates of NOMA for the strong user and the weak user presented in Section 1.3.1 |
| $R_{b,g}^{(i)\to(k)}, \mathrm{SINR}_{b,g}^{(i)\to(k)}$ | the attainable rate and SINR at the strong user $k$ for decoding the signal of the weak user $i$ |
| $R_{b,g}^{(i)}, \mathrm{SINR}_{b,g}^{(i)}$ | the rate and SINR obtained at user $i$ for decoding its own signal |
| $S_L$ | the total number of neurons in the output layer |
| $S_l$ | the number of neurons in hidden layer $l$ |
| $S_L^{(n)}$ | the number of neurons in each sublayer $n$ belonging to the output layer |
| $t_{b,g}$ | the slack variable introduced to design the corresponding $\mathbf{w}_{b,g}$ |
| $X$ | the metric space with beam gain $G(\cdot)$ |
| $x_{b,g}^{(i)}$ | the signal transmitted to user $i$ in beam $b$ served by GW $g$ |
| $y_{b,g}^{(i)}$ | the signal received at user $i$ in beam $b$ served by GW $g$ |
| $z_{\mathrm{I}}^{j}, \mho_j$ | the slack variable and auxiliary variable introduced to design the $j$-th inner BF $\mathbf{w}_{\mathrm{I}}^{j}$ |
| $z_{b,g}^{(i)}$ | the additive white Gaussian noise received at user $i$ in beam $b$ served by GW $g$ |

| | |
|---|---|
| $(\epsilon_{b,g}^{(i)})^*$ | the optimal $\epsilon_{b,g}^{(i)}$ |
| $(\epsilon_{b,g}^{(i)})_{\text{MMSE}}^{\min}$ | the minimum mean-square error estimation |
| $\bar{\mathbf{w}}_{b,g}$ | the auxiliary vector introduced to design the corresponding $\mathbf{w}_{b,g}$ |
| $\tilde{\mathbf{H}}_k$ | the auxiliary channel $\left[\mathbf{H}_1^{(R),T}, \cdots \mathbf{H}_{k-1}^{(R),T}, \mathbf{H}_{k+1}^{(R),T}, \cdots \mathbf{H}_N^{(R),T}\right]^T$ |
| $\mathbf{A}_{\text{rotation}}$ | the rotation of an angle $\vartheta_b$ around the $y-$axis |
| $\mathbf{B}_{\text{rotation}}$ | the rotation of an angle $\phi_b$ around the $x-$axis |
| $\mathbf{H}$ | the entire channel matrix of the selected users |
| $\mathbf{H}^{(j)}$ | the $j$-th training sample of $\mathbf{H}$ |
| $\mathbf{H}^{(R)}$ | the regularized channel $\mathbf{H}\mathbf{H}^H + \frac{NN_u}{P_{\text{total}}}\mathbf{I}$ |
| $\mathbf{H}_k^{(R)}$ | the auxiliary channel created by removing the corresponding column $k$ of the regularized channel $\mathbf{H}^{(R)}$ |
| $\mathbf{h}_{b,g}^{(i)}$ | the aggregate channel vector between GW $g$ and its served user $i$ |
| $\mathbf{H}_{\text{eff}}^j$ | the $j$-th effective channel or beam radiation pattern formed with the $j$-th outer BF $\mathbf{W}_{\text{O}}^j$ |
| $\mathbf{v}_b$ | the center point of the beam radiated from feed $b$ |
| $\mathbf{W}_{\text{final}}$ | the aggregated BF matrix for the selected users |
| $\mathbf{W}_{\text{final}}^{(j)}$ | the $j$-th training sample of $\mathbf{W}_{\text{final}}$ |
| $\mathbf{W}_{\text{full}}$ | the full BF matrix |
| $\mathbf{W}_g$ | the BF matrix processed at GW $g$ |
| $\mathbf{W}_{\text{O}}, \mathbf{W}_{\text{I}}$ | the outer BF and the inner BF in a centralized multilayer BF strategy |
| $\mathbf{X}_p^{\text{Circle}}, \mathbf{Y}_p^{\text{Circle}}, \mathbf{Z}_p^{\text{Circle}}$ | the satellite-centered coordinates for a reference point $p$ of the circle |
| $\bar{\text{Ł}}_{b,g}$ | the $j$-th entity from the set $\{\bar{\mathbf{P}}_{b,g}\bar{\mathbf{D}}_{b,g}^{\frac{1}{2}}\bar{\mathsf{T}}_{b,g},\ \forall g, b\}$ |

| | |
|---|---|
| $\text{Ł}_j$ | the $j$-th entity from the set $\{\mathbf{P}_j\mathbf{D}_j^{\frac{1}{2}}\mathsf{T}_j,\ \forall j\}$ |
| $f_{b,g}^{(i)}$ | the linear equalizer employed at user $i$ in beam $b$ served by GW $g$ |
| $P_{\text{Total}}$ | the total power on board |
| $P_{g,m}$ | the power budget of the feed $m$ belonging to GW $g$ |
| $T_b^{(i)}$ | the receiver noise temperature |
| $U_{\text{Total}}$ | the total number of users |
| $\hat{\mathbf{w}}_{b,g}$ | the distributed BF vector for beam $b$ served by GW $g$ |
| $\overline{\mathbf{X}}_p^{\text{Circle}}, \overline{\mathbf{Y}}_p^{\text{Circle}}, \overline{\mathbf{Z}}_p^{\text{Circle}}$ | the satellite-centered coordinates for a reference point $p$ of the circle after the rotation |
| $\mathbf{H}_g^g$ | the auxiliary submatrix located at row partition $g$ and column partition $g$ of the entire channel matrix $\mathbf{H}$ |
| $\mathbf{W}_{\text{O}}^l, \mathbf{w}_{\text{I}}^l$ | the $l$-th column of the outer BF $\mathbf{W}_{\text{O}}$ and the inner BF $\mathbf{W}_{\text{I}}$ |
| $\mathbf{w}_{b,g}$ | the direction of distributed BF vector $\hat{\mathbf{w}}_{b,g}$ |
| $\frac{\partial \epsilon_{b,g}^{(i)}}{\partial f_{b,g}^{(i)}}$ | the derivative of $\epsilon_{b,g}^{(i)}$ with respect to $f_{b,g}^{(i)}$ |
| $L_{b,g}^{(i)}$ | the total leakage from user $i$ in beam $b \in \mathcal{N}_g$ to the users of all other beams |
| $:=$ | a definition |
| $\mathbb{C}$ | the set of complex numbers |
| $\mathbb{R}$ | the set of real numbers |
| $\mathcal{O}$ | the big-O |

# Chapter 1

# Introduction

*This chapter first describes the background (i.e. characteristics and development) of satellite communication (SATCOM) systems. Next, I discuss the evolution of the high throughput satellite, which is provisioned by deploying beamforming or precoding in multibeam SATCOM. To improve the throughput of the beamforming (BF) based SATCOM, the non-orthogonal multiple access (NOMA-based) SATCOM investigated in this work has the advantage of bandwidth efficiency. Then, the state-of-the-art power-domain (PD) NOMA techniques in both terrestrial and satellite systems are reviewed. Finally, I summarize the contribution and structure of this dissertation.*

## 1.1 Background

To sustain a flexible and ubiquitously available network for backhaul and access in beyond-fifth-generation (B5G) systems, the exploitation of higher frequency bands and the adoption of novel technologies is required [1]. The cost-effective terrestrial systems are generally deployed to only cover the urban centers. As a result, the connectivity service is still lacking in rural and remote areas. Thus, it is predicted that the B5G services are not continuously available for the customers via the terrestrial systems. As a complementary solution for next-generation communication networks, there is a resurgent interest in supplying wireless access from space. Particularly, the 3rd Generation Partnership Project (3GPP) investigates the integration of SATCOM in B5G to provide customers service ubiquity, continuity and scalability from space [2].

As a fact, the ambition of providing satellite-based connectivity is originated in science fiction, which is written by Arthur C. Clarke in 1945. This concept is demonstrated by the launch of the first low-Earth-orbit (LEO) satellite *Sputnik* (October 1957) [3]. Following that, various LEO SATCOM experiments are performed with

passive space-based reflector, store-and-forward and space-based repeater technologies, which are listed in the satellite launch order as: the *SCORE* (1958), the *ECHO-I* (August, 1960), the *COURIER* (October, 1960), *TELSTAR-I* (July, 1962), *RELAY-I* (December, 1962), *WESTFORD* (May, 1963), *TELSTAR-II* (May, 1963), and the *ECHO-II* (January, 1964), *RELAY-II* (January, 1964) [4, 5].

As the basis of today's SATCOM industry, the first operational geostationary-earth-orbit (GEO) satellite *SYNCOM-II* is launched in July 1963[1]. It is followed by the GEO satellite *SYNCOM-III*, which is in service in July 1964. The first commercial GEO satellite *EARLY-BIRD* (or *INTELSAT-I*) is in orbit in April 1965, which initiates the INTELSAT fleet [4, 5]. Up to now, the mass of commercial GEO satellites supply the majority of global communications services, due to their high coverage capability. Generally, a GEO satellite of large size can cover 33% of the earth's surface. To achieve high capacity with a GEO satellite, the advanced signal processing technologies in the GEO SATCOM is explored, including (i) the evolution of SATCOM television standards: digital video broadcasting-satellite (DVB-S), the second-generation standard for digital video broadcasting over satellite (DVB-S2), and the DVB-S2's extension (DVB-S2X); (ii) the utilization of efficient adaptive coding and modulation (ACM) schemes; (iii) employment of high power gateways and high gain directional antennas at ground stations [4, 5].

To satisfy the increasing demand for broadband and higher data rate, the employment of high-throughput satellite (HTS) communication systems with wideband spectrum (e.g. the Ku and Ka-band) is indispensable. Next, I introduce a brief overview of HTS in terms of its development and the advances of precoding and beamforming in HTS systems.

## 1.2 Evolution to High Throughput SATCOM

The advancement in the deployment of HTS systems is driven by the increasing requirement for Internet services and quality of services, where multiple spot beams are generated and each spot beam serves one unique area. In general, the coverage area of each spot beam is only $1 \sim 2\%$ the size of a global beam (i.e. each spot beam is typical $70 \sim 100$ miles in diameter) [4, 5]. By distributed multiple spot beams over the satellite service area, the HTS system captures a two-fold advantage: **(i) frequency reuse** is implemented across the coverage area, where several beams reuse the same frequency resource to increase the system capacity; **(ii) high**

[1] *SYNCOM-I* failed to launch

**satellite/user antenna gain** is obtained with the very narrow beams (i.e. the antenna gain and its beamwidth are inversely proportional to each other), where the link budget is increased for small user terminals such that higher-order modulation and coding (MODCOD) is leveraged to capture a higher data rate. Consequently, HTS systems can support a multitude of diverse customer requirements by delivering higher system capacity at a lower cost of service [6].

In 2004, the first HTS *Anik-F2* is launched to provide high-speed Internet services in North America. In the following, a large of HTS systems are deployed to offer high-speed interactive services. The HTS systems can be classified by their capacity: the first generation offered up to a few tens of Gbit/s, such as *Anik-F2*, *Thaicom-4* (or *IPSTAR*), *WildBlue-1*, *Astra-1L* (or *SpaceWay-3*), *KIZUNA* (or *WINDS*), *Ciel-2*, *LHYLAS-1*, *LHYLAS-2* [4–6]. The second generation offers approximately 100 Gbit/s, for example *KA-SAT*, *Viasat-1*, *Yahsat-1B*, *Echostar-17* (or *Jupiter-1*) [4–6]. Due to a growing traffic requirement for multimedia services, the next-generation SATCOM systems must improve their capacity to keep competitive in the space internet market. Therefore, the possibility to design SATCOM systems with 1 terabit per second of capacity has been considered between 2020 and 2025 [7]. As a result of the salient evolution from HTS systems, very high throughput satellites (VHTS) [8] is suitable satellite solutions to meet such high spectral efficiency demand and decrease the cost per bit. VHTS is usually a geostationary satellite with an increased number of spot beams and multiple ground gateways. For instance, a VHTS *Konnect* is launched to provide the European fixed broadband services and in-flight connectivity businesses. *Konnect* offered a few hundreds of Gigabit per second by the key techniques, such as allocation flexibility, optimal spectrum use, and progressive ground network deployment [7]. In the forward link of such a system, terabits per second of throughput are achieved with thousands of spot beams. The launched and planned high or very high throughput satellites (H/VHTS) are selectively listed in Table 1.1, where the Ku-band and Ka-band are typically used[2].

In the SATCOM literary, the advanced optimization processing methodology for VHTS systems is introduced in [9], where some key parameters design of SATCOM systems are analyzed. As the crucial transmission strategy for implementing VHTS systems, precoding and beamforming-based interference mitigation techniques adopt a full frequency reuse (FFR) scheme to enable terabit capacity, which will be shortly discussed in the next subsection.

[2]It is Remarkable that certain HTS systems have hybrid payloads, where Ku and Ka-band coexist with other frequency bands such as C, Q, and V-band.

Table 1.1: Timeline of HTS milestones.

| Year | System | Coverage area | Frequency band |
|---|---|---|---|
| Jul. 2004 | *Anik-F2* | North America | C, Ku, Ka |
| Aug. 2005 | *Thaicom-4* (or *IPSTAR*) | Asia-Pacific | Ku |
| Dec. 2006 | *WildBlue-1* | the United States | Ka |
| May 2007 | *Astra-1L* | Europe | Ku, Ka |
| Aug. 2007 | *Spaceway-3* | North America | Ka |
| Feb. 2008 | *KIZUNA* (or *WINDS*) | Asia-Pacific | Ka |
| Dec. 2008 | *Ciel-2* | North America | Ku |
| Nov. 2010 | *LHYLAS-1* | Europe | Ku, Ka |
| Dec. 2010 | *KA-SAT* | Europe, Africa | Ka |
| Oct. 2011 | *Viasat-1* | North America | Ka |
| Apr. 2012 | *Yahsat-1B* | Europe, Africa | C, Ku, Ka |
| Jul. 2012 | *Echostar-17* (or *Jupiter-1*) | North America | Ka |
| Aug. 2012 | *LHYLAS-2* | Europe, Middle East, Africa | Ka |
| Sep. 2012 | *Astra-2F* | Europe, Middle East, Africa | Ku, Ka |
| Feb. 2013 | *Amazonas-3* (or *AMZ3*) | North America, Latin America | C, Ku, Ka |
| Sep. 2013 | *Astra-2E* | Europe, Middle East | Ku, Ka |
| Dec. 2013 | *Inmarsat-5-F1* (or *GX-1*) | Europe, Middle East, Africa | Ka |
| Feb. 2015 | *Inmarsat-5-F2* (or *GX-2*) | North America, South America | Ka |
| Aug. 2015 | *Inmarsat-5-F3* (or *GX-3*) | Asia-Pacific, West Americas | Ka |
| Sep. 2015 | *Sky-Muster-1* (or *NBN-Co-1A*) | Australia mainland, some overseas territories | Ka |
| Nov. 2015 | *Badr-7* (or *Arabsat-6B*) | Africa, Asia | Ku, Ka |
| Jan. 2016 | *Intelsat-29e* (or *IS-29e*) | North America, Latin America, | C, Ku, Ka |
| Aug. 2016 | *Intelsat-33e* (or *IS-33e*) | Europe, Africa, Asia | C, Ku, Ka |
| Oct. 2016 | *Sky-Muster-2* (or *NBN-Co-1B*) | Australia mainland, some overseas territories | Ka |
| Feb. 2017 | *Intelsat-32e* (or *IS-32e*) | Caribbean, North Atlantic | Ku |
| Apr. 2017 | *Chinasat-16* (or *Shijian-13*) | China | Ka |
| May 2017 | *Inmarsat-5-F4* (GX-4) | Europe | Ka |
| May 2017 | *SES-15* | North America | Ku, Ka |
| Jun. 2017 | *Eutelsat-172B* | Asia-Pacific | C, Ku |
| Jun. 2017 | *Viasat-2* | North America, South America | Ka |
| Jun. 2017 | *GSAT-19* | India | Ku, Ka |
| Jul. 2017 | *Intelsat-35e* (or *IS-35e*) | Europe, Sub-Sahara Africa, Americas, | C, Ku |
| Sep. 2017 | *Intelsat-37e* (or *IS-37e*) | Europe, Africa, Americas | C, Ku, Ka |
| Jan. 2018 | *SES-14* | Americas, the North Atlantic | C, Ku, Ka |
| Jan. 2018 | *Yahsat-3* | Africa, Brazil | Ka |
| Apr. 2018 | *LHYLAS-4* | Europe, Africa | Ka |
| Jun. 2018 | *SES-12* | South Asia, Asia-Pacific | Ku |
| Nov. 2018 | *GSAT-29* | India | Q, V, Ku, Ka |
| Feb. 2019 | *Nusantara-Satu* | Indonesian archipelago, South East Asia | C, Ku |
| Apr. 2019 | *Arabsat-6A* | Middle East and North Africa | Ku, Ka |
| Aug. 2019 | *LHYLAS-3* | Africa, Asia | Ka |
| Aug. 2019 | *ChinaSat-18* | China | Ka |
| Nov. 2019 | *Inmarsat-5-F5* (*GX-5*) | Global Xpress | Ka |
| Dec. 2019 | *Kacific-1* | South East Asia, Pacific Islands | Ka |
| Jan. 2020 | *Eutelsat-Konnect* | Europe, Africa | Ka |
| Oct. 2021 | *SES-17* | Americas, the Atlantic Ocean | Ka |

### 1.2.1 Beamforming and Precoding for Multibeam SATCOM

Currently, the multibeam SATCOM[3] systems employ a partial four-color frequency reuse scheme to combat inter-beam interference (IBI), where adjacent beams perform on orthogonal polarizations or different subbands [10]. It is reported that the HTS Viasat-1 implements about 140 Gbit/s transmissions, where the whole bandwidth is divided into two different subbands and two orthogonal polarizations [11]. Unfortunately, the scarce frequency resources are not efficiently employed. VHTS networks are anticipated to provide terabit capacity by fully reusing frequency resources. However, the performance of such terabit VHTS deteriorates by the increasing inter-beam and inter-gateway interference in the forward (FWD) link. To this end, multi-user multiple-input and multiple-output (MU-MIMO) beamforming and precoding [12] are studied as powerful interference mitigation techniques for the implementation of the terabit VHTS. In this direction, the following subsection reviews advances in precoding and beamforming-based interference suppression techniques for the FWD link of multibeam satellite systems.

#### A. Unicast Beamforming

Typically, multibeam satellite systems rely on deploying a phased array antenna (e.g., an array-fed reflector) on the satellite so that multiple beams[4] are accordingly generated to achieve the spatial division multiplexing access (SDMA) [13, 14]. Thus, multiple spatially multiplexed data streams can be transmitted to multiple users, resulting in high throughput of SATCOM systems with full frequency reuse [10]. The challenge is that the increased interference level usually exists at the receiver side. To address the limitation of the increased interference, precoding and BF techniques are investigated in the FWD of multi-beam SATCOM systems. In multibeam SATCOM systems, precoding is equivalently referred to as BF in the case of a single-feed-per-beam architecture [12]. However, BF sometimes specifies the formation of beam-radiation patterns in the configuration of multiple feeds per beam, which is implemented by a feed-domain BF network[5] [15, 16]. It is worth mentioning that such BF only considers the geographical coverage of the satellite beams, whereas

[3]Multibeam SATCOM and HTS communication are used interchangeably throughout.

[4]Regarding the beam implementation, there are two possible methods: single feed per beam and multiple feeds per beam.

[5]This BF network is implemented by either an on-board (i.e., payload) or an on-ground (i.e., gateway) BF network [15, 16]. For instance, the onboard phased-array antennas are electronically steered via using on-ground beamforming (OGB) in the realistic geo-satellite systems [23]. Herein, the OGB methods enable a digital payload processor [23].

user positions are ignored [17]. Thus, beam-domain precoding is further required for the interference mitigation involving the user positions [17–19]. Remarkably, the hybrid scheme of beam-domain precoding and feed-domain BF is similar to the hybrid digital-analog BF scheme in millimeter-wave terrestrial systems [20]. Since existing HTS systems implement a single-feed-per-beam architecture [21,22], my work assumes that a single beam is generated by a single feed (i.e., there is no BF network). In the following, the focus is to be put on presenting state-of-the-art on-ground BF (or equivalently called precoding)[6] in the uniquely existing feed domain. Specifically, the MU-MIMO precoding is studied as an advanced interference mitigation technique for the FWD link by the European Space Agency (ESA) [24]. The results of the ESA studies confirm that the throughput of SATCOM systems is improved by precoding technique with the cost of more complicated processing to the gateway (GW) [24]. Moreover, nonlinear Tomlinson-Harasima precoding (THP) [25], [26] and linear precoding (e.g. zero-forcing (ZF) [25], and regularized ZF [24]) schemes are applied in multibeam SATCOM systems. Additionally, opportunistic BF [27], generic precoding optimization [28], and energy-efficient minimum mean square error (MMSE) BF [29] are studied for multibeam SATCOM systems. Apart from the abovementioned multiuser unicast BF, much work is done for multicast BF, which is presented in the sequel.

### B. Multicast Beamforming

Due to the ubiquitous accessibility provided by multibeam SATCOM systems, the same transmitted data from the satellite can be easily captured by multiple users within the footprint of a multibeam satellite. Therefore, multicast (i.e. the same data is transmitted to multiple users) scenario appears in multibeam SATCOM systems. Much fundamental work has been done for providing multicast service via multibeam satellites. Particularly, in advanced satellite standards, such as DVB-S2X standardization, the precoding-based interference management technique is supported [11,30]. Considering that the SATCOM standard DVB-S2X has a framing structure, physical layer multigroup multicast BF techniques are investigated in multibeam SATCOM systems, such as the average MMSE scheme [31], the frame-based precoding [12] and the block singular-value decomposition-based (SVD-based) precoding [32]. In addition, several challenges for the implementation of beamforming in multibeam satellite systems are pointed out. I describe them in detail in the following section.

[6]The terms “precoding” and “beamforming” are used as synonyms throughout this work.

### C. Multigateway Beamforming

Considering that the vast amount (>1000) of beams will be deployed in the forthcoming VHTS systems, the required feeder link bandwidth is accordingly increased [10]. Thus, one limitation of the BF-based VHTS systems is the increasing requirement of the feeder link bandwidth. To satisfy the requirements of feeder link bandwidth, one considered solution is to gradually move the feeder link to the higher-frequency bands like Q/V/W-bands [10]. Although these bands have a larger available bandwidth than the Ka-band operated at present, the performance at the bands is more susceptible to atmospheric attenuations [33]. Alternatively, it is feasible that multiple GWs can be utilized in the feeder link, where the Ka-band available bandwidth is reused among all the GWs with directive antennas[7]. In contrast to the single-GW BF, multiple-GW BF suffers from two-fold constraints. The first constraint is that each GW can only access a group of feeds to perform the BF-based interference mitigation. Then, the resulting block-diagonal BF matrix is resigned in the multiple-GW configuration [32]. The second constraint is that high signaling overheads for sharing the data and the channel state information (CSI) are demanded across the fully cooperative GWs. Thus, transferring to the multigateway (multi-GW) architecture, multi-GW BF develops more imperative to suppress the increased intergateway and interbeam interference, resulting from the additional execution of multiple GWs.

For multi-GW BF, a linear mean square error (LMMSE) precoding scheme via power optimization is proposed, where both full and partial cooperations among the GWs are discussed for a unicast multibeam SATCOM system [34]. Moreover, ZF-based precoding on ground is applied to eliminate inter-GW and intra-GW interference in a unicast multibeam SATCOM system [35], further complemented via devising a per-feed available power-constraint-based (PAPC-based) ZF precoding problem in [36]. Additionally, a fully on-ground multi-GW precoding via a block SVD method is proposed in [33] for both unicast and multicast multibeam SATCOM systems without a BF network. Recently, a cooperative multi-GW precoding using a multigroup multicast optimization method is investigated for multicast multibeam SATCOM systems [37]. However, the full CSI still is required to exchange between the GWs in [37]. Alternatively, the subject of a distributed precoder design is necessary in the case that full cooperation among the GWs cannot be supposed. Moreover, distributed BF processing algorithms are conventionally promising for the realization of coordinated multi-GW systems [14]. In line with this strategy, a distributed uni-

[7]This work supposes that a sufficient distance is between multiple GWs and the interference among feeder links is ignored.

cast BF[8] is proposed for the autonomous operation of the GWs [14]. However, the paper does not investigate the user fairness optimality, along with other practical constraints on SATCOM, such as the PAPC on board. Thus, unlike the previous works, a distributed multi-GW BF obeying practical SATCOM constraints is to be computed and processed in geographically decentralized areas for forthcoming multi-GW VHTS systems. This work will investigate this point.

### 1.2.2 Toward Nonorthogonal Transmission

Since satellites are essential components of B5G systems to provide seamless service everywhere, the spectrum efficiency of next-generation SATCOM systems becomes a crucial performance criterion, which needs to be further improved by efficiently exploring the scarce bandwidth. In the 5G systems, the PD NOMA is identified as a key technique to improve the throughput and the user connectivity [38, 39], such as the developed MIMO-NOMA scheme [40]. In the industry, the downlink multiuser superposition transmission (MUST) as a special case of PD-NOMA is standardized in 3rd Generation Partnership Project Long-Term Evolution (3GPP-LTE) Release-13 [41]. Furthermore, the evolutional NOMA is studied as a potential multiple access technique for the B5G wireless networks [39]. In this conceptual framework, it becomes imperative to investigate whether the integration of NOMA in BF-based FWD links can further increase the system throughput. This work will address this question.

## 1.3 NOMA Basics

In general, NOMA techniques are classified into two main categories: PD-NOMA and code-domain NOMA [41–43]. In this work, the PD-NOMA is investigated and incorporated into the FWD multibeam SATCOM. The key idea of PD-NOMA is to multiplex multiple users at the transmitter with superposition coding (SC) technique in power domain and perform multiple users detection at the user side with successive interference cancellation (SIC) technique. Contrast to traditional orthogonal multiple access (OMA), PD-NOMA[9] can support multiple access within a given time and frequency resource, by using different power levels. Therefore, the NOMA techniques including SC and SIC are mainly discussed in the following.

[8]The fully on-ground BF from [14] is to be investigated as a reference in this work, where the instantaneous CSI is adopted for a fair comparison.

[9]Henceforth, I simply term PD-NOMA as NOMA in this work

### 1.3.1 Key Features of NOMA

NOMA is studied for decades under a different expression of "SC with SIC" [44]. This different words in the work of Cover *et al.* [44] explicitly manifests that the SC and SIC techniques are key techniques of NOMA to achieve superior performance over OMA. To present the key features of NOMA, a generic two-user NOMA scheme is adopted in the FWD link of a single spot beam satellite system, as shown in Figure 1-2.

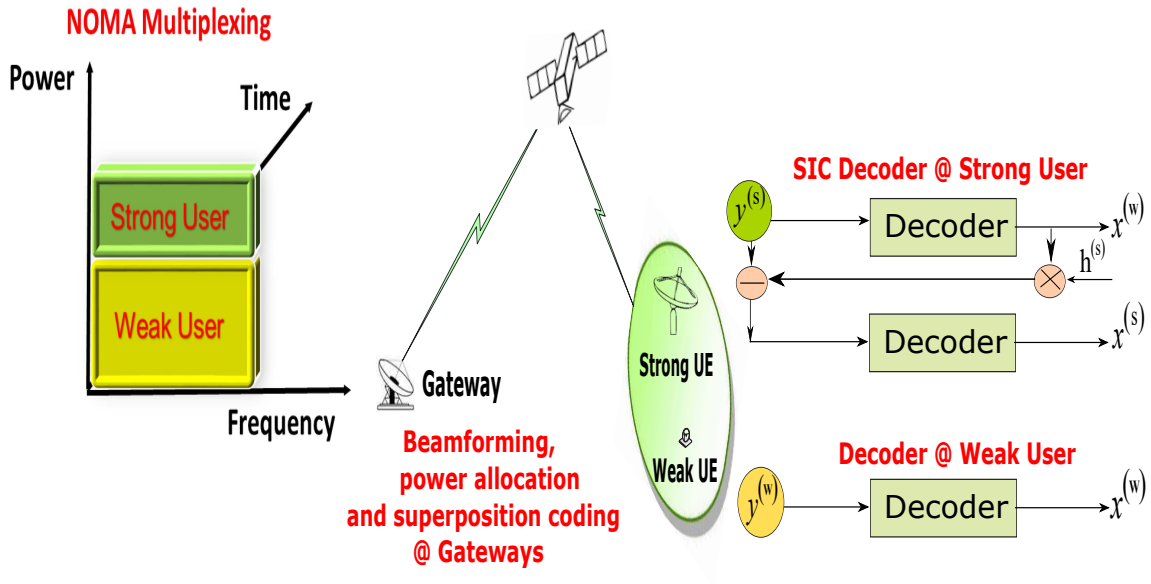


Figure 1-1: The illustration of NOMA.

#### A. SC at Transmitter

In NOMA, the transmitter superposes multiple signals in the power domain and conveys simultaneously the superposed signals over the same resources. Here, two users within a spot beam are simultaneously scheduled over the same frequency and time resource block, and their signals are multiplied with different power levels. Based on the SC principle, a two-step procedure [41–43] is performed in the following:

- *Step 1:* the signal of a weak user is encoded with a small data rate;
- *Step 2:* the signal of another strong user is superimposed with the signal of the weak user.

In this way, the orthogonality of transmit signal can be broken via exploring the power domain and result in high system throughput.

### B. SIC at Receiver

Correspondingly, the receiver (i.e. on-ground user) applies SIC to mitigate the interference caused by superposition coding. Specifically, the SIC algorithm is conducted at the strong user to accomplish interference elimination as follows [41–43]:

- *Step 1:* a strong user decodes the signal of a weak user, which is allocated with high power;
- *Step 2:* the strong user regenerates and deduct the weak user's signal from its received signal;
- *Step 3:* the strong user decodes its signal from the remaining signal after the subtraction (or interference mitigation).

Therefore, the SIC technique can reach the theoretical Shannon capacity boundary of both broadcast and multiple access channels.

### C. The Superiority of NOMA Over OMA

Then, the analysis is provided to characterize the NOMA and OMA performance with the information-theoretic method. More specifically, in Figure 1-2, the NOMA transmission is applied to the FWD link of a multibeam geo-satellite system. Let the channel coefficient of the strong user and the weak user be $h^{(\mathrm{s})}$ and $h^{(\mathrm{w})}$, respectively. For ease of illustration, four assumptions are made as follows: (1) two users are simultaneously served by a single spot beam; (2) there is no interference between the beams since the four-color frequency scheme is adopted; (3) the channel gain of a strong user is greater than that of a weak user (i.e., $|h^{(\mathrm{s})}| > |h^{(\mathrm{w})}|$); (4) the normalized signal bandwidth is 1. Following the Shannon-Hartley theorem, the data rates of the strong and weak NOMA users can be respectively given as [45]

$$R_{\mathrm{NOMA}}^{(\mathrm{s})} = \log_2(1 + \rho P|h^{(\mathrm{s})}|^2) \tag{1.1}$$

and

$$R_{\mathrm{NOMA}}^{(\mathrm{w})} = \log_2(1 + \frac{(1-\rho)P|h^{(\mathrm{w})}|^2}{\rho P|h^{(\mathrm{w})}|^2 + 1}) \tag{1.2}$$

where $\rho$ is power allocation (PA) coefficients for the strong user. $P$ is the total transmit power at the GW. The noise is normalized to 1. From equations (1.1) and

(1.2), it can be indicated that the user data rates can be controlled by allocating the power among the users. Furthermore, it is proofed through rigorous mathematical derivations that NOMA is superior to OMA with optimum PA [45]. In addition, it is well known [45] that (1) the NOMA scheme can achieve better performance than the OMA scheme in the existence of the user channel gain difference; (2) the NOMA scheme has the same performance as the OMA scheme when both user channel gains are similar or the same, as indicated in Figure 1-2. Specifically, in Figure 1-2, a channel gain difference of 11 dB is between the strong and weak users in the left subfigure, while the identical channel gain is among two strong users in the right subfigure. Without a loss of generality, the difference user channel gain of 11 dB is hence assumed to successfully apply NOMA in BF satellite systems, as further illustrated in 1.3.4.

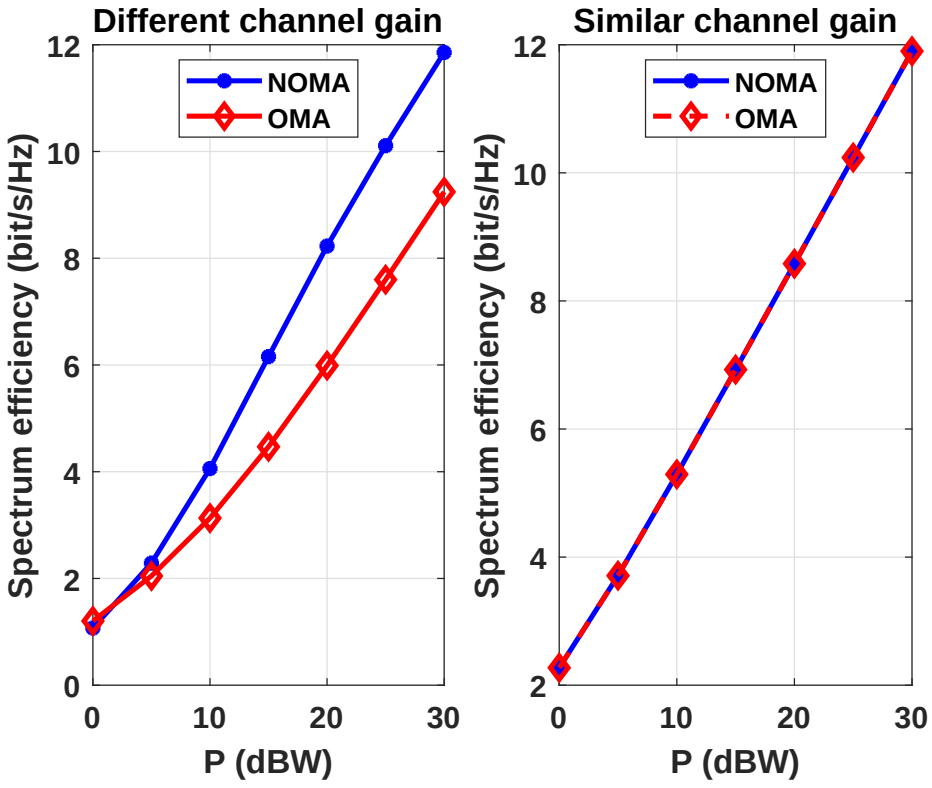


Figure 1-2: The performance of two-user NOMA: the spectrum efficiency against the transmit power. Here, a channel gain difference of 11 dB is among a strong user and a weak user in the left subfigure, while an identical channel gain is among two strong users in the right subfigure.

### 1.3.2 Terrestrial MIMO-NOMA Contributions

Since the multibeam SATCOM adopts MIMO techniques (e.g. precoding and BF) to suppress the multibeam interference, this section summarizes the compatibility of NOMA with multiple antenna technologies in the terrestrial wireless networks. More specifically, the general principles of downlink MIMO-NOMA are developed for terrestrial systems [40,46], where the orthogonal beams are generated via employing MIMO precoding techniques and interference alignment concept [47], and the NOMA principle is successively adopted to guarantee that multiple users can be served by each of the generated beams. Hence, the incorporation of NOMA into the MIMO precoding scheme can improve the system throughput. The authors in [48] consider a clustering and PA algorithm based on the correlation among users and channel gain difference, where equal PA for different groups is assumed, and intra-group PA is optimized to maximize the achievable sum rate. The intra-group power optimization method is investigated in [49], where a convex optimization algorithm is employed to obtain the closed-form solution for PA. Thus far, the aforementioned MIMO-NOMA schemes employ the same BF vectors for multiple users within each beam (it is termed as the cluster-based MIMO-NOMA scheme). Except for the cluster-based MIMO-NOMA schemes, the BF-based MIMO-NOMA schemes are also investigated in NOMA literature, where each user adopts a unique BF vector [50–52]. For the compatibility and incorporation of NOMA with the existing BF SATCOM, this work supposes that the multiple users share the same BF vectors within each beam (i.e., a cluster-based MIMO-NOMA scheme).

### 1.3.3 State-of-the-Art Nonorthogonal SATCOM

The state-of-the-art works related to non-orthogonal SATCOM are presented in this section. The application of nonorthogonal transmission to multibeam satellite systems is first studied in the work of Caus *et al.* [53], where low complexity scheduling algorithms are designed for supporting two users per beam. Recently, a general overview of how NOMA can be applied to multibeam SATCOM has been introduced [54]. Moreover, an overview for integrating NOMA in cognitive and cooperative satellite-terrestrial systems has been provided [55]. The existing research efforts also investigate the possibility of applying nonorthogonal transmission to multibeam SATCOM systems by efficiently using available network resources [56–61]. Additionally, optimization design is discussed for NOMA-based satellite-terrestrial integrated networks in [62, 63].

### 1.3.4 NOMA Application Scenario

In a non-orthogonal multi-beam SATCOM system as shown in 1-3, the users within each beam have different receiver gain to noise temperatures (G/T) towards the satellite since the users have different sizes of the antennas or different noise factors of the power amplifiers.

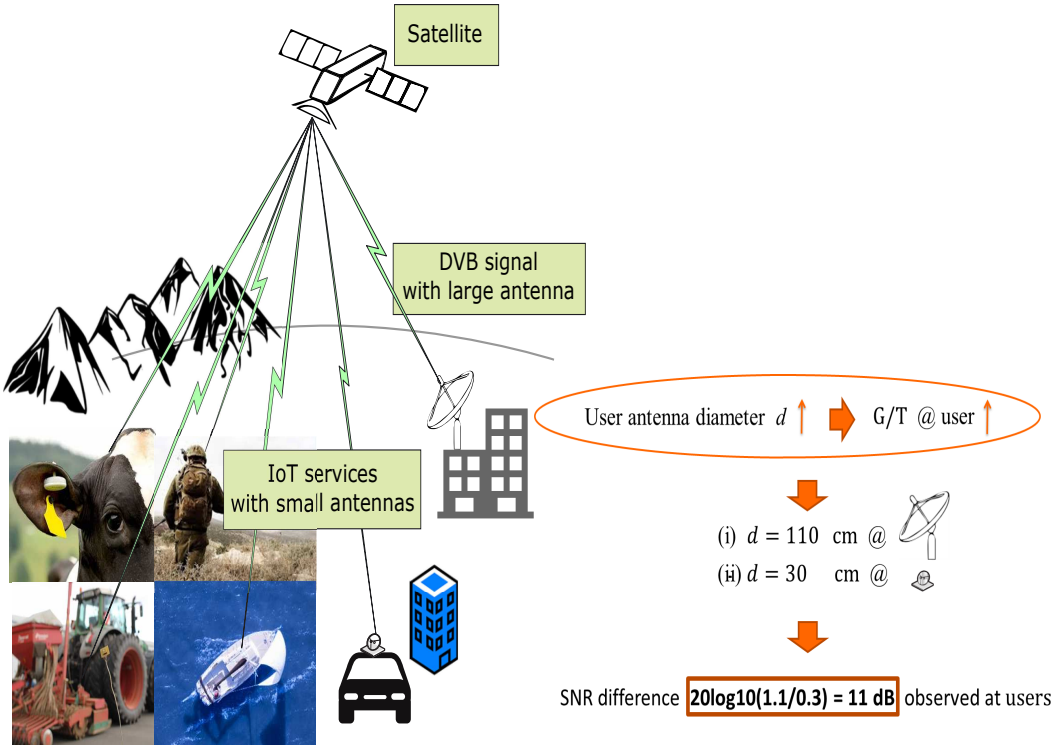


Figure 1-3: The NOMA application scenario.

For instance, this work considers two potential cases for non-orthogonal FWD transmissions:

- (i) digital video broadcasting (DVB) services via geo-satellites;
- (ii) internet of things (IoT) services, which are linked through geo-satellites.

Accordingly, a general Ka-band satellite serves two classes of users:

- (i) one class of users with an antenna diameter of 110 centimeters (cm);
- (ii) another class of users with an antenna diameter of 30 cm.

Thus, there is a large SNR imbalance $20\log_{10}(1.1/0.3) = 11$ dB between the FWD links, which are characterized by satellite line of sight (LOS-dominant) channels (i.e.

equation (2.26) in Section 2.3.4) in this work. Remarkably, a satellite LOS channel with even a small parabolic reflector antenna is observed and identified through wideband channel measurement experiments, which are conducted by Institute of Information Technology, University of the Bundeswehr Munich [64]. Ultimately, the presence of a large SNR imbalance between the wideband FWD links (e.g., at Ka-band) [10] facilitates the successful application of NOMA, as already illustrated in 1-2.

## 1.4 Motivation, Methodology and Contributions

This work is related to next-generation multiple access for SATCOM. This subsection successively introduces the motivation, methodology and contributions of this work.

***Motivation:*** *This work is to investigate the incorporation of NOMA with existing BF SATCOM* [11] *for higher throughput and more user connectivity.*

***Methodology:*** To address the above question, my methodology is to integrate and comply with the existing BF SATCOM system such that conventional infrastructures and the original capacity of the existing BF satellite are preserved, while the throughput and the amount of connectivity are improved via NOMA. First, the BF strategy under practical SATCOM constraints is compatible with the existing beam-specific BF SATCOM. After applying the BF direction-optimization strategy, I proceed to consider the additional combination of flexible payload power resources-allocation strategy, depending on the actual traffic demand. The benefits of my methodology can be summarized as follows:

- *Incorporation for higher throughput:* NOMA raises the throughput with SC and SIC;
- *Good compatibility:* NOMA is also counted as an encouraging "add-on" method for the existing BF satellite;
- *Advantage of compatibility:* preserve conventional infrastructures and original capacity of existing BF satellite, while improving the throughput and the amount of connectivity via NOMA. The compatibility and advantage are further detailed in Section 2.3.1.

[10] In addition, the coded and modulation schemes in DVB-S2X can generate different SNR values (in range from $-10$ dB to 20 dB) for the FWD links.

[11] Herein, the existing BF SATCOM is featured with (i) the fully on-ground BF for interference mitigation, and (ii) flexible on-board power resource reconfiguration for inherently uncertain traffic demand.

***Contributions:*** Due to the instinctive attributes of the VHTS system and its non-degraded multiple input single output (MISO) broadcast channel, it is nontrivial to effectively integrate NOMA into the BF SATCOM. Thus, the contributions are described in the following.

Though a non-degraded MISO broadcast channel exists in overloaded BF satellite systems, SATCOM is a relevant candidate scenario for NOMA. From an information theoretic perspective, the benefit of NOMA would not be achieved unless the MISO broadcast channel is efficiently degraded (i.e., multiple equivalent single-input-single-output channels are generated after BF for the successful SC and SIC). To achieve it, the scheduled user channels must be highly correlated within each beam to enable the effective BF-based interference mitigation. Fortunately, the users can be scheduled with the high channel correlation within each beam, which is generally inherent in a multibeam geo-satellite system (i.e., Figure 3-4 in Chapter 3). Therefore, BF SATCOM is a relevant candidate scenario for NOMA.

To this end, in overloaded BF satellite systems, *the distributed user scheduling strategy* is first developed with the customized beam association and K-means methods. Especially, the beam-association method is adopted to gather such high user channel correlation within each beam. As a result, the scheduled users are with higher channel correlation and the BF strategy can efficiently manage intergateway and interbeam interference. On the other hand, the underlying difference of the users' channel gains is exploited by the K-means algorithm to unleash the full potential of NOMA.

Furthermore, the application of NOMA to BF-based FWD links is restrained by the instinctive attributes of future VHTS systems, such as (i) the complexity proportional to the significant number ($> 1000$) of beams; (ii) PAPC on the payload; (iii) the feeder-link limitations [65]. Consequently, a direct application of terrestrial NOMA strategies to multibeam SATCOM systems is not appropriate. Hence, a nonorthogonal SATCOM demands specific signal optimization processing strategies, including *alternative multi-GW BF optimization techniques* and *flexible payload power resource allocation strategies*. As the data traffic is locally generated at each GW, the resource optimization shall be correspondingly executed at each GW in a distributed manner to facilitate its realization. Accordingly, distributed resource optimization can reduce the CSI and signaling exchange among the GWs, and distributed resource-optimization problems with small channel dimensions are more tractable. Therefore, distributed resource-optimization strategies are investigated to efficiently integrate the nonorthogonal transmission with BF in the upcoming multi-GW SATCOM. This

is different from the aforementioned works focused on a single-GW nonorthogonal SATCOM.

More specifically, regarding *multi-GW BF optimization strategies*, this work aims to apply the beam-specific BF strategy (i.e., all users per beam share the same BF vector) in the nonorthogonal SATCOM for interference management, compatible with the existing beam-specific BF SATCOM in Section 1.2.1 and is different from the user-specific BF strategy (i.e., each user per beam uses a unique BF vector) in the existing nonorthogonal SATCOM [53]. Particularly, user fairness optimization is not well investigated by the existing user-specific BF strategy while user fairness is really relevant to NOMA. Hence, a distributed multi-GW BF optimization technique is proposed in this work to optimize user fairness. As a performance limit, a centralized multilayer BF technique is also investigated to optimize user fairness.

After BF direction optimization strategy, I proceed to consider the additional combination of *the flexible payload power resources-allocation strategy* to further enhance the system's performance. In this regard, beam-level payload PA is typically optimized and intrinsically realized via multiport amplifiers (MPAs) or flexible traveling wave tube amplifiers (TWTAs) for further performance enhancement of existing BF SATCOM systems [24, 66], while the joint optimization of beam-level and user-level PA is investigated in this work. To achieve it, the power-resources allocation can be optimized with various rules, including the maximization of the sum rate (i.e., throughput) and the maximization of the minimum rate. The former has no user fairness since users with poor channel conditions are not completely served. The latter may result in maximum user fairness, in which all users are assigned the same power/rate [24]. Then, the same power/rate allocation would prevent a poor user from acquiring more power resources than any strong user and would result in an unsuccessful SIC decoding at each strong user in the nonorthogonal strategy. Thus, as a compromise between the maximization of minimum rate and maximum throughput, the weighted sum rate maximization (WSRM) is applied to accomplish the objective of the PA strategy. In addition to a mathematical optimization based strategy, a machine learning-based flexible payload power resources-allocation strategy is investigated to further avoid the signaling and real-time information exchange among the GWs and the users. Last, the contributions can be summarized as follows:

- Mathematical modeling is used to generate an exemplary multi-GW multibeam SATCOM scenario, including the near-to-real footprints, the generation of users, and interference. The system throughput and user fairness are superior comparing with state-of-the-art counterparts.

- A distributed user scheduling is developed with the customized beam association and K-means methods and is conducted locally at each decentralized GW to substantially improve the system throughput, extended from my previous work [56] to a multi-GW scenario.

- A distributed PAPC constrained BF via maximization of the worst-user signal-to-leakage-and-noise ratio (SLNR) is mathematically derived based on a semidefinite relaxation (SDR) to be locally processed in geographically decentralized GWs, aiming to optimize user fairness.

- As a performance limit, a centralized multilayer BF via the block SVD [32] is explored at a central unit with full CSI and data sharing. Unlike [32], I propose to employ SDR to optimize the worst-user signal-to-noise ratio (SNR) after interbeam interference suppression.

- This investigation of a flexible payload PA-optimization strategy by formulating a weighted sum-rate maximization (WSRM) optimization problem under practical SATCOM and successful SIC decoding constraints provides new insights. Through rigorous mathematical derivations, I further reformulate a nonconvex WSRM optimization to a weighted sum mean square error minimization (WMMSE) counterpart with a distributed optimization.

- Alternatively, a tailored deep neural network (DNN) architecture with a customized loss function is trained to intelligently allocate payload-power resources among both the beams and users, by learning the undercover structure of its input (i.e., unsupervised learning as a field of machine learning). Moreover, the DNN-based scheme can be trained as a universal approximator of the payload power resources-allocation agent for any unseen satellite channel and has the potential for a real-time operation without signaling and real-time information exchange between the GWs and the users.

## 1.5 Structure of this Dissertation

The rest of this dissertation is structured as follows:

**Chapter 2:** The multi-GW multibeam satellite system architecture and model are introduced. A practical multi-GW multibeam SATCOM scenario is regenerated via mathematical modeling, which is further applied to design and validate

distributed resource optimization strategies. Herein, the distributed resource control framework is presented.

**Chapter 3:** A distributed user-scheduling strategy is described. Then, the computational complexity and the spectrum efficiency performance of the user-scheduling strategy is analyzed.

**Chapter 4:** The optimization strategies for multi-GW BF are presented, where a distributed BF scheme is focused on. As an upper-bound performance benchmark, the centralized BF is also investigated. The computational complexity and the spectrum efficiency performance of the BF strategies is analyzed.

**Chapter 5:** The payload PA-optimization strategies based on mathematical optimization and machine learning are applied to flexibly allocate payload power resources among the beams and users, depending on the actual traffic demand. The computational complexity of the payload PA-optimization strategies is evaluated. Furthermore, the performance of the payload PA-optimization strategies is analyzed through simulation results.

**Chapter 6:** The conclusions and the future research are provided.

In summary, the relation among the main chapters 2, 3, 4, and 5 is depicted in Figure 1-4.

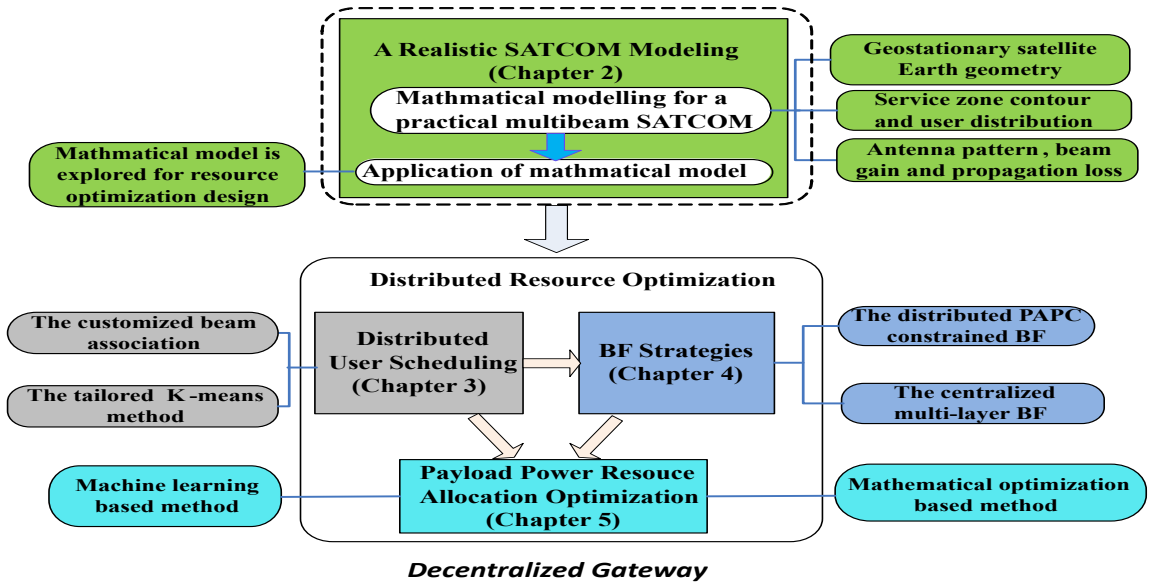


Figure 1-4: The relation among the main chapters.

## 1.6 Publications

All research results are either presented or accepted for publication as listed as follows.

### 1.6.1 Conference Publications

- Y. Zhu, T. Delamotte, A. Knopp, "Geographical NOMA Beamforming in Multi-Beam Satellite-Based Internet of Things," in *Proc. 2019 IEEE GLOBECOM*, Dec. 2019, pp. 2239-2251.
- Y. Zhu, C. A. Hofmann, A. Knopp, "Performance Optimization for Multi-Gateway NOMA-Beamforming in Multi-Beam SATCOM," in *Proc. 2021 IEEE ICC*, pp. 1-6, Jun. 2021.
- Y. Zhu, C. A. Hofmann, A. Knopp, "Machine Learning-based Flexible Payload Power Resource Allocation for Non-orthogonal SATCOM," accepted in *Proc. 2022 IEEE GLOBECOM*, Dec. 2022.

### 1.6.2 Journal Publication

- Y. Zhu, C. A. Hofmann and A. Knopp, "Distributed Resource Optimization for NOMA Transmission in Beamforming SATCOM," *IEEE Journal on Selected Areas in Communications*, vol. 40, no. 4, pp. 1190-1209, Apr. 2022.

### 1.6.3 Publication outside the dissertation

To remain consistent within this dissertation, the below work is not contained:

- Y. Zhu, C. A. Hofmann, A. Knopp, "Ka-Band LEO Satellite Internet of Things Channel Characterization: Survey and Measurement," in *Proc. 2022 IEEE ICC*, pp. 1-6, May 2022.

# Chapter 2

# Background of Forward Link Non-orthogonal SATCOM

*In this chapter, the incorporation and compatibility of the nonorthogonal transmission with the existing BF SATCOM system are depicted. First, the architecture and the mathematical modeling for a multi-GW SATCOM system are presented. Based on the system architecture, I present a distributed resource control framework, where resource optimization strategies for beamforming, flexible payload power resource allocation, and user scheduling are consecutively conducted at the GWs. Then, the system signal and channel model are described, respectively.*

## 2.1 System Architecture

In this section, the architecture of multi-GW SATCOM systems is detailed. First, an overview of a multi-GW SATCOM system is described. Then, the illustration of multi-GW SATCOM scenarios is provided.

### 2.1.1 Overview

As shown in Figure 2-1, this work considers the FWD link of a multibeam geostationary satellite (geo-satellite) system deployed with a transparent bent-pipe payload and served by geographically separated GWs. The system architecture incorporates nonorthogonal transmission into BF-based FWD links. The system conducts in Ka-band, which is composed of the GW segment, the space segment, and the user segment.

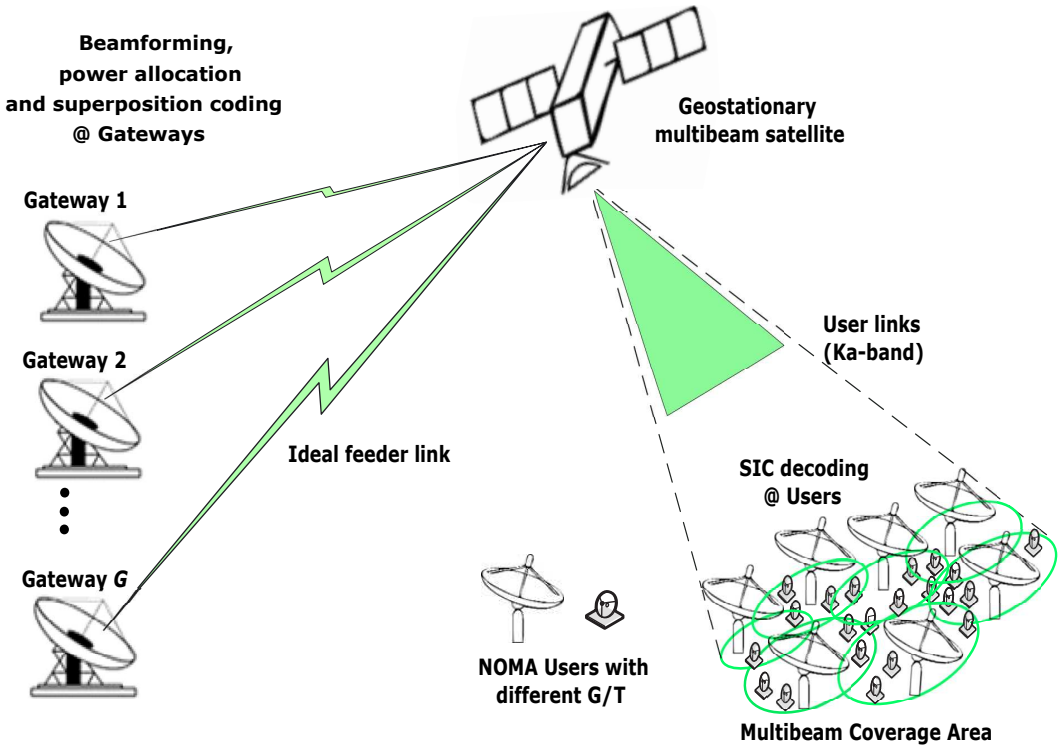


Figure 2-1: The multiple-GW satellite system architecture.

### 2.1.2 Gateway Segment

In general, an on-ground GW station has the function of connecting a satellite to other parts of the network. Considering the feeder link limitations, the large number of beams need to be supplied by multiple GWs in HTS systems. In the FWD link with the full frequency reuse factor, the signals are conveyed from the GWs to the corresponding users, suffering the increasing interbeam and inter-GW interference. Under such circumstances, on-ground BF is employed at each GW to manage the intra-system interference in the FWD link. Furthermore, I incorporate the non-orthogonal transmission into the BF-based FWD link. Accordingly, NOMA is applied in the BF FWD link with full-frequency reuse [56], [61], where multiple users are simultaneously served by each beam in each time slot. Within each beam, NOMA is explored by performing superposition coding and SIC at the GW and the user side, respectively. The additional superposition coding operation on multiple users' signals per beam is only performed at each GW for generating the superimposed signals. Sequentially, the BF strategy is applied to suppress the intrasystem interference. Following the BF strategy, an flexible payload power resource-allocation strategy can then be computed and processed at the GW segment, based on actual user traffic demand. After that, the processed data signals are transmitted to the satellite

through the ideal feeder links.

### 2.1.3 Space Segment

The space segment (i.e. satellite) comprises the spacecraft platform and the communication payloads. Typically, the spacecraft platform mainly consists of the on-broad control bus, power, thermal control, altitude control. On the other hand, an exemplary transparent (i.e. bent pipe) payload involves transceiver antennas and other components in the transponder (i.e. channel). More particularly, the receiver antenna with BF capacity first offers a certain antenna gain, which is adapted to discover the expected signal over the captured signals and noise. Next, the received RF signal is amplified at the lower-noise amplifier (LNA) to capture the effective noise temperature. Sequentially, the generated RF signal is transferred across a frequency down-converter, the outcome of which is a low-frequency baseband signal. Then, the baseband signal is amplified at the high-power amplifier (HPA), which is followed by the satellite transmit antenna. In the sequel, the crucial electronic equipments and associated technique concepts are further presented for a multibeam transparent transponder.

- *Multiple Beam Satellite Antenna:* in general, a multiple beam antenna payload generates several antenna beam patterns in the whole service coverage. In this case, the satellite antenna is a critical component for the formation of the spot beam pattern. There are two types of multibeam satellite antenna architectures: (1) single feed per beam antenna; (2) multiple feeds per beam antenna. In the configuration of the single feed per beam antenna architecture, a single beam is generated by exploiting a single feed horn. To capture the overlapping beams, either a single shaped reflector or a lens is additionally utilized [67]. In the configuration of the multiple feeds per beam antenna architecture, each spot beam is created by using a subarray feed horn system. Additionally, a complex beamforming network is required to reconfigure the patterns of spot beams. Thus, it is concluded that a specific beam can be shaped by either using a shaped reflector or a BF network. The former is a simpler method to attain the specific beam pattern, which is not flexibly varied on board. On the contrary, the latter is in the capacity of dynamically altering beam pattern and coverage, which is accomplished at the cost of the high complexity. Since the single feed per beam antenna is superior to the multiple feeds per beam antennas in terms of gain performance, existing large spacecraft (e.g. Eurostar E3000 [68]) is

equipped with the single feed per beam antenna. Therefore, this work considers a geo-satellite scenario with the single-feed-per-beam antenna configuration.

- *Multiple Beam Pattern:* By using the above-mentioned multiple beam satellite antenna, multiple spot beam patterns can be generated, instead of a global beam. The resulting footprint of each spot beam appears an oval on the Earth, featured with the small-scale size of a few kilometers. The spot beam pattern brings some benefits summarized as follows: (1) the whole bandwidth is improved by the number of spot beams since the bandwidth is invariable within each beam coverage; (2) spot beam pattern permits to simultaneously transmit distinct data streams in the same frequency resource (i.e. the same frequency is reused in different geographical regions); (3) each spot beam from the satellite offers an enhanced antenna gain, due to the reduced angular width of the corresponding radiation beam. Thus, the small compact antenna is allowed to be equipped for the on-ground user.

- *Flexible Power Amplifier:* The on-board flexible PA is implemented via adopting flexible payload power amplifiers such as multiport MPAs and flexible TWTAs. Using MPAs, the total available power on payload can be adjustably and flexibly assigned between different beams in the user link. Within the whole beam coverage area, antenna gains are dissimilar between corresponding beam service regions, since the view angles are different from a satellite perspective. Thus, the resulting diverse performance among different beams can be dynamically balanced by the means of the MPA techniques. The MPAs can be also reconfigured to satisfy the unique traffic demand of each served user. Alternatively, the flexible payload power resource allocation is achieved via employing flexible TWTAs. Herein, on the flexible payload, the available on-broad power is dynamically redistributed among multiple service beams to meet diverse traffic demands over time. Using a telecommand, the flexible TWTAs control bias conditions of payload amplifiers to tune the output power range of a traveling wave tube (TWT). More specifically, a diverse output power of TWT is created via adjusting the anode voltage to produce a distinct cathode current [69]. In this work, flexible payload power amplifiers (such as MPAs and flexible TWTAs) are applied to flexibly reconfigure the non-uniform user traffic rate/power within each beam coverage. Consequently, the multiple users' signals can be multiplexed in the power domain.

### 2.1.4 User Segment

The user segment covers the electronic devices for communicating with the satellite over the FWD link, where the fixed user devices are equipped with directional dish antennas. The antenna output is conveyed through a feed (or transmission) line. The output signal of the transmission line is then provided to a preamplifier (e.g., an LNA). Consecutively, the output signal of the preamplifier goes through the rest components including a digital downconverter and an intermediate frequency amplifier. Remarkably, due to a large free space loss (typically an order of 200 dB), a weak satellite signal is usually captured at the user side. In order to detect the desired signal, the average power of the noises created by these components (i.e., the preamplifier, the digital downconverter and the intermediate frequency amplifier) should be kept at a low threshold value compared to the amplified power of the desired signal. On the other hand, since this work considers a non-orthogonal multi-beam SATCOM scenario in Section 1.3.4, the users within each beam thus have different receiver G/T towards the satellite. As a result, there is a large SNR imbalance between the FWD links, which facilitates the successful application of NOMA in each beam.

## 2.2 Mathematical Modeling for SATCOM

This section describes the mathematical modeling for an exemplary multi-GW multi-beam SATCOM scenario. First, the overall architecture of the mathematic model is presented. Then, the crucial traits of the mathematical model are detailed.

### 2.2.1 Overall

A realistic and flexible multibeam SATCOM scenario is regenerated via mathematical modeling, which is composed of multi-beam satellite position, beam pattern, service zone contour, user distributions, the FWD link budget design and co-channel interference[1]. The mathematical modeling involves the link budget design, where beam patterns and transmission attenuation from free-space path loss and rainfall are accounted for. Moreover, these system parameters (e.g., link budgets and beam gain) are further related to the characterization of a FWD link channel. To accomplish these functions, the general architecture of the mathematical model is classified into three blocks: the input, the processing, and the output blocks. The input block produces geographical positions (i.e., latitudes and longitudes in geographical coordinate

[1] A fixed allocation of spot beams to on-ground GWs is considered in this work.

systems) for a GEO-satellite, 96 spot beam centers and preliminary dropped users beyond the whole beam coverage. These geographical positions are then adopted as the input of the processing block. Finally, the processing block produces the beam association and channel matrix to the output block. This mathematical model is developed with Matlab. The general schematic of the mathematical model is depicted in Figure 2-2. The key features of the mathematical model are described in the following.

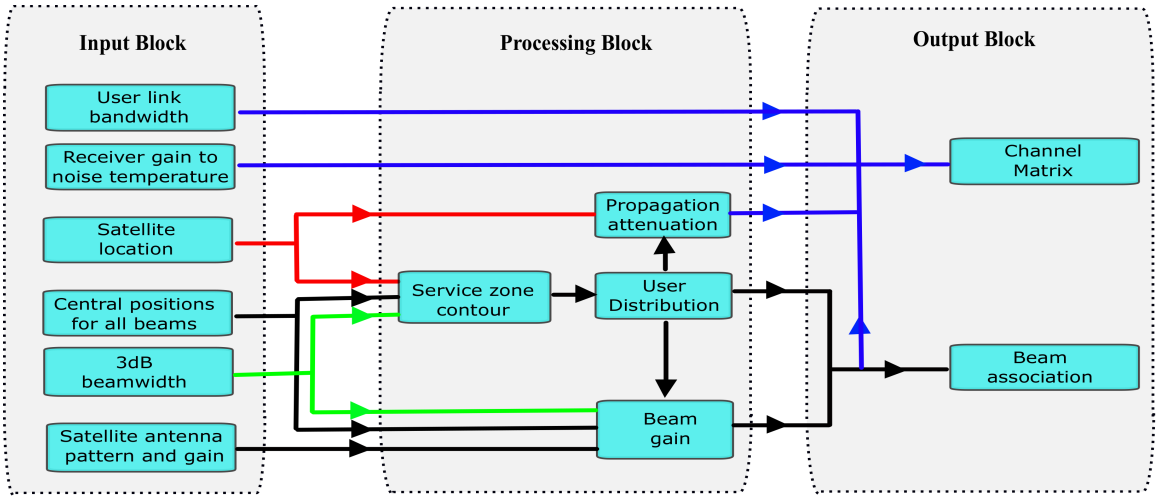


Figure 2-2: The general schematic of the mathematical model.

### 2.2.2 Geostationary Satellite-Earth Geometry

In this work, a practical Ka-band (20 GHz) transparent SATCOM system is regenerated, where a multi-beam geostationary satellite is located on the orbit at 19.2° east [70, 71]. The European region on the Earth's surface is covered by the footprints of 96 spot beams with 0.2 degrees, and satellite antenna gain is 53.23 dBW. Here, the overall satellite system architecture consists of multiple GWs, each of which supplies a subgroup of beams due to the feeder link limitations. Multiple users are simultaneously served by the same beam. Since the service zone contour of each spot beam (i.e., beam footprint) is determined from the perspective of true satellite antenna view angles, which are herein clarified in the satellite-centered coordinate system. The geometry for a geo-satellite is illustrated in Figure 2-3, where the points O and S denote the center of the Earth and the satellite, respectively. The point B is the beam point of beam $b$. The points A and $B$ are on round secant to the equatorial plane. The points C and D are the project of the points A and B on the equatorial plane, respectively. The angles $\vartheta_b$ and $\phi_b$ represent the true view angles of satellite antenna

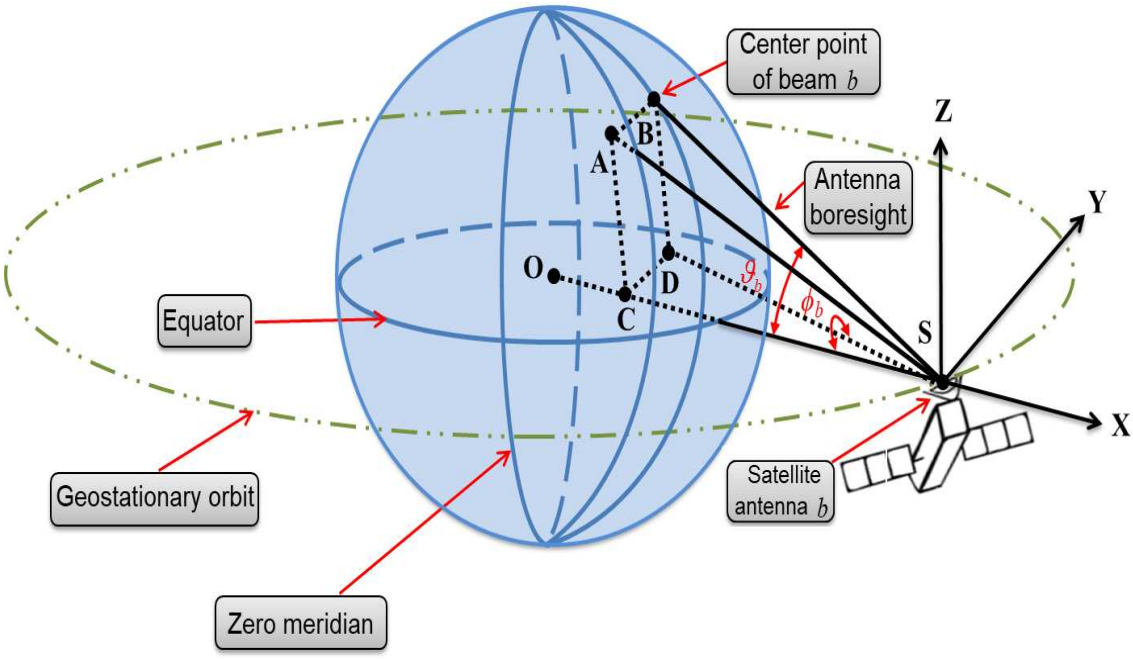


Figure 2-3: The geometry for a geo-satellite.

feed $b$, respectively. More specifically, $\vartheta_b$ is the nadir angle between the direction of the Earth center (i.e. SC) and the boresight direction of satellite antenna $b$ (i.e. SB); $\phi_b$ denotes an angle among the equatorial reference plane and the plane established by the antenna boresight (i.e. SB) and the line between the satellite and the Earth center (i.e. SC) [5]. In the next section, the geostationary satellite-Earth geometry with the satellite-centered coordinate system and true view angles are further used to determine the spot beam projection of a satellite antenna pattern on the Earth (i.e. service zone contour of a spot beam).

### 2.2.3 Service Zone Contour and User Distribution

This part explains the approaches for determining the service zone contour of each spot beam and generating on-ground users. Each beam footprint is related to the beam pattern model (e.g. the Bessel antenna pattern model [4] is assumed in Section 2.2.4), geographical positions of the satellite and each beam center. This work supposes the same geographical positions for the satellite and each spot beam center as those in [71]. To obtain the service zone contour of a spot beam, a satellite-centered coordinate system and the true satellite antenna view angles are generally adopted. For ease of presentation, the service zone contour determination procedure is explained in an exemplary scenario such as:

- The east satellite altitude above the equator $r_o$ is 35786 km.
- The radius of the Earth $r_e$ is 6378.137 km.
- Location vector of the satellite is denoted as a (latitude, longitude) pair ($\varphi_b = 0°$, $\lambda_b = 19.2°$) in geographical coordinate systems.
- Location vector $\mathbf{v}_b$ of the center point of the beam $b$ (i.e. beam center) radiated from the served feed $b$ is defined by a (latitude, longitude) pair $(\varphi_b, \lambda_b)$ in geographical coordinate systems.
- The 3dB beamwidth $\theta_{3\text{dB}}$ is 0.2°.

Specifically, the service zone contour of a spot beam $b$ is determined through five steps as follows:

*Step 1.* Knowing geographical position vector $\mathbf{v}_b$ of the beam center radiated from the served feed $b$, the corresponding satellite-centred coordinates of spot beam $b$ are given in **Definition 1**.

**Definition 1** A beam center $\mathbf{v}_b$ on the surface of the Earth has corresponding satellite-centered coordinates $\{\mathbf{X}_{\mathbf{v}_b}, \mathbf{Y}_{\mathbf{v}_b}, \mathbf{Z}_{\mathbf{v}_b}\}$ in a satellite-centered coordinate system [4], which are respectively calculated as

$$\begin{aligned} \mathbf{X}_{\mathbf{v}_b} &= r_e \cos(\varphi_b) \cos(\lambda_b - \lambda_s) - (r_e + r_o) \\ \mathbf{Y}_{\mathbf{v}_b} &= r_e \cos(\varphi_b) \sin(\lambda_b - \lambda_s) \\ \mathbf{Z}_{\mathbf{v}_b} &= r_e \sin(\varphi_b) \end{aligned} \tag{2.1}$$

In an example, the longitude of the satellite is $\lambda_s = 19.2$ degree.

*Step 2.* The coordinates of the beam center $\mathbf{v}_b$ are re-expressed from the perspective of satellite feed view angles; that is:

$$\begin{aligned} \mathbf{X}_{\mathbf{v}_b} &= -r_b^{'} \cos(\vartheta_b) \\ \mathbf{Y}_{\mathbf{v}_b} &= r_b^{'} \sin(\vartheta_b) \cos(\phi_b) \\ \mathbf{Z}_{\mathbf{v}_b} &= r_b^{'} \sin(\vartheta_b) \sin(\phi_b) \end{aligned} \tag{2.2}$$

where $r_b^{'}$ is the slant range (i.e. the distance between the satellite and the the spot beam center $b$ on the Earth). The angles $\vartheta_b$ and $\phi_b$ denote the true view angles of the satellite antenna feed $b$, which are defined in Section 2.2.2.

*Step 3.* From equation (2.1) and (2.2), true view angles of the satellite antenna feed $b$ are then determined as

$$\begin{aligned} \vartheta_b &= \arccos\left(\frac{-\mathbf{X}_{\mathbf{v}_b}}{r_b^{'}}\right) \\ \phi_b &= \arctan\left(\frac{\mathbf{Z}_{\mathbf{v}_b}}{\mathbf{Y}_{\mathbf{v}_b}}\right) \end{aligned} \tag{2.3}$$

with $r_b^{'} = \sqrt{\mathbf{X}_{\mathbf{v}_b}^2 + \mathbf{Y}_{\mathbf{v}_b}^2 + \mathbf{Z}_{\mathbf{v}_b}^2}$.

*Step 4.* Furthermore, an auxiliary circle pattern (i.e. the satellite antenna pattern) is assumed to be situated at the length $r_b^{'}$ from the satellite to the center of the Earth, which is depicted as Figure 2-4.

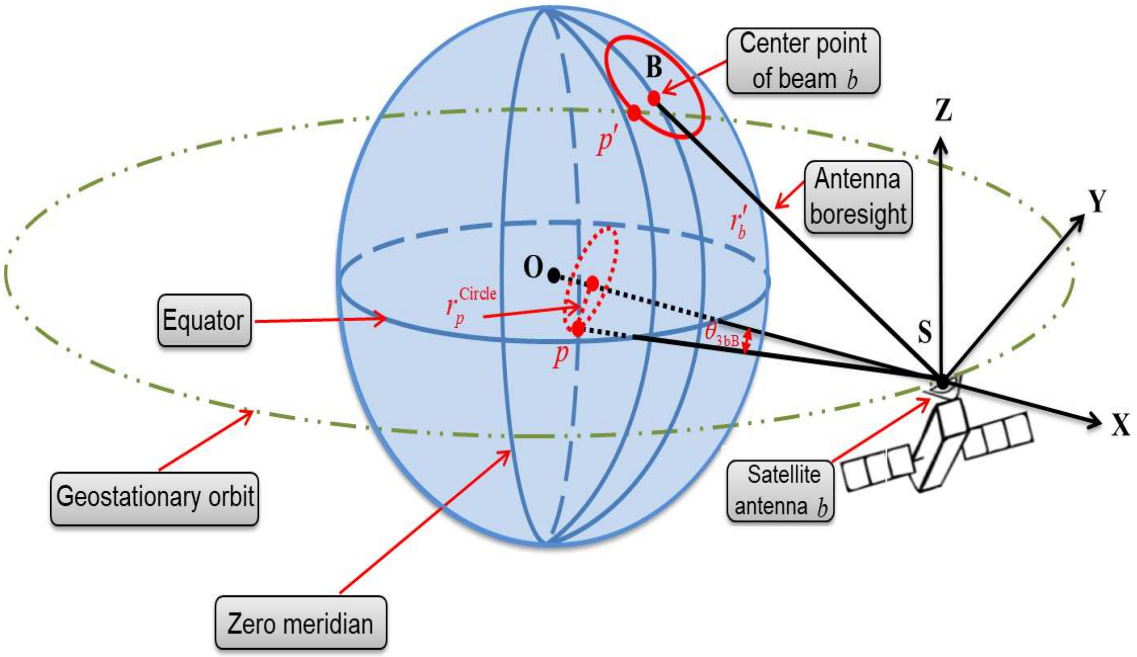


Figure 2-4: The auxiliary circle pattern with the GEO-satellite geometry.

Then, a reference point $p$ of the circle is defined in the satellite-centered coordinate system as

$$\begin{aligned} \mathbf{X}_p^{\text{Circle}} &= -r_b^{'} \\ \mathbf{Y}_p^{\text{Circle}} &= r_p^{\text{Circle}} \cos(\beta) \\ \mathbf{Z}_p^{\text{Circle}} &= r_p^{\text{Circle}} \sin(\beta) \end{aligned} \tag{2.4}$$

where the circle radius $r_p^{\text{Circle}} = r_b^{'} \tan(\theta_{3\text{dB}})$ is obtained based on the geometry in

Figure 2-4. The auxiliary parameter $\beta \in [0, 2\pi]$ is assumed.

*Step 5.* Within each spot beam coverage, the boresight of satellite antenna feed points to the center of spot beam coverage on the surface of the Earth, instead of the Earth's center. Next, the above satellite-centered coordinates $(\mathbf{X}_p^{\text{Circle}}, \mathbf{Y}_p^{\text{Circle}}, \mathbf{Z}_p^{\text{Circle}})$ for the auxiliary circle pattern is further rotated in terms of the true view angles (i.e. $\vartheta_b$ and $\phi_b$) of the satellite antenna feed $b$. Correspondingly, the circle center is rotated to the beam center $\mathbf{v}_b$. The resultant coordinate system is given as

$$\begin{pmatrix} \overline{\mathbf{X}}_p^{\text{Circle}} \\ \overline{\mathbf{Y}}_p^{\text{Circle}} \\ \overline{\mathbf{Z}}_p^{\text{Circle}} \end{pmatrix} = \mathbf{A}_{\text{rotation}} \mathbf{B}_{\text{rotation}} \begin{pmatrix} \mathbf{X}_p^{\text{Circle}} \\ \mathbf{Y}_p^{\text{Circle}} \\ \mathbf{Z}_p^{\text{Circle}} \end{pmatrix} \tag{2.5}$$

where the rotation of an angle $\vartheta_b$ around the $y-$axis is described as

$$\mathbf{A}_{\text{rotation}} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos(\vartheta_b) & \sin(\vartheta_b) \\ 0 & -\sin(\vartheta_b) & \cos(\vartheta_b) \end{pmatrix} \tag{2.6}$$

and the rotation of an angle $\phi_b$ around the $x-$axis is denoted as

$$\mathbf{B}_{\text{rotation}} = \begin{pmatrix} \cos(\phi_b) & 0 & -\sin(\phi_b) \\ 0 & 1 & 0 \\ \sin(\phi_b) & 0 & \cos(\phi_b) \end{pmatrix} \tag{2.7}$$

From equation (2.3) and (2.5), true view angles are then captured for a reference point $p$ of the circle after the rotation; that is:

$$\begin{aligned} \vartheta_p^{\text{Circle}} &= \arccos\left(\frac{-\overline{\mathbf{X}}_p^{\text{Circle}}}{r_p}\right) \\ \phi_p^{\text{Circle}} &= \arctan\left(\frac{\overline{\mathbf{Z}}_p^{\text{Circle}}}{\overline{\mathbf{Y}}_p^{\text{Circle}}}\right) \end{aligned} \tag{2.8}$$

where $r_p = \sqrt{(\overline{\mathbf{X}}_p^{\text{Circle}})^2 + (\overline{\mathbf{Y}}_p^{\text{Circle}})^2 + (\overline{\mathbf{Z}}_p^{\text{Circle}})^2}$. Remarkably, it can be found from the true view angles definition that $\vartheta_p^{\text{Circle}} = \vartheta_b$. Eventually, the rotated coordinate

system in (2.5) and the true view angles in (2.8) are employed to determine the beam projection of a satellite antenna $b$ on the Earth (i.e. the service zone contour of a spot beam $b$). Mathematically, the service zone contour of a spot beam $b$ is determined by computing the geographical positions for each reference point $p^{'}$ of the service zone contour; that is:

$$\begin{aligned} \vartheta_{p'} &= \arcsin\left(\frac{r_{p'}\sin(\vartheta_p^{\text{Circle}})\cos(\phi_p^{\text{Circle}})}{r_e}\right) \\ \phi_{p'} &= \arcsin\left(\frac{\tan(\vartheta_{p'})}{\tan(\phi_p^{\text{Circle}})}\right) + 19.2 \end{aligned} \tag{2.9}$$

where $r_{p'}$ is the minimum solution to the equation $r_{p'}^2 - 2(r_e + r_o)\cos(\vartheta_p^{\text{Circle}})r_{p'} + r_o^2 + 2r_e r_o = 0$.

After the service zone contour of each spot beam is ascertained, on-ground users are distributed on the whole beam coverage. As the subscribed users are randomly altered in a realistic SATCOM system, this work randomly generates the on-ground users in the whole beam coverage (i.e. the geographical positions of the dropped user are unknown during the simulation). The user generation function (or process) is described as

- *First*, defining the latitude and longitude coordinates for each randomly generated user as the input of the user generation function.
- *Second*, determining and only reserving each user geographical position lies inside or on the edge of multiple overlapped beam footprints.
- *Third*, collecting the reserved user geographical positions as the output of the user generation function.

After the users are irregularly spread over the entire beam coverage, each beam is ready to be associated with its corresponding users. It is critical to choose an appropriate beam association metric and to assure that the users are severed by the corresponding beams. In Chapter 3, each beam is to be associated with the on-ground users based on the maximum beam gain metric.

### 2.2.4 Antenna Pattern and Beam Gain

In the preceding section, the service zone contour and user distribution are confirmed. Consecutively, beam gain is computed for each on-ground user as a function of user

locations (i.e. user locations are herein expressed utilizing satellite feed view angles) and antenna pattern. Before the beam gain calculation, the antenna radiation pattern is first modeled in this part.

- *Satellite Antenna Pattern:* Assuming that a satellite is equipped with a conventional circular aperture reflector antenna, the Bessel antenna pattern model is generally adopted to characterize the radiation pattern of the satellite antenna [4]; that is:

$$G(u) = G^{\max}\left(\frac{J_1(u)}{2u} + 36\frac{J_3(u)}{(u)^3}\right)^2 \tag{2.10}$$

where $G^{\max}$ is the satellite feed gain at the center of the spot beam. $J_1$ and $J_3$ are the order one and three Bessel functions of the first kind. $u = 2.07123\frac{\sin(\theta)}{\sin(\theta_{3\text{dB}})}$ and $\theta$ is the angle from the satellite point of view over the feed boresight. Then, based on the antenna model (i.e. the equation (2.10)), the radiation pattern is simulated as the function of the satellite view angle. The following Figure 2-5 illustrates the resulting antenna pattern with $G^{\max} = 53.23$ dB.

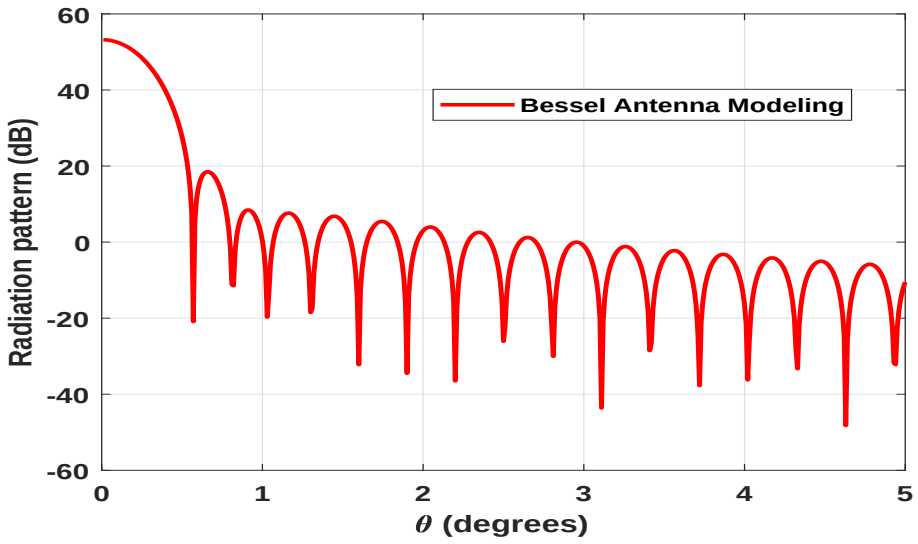


Figure 2-5: The exemplary satellite antenna pattern.

- *Satellite Beam Gain:* Subsequently, a procedure is presented to compute the bream gain from satellite feed $b$ to a user $\mathbf{u}_j$ on the surface of the Earth. Since **Definition 1** gives the geographical position vector $\mathbf{v}_b$ of the beam center ra-

diated from the served feed $b$, let us continue to denote the satellite-centered coordinates for a on-ground user $\mathbf{u}_j$ in **Definition 2**.

**Definition 2** Location $\mathbf{u}_j$ on the surface of the Earth has corresponding satellite-centered coordinates $\{\mathbf{X}_{\mathbf{u}_j}, \mathbf{Y}_{\mathbf{u}_j}, \mathbf{Z}_{\mathbf{u}_j}\}$ in a satellite-centered coordinate system [4], which are respectively calculated as

$$\begin{aligned} &\{\mathbf{X}_{\mathbf{u}_j}, \mathbf{Y}_{\mathbf{u}_j}, \mathbf{Z}_{\mathbf{u}_j}\} = \\ &\{r_e \cos(\varphi_j) \cos(\lambda_j - \lambda_s) - (r_e + r_o), r_e \cos(\varphi_j) \sin(\lambda_j \\ &-\lambda_s), r_e \sin(\varphi_j)\} \end{aligned} \tag{2.11}$$

Thus, the angle $\theta(\mathbf{u}_j, \mathbf{v}_b) \in [0, 2\pi]$ between location vectors $\mathbf{u}_j$ and $\mathbf{v}_b$, which indicates the off-axis angle of user $\mathbf{u}_j$ with respect to the boresight of the beam radiated by feed $b$, is given by

$$\theta(\mathbf{u}_j, \mathbf{v}_b) = \arctan\left(\frac{\|\mathbf{u}_j \times \mathbf{v}_b\|}{\mathbf{u}_j \cdot \mathbf{v}_b}\right) \tag{2.12}$$

where $\mathbf{u}_j \times \mathbf{v}_b$ and $\mathbf{u}_j \cdot \mathbf{v}_b$ are the cross and dot products of $\mathbf{u}_j$ and $\mathbf{v}_b$, respectively. Here, $\mathbf{u}_j \times \mathbf{v}_b$ is expressed as

$$\mathbf{u}_j \times \mathbf{v}_b = \sqrt{F^2 + \beth^2 + \gimel^2} \tag{2.13}$$

where

$$\begin{aligned} F &= \mathbf{Y}_{\mathbf{u}_j}\mathbf{Z}_{\mathbf{v}_b} - \mathbf{Z}_{\mathbf{u}_j}\mathbf{Y}_{\mathbf{v}_b} \\ \beth &= \mathbf{Z}_{\mathbf{u}_j}\mathbf{X}_{\mathbf{v}_b} - \mathbf{X}_{\mathbf{u}_j}\mathbf{Z}_{\mathbf{v}_b} \\ \gimel &= \mathbf{X}_{\mathbf{u}_j}\mathbf{Y}_{\mathbf{v}_b} - \mathbf{Y}_{\mathbf{u}_j}\mathbf{X}_{\mathbf{v}_b} \end{aligned}$$

Furthermore, $\mathbf{u}_j \cdot \mathbf{v}_b$ is denoted as

$$\mathbf{u}_j \cdot \mathbf{v}_b = \mathbf{X}_{\mathbf{u}_j}\mathbf{X}_{\mathbf{v}_b} + \mathbf{Y}_{\mathbf{u}_j}\mathbf{Y}_{\mathbf{v}_b} + \mathbf{Z}_{\mathbf{u}_j}\mathbf{Z}_{\mathbf{v}_b} \tag{2.15}$$

Let $G(\mathbf{u}_j, \mathbf{v}_b)$ be the beam gain from feed $b$ to user $\mathbf{u}_j$. Then, recalling the equation (2.10), beam gain $G(\mathbf{u}_j, \mathbf{v}_b)$ can be ultimately expressed as the function of

variable $\mathbf{u}_j$ and $\mathbf{v}_b$; that is:

$$G\left(\mathbf{u}_j, \mathbf{v}_b\right) = G^{\max}\left(\frac{J_1\left(u_b^{(j)}\right)}{2u_b^{(j)}} + 36\frac{J_3\left(u_b^{(j)}\right)}{\left(u_b^{(j)}\right)^3}\right)^2 \tag{2.16}$$

where $u_b^{(j)} = 2.07123\frac{\sin(\theta(\mathbf{u}_j,\mathbf{v}_b))}{\sin(\theta_{3\text{dB}})}$. Then, beam gain outcome is to be adopted in the below link budget design and a beam association operation in Chapter 3, where the beam association is involved in user scheduling mechanism.

### 2.2.5 Propagation Attenuation

In addition to the aforementioned beam gain, I proceed to describe propagation losses due to free-space path loss and atmospheric loss such as rainfall since these parameters are further employed to describe the satellite LOS channel in Section 2.3.4.

- *Free-space path loss:* The free space path loss of the signal is the highest attenuation in a SATCOM system. To be specific, the free space path loss can be generally denoted as the function of carrier frequency and slant range between the satellite and on-ground terminal. Mathematically, the free-space path loss (FSPL) is given in decibel format as

$$FSPL(d, f) = 20\log_{10}(d) + 20\log_{10}(f) + 92.45 \tag{2.17}$$

where $f$ is the signal frequency in unit of gigahertz (GHz) and $d$ is the slant range in unit of kilometer (Km). To facilitate the computation of slant range, an exemplary slant range for a on-ground terminal is depicted in Figure 2-6. Specifically, the slant range $d$ between the satellite and on-ground terminal is [72]

$$d = \sqrt{r_e^2\sin^2\theta + h_0^2 + 2h_0r_e} - r_e\sin\theta \tag{2.18}$$

where $h_0$ is the satellite altitude. $\theta$ is satellite elevation angle.

- *Atmospheric Loss:* As the most significant propagation phenomena, tropospheric impairments are commonly suffered in a slant path (e.g. a FWD link) operating in the Ka-band. In particular, there are distinct tropospheric impairments (e.g. atmospheric gaseous attenuation, clouds attenuation, and rain

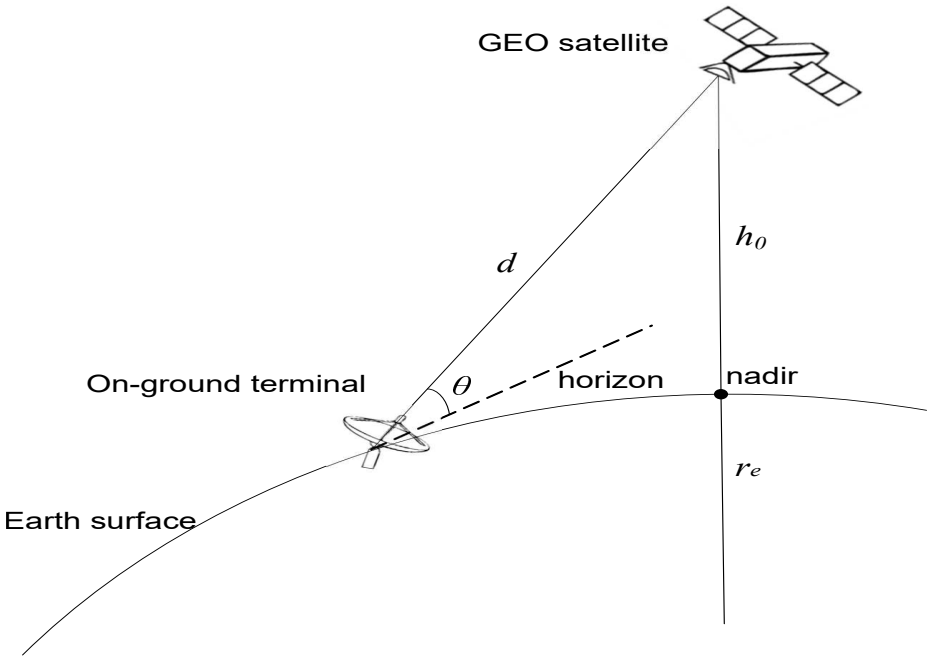


Figure 2-6: The exemplary slant range, which is edited from Figure 6.6.2-1 in [72].

attenuation) on the signal transmitted over the FWD links. Among them, rain attenuation has a dominant effect on Ka-band FWD links. Herein, the dominant atmospheric impairment that occurred in the rain event is then discussed. Regularly, the radio wave is impaired by the water vapor (the impairment is termed as rain attenuation), propagating through the cloud. Different kinds of rain attenuation models [73] exist including empirical-based model (e.g. Crane model [74]), statistical-based model (e.g. ITU-R rain attenuation model [75]), just to name a few. In general, rain attenuation is obtained from either propagation experimental data or the rain fade estimation approach in ITU-R rain attenuation model (i.e. recommendation ITU-R P.618) [75]. In this work, the outstanding ITU-R rain attenuation model [75]) is adopted to predict the rain attenuation for a FWD link operating in Ka-band. Following the ITU-R rain attenuation model [75], the required input parameters are illustrated as

- A given specific percentage of an average year (in the range from 0.001% to 5%);
- The height above mean sea level earth station (kilometers);
- The earth station's latitude and longitude (degrees);
- The carrier frequency (GHz);

– The radius of the Earth $r_e$;

– The elevation angle toward the satellite (degrees)

Then, the protocol of the ITU-R rain attenuation model [75] is used to estimate the long-term rain attenuation, which is summarized as

| **The Protocol of ITU-R Rain Attenuation Model [75].** |
|---|
| **Input:** The above-required parameters |
| ***Step 1.*** Estimating the rain height, according to recommendation ITU-R P.839 [76]; |
| ***Step 2.*** Calculating the slant path length; |
| ***Step 3.*** Computing horizontal projection of the slant path length; |
| ***Step 4.*** Capturing rain attenuation for a specific percentage of an average year, based on recommendation ITU R P.837 [77]; |
| ***Step 5.*** Acquiring specific attenuation, based on recommendation ITU R P.838 [78]; |
| ***Step 6.*** Counting the horizontal reduction factor; |
| ***Step 7.*** Reckoning the vertical adjustment factor; |
| ***Step 8.*** Obtaining the effective path length; |
| ***Step 9.*** Predicting rain attenuation for a specific percentage of an average year. |
| **Output:** The long-term rain attenuation |

In the next section, the system model under study is presented, wherein the application of the mathematical model is explored (i.e. this mathematical model is further adopted to devise resource optimization strategies for efficiently applying non-orthogonal in BF SATCOM.)

## 2.3 System Model

According to the system architecture in Section 2.1, the accommodation of non-orthogonal transmission in the existing BF SATCOM system is a foreseeable paradigm to boost the system throughput. Thus, facilitated by the mathematical model developed in Section 2.2, distributed resource control is investigated for efficiently incorporating non-orthogonal transmission in the conventional BF SATCOM system. In what follows, the system model is presented, where the incorporation and compatibility of the nonorthogonal transmission with existing BF SATCOM systems are delineated, and distributed resource-control strategies at the GWs are considered. As shown in Figure 2-1, this work considers the FWD link of a multibeam geostationary satellite (geo-satellite) system deployed with a transparent bent-pipe payload and served by geographically separated GWs. In the framework of this system architecture, each user is equipped with a single antenna, and the satellite has an array-fed reflector

antenna with $N$ feeds. The multibeam radiation pattern consisting of $K$ beams is formed by combining the corresponding feed signals.

I further embody a system with $G$ GWs and denote the total set of GWs as $\mathcal{G} = \{1, \cdots, G\}$. It is supposed that the feed elements at the satellite are consecutively allocated for each GW. Let $\mathcal{N}_g = \{N_g(g-1)+1, \cdots, N_g(g-1)+N_g\}$ be the subset of feeds, which are used by GW $g \in \mathcal{G}$, with $|\mathcal{N}_g| = N_g$ and $\sum_{g\in\mathcal{G}} N_g = N$. Thus, the set of beams (i.e., the beam cluster) served by GW $g$ is denoted by $\mathcal{N}_g$. There are, in total, $K = N$ beams generated in one feed per beam architecture. $\Upsilon_{b,g}$ is a group of users belonging to beam $b \in \mathcal{N}_g$ served by GW $g$. In practice, due to a large coverage area (e.g., $>$ 1000 kilometers wide area) of geo-satellite systems, the number of subscribed users is greater than the number of feeds located on the satellite payload. Thus, before actual data transmissions, a subset of active users is selected at each GW via a user-scheduling strategy. Consequently, $\mathcal{S}_{b,g} \subset \Upsilon_{b,g}$ ($|\mathcal{S}_{b,g}| = N_u$, $g \in \mathcal{G}$, $b \in \mathcal{N}_g$) is the group of selected users in beam $b \in \mathcal{N}_g$.

To boost the system throughput, the integration of nonorthogonal transmission in the existing BF SATCOM system is a foreseeable paradigm. Accordingly, NOMA is applied in the BF FWD link with full-frequency reuse [56], [61], where $N_u$ users are simultaneously served by each beam in each time slot. Within the beam $b$ served by GW $g$, NOMA is explored by performing superposition coding and SIC at the GW and the user side, respectively. The additional superposition coding operation on $N_u$ users' signals per beam is only performed at each GW for generating the superimposed signals, which are sequentially processed by the BF strategy. Specifically, compatible with the existing beam-specific BF SATCOM, distributed beam-specific BF signal processors at geographically separated GWs utilize the local CSI to suppress the increased intergateway and interbeam interference, resulting from the full-frequency reuse and multi-GW architecture. Hence, a beam-specific BF vector $\mathbf{w}_{b,g} \in \mathbb{C}^{N_g \times 1}$ is dedicated to all the users served by beam $b$ ($b \in \mathcal{N}_g$). After BF, an optimal payload power resource-allocation strategy can then be computed and processed at geographically separated GWs, based on the users' effective CSI. After that, the processed data signals are conveyed to the satellite via the ideal feeder links. Subsequently, the satellite payload executes a frequency conversion and transmits the processed signals to on-ground users after realizing the satellite payload power resource-allocation over the antenna array elements.

Next, for the on-ground reception segment, the user with poorer channel condition directly decodes its own signal. Meanwhile, the user with better channel condition performs the SIC decoding to detect its desired signal. Without a loss of generality,

this work assumes that $|\mathbf{h}_{b,g}^{(1)}\mathbf{w}_{b,g}| \geq |\mathbf{h}_{b,g}^{(2)}\mathbf{w}_{b,g}| \cdots \geq |\mathbf{h}_{b,g}^{(i)}\mathbf{w}_{b,g}| \cdots \geq |\mathbf{h}_{b,g}^{(|\mathcal{S}_{b,g}|)}\mathbf{w}_{b,g}|$, where $\mathbf{h}_{b,g}^{(i)}$ is the $1 \times N_g$ channel vector between GW $g$ and user $i$ in the corresponding beam $b \in \mathcal{N}_g$. Within the serving GW $g$, user $i$ in beam $b \in \mathcal{N}_g$ executes SIC to cancel the interference from user $j$ (for all $j > i$) in beam $b \in \mathcal{N}_g$ [56], [61]. Accordingly, the resulting signal received at user $i$ in beam $b \in \mathcal{N}_g$ is given as

$$
\begin{aligned}
y_{b,g}^{(i)} = & \underbrace{z_{b,g}^{(i)}}_{\text{noise}} + \underbrace{\mathbf{h}_{b,g}^{(i)}\mathbf{w}_{b,g}\sqrt{p_{b,g}^{(i)}}x_{b,g}^{(i)}}_{\text{desired-signal}} \\
& + \underbrace{\mathbf{h}_{b,g}^{(i)}\mathbf{w}_{b,g}\sum_{j=1}^{i-1}\sqrt{p_{b,g}^{(j)}}x_{b,g}^{(j)}}_{\text{intra-beam-interferences}} \\
& + \underbrace{\mathbf{h}_{b,g}^{(i)}\sum_{l\in\mathcal{N}_g\setminus\{b\}}\sum_{j\in\mathcal{S}_{l,g}}\mathbf{w}_{l,g}\sqrt{p_{l,g}^{(j)}}x_{l,g}^{(j)}}_{\text{intra-cluster-interferences}} \\
& + \underbrace{\mathbf{h}_{b,m}^{(i)}\sum_{m\in\mathcal{G}\setminus\{g\}}\sum_{l\in\mathcal{N}_m}\sum_{j\in\mathcal{S}_{l,m}}\mathbf{w}_{l,m}\sqrt{p_{l,m}^{(j)}}x_{l,m}^{(j)}}_{\text{inter-cluster-interferences}}
\end{aligned}
\tag{2.19}
$$

where $z_{b,g}^{(i)}$ denotes the zero-mean additive white Gaussian noise with unit variance. $\left\{x_{b,g}^{(i)}\right\}_{g\in\mathcal{G},\ b\in\mathcal{N}_g}^{i\in\mathcal{S}_{b,g}}$ are mutually uncorrelated transmit signals. The transmit power $p_{b,g}^{(i)}$ satisfies $\sum_{g\in G}\sum_{b\in N_g}\sum_{i\in S_{b,g}} p_{b,g}^{(i)} \leq P_{\text{total}}$. The entire transmit power budget is $P_{\text{total}}$. Thus, the rate is obtained at user $i$ for decoding its own signal as

$$
R_{b,g}^{(i)} = \log_2(1 + \text{SINR}_{b,g}^{(i)}) \tag{2.20}
$$

where the SINR at the user is given as

$$
\begin{aligned}
\text{SINR}_{b,g}^{(i)} = & |\mathbf{h}_{b,g}^{(i)}\mathbf{w}_{b,g}|^2 p_{b,g}^{(i)} \times \\
& (|\mathbf{h}_{b,g}^{(i)}\mathbf{w}_{b,g}|^2 \sum_{j=1}^{i-1} p_{b,g}^{(j)} + \sum_{l\in\mathcal{N}_g\setminus\{b\}} |\mathbf{h}_{b,g}^{(i)}\mathbf{w}_{l,g}|^2 \sum_{j\in\mathcal{S}_{l,g}} p_{l,g}^{(j)} \\
& + \sum_{m\in\mathcal{G}\setminus\{g\}}\sum_{l\in\mathcal{N}_m} |\mathbf{h}_{b,m}^{(i)}\mathbf{w}_{l,m}|^2 \sum_{j\in\mathcal{S}_{l,m}} p_{l,m}^{(j)} + 1)^{-1}
\end{aligned}
$$

### 2.3.1 Compatibility with existing BF SATCOM

The incorporated nonorthogonal transmission solution is fully compatible with the existing BF SATCOM because it does not require any complexity constraints or modifications on the existing multibeam satellite payload. Although it increases the signal-processing complexity of the GW and the strong users on the ground, it is not a foremost challenge, considering that the additional complexity occurs on the ground. Consequently, the conventional feeder-link infrastructures between GWs and satellite also can preserve its original capacity for the additional accommodation of the nonorthogonal transmission in the existing BF SATCOM systems. Furthermore, to incorporate nonorthogonal transmission solutions into the existing BF-based SATCOM systems, the practical single-shot transmission policy (i.e., user scheduling, BF, and power resource-allocation strategies are separately processed at the different times) is to be concentrated, due to stagnant scheduling demand and the uncertainty of the actual traffic intrinsic in the multibeam SATCOM [66].

### 2.3.2 Distributed Resource Control

The block diagram of the distributed resource-control strategies at geographically separated GWs is illustrated as Figure 2-7, where the user scheduling, BF, and power resource-allocation strategies are successively conducted, with an emphasis on studying the distributed implementations of multi-GW resource optimization. Remarkably, if a joint BF and PA optimization strategy is supposed for the nonorthogonal SATCOM system, it cannot integrate and comply with the existing BF SATCOM system (i.e., the fully on-ground BF transmission system with flexible onboard power resource allocation). This can be explained by the fact that the difference in the system model leads to two distinguishing optimization objective functions with two coupled optimization variables (i.e., the coupled BF matrix and PA coefficient vector) under different constraints. Therefore, this work explores the incorporation and compatibility of the nonorthogonal transmission with existing BF SATCOM systems, which can be leveraged by my proposal. That is, in an existing multibeam SATCOM system with inherently uncertain traffic demand, a fully flexible payload PA strategy allows us to preserve BF directions and reconfigure the available onboard power resources among both beams and users, depending on the actual traffic demand [24,66]. Accordingly, after the BF direction optimization, a fully flexible allocation of payload power resources is leveraged within my separate design.

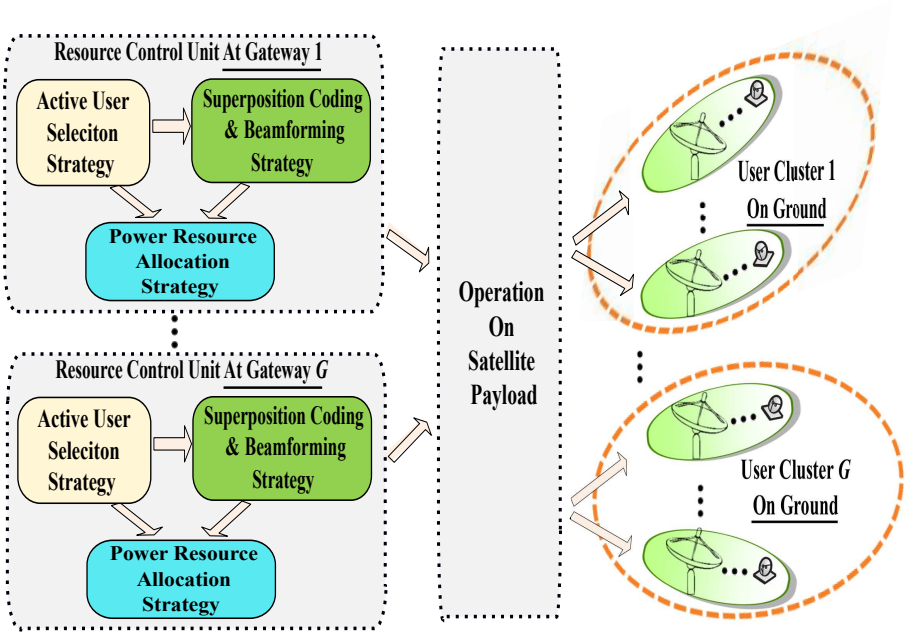


Figure 2-7: A block diagram for distributed resource control strategies.

### 2.3.3 Rate Conditions of Successful SIC Decoding

Furthermore, the signal of the weak user $i$, $\forall i$ can be also detected at the strong user $k$ ($k < i < |\mathcal{S}_{b,g}|$) within the same beam $b \in \mathcal{N}_g$ by performing successful SIC decoding. Consequently, the attainable rate at the strong user $k$ for decoding the signal of the weak user $i$ is captured as

$$R_{b,g}^{(i)\to(k)} = \log_2(1 + \mathrm{SINR}_{b,g}^{(i)\to(k)}) \tag{2.22}$$

where

$$\begin{aligned}\mathrm{SINR}_{b,g}^{(i)\to(k)} &= |\mathbf{h}_{b,g}^{(k)}\mathbf{w}_{b,g}|^2 p_{b,g}^{(i)} \times \\ &(|\mathbf{h}_{b,g}^{(k)}\mathbf{w}_{b,g}|^2 \sum_{j=1}^{i-1} p_{b,g}^{(j)} + \sum_{l\in\mathcal{N}_g\setminus\{b\}} |\mathbf{h}_{b,g}^{(k)}\mathbf{w}_{l,g}|^2 \sum_{j\in\mathcal{S}_{l,g}} p_{l,g}^{(j)} \\ &+ \sum_{m\in\mathcal{G}\setminus\{g\}} \sum_{l\in\mathcal{N}_m} |\mathbf{h}_{b,m}^{(k)}\mathbf{w}_{l,m}|^2 \sum_{j\in\mathcal{S}_{l,m}} p_{l,m}^{(j)} + 1)^{-1}\end{aligned}$$

To guarantee that the strong user $k$ ($k < i < |\mathcal{S}_{b,g}|$) successfully conducts SIC to decode the signal of the weak user $i$, $\forall i$ within the same beam $b \in \mathcal{N}_g$, the attainable rate $R_{b,g}^{(i)\to(k)}$ at the strong user $k$ for decoding the signal of the weak user $i$ should be not less than the received rate $R_{b,g}^{(i)}$ at the weak user $i$ for decoding its own signal,

according to the Shannon theorem [79]. Thus, following [40], the signal of the weak user $i,\ \forall i$ can be detected at each strong user $k$ $(k < i < |\mathcal{S}_{b,g}|)$ within the same beam $b \in \mathcal{N}_g$ by performing successful SIC decoding, when rate constraints are satisfied as

$$\begin{aligned} &R_{b,g}^{(i)\to(1)} \geq R_{b,g}^{(i)}, \\ &R_{b,g}^{(i)\to(2)} \geq R_{b,g}^{(i)}, \\ &\quad\vdots \\ &\underbrace{R_{b,g}^{(i)\to(i-1)} \geq R_{b,g}^{(i)}} \end{aligned} \tag{2.24}$$

$$\Downarrow$$

$$\begin{aligned} \mathrm{SIC}_{b,g}^{(i)\to(k)} := &|\mathbf{h}_{b,g}^{(k)}\mathbf{w}_{b,g}|^2 \times \Big(\sum_{l\in\mathcal{N}_g\setminus\{b\}} |\mathbf{h}_{b,g}^{(i)}\mathbf{w}_{l,g}|^2 \sum_{j\in\mathcal{S}_{l,g}} p_{l,g}^{(j)} \\ &+ \sum_{m\in\mathcal{G}\setminus\{g\}} \sum_{l\in\mathcal{N}_m} |\mathbf{h}_{b,m}^{(i)}\mathbf{w}_{l,m}|^2 \sum_{j\in\mathcal{S}_{l,m}} p_{l,m}^{(j)} + 1\Big) \\ &- |\mathbf{h}_{b,g}^{(i)}\mathbf{w}_{b,g}|^2 \times \Big(\sum_{l\in\mathcal{N}_g\setminus\{b\}} |\mathbf{h}_{b,g}^{(k)}\mathbf{w}_{l,g}|^2 \sum_{j\in\mathcal{S}_{l,g}} p_{l,g}^{(j)} \\ &+ \sum_{m\in\mathcal{G}\setminus\{g\}} \sum_{l\in\mathcal{N}_m} |\mathbf{h}_{b,m}^{(k)}\mathbf{w}_{l,m}|^2 \sum_{j\in\mathcal{S}_{l,m}} p_{l,m}^{(j)} + 1\Big) \geq 0, 1 \leq k < i,\ \forall g, b, i \end{aligned} \tag{2.25}$$

where the symbols $\mathrm{SIC}_{b,g}^{(i)\to(k)}$ and $:=$ denote a variable and a definition, respectively. The rate constraints (2.25) for ensuring successful SIC decoding will be fulfilled in the power resource-allocation strategy of Chapter 5.

### 2.3.4 Multibeam Satellite Channel Model

The channel coefficient from the satellite antenna $n \in \{1, \cdots, N\}$ to the $i$-th user in beam $b \in \{1, \cdots, N\}$ is [32]

$$h_{b,n}^{(i)} = \left(\frac{c_0}{4\pi f d_b^{(i)}}\right) \sqrt{\frac{(G_R)_b^{(i)} G_{b,n}^{(i)}}{\kappa B T_b^{(i)} A_b^{(i)}}} e^{-j\frac{2\pi f d_b^{(i)}}{c_0}} \tag{2.26}$$

where $f$, $c_0$, $\kappa$, $B$ and $T_b^{(i)}$ are the downlink operating frequency, speed of light, Boltzmann constant, carrier bandwidth and receiver noise temperature, respectively.

$\kappa B T_b^{(i)}$, $\frac{c_0}{4\pi f d_b^{(i)}}$ and $(G_R)_b^{(i)}$ are the noise power at the user, the free space loss, and the receiver antenna gain, respectively. The distance $d_b^{(i)}$ is measured between the satellite and user $i$ in beam $b$. Rain attenuation is denoted by $A_b^{(i)}$ for user $i$ in beam $b$. I express the power gain distribution (i.e., $(A_b^{(i)})_{\text{dB}} = 20\log_{10}(A_b^{(i)})$) as $\ln((A_b^{(i)})_{\text{dB}}) \sim \mathcal{N}(\mu, \sigma)$, which is related to polarization, the carrier frequency, the elevation angle, and the user location [28]. $G_{b,n}^{(i)}$ models radiation pattern from the $n$-th antenna toward user $i$ in beam $b$ as

$$G_{b,n}^{(i)} = G^{\max}\left(\frac{J_1\left(u_{b,n}^{(i)}\right)}{2u_{b,n}^{(i)}} + 36\frac{J_3\left(u_{b,n}^{(i)}\right)}{\left(u_{b,n}^{(i)}\right)^3}\right)^2 \tag{2.27}$$

where $u_{b,n}^{(i)} = 2.07123\frac{\sin\left(\theta_{b,n}^{(i)}\right)}{\sin(\theta_{3\text{dB}})}$, where the variable $\theta_{b,n}^{(i)}$ indicates the off-axis angle of user $i$ in beam $b$, with respect to the axis of maximum power radiated by antenna feed $n$. $\theta_{3\text{dB}}$ is the 3dB angle. Then, the aggregate channel vector between GW $g \in \mathcal{G}$ (i.e. the feeds $\{n | n \in \mathcal{N}_g = \{(g-1)+1, \cdots, (g-1)+N_g\}\}$) to user $i$ in the corresponding beam $b \in \mathcal{N}_g$ can be expressed as

$$\mathbf{h}_{b,g}^{(i)} = (h_{b,(g-1)+1}^{(i)}, \cdots, h_{b,(g-1)+N_g}^{(i)}) \in \mathbb{C}^{1\times N_g} \tag{2.28}$$

It is assumed that each GW perfectly knows the local CSI among itself and users in all clusters. Particularly, in the multi-GW multibeam SATCOM system, the local CSI acquisition at each GW involves the instantaneous CSI exchange among the GWs [14, 80]. Additionally, I assume a nonorthogonal multibeam SATCOM system, where the users within each beam have different sizes of the antennae or different noise factors of the power amplifiers. Thus, as discussed in Sections 1.3.4 and 2.1.4, there is a large SNR imbalance between the FWD links, which is required for a successful application of NOMA.

# Chapter 3

# User Scheduling Mechanism

*In general, practical multibeam SATCOM systems cover vast geographical regions to supply access to a great number of subscribed users. Therefore, an appropriate user-scheduling strategy typically is required to provide services efficiently with an existing SATCOM system [30]. Furthermore, I previously found that user scheduling influences the throughput performance of single GW nonorthogonal SATCOM systems [56]. Thus, the aim of this chapter is to further capture the attainable throughput performance gain by employing a suitable user-scheduling strategy in the considered multi-GW SATCOM scenario. Due to the significant system dimensions and the high computational complexity, a heuristic user scheduling strategy is adopted as follows.*

## 3.1 Overview of User Scheduling Procedure

For NOMA in BF SATCOM, a given user schedule is regularly considered. Though low complexity scheduling algorithms are studied for NOMA SATCOM in the work of Caus *et al.* [53], no (random) beam association is assumed in user scheduling schemes. As a result, the high user channel correlation is not fully explored. To address it, a user scheduling is thus proposed to fully explore the high user channel correlation within each beam, enabling the effective BF-based interference mitigation and the successful NOMA in each beam.

To be specific, each user from the initial user pool is first associated with a specific beam based on the highest beam gain metric, a function of the geographical positions of the users, beam centers, and satellite (it is termed as beam association). Consequently, the beam association gathers the high user channel correlation within each beam, which is inherent in a multibeam geo-satellite system. Based on the output of beam association, the users per beam can be scheduled with high channel correlation.

Next, a NOMA user is randomly selected per beam. Then, NOMA user candidates are preliminarily searched per beam, where users (whose channel correlation with the first NOMA user is larger than a predefined threshold $\delta^0$) are considered as NOMA user candidates per beam. Sequentially, the K-means algorithm [81] is tailored to divide the NOMA user candidates into $N_u$ groups per beam, so that every user within a group has similar channel gain and the underlying channel gain patterns among the groups are discovered. In this way, the underlying difference of the users' channel gains is exploited by the K-means algorithm to unleash the full potential of NOMA.

Finally, NOMA users per beam are successively scheduled from each group. In summary, the schematic diagram of the user scheduling procedure is composed of four steps, as represented in Figure 3-1.

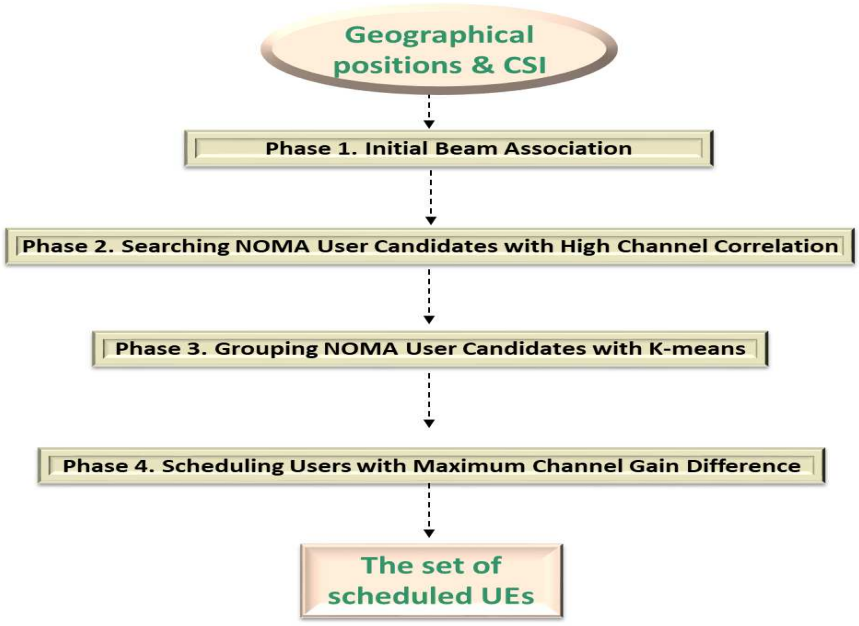


Figure 3-1: The schematic diagram of the user scheduling procedure.

## 3.2 Initial Beam Association

In the first phase, the users within the initial user pool are associated with appropriate satellite beams (i.e., beam association) in Ka-band multibeam satellite systems, based on the highest beam gain metric. To this end, I regenerate a practical multibeam geostationary satellite system [70, 71], where the European region on the Earth's surface is covered by the antenna footprints of 96 spot beams with $\theta_{3\text{dB}} = 0.2$ degree, and a satellite antenna gain is of 53.23 dBW. Before presenting the beam association,

I first state the following useful interpretation.

Within the geographical service area of GW $g \in \mathcal{G}$, let $\mathbf{u}_j$ and $\mathbf{v}_b$ denote the location vectors of the served user $j$ ($j = 1, \cdots U_{Total}$) and the center point of the beam (i.e., beam center) radiated from the served feed $b$ ($b \in \mathcal{N}_g$), respectively, correspondingly defined by (latitude, longitude) pairs $(\varphi_j, \lambda_j)$ and $(\varphi_b, \lambda_b)$ in geographical coordinate systems. In a slight abuse of notation, let feed index $b$ also denote the corresponding beam index, due to the consideration of a single-feed-per-beam case.

Then, let $G(\mathbf{u}_j, \mathbf{v}_b)$ be the beam gain from feed $b$ to user $\mathbf{u}_j$. Recalling the equation (2.16) of Chapter 2, beam gain $G(\mathbf{u}_j, \mathbf{v}_b)$ can be ultimately expressed as the function of variable $\mathbf{u}_j$ and $\mathbf{v}_b$ before beam association; that is:

$$G(\mathbf{u}_j, \mathbf{v}_b) = G^{\max} \left( \frac{J_1\left(u_b^{(j)}\right)}{2u_b^{(j)}} + 36\frac{J_3\left(u_b^{(j)}\right)}{\left(u_b^{(j)}\right)^3} \right)^2 \tag{3.1}$$

where $u_b^{(j)} = 2.07123\frac{\sin(\theta(\mathbf{u}_j, \mathbf{v}_b))}{\sin(\theta_{3\text{dB}})}$. Note that (2.27) is defined based on a given beam association outcome. Next, the beam association with the highest beam gain is introduced.

***The highest beam gain metric:*** Given a set of users $\{\mathbf{u}_j\}_{j=1}^{U_{\text{Total}}}$ as initial user pool on ground, and a tuple of beam centers $\{\mathbf{v}_b\}_{b\in\mathcal{N}_g}$ radiated from the corresponding satellite feeds, which are served by GW $g \in \mathcal{G}$. Then, within the service area of GW $g \in \mathcal{G}$, each corresponding beam center $\mathbf{v}_b$ ($b \in \mathcal{N}_g$) is a geographical spot on the Earth surface, and its corresponding beam association $\Upsilon_{b,g}$ consists of every user whose beam gain with respect to $\mathbf{v}_b$ ($b \in \mathcal{N}_g$) is greater than or equal to its beam gain with respect to any other $\mathbf{v}_l$ ($l \in \mathcal{N}_g, l \neq b$). In mathematics, let $X$ be a metric space with beam gain $G(\cdot)$ given in (3.1). Let $\{\mathbf{u}_j\}_{j=1}^{U_{\text{Total}}}$ and $\{\mathbf{v}_b\}_{b\in\mathcal{N}_g}$ be a set of users and a tuple of beam centers in the space $X$, respectively. Then, for beam center $\mathbf{v}_b$ ($b \in \mathcal{N}_g$), its beam association $\Upsilon_{b,g}$ consists of all users in $X$ whose beam gain with respect to $\mathbf{v}_b$ ($b \in \mathcal{N}_g$) is not less than their beam gain with respect to any other beam center $\mathbf{v}_l$ ($l \in \mathcal{N}_g, l \neq b$) (i.e., the highest beam gain metric, expressed as follows):

$$\Upsilon_{b,g} = \left\{ \mathbf{u}_j \in X \middle| G(\mathbf{u}_j, \mathbf{v}_b) \geq G(\mathbf{u}_j, \mathbf{v}_l), \begin{array}{c} b, l \in \mathcal{N}_g, l \neq b \\ j = 1, \cdots U_{\text{Total}} \end{array} \right\} \tag{3.2}$$

It is worth noting that the above notation is used to conveniently clarify the proposed

beam association procedure. After beam association, I let $\mathbf{u}_{b,g}^{(i)}$ denote user $i$ ($i = 1, \cdots |\Upsilon_{b,g}|$) associated with beam $b \in \mathcal{N}_g$ served by GW $g \in \mathcal{G}$. Thus, the set of associated users with beam $b \in \mathcal{N}_g$ is explicitly expressed by $\Upsilon_{b,g} = \left\{\mathbf{u}_{b,g}^{(i)}\right\}$. The beam association outcome is adopted in the next phase. Regarding the implementation of beam association, I consider that the beam pattern gain characteristics of each feed antenna is a function of the feed configuration and operating frequency. Additionally, the location of the antenna-gain contour on the surface of the Earth is a function of the satellite geographical location and antenna orientation. When the parameters are captured by the on-ground GWs, the geographical distance or antenna gain are calculated with the Earth surface. Thus, the on-ground GWs successfully allocate users to a certain beam.

## 3.3 Searching NOMA User Candidates with High Channel Correlation

After the beam association, each GW $g \in \mathcal{G}$ first selects an arbitrary NOMA user $\mathbf{u}_{b,g}^{(i^*)} \in \mathcal{S}_{b,g}$ from $\Upsilon_{b,g}$ within the second phase. Next, I proceed to seeking users from $\Upsilon_{b,g} \backslash \mathbf{u}_{b,g}^{(i^*)}$, whose channel correlations with the first NOMA user are larger than a threshold $\delta^0$ as NOMA user candidates for beam $b$ ($b \in \mathcal{N}_g$). Mathematically, NOMA user candidates $\mathcal{M}_{b,g}$ for beam $b$ ($b \in \mathcal{N}_g$) are found as follows:

$$\begin{aligned} \mathcal{M}_{b,g} &= \left\{\mathbf{u}_{b,g}^{(j^*)}\right\} \\ &= \underset{\mathbf{u}_{b,g}^{(j)} \in \{\Upsilon_{b,g} \backslash \mathbf{u}_{b,g}^{(i^*)}\}}{\text{find}} \left( \frac{\left| \mathbf{h}_{b,g}^{(i^*)} \left(\mathbf{h}_{b,g}^{(j)}\right)^H \right|}{\left\| \mathbf{h}_{b,g}^{(i^*)} \right\|_2 \left\| \mathbf{h}_{b,g}^{(j)} \right\|_2} \geq \delta^0 \right) \end{aligned} \tag{3.3}$$

where $\Upsilon_{b,g} \backslash \mathbf{u}_{b,g}^{(i^*)}$ is the updated beam association excepting the first NOMA user $\mathbf{u}_{b,g}^{(i^*)}$. A threshold $\delta^0$ is used to measure the channel correlations of the users per beam.

## 3.4 Grouping NOMA Users with K-means

In the third phase, the K-means algorithm [81] is further employed to divide the channel gains of a user set $\mathbf{u}_{b,g}^{(i^*)} \cup \mathcal{M}_{b,g}$ into $N_u$ groups per beam. In other words, given the channel gains of a user set $\mathbf{u}_{b,g}^{(i^*)} \cup \mathcal{M}_{b,g}$, the objective of the K-means

algorithm is to group users with similar channel gains (i.e., data points) together and to discover the underlying channel-gain patterns among the groups of each beam.

Thus, the number $N_u$ of user groups $\mathcal{C}_{b,g}^{(j)}$ $(j = 1, \cdots N_u)$ can be obtained for beam $b$ $(b \in \mathcal{N}_g)$ as the ultimate output of K-means algorithm, which is presented as Algorithm 1.

**Algorithm 1.**

**Input:**

The number of user groups $N_u$;

The channel gains of user set $\mathbf{u}_{b,g}^{(i^*)} \cup \mathcal{M}_{b,g}$ for beam $b$ $(b \in \mathcal{N}_g)$

**Output:**

User groups $\left\{\mathcal{C}_{b,g}^{(j)}\right\}_{j=1}^{N_u}$ for beam $b$ $(b \in \mathcal{N}_g)$

**Begin**

*Step 1:* Choose a group number $N_u$ and obtain the channel gains of user set $\mathbf{u}_{b,g}^{(i^*)} \cup \mathcal{M}_{b,g}$ for beam $b$ $(b \in \mathcal{N}_g)$

*Step 2:* Randomly place the centroid $c_{b,g}^{(j)}$ of data points (i.e., channel gains of the users) in user group $\mathcal{C}_{b,g}^{(j)}$ $(j = 1, \cdots N_u)$ for beam $b$ $(b \in \mathcal{N}_g)$

*Step 3:* For each data point $\left|\mathbf{h}_{b,g}^{(i)}\right|$ (i.e., channel gain of each user $\mathbf{u}_{b,g}^{(i)} \in \mathbf{u}_{b,g}^{(i^*)} \cup \mathcal{M}_{b,g}$) in beam $b$ $(b \in \mathcal{N}_g)$:

**Find** the nearest centroid based on one-dimension Euclidean distance metric; that is:

$$j^\circ = \arg\min_{j=1,\cdots N_u} \left|\left|\mathbf{h}_{b,g}^{(i)}\right| - c_{b,g}^{(j)}\right|,$$

for the user $\mathbf{u}_{b,g}^{(i)} \in \mathbf{u}_{b,g}^{(i^*)} \cup \mathcal{M}_{b,g}, b \in \mathcal{N}_g$

**Correspondingly**, assign user $\mathbf{u}_{b,g}^{(i)}$ to user group $j^\circ$ as $\mathbf{u}_{b,g}^{(i)} \in \mathcal{C}_{b,g}^{(j^\circ)}$.

*Step 4:* For each user group $j = 1, \cdots N_u$:

**New** centroid is the mean of data points (i.e. channel gains of the users) assigned to that user group; that is:

$$c_{b,g}^{(j)} = \frac{1}{\left|\mathcal{C}_{b,g}^{(j)}\right|} \sum_{\mathbf{u}_{b,g}^{(i)} \in \mathcal{C}_{b,g}^{(j)}} \left|\mathbf{h}_{b,g}^{(i)}\right|$$

*Step 5:* Repeat Steps 3 and 4 until the end of a fixed number of iterations

**End**

## 3.5 Scheduling Users with Maximum Channel Gain Difference

In the last phase, each GW $g \in \mathcal{G}$ successively schedules users that belong to different user groups to maximize corresponding channel-gain differences from the first NOMA user $\mathbf{u}_{b,g}^{(i^*)} \in \mathcal{S}_{b,g}$ for each served beam. Mathematically, the $j^\dagger$-th NOMA user $\mathbf{u}_{b,g}^{(j^\dagger)} \in \mathcal{S}_{b,g}$ in beam $b$ $(b \in \mathcal{N}_g)$ is consecutively selected from the corresponding user group

as

$$\mathbf{u}_{b,g}^{(j^\dagger)} = \underset{\substack{\mathbf{u}_{b,g}^{(j)} \in \mathcal{C}_{b,g}^{(j)}, \\ j=1\cdots i-1,i+1,\cdots N_u}}{\arg\max} \left|\left|\mathbf{h}_{b,g}^{(j)}\right| - \left|\mathbf{h}_{b,g}^{(i^*)}\right|\right|, \text{for } b \ (b \in \mathcal{N}_g) \tag{3.4}$$

From simulation results described in Section 3.7, the substantial performance gain can be captured by exploiting this scheduling process.

## 3.6 Complexity Analysis

I analyze the computational complexity of the user-scheduling strategy at GW $g \in \mathcal{G}$ as follows:

- *In phase 1:* Initial beam association involves the complexity of $\mathcal{O}(N_g U_{\text{Total}}^2)$.
- *In phase 2:* Finding $|\mathcal{M}_{b,g}|$ NOMA user candidates from the $|\Upsilon_{b,g}| - 1$ users in beam $b$ ($b \in \mathcal{N}_g$) has the complexity of $\mathcal{O}((|\Upsilon_{b,g}| - 1)^2)$. Therefore, the corresponding complexity for the beams served by GW $g \in \mathcal{G}$ is computed as $\mathcal{O}(\sum_{b\in\mathcal{N}_g}(|\Upsilon_{b,g}| - 1)^2)$.
- *In phase 3:* Let $I_{\text{Kmeans}}$ be the amount of iterations required by K-means algorithm, the complexity of the K-means algorithm is $\mathcal{O}(\sum_{b\in\mathcal{N}_g}(I_{\text{Kmeans}} N_u(|\mathcal{M}_{b,g}|+1)))$ [82]. As the K-means algorithm is conducted to divide the users per beam into $N_u = 2$ or 3 user groups based on the one-dimension Euclidean distance metric, the computational complexity of the K-means algorithm is sufficiently low.
- *In phase 4:* Consecutively selecting the maximum value from the corresponding user group holds the complexity of $\mathcal{O}(\sum_{b\in\mathcal{N}_g}\sum_{j\in\{[1,N_u]\setminus i\}}|\mathcal{C}_{b,g}^{(j)}|^2)$.

Finally, the entire complexity of the user-scheduling strategy at GW $g \in \mathcal{G}$ is $\mathcal{O}(N_g U_{\text{Total}}^2)$, since the complexity of phase 1 is dominant compared to that of other phases.

## 3.7 Performance Analysis

To validate the performance of my proposal, a representative multi-GW multibeam SATCOM scenario is simulated based on mathematic modeling with Matlab. Figure 3-2 depicts the European region on the Earth's surface covered by the antenna

footprints of 96, 0.2 degree spot beams, all of which originate from a single geostationary satellite located at 19.2 degrees East [70, 71]. The elliptic curves denote the 3 dB contours intersecting the Earth's surface. Without a loss of generality, this work considers a subgroup $N = 21$ from all 96 beams to simulate an interference-limited SATCOM scenario, since the intrasystem interference predominantly comes from the neighboring beams and beam clusters [12, 37]. The system has $G = 3$ GWs (i.e., the beam-coverage areas served by the three GWs are highlighted with the colors red, black, and pink, respectively), where every GW manages a set of $N_g = 7, g \in \{1, 2, 3\}$ beams. In Figure 3-2, an exemplary map of the carrier-to-interference ratio (C/I)[1] is produced, where the C/I at user $i$ in beam $b$ radiated from the served feed $b$ is expressed as

$$(\mathrm{C/I})_{b,b}^{(i)} = \frac{G_{b,b}^{(i)}}{\sum\limits_{b' \in [1,N] \backslash b} G_{b,b'}^{(i)}} \tag{3.5}$$

It is shown in Figure 3-2 that the C/I value reduces from the beam center to the beam edge (i.e., the intrasystem interference mainly originates from the neighboring beams and beam clusters).

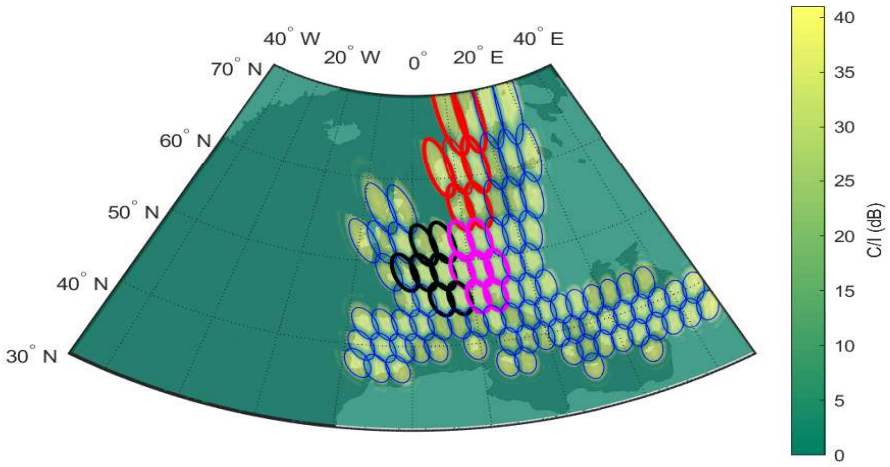


Figure 3-2: The footprints of a realistic 96-beams satellite [70, 71].

Notably, large-scale simulations with an optimization algorithm are typically infeasible in multibeam SATCOM [12, 37, 80], due to two facts: (1) the computational complexity of the optimization algorithm (e.g., the interior-point algorithm) scales

[1]This work adopts single carrier.

with the number of beams; and (2) since the excellent beam isolation is formed by highly directional antennae on the satellite payload, the high-dimensional channel matrix is ill conditioned, making the optimization problem impossible to be solved exactly or even intractable. Herein, an ill-conditioned channel matrix with a high dimension is further indicated by the C/I map of Figure 3-2 that the co-channel interference is dominated by the interference of neighboring beams and beam clusters. To overcome this issue, following the solid method in [12, 37, 80], the optimization variable should be a low-dimensional vector variate to execute the optimization. The resulting optimization solution accounts for neighboring beams and beam clusters by performing small-scale simulations in this work. Then, the attained spectrum efficiency is averaged over the number of beams (this work terms it as "the spectrum efficiency per beam"). The spectrum efficiency per beam can further assist the extension of small-scale simulation results to a large-scale multibeam SATCOM system, as the spectrum efficiency of a multibeam SATCOM system approximately increases in proportion to the number of beams, rationalized and verified in [83]. Thus, small-scale simulation results can provide an evident indication of whether my resource optimization strategies are efficient. According to the link budget parameters defined in [32], I summarize the considered parameters in TABLE 3.1.

Table 3.1: Considered Link Parameters

| **Parameter** | **Value** |
|---|---|
| The location of satellite | 19.2°E |
| Number of beams $K$ | 21 |
| The distribution of users | Uniformly distributed |
| Frequency band | 20 GHz (Ka) |
| Atmospheric effect | Rain attenuation |
| Receiver gain to noise temperature | 25.32 and 11.32 (dB/K) |
| Satellite feed gain | 53.23 (dBW) |
| User link bandwidth $B$ | 500 (MHz) |
| 3dB beamwidth | 0.2° |

### 3.7.1 Exemplary Outcome of User Scheduling

In this subsection, an exemplary outcome of the user scheduling strategy is exhibited. In the scenario under study, the antenna gain pattern points away from the subsatellite point. Thus, the footprint of a circular aperture appears elliptical as it intersects the surface of the Earth, which is approximately spherical: (Figure 3-2). It

is worth noting that (i) the boresights are oriented slightly less than 0.2 degree apart to prevent gaps in coverage, resulting in slightly overlapped 0.2 degree 3 dB footprint contours; and (ii) the 3 dB footprint contours become more elongated as they move from the equator to the polar region. Consequently, the intersected or overlapped areas between the neighboring three-footprint contours show irregular shapes. This region of intersection of three-footprint contours is within the field of view (FOV) of three corresponding feeds. It is reasonable that an interference (or signal) from one feed is seen simultaneously by all users with full frequency use in this intersection region. Therefore, it is important to perform an initial beam association in the beam-overlapped areas. Hence, users within an initial user pool are first associated with appropriate satellite beams, which is operated at the GWs based on its highest beam gain with respect to center points of beams covering a specific geographical region, such as Europe. In the following, I present the outcome of beam association and the achievable throughput performance of the proposed user scheduling.

Once each beam coverage (i.e., each beam footprint) on the Earth is known, the users can be distributed in the manner of a grid mapped over the Earth. Based on this deployment, Figure 3-3 shows the beam association outcome with the highest beam-gain methodology, where the users belonging to the same beam coverage are associated together with the highest beam-gain metric. Orange dots represent the original users within the users' pool. Blue circles denote the users associated with beam 5. Green circles indicate the users associated with beam 6. Since the users originated from the same beam coverage are indeed associated together, each feed is facilitated to only schedule the users from its beam association. As a result, the high user channel similarity after beam association is attainable, which is to observed in Figure 3-4.

### 3.7.2 Highly-correlated Channels Inherent within Each Beam

Importantly, a high user-channel correlation within each beam is inherent in a multi-beam geo-satellite system. To illustrate this, an analysis of channel correlation is performed for a multibeam geo-satellite system with the help of numeric simulations.

For instance, it is demonstrated in Figure 3-4 that the maximum correlation coefficient, the minimum correlation coefficient, and the percentage of the users with the channel correlation $\delta^0 \geq 0.94$ (versus the total number of users per beam) are 0.9998, 0.8373, $\frac{609}{905} \times 100\% = 67.29\%$ and 0.9999, 0.8075, $\frac{588}{855} \times 100\% = 68.77\%$ for beams 1 and 2, respectively. During each simulation run, a reference user $\mathbf{u}_{b,g}^{(i^\diamond)}$ is randomly se-

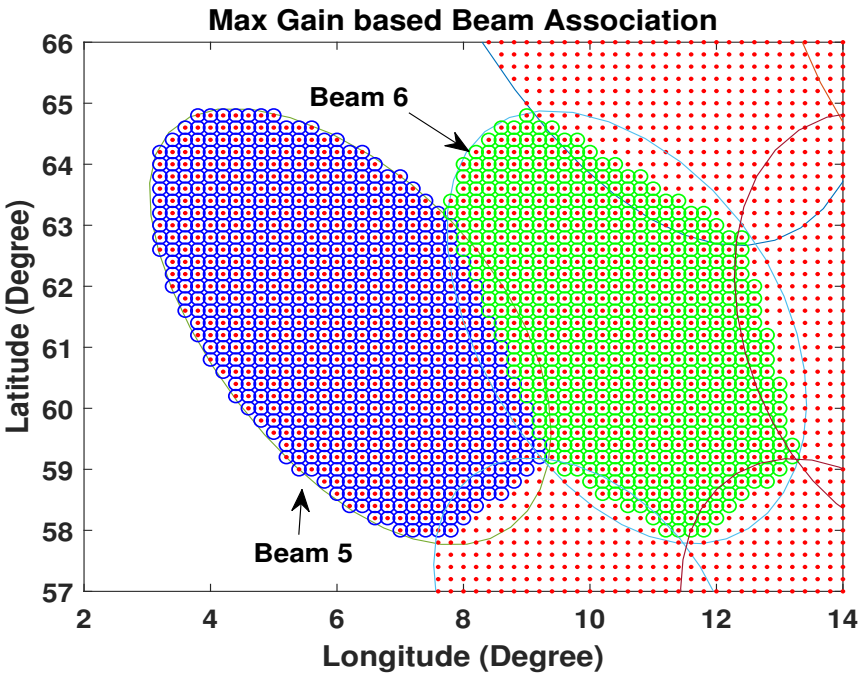


Figure 3-3: The outcome of a maximum beam gain-based beam association.

lected from its beam association $\Upsilon_{b,g}$ (or its beam coverage area). That is, a user $\mathbf{u}_{b,g}^{(i^\diamond)}$ is randomly selected from the $|\Upsilon_{1,1}| = 905$ and $|\Upsilon_{2,1}| = 855$ users within the beam $b = 1$ and beam $b = 2$ served by GW $g = 1$, respectively. Subsequently, the channel correlations of the remaining users per beam are computed with respect to the chosen reference user as $\left|\mathbf{h}_{b,g}^{(i^\diamond)}\left(\mathbf{h}_{b,g}^{(j)}\right)^H\right| \Big/ \left\|\mathbf{h}_{b,g}^{(i^\diamond)}\right\|_2 \left\|\mathbf{h}_{b,g}^{(j)}\right\|_2$, $j \in \{\Upsilon_{b,g} \backslash \mathbf{u}_{b,g}^{(i^\diamond)}\}$, $b \in \{1, 2\}$, $g = 1$. Thus, it is concluded in the above illustrated example that (1) high channel correlation among the users per beam emerges in the general case; and (2) the maximum channel correlation is always close to 1. It is also noticed that the minimum channel correlation may be larger or smaller for another simulation run since the first reference user is randomly chosen during each simulation run. Moreover, it is further unveiled through numerical simulations that the required threshold of channel correlation ($\delta^0 \geq 0.94$) is identified to leverage BF-based interference suppression for the feasibility of the upcoming optimization problem (5.1). Herein, the beam-association method is adopted to gather such high user channel correlation, resulting in a substantial throughput gain as shown in the following section.

### 3.7.3 The Performance Evaluation

Based on the above simulation model, the system throughput performance of the following three schemes is first compared:

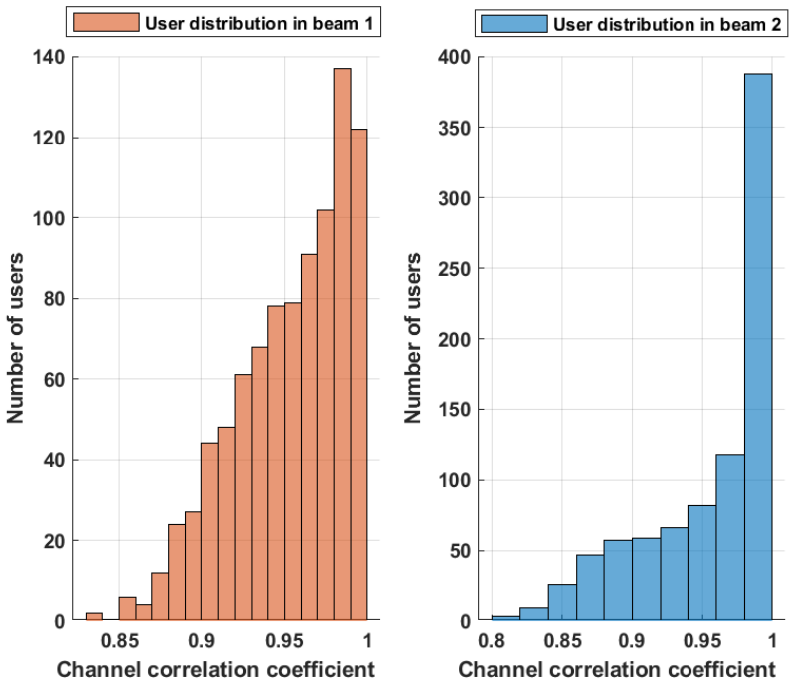


Figure 3-4: An exemplar of channel correlation for beam 1 and beam 2.

- "**Distributed BF**", where $\mathbf{w}_{b,g}$ from Section 4.3 is used. The WSRM PA from Chapter 5 and the user scheduling from this chapter are performed.
- "**User Scheduling, [53]**", where user scheduling from [53] is adopted. $\mathbf{w}_{b,g}$ from Section 4.3 and the WSRM PA from Chapter 5 are executed.
- "**OMA-BF**", where $N_u$ time blocks are considered in the orthogonal multiple access BF scheme. During every time block, each of $N$ users is from different beams and $N$ users are supplied together.

Following the MIMO literature [84–87], the spectrum efficiency per beam is evaluated in terms of the weighted sum rate, since the optimization objective of the problem (5.1) is the maximization of weighted sum rate. To quantify the potential benefits of the strategies, the spectrum efficiency per beam (or equivalently system throughput) is then calculated as

$$\sum_{g\in G}\sum_{b\in N_g}\sum_{i\in S_{b,g}} \varsigma_{b,g}^{(i)} R_{b,g}^{(i)}/(N)^{-1} \tag{3.6}$$

Then, the effect of user scheduling on the system throughput performance is evaluated by comparing the performance of the proposed user scheduling with the solution of the

"User Scheduling, [53]". It is depicted in Figure 3-5 that the proposed user scheduling outperforms the "User Scheduling, [53]" scheme, since the performance benefit comes from: 1) the scheduled users with higher channel correlation are achieved by adopting the highest gain-based beam association, compared to the counterpart by using random (or no) beam association. As a result, the BF strategy in Chapter 4 can efficiently manage intergateway and interbeam interference in this study; and 2) on the other hand, the underlying difference of the users' channel gains is exploited by the K-means algorithm in the last two phases of the user scheduling to unleash the full potential of NOMA. In this figure, the orthogonal transmission (i.e., "OMA-BF" scheme) is also simulated as the performance benchmark.

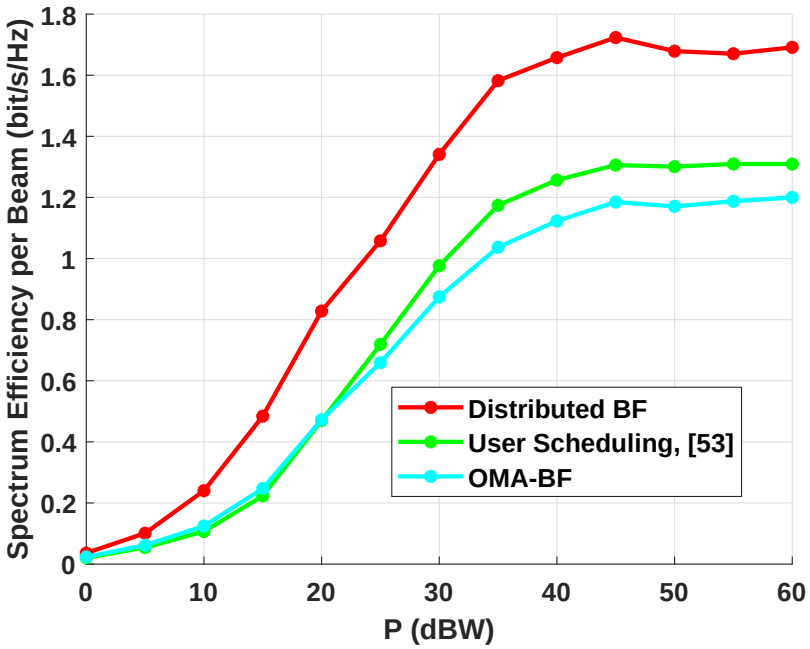


Figure 3-5: Spectrum efficiency against transmit power with the use case $\varsigma_{b,g}^{(1)} = 0.5$, $\varsigma_{b,g}^{(2)} = 0.9$ and $\delta^0 = 0.98$.

Furthermore, I evaluate target data rates of the scheduled users in the "**Distributed BF**" scheme to check if the scheduled users are well served, where the parameters are set to: $N_u = 2$, target data rate $\bar{R}_{b,g}^{(i)} = R_{b,g}^{(i),\mathrm{OMA}}$, $\varsigma_{b,g}^{(1)} = 0.3$, $\varsigma_{b,g}^{(2)} = 0.9$ and $\delta^0 = 0.98$. As shown in Figure 3-6, the targeted data rates for the scheduled users are strictly satisfied with $P_{\mathrm{Total}} = 24$ dBw. Here, the achievable rates are instantaneous data rates from one simulation run for the scheduled users. $R_{b,g}^{(i),\mathrm{OMA}}$ is the corresponding user data rate achieved by the "OMA-BF" scheme.

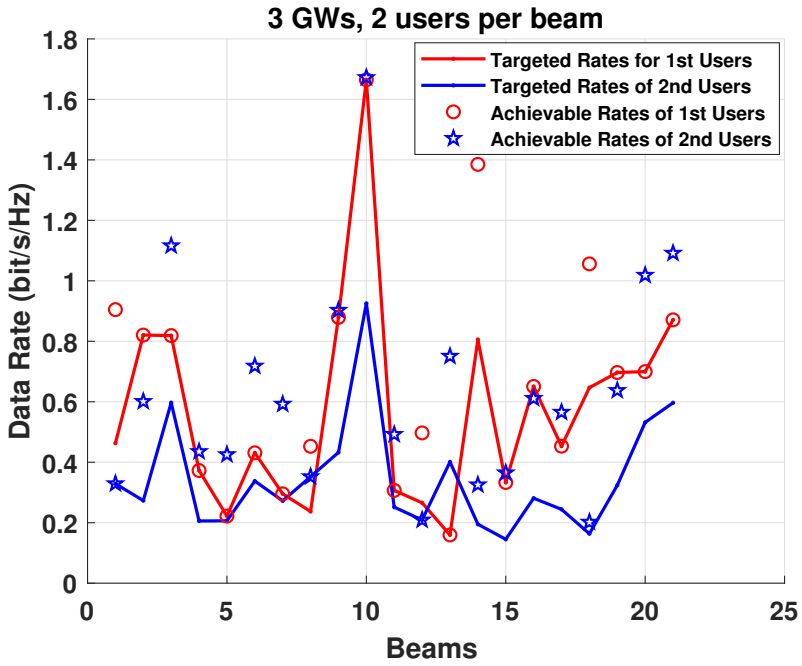


Figure 3-6: The achievable data rates for the scheduled users.

## 3.8 Summary

A distributed user scheduling is developed with the customized beam association and K-means method. The system performance of the distributed user scheduling scheme is evaluated under an exemplary multi-GW multibeam SATCOM scenario. The scientific discoveries of this chapter are outlined in the following:

- The maximum gain-based beam association method is adopted to gather high user channel correlation within each beam, which is inherent in a multibeam geo-satellite system.
- The K-means algorithm is utilized in the last two phases of the user scheduling to unleash the full potential of NOMA, where the users with similar channel gains are grouped and the underlying difference of the users' channel gains is discovered.
- It is identified from the simulation results that the feasibility of the optimization problem ( 5.1) is achievable when the users are scheduled with high channel correlation $\delta^0 \geq 0.94$. It indicates that the BF strategy can efficiently manage intergateway and interbeam interference when the scheduled users are with higher channel correlation, resulting in the substantial system-throughput gain and the satisfied user traffic demand.

# Chapter 4

# Multiple Gateway On-ground Beamforming

*In this chapter, the optimization processing strategies for multi-GW on-ground BF are presented. Particularly, a distributed optimization of PAPC-constrained BF is determined via maximization of the worst-user SLNR, where BF matrix is to be computed and processed in geographically distributed areas with local CSI for a distributed operation of GWs. Before proceeding to the distributed multi-GW BF, a centralized multilayer BF is optimized via maximization of the worst-user SNR, regarded as an upper-bound performance. Then, the computational complexity for the BF strategies is analyzed. Last, the performance of the considered multiple-GW BF strategies is evaluated in terms of spectrum efficiency and user fairness.*

## 4.1 State-of-the-Art

In this chapter, the fully on-ground BF schemes (including the distributed BF and the centralized BF) are investigated for the inference mitigation. With respect to the fully on-ground BF, this chapter compares the proposed BF optimization strategies with the state-of-the-art counterparts in [14] and [32].

Especially, for the distributed BF, the state-of-the-art proposal in [14] has the best throughput performance. Furthermore, as an upper-bound performance limit, the centralized BF is also investigated. For the centralized BF, it is concluded in [88] that the proposal in [32] (i.e. the Block-SVD scheme) combined with user scheduling has the best throughput performance. However, state-of-the-art counterparts in [14] and [32] does not investigate the user fairness optimality, along with other practical constraints on SATCOM, such as the PAPC on board. To solve the problems, the BF

strategies in this work aim to optimize the user fairness compared to the state-of-the-art counterparts in [14] and [32], respectively. The following contents are dedicated to clearly presenting my proposals.

## 4.2 Centralized BF with Full CSI

In this subsection, centralized multilayer BF is calculated and applied to all beams in a central unit with full CSI and data sharing. In such a system acting as a single macro GW, data are synchronously conveyed by all GWs, and all feeds of the GWs cooperatively supply every user. Following the proposal in [32], a single-GW multilayer BF is designed in two steps[1]: (i) outer BF $\mathbf{W}_{\mathrm{O}}$ is first conceived via block SVD processing for mitigating the interference among the beams; and (ii) in the second step, I design the inner BF $\mathbf{W}_{\mathrm{I}}$ to maximize the worst-user SNR on the effective interference-free channel, instead of maximizing the average user SNR per beam considered in [32]. Thus, the full BF matrix is given as

$$\mathbf{W}_{\text{full}} = \mathbf{W}_{\mathrm{O}}\mathbf{W}_{\mathrm{I}} \tag{4.1}$$

where

$$\mathbf{W}_{\mathrm{O}} = \left[ \begin{array}{ccc} \mathbf{W}_{\mathrm{O}}^{1} & \cdots & \mathbf{W}_{\mathrm{O}}^{N} \end{array} \right] \in \mathbb{C}^{N \times NN_u}$$

and

$$\mathbf{W}_{\mathrm{I}} = \mathbf{blkdiag}[\ \mathbf{w}_{\mathrm{I}}^{1},\ \ \mathbf{w}_{\mathrm{I}}^{2},\ \ \cdots,\ \ \mathbf{w}_{\mathrm{I}}^{N}\ ] \in \mathbb{C}^{NN_u \times N}.$$

where the symbol $\mathbf{blkdiag}[\cdot]$ denotes block diagonal concatenation of matrix input arguments.

Specifically, the entire channel matrix of the selected users is denoted by $\mathbf{H} \in \mathbb{C}^{NN_u \times N}$, where each submatrix $\mathbf{H}_g^g$ located at row partition $g$ and column partition $g$ of block matrix $\mathbf{H}$ is a $N_g N_u \times N_g$ channel matrix from GW $g \in \mathcal{G}$ to its serving users. The row $(b-1)N_u + i$ of $\mathbf{H}_g^g$ is the channel vector $\mathbf{h}_{b,g}^{(i)}$ for each user $i \in [1, N_u]$ in beam $b \in \mathcal{N}_g$ served by GW $g$, which is already given by (2.28). Then, I define $\tilde{\mathbf{H}}_k = \left[\mathbf{H}_1^{(R),T}, \cdots \mathbf{H}_{k-1}^{(R),T}, \mathbf{H}_{k+1}^{(R),T}, \cdots \mathbf{H}_N^{(R),T}\right]^T \in \mathbb{C}^{(N-1)N_u \times NN_u}$, where $\mathbf{H}_k^{(R)} \in$

[1]Since this work applies a separate BF and PA optimization, I cannot investigate a BF optimization, essentially converted or related to a corresponding PA optimization with a different objective function in a traditional multicast system [12]. Furthermore, it is concluded in [88] that the proposal in [32] combined with user scheduling has the best throughput performance compared with other works [12].

$\mathbb{C}^{N_u \times N N_u}$ is created by removing the corresponding columns of the regularized channel $\mathbf{H}^{(R)} = \mathbf{H}\mathbf{H}^H + \frac{N N_u}{P_{\text{total}}}\mathbf{I}$. Thus, $\tilde{\mathbf{H}}_k = \tilde{\mathbf{U}}_k \tilde{\mathbf{\Sigma}}_k \tilde{\mathbf{V}}_k^H$ is obtained via the SVD of $\tilde{\mathbf{H}}_k$. In the first step, the goal of the outer layer BF is to only suppress interbeam interference. Consequently, based on the block-SVD method [32], the outer BF $\mathbf{W}_{\text{O}}^j \in \mathbb{C}^{N \times N_u}$ for beam $j \in [1, N]$ is given by

$$\mathbf{W}_{\text{O}}^j = \mathbf{H}^H \tilde{\mathbf{V}}_k \left(:, (N-1)\, N_u + 1 : N N_u\right) \tag{4.4}$$

where the notation of Matlab is adopted for the ease of explanation.

After performing outer BF $\mathbf{W}_{\text{O}}^j$, the effective channel (i.e. the $j$-th beam radiation pattern) is formed as

$$\mathbf{H}_{\text{eff}}^j = \mathbf{H}\left((j-1)+1 : (j-1)+N_u, j\right) \mathbf{W}_{\text{O}}^j \in \mathbb{C}^{N_u \times N_u} \tag{4.5}$$

In the step two, according to the equation (4.5), the inner BF $\mathbf{w}_{\text{I}}^j \in \mathbb{C}^{N_u \times 1}$ aims to maximize the worst-user SNR within the $j$-th beam. Mathematically, $\mathbf{w}_{\text{I}}^j$ is selected to optimize the following max-min fair problem

$$\begin{aligned} &\max_{\mathbf{w}_{\text{I}}^j} \min_{k \in [1, N_u]} |\mathbf{H}_{\text{eff}}^j(k,:)\mathbf{w}_{\text{I}}^j|^2 \\ &\text{s.t.} \|\mathbf{w}_{\text{I}}^j\|_2 \leq 1 \end{aligned} \tag{4.6}$$

which is efficiently resolved via using SDR methodology [89, 90]. That is, by introducing a slack variable $z_{\text{I}}^j, \forall j$ and applying the SDR $\mho_j = \mathbf{w}_{\text{I}}^j \mathbf{w}_{\text{I}}^{jH}, \forall j$, problem (4.6) is relaxed and reformulated to a semidefinite programming (SDP) problem

$$\begin{aligned} &\max_{\mho_j \in \mathbb{C}^{N_u \times N_u}, z_{\text{I}}^j \in \mathbb{R}} \quad z_{\text{I}}^j \\ &\text{s.t.} \begin{cases} C_1 : z_{\text{I}}^j \leq \text{Tr}(\mho_j (\mathbf{H}_{\text{eff}}^j(k,:))^H \mathbf{H}_{\text{eff}}^j(k,:)), \ \forall j, k \\ C_2 : \text{Tr}(\mho_j) \leq 1, \ \forall j \\ C_3 : \mho_j \succeq 0, \ \forall j \end{cases} \end{aligned} \tag{4.7}$$

which is a convex optimization problem and can be efficiently solved by using the CVX optimization toolbox in Matlab. Since the rank-one constraint on the $\mho_j, \forall j$ is relaxed for a tractable solution of problem (4.7), it is challenging to precisely extract

$\mathbf{w}_{\mathrm{I}}^{j}$ from the optimal solution $\mho_j$ (denoted as $\mho_j^*$) of problem (4.7). To this end, let us express $\mho_j^*$ as $\mho_j^* = \mathbf{P}_j\mathbf{D}_j(\mathbf{P}_j)^H$ by conducting its eigenvalue decomposition. When $\mho_j^*$ is a matrix of rank 1 (i.e. $\mathrm{rank}(\mho_j^*) = 1$), $\mathbf{w}_{\mathrm{I}}^{j}$ is selected as

$$\mathbf{w}_{\mathrm{I}}^{j} = \frac{\mathbf{P}_j(:,i)}{||\mathbf{W}_{\mathrm{O}}^{j}\mathbf{P}_j(:,k)||_2}, \text{ if } \mathrm{rank}(\mho_j^*) = 1, \ \forall j,k \tag{4.8}$$

where $\mathbf{P}_j(:,i)$ is the $i$-th eigenvector with the primary eigenvalue (i.e., the nonzero eigenvalue). Since the separate BF and PA optimization design is applied in this work, the normalization operation of the BF vector is herein conducted to prevent the power recalculation in the subsequent PA strategy. When $\mho_j^*$ is a higher-rank matrix (i.e., $\mathrm{rank}(\mho_j^*) > 1$), a modified version of the Gaussian randomization approach [89, 91] is applied to approach a rank-one BF vector as $\mathbf{P}_j\mathbf{D}_j^{\frac{1}{2}}\mathsf{T}_j$ from $\mho_j^*$, and the resulting $\mathbf{w}_{\mathrm{I}}^{j}$ is constructed as

$$\mathbf{w}_{\mathrm{I}}^{j} = \frac{\mathbf{P}_j\mathbf{D}_j^{\frac{1}{2}}\mathsf{T}_j}{||\mathbf{W}_{\mathrm{O}}^{j}\mathbf{P}_j\mathbf{D}_j^{\frac{1}{2}}\mathsf{T}_j||_2}, \text{ if } \mathrm{rank}(\mho_j^*) > 1, \ \forall j \tag{4.9}$$

where $\mathsf{T}_j, \ \forall j$ is a complex Gaussian random vector with zero mean and unit variance. Here, the best $\mathsf{T}_j, \ \forall j$ can be captured through the following steps: (i) generating a set of the candidates $\{\mathsf{T}_j, \ \forall j\}$; (ii) scaling each candidate to satisfy the constraint $C_2$ of problem (4.7); (iii) choosing the best $\mathsf{T}_j, \ \forall j$ from the candidates to

$$\max \min_{k\in[1,N_u]} \mathrm{Tr}(Ł_j {Ł_j}^H(\mathbf{H}_{\mathrm{eff}}^{j}(k,:))^H\mathbf{H}_{\mathrm{eff}}^{j}(k,:)), \ \forall j,k \tag{4.10}$$

where $Ł_j$ is each entity from the set $\{\mathbf{P}_j\mathbf{D}_j^{\frac{1}{2}}\mathsf{T}_j, \ \forall j\}$ and is introduced only for the ease of notation. As explained in Section 2.3.2, this section focuses on the BF direction optimization (i.e. the normalization operations in (4.8) and (4.9)). I leave the payload power resource allocation to be further optimized for performance enhancement in Chapter 5. In this case, I simplify the extraction of $\mathbf{w}_{\mathrm{I}}^{j}$ from $\mho_j^*$ by ignoring the above scaling operation of Step ii, which can be compensated by a subsequential payload power resource-allocation strategy.

## 4.3 Distributed BF with Local CSI

The BF described above requires a central processing unit with the global CSI and data sharing, resulting from the full cooperation between the GWs. To decouple the mandatory, expensive cooperation among the GWs, a distributed BF strategy is investigated with local CSI, inspired by the work of Mosquera *et al.* [14] and Sadek *et al.* [92]. Since each GW lacks the instantaneous CSI for the corresponding out-of-cluster leakage channels, the GWs must exchange the instantaneous CSI before designing the BF. Because only a group of feeds is accessed by each GW in a distributed manner, the BF matrix $\mathbf{W}_{\text{full}}$ becomes a block-diagonal matrix; that is:

$$\mathbf{W}_{\text{full}} \Rightarrow \mathbf{blkdiag}[\ \mathbf{W}_1,\ \ \mathbf{W}_2,\ \ \cdots,\ \ \mathbf{W}_G\ ] \tag{4.11}$$

where each GW can separately compute and process its BF submatrix as

$$\mathbf{W}_g = [\mathbf{w}_{(g-1)+1,g}, \cdots, \mathbf{w}_{(g-1)+N_g,g}] \in \mathbb{C}^{N_g \times N_g},\ g \in [1, G]$$

Here, each subvector $\mathbf{w}_{b,g}$ ($b \in \mathcal{N}_g = \{(g-1)+1, \cdots, (g-1)+N_g\}$) of $\mathbf{W}_g$ denotes the direction of the BF-weighted subvector for the corresponding beam $b$ served by GW $g$. I further express the BF normalization by $\mathbf{w}_{b,g} = \frac{\hat{\mathbf{w}}_{b,g}}{\|\hat{\mathbf{w}}_{b,g}\|_2}$, where the BF subvector $\hat{\mathbf{w}}_{b,g}$ is designed to maximize the worst-user SLNR in beam $b \in \mathcal{N}_g$, ensuring realistic PAPC. Recalling (2.19), the total leakage from user $i$ in beam $b \in \mathcal{N}_g$ to the users of all other beams is then given as

$$\begin{aligned} L_{b,g}^{(i)} = &\sum\nolimits_{l \in \mathcal{N}_g \setminus \{b\}} \sum\nolimits_{j \in \mathcal{S}_{l,g}} |\mathbf{h}_{l,g}^{(j)} \hat{\mathbf{w}}_{b,g}|^2 \\ &+ \sum\nolimits_{m \in \mathcal{G} \setminus \{g\}} \sum\nolimits_{t \in \mathcal{N}_m} \sum\nolimits_{j \in \mathcal{S}_{t,m}} |\mathbf{h}_{t,g}^{(j)} \hat{\mathbf{w}}_{b,g}|^2 \end{aligned}$$

Let us denote PAPC as $\sum_{b \in \mathcal{N}_g} \wp_m^T \hat{\mathbf{w}}_{b,g} \hat{\mathbf{w}}_{b,g}^H \wp_m \leq P_{g,m}, \forall g \in \mathcal{G}, m \in \mathcal{N}_g$, where $\wp_m$ is the $m$-th column of the $N_g \times N_g$ identity matrix. $P_{g,m}$ is the power budget of the feed $m$ belonging to GW $g$. For convenience of explication, it is herein supposed that the initial powers assigned to each user at each feed are equal, i.e. $\wp_m^T \hat{\mathbf{w}}_{b,g} \hat{\mathbf{w}}_{b,g}^H \wp_m \leq P_{g,m}(N_g)^{-1}$, which is left to be further optimized by satisfying the PAPC $C_3$ of PA optimization problem $\mathcal{P}1$ in Chapter 5. Mathematically, BF design with PAPC for users in beam $b \in \mathcal{N}_g$ is then expressed by using the worst-user SLNR maximization

model as

$$\begin{aligned} &\max_{\hat{\mathbf{w}}_{b,g}} \min_{i\in\mathcal{S}_{b,g}} |\mathbf{h}_{b,g}^{(i)}\hat{\mathbf{w}}_{b,g}|^2\times(L_{b,g}^{(i)}+1)^{-1} \\ &\text{s.t.} \quad \wp_m^T\hat{\mathbf{w}}_{b,g}\hat{\mathbf{w}}_{b,g}^H\wp_m \le P_{g,m}(N_g)^{-1}, \forall g,b,m \end{aligned} \tag{4.14}$$

To handle the nonconvex problem (4.14), I use Charnes-Cooper transformation [93] to get

$$\begin{aligned} &\mathbf{\Psi}_{b,g} = \sum_{l\in\mathcal{N}_g\setminus\{b\}}\sum_{j\in\mathcal{S}_{l,g}} \mathbf{h}_{l,g}^{(j),H}\mathbf{h}_{l,g}^{(j)} + \sum_{m\in\mathcal{G}\setminus\{g\}}\sum_{t\in\mathcal{N}_m}\sum_{j\in\mathcal{S}_{t,m}} \mathbf{h}_{t,g}^{(j),H}\mathbf{h}_{t,g}^{(j)} + \|\hat{\mathbf{w}}_{b,g}\|_2^{-1}\mathbf{I} \\ &\hat{\mathbf{w}}_{b,g} = \mathbf{\Psi}_{b,g}^{\frac{-1}{2}}\bar{\mathbf{w}}_{b,g} \end{aligned} \tag{4.15}$$

By introducing a slack variable $t_{b,g}, \forall g,b$ and applying the SDR $\mathbf{\Theta}_{b,g} = \bar{\mathbf{w}}_{b,g}\bar{\mathbf{w}}_{b,g}^H, \forall g,b$, problem (4.14) is then reconstructed into the below problem

$$\begin{aligned} &\max_{t_{b,g},\mathbf{\Theta}_{b,g}} \ t_{b,g} \\ &\text{s.t.} \begin{cases} C_1: t_{b,g} \le \mathrm{Tr}(\mathbf{\Theta}_{b,g}\mathbf{\Psi}_{b,g}^{\frac{-1}{2},H}\mathbf{h}_{b,g}^{(i),H}\mathbf{h}_{b,g}^{(i)}\mathbf{\Psi}_{b,g}^{\frac{-1}{2}}), \forall g,b,i \\ C_2: \mathrm{Tr}(\wp_m^T\mathbf{\Psi}_{b,g}^{\frac{-1}{2}}\mathbf{\Theta}_{b,g}\mathbf{\Psi}_{b,g}^{\frac{-1}{2},H}\wp_m) \le \dfrac{P_{g,m}}{N_g}, \forall g,b,m \\ C_3: \mathrm{Tr}(\mathbf{\Theta}_{b,g}) = 1, \forall g,b \\ C_4: \mathrm{rank}(\mathbf{\Theta}_{b,g}) = 1, \forall g,b \\ C_5: \mathbf{\Theta}_{b,g} \succeq 0, \forall g,b \end{cases} \end{aligned} \tag{4.16}$$

According to the SDP method, I further recast the nonconvex problem (4.16) as a convex SDP problem by relaxing the rank-one constraint $C_4$ in (4.16); that is:

$$\begin{aligned} &\max_{t_{b,g},\mathbf{\Theta}_{b,g}} \ t_{b,g} \\ &\text{s.t.} \begin{cases} C_1: t_{b,g} \le \mathrm{Tr}(\mathbf{\Theta}_{b,g}\mathbf{\Psi}_{b,g}^{\frac{-1}{2},H}\mathbf{h}_{b,g}^{(i),H}\mathbf{h}_{b,g}^{(i)}\mathbf{\Psi}_{b,g}^{\frac{-1}{2}}), \forall g,b,i \\ C_2: \mathrm{Tr}(\wp_m^T\mathbf{\Psi}_{b,g}^{\frac{-1}{2}}\mathbf{\Theta}_{b,g}\mathbf{\Psi}_{b,g}^{\frac{-1}{2},H}\wp_m) \le \dfrac{P_{g,m}}{N_g}, \forall g,b,m \\ C_3: \mathrm{Tr}(\mathbf{\Theta}_{b,g}) = 1, \forall g,b \\ C_4: \mathbf{\Theta}_{b,g} \succeq 0, \forall g,b \end{cases} \end{aligned} \tag{4.17}$$

Resorting to the CVX optimization toolbox in Matlab, I solve the convex problem

(4.17) to get the optimal $\mathbf{\Theta}_{b,g}$ (described as $\mathbf{\Theta}_{b,g}^*$) with $\|\hat{\mathbf{w}}_{b,g}\|_2 = 1$, which is satisfied later by a normalization postprocessing. By the extraction processing over inverse-substitution $\hat{\mathbf{w}}_{b,g}\hat{\mathbf{w}}_{b,g}^H = \mathbf{\Psi}_{b,g}^{\frac{-1}{2}}\mathbf{\Theta}_{b,g}^*\mathbf{\Psi}_{b,g}^{\frac{-1}{2},H}$ and subsequent normalization postprocessing, one solution of (4.14) is thus acquired as $\frac{\hat{\mathbf{w}}_{b,g}}{\|\hat{\mathbf{w}}_{b,g}\|_2}$, which is certified as follows. Specifically, each relevant term $\mathbf{\Theta}_{b,g}^*$ in (4.17) is substituted by the transformation $\frac{\mathbf{\Theta}_{b,g}^*}{\|\hat{\mathbf{w}}_{b,g}\|_2^2}$ of the term $\frac{\hat{\mathbf{w}}_{b,g}\hat{\mathbf{w}}_{b,g}^H}{\|\hat{\mathbf{w}}_{b,g}\|_2^2}$. Consequently, it is evident that $\frac{\mathbf{\Theta}_{b,g}^*}{\|\hat{\mathbf{w}}_{b,g}\|_2^2}$ is also one solution of (4.17), since $\mathbf{\Theta}_{b,g}^*$ is its solution. Note that the term $\|\hat{\mathbf{w}}_{b,g}\|_2$ in $\mathbf{\Psi}_{b,g}$ of (4.17) is equivalent to 1, as there is a normalization postprocessing $\frac{\hat{\mathbf{w}}_{b,g}}{\|\hat{\mathbf{w}}_{b,g}\|_2}$. Considering the term $\|\hat{\mathbf{w}}_{b,g}\|_2 = 1$ in $\mathbf{\Psi}_{b,g}$, problem (4.14) and (4.17) are equivalent. Thus, $\frac{\hat{\mathbf{w}}_{b,g}}{\|\hat{\mathbf{w}}_{b,g}\|_2}$ solves (4.14), since $\frac{\mathbf{\Theta}_{b,g}^*}{\|\hat{\mathbf{w}}_{b,g}\|_2^2}$ maximizes the worst-user SLNR in (4.17).

Next, I continue to explicitly express $\mathbf{w}_{b,g}$ as

$$\mathbf{w}_{b,g} = \frac{\hat{\mathbf{w}}_{b,g}}{\|\hat{\mathbf{w}}_{b,g}\|_2} \overset{(a)}{=} \frac{\mathbf{\Psi}_{b,g}^{\frac{-1}{2}}\bar{\mathbf{w}}_{b,g}}{\|\mathbf{\Psi}_{b,g}^{\frac{-1}{2}}\bar{\mathbf{w}}_{b,g}\|_2}, \ \forall g, b \tag{4.18}$$

where Step (a) is from $\hat{\mathbf{w}}_{b,g} = \mathbf{\Psi}_{b,g}^{\frac{-1}{2}}\bar{\mathbf{w}}_{b,g}$. $\bar{\mathbf{w}}_{b,g}$ is extracted from $\mathbf{\Theta}_{b,g}^*$, following the same approach as Section 4.2. Specifically, the eigenvalue decomposition of $\mathbf{\Theta}_{b,g}^*$ is first conducted as $\mathbf{\Theta}_{b,g}^* = \bar{\mathbf{P}}_{b,g}\bar{\mathbf{D}}_{b,g}(\bar{\mathbf{P}}_{b,g})^H$. When $\mathbf{\Theta}_{b,g}^*$ is a rank-one matrix, I get $\bar{\mathbf{w}}_{b,g} = \bar{\mathbf{P}}_{b,g}(:,k)$, where $\bar{\mathbf{P}}_{b,g}(:,k)$ is the $k$-th eigenvector with the primary eigenvalue. Then, $\mathbf{w}_{b,g}$ is denoted as

$$\mathbf{w}_{b,g} = \frac{\mathbf{\Psi}_{b,g}^{\frac{-1}{2}}\bar{\mathbf{P}}_{b,g}(:,k)}{||\mathbf{\Psi}_{b,g}^{\frac{-1}{2}}\bar{\mathbf{P}}_{b,g}(:,k)||_2}, \ \text{if } \operatorname{rank}(\mathbf{\Theta}_{b,g}^*) = 1, \ \forall g, b, k \tag{4.19}$$

When $\mathbf{\Theta}_{b,g}^*$ is a higher rank matrix, a rank-one BF vector $\bar{\mathbf{w}}_{b,g}$ is approximated as $\bar{\mathbf{P}}_{b,g}\bar{\mathbf{D}}_{b,g}^{\frac{1}{2}}\bar{\mathsf{T}}_{b,g}$ from $\mathbf{\Theta}_{b,g}^*$, and the resulting $\mathbf{w}_{b,g}$ is given as

$$\mathbf{w}_{b,g} = \frac{\mathbf{\Psi}_{b,g}^{\frac{-1}{2}}\bar{\mathbf{P}}_{b,g}\bar{\mathbf{D}}_{b,g}^{\frac{1}{2}}\bar{\mathsf{T}}_{b,g}}{||\mathbf{\Psi}_{b,g}^{\frac{-1}{2}}\bar{\mathbf{P}}_{b,g}\bar{\mathbf{D}}_{b,g}^{\frac{1}{2}}\bar{\mathsf{T}}_{b,g}||_2}, \ \text{if } \operatorname{rank}(\mathbf{\Theta}_{b,g}^*) > 1, \ \forall g, b \tag{4.20}$$

where $\bar{\mathsf{T}}_{b,g}, \ \forall g, b$ is a complex Gaussian random vector with zero mean and unit variance. The same processing procedure of Section 4.2 can be adopted to achieve the best $\bar{\mathsf{T}}_{b,g}, \ \forall g, b$, except for the last step. In particular, the last step selects the

best $\bar{\mathsf{T}}_{b,g},\ \forall g,b$ from the candidates $\{\bar{\mathsf{T}}_{b,g},\ \forall g,b\}$ as

$$\max \min_{i\in[1,N_u]} \mathrm{Tr}(\bar{\mathbb{L}}_{b,g}\bar{\mathbb{L}}_{b,g}^{H}\boldsymbol{\Psi}_{b,g}^{\frac{-1}{2},H}\mathbf{h}_{b,g}^{(i),H}\mathbf{h}_{b,g}^{(i)}\boldsymbol{\Psi}_{b,g}^{\frac{-1}{2}}),\ \forall g,b \tag{4.21}$$

where $\bar{\mathbb{L}}_{b,g}$ is each entity from the set $\{\bar{\mathbf{P}}_{b,g}\bar{\mathbf{D}}_{b,g}^{\frac{1}{2}}\bar{\mathsf{T}}_{b,g},\ \forall g,b\}$ and is introduced only for the ease of notation. Thus far, the distributed BF does not account for an autonomous operation at each GW, as the local CSI acquisition at each GW involves the instantaneous CSI exchange among the GWs. It is worth noting that if the instantaneous CSI is unavailable, the channel statistics can be instead employed as input parameters of the optimization problem (4.17) (i.e., $\mathbf{h}_{b,g}^{(i),H}\mathbf{h}_{b,g}^{(i)} \approx \mathbb{E}[\mathbf{h}_{b,g}^{(i),H}\mathbf{h}_{b,g}^{(i)}], \forall g,b,i$ can be adopted in the optimization problem (4.17)), first proposed in the work of Mosquera *et al.* [14] for the autonomous operation of the GWs in a multibeam SATCOM system.

## 4.4 Complexity Evaluation

Based on the complexity of the interior point methods, the computational complexity is evaluated for the BF strategies. That is, a worst-case complexity of solving an SDP problem with the interior point method is [90]

$$\mathcal{O}(\max\{m_{\mathrm{SDP}}, n_{\mathrm{SDP}}\}^4 (n_{\mathrm{SDP}})^{\frac{1}{2}} \log\frac{1}{\varepsilon})$$

where $m_{\mathrm{SDP}}$, $n_{\mathrm{SDP}}$ and $\varepsilon$ are the dimensions of a semidefinite matrix variable, the number of linear constraints and a given solution accuracy, respectively. Specifically, I assess the corresponding complexity for each BF strategy as follows:

- *In the centralized BF strategy:* The computational complexity of the outer BF for each beam is dominated by an eigenvector decomposition of a matrix of size $N \times N$ (i.e., the computational complexity involved in the first step is $\mathcal{O}(N^3)$). In the step two, the complexity of the inner BF for each beam primarily arises from using the interior point method to find the optimal solution to the SDP problem (4.7). Because the SDP problem (4.7) has one matrix variable of size $N_u \times N_u$ and $N_u$ linear constraints, the interior point method requires a worst-case complexity of $\mathcal{O}(N_u^{4.5}\log\frac{1}{\varepsilon_1})$ to seek the $\varepsilon_1$-optimal solution. Next, an eigenvector decomposition operation and a Gaussian randomization process

are successively applied to the preceding $\varepsilon_1$-optimal solution for each beam. Herein, an eigenvector decomposition operation of a $N_u \times N_u$ matrix has the complexity of $\mathcal{O}(N_u^3)$. The complexity of a Gaussian randomization process is $\mathcal{O}(N_u^2 I_{\text{rand}})$ [91], where $I_{\text{rand}}$ is the amount of randomization. In actuality, the complexity of a Gaussian randomization process is insignificant for modest $I_{\text{rand}}$. Next, the whole complexity of the centralized BF strategy for the $N$ beams can be counted as $\mathcal{O}(N(N^3 + N_u^3 + N_u^{4.5} \log \frac{1}{\varepsilon_1})) \sim \mathcal{O}(N^4 + N N_u^{4.5} \log \frac{1}{\varepsilon_1})$.

- *In the distributed BF strategy:* The main computational complexity of the BF vector at GW $g \in \mathcal{G}$ originates from seeking the $\varepsilon_2$-optimal solution of the standard SDP problem (4.17). As the SDP problem (4.17) has one matrix variable of size $N_g \times N_g$ and $N_u + N_g$ linear constraints, the interior point method requires a worst-case complexity of $\mathcal{O}((N_u + N_g)^{4.5} \log \frac{1}{\varepsilon_2})$ to find the $\varepsilon_2$-optimal solution. Following the aforementioned evaluation approach, an eigenvector decomposition operation of a $N_g \times N_g$ matrix has the complexity of $\mathcal{O}(N_g^3)$. The complexity of a Gaussian randomization process with $\dot{I}_{\text{rand}}$ randomization is $\mathcal{O}(N_g^2 \dot{I}_{\text{rand}})$. As a result, GW $g \in \mathcal{G}$ has the overall computation complexity of the distributed BF strategy for the $N_g$ BF vectors as $\mathcal{O}(N_g(N_g^3 + (N_u + N_g)^{4.5} \log \frac{1}{\varepsilon_2})) \sim \mathcal{O}(N_g(N_u + N_g)^{4.5} \log \frac{1}{\varepsilon_2})$.

## 4.5 Performance Evaluation

This section assesses the system throughput and user fairness performance for the proposed BF strategies. On the basis of the aforementioned simulation model, the performance for the following schemes are evaluated.

- "**Centralized BF**", where $\mathbf{W}_{\text{full}}$ from Section 4.2 is used. The WSRM PA from Chapter 5 and the user scheduling from Chapter 3 are performed.
- "**Centralized, [32]**", where $\mathbf{W}_{\text{full}}$ is replaced by that from [32]. The WSRM PA from Chapter 5 and the user scheduling from Chapter 3 are used.
- "**Distributed BF**", where $\mathbf{w}_{b,g}$ from Section 4.3 is used. The WSRM PA from Chapter 5 and the user scheduling from Chapter 3 are performed.
- "**Distributed, [14]**", where $\mathbf{w}_{b,g}$ is replaced by that from [14][2]. The WSRM PA from Chapter 5 and the user scheduling from Chapter 3 are performed.

[2]To apply the distributed unicast BF [14] in the NOMA context, an equivalent channel vector for each beam is generally represented by the channel vector of the user with maximum channel gain.

- "**OMA-BF**", where $N_u$ time blocks are considered in the orthogonal multiple access BF scheme. During every time block, each of $N$ users is from different beams and $N$ users are supplied together.

### 4.5.1 Numerical Results on Throughput Performance

Based on the same simulation model, this part manifests the improved performance achieved by the BF strategies through numerical simulations. The spectrum efficiency per beam versus total power on board $P_{\text{total}}$ for all schemes is shown in the following.

In Figure 4-1, the "Centralized BF" scheme can obtain higher spectrum efficiency than the "OMA-BF" scheme for two users per beam (i.e., $N_u = 2$), resulting from the nonorthogonal strategy. It is also observed that the system spectral efficiency increases as long as the correlation $\delta^0$ of the scheduled users becomes higher. The performance gain results from the interference being efficiently suppressed by the BF strategy when the users with higher correlations are scheduled. Moreover, the "Centralized BF" scheme and the "Centralized, [32]" scheme provide a similarly high level of spectral efficiency. It is represented that the "Centralized BF" scheme does not sacrifice the spectrum efficiency to increase the worst-user SINR. The improved user fairness in the "Centralized BF" scheme is to be depicted in Section 4.5.2.

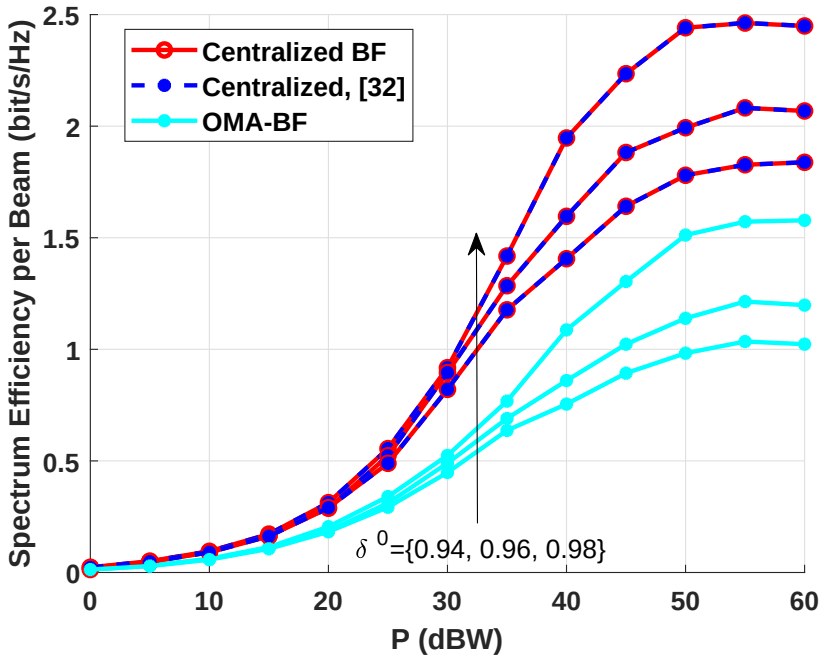


Figure 4-1: Spectrum efficiency against transmit power with the relative user priorities $\varsigma_{b,g}^{(1)} = 0.1$, $\varsigma_{b,g}^{(2)} = 0.9$.

Furthermore, as shown in Figure 4-2, the "Centralized BF" scheme can repeatedly achieve the same spectrum efficiency as the "Centralized, [32]" scheme and certainly act as the upper-bound performance of the "Distributed BF" scheme for two users per beam (i.e., $N_u = 2$). Herein, the performances of the "Centralized BF" scheme in Figures 4-1 and 4-2 are significantly different, since the different user priorities $\varsigma_{b,g}^{(1)} = 0.1$, $\varsigma_{b,g}^{(2)} = 0.9$ and $\varsigma_{b,g}^{(1)} = 0.5$, $\varsigma_{b,g}^{(2)} = 0.9$ are set for Figures 4-1 and 4-2, respectively. It is implied that the data rate of the strong user dominates the spectrum efficiency of the system. Moreover, the "Distributed BF" scheme has a spectrum efficiency gain per beam of 1.46 times in contrast to the "OMA-BF" scheme. It is also noticed in Figure 4-2 that the "Distributed, [14]" scheme is superior to the "Distributed BF" scheme for $N_u = 2$. Nevertheless, since the goal of the "Distributed BF" scheme is to obtain a certain fairness optimality as illustrated in Section 4.5.2, it is precious for the "Distributed BF" scheme to capture the satisfied throughput performance.

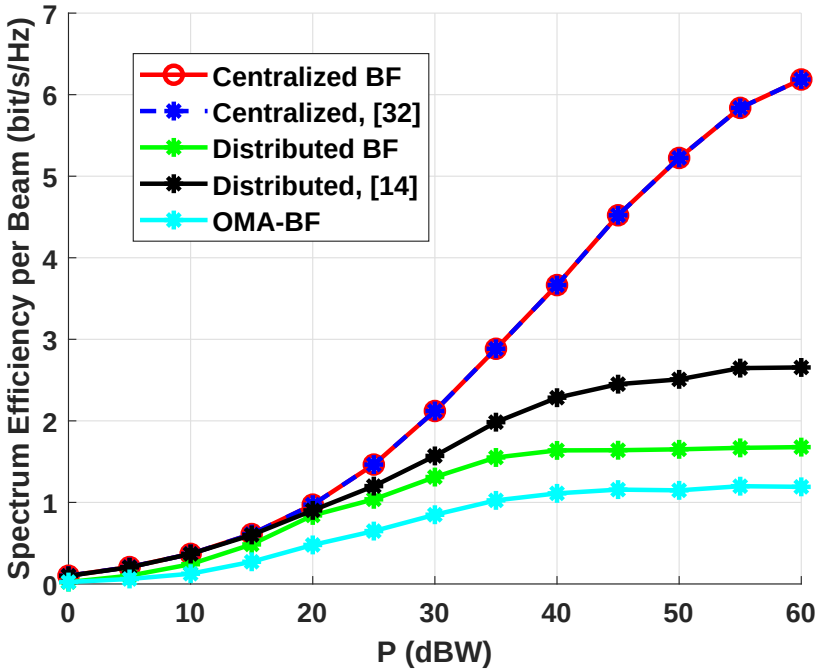


Figure 4-2: Spectrum efficiency against transmit power with the use case $\varsigma_{b,g}^{(1)} = 0.5$, $\varsigma_{b,g}^{(2)} = 0.9$ and $\delta^0 = 0.98$.

### 4.5.2 Numerical Results on User Fairness

This subsection provides results considering the user fairness when the multi-GW BF optimization strategies are adopted. To this end, the user SINR distributions over the whole beam coverage are illustrated.

Specifically, the "Centralized BF" scheme can optimize the worst-user performance per beam by equation (4.6). Herein, the exemplary use cases are described as follows: $N_u = 2$, $\delta^0 = \{0.94,\ 0.96\}$, $\varsigma_{b,g}^{(1)} = 0.1$, $\varsigma_{b,g}^{(2)} = 0.9$. The simulation results of Figures 4-3 and 4-4 clearly demonstrate that the worst-user SINR is improved in the "Centralized BF" scheme, which is shown exemplarily by assessing the empirical cumulative distribution function (CDF) of user SINR[3] in the entire footprints under study. Specifically, it is observed in Figures 4-3 and 4-4 that the worst-user SINR is increased by 31.53 dB and 3.34 dB, compared to that of the "Centralized, [32]" scheme, respectively. Additionally, it is also found in Figures 4-3 and 4-4 that the worst-user SINR is improved when the correlation $\delta^0$ of the scheduled users increases. Here, the higher user correlation is utilized by the outer layer BF (i.e., formula (4.4)) to efficiently suppress the co-channel interference, resulting in the improvement of the worst-user SINR. Furthermore, it is shown from Figures 4-3 and 4-4 that the discrepancy between user SINR distributions with the correlation $\delta^0 = 0.94$ and $\delta^0 = 0.96$ is small in the "Centralized BF" scheme, owing to its user fairness optimization. Remarkably, the higher correlation of the scheduled users gathers the higher user SINR. It indicates that higher user correlation $\delta^0$ provides higher spectral efficiency. This observation coincides well with the result of Figure 4-1 shown in Section 4.5.1.

Last, the "Distributed BF" scheme substantially optimizes the worst-user performance per beam in the exemplary use cases: $N_u = 2$, $P_{\text{total}} = \{25, 35\}$ dBW, $\delta^0 = 0.98$, $\varsigma_{b,g}^{(1)} = 0.5$, $\varsigma_{b,g}^{(2)} = 0.9$. This advantage of the "Distributed BF" scheme over the counterpart is evident in Figures 4-5 and 4-6, where the minimum user SINR is increased by 13.67 dB and 18 dB, compared to that of the "Distributed, [14]" scheme, respectively. Moreover, from Figures 4-5 and 4-6, it is found that the relatively small user SINR becomes lower as long as the total power budget $P_{\text{total}}$ available on payload grows higher. It implies that the high $P_{\text{total}}$ causes the increased co-channel interference, resulting in the lower user SINR. Furthermore, it is unveiled in Figures 4-5 and 4-6 that when the total payload power budget $P_{\text{total}}$ is increased, the user SINR is always balanced via the worst-user SINR optimization in the "Distributed BF" scheme, and the improvement of the worst-user SINR is enlarged. Addition-

[3]In this dissertation, SNR is used as a quantity of interference-free communication link, and SINR is utilized to measure the quality of interference-limited communication link.

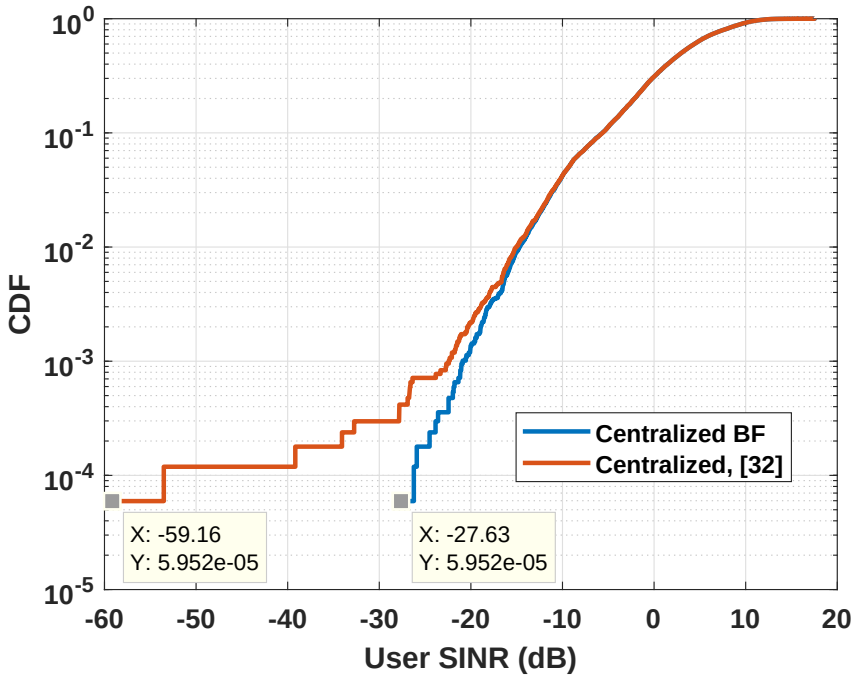


Figure 4-3: CDF of user SINR over the coverage with $\delta^0 = 0.94$, for single GW, central operation at $P_{\text{total}} = 35$ dBW.

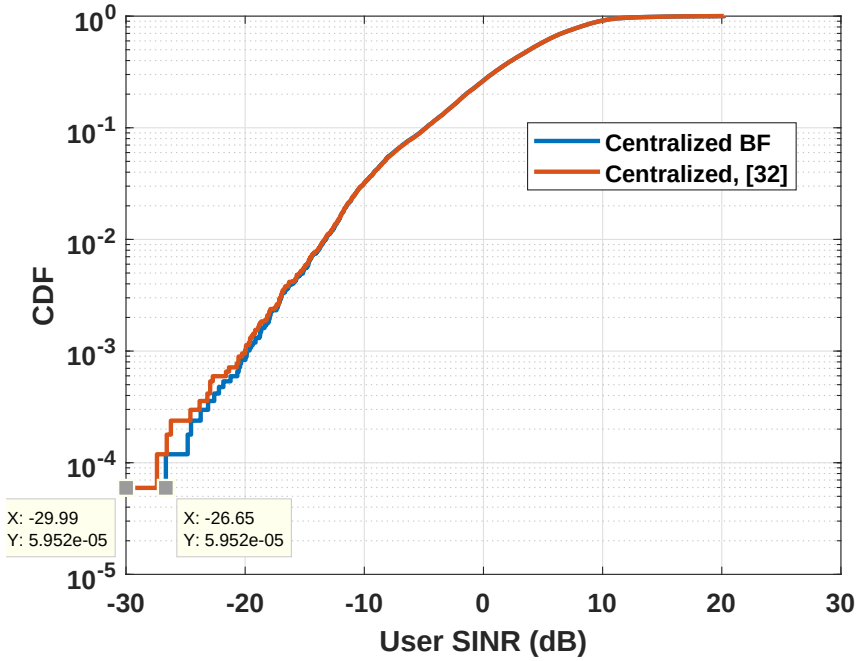


Figure 4-4: CDF of user SINR over the coverage with $\delta^0 = 0.96$, for single GW, central operation at $P_{\text{total}} = 35$ dBW.

ally, it is noteworthy that in Figures 4-5 and 4-6, the relatively high user SINR is significantly increased in the "Distributed, [14]" scheme due to the increased $P_{\text{total}}$, resulting in a distinct system throughput gain at high $P_{\text{total}}$ in the "Distributed, [14]" scheme. This discovery gives an illuminating insight into the results of Figure 4-2 shown in Section 4.5.1. It is interpreted by the truth that the "Distributed, [14]" scheme concentrates on the throughput maximization, ignoring the worst-user SINR optimization.

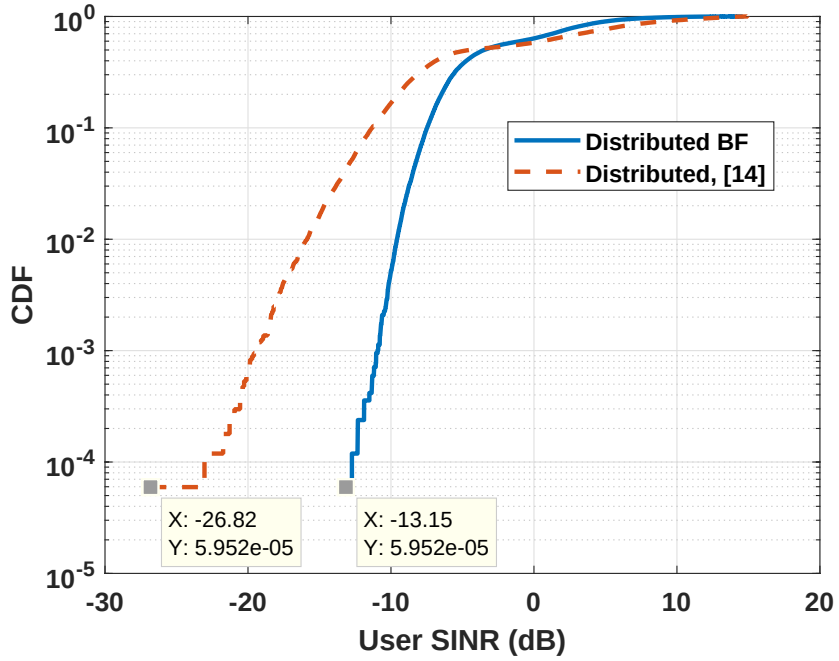


Figure 4-5: CDF of user SINR over the coverage, for 3 GWs, 2 users per frame, distributed operation with $P_{\text{total}} = 25$ dBW.

## 4.6 Summary

The BF strategies for the interference mitigation are proposed, aiming to optimize user fairness. The system throughput and user fairness performance for the BF strategies are assessed. The main results of this chapter are figured out as follows:

- A distributed PAPC constrained BF strategy is proposed via maximizing the worst-user SLNR to be locally processed in geographically decentralized GWs.
- As the throughput performance limit, the centralized multilayer BF strategy based on the block SVD [32] is antecedently explored at a central unit with full

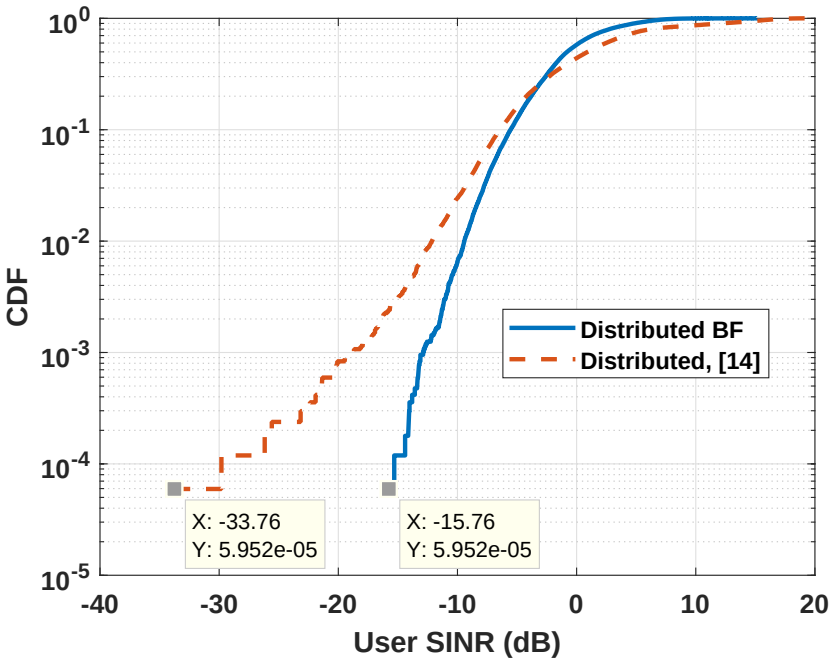


Figure 4-6: CDF of user SINR over the coverage, for 3 GWs, 2 users per frame, distributed operation with $P_{\text{total}} = 35$ dBW.

CSI and data sharing.

The simulation results indicate that the BF strategies can efficiently improve the user fairness, while achieving the satisfactory throughput performance.

# Chapter 5

# Flexible Payload Power Allocation

*After the BF strategy, this chapter proceeds to consider the additional combination of flexible payload power resources-allocation strategy to meet the actual traffic demand. First, the optimization problem is formulated via WSRM under the practical satellite constraints. Sequentially, the nonconvex WSRM problem is addressed by the mathematical optimization-based method and the DNN-based method, respectively. Last, the complexity analysis and performance evaluation for the payload PA-optimization strategies are provided.*

## 5.1 State-of-the-Art

In the existing nonorthogonal single-GW SATCOM, the perfect SIC decoding constraints are simply ignored (or relaxed) in the payload power-resources optimization problem. However, the corresponding optimization objective function certainly accounts for the perfect SIC decoding (i.e., perfect intrabeam interference mitigation within each beam). In actuality, for multibeam SATCOM systems, these successful SIC decoding conditions cannot be automatically satisfied by ordering the users in terms of their effective channel gains or approximate signal-to-interference-plus-noise ratios (SINRs). From the optimization perspective, the resulting power resources-allocation optimization solution is not rigorous.

Hence, there is still a lack of a suitable payload power resources-allocation optimization solution for the nonorthogonal multi-GW SATCOM. Thus, in the nonorthogonal SATCOM, this work aims to investigate the distributed and flexible payload power resources allocation optimization with the assurance of successful SIC decoding constraints and the fulfillment of practical satellite constraints (e.g. the multi-GW architecture). In the following, my proposals are detailed.

## 5.2 Optimization Problem Formulation

To achieve the trade-off between maximum throughput (resulting in no service for weak users) and maximum fairness (resulting in unfavorably equal user power resource allocation for SIC decoding), the payload PA-optimization problem is formulated with the WSRM objective, regularly classified as a transmit power/rate control policy [94]. Mathematically, the WSRM PA problem is formulated as:

$$\mathcal{P}1: \max_{\{p_{b,g}^{(i)}\}} \sum_{g\in\mathcal{G}} \sum_{b\in\mathcal{N}_g} \sum_{i\in\mathcal{S}_{b,g}} \varsigma_{b,g}^{(i)} R_{b,g}^{(i)}$$
$$\text{s.t.} \begin{cases} C_1: R_{b,g}^{(i)} \geq \bar{R}_{b,g}^{(i)}, \quad \forall g, b, i \\ C_2: \sum\limits_{g\in\mathcal{G}} \sum\limits_{b\in\mathcal{N}_g} \sum\limits_{i\in\mathcal{S}_{b,g}} p_{b,g}^{(i)} \leq P_{\text{total}} \\ C_3: \sum\limits_{b\in\mathcal{N}_g} \sum\limits_{i\in\mathcal{S}_{b,g}} p_{b,g}^{(i)} \wp_m^T \mathbf{w}_{b,g} \mathbf{w}_{b,g}^H \wp_m \leq P_{g,m}, \forall g, m \\ C_4: \text{SIC}_{b,g}^{(i)\to(k)} \geq 0, \quad 1 \leq k < i, \quad \forall g, b, i \end{cases} \tag{5.1}$$

where $\{\varsigma_{b,g}^{(i)}\}$ are the user-priority factors employed for obtaining a certain user fairness degree (i.e., the user priority factors are fine-tuned to obtain a certain compromise performance among maximum throughput and maximum fairness) [95]. $R_{b,g}^{(i)}$ is the rate of user $i$ in beam $b \in \mathcal{N}_g$ (i.e. (2.20) derived in Chapter 2). Constraint $C_1$ denotes the user's target rate $\bar{R}_{b,g}^{(i)}$ constraint. Constraint $C_2$ indicates the constraint on the total transmit-power budget. Constraint $C_3$ is PAPC. Constraints $C_4$ ensures that each strong user $k$ ($1 \leq k < i$) can detect the signal of the weak user $i$ within the same beam $b \in \mathcal{N}_g$ by performing successful SIC decoding. To address the nonconvex problem $\mathcal{P}1$, the mathematical optimization-based and the DNN-based methods are investigated in this chapter.

## 5.3 Mathematical optimization-based Solution

To address the problem $\mathcal{P}1$, this work employs the equivalent WMMSE method [96,97]. As the terrestrial WMMSE problems are designated for the joint optimization of BF direction and power resource allocation in the conventional MU-MIMO wireless networks [96, 97], the thorough mathematical derivations of an equivalent WMMSE problem are necessitated for a WSRM based payload power resource-allocation optimization problem in a nonorthogonal multi-GW multibeam SATCOM system [61].

Furthermore, the preliminary inspection method from my earlier work [61] is extended to solve the problem $\mathcal{P}1$ with the additional assurance of successful SIC-decoding constraints. In the following, I proceed to derive such an equivalence.

### 5.3.1 Reformulation and Equivalence via WMMSE

A linear equalizer $f_{b,g}^{(i)}$ is employed at user $i$ in beam $b \in \mathcal{N}_g$ to process the accepted signal $y_{b,g}^{(i)}$, after the perfect SIC decoding is supposed to be accomplished. Thus, $\tilde{x}_{b,g}^{(i)}$ is the estimated data symbol by scaling $y_{b,g}^{(i)}$ with $f_{b,g}^{(i)} \in \mathbb{C}$; that is: $\tilde{x}_{b,g}^{(i)} = f_{b,g}^{(i)} y_{b,g}^{(i)}$. The mean-square estimation error $\epsilon_{b,g}^{(i)}$ of estimating the desired signal $x_{b,g}^{(i)}$ is

$$\begin{aligned}\epsilon_{b,g}^{(i)} = \mathbb{E}[(\tilde{x}_{b,g}^{(i)} - x_{b,g}^{(i)})^2] = |\sqrt{p_{b,g}^{(i)}} f_{b,g}^{(i)} \mathbf{h}_{b,g}^{(i)} \mathbf{w}_{b,g} - 1|^2 + \\ |f_{b,g}^{(i)}|^2 \left(|\mathbf{h}_{b,g}^{(i)} \mathbf{w}_{b,g}|^2 \sum_{j=1}^{i-1} p_{b,g}^{(j)} + \sum_{l \in \mathcal{N}_g \setminus \{b\}} |\mathbf{h}_{b,g}^{(i)} \mathbf{w}_{l,g}|^2 \sum_{j \in \mathcal{S}_{l,g}} p_{l,g}^{(j)} \right. \\ \left. + \sum_{m \in \mathcal{G} \setminus \{g\}} \sum_{l \in \mathcal{N}_m} |\mathbf{h}_{b,m}^{(i)} \mathbf{w}_{l,m}|^2 \sum_{j \in \mathcal{S}_{l,m}} p_{l,m}^{(j)} + 1 \right), \ \forall g, b, i\end{aligned} \tag{5.2}$$

By minimizing the mean-square estimation error $\epsilon_{b,g}^{(i)}$, I can achieve an optimal MMSE equalizer as

$$(f_{b,g}^{(i)})_{\text{MMSE}}^{\text{opt}} = \arg\min \epsilon_{b,g}^{(i)}, \ \forall g, b, i \tag{5.3}$$

Furthermore, I can obtain a closed-form formula for the optimal MMSE receiver by inspecting the first order necessary condition of $\epsilon_{b,g}^{(i)}$; that is:

$$\begin{aligned}\frac{\partial \epsilon_{b,g}^{(i)}}{\partial f_{b,g}^{(i)}} = 0 \Rightarrow (f_{b,g}^{(i)})_{\text{MMSE}}^{\text{opt}} = (\sqrt{p_{b,g}^{(i)}} \mathbf{h}_{b,g}^{(i)} \mathbf{w}_{b,g})^* \left(|\mathbf{h}_{b,g}^{(i)} \mathbf{w}_{b,g}|^2 p_{b,g}^{(i)} + \right. \\ |\mathbf{h}_{b,g}^{(i)} \mathbf{w}_{b,g}|^2 \sum_{j=1}^{i-1} p_{b,g}^{(j)} + \sum_{l \in \mathcal{N}_g \setminus \{b\}} |\mathbf{h}_{b,g}^{(i)} \mathbf{w}_{l,g}|^2 \sum_{j \in \mathcal{S}_{l,g}} p_{l,g}^{(j)} + \\ \left. \sum_{m \in \mathcal{G} \setminus \{g\}} \sum_{l \in \mathcal{N}_m} |\mathbf{h}_{b,m}^{(i)} \mathbf{w}_{l,m}|^2 \sum_{j \in \mathcal{S}_{l,m}} p_{l,m}^{(j)} + 1 \right)^{-1}, \ \forall g, b, i\end{aligned} \tag{5.4}$$

Correspondingly, via substituting the optimum values $(f_{b,g}^{(i)})_{\text{MMSE}}^{opt}$ into (5.2), the

minimum mean-square error estimation is attained as

$$
\begin{aligned}
&(\epsilon_{b,g}^{(i)})_{\mathrm{MMSE}}^{\min} = \min_{(f_{b,g}^{(i)})_{\mathrm{MMSE}}^{\mathrm{opt}}} \epsilon_{b,g}^{(i)} \\
&= (1+|\mathbf{h}_{b,g}^{(i)}\mathbf{w}_{b,g}|^2 p_{b,g}^{(i)} \times \\
&(|\mathbf{h}_{b,g}^{(i)}\mathbf{w}_{b,g}|^2 \sum_{j=1}^{i-1} p_{b,g}^{(j)} + \sum_{l\in\mathcal{N}_g\setminus\{b\}} |\mathbf{h}_{b,g}^{(i)}\mathbf{w}_{l,g}|^2 \sum_{j\in\mathcal{S}_{l,g}} p_{l,g}^{(j)} \\
&+ \sum_{m\in\mathcal{G}\setminus\{g\}} \sum_{l\in\mathcal{N}_m} |\mathbf{h}_{b,m}^{(i)}\mathbf{w}_{l,m}|^2 \sum_{j\in\mathcal{S}_{l,m}} p_{l,m}^{(j)} + 1)^{-1})^{-1}, \ \forall g,b,i
\end{aligned}
\tag{5.5}
$$

Comparing (2.20) and (5.5), the data rate for user $i$ in beam $b \in \mathcal{N}_g$ is reexpressed as

$$
R_{b,g}^{(i)} = -\log_2((\epsilon_{b,g}^{(i)})_{\mathrm{MMSE}}^{\min}), \ \forall g,b,i \tag{5.6}
$$

Next, I use the representation in (5.6) to demonstrate that the WSRM problem $\mathcal{P}1$ can be recast to an equivalent WMMSE problem in my setup. In particular, by introducing additional optimization variables $\{\varpi_{b,g}^{(i)}\}$ [97], it can be inferred from the proof of **Proposition 1** that $\mathcal{P}1$ and WMMSE optimization problem $\mathcal{P}2$ have exactly the identical solution:

$$
\begin{aligned}
\mathcal{P}2: &\min_{\left\{\substack{\varpi_{b,g}^{(i)},\\ p_{b,g}^{(i)}, f_{b,g}^{(i)}}\right\}} \sum_{g\in\mathcal{G}}\sum_{b\in\mathcal{N}_g}\sum_{i\in\mathcal{S}_{b,g}} \varsigma_{b,g}^{(i)}(\varpi_{b,g}^{(i)}\epsilon_{b,g}^{(i)} - \log_2(\varpi_{b,g}^{(i)})) \\
&\text{s.t.} \begin{cases}
C_1: R_{b,g}^{(i)} \geq \bar{R}_{b,g}^{(i)}, \quad \forall g,b,i \\
C_2: \sum_{g\in\mathcal{G}}\sum_{b\in\mathcal{N}_g}\sum_{i\in\mathcal{S}_{b,g}} p_{b,g}^{(i)} \leq P_{\mathrm{total}} \\
C_3: \sum_{b\in\mathcal{N}_g}\sum_{i\in\mathcal{S}_{b,g}} p_{b,g}^{(i)} \wp_m^T \mathbf{w}_{b,g}\mathbf{w}_{b,g}^H \wp_m \leq P_{g,m}, \forall g,m \\
C_4: \mathrm{SIC}_{b,g}^{(i)\to(k)} \geq 0, \quad 1\leq k<i, \quad \forall g,b,i
\end{cases}
\end{aligned}
\tag{5.7}
$$

where the successful SIC decoders are assured by constraints $C_4$, instead of previously supposing the perfect SIC decoders. Moreover, $f_{b,g}^{(i)}$ can be an arbitrary linear equalizer, instead of a linear MMSE equalizer.

**Proposition 1** *For my system setup, there are identical solutions* $\left\{p_{b,g}^{(i)}\right\}$ *between the WSRM problem* $\mathcal{P}1$ *and a suitably defined WMMSE problem* $\mathcal{P}2$.

*Proof:* The proof of **Proposition 1** is provided in Appendix A.

### 5.3.2 Distributed Power Resource Allocation Optimization

Here, I present a distributed optimization-processing protocol to calculate an optimal and flexible payload power resource-allocation solution based on the equivalent WMMSE problem $\mathcal{P}2$ as previously derived. From the perspective of mathematical optimization, I can find a convex-cost function in $\mathcal{P}2$ over every block of the optimization parameters $\left\{\varpi_{b,g}^{(i)}\right\}$, $\left\{p_{b,g}^{(i)}\right\}$ and $\left\{f_{b,g}^{(i)}\right\}$. Thus, $\mathcal{P}2$ can be solved via using the block-coordinate descent method [98], which is a basic optimization framework for large-dimension optimization problems [96, 97]. In particular, by first fixing $\left\{\varpi_{b,g}^{(i)}, p_{b,g}^{(i)}\right\}$ in $\mathcal{P}2$, $f_{b,g}^{(i)}$ and $\epsilon_{b,g}^{(i)}$ are updated as (5.4) and (5.5), respectively. Next, by fixing $\{p_{b,g}^{(i)}, f_{b,g}^{(i)}\}$, each user can independently solve

$$\min_{\varpi_{b,g}^{(i)}} \varsigma_{b,g}^{(i)}(\varpi_{b,g}^{(i)}\epsilon_{b,g}^{(i)} - \log_2(\varpi_{b,g}^{(i)})), \ \forall g, b, i \tag{5.8}$$

and update

$$\varpi_{b,g}^{(i)} = \frac{1}{(\epsilon_{b,g}^{(i)})_{\text{MMSE}}^{\min}}, \ \forall g, b, i \tag{5.9}$$

which is from plugging $(\epsilon_{b,g}^{(i)})^*$ as (A.4) into (A.1). $\left\{p_{b,g}^{(i)}\right\}$ is updated by solving $\mathcal{P}2$, while keeping the the user variables $\{\varpi_{b,g}^{(i)}, f_{b,g}^{(i)}\}$ fixed. The problem $\mathcal{P}2$ under practical satellite constraints and successful SIC decoding constraints decouples over the GWs and is resolved by employing the interior-point algorithm [93]. This iterative process is terminated when a maximum number of specified iterations is met.

More specifically, the most demanding step within the above iterative process (i.e., Step 4 of Algorithm 2 summarized later) is to optimize the payload power resource allocation $\{p_{b,g}^{(i)}\}$, where the considered interior-point optimization algorithm numerically solves the Karush-Kuhn-Tucker (KKT) conditions [93] for the large-dimension optimization problem $\mathcal{P}2$ with multiple constraints (e.g., practical PAPC on payload and successful SIC decoding constraints). To accomplish the above mathematical optimization task, this work generalizes the Lagrange multipliers method [96], [97] (i.e., only counting one equality constraint such as the total power budget constraint) to the KKT method [93], allowing multiple inequality constraints in the mathematical optimization model. As numerous simulation software tools (e.g., Mat-

lab and Mathematica) accommodate the interior-point optimization algorithm and its variants, this work directly implores the interior-point algorithm function within Optimization Toolbox$^{\text{TM}}$ of Matlab to facilitate the acquisition of the optimal $\{p_{b,g}^{(i)}\}$. Furthermore, a stationary point can be found through using block-coordinate descent over the blocks of variables $\{\varpi_{b,g}^{(i)}\}$, $\{p_{b,g}^{(i)}\}$ and $\{f_{b,g}^{(i)}\}$ [97]. Due to the equivalence of the problem $\mathcal{P}2$ and $\mathcal{P}1$ established by **Proposition 1**, the block-coordinate descent of $\mathcal{P}2$ also obtains the convergence for a steady point of $\mathcal{P}1$ [99]. Accordingly, the distributed power resource-allocation-optimization protocol can be summarized as Algorithm 2.

**Algorithm 2: The WMMSE-based PA.**

**Step 1. On-ground GW** $g$ $(\forall g)$ **initializes** $p_{b,g}^{(i)}$ $(\forall g, b, i)$**:**

$$\frac{p_{b,g}^{(i)}}{p_{b,g}^{(i-1)}} = \left( \frac{\left|\mathbf{h}_{b,g}^{(1)}\mathbf{w}_b\right|^2}{\left|\mathbf{h}_{b,g}^{(1)}\mathbf{w}_{b,g}\right|^2 - \left|\mathbf{h}_{b,g}^{(i)}\mathbf{w}_{b,g}\right|^2} \right)^{i-1}, \ \forall g, b, i$$

**Repeat**

**Step 2. On-ground user** $i$ $(\forall i)$ **in beam** $b \in \mathcal{N}_g$ $(\forall g, b)$ **optimizes its** $f_{b,g}^{(i)}$ $(\forall g, b, i)$**:** for given $p_{b,g}^{(i)}$ $(\forall g, b, i)$, $f_{b,g}^{(i)}$ is separately optimized as $f_{b,g}^{(i)} = (f_{b,g}^{(i)})_{\text{MMSE}}^{\text{opt}}, \forall g, b, i$

**Step 3. On-ground user** $i$ $(\forall i)$ **in beam** $b \in \mathcal{N}_g$ $(\forall g, b)$ **optimizes its** $\varpi_{b,g}^{(i)}$ $(\forall g, b, i)$**:** for given $p_{b,g}^{(i)}, f_{b,g}^{(i)}$ $(\forall g, b, i)$, $\varpi_{b,g}^{(i)}$ is locally optimized as $\varpi_{b,g}^{(i)} = \frac{1}{(\epsilon_{b,g}^{(i)})_{\text{MMSE}}^{\min}}$ $(\forall g, b, i)$

**Step 4. On-ground GW** $g$ $(\forall g)$ **optimizes** $p_{b,g}^{(i)}$ $(\forall g, b, i)$**:** for given $\varpi_{b,g}^{(i)}, f_{b,g}^{(i)}$ $(\forall g, b, i)$, $p_{b,g}^{(i)}$ $(\forall g, b, i)$ is locally optimized at the GW $g$ $(\forall g)$ by solving $\mathcal{P}2$.

**Until** meeting a maximum number of iterations

Remarkably, during the iterative procedure, the information such as effective CSI and weights must be exchanged among the users and the GWs by employing the actual CSI acquisition mechanism, including effective channel estimation, feedback, and GW backhaul signaling. Thus, in addition to the payload power resource-allocation-optimization strategy, a relative CSI acquisition mechanism is another interesting research topic. As an extension of the mathematical optimization Algorithm 2, a tailored DNN architecture with a customized loss function is trained to intelligently allocate payload power resources among both the beams and users, by learning the undercover structure of its input (i.e., unsupervised learning). Since the DNN-based scheme doesn't need signaling and real-time information exchange between the GWs and the users, it can significantly decrease the implementation complexity by employ-

ing the channel statistics of users in multibeam SATCOM. Moreover, the DNN-based scheme can be trained as a universal approximator of the payload power resource-allocation agent for any unseen satellite channel and has the potential for a real-time operation with reduced implementation complexity, compared to the mathematical optimization-based scheme (i.e., the WMMSE-based PA in Algorithm 2).

## 5.4 DNN-based Solution

Typically, machine learning based on the neural network (NN) can be classified into supervised learning, unsupervised learning, and reinforcement learning (RL) [100]. Due to stagnant scheduling demand and the uncertainty of the actual traffic intrinsic in the multibeam SATCOM [101], the practical single-shot transmission policy (i.e., user scheduling, BF, and power resource-allocation strategies are separately processed with independent channel information at different periods) is concentrated. Thus, the RL techniques based on the Markov decision process (MDP) are not suitable, which generally require a correlation among multiple consecutive time slots. Additionally, regarding DNN-based power resource-allocation schemes, supervised learning-based DNN schemes [102,103] and unsupervised learning-based DNN schemes [104–106] are proposed with very similar performance and lower computational complexity compared to the original WMMSE scheme. However, existing DNN-based power allocation schemes from terrestrial systems don't focus on the NOMA system [102–106] and don't deal with the practical constraints on SATCOM such as i) the total payload power budget constraint; ii) PAPC. Thus, a machine learning-based flexible payload power resource-allocation scheme is investigated to obey these constraints in non-orthogonal SATCOM.

In the following, a tailored DNN architecture is first introduced to address problem $\mathcal{P}1$ for intelligently reconfiguring payload power resources in real-time, instead of a mathematical optimization-based scheme (i.e., the WMMSE-based payload PA in Algorithm 2). Then, the deep-learning protocol is presented including the data preprocessing, training, and testing phases.

### 5.4.1 Neural Network Structure

A fully-connected DNN with unsupervised learning is adopted to intelligently implement payload power-resource reconfiguration in real time. As shown in Figure 5-1, the whole NN architecture comprises two feedforward DNN modules: (1) a **B**eam-

level payload **p**ower resource-**a**llocation **Mod**ule (**BpaMod**); (2) a **U**ser-level payload **p**ower resource-**a**llocation **Mod**ule (**UpaMod**). Particularly, the **BpaMod** is composed of:

- One input layer including $2N_u \cdot N^2$ neurons;
- $L-1$ hidden layers, where each layer $l$ has $S_l$ neurons, $l \in [1, L-1]$;
- One output layer consisting of $N$ neurons.

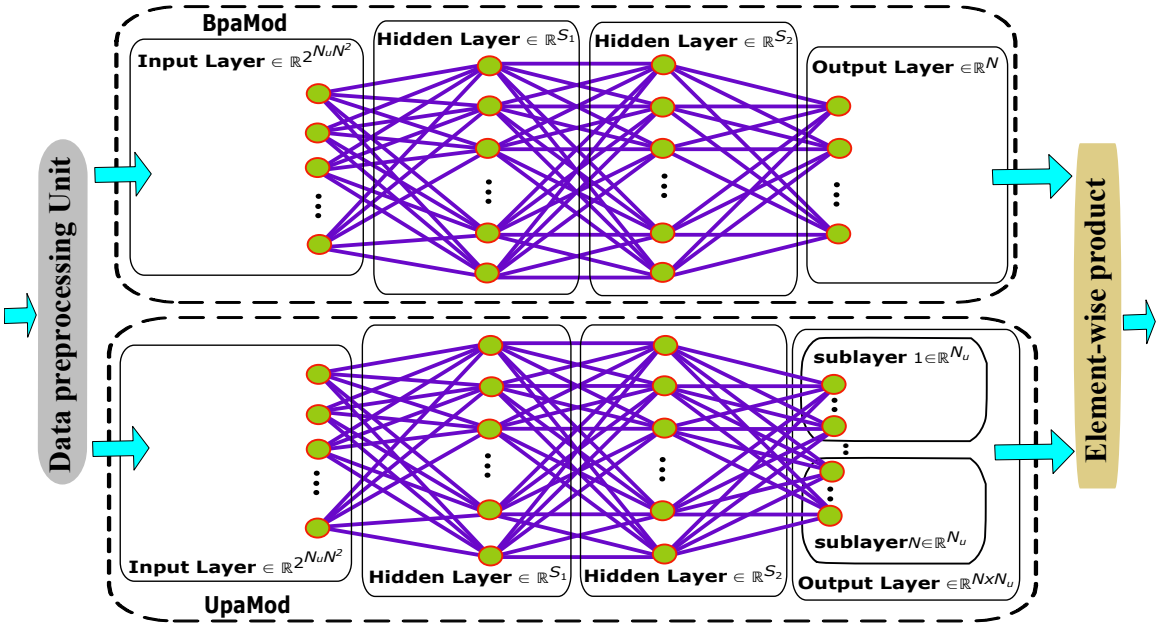


Figure 5-1: The exemplary four-layers DNN-based structure.

In the **BpaMod**, the multiple layers of neurons are labeled from 0 to $L$. Herein, the input layer of neurons receives input directly outside the network and its output vector $\mathbf{a}_0$ has the size $(2N_u \cdot N^2) \times 1$. Next, the output vector $\mathbf{a}_0$ of the input layer is fed to the first hidden layer. Each hidden layer possesses the rectified linear unit (ReLU) function $f_{\text{Relu}}(x) = max(x, 0)$ as the activation function of each neuron, which provides the desired nonlinearity to the neural network for better performance. Then, the abbreviated neuron output vector $\mathbf{a}_l \in \mathbb{R}^{S_l \times 1}$ of the $l$-th hidden layer is calculated as

$$\mathbf{a}_l = f_{\text{Relu}}(\mathbf{\Psi}_l \mathbf{a}_{l-1} + \mathbf{b}_l), \ l \in [1, L-1] \tag{5.10}$$

where $\mathbf{\Psi}_l \in \mathbb{R}^{S_l \times S_{l-1}}$ and $\mathbf{b}_l \in \mathbb{R}^{S_l \times 1}$ are the abbreviated neuron weight matrix and the abbreviated neuron bias vector of the $l$-th hidden layer, respectively. In general, the number of the inputs to the first hidden layer does not definitely match the number

of neurons of the first hidden layer (i.e., $S_0 \neq S_1$). Accordingly, $\mathbf{\Psi}_1 \in \mathbb{R}^{S_1 \times S_0}$ has the column dimension $S_0 = 2N_u \cdot N^2$, which is the length of the output from the input layer.

Then, the output layer acts as a function approximator and determines how to allocate the total payload power budget between multiple beams (i.e., it is named as the beam-level payload power resource-allocation in this work). Herein, the linear transfer function $f_{\text{Relu}}(x)$ is used as the activation function of each neuron for the output layer. In abbreviated notation, the output vector $\mathbf{a}_L \in \mathbb{R}^{N \times 1}$ (corresponding to the deep-learning beam-level payload power resource-allocation result) of the output layer is computed as

$$\mathbf{a}_L = f_{\text{Relu}}(\mathbf{\Psi}_L \mathbf{a}_{L-1} + \mathbf{b}_L) \tag{5.11}$$

Remarkably, the total payload transmit budget-constraint on the output $\mathbf{a}_L$ can be automatically met by adding the penalty function of the corresponding constraint violation to the customized loss function. This is different from both the work of Sanguinetti *et al.* [102] and the work of Sun *et al.* [103], where (1) the linear transfer function $f_{\text{Linear}}(x) = x$ is adopted as the activation function of each neuron for the output layer; (2) the total power budget constraint is fulfilled by either additional DNN structure design [102] or the postprocessing [103].

In parallel, by feeding the same input to another concurrent DNN module **UpaMod**, I aim to assign a given beam-level payload power resource among all selected users per beam (i.e., it is termed as the user-level payload power resource-allocation in this work). To achieve this goal, the same network structure as the parallel **BpaMod** can be adopted to the **UpaMod** except for the output layer. In particular, the output layer has $N$ number of sublayers. Each sublayer of the output layer has $N_u$ number of neurons and employs the softmax transfer function as its activation function; that is: $f_{\text{Softmax}}(x_i) = \frac{e^{x_i}}{\sum_{j=1}^{N_u} e^{x_j}}$, $\forall i$, where $(x_1, x_2, \cdots x_{N_u})$ is an input vector to a softmax transfer function. Considering that a softmax transfer function outputs non-integer probabilities (between 0 and 1) and the sum of the output probabilities is 1, the factor of the user-level payload power resource-allocation can be well mimicked by a softmax transfer function. Accordingly, in the **UpaMod**, the abbreviated neuron output vector $\tilde{\mathbf{a}}_L^{(n)} \in \mathbb{R}^{N_u \times 1}$ of each sublayer $n \in [1, N]$ belonging to the output layer is given as

$$\tilde{\mathbf{a}}_L^{(n)} = f_{\text{Softmax}}(\tilde{\mathbf{\Psi}}_L^{(n)} \tilde{\mathbf{a}}_{L-1} + \tilde{\mathbf{b}}_L^{(n)}), \ \forall n \in [1, N] \tag{5.12}$$

where $\tilde{\mathbf{\Psi}}_L^{(n)} \in \mathbb{R}^{S_L^{(n)} \times S_{L-1}}$ and $\tilde{\mathbf{b}}_L^{(n)} \in \mathbb{R}^{S_L^{(n)} \times 1}$ are the abbreviated neuron weight matrix

and the abbreviated neuron bias vector of each sublayer $n$ belonging to the output layer. The output layer has $S_L = \sum_{n=1}^{N} S_L^{(n)} = N \cdot N_u$ neurons, since each sublayer $n$ has $S_L^{(n)} = N_u$ neurons. Subsequently, a set of sublayer's output vector $\{\tilde{\mathbf{a}}_L^{(n)} \in \mathbb{R}^{N_u \times 1}\}_{n=N}$ is converted to an output matrix $\tilde{\mathbf{A}}_L$ with the size $N \times N_u$ (corresponding to the deep-learning user-level payload power resource-allocation result), which is the output of the **UpaMod**.

Eventually, by conducting the element-wise product of the output of the **BpaMod** and the output of the **UpaMod**, the payload power resource allocation among both the beams and users (i.e., $\left\{p_{b,g}^{(i)}\right\}$) can be captured in abbreviated notation as $\mathbf{a}_L \odot \tilde{\mathbf{A}}_L$.

### 5.4.2 Data Preprocessing

During this phase, a practical 96-beams satellite is first realized, serving the geographical region of Europe [101]. A subgroup $N = 21$ from all 96 beams is considered. The system has $G = 3$ GWs, each of which controls a group of $N_g = 7,\ g \in \{1, 2, 3\}$ beams. The users are uniformly distributed within the footprints and are scheduled by following the strategy in Chapter 3. Then, the entire channel matrix $\mathbf{H} \in \mathbb{C}^{N \times NN_u}$ for the selected users is generated, where each channel vector $\mathbf{h}_{b,g}^{(i)}$ of $\mathbf{H}$ is already given by (2.28). Moreover, the aggregated BF matrix $\mathbf{W}_{\text{final}} \in \mathbb{C}^{N \times NN_u}$ for the selected users is created with the BF strategy in Section 4.3, where the $N_u$ users within beam $i,\ \forall i \in [1, N]$ share the same BF vector $i$ of the full BF matrix $\mathbf{W}_{\text{full}} \in \mathbb{C}^{N \times N}$.

Since the training algorithm learns representations from different sampled data, the output of the neural networks could be determined by some sampled data with significant amplitude, instead of all sampled data. Therefore, the sampled channel matrices and BF matrices are preprocessed with the standardization operation before the training, which helps to converge the training process of the DNN. Accordingly, the training data set $\{\mathbf{H}^{(j)}\}_{j \in \mathcal{T}}$ and $\{\mathbf{W}_{\text{final}}^{(j)}\}_{j \in \mathcal{T}}$ are first converted to be in the dB domain and sequentially standardized to have zero mean and unit variance in the dB domain, where $\mathbf{H}^{(j)}$ and $\mathbf{W}_{\text{final}}^{(j)}$ are the $j$-th training sample and $\mathcal{T}$ represents the training set. During the standardization process of $\{\mathbf{W}_{\text{final}}^{(j)}\}_{j \in \mathcal{T}}$, $log(10^{-8})$ is used to dismiss the $log(0)$ problem. Afterward, the training data sets $\{\mathbf{H}^{(j)}\}_{j \in \mathcal{T}}$ and $\{\mathbf{W}_{\text{final}}^{(j)}\}_{j \in \mathcal{T}}$ are vertically stacked into a concatenated data set $\{\begin{bmatrix} \mathbf{H}^{(j)} \\ \mathbf{W}_{\text{final}}^{(j)} \end{bmatrix}\}_{j \in \mathcal{T}}$, where each training sample $\begin{bmatrix} \mathbf{H}^{(j)} \\ \mathbf{W}_{\text{final}}^{(j)} \end{bmatrix},\ \forall j$ is sequentially reshaped to a column vector as the input of the DNN.

### 5.4.3 Training and Testing

Without labeling ground truth, the neuron weights and the neuron biases in the applied DNN are first tuned by learning the undercover structure of its input. After the neuron weights and the neuron biases are optimized, the trained DNN is used to intelligently predict a payload power resource-allocation strategy with the testing data set.

***Network Training:*** During the training process, the neuron weight matrices and neuron bias vectors of the applied DNN need to be well tuned to optimize the DNN performance, as defined by the network performance function (i.e., it is termed as the loss function in this work). To this end, a loss function is customized for the applied DNN and derived by simultaneously considering the objective function and all the constraints of the optimization problem $\mathcal{P}1$. Mathematically, the customized loss function of the applied DNN is formulated as

$$\begin{aligned}
\text{Loss} = & -\sum_{g\in\mathcal{G}}\sum_{b\in\mathcal{N}_g}\sum_{i\in\mathcal{S}_{b,g}} \varsigma_{b,g}^{(i)} R_{b,g}^{(i)} \\
& + \lambda_1 \sum_{g\in\mathcal{G}}\sum_{b\in\mathcal{N}_g}\sum_{i\in\mathcal{S}_{b,g}} f_{\tanh}\left(\frac{f_{\text{Relu}}(\bar{R}_{b,g}^{(i)} - R_{b,g}^{(i)})}{\bar{R}_{b,g}^{(i)} + \tau}\right) \\
& + \lambda_2 f_{\tanh}\left(\frac{f_{\text{Relu}}(\sum_{g\in\mathcal{G}}\sum_{b\in\mathcal{N}_g}\sum_{i\in\mathcal{S}_{b,g}} p_{b,g}^{(i)} - P_{\text{total}})}{P_{\text{total}} + \tau}\right) \\
& + \lambda_3 \sum_{g\in\mathcal{G}}\sum_{m\in\mathcal{N}_g}\sum_{b\in\mathcal{N}_g}\sum_{i\in\mathcal{S}_{b,g}} f_{\tanh}\left(\frac{f_{\text{Relu}}(\text{PAPC}_{b,g}^{(i),m})}{P_{g,m} + \tau}\right) \\
& + \lambda_4 \sum_{g\in\mathcal{G}}\sum_{b\in\mathcal{N}_g}\sum_{i\in\mathcal{S}_{b,g}}\sum_{k\in[1,i)} f_{\tanh}\left(\frac{f_{\text{Relu}}(-\text{SIC}_{b,g}^{(i)\to(k)})}{\tau}\right)
\end{aligned} \tag{5.13}$$

where the objective function of the corresponding optimization problem $\mathcal{P}1$ is set to its negative counterpart since the loss function is to be minimized by the training optimizer. Other penalty functions are also added to the objective function. Each penalty function comprises a penalty coefficient $\lambda_i$ ($i \in [1, 4]$) multiplied by an estimate of the violation of the corresponding constraint (i.e., $C_i$ ($i \in [1, 4]$) of the optimization problem $\mathcal{P}1$). When the constraint $C_i$ ($i \in [1, 4]$) is violated, the corresponding penalty function is increased such that the neuron weights and neuron biases of the training DNN are enforced to be tuned in the right track where the constraint is met. The auxiliary $\text{PAPC}_{b,g}^{(i),m} = p_{b,g}^{(i)} \wp_m^T \mathbf{w}_{b,g} \mathbf{w}_{b,g}^H \wp_m - P_{g,m}$ is introduced only

for the ease of notation. $f_{\tanh}(\cdot)$ denotes a Tanh function as $f_{\tanh}(x) = \frac{e^{x}-e^{-x}}{e^{x}+e^{-x}}$, which maps a penalty function value as a strongly positive value. The constant $\tau = 10^{-8}$ is adopted to eliminate the zero denominator problem.

After settling the customized loss function, the applied DNN is ready to be trained to achieve the desired network performance (i.e., (5.13)). To train the applied DNN, the off-the-shelf numerical optimization algorithms (e.g., either gradient- or Jacobian-based algorithms) are available to optimize the performance of the applied DNN. In this work, the gradient descent algorithm is adopted as the training algorithm or optimizer for training the applied DNN.

Herein, the appropriate batch size and learning rate are chosen by rotation estimation. Based on the gradient descent algorithm, the neuron weight matrices and the neuron bias vectors are updated per iteration in the direction where the customized loss function rapidly decreases (i.e., the opposite direction of the gradient[1]). Ultimately, the DNN after training is ready to deal with minimizing the loss function (5.13) on an unseen data set.

***Network Testing:*** Once the neuron weights and the neuron biases are well-tuned, the proper payload power resource allocation for both the users and the beams is intelligently approximated by feeding the current channel matrices and BF vectors (i.e., the testing data set) to the trained DNN, where the testing data set is preprocessed in the same manner as described in Section 5.4.2.

## 5.5 Complexity Analysis

In the following, I first analyze the computational complexity of the WMMSE-based payload PA scheme as:

- *In Step 1:* PA initialization for all the users has the complexity of $\mathcal{O}(N_u N)$.
- *In Step 2:* During each iteration, calculating $f_{b,g}^{(i)}$ at user $i$ in beam $b$ ($b \in \mathcal{N}_g$) requires the complexity $\mathcal{O}(1)$. Accordingly, the demanded complexity for all the users is $\mathcal{O}(N_u N)$.
- *In Step 3:* During each iteration, the complexity for computing $\varpi_{b,g}^{(i)}$ at user $i$ in beam $b$ ($b \in \mathcal{N}_g$) is $\mathcal{O}(1)$. Correspondingly, the needed complexity for all the users is $\mathcal{O}(N_u N)$.

[1]The gradient can be calculated with a back-propagation method. Note that the back-propagation method occasionally refers to the gradient descent algorithm in the context of DNN training [105].

- *In Step 4:* During each iteration, following the assessment method in [90], the interior-point method needs a worst-case complexity of $\mathcal{O}((N_u N)^{4.5} \log \frac{1}{\varepsilon_3})$ to seek the $\varepsilon_3$-optimal solution of $\mathcal{P}2$.

Therefore, the total computational complexity of Algorithm 2 is $\mathcal{O}(N_u N + I_{\text{PA}}(2N_u N + (N_u N)^{4.5} \log \frac{1}{\varepsilon_3})) \sim \mathcal{O}(I_{\text{PA}}(N_u N)^{4.5} \log \frac{1}{\varepsilon_3})$, where $I_{\text{PA}}$ is the maximum number of iterations required in Algorithm 2.

Next, the computational complexity of the DNN-based payload PA scheme is analyzed as:

- *In the phases of data preprocessing:* the total computation complexity of the data preprocessing for both training and testing is $\mathcal{O}(2N_u N^2)$.

- *In the phases of training and testing:* the computation of **BpaMod** in the phases of training and testing is dominated by the matrix computation; that is: $\mathcal{O}(2N_u N^2 S_l + 2S_l^2(L-1) + 2S_l N)$, where the computation complexity of training and testing is treated as the same. Similarly, the computation of **UpaMod** in the phases of training and testing is calculated as $\mathcal{O}(2N_u N^2 S_l + 2S_l^2(L-1) + 2S_l N_u)$. Next, the element-wise product is conducted for the output of the **BpaMod** and the output of the **UpaMod**, which has the complexity of $\mathcal{O}(N_u N)$. Eventually, the computation in phases of training and testing is given as $\mathcal{O}(4N_u N^2 S_l + 4S_l^2(L-1) + 2S_l(N_u + N) + N_u N)$.

Hence, the total computational complexity of the DNN-based payload PA scheme is $\mathcal{O}(4N_u N^2 S_l + 4S_l^2(L-1) + 2S_l(N_u + N) + N_u N + 2N_u N^2) \sim \mathcal{O}(4N_u N^2 S_l + 4S_l^2 L)$.

## 5.6 Performance Evaluation

Based on the same simulation model, this part manifests the performance achieved by the flexible payload power-resource-optimization strategies through numerical simulations. Herein, the performance for the following schemes is estimated.

- "**WMMSE-based payload PA**", where $\mathbf{w}_{b,g}$ from Section 4.3 is used. The WMMSE-based PA from Section 5.3 and the user scheduling from Chapter 3 are performed.

- "**DNN-based payload PA**", where $\mathbf{w}_{b,g}$ from Section 4.3 is used. The DNN-based PA from Section 5.4 and the user scheduling from Chapter 3 are performed.

- "**Static payload PA**", where $\mathbf{w}_{b,g}$ from Section 4.3 is adopted. The static PA and the user scheduling from Chapter 3 are conducted.

The performance of different schemes is assessed by conducting numerical simulations. The simulation platform is set up with a 2-core central processing unit (CPU) (i.e., Intel(R) Core(TM) i7 − 7500U) and a running frequency 2.70 GHz. The applied DNN model is programmed, trained, and tested with Python 3.7 and TensorFlow. The interior-point algorithm function within Optimization Toolbox$^{\text{TM}}$ of Matlab is employed to obtain the optimal payload power resource-reconfiguration $\{p_{b,g}^{(i)}\}$ of problem $\mathcal{P}1$. Herein, I consider the same parameters as summarized in TABLE 3.1. In principle, a feedforward DNN with more layers generally approximates an unknown nonlinear mapping more quickly. Accordingly, this work starts with the three-layers DNN and further expands the DNN to four layers unless the performance of the three-layers DNN is not satisfactory. Thus, a four-layers DNN is determined.

As the performance assessment metric, the cumulative distribution function (CDF) of the spectrum efficiency per beam is considered for different schemes since the WSR is the optimization objective. Herein, the spectrum efficiency per beam is already defined as the equation (3.6) in 3.7.3. Without a loss of generality, the predefined minimum user rate per beam $\bar{R}_{b,g}^{(i)}$ is the data rate achieved by an orthogonal transmission scheme. The relative user priorities are $\varsigma_{b,g}^{(1)} = 0.5$, $\varsigma_{b,g}^{(2)} = 0.9$ for $N_u = 2$. The total payload power budget is set as $P_{total} = \{25,\ 30\}$ dBW. It is illustrated in Figures 5-2 and 5-3 that the DNN-based payload PA scheme can achieve satisfactory performance compared to the WMMSE-based payload PA scheme [101], where the DNN-based payload PA scheme is trained with penalty coefficients $\lambda_1 = 1.72$, $\lambda_2 = 1.4$, $\lambda_3 = 0.7$, $\lambda_4 = 0.55$ and $\lambda_1 = 0.35$, $\lambda_2 = 0.45$, $\lambda_3 = 0.55$, $\lambda_4 = 0.35$ for $P_{total} = 25$ dBW and $P_{total} = 30$ dBW, respectively. Specifically, it is observed from Figures 5-2 and 5-3 that the DNN-based payload PA scheme and the WMMSE-based payload PA scheme [101] have nearly identical perform on average, while the results for the DNN-based payload PA scheme have a slightly larger spreading. Moreover, as anticipated, both the DNN-based payload PA and WMMSE-based payload PA schemes are superior to the static payload PA scheme. Furthermore, it is implied from Figures 5-2 and 5-3 that the spectrum efficiency performance for all the schemes is increased as the the total payload power budget $P_{total}$ is raised, which well matches the behavior (i.e., Figure 4-2) of the WMMSE-based payload PA scheme. Additionally, it is worthy mentioning from Figures 5-2 and 5-3 that the DNN-based payload PA scheme performs slightly better than the WMMSE-based payload PA scheme in some simulation runs. It is explained by the fact that since my network

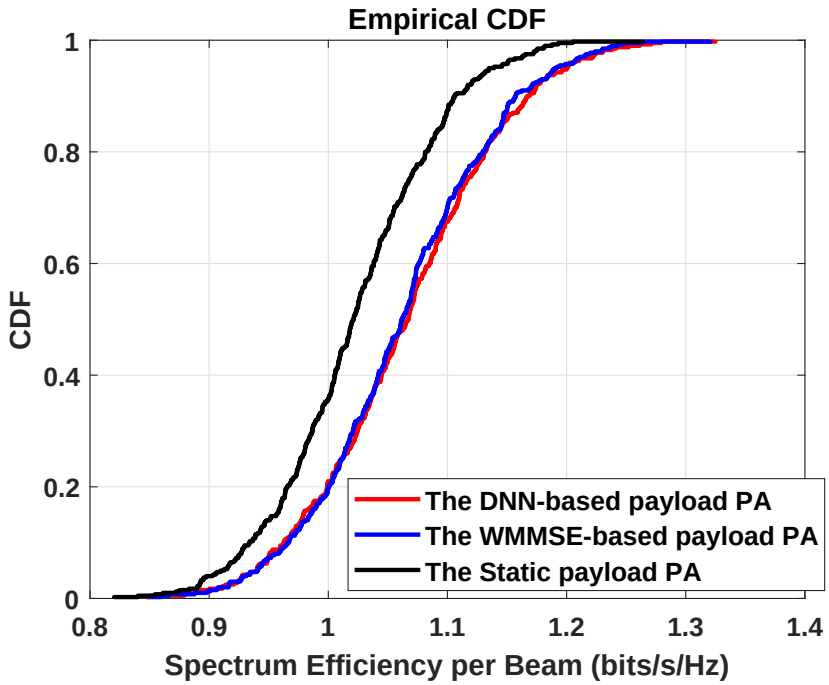


Figure 5-2: CDF of the spectrum efficiency per beam for the different schemes with $P_{\text{Total}} =$ 25 dBW.

is trained via a unsupervised-learning method, the global optima solutions can be learned by DNN-based payload PA scheme in some simulation runs [106]. However, from the mathematic optimization perspective, the local optima solutions are always determined by the WMMSE-based payload PA scheme [96, 97].

Next, each penalty coefficient $\lambda_i$ ($i \in [1, 4]$) weighs the corresponding penalty term in the loss function and is limited exclusively to be tuned based on the numerical results [106]. Thus, the influence of each penalty coefficient $\lambda_i$ ($i \in [1, 4]$) on the performance is presented, where the spectrum efficiency performance of the DNN-based payload PA scheme is averaged over 400 channel realizations. Herein, the exemplary use cases are described as follows: (1) $P_{total} = 30$ dBW; (2) one penalty coefficient $\lambda_i$ ($i \in [1, 4]$) is varying in the range $0 \sim 1$, while the remaining penalty coefficients are correspondingly chose from $\{\lambda_1 = 0.35,\ \lambda_2 = 0.45,\ \lambda_3 = 0.55,\ \lambda_4 = 0.35\}$. It is shown from Figure 5-4 that the performance of the DNN-based payload PA scheme is degrading when one penalty coefficient $\lambda_i$ ($i \in [1, 3]$) is increasing and others are fixed. The reason is that the training optimizer makes more effects on the corresponding constraint $C_i$ ($i \in [1, 3]$), with the expense of sacrificing the spectrum efficiency performance. Thus, it infers that each penalty coefficient $\lambda_i$ ($i \in [1, 3]$) has a noteworthy influence on the DNN training and can be flexibly chosen with the desired target

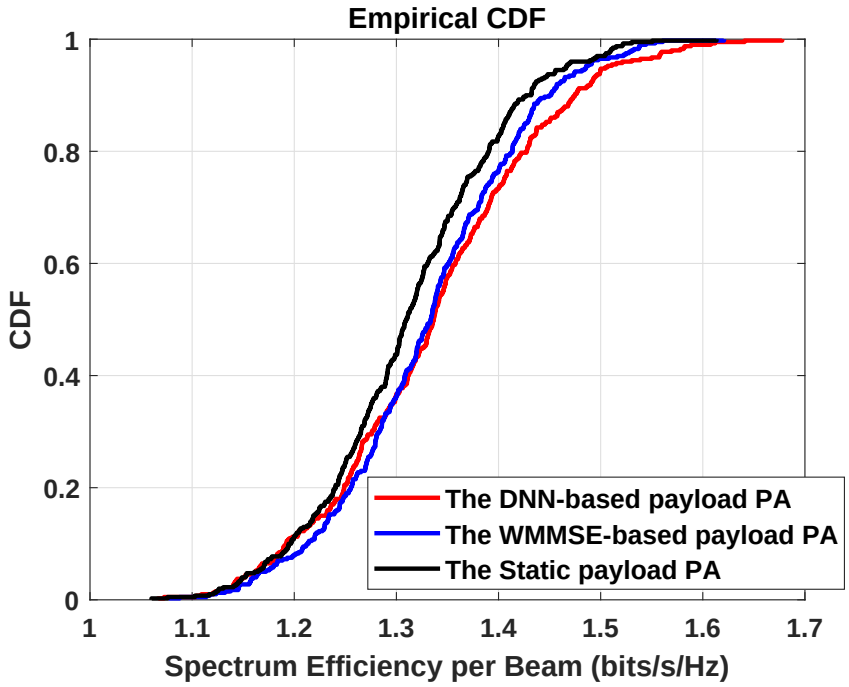


Figure 5-3: CDF of the spectrum efficiency per beam for the different schemes with $P_{\text{Total}} =$ 30 dBW.

performance. Furthermore, the varied $\lambda_4$ doesn't change the performance of the DNN-based payload PA scheme. It can be explained from the equation (2.25) that the SIC constraint $C_4$ is met in the case of $|\mathbf{h}_{b,g}^{(k)}\mathbf{w}_{b,g}|^2 \geq |\mathbf{h}_{b,g}^{(i)}\mathbf{w}_{b,g}|^2,\ 1 \leq k < i, \forall g, b, i$, which is irrelative to the allocated power and is already achieved by the predefined user decoding order. In inference-limited SATCOM systems, the predefined user decoding order can be only captured via scheduling highly-correlated users and designing effective BF (i.e., when the interference is effectively suppressed) [101]. Remarkably, even though $\lambda_4$ does not influence the PA solution (or the resulting performance of the DNN), it is a necessary constraint for optimization problem $\mathcal{P}1$ (or loss function (5.13)) since the objective function supposes a successful SIC decoding. In practice, without the SIC constraint $C_4$, random user channel and inefficient BF matrices could be used as the input of the DNN (or problem $\mathcal{P}1$) to produce fake and unrealistic payload power resource-allocation solutions. Additionally, although Figure 5-4 shows that $\lambda_1$ and $\lambda_3$ should be set to zero for maximum WSR, it cannot generate a realistic payload power resource-allocation solution since constraints on the demanded user rate (i.e., $C_1$) and the PAPC (i.e., $C_3$) are not met.

Last but not least, for the computation time, the DNN-based payload PA scheme takes less than 6 minutes on the CPU-based platform to complete the protocol in-

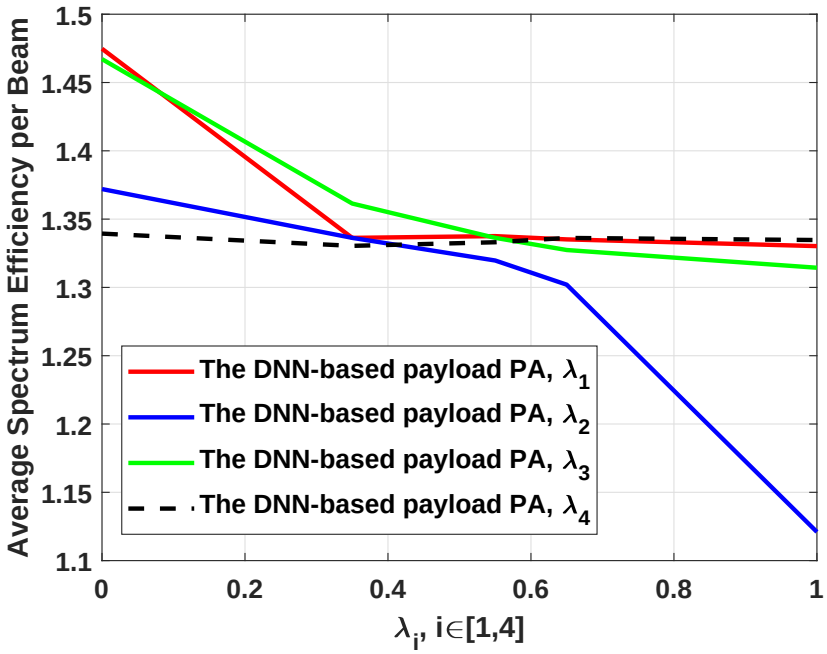


Figure 5-4: Performance influence of the penalty coefficient $\lambda_i$.

cluding the data-preprocessing, training, and testing over the given training and testing data sets. Thus far, the DNN-based payload PA scheme is simulated with the CPU-based platform and the full CSI. In practice, the training of DNN is assumed to be conducted online by utilizing the potential of unsupervised learning, where a high-performance graphical processing unit (GPU) with multiple running threads is employed to train the DNN at the resource control center of each GW. Thus, the computation time of the DNN-based scheme is to be rapidly accelerated. It implies that the DNN-based payload PA scheme has the potential to accomplish a real-time operation with online training and reduced implementation complexity since the real-time information exchange between the GWs and the users is completely avoided.

## 5.7 Summary

The payload PA-optimization strategies are proposed to satisfy the actual traffic demand. The system throughput performance for the payload PA-optimization strategies is evaluated. The key conclusions are given for this chapter in the following:

- The mathematical optimization-based (i.e., the WMMSE-bassed) payload PA scheme is developed through rigorous mathematical derivations.

- The DNN-based payload PA scheme can be trained as a universal approximator of the payload PA agent for any unknown satellite channel and has the potential for a real-time operation without signaling and real-time information exchange between the GWs and the users.

The simulation results imply that the payload PA-optimization strategies can flexibly reconfigure onboard power resources among both the users and beams to meet actual traffic demand, while achieving better throughput performance compared to the static payload PA scheme.

# Chapter 6

# Conclusions and Future Research

## 6.1 Conclusions

This Ph.D. dissertation has given new insights into the application and integration of NOMA techniques to BF SATCOM. For the overloaded multibeam SATCOM, NOMA is a candidate transmission technology to improve system throughput and user connectivity, due to its high compatibility. Hence, the NOMA transmission technique can be integrated and complied with the existing BF SATCOM system such that conventional infrastructures and the original capacity of the existing BF satellite are preserved, while spectrum efficiency and the amount of connectivity are improved via NOMA. From an information theoretic perspective, NOMA is not always capacity-achieving for BF-based FWD links. Thus, to effectively incorporate NOMA into BF-based SATCOM, this Ph.D. dissertation has identified and addressed the crucial obstacles to incorporate NOMA with BF-based SATCOM. In particular, the signal processing and optimization techniques are proposed in the form of the distributed resource optimization framework.

- Chapter 2 develops mathematical modeling to regenerate an exemplary multi-GW multibeam SATCOM scenario, including the near-to-real footprints, the generation of users, and interference, which is further applied to design the distributed resource optimization strategies in Chapters 3, 4 and 5.

- Chapter 3 proposes a distributed user scheduling with the customized beam association and K-means methods, which is conducted locally at each decentralized GW to substantially improve the system throughput. Herein, the performance benefit comes from 1) the scheduled users with higher channel correlation ($\delta^0 \geq 0.94$) are achieved by adopting the highest gain-based beam association,

compared to the counterpart by using random (or no) beam association. As a result, the BF strategy in Chapter 4 can efficiently manage intergateway and interbeam interference, given the high user channel correlation ($\delta^0 \geq 0.94$); 2) on the other hand, the underlying difference of the users' channel gains is exploited by the K-means algorithm in the last two phases of the user scheduling to unleash the full potential of NOMA.

- Chapter 4 optimizes the BF strategies under the practical SATCOM constraints, aiming to optimize user fairness. The proposed BF strategies are compatible with the existing beam-specific BF SATCOM. For the distributed BF strategy, the previous works do not investigate the user fairness optimality, along with other practical constraints on SATCOM, such as the PAPC on board. Unlike the previous works, a distributed PAPC constrained BF via maximization of the worst-user SLNR is mathematically derived based on SDR to be locally processed in geographically decentralized GWs. As a performance benchmark, a centralized multilayer BF via the SDR method is proposed to optimize the worst-user SNR after interbeam interference suppression.

- Chapter 5 proceeds to consider the additional combination of flexible payload power resources-allocation strategies to meet the actual traffic demand. Regarding payload power resources optimization, there is still a lack of a suitable payload power resources-allocation optimization solution for nonorthogonal multi-GW SATCOM, since successful SIC decoding is not ensured in previous works. This investigation of flexible payload PA-optimization strategies by formulating a WSRM optimization problem under practical SATCOM and successful SIC decoding constraints provides new insights.

## 6.2 Future Research

As the investigation of next-generation multiple access for SATCOM is in the technology readiness stage, several possible research topics are to be further addressed for applying and integrating NOMA in practical and flexible SATCOM deployments. The future work is summarized as:

- an intelligent user scheduling could employ machine learning to emulate and approximate the functions of the customized beam association and K-means methods, by training a DNN with the geographical locations of on-ground users, beam centers, and satellite.

- the distributed BF strategy should be further investigated with the channel statistics (i.e., $\mathbf{h}_{b,g}^{(i),H}\mathbf{h}_{b,g}^{(i)} \approx \mathbb{E}[\mathbf{h}_{b,g}^{(i),H}\mathbf{h}_{b,g}^{(i)}], \forall g, b, i$ can be adopted in the optimization problem (4.17) for the autonomous operation of the GWs in a multibeam SATCOM system). Furthermore, a machine learning-based BF optimization strategy is worth studying for an intelligent beam optimization operation.

- The DNN-based payload PA strategy needs to further optimize the DNN structure for better performance. Thus far, the DNN is separately trained for each specified value of the total payload power budget. A universal DNN should be further investigated and trained for any value of the total payload power budget.

# Appendix A

# The proof for Proposition 1

From (5.7), a convex cost function is perceived over $\varpi_{b,g}^{(i)}$. Thus, I inspect the first order necessary condition of the cost function in (5.7) for $\varpi_{b,g}^{(i)}$ to find the optimal $\varpi_{b,g}^{(i)}$ (expressed as $(\varpi_{b,g}^{(i)})^*$)

$$(\varpi_{b,g}^{(i)})^* = (\epsilon_{b,g}^{(i)})^{-1}, \ \forall g, b, i \tag{A.1}$$

Plugging these optimal values in (5.7) gives the following equivalent problem

$$\begin{aligned} &\min_{\left\{p_{b,g}^{(i)}, f_{b,g}^{(i)}\right\}} \sum_{g\in\mathcal{G}} \sum_{b\in\mathcal{N}_g} \sum_{i\in\mathcal{S}_{b,g}} \varsigma_{b,g}^{i}(\log_2(\epsilon_{b,g}^{i})) \\ &\text{s.t.} \begin{cases} C_1 : R_{b,g}^{(i)} \geq \bar{R}_{b,g}^{(i)}, \quad \forall g, b, i \\ C_2 : \sum_{g\in\mathcal{G}} \sum_{b\in\mathcal{N}_g} \sum_{i\in\mathcal{S}_{b,g}} p_{b,g}^{(i)} \leq P_{\text{total}} \\ C_3 : \sum_{b\in\mathcal{N}_g} \sum_{i\in\mathcal{S}_{b,g}} p_{b,g}^{(i)} \wp_m^T \mathbf{w}_{b,g} \mathbf{w}_{b,g}^H \wp_m \leq P_{g,m}, \quad \forall g, m \\ C_4 : \text{SIC}_{b,g}^{(i)\to(k)} \geq 0, \quad 1 \leq k < i, \quad \forall g, b, i \end{cases} \end{aligned} \tag{A.2}$$

Considering the convex cost function over $f_{b,g}^{(i)}$ in (A.2), the optimal $f_{b,g}^{(i)}$ (described as $(f_{b,g}^{(i)})^*$) is obtained via inspecting the first-order necessary condition for this cost function; that is:

$$(f_{b,g}^{(i)})^* = (f_{b,g}^{(i)})_{\text{MMSE}}^{\text{opt}}, \ \forall g, b, i \tag{A.3}$$

Accordingly, based on the equation (5.5), the optimal $\epsilon_{b,g}^{(i)}$ (denoted as $(\epsilon_{b,g}^{(i)})^*$) is

$$(\epsilon_{b,g}^{(i)})^* = (\epsilon_{b,g}^{(i)})_{\text{MMSE}}^{\min}, \ \forall g, b, i \tag{A.4}$$

Last, by substituting these optimum values $(f_{b,g}^{(i)})^*$ or equivalently $(\epsilon_{b,g}^{(i)})^*$ into (A.2) and using (5.6), the equivalence of problem $\mathcal{P}1$ and $\mathcal{P}2$ is obvious.

# Bibliography


[1] M. Giordani *et al.*, "Towards 6G Networks: Use Cases and Technologies," *IEEE Commun. Magazine*, vol. 58, no. 3, pp. 55-61, Mar. 2020. (Cited on page 1.)

[2] 3GPP T R 38.811, "Study on New Radio (NR) to Support Non - terrestrial Networks," V15.0.0, August 2018. Available at http://www.3gpp.org/ftp//Specs/archive/38_series/38.811/38811-f00.zip, [Online; Last accessed on 03-25-2019]. (Cited on page 1.)

[3] J. N. Pelton, "The start of commercial satellite communications [History of communications]," *IEEE Communications Magazine*, vol. 48, no. 3, pp. 24-31, Mar. 2010. (Cited on page 1.)

[4] G. Maral and M. Bousquet, "Satellite Communications Systems: Systems, Techniques and Technology," *Singapore, NJ: Wiley*, 2009. (Cited on pages 2, 3, 27, 28, 32 and 33.)

[5] E. Lutz *et al.*, "Satellite Systems for Personal and Broadband Communications," *Springer-Verlag*, 2000. (Cited on pages 2, 3 and 27.)

[6] O. Vidal, G. Verelst, J. Lacan, E. Alberty, J. Radzik, and M. Bousquet, "Next generation High Throughput Satellite system," in *2012 IEEE First AESS Eur. Conf. Satell. Telecommun. IEEE*, oct 2012, pp. 1-7. (Cited on page 3.)

[7] H. Fenech *et al.*, "Future High Throughput Satellite Systems," in *2012 IEEE First AESS Euro. Conf. Satellite Telecommun.*, Oct. 2012, pp 1-7. (Cited on page 3.)

[8] D. Serrano-Velarde *et al.*, "Novel Dimensioning Method for High-Throughput Satellites: Forward Link," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 50, no. 3, pp. 2146-2163, Jul. 2014. (Cited on page 3.)

[9] H. Fenech, S. Amos, A. Hirsch and V. Soumpholphakdv, "High throughput satellite systems: An analytical approach," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 51, no. 1, pp. 192-202, Jan. 2015 (Cited on page 3.)

[10] A. I. Pérez-Neira, M. Á. Vázquez, M. R. B. Shankar, S. Maleki, and S. Chatzinotas, "Signal processing for high-throughput satellites: Challenges in new interference-limited scenarios," *IEEE Signal Process. Mag.*, vol. 36, no. 4, pp. 112-131, Jul. 2019. (Cited on pages 5 and 7.)

[11] P. D. Arapoglou *et al.*, "DVB-S2X-enabled precoding for high throughput satellite systems," *Int. J. Satellite Commun. Netw.*, vol. 34, no. 3, pp. 439-455, Jun. 2015. (Cited on pages 5 and 6.)

[12] D. Christopoulos, S. Chatzinotas, and B. Ottersten, "Multicast multigroup precoding and user scheduling for frame-based satellite communications," *IEEE Trans. Wireless Commun.*, vol. 14, no. 9, pp. 4695-4707, Sep. 2015. (Cited on pages 5, 6, 49, 50 and 58.)

[13] P. Angeletti, G. Gallinaro, M. Lisi, and A. Vernucci, "Hybrid space/ground beamforming techniques for satellite telecommunications," in *32nd ESA Antenna Workshop on Antennas for Space Applications, Noordwijk*, The Netherlands, pp 5-8 Oct. 2010. (Cited on page 5.)

[14] C. Mosquera *et al.*, "Distributed Precoding Systems in Multi-Gateway Multibeam Satellites: Regularization and Coarse Beamforming," *IEEE Trans. on Wireless Commun.*, vol. 17, no. 10, pp. 6389-6403, Oct. 2018. (Cited on pages 5, 7, 8, 42, 57, 58, 61, 64, 65, 67, 68 and 70.)

[15] J. Tronc *et al.*, "Overview and comparison of on-ground and on-board beamforming techniques in mobile satellite service applications," *International Journal of Satellite Communications and Networking*, vol. 32, no. 4, pp. 291-308, 2014. (Cited on page 5.)

[16] P. Angeletti *et al.*, "Onground digital beamforming techniques for satellite smart antennas," in *32nd ESA Antenna Workshop on Antennas for Space Applications, Noordwijk*, The Netherlands, Oct. 2010, pp. 1-8. (Cited on page 5.)

[17] S. Chatzinotas, and B. Ottersten and R. D. Gaudenzi, "Cooperative and Cognitive Satellite Systems," *Elsevier*, May 2015. (Cited on page 6.)

[18] J. Arnau *et al.*, "Performance study of multiuser interference mitigation schemes for hybrid broadband multibeam satellite architectures," *EURASIP J. Wirel. Commun. Network*, vol. 2012, no. 132, pp. 1-19, 2012. (Cited on page 6.)

[19] B. Devillers, A. I. Pérez-Neira, and C. Mosquera, "Joint linear precoding and beamforming for the forward link of multi-beam broadband satellite systems," in *Proc. IEEE Global Telecommun. Conf. (GLOBECOM 2011)*, Dec. 2011, pp. 1-6. (Cited on page 6.)

[20] Y. Zhu *et al.*, "Low-complexity hybrid precoding with dynamic beam assignment in mmwave OFDM systems," *IEEE Trans. Veh. Technol.*, vol. 67, no. 4, pp. 3685-3689, Apr. 2018. (Cited on page 6.)

[21] E. Amyotte, Y. Demers, L. Hildebrand, M. Forest, S. Riendeau, S. Sierra-Garcia, J. Uher, "Recent developments in Ka-band satellite antennas for broadband communications," in *Proc. of Europ. Conf. on Antennas and Propagation (EuCAP), IEEE*, Apr. 2010, pp. 1-5. (Cited on page 6.)

[22] T. De Cola *et al.*, "Network and protocol architectures for future satellite systems," *Foundations and Trends R in Networking*, vol. 12, no. 1-2, pp. 1-161, 2017. (Cited on page 6.)

[23] D. L. Brandel, W. A. Watson, A. Weinberg, "NASA's advanced tracking and data relay satellite system for the years 2000 and beyond," *Proceedings of the IEEE*, Vol. 78, No. 7, pp 1141-1151, Jul. 1990. (Cited on page 5.)

[24] L. Cottatellucci *et al.*, "Interference mitigation techniques for broadband satellite systems," in *Proc. 24th AIAA Int. Commun. Satell. Syst. Conf., ICSSC 2006*, Jun. 2006, pp. 1-13. (Cited on pages 6, 16 and 39.)

[25] M. Diaz, N. Courville, C. Mosquera, G. Liva, and G. Corazza, "Nonlinear interference mitigation for broadband multimedia satellite systems," in *Proc. 2007 International Workshop on Satellite and Space Communications*, Sep. 2007, pp. 61-65. (Cited on page 6.)

[26] M. Poggioni, M. Berioli, and P. Banelli, "BER performance of multibeam satellite systems with Tomlinson-Harashima precoding," in *Proc. 2009 IEEE International Conference on Communications*, pp. 1-6. (Cited on page 6.)

[27] N. Zorba, M. Realp, A. I. Pérez-Neira, "An improved partial CSIT random beamforming for multibeam satellite systems," in *10th Int. Work. Signal. Process. Space Commun., SPSC 2008*, Oct. 2008, pp. 1-8. (Cited on page 6.)

[28] G. Zheng, S. Chatzinotas, and B. Ottersten, "Generic optimization of linear precoding in multibeam satellite systems," *IEEE Trans. Wireless Commun.*, vol. 11, no. 6, pp. 2308-2320, Jun. 2012. (Cited on pages 6 and 42.)

[29] S. Chatzinotas, G. Zheng, B. Ottersten, "Energy-efficient MMSE beamforming and power allocation in multibeam satellite systems," in *Proceedings of the 45th Asilomar Conference on Signals, Systems, and Computers (2011)*, Nov. 2011, pp. 1081-1085 (Cited on page 6.)

[30] Second Generation Framing Structure, Channel Coding and Modulation Systems for Broadcasting Interative Services, News Gathering and Other Broadband Satellite Applications; Part 2: DVB-S2 Extensions (DVB-S2X), document ETSI EN 302-307-2 V1.1.1, European Broadcasting Union (EBU), Grand-Saconnex, Switzerland, Oct. 2014. (Cited on pages 6 and 43.)

[31] G. Taricco, "Linear precoding methods for multi-beam broadband satellite systems," in *Proc. 20th Eur. Wireless Conf. Eur. Wireless*, May 2014, pp. 1-6. (Cited on page 6.)

[32] V. Joroughi *et al.*, "Generalized multicast multibeam precoding for satellite communications," *IEEE Trans. Wireless Commun.*, vol. 16, no. 2, pp. 952-966, Feb. 2017. (Cited on pages 6, 7, 17, 41, 50, 57, 58, 59, 65, 66, 67, 68 and 70.)

[33] V. Joroughi *et al.*, "Precoding in multigateway multibeam satellite systems," *IEEE Trans. Wireless Commun.*, vol. 15, no. 7, pp. 4944-4956, Jul. 2016. (Cited on page 7.)

[34] G. Zheng, S. Chatzinotas, and B. Ottersten, "Multi-Gateway Cooperation in Multibeam Satellite Systems," in *Proc. 23rd IEEE Symp. Personal Indoor Mobile Radio Commun.*, Sep. 2012, pp. 1360-1364. (Cited on page 7.)

[35] B. Devillers and A. I. Pérez-Neira, "Advanced interference mitigation techniques for the forward link of multi-beam broadband satellite systems," in *Signals, Systems and Computers (ASILOMAR)*, Nov. 2011, pp. 1810-1814. (Cited on page 7.)

[36] V. Joroughi, M. Á. Vázquez and A. I. Pérez-Neira, "Multiple Gateway Precoding with Per Feed Power Constraints for Multibeam Satellite Systems," in *Proceedings of 20th European Wireless Conference*, pp. 1-7, May 2014. (Cited on page 7.)

[37] D. Christopoulos, H. Pennanen, S. Chatzinotas and B. Ottersten, "Multicast multigroup precoding for frame-based multi-gateway satellite communications," in *Proc. 8th Advanced Satellite Multimedia Systems Conf. and 14th Signal Processing Space Communications Workshop (ASMS/SPSC)*, Sep. 2016, pp. 1-6. (Cited on pages 7, 49 and 50.)

[38] Y. Chen *et al.*, "Toward the Standardization of Non-Orthogonal Multiple Access for Next Generation Wireless Networks," *IEEE Communications Magazine*, vol. 56, no. 3, pp. 19-27, Mar. 2018. (Cited on page 8.)

[39] Z. Ding, R. Schober and H. V. Poor, "Unveiling the Importance of SIC in NOMA Systems Part 1: State of the Art and Recent Findings," *IEEE Communications Letters*, vol. 24, no. 11, pp. 2373-2377, Nov. 2020. (Cited on page 8.)

[40] Z. Ding *et al.*, "A general MIMO framework for NOMA downlink and uplink transmission based on signal alignment," *IEEE Trans. Wireless Commun.*, vol. 15, no. 6, pp. 4438-4454, Jun. 2016. (Cited on pages 8, 12 and 41.)

[41] Y. Liu, Z. Qin, M. Elkashlan, Z. Ding, A. Nallanathan, and L. Hanzo, "Non-orthogonal multiple access for 5G and beyond," *Proc. IEEE*, vol. 105, no. 12, pp. 2347-2381, Dec. 2017. (Cited on pages 8, 9 and 10.)

[42] Y. Saito, Y. Kishiyama, A. Benjebbour, T. Nakamura, A. Li, and K. Higuchi, "Non-orthogonal multiple access (NOMA) for cellular future radio access," in *Proc. IEEE 77th Vehicular Technology Conference (VTC Spring)*, Jun. 2013, pp. 1-5. (Cited on pages 8, 9 and 10.)

[43] S. R. Islam, N. Avazov, O. A. Dobre, and K.-S. Kwak, "Power-domain non-orthogonal multiple access (NOMA) in 5G systems: potentials and challenges," *IEEE Commun. Surveys Tuts.*, vol. 19, no. 2, pp. 721-742, Oct. 2017. (Cited on pages 8, 9 and 10.)

[44] T. Cover, "Broadcast channels," *IEEE Trans. Inf. Theory.*, vol. 18, no.1, pp. 2-14, Jan. 1972 (Cited on page 9.)

[45] Z. Chen, Z. Ding, X. Dai and R. Zhang, "An Optimization Perspective of the Superiority of NOMA Compared to Conventional OMA," *IEEE Transactions on Signal Processing*, vol. 65, no. 19, pp. 5191-5202, Oct, 2017. (Cited on pages 10 and 11.)

[46] Z. Ding, F. Adachi, and H. V. Poor, "The application of MIMO to non-orthogonal multiple access," *IEEE Trans. Wireless Commun.*, vol. 15, no. 1, pp. 537-552, Jan. 2016. (Cited on page 12.)

[47] Y. Zhu et *et al.*, "Near-optimal practical convergent method for interference alignment in MIMO interference channels," *Electronics Letters*, vol. 52, no. 25, pp. 2074-2076, Dec. 2016, (Cited on page 12.)

[48] B. Kim et al., "Non-orthogonal multiple access in a downlink multiuser beamforming system," in *Proc. IEEE Military Commun. Conf. (MILCOM)*, Nov. 2013, pp. 1278-1283. (Cited on page 12.)

[49] M. S. Ali, E. Hossain, and D. I. Kim, "Non-orthogonal multiple access (NOMA) for downlink multiuser MIMO systems: User clustering, beamforming, and power allocation," *IEEE Access*, vol. 5, pp. 565-577, Dec. 2016. (Cited on page 12.)

[50] Z. Chen, Z. Ding, P. Xu, and X. Dai, "Optimal precoding for a QoS optimization problem in two-user MISO-NOMA downlink," *IEEE Commun. Lett.*, vol. 20, no. 6, pp. 1263-1266, Jun. 2016. (Cited on page 12.)

[51] Z. Chen, Z. Ding, X. Dai, and G. K. Karagiannidis, "On the application of quasi-degradation to MISO-NOMA downlink," *IEEE Trans. Signal Process.*, vol. 64, no. 23, pp. 6174-6189, Dec. 2016. (Cited on page 12.)

[52] M. F. Hanif, Z. Ding, T. Ratnarajah, and G. K. Karagiannidis, "A minorization-maximization method for optimizing sum rate in non-orthogonal multiple access systems," *IEEE Trans. Signal Process.*, vol. 64, no. 1, pp. 76-88, Jan. 2016. (Cited on page 12.)

[53] M. Caus, M. Á. Vázquez and A. I. Pérez-Neira, "NOMA and interference limited satellite scenarios," in *2016 50th Asilomar Conference on Signals, Systems and Computers, 2016*, Nov. 2016, pp. 497-501. (Cited on pages 12, 16, 43, 53 and 54.)

[54] A. I. Pérez-Neira, M. Caus, and M. Á. Vázquez, "Non-orthogonal transmission techniques for multibeam satellite systems," *IEEE Commun. Mag.*, vol. 57, no. 12, pp. 58-63, Dec. 2019. (Cited on page 12.)

[55] X. Yan et al., "The Application of Power-Domain Non-Orthogonal Multiple Access in Satellite Communication Networks," *IEEE Access*, vol. 7, 2019, pp. 63531-63539. (Cited on page 12.)

[56] Y. Zhu, T. Delamotte, A. Knopp, "Geographical NOMABeamforming in Multi-Beam Satellite-Based Internet of Things," in *2019 IEEE GLOBECOM*, Dec. 2019, pp. 2239-2251. (Cited on pages 12, 17, 22, 37, 38 and 43.)

[57] T. Ramírez and C. Mosquera, "Resource Management in the Multibeam NOMA-based Satellite Downlink," in *2020 IEEE ICASSP*, May 2020, pp. 8812-8816. (Cited on page 12.)

[58] T. Ramírez et al., "Study on the Application of NOMA Techniques for Heterogeneous Satellite Terminals," in *10th ASMS and 16th SPSC*, Oct. 2020, pp. 1-8. (Cited on page 12.)

[59] S. Majidi, M. Caus, M. Á. Vázquez, M. Reza Soleymani, Y. R. Shayan, A. I. Pérez-Neira, "Power Allocation and User Clustering in Multicast NOMA based Satellite Communication Systems," in *2020 IEEE ICC*, Jun. 2020, pp. 1-6. (Cited on page 12.)

[60] A. Wang *et al.*, "NOMA-Enabled Multi-Beam Satellite Systems: Joint Optimization to Overcome Offered-Requested Data Mismatches," *IEEE Trans. on Vehicular Technology.*, vol. 70, no. 1, pp. 900-913, Jan. 2021. (Cited on page 12.)

[61] Y. Zhu, C. A. Hofmann, A. Knopp, "Performance Optimization for Multi-Gateway NOMA-Beamforming in Multi-Beam SATCOM," in *2021 IEEE ICC*, Jun. 2021, pp. 1-6. (Cited on pages 12, 22, 37, 38, 74 and 75.)

[62] X. Zhu, C. Jiang, L. Kuang, N. Ge and J. Lu, "Non-Orthogonal Multiple Access Based Integrated Terrestrial-Satellite Networks," *IEEE J. on Sel. Areas in Communications*, vol. 35, no. 10, pp. 2253-2267, Oct. 2017. (Cited on page 12.)

[63] Z. Lin *et al.*, "Joint beamforming and power allocation for satellite-terrestrial integrated networks with non-orthogonal multiple access," *IEEE J. Sel. Topics Sig. Proc.*, vol. 13, no. 3, pp. 657-670, Jun. 2019. (Cited on page 12.)

[64] C. A. Hofmann, R. T. Schwarz and A. Knopp, "SOTM Measurements for the Characterization of the Wideband Mobile Satellite Channel at Ku-Band," in *SCC 2015; 10th International ITG Conference on Systems, Communications and Coding*, Feb. 2015, pp. 1-8. (Cited on page 14.)

[65] A. I. Pérez-Neira, *et al.*, "Signal processing for high-throughput satellites: Challenges in new interference-limited scenarios," *IEEE Signal Process. Mag.*, vol. 36, no. 4, pp. 112-131, Jul. 2019. (Cited on page 15.)

[66] J. Lizarraga, P. Angeletti, N. Alagha and M. Aloisio, "Flexibility performance in advanced Ka-band multibeam satellites," in *Proc. IEEE International Vacuum Electronics Conference*, Apr. 2014, pp. 45-46. (Cited on pages 16 and 39.)

[67] M. Schneider, C. Hartwanger and H. Wolf, "Antennas for multiple spot beam satellites," *CEAS Space Journal*, vol. 2, pp. 59-66, Aug. 2011 (Cited on page 23.)

[68] D. Minoli, Innovations in Satellite Communication and Satellite Technology, *Wiley*, 2015, pp.413-420. (Cited on page 23.)

[69] W. Q. Lohmeyer, R. J. Aniceto, and K. L. Cahoy, "Communication satellite power amplifiers: current and future SSPA and TWTA technologies," *Int. J. Satell. Commun. Network.*, vol. 34, pp: 95-113, Mar. 2015 (Cited on page 24.)

[70] C. Haardt *et al.*, "Internet switching by satellite: an ultra fast processor with radio burst switching," in *Proc. Disruption in Space*, May 2005, pp. 1-17. (Cited on pages xv, 26, 44 and 49.)

[71] V. Boussemart, "On the Benefits of Cross-Layer - Assisted Interference, Resource and Application Management in Satellite-Based Communications," Ph.D. Thesis, ISAE (Toulouse, France) and DLR (Munich, Germany), Apr. 2013 (Cited on pages xv, 26, 27, 44 and 49.)

[72] $3GPP\ TR38.811$, "Study on New Radio (NR) to Support Non - terrestrial Networks." $V15.0.0$, August 2018. Available at http://www.3gpp.org/ftp//Specs/archive/38_series/38.811/38811-f00.zip. [Online; Last Accessed on 03-25-2019]. (Cited on pages xv, 34 and 35.)

[73] E. Michailidis, A. Kanatas, "Radio Wave Propagation and Channel Modeling for Earth-Space Systems," *Wiley*, 2016. (Cited on page 35.)

[74] R. Crane, "Prediction of Attenuation by Rain," *IEEE Transactions on Communications*, vol. 28, no. 9, pp. 1717-1733, Sep. 1980. (Cited on page 35.)

[75] $ITU-R\ P.618-13$, "Propagation Data and Prediction Methods Required for the Design of Earth-Space Telecommunication Syste." *Report*, International Telecommunication Union: Geneva, Switzerland, 2011. (Cited on pages 35 and 36.)

[76] $ITU-R\ P.839-4$, "Rain height model for prediction methods." *Report*, International Telecommunication Union: Geneva, Switzerland, 2013. (Cited on page 36.)

[77] $ITU-R\ P.837-7$, "Characteristics of precipitation for propagation modeling." *Report*, International Telecommunication Union: Geneva, Switzerland, 2017. (Cited on page 36.)

[78] $ITU-R\ P.838-3$, "Specific attenuation model for rain for use in prediction methods." *Report*, International Telecommunication Union: Geneva, Switzerland, 2005. (Cited on page 36.)

[79] D. Tse *et al.*, "Fundamentals Wireless Communication," *Cambridge, U.K.: Cambridge Univ. Press*, 2005. (Cited on page 41.)

[80] D. Christopoulos *et al.*, "Linear and nonlinear techniques for multibeam joint processing in satellite communications," *J Wireless Com Network*, 162, pp. 1-13, May 2012. (Cited on pages 42, 49 and 50.)

[81] H. -P. Kriegel, E. Schubert, A. Zimek, "The (black) art of runtime evaluation: Are we comparing algorithms or implementations?," *Knowledge and Information Systems*, pp. 1-38, 2016. (Cited on pages 44 and 46.)

[82] J. A. Hartigan and M. A. Wong, "Algorithm AS 136: A K-Means Clustering Algorithm," *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, vol. 28, no. 1, pp. 100-108, 1979. (Cited on page 48.)

[83] D. Mignolo *et al.*, "Approaching Terabit/s satellite capacity: A system analysis," in *Proc. 17th Ka and Broadband Communications Conference*, Jan. 2011, pp. 1-10. (Cited on page 50.)

[84] X. Mu, Y. Liu, L. Guo, J. Lin and R. Schober, "Joint Deployment and Multiple Access Design for Intelligent Reflecting Surface Assisted Networks," *IEEE Transactions on Wireless Communications*, vol. 20, no. 10, pp. 6648-6664, Oct. 2021. (Cited on page 53.)

[85] H. Guo, Y. Liang, J. Chen and E. G. Larsson, "Weighted Sum-Rate Maximization for Reconfigurable Intelligent Surface Aided Wireless Networks," *IEEE Transactions on Wireless Communications*, vol. 19, no. 5, pp. 3064-3076, May 2020. (Cited on page 53.)

[86] Y. I. Choi, J. W. Lee, M. Rim and C. G. Kang, "On the Performance of Beam Division Nonorthogonal Multiple Access for FDD-Based Large-Scale Multi-User MIMO Systems," *IEEE Transactions on Wireless Communications*, vol. 16, no. 8, pp. 5077-5089, Aug. 2017. (Cited on page 53.)

[87] J. Wang, D. Wang and Y. Liu, "Weighted Sum Rate-Based Coordinated Beamforming in Multi-Cell Multicast Networks," *IEEE Communications Letters*, vol. 20, no. 8, pp. 1567-1570, Aug. 2016. (Cited on page 53.)

[88] M. Á. Vázquez *et al.*, "Precoding in Multibeam Satellite Communications: Present and Future Challenges," *IEEE Wireless Communications*, vol. 23, no. 6, pp. 88-95, Dec. 2016. (Cited on pages 57 and 58.)

[89] N. Sidiropoulos *et al.*, "Transmit beamforming for physical-layer multicasting," *IEEE Trans. Sig. Proc.*, vol. 54, no. 6, pp. 2239–2251, Jun. 2006. (Cited on pages 59 and 60.)

[90] Z.-Q. Luo *et al.*, "Semidefinite relaxation of quadratic optimization problems," *IEEE Sig. Proc. Mag.*, vol. 27, no. 3, pp. 20–34, May 2010. (Cited on pages 59, 64 and 85.)

[91] W.-K. Ma, T.-N. Davidson, K.-M. Wong, Z.-Q. Luo, and P.-C. Ching, "Quasi-ML multiuser detection using semi-definite relaxation with application to synchronous CDMA," *IEEE Trans. Signal Process.*, vol. 50, no. 4, pp. 912-922, Apr. 2002. (Cited on pages 60 and 65.)

[92] M. Sadek *et al.*, "A leakage-based precoding scheme for downlink multi-user MIMO channels," *IEEE Trans. on Wireless Commun.*, vol. 6, no. 5, pp. 1711-1721, May 2007. (Cited on page 61.)

[93] S. Boyd *et al.*, "Convex optimization," *Cambridge Univ. Press*, 2004. (Cited on pages 62 and 77.)

[94] P. C. Weeraddana *et al.* "Weighted Sum-Rate Maximization in Wireless Networks: A Review," *Foundations and Trends in Networking*, vol. 6, no. 1-2, pp. 1-163, Oct. 2012. (Cited on page 74.)

[95] F. P. Kelly *et al.*, "Rate control for communication networks: Shadow prices, proportional fairness and stability," *J. Operat. Research Soc.*, vol. 49, no. 3, pp. 237-252, 1998. (Cited on page 74.)

[96] S. S. Christensen *et al.*, "Weighted sum-rate maximization using weighted MMSE for MIMO-BC beamforming design," *IEEE Trans. Wireless Commun.*, vol. 7, no. 12, pp. 4792-4799, Dec. 2008. (Cited on pages 74, 77 and 87.)

[97] Q. Shi *et al.*, "An iteratively weightedMMSE approach to distributed sum-utility maximization for a MIMO interfering broadcast channel," *IEEE Trans. Signal Process.*, vol. 59, no. 9, pp. 4331-340, Sep. 2011. (Cited on pages 74, 76, 77, 78 and 87.)

[98] S. J. Wright, "Coordinate descent algorithms," *Math. Program*, ser. B, no. 151, pp. 3-34, Mar. 2015. (Cited on page 77.)

[99] M. Razaviyayn *et al.*, "A unified convergence analysis of block successive minimization methods for nonsmooth optimization," *SIAM J. Optimization*, vol. 23, no. 2, pp. 1126-1153, 2013. (Cited on page 78.)

[100] O. Simeone, "A Very Brief Introduction to Machine Learning With Applications to Communication Systems," *IEEE Transactions on Cognitive Communications and Networking*, vol. 4, no. 4, pp. 648-664, Dec. 2018. (Cited on page 79.)

[101] Y. Zhu, C. A. Hofmann and A. Knopp, "Distributed Resource Optimization for NOMA Transmission in Beamforming SATCOM," *IEEE Journal on Sel. Areas in Com.*, vol. 40, no. 4, pp. 1190-1209, Apr. 2022. (Cited on pages 79, 82, 86 and 88.)

[102] L. Sanguinetti *et al.*, "Deep Learning Power Allocation in Massive MIMO," in *Proc. 2018 52nd ACSSC*, Feb. 2018, pp. 1257-1261. (Cited on pages 79 and 81.)

[103] H. Sun *et al.*, "Learning to Optimize: Training Deep Neural Networks for Interference Management," *IEEE Transactions on Signal Processing*, vol. 66, no. 20, pp. 5438-5453, Oct. 2018. (Cited on pages 79 and 81.)

[104] N. Rajapaksha *et al.*, "Deep Learning-based Power Control for Cell-Free Massive MIMO Networks," in *ICC 2021*, Jun. 2021, pp. 1-7. (Cited on page 79.)

[105] W. Lee *et al.*, "Intelligent Resource Allocation in Wireless Communications Systems," *IEEE C. M.*, vol. 58, no. 1, pp. 100-105, Jan. 2020. (Cited on pages 79 and 84.)

[106] F. Liang *et al.*, "Towards Optimal Power Control via Ensembling Deep Neural Networks," *IEEE T. on C.*, vol. 68, no. 3, pp. 1760-1776, Mar. 2020. (Cited on pages 79 and 87.)