Distributed Precoding Systems in Multi-Gateway Multibeam Satellites: Regularization and Coarse Beamforming

Carlos Mosquera, *Senior Member, IEEE*, Roberto López-Valcarce, *Member, IEEE*, Tomás Ramírez, Vahid Joroughi

***Abstract*—This paper deals with the problem of beamformer design in a multibeam satellite, which is shared by several clusters of terminals, each served by an Earth station, or gateway. Each gateway precodes the symbols addressed to its respective users; the design follows an MMSE criterion, and a judiciously chosen regularization factor allows to account for the presence of mutually interfering clusters, extending more classical results applicable to the case of a single centralized station. More importantly, channel statistics can be used instead of instantaneous channel state information, avoiding the exchange of information among gateways through backhaul links. The on-board satellite beamforming weights are designed to exploit the degrees of freedom of the satellite antennas to minimize the noise impact and the interference to some specific users. On-ground beamforming results are provided as a reference to compare the joint performance of MMSE precoders and on-board beamforming network. A non-adaptive design complements the results and makes them more amenable to practical use by designing a coarse beamforming network.**

I. Introduction

Bent-pipe communication satellites can be considered as non-regenerative relays [1], essentially filtering and amplifying signals, although they are very complex communication systems and handle simultaneously many streams of information. The object of this study is a multibeam satellite which relays the signals coming from M ground stations (gateways) to convey their communication with single-antenna terminals. The foot-print of a multibeam satellite is made of many spot-beams, hundreds in some specific commercial cases, which are synthesized by the on-board beamforming network (BFN) in combination with the radiation pattern of the antennas. Two implementation approaches are possible: single feed per beam and multiple feeds per beam. For the purpose of this paper, it is of specific interest the case of multiple feeds per beam, for which small subarrays are used for each spot-beam[1], and adjacent spot-beams share some of the array elements. This technology has some advantages since individual beams can overlap and a single reflector antenna served by several feeds can cover a larger area [1]. The on-board beamforming (OBBF) process contributes some flexibility to the shaping of the beams, although the configurability is in most cases quite limited, and real-time adaptation in the range of milliseconds is rarely feasible. Remarkably, a technology known as On-Ground Beamforming (OGBF) has emerged as an alternative solution for some specific cases, to avoid the need for a complex on-board digital processor. This technology has been used in some recent multibeam mobile satellite systems [2], and requires the exchange of all the feed signals between the satellite and the gateway, increasing the bandwidth demands on the feeder link due to the higher number of information streams. In addition, OGBF poses some specific calibration issues, in particular for the multi-gateway case [3].

Fig. 1 depicts the conceptual abstraction of the multi-beam satellite operation, with the following features to be highlighted: (i) the feeder links, from the gateways to the satellites, can be assumed transparent, whereas the user links are frequency non-selective; (ii) there is no interference between feeder links and user links, since the communication takes place on different frequency bands; (iii) a given cluster comprises several beams (see Fig. 2), with one user per beam served at a time by a given frequency carrier; (iv) the user link frequency carriers are made available to all beams and clusters, in what is known as full-frequency reuse.

One major challenge for multibeam satellites is the large spectral demand on the feeder link between the satellite and the operator stations, since it has to aggregate the traffic from all beams. Technology has contributed to a steady increase of this traffic during the last years due to, among other things, a more efficient reuse of spectrum across the different beams [4]. The use of different gateways can generate several parallel channels provided that the antennas guarantee the required spatial isolation, which is the case for frequencies in Ka-band and Q/V-band. Thus, the different feeder links can reuse the whole available bandwidth while avoiding mutual interference.

In contrast, the user links from satellite to user terminals are significantly affected by cross-interference caused by the side-lobes of the different radiating elements together with the adoption of full-frequency reuse. Mitigating this interference becomes then necessary, and this is usually done by means of preprocessing techniques. The preprocessing of signals to communicate a multi-antenna transmitter with many users simultaneously is supported by theoretical bounds

Carlos Mosquera, Roberto López-Valcarce and Tomás Ramírez are with the Signal Theory and Communications Department, University of Vigo, Galicia, Spain. (e-mail: {mosquera, valcarce, tramirez}@gts.uvigo.es). Vahid Joroughi was with the Signal Theory and Communications Department, University of Vigo, Galicia, Spain. This work was partially funded by the Agencia Estatal de Investigación (Spain) and the European Regional Development Fund (ERDF) through the Projects MYRADA under Grant TEC2016-75103-C2-2-R, WINTER under Grant TEC2016-76409-C2- 2-R, and COMONSENS under Grant TEC2015-69648-REDC. Also funded by the Xunta de Galicia (Agrupación Estratéxica Consolidada de Galicia accreditation 2016-2019; Red Temática RedTEIC 2017-2018) and the European Union (European Regional Development Fund - ERDF).

[1]We will use spot-beam and beam as equivalent terms in this paper.

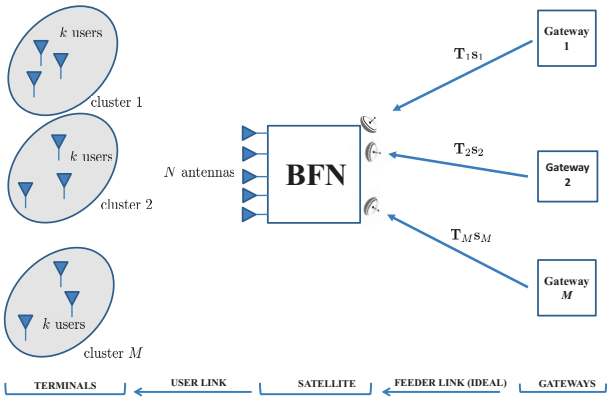


Fig. 1: Satellite shared by a number of ground stations.

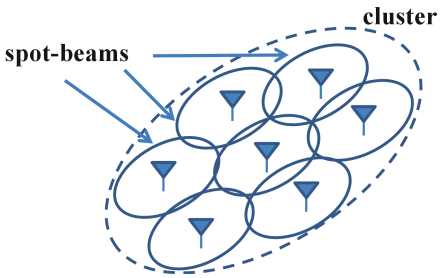


Fig. 2: Cluster of beams.

and practical schemes presented in many references. In the context of terrestrial systems, new insights to the problem have been obtained in the past few years, especially for a large number of antennas, see, among others, [5], [6], [7], [8]. In the particular case of linear precoding, the seminal paper [9] analyzes the regularization of channel inversion at the transmitter to maximize the signal to interference and noise ratio (SINR), with specific focus on Rayleigh channels. Multibeam satellite channels need line-of-sight links to guarantee a minimum received power; several propagation paths are created from the different radiating elements -feeds- to a given Earth terminal. These propagation paths can also be exploited by preprocessing at the transmit side to mitigate the co-channel interference. In particular, precoding for multibeam satellites has been extensively explored in the literature to fight interbeam interference in the case of a single gateway, see, e.g., [10] and [11] among others. Potential gain and calibration requirements have been properly identified by these works. In contrast, results for multiple gateways are still incomplete; a centralized multi-gateway resource management, which for mathematical purposes can be assumed, is far from being realistic in practice [4], due to the high-bandwidth backhaul links that would be needed to connect all the gateways[2]. Some precoding schemes for multiple gateways without BFN are presented in [13], which assume the exchange of information for the design of their respective precoders. However, no comprehensive approach is known conciliating the design of an on-board BFN and the use of several gateways, required to channelize the high throughput for an aggressive use of the user link spectrum. The mapping of groups of beams to different gateways precludes a centralized management, making inter-cluster interference more difficult to control.

In this work we try to keep cooperation at a minimum, so that no information symbols are exchanged among the terrestrial gateways, each communicating with the terminals operating on its cluster. Initially, each gateway is assumed to have the channel state information (CSI) of the links originating from itself, including inter-cluster links. Later on, we will show that the use of channel statistics (instead of CSI) can completely avoid interaction among clusters without significantly degrading performance. The global interplay of distributed precoders operating at the different gateways and the BFN on the satellite can be designed in such a way that different solutions accrue, each fitting the available information and flexibility of the involved subsystems. Specific regularization rules for the distributed precoders are obtained as an extension of the results for one cluster. As a reference, we compute the performance of the fully flexible OGBF with perfect CSI at both the gateways (CSIT) and receive terminals (CSIR). Starting from that, separate on-board BFN and precoders are derived. The performance indicator is the Mean Square Error (MSE) for three different settings: (i) intra-cluster interference driven; (ii) cancellation of interference leaked to specific off-cluster users; (iii) coarse fixed BFN designed as a trade-off solution. The last scheme is particularly important, since it allows to fix the on-board BFN and confine the flexibility to the gateway precoders: naturally, this less complex satellite payload incurs in some performance loss.

The major contributions of this paper are the following, all applied to a multi-gateway multibeam full frequency reuse system.

- For an on-ground beamforming setting, a scheme to obtain adaptively the coefficients for each gateway is presented under the minimum MSE (MMSE) criterion.
- For an on-board beamforming setting, we obtain the expression of the ground multi-gateway MMSE precoders. We also provide rules for the regularization factor of the ground MMSE precoders, which make them amenable to operate in multi-cluster environments. These rules are such that autonomous operation is achieved, based only on second-order statistics of the global system.
- The on-board beamforming network is obtained under different premises. Full adaptive solutions are obtained, together with fixed coarse solutions based on the Gramians of the involved channels.

After detailing the satellite relaying operation in Sec. II, we derive the optimal beamforming weights and gateway precoders, first for the OGBF case in Sec. III and then for

[2]If this complexity was affordable, a CRAN-like approach, with CRAN standing for Cloud or Centralized Radio Access Network [12], could yield the performance of a single gateway network.

separate on-board beamforming and ground precoders in Sec. IV. Both fixed and adaptive BFN weights are object of the study, with performance tested in the simulations shown before the conclusions.

Notation: Upper (lower) boldface letters denote matrices (vectors). $(.)^H$, $(.)^T$, $(.)^\dagger$, $\mathrm{tr}\{\cdot\}$, $\mathbf{I}_N$, $\mathbf{0}$, $\mathrm{diag}\{\cdot\}$ denote Hermitian transpose, transpose, pseudoinverse, matrix trace operator, $N \times N$ identity matrix, all-zero matrix, and diagonal matrix, respectively. $\mathbb{E}[\cdot]$ is the expected value operator.

II. Satellite Relaying Operation

The satellite serves K terminals at each channel use[3]. All K users get access to the same frequency spectrum, thus giving rise to both intra-cluster and inter-cluster interference. The satellite has N radiation elements, or feeds, with $N \geq K$. The $K \times 1$ vector comprising the values received by the K users at a given time instant is written as

$$\mathbf{y} = \widetilde{\mathbf{H}}\widetilde{\mathbf{B}}\mathbf{x} + \mathbf{n} \tag{1}$$

with $\mathbf{x} \in \mathbb{C}^K$ the vector of transmitted symbols, and $\mathbf{n} \in \mathbb{C}^K$ zero-mean unit variance Additive White Gaussian Noise (AWGN), such that $\mathbb{E}\left[\mathbf{n}\mathbf{n}^H\right] = \mathbf{I}_K$. The BFN weights are included in matrix $\widetilde{\mathbf{B}} \in \mathbb{C}^{N\times K}$. $\widetilde{\mathbf{H}} \in \mathbb{C}^{K\times N}$ is the overall user link channel matrix whose element $[\widetilde{\mathbf{H}}]_{ij}$ represents the gain of the link between the i-th user and the j-th satellite feed.

As shown in Fig. 1, the number of transmit ground stations is M, each sending k signal streams simultaneously (in different frequency slots, for example) to the satellite, which makes use of n antenna feeds to send those symbols to the k users in the mth cluster, with $k \leq n \leq N$. The groups of n feeds are not necessarily disjoint: in practical deployments, some of the feeds can be shared between adjacent beams and, by extension, by adjacent clusters [14], [15]. Feed sharing has some practical beneficial effects in the reduction of inter-beam interference due to the increase of the effective primary feeding apertures.

The information transmitted from each ground station is written as $\mathbf{x}_m = \mathbf{T}_m\mathbf{s}_m$, with $\mathbf{T}_m \in \mathbb{C}^{k\times k}, m = 1,\ldots,M$, a set of distributed precoding matrices, and $\mathbf{s}_m \in \mathbb{C}^{k\times 1}, m = 1,\ldots,M$ the symbols transmitted by the mth gateway. The initial model (1) can be detailed as

$$\mathbf{y} = \widetilde{\mathbf{H}}\begin{bmatrix} \widetilde{\mathbf{B}}_1 & \cdots & \widetilde{\mathbf{B}}_M \end{bmatrix}\begin{bmatrix} \mathbf{T}_1\mathbf{s}_1 \\ \vdots \\ \mathbf{T}_M\mathbf{s}_M \end{bmatrix} + \mathbf{n} \tag{2}$$

where the tall matrices $\widetilde{\mathbf{B}}_m \in \mathbb{C}^{N\times k}$ contain the BFN weights assigned to gateway m, and $k \cdot M = K$. The transmit power is normalized as $\mathbb{E}\left[\mathbf{s}_m\mathbf{s}_m^H\right] = \mathbf{I}_k$. The goal of the precoder at each transmitter is mainly to mitigate the intra-cluster interference, while trying to reduce the negative impact of its interference on other clusters. The BFN should exploit the additional degrees of freedom to gain inter-cluster interference and/or noise resilience, preferably in a robust way against the uncertain location of the users.

The fact that only a subset of feeds is used to give service to any given cluster imposes certain structure on the matrices $\widetilde{\mathbf{B}}_m$. Since n denotes the number of feeds serving each cluster, the weights with content in the BFN can be collected by the tall submatrices $\mathbf{B}_m \in \mathbb{C}^{n\times k}, m = 1,\ldots,M$, with $k \leq n \leq N$. Each matrix $\widetilde{\mathbf{B}}_m$ in (2) only has n non-zero rows, so we can write

$$\widetilde{\mathbf{B}}_m = \mathcal{S}_m\mathbf{B}_m \tag{3}$$

where $\mathcal{S}_m$ comprises n columns of $\mathbf{I}_N$, in particular those with the indices of the feeds used by gateway m. Note that $\mathcal{S}_m$ and $\mathcal{S}_{m'}$, $m \neq m'$, may share common columns, i.e., a given feed may be used by more than one gateway. With $n > k$, there are extra degrees of freedom to fight the inter-cluster interference and gain noise resilience without increasing the bandwidth of the user link.

If we decompose the received signal and noise vectors in (1) into their respective vectors per cluster, $\mathbf{y}_m \in \mathbb{C}^k$ and $\mathbf{n}_m \in \mathbb{C}^k$, respectively, and denoting by $\mathbf{H}_{mp} \in \mathbb{C}^{k\times n}$ the channel between the n feeds operated by the pth gateway and the mth cluster, then (2) reads as

$$\begin{pmatrix} \mathbf{y}_1 \\ \vdots \\ \mathbf{y}_M \end{pmatrix} = \underbrace{\begin{pmatrix} \mathbf{H}_{11} & \ldots & \mathbf{H}_{1M} \\ \vdots & \ddots & \vdots \\ \mathbf{H}_{M1} & \ldots & \mathbf{H}_{MM} \end{pmatrix}}_{\mathbf{H}} \begin{pmatrix} \mathbf{B}_1 & & \\ & \ddots & \\ & & \mathbf{B}_M \end{pmatrix} \begin{pmatrix} \mathbf{T}_1\mathbf{s}_1 \\ \vdots \\ \mathbf{T}_M\mathbf{s}_M \end{pmatrix} + \begin{pmatrix} \mathbf{n}_1 \\ \vdots \\ \mathbf{n}_M \end{pmatrix}. \tag{4}$$

Note that, in this cluster-oriented notation, the channel matrix $\mathbf{H} \in \mathbb{C}^{K\times nM}$ does not coincide with $\widetilde{\mathbf{H}}$ in (2). If $\mathbf{H}$ is decomposed as $\begin{bmatrix} \mathbf{H}_1 & \cdots & \mathbf{H}_M \end{bmatrix}$, and since both expressions (4) and (2) need to be equivalent, we can readily conclude that $\mathbf{H}_m = \widetilde{\mathbf{H}}\mathcal{S}_m$, with $\mathbf{H}_m \in \mathbb{C}^{K\times n}$.
The vector of samples received by users in cluster m is decomposed as

$$\mathbf{y}_m = \underbrace{\mathbf{H}_{mm}\mathbf{B}_m\mathbf{T}_m\mathbf{s}_m}_{\text{intra-cluster}} + \underbrace{\sum_{p\neq m}\mathbf{H}_{mp}\mathbf{B}_p\mathbf{T}_p\mathbf{s}_p}_{\text{inter-cluster}} + \underbrace{\mathbf{n}_m}_{\text{noise}}. \tag{5}$$

To some extent, this is a downlink multi-user MIMO model similar to that applied for cellular terrestrial communications, where different multi-antenna base stations communicate with several co-channel users, and transmission schemes are required to suppress co-channel interference [16], [17]. Matrices $\mathbf{H}_{mm}$ are described by different models in both satellite and terrestrial communications, and the role of the multi-antenna satellite as a relay poses a major difference with cellular commmunications. As exposed later, the location of beamforming weights $\mathbf{B}_m$ and precoding coefficients $\mathbf{T}_m$ in the satellite payload and gateways, respectively, have practical relevance as to their degree of adaptation.

As end users are not allowed to cooperate, we consider a receiver of the form $\hat{\mathbf{s}}_m = \mathbf{D}_m\mathbf{y}_m$, where the matrix $\mathbf{D}_m$ is

[3]The focus of this study is on the forward link from gateways to user terminals, without precluding the support to the return link.

$k \times k$ diagonal. The particular case in which the same scaling is used across the cluster will be also considered later, with $\mathbf{D}_m = \frac{1}{\sqrt{t_m}}\mathbf{I}_k$.
As performance metric we use the aggregated MSE, or Sum MSE (SMSE), given by

$$\text{SMSE} = \sum_{m=1}^{M} \text{tr}\left\{\mathbf{E}_m\right\}, \tag{6}$$

with

$$\mathbf{E}_m \triangleq \mathbb{E}\left[(\mathbf{s}_m - \hat{\mathbf{s}}_m)(\mathbf{s}_m - \hat{\mathbf{s}}_m)^H\right]. \tag{7}$$

The expectation is computed with respect to the symbols and the noise for a fixed channel, and reads as

$$\begin{aligned}\text{SMSE} = \\ \sum_{m=1}^{M} \text{tr}\Big\{\mathbf{I}_k - \mathbf{D}_m\mathbf{H}_{mm}\mathbf{B}_m\mathbf{T}_m - \mathbf{T}_m^H\mathbf{B}_m^H\mathbf{H}_{mm}^H\mathbf{D}_m^H \\ +\mathbf{T}_m^H\mathbf{B}_m^H\left(\sum_{p=1}^{M}\mathbf{H}_{pm}^H\mathbf{D}_p^H\mathbf{D}_p\mathbf{H}_{pm}\right)\mathbf{B}_m\mathbf{T}_m + \mathbf{D}_m\mathbf{D}_m^H\Big\},\end{aligned} \tag{8}$$

written in such a way that the impact of $\mathbf{T}_m$ and $\mathbf{B}_m$ on the overall error is limited to the mth term of the summation. This way of dealing together with the interference posed on the same cluster and leaked to other clusters have been explored in other works such as [16], where the Signal to Leakage and Noise Ratio (SLNR) was maximized for a single transmitter. Further, [18] showed that the minimization of the MSE and the maximization of SLNR lead to equivalent solutions for equal allocation of power for all users, a single base station, and single-antenna terminals. It is important to remark that the minimization of SMSE and the maximization of sum capacity are related, although they can suffer from lack of fairness issues with less favored users [19].
For convenience, we define

$$\mathbf{A}_m \triangleq \sum_{p=1}^{M}\mathbf{H}_{pm}^H\mathbf{D}_p^H\mathbf{D}_p\mathbf{H}_{pm} \in \mathbb{C}^{n\times n} \tag{9}$$

$$\mathbf{X}_m \triangleq \mathbf{H}_{mm}^H\mathbf{D}_m^H \in \mathbb{C}^{n\times k}, \tag{10}$$

so that

$$\begin{aligned}\text{SMSE} = \sum_{m=1}^{M} \text{tr}\{\mathbf{I}_k - \mathbf{X}_m^H\mathbf{B}_m\mathbf{T}_m - \mathbf{T}_m^H\mathbf{B}_m^H\mathbf{X}_m \\ + \mathbf{T}_m^H\mathbf{B}_m^H\mathbf{A}_m\mathbf{B}_m\mathbf{T}_m + \mathbf{D}_m\mathbf{D}_m^H\}.\end{aligned} \tag{11}$$

Note that $\mathbf{A}_m$ is positive (semi)definite and can be written as $\mathbf{A}_m = \mathbf{X}_m\mathbf{X}_m^H + \mathbf{K}_m$, with $\mathbf{K}_m \triangleq \sum_{p\neq m}\mathbf{H}_{pm}^H\mathbf{D}_p^H\mathbf{D}_p\mathbf{H}_{pm}$, which is also positive (semi)definite.

The SMSE in (8), or equivalently (11), is the starting point to explore several solutions for the precoding matrices $\{\mathbf{T}_m\}_{m=1}^M$ and BFN weights $\{\mathbf{B}_m\}_{m=1}^M$, each targetting different constraints. Under the proposed aggregated MSE framework, all the involved coefficients in the transmission process would be the result of minimizing the overall MSE:

$$\begin{aligned}\text{(P1)} \quad \{\mathbf{T}_m, \mathbf{B}_m, \mathbf{D}_m\}_{m=1}^M &= \arg\min \sum_{m=1}^{M}\text{tr}\{\mathbf{E}_m\} \\ \text{s. to } &\text{tr}\{\mathbf{B}_m\mathbf{T}_m\mathbf{T}_m^H\mathbf{B}_m^H\} \leq P_m,\end{aligned} \tag{12}$$

with P_m the power allocated to the m-th cluster. This is the power limit for each group of n antenna feeds, with the overall available power at the satellite is $P = \sum_{m=1}^M P_m$. This power restriction is based on an active antenna array with multi-port amplifiers [20], so that the power can be shared among those amplifiers serving a given cluster. As opposed, fixed power per feed restrictions should be imposed if independent amplifiers are used instead. Following [21], the case of power per-feed constraints would give rise to n restrictions in (12) rather than one. In case feeds are shared among more than one cluster, a conservative approach would impose a power cap for those feeds equal to the maximum power per feed divided by the number of clusters making use of that feed. Thus, problem (P1) can still be decoupled into M different optimization problems, avoiding the interplay of restrictions. In the remaining of the paper we will consider the power per cluster restriction as in (12).

We are interested in addressing the separate optimization of $\{\mathbf{T}_m\}_{m=1}^M$ and $\{\mathbf{B}_m\}_{m=1}^M$, since the flexibility and amount of CSI is not necessarily the same on-board the satellite and on-ground. Only for the OGBF case, with all weights operated at the gateways, a joint $\{\mathbf{B}_m\mathbf{T}_m\}_{m=1}^M$ matrix will be considered.

No closed-form expressions seem to be available for variables $\{\mathbf{D}_m, \mathbf{T}_m, \mathbf{B}_m\}_{m=1}^M$ minimizing the SMSE in (12) under transmit power constraints, so a multistage approach is explored. Initially we will assume that perfect knowledge is available to obtain the optimum weights, with practical constraints being imposed later to come up with a fixed BFN and no exchange of signalling information among gateways.

III. On-Ground Beamforming

For setting a reference we start with the most favorable case, for which CSIT is perfectly known and all coefficients can be correspondingly adjusted. Joint adaptation of precoding and beamforming coefficients can be applied if their combined operation takes place at the ground stations, by using the OGBF technology detailed in the Introduction. All coefficients can be directly manipulated on the ground, at the price of a higher number of exchanged signals with the satellite, one per feed managed by the corresponding gateway. This joint on-ground design is also known as precoding in the feed space [11], as opposed to beam space precoding; the latter is the type of precoding considered in this paper when combined with on-board beamforming.

The grouping of $\mathbf{B}_m$ and $\mathbf{T}_m$ in (8) leads us to write $\mathbf{F}_m \triangleq \mathbf{B}_m\mathbf{T}_m$, and optimize directly with respect to these $\mathbf{F}_m$ matrices. Since no closed-form solution seems to be available, we optimize cyclically with respect to $\{\mathbf{F}_m\}_{m=1}^M$ keeping $\{\mathbf{D}_m\}_{m=1}^M$ fixed, and then with respect to $\{\mathbf{D}_m\}_{m=1}^M$

keeping $\{\mathbf{F}_m\}_{m=1}^M$ fixed. In this way, convergence in the cost is guaranteed [22]. With fixed $\{\mathbf{D}_m\}_{m=1}^M$, the optimization decouples into M separate problems, for $m = 1, \ldots, M$:

$$\begin{aligned}\text{(P2)} \quad \mathbf{F}_m \;=\; & \arg\min \mathrm{tr}\{\mathbf{I}_k - \mathbf{X}_m^H\mathbf{F}_m \\ & -\mathbf{F}_m^H\mathbf{X}_m + \mathbf{F}_m^H\mathbf{A}_m\mathbf{F}_m + \mathbf{D}_m\mathbf{D}_m^H\} \\ & \text{s. to } \mathrm{tr}\{\mathbf{F}_m\mathbf{F}_m^H\} \le P_m. \end{aligned} \tag{13}$$

The minimization of (13) reduces to a Least Squares problem with a quadratic inequality constraint [23]. The solution is found as follows, for $m = 1, \ldots, M$ and $M > 1$:

- First, check whether the unconstrained solution $\mathbf{F}_m = \mathbf{A}_m^\dagger\mathbf{X}_m$ is feasible; if so, stop.
- Otherwise, the constraint is satisfied with equality; the solution is $\mathbf{F}_m = (\mathbf{A}_m + \nu_m\mathbf{I}_n)^{-1}\mathbf{X}_m$, where the Lagrange multiplier ν_m has to be numerically computed to meet the power constraint $\mathrm{tr}\{\mathbf{F}_m\mathbf{F}_m^H\} = P_m$, see Appendix A.

Now we fix the matrices $\{\mathbf{F}_m\}_{m=1}^M$ and need to find $\{\mathbf{D}_m\}_{m=1}^M$. We rewrite the SMSE (8) as

$$\begin{aligned}\mathrm{SMSE} = \sum_{m=1}^M \mathrm{tr}\Big\{&\mathbf{I}_k - \mathbf{D}_m\mathbf{H}_{mm}\mathbf{F}_m - \mathbf{F}_m^H\mathbf{H}_{mm}^H\mathbf{D}_m^H \\ &+\mathbf{D}_m\left(\sum_{p=1}^M \mathbf{H}_{mp}\mathbf{F}_p\mathbf{F}_p^H\mathbf{H}_{mp}^H\right)\mathbf{D}_m^H + \mathbf{D}_m\mathbf{D}_m^H\Big\}\end{aligned} \tag{14}$$

and define

$$\mathbf{C}_m \triangleq \mathbf{I}_k + \sum_{p=1}^M \mathbf{H}_{mp}\mathbf{F}_p\mathbf{F}_p^H\mathbf{H}_{mp}^H \in \mathbb{C}^{k\times k} \tag{15}$$

$$\mathbf{G}_m \triangleq \mathbf{F}_m^H\mathbf{H}_{mm}^H \in \mathbb{C}^{k\times k}, \tag{16}$$

so that we obtain the following compact expression for the SMSE:

$$\mathrm{SMSE} = \sum_{m=1}^M \mathrm{tr}\{\mathbf{I}_k - \mathbf{D}_m\mathbf{G}_m^H - \mathbf{G}_m\mathbf{D}_m^H + \mathbf{D}_m\mathbf{C}_m\mathbf{D}_m^H\}. \tag{17}$$

The minimization of (17) subject to $\mathbf{D}_m$ being diagonal is straightforward: if we let $\mathbf{D}_m = \mathrm{diag}\{\ d_1^{(m)} \ \cdots \ d_k^{(m)}\ \}$, then

$$d_j^{(m)} = \frac{[\mathbf{G}_m]_{jj}}{[\mathbf{C}_m]_{jj}}, \qquad j = 1, \ldots, k, \quad m = 1, \ldots, M. \tag{18}$$

The **single gateway case** ($M = 1$) offers the best possible performance since all streams can have access to all feeds, achieving a better attenuation of the co-channel interference. Even further, if the scaling parameter is the same for all terminals, with $\mathbf{D} = (1/\sqrt{t})\mathbf{I}_K$, then the previous mathematical derivations can be simplified. Thus, $\mathbf{F}$ in (13) is simply given by

$$\mathbf{F} = \sqrt{t}(\mathbf{H}^H\mathbf{H} + \gamma\mathbf{I}_N)^{-1}\mathbf{H}^H \tag{19}$$

and $\gamma = K/P$. This result is already reported in [10], and can be proved by using the eigen-value decomposition of the channel Gramian $\mathbf{H}^H\mathbf{H} = \mathbf{U}\mathbf{S}\mathbf{U}^H$ and similar steps to those exposed in the next section.

IV. On-Board Beamforming

Satellites with on-board beamforming rather than on-ground beamforming are very common in practice: one stream per beam, not per feed, needs to be exchanged with the ground stations. The flexibility degree of the BFN may differ from one case to another. Fully adaptive OBBF weights turn out to be highly challenging from the implementation point of view and, as a general rule, the adaptation time scale of BFN weights is more constrained than that of ground precoding weights [2]. This is why we address separately the adaptation of BFN and precoding weights in this section, in an effort to leverage their separate roles and eventually design a fixed BFN or with a limited degree of programmability.
The complexity of (P1) is such that no closed-form expressions can be jointly obtained. For practical power considerations, we assume that the beamforming matrices are semi-unitary, with orthonormal columns:

$$\mathbf{B}_m^H\mathbf{B}_m = \mathbf{I}_k, m = 1, \ldots, M. \tag{20}$$

This condition is required to keep the power constant between the input and the output of the BFN, and as such is known as a lossless condition [15]. With this, the power constraint in (P1) can be written as $\mathrm{tr}\{\mathbf{T}_m\mathbf{T}_m^H\} \le P_m$.

We need to highlight that any rank-k $\mathbf{F}_m \in \mathbb{C}^{n\times k}$ can be non-uniquely factorized as $\mathbf{F}_m = \mathbf{B}_m\mathbf{T}_m$, with invertible $\mathbf{T}_m \in \mathbb{C}^{k\times k}$ and $\mathbf{B}_m \in \mathbb{C}^{n\times k}$ with orthonormal columns. In consequence, under full adaptation capabilities and transparent feeder link, the optimal performance of OGBF and OBBF would be the same: based on the optimal OGBF design from the previous section, it would suffice to pick $\mathbf{B}_m$ and $\mathbf{T}_m$ from such factorization of $\mathbf{F}_m$ as the BFN matrix and the precoding matrix, respectively. However, this approach is not amenable to limited adaptation capabilites of the BFN, since a fixed $\mathbf{B}_m$ and a fully adaptive $\mathbf{T}_m$ cannot be conciliated under an adaptive $\mathbf{F}_m$. Since the ultimate goal is to obtain a coarse fixed BFN, we decouple the design of the BFN $\mathbf{B}_m$ and the precoder $\mathbf{T}_m$, at the price of a suboptimal performance in the fully adaptive case.

For a given set of fixed beamforming weights $\mathbf{B}_m$, the ground precoders $\mathbf{T}_m$ can be particularized from Sec. III, by using $\mathbf{B}_m^H\mathbf{A}_m\mathbf{B}_m$ and $\mathbf{B}_m^H\mathbf{X}_m$ in lieu of $\mathbf{A}_m$ and $\mathbf{X}_m$, respectively:

$$\mathbf{T}_m = \left(\mathbf{B}_m^H\mathbf{A}_m\mathbf{B}_m + \nu_m\mathbf{I}_k\right)^{-1}\mathbf{B}_m^H\mathbf{H}_{mm}^H\mathbf{D}_m^H. \tag{21}$$

Note that the different scaling matrices $\{\mathbf{D}_m\}_{m=1}^M$ are also embedded in $\mathbf{A}_m$ as per (9). Again, a cyclic scheme can be used to compute $\{\mathbf{T}_m\}_{m=1}^M$ and $\{\mathbf{D}_m\}_{m=1}^M$ iteratively until convergence, by optimizing cyclically with respect to $\{\mathbf{T}_m\}_{m=1}^M$ keeping $\{\mathbf{D}_m\}_{m=1}^M$ fixed, and then with respect to $\{\mathbf{D}_m\}_{m=1}^M$ keeping $\{\mathbf{T}_m\}_{m=1}^M$ fixed. In the particular case in which $\mathbf{D}_m = \frac{1}{\sqrt{t_m}}\mathbf{I}_k$, i.e., the same scalar applies for all users belonging to the same cluster, then $\mathbf{T}_m$ is given by

$$\mathbf{T}_m = \sqrt{t_m}\left(\mathbf{B}_m^H(t_m\mathbf{A}_m)\mathbf{B}_m + \nu'_m\mathbf{I}_k\right)^{-1}\mathbf{B}_m^H\mathbf{H}_{mm}^H \tag{22}$$

where $\nu'_m = t_m \nu_m$. From (9), $\mathbf{A}_m$, the sum of the channel Gramians from those feeds managed by the mth gateway to users in all clusters, can be expressed as

$$\mathbf{A}_m = (1/t_m)(\mathbf{H}_{mm}^H \mathbf{H}_{mm} + \mathbf{\Sigma}_m) \tag{23}$$

with

$$\mathbf{\Sigma}_m \triangleq \sum_{\substack{p=1 \\ p \neq m}}^{M} \frac{t_m}{t_p} \mathbf{H}_{pm}^H \mathbf{H}_{pm}. \tag{24}$$

The first term in $\mathbf{A}_m$ corresponds to the intra-cluster channel, whereas the second collects the leakage channel to all other clusters. For practical reasons, the acquisition of the inter-cluster channels to make $\mathbf{\Sigma}_m$ available to gateway m is difficult to guarantee in practice. Even in the case that $\mathbf{H}_{pm}$ were known, the scaling parameters $\{t_m\}$ present in $\mathbf{\Sigma}_m$ would need coordination for their computation; a cyclic process, for instance, would obtain $\{\nu'_m\}$ for fixed values of $\{t_m\}$, then $\{t_m\}$ would be recomputed for the obtained values of $\{\nu'_m\}$, and so on. Although message passing algorithms such as in [24] could be devised for this process, they will not be considered in the sequel, since our main focus is on the autonomous operation of the gateways. The general expression (22), written as

$$\begin{aligned} \mathbf{T}_m = \; & \sqrt{t_m} \left(\mathbf{B}_m^H \mathbf{H}_{mm}^H \mathbf{H}_{mm} \mathbf{B}_m + \right. \\ & \left. \mathbf{B}_m^H \mathbf{\Sigma}_m \mathbf{B}_m^H + \gamma_m \mathbf{I}_k \right)^{-1} \mathbf{B}_m^H \mathbf{H}_{mm}^H, \end{aligned} \tag{25}$$

was first introduced in [25], where the optimum factor for γ_m was shown to be equal to $\gamma_m = k/P_m$. This multi-cluster precoder accounts for the contamination from other clusters by means of the term $\mathbf{B}_m^H \mathbf{\Sigma}_m \mathbf{B}_m^H$. It can be assimilated to a classic single MMSE precoder if we approximate $\mathbf{B}_m^H \mathbf{\Sigma}_m \mathbf{B}_m$ by $c_m \cdot \mathbf{I}_K$, with c_m a constant which can be absorbed by the regularization factor. Thus, by reducing $\mathbf{B}_m^H (t_m \mathbf{A}_m) \mathbf{B}_m$ in (22) to $\mathbf{B}_m^H \mathbf{H}_{mm}^H \mathbf{H}_{mm} \mathbf{B}_m + c_m \mathbf{I}_k$, the precoder of the mth gateway is

$$\mathbf{T}_m = \sqrt{t_m} \left(\mathbf{B}_m^H \mathbf{H}_{mm}^H \mathbf{H}_{mm} \mathbf{B}_m + \gamma_m \mathbf{I}_k \right)^{-1} \mathbf{B}_m^H \mathbf{H}_{mm}^H, \tag{26}$$

with $\gamma_m = \nu'_m + c_m$. The **regularization factor** γ_m needs to be obtained so that the contribution of the m-th gateway to the SMSE is minimized, as addressed in the following sections. The proper design of the regularization factor γ_m, as detailed next, allows the application of expressions obtained for single-cluster precoders to mitigate the impact of inter-cluster interference. If we impose $\gamma_m = k/P_m$, then the intra-cluster MMSE precoder from [10] is obtained.

As stated earlier, we need to point out that the separated optimization of the ground precoders $\{\mathbf{T}_m\}_{m=1}^M$ and the BFN $\{\mathbf{B}_m\}_{m=1}^M$ has as ultimate goal to fix the BFN and let the precoders adapt to the channel variations. This is why we do not pursue the full optimization of $\mathbf{B}_m$ in (P1) for fixed $\mathbf{T}_m$ and then enter into a sequential minimization process as that in Sec. III, but instead try to decouple the derivation of $\mathbf{B}_m$ from $\mathbf{T}_m$. Next we illustrate how to obtain the scaling t_m and the regularization factor γ_m in (26) under different restrictions on $\mathbf{B}_m$.

A. Pre-fixed BFN

The function of the BFN is to provide proper phase and amplitude excitations to the different antenna feeds [15]. These excitations are static in many practical cases, and designed under criteria not necessarily aligned with those developed in this work. If the $\{\mathbf{B}_m\}$ weights are already in-place and cannot be altered, then the parameters to optimize are $\{t_m, \gamma_m\}$ of the ground precoders in (26). The optimization problem can be posed now as

$$\begin{aligned} \text{(P3)} \quad \{t_m, \gamma_m\}_{m=1}^M = \; & \arg\min \operatorname{tr}\{\mathbf{E}_m\} \\ & \text{s. to } \operatorname{tr}\{\mathbf{T}_m \mathbf{T}_m^H\} \leq P_m. \end{aligned} \tag{27}$$

The error used for the cost function is written as

$$\begin{aligned} & \operatorname{tr}\{\mathbf{E}_m\} = \operatorname{tr}\{\mathbf{I}_k\} \\ & - 2 \operatorname{tr}\{(\mathbf{B}_m^H \mathbf{H}_{mm}^H \mathbf{H}_{mm} \mathbf{B}_m + \gamma_m \mathbf{I}_k)^{-1} \mathbf{B}_m^H \mathbf{H}_{mm}^H \mathbf{H}_{mm} \mathbf{B}_m\} \\ & + \operatorname{tr}\{(\mathbf{B}_m^H \mathbf{H}_{mm}^H \mathbf{H}_{mm} \mathbf{B}_m + \gamma_m \mathbf{I}_k)^{-1} \mathbf{B}_m^H (\mathbf{H}_{mm}^H \mathbf{H}_{mm} + \mathbf{\Sigma}_m) \\ & \cdot \mathbf{B}_m (\mathbf{B}_m^H \mathbf{H}_{mm}^H \mathbf{H}_{mm} \mathbf{B}_m + \gamma_m \mathbf{I}_k)^{-1} \cdot \mathbf{B}_m^H \mathbf{H}_{mm}^H \mathbf{H}_{mm} \mathbf{B}_m\} \\ & + \operatorname{tr}\left\{ \frac{1}{t_m} \mathbf{I}_k \right\}. \end{aligned}$$

Note that all M minimization problems in (P3), one per gateway, are coupled through the inter-cluster term $\mathbf{\Sigma}_m$ given by (24). They can be decoupled if we assume that all $\{t_m\}$ are similar, so that message exchange among the gateways can be avoided during the optimization phase:

$$\mathbf{\Sigma}_m \approx \sum_{\substack{p=1 \\ p \neq m}}^{M} \mathbf{H}_{pm}^H \mathbf{H}_{pm}. \tag{28}$$

This looks like a reasonable assumption for a large number of users, as the results in Sec. V will show. As a consequence, the power constraints in (27) become active, with t_m taking the largest possible value. With the M problems decoupled, the regularization factor γ_m at each gateway precoder (26) can be designed so that the contribution $\operatorname{tr}\{\mathbf{E}_m\}$ to the SMSE is minimized. If this can be effectively applied, then the resulting precoder will be inter-cluster aware. The optimum value is given by the following lemma, which is proved in Appendix B.

Lemma 1: If we write the eigen-decomposition

$$\mathbf{B}_m^H \mathbf{H}_{mm}^H \mathbf{H}_{mm} \mathbf{B}_m = \mathbf{U}_m \mathbf{S}_m \mathbf{U}_m^H \tag{29}$$

with $\mathbf{S}_m = \operatorname{diag}\{ \; \lambda_1^{(m)} \quad \cdots \quad \lambda_k^{(m)} \; \}$, then the regularization factor γ_m minimizing $\operatorname{tr}\{\mathbf{E}_m\}$ is the solution of the following equation:

$$\sum_{i=1}^{k} \frac{\lambda_i^{(m)}}{(\lambda_i^{(m)} + \gamma_m)^3} \left(\gamma_m - \sigma_{ii}^{(m)} - \frac{k}{P_m} \right) = 0 \tag{30}$$

with $\sigma_{ii}^{(m)}$ the ith diagonal entry of $\mathbf{U}_m^H \mathbf{B}_m^H \mathbf{\Sigma}_m \mathbf{B}_m \mathbf{U}_m$. The corresponding scaling parameter of the precoder and receiver is given by

$$t_m = P_m / \sum_{i=1}^{k} \frac{\lambda_i^{(m)}}{(\lambda_i^{(m)} + \gamma_m)^2}. \tag{31}$$

If all $\sigma_{ii}^{(m)}$ are equal to zero, i.e., there is no inter-cluster interference, then the solution to (30) is trivially seen to be $\gamma_m = k/P_m$, similarly to [9]. More generally, it can be readily seen to lie in the interval $[k/P_m, k/P_m + \max(\sigma_{ii}^{(m)})]$. However, its derivation relies on the knowledge of $\mathbf{\Sigma}_m$, which participates in (30) through $\sigma_{ii}^{(m)}$. Even though we use the approximation in (28), the lack of coordination among gateways prevents the acquisition of the channel response from feeds serving cluster m to terminals in all other clusters; we propose instead to make use of the expected leakage channel Gramians, thus avoiding their instantaneous acquisition. The computation of the expectations can be performed off-line, by means of proper simulations which collect all the relevant random inputs such as users' locations or phases of RF chains. With this, we approximate (24) as

$$\mathbf{\Sigma}_m \approx \hat{\mathbf{\Sigma}}_m = \sum_{\substack{p=1 \\ p \neq m}}^{M} \mathbb{E}\left[\mathbf{H}_{pm}^H \mathbf{H}_{pm}\right]. \tag{32}$$

In addition to the numerical solution of (30), we will also test in the simulations the following approximation:

$$\gamma_m = k/P_m + \mathrm{tr}\{\mathbf{B}_m^H \hat{\mathbf{\Sigma}}_m \mathbf{B}_m\}/k. \tag{33}$$

The first term k/P_m is the regularization factor for intra-cluster precoders; the second term comes from approximating $\mathbf{B}_m^H \mathbf{\Sigma}_m \mathbf{B}_m$ by $c_m \mathbf{I}_k$ in (26), in such a way that the trace of both matrices is the same (for identical $\{t_m\}$ values in $\mathbf{\Sigma}_m$). The precoder computed in this way is still inter-cluster aware, having an edge with respect to intra-cluster precoders. We will show the validity of this approach in the simulations.

With the results in this section we have extended some well-known previous studies on regularization of precoders [9] to a different context, with mutually interfering clusters of beams, in what can be considered as multi-cluster precoders. The relevance of the proper tuning of the regularization factor will be highlighted in the numerical results.

B. Adaptive BFN

For those cases for which the satellite BFN weights can be optimized, at first sight we should choose $\mathbf{B}_m$ under the SMSE criterion to minimize the error term $\mathrm{tr}\{\mathbf{E}_m\}$ in (P1) as

$$\begin{aligned} \text{(P4)} \quad \{\mathbf{B}_m\}_{m=1}^M = \arg\min \mathrm{tr}\{&\mathbf{I}_k - \frac{1}{\sqrt{t_m}}\left(\mathbf{H}_{mm}\mathbf{B}_m\mathbf{T}_m\right. \\ &\left. + \mathbf{T}_m^H\mathbf{B}_m^H\mathbf{H}_{mm}^H\right) + \mathbf{T}_m^H\mathbf{B}_m^H\mathbf{A}_m\mathbf{B}_m\mathbf{T}_m \\ &+ \frac{1}{t_m}\mathbf{I}_k\} \\ \text{s. to} \quad &\mathbf{B}_m^H\mathbf{B}_m = \mathbf{I}_k. \end{aligned} \tag{34}$$

Desirably, we would like to decouple the derivation of BFN $\mathbf{B}_m$ and the precoder $\mathbf{T}_m$ as much as possible to simplify the practical implementation, so that different degrees of flexibility can be accommodated. No closed-form seems to be feasible for $\mathbf{B}_m$ minimizing inter-cluster and intra-cluster interference together with noise. For the zero-forcing version of the precoder $\mathbf{T}_m$, that is, with $\gamma_m = 0$ in (26), the intra-cluster contribution in (P4) becomes independent of $\mathbf{B}_m$, and only the inter-cluster interference and additive noise components remain. If $\mathbf{B}_m$ is designed to minimize the last term in (P4), then t_m needs to be maximized. This problem is written now as

$$\begin{aligned} \text{(P5)} \quad \{\mathbf{B}_m, t_m\}_{m=1}^M = \arg\min \mathrm{tr}&\left\{\frac{1}{t_m}\mathbf{I}_k\right\} \\ \text{s. to} \quad &\begin{cases} \mathbf{B}_m^H\mathbf{B}_m = \mathbf{I}_k, \\ \mathrm{tr}\{\mathbf{T}_m\mathbf{T}_m^H\} \leq P_m \end{cases} \end{aligned} \tag{35}$$

with $\mathbf{T}_m = \sqrt{t_m}(\mathbf{B}_m^H\mathbf{H}_{mm}^H\mathbf{H}_{mm}\mathbf{B}_m)^{-1}\mathbf{B}_m^H\mathbf{H}_{mm}^H = \sqrt{t_m}(\mathbf{H}_{mm}\mathbf{B}_m)^{-1}$. It can be readily seen that, at the optimum point, the power constraint must hold with equality, and

$$t_m = P_m / \mathrm{tr}\{(\mathbf{B}_m^H\mathbf{H}_{mm}^H\mathbf{H}_{mm}\mathbf{B}_m)^{-1}\}. \tag{36}$$

This design scheme effectively decouples the derivation of BFN and precoder, while exploiting the degrees of freedom available at the satellite to increase the resilience against the noise. As a remark, the inter-cluster leakage will be again addressed by the proper design of the regularization factor as shown later. As noted in the previous paragraph for fixed $\mathbf{B}_m$, the regularization factor γ_m in (26) has a non-trivial dependence on the channel and the BFN.

As proved in the Appendix C, the optimal precoder matrix reads as

$$\mathbf{T}_m = \sqrt{t_m}\left(\mathbf{S}_{H,m} + \gamma_m\mathbf{I}_k\right)^{-1}\mathbf{B}_m^H\mathbf{H}_{mm}^H \tag{37}$$

with $\mathbf{S}_{H,m}$ the diagonal matrix containing the k non-zero eigenvalues of the channel Gramian $\mathbf{H}_{mm}^H\mathbf{H}_{mm}$.

Note that, although we have assumed a zero forcing precoder to solve (P4), since the BFN design is thus decoupled from the precoder, a more general regularized precoder is used instead at the gateway for additional gain. As in the case for fixed $\mathbf{B}_m$, the regularization factor γ_m at each gateway precoder (37) can be designed so that its contribution $\mathrm{tr}\{\mathbf{E}_m\}$ to the SMSE is minimized. Again, the solution for t_m and γ_m is that for the fixed case in (30) and (31). With respect to the solution for an isolated cluster, k/P_m, the optimized regularization factor is higher to account for the inter-cluster leakage. We will see in the simulations that the properly chosen increment of the regularization factor is critical for the performance of the system. Again, we propose to resort to the approximation (32) to avoid the communication among gateways.

C. Null steering

Some or all the degrees of freedom of $\mathbf{B}_m$ can be used to cancel the interference posed by the mth gateway on some given off-cluster users. Inter-cluster cancellation was also addressed in [26], in this case from the ground in the absence of on-board BFN. In our setting the ground precoders can follow the design in the previous sections, and the on-board BFN can create nulls in some specific locations. These off-cluster locations to preserve interference-free could be fixed or time-varying provided that some mechanism exists

to track the corresponding channels.

If $\bar{k}$ denotes the number of users which must be protected, then $\bar{k} \leq n-k$. The rows of $\mathbf{H}$ containing the channel from the feeds allocated to the mth gateway to those selected $\bar{k}$ users are collected in $\bar{\mathbf{H}}_{mm} \in \mathbb{C}^{\bar{k}\times n}$, assumed to have full rank $\bar{k}$. (P5) reads now as

$$\text{(P6)} \quad \{\mathbf{B}_m\}_{m=1}^M = \arg\min \mathrm{tr}\{(\mathbf{B}_m^H \mathbf{H}_{mm}^H \mathbf{H}_{mm}\mathbf{B}_m)^{-1}\} \quad \text{s. to} \quad \begin{cases} \bar{\mathbf{H}}_{mm}\mathbf{B}_m = \mathbf{0}, \\ \mathbf{B}_m^H\mathbf{B}_m = \mathbf{I}_k. \end{cases} \tag{38}$$

Let the singular value decomposition of $\bar{\mathbf{H}}_{mm}$ be expressed as

$$\bar{\mathbf{H}}_{mm} = \bar{\mathbf{U}}_m \bar{\mathbf{S}}_m \bar{\mathbf{V}}_m^H, \tag{39}$$

with $\bar{\mathbf{U}}_m \in \mathbb{C}^{\bar{k}\times\bar{k}}$, $\bar{\mathbf{V}}_m \in \mathbb{C}^{n\times n}$, and $\bar{\mathbf{S}}_m = \left(\begin{array}{c|c} \ddots & \\ & \mathbf{0} \end{array} \right) \in \mathbb{C}^{\bar{k}\times n}$. The last $n-\bar{k}$ columns of $\bar{\mathbf{V}}_m$ span the null space of $\bar{\mathbf{H}}_{mm}$; let $\bar{\mathbf{V}}_m^0 \in \mathbb{C}^{n\times(n-\bar{k})}$ comprise those columns, so the cancellation can be achieved via null-space projection [17], [27], building $\mathbf{B}_m$ as

$$\mathbf{B}_m = \bar{\mathbf{V}}_m^0 \mathbf{B}_m^0. \tag{40}$$

Note the reduction in degrees of freedom: the number of rows of $\mathbf{B}_m^0 \in \mathbb{C}^{(n-\bar{k})\times k}$ is now $n-\bar{k}$ rather than n. If we define $\mathbf{Q}_m \triangleq \mathbf{H}_{mm}\bar{\mathbf{V}}_m^0 \in \mathbb{C}^{k\times(n-\bar{k})}$, then (P6) is rephrased as

$$\text{(P7)} \quad \{\mathbf{B}_m^0\}_{m=1}^M = \arg\min \mathrm{tr}\{((\mathbf{B}_m^0)^H \mathbf{Q}_m^H \mathbf{Q}_m \mathbf{B}_m^0)^{-1}\} \quad \text{s. to} \quad (\mathbf{B}_m^0)^H\mathbf{B}_m^0 = \mathbf{I}_k, \tag{41}$$

and the derivation of $\mathbf{B}_m^0$ follows, *mutatis mutandis*, the sequence in problem (P5), by working with $\mathbf{Q}_m$ and $n-\bar{k}$ instead of $\mathbf{H}_{mm}$ and n, respectively.

The use of the null space as design tool is quite standard, and has been used also in the satellite literature, for example in [26]. This nulling beamformer, in combination with a properly designed set of ground MSE precoders, performs significantly better than the zero-forcing solution of [26], which in any case must be credited to be the first reference, to the authors' knowledge, to apply these ideas to a multi-gateway satellite system. In [13] a set of ground precoders were also designed to align with the null space of some inter-cluster channels; the proposed precoders are not optimal under any prescribed metric, and require the exchange of real-time information among the gateways.

D. Coarse BFN

An accurate tracking of the channels to synthesize any of the on-board beamforming solutions exposed above is not easy to implement, especially when it is required for a permanent adaptation of the BFN weights on-board the satellite. The changes in the channel matrix $\mathbf{H}$ are due, to a large extent, to the random relative location of the users within satellite coverage. We fix the weights of the satellite BFN so that the precoders at the gateways undertake all the effort to adapt to the varying CSI, at least partially. The regularization factor of the precoders, as shown earlier, can also be judiciously chosen to avoid the exchange of information among clusters and simplify the implementation. Based on two premises, (i) the need to fix the BFN weights, and (ii) the absence of exchange of structured information among clusters, we apply the results of previous sections by using the expected behavior of the channels when needed. The following logic sequence describes the proposed design:

1) Obtain the fixed BFN as

$$\text{(P8)} \quad \{\mathbf{B}_m\}_{m=1}^M = \arg\min \mathrm{tr}\{(\mathbf{B}_m^H \mathbb{E}\left[\mathbf{H}_{mm}^H\mathbf{H}_{mm}\right]\mathbf{B}_m)^{-1}\} \quad \text{s. to} \quad \mathbf{B}_m^H\mathbf{B}_m = \mathbf{I}_k. \tag{42}$$

   The solution follows the steps in Sec. IV-B for the adaptive case, with $\mathbb{E}\left[\mathbf{H}_{mm}^H\mathbf{H}_{mm}\right]$ playing the role of $\mathbf{H}_{mm}^H\mathbf{H}_{mm}$.
2) Adapt the gateway precoders to the channel changes as

$$\mathbf{T}_m = \sqrt{t_m}\left(\mathbf{B}_m^H\mathbf{H}_{mm}^H\mathbf{H}_{mm}\mathbf{B}_m + \gamma_m\mathbf{I}_k\right)^{-1}\mathbf{B}_m^H\mathbf{H}_{mm}^H. \tag{43}$$

3) The regularization factor γ_m is computed as the solution of

$$\sum_{i=1}^{k} \frac{\lambda_i^{(m)}}{(\lambda_i^{(m)}+\gamma_m)^3}\left(\gamma_m - \sigma_{ii}^{(m)} - \frac{k}{P_m}\right) = 0 \tag{44}$$

   with $\sigma_{ii}^{(m)}$ the ith diagonal entry of $\mathbf{U}_m^H\mathbf{B}_m^H\hat{\mathbf{\Sigma}}_m\mathbf{B}_m\mathbf{U}_m$, and $\hat{\mathbf{\Sigma}}_m$ defined in (32). $\{\lambda_i^{(m)}\}$ are the k non-null eigenvalues of $\mathbf{B}_m^H\mathbf{H}_{mm}^H\mathbf{H}_{mm}\mathbf{B}_m$. The scaling parameter t_m is such that $\mathrm{tr}\{\mathbf{T}_m\mathbf{T}_m^H\} = P_m$.

Tables I and II compile the expressions of the proposed transmit precoders and satellite beamforming weights for the OBBF case. The coarse OBBF solution, computed as detailed in the previous steps, is such that the satellite BFN weights are fixed, and the ground transmitters adapt to cope with the intra-cluster interference among their respective users, based on MMSE precoders with a regularization factor tuned to reduce the leakage onto other clusters. Thus, the use of the channel Gramians reveals as instrumental to avoid the exchange of information among the ground precoders and fix the satellite BFN weights.

The implementation complexity is higher for the adaptive solutions, i.e., OBBF-adaptive and OBBF-nulling, since extra signaling and hardware capacity are required to adapt the BFN to the changes in the channel response. The ability to adapt to the instantaneous channel variations offers, as a result, better performance, as shown in the following numerical results.

V. Numerical Results

We have tested the performance of the different schemes in a Monte Carlo simulation for the specifications of a multibeam satellite antenna which uses a fed reflector antenna array with $N = 155$ feeds to exchange signals with the users. First, we tested adaptive OBBF (labeled as OBBF-adaptive), OBBF with coarse BFN (labeled as OBBF-coarse) and OBBF with pre-fixed matrix (labeled as OBBF-pre-fixed), with the OGBF solution from Sec. III as reference. As representative

TABLE I: BFN and distributed precoders

OBBF-adaptive	$\mathbf{B}_m$ built as the first k left singular vectors of $\mathbf{H}_{mm}^H \mathbf{H}_{mm}$
OBBF-nulling	$\mathbf{B}_m = \bar{\mathbf{V}}_m^0 \mathbf{B}_m^0$, with $\bar{\mathbf{V}}_m^0$ the null space of $\bar{\mathbf{H}}_{mm}$, and $\mathbf{B}_m^0$ built as the first k left singular vectors of $\bar{\mathbf{V}}_m^{0,H} \mathbf{H}_{mm}^H \mathbf{H}_{mm} \bar{\mathbf{V}}_m^0$
OBBF-coarse	$\mathbf{B}_m$ built as the first k left singular vectors of $\mathbb{E}\left[\mathbf{H}_{mm}^H \mathbf{H}_{mm}\right]$
Full precoder	$\mathbf{T}_m = \sqrt{t_m}\left(\mathbf{B}_m^H \mathbf{H}_{mm}^H \mathbf{H}_{mm} \mathbf{B}_m + \mathbf{B}_m^H \mathbf{\Sigma}_m \mathbf{B}_m^H + \frac{k}{P_m}\mathbf{I}_k\right)^{-1} \mathbf{B}_m^H \mathbf{H}_{mm}^H$
Regularized precoder	$\mathbf{T}_m = \sqrt{t_m}\left(\mathbf{B}_m^H \mathbf{H}_{mm}^H \mathbf{H}_{mm} \mathbf{B}_m + \gamma_m \mathbf{I}_k\right)^{-1} \mathbf{B}_m^H \mathbf{H}_{mm}^H$

TABLE II: Transmitter regularization factor and receiver gain

γ_m is the solution of $\sum_{i=1}^k \frac{\lambda_i^{(m)}}{(\lambda_i^{(m)}+\gamma_m)^3}\left(\gamma_m - \sigma_{ii}^{(m)} - \frac{k}{P_m}\right) = 0$	
$\sigma_{ii}^{(m)}$ is the ith diagonal entry of $\mathbf{U}_m^H \mathbf{B}_m^H \hat{\mathbf{\Sigma}}_m \mathbf{B}_m \mathbf{U}_m$	
OBBF-adaptive	$\lambda_i^{(m)}$ is the ith eigenvalue of $\mathbf{B}_m^H \mathbf{H}_{mm}^H \mathbf{H}_{mm} \mathbf{B}_m$
OBBF-nulling	$\lambda_i^{(m)}$ is the ith eigenvalue of $\mathbf{B}_m^H \bar{\mathbf{V}}_m^{0,H} \mathbf{H}_{mm}^H \mathbf{H}_{mm} \bar{\mathbf{V}}_m^0 \mathbf{B}_m$
OBBF-coarse	$\lambda_i^{(m)}$ is the ith non-null eigenvalue of $\mathbf{B}_m^H \mathbf{H}_{mm}^H \mathbf{H}_{mm} \mathbf{B}_m$
$t_m = P_m / \sum_{i=1}^k \frac{\lambda_i^{(m)}}{(\lambda_i^{(m)}+\gamma_m)^2}$	

TABLE III: User link budget parameters

Parameter	Value
Satellite height	35,786 Km
Number of feeds	155
Number of beams	100
Number of simultaneous users	100 (one per beam)
Number of gateways	10
Carrier frequency	20 GHz
Bandwidth	500 MHz
Number of polarizations	1
Terminal G/T	17.68dB/K

example we have chosen the radiation pattern provided by the European Space Agency (ESA) and used in different projects and publications by researchers cooperating in Europe with ESA, see, e.g., [10] and [13]. This radiation pattern is designed to limit the level of interference among users in systems with conservative frequency reuse and a single gateway. As opposed to this, we assume that the whole available bandwidth is used by all beams, resulting in high intra-cluster and inter-cluster interference levels. The most relevant parameters of the user link budget are detailed in Table III.

For the BFN provided by ESA, Fig. 3 shows the empirical probability density function (pdf) of the signal to interference ratio (SIR) without precoding for full-frequency reuse, obtained from evaluating the interference for the different users and 100 realizations. For each realization the channel response to 100 randomly located users, one per cluster, is generated. As expected, many users suffer from high interference, especially those which happen to be near the edge of the corresponding beam, given that this BFN is suited for the low co-channel interference associated to a conservative frequency reuse across beams. In the setting under study, the feeder link is shared by $M = 10$ gateways, with the corresponding clusters shown in Fig. 4. Clusters are groups of ten beams ($k = 10$). Each gateway uses only a subset of n feeds, which is chosen by maximizing the average gain for all users in the cluster; this is performed by maximizing the Frobenius norm of $\mathbb{E}\left[\mathbf{H}_{mm}\right]$ for cluster m. The allocated power to all clusters is the same,

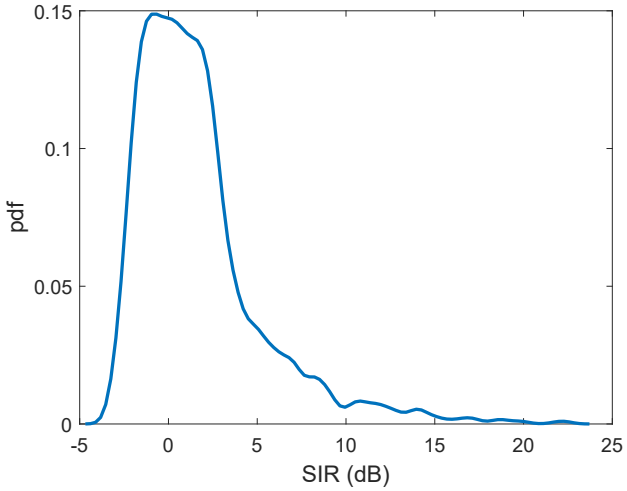


Fig. 3: Empirical SIR pdf without precoding, pre-fixed BFN.

$P_m = P/M$, with P the satellite available transmit power. We assume that the different feeder links are transparent, neglecting the possible impairments in the communication between the gateways and the satellite. The randomness of the Monte Carlo simulation comes from the location of the users at the $K = 100$ spot-beams; these locations are chosen from independent uniform distributions inside the different beams, with 100 users being served at each realization, and independently across realizations.

In order to compare the performance of different schemes the operation point needs to be calibrated. This is set by defining the signal to noise ratio (SNR) as

$$\mathrm{SNR} = \mathbb{E}\left[\mathrm{tr}\{\mathbf{H}\mathbf{F}\mathbf{F}^H\mathbf{H}^H\}\right]/K \tag{45}$$

and $\mathbf{F}$ the transmit beamforming matrix $\mathbf{F} = \frac{\sqrt{P}}{\sqrt{\mathrm{tr}\{\mathbf{H}^H\mathbf{H}\}}}\mathbf{H}^H$. Fig. 5 presents the overall achievable rate, aggregated across all beams and averaged after 200 realizations. The rate is computed as the product of the available bandwidth and

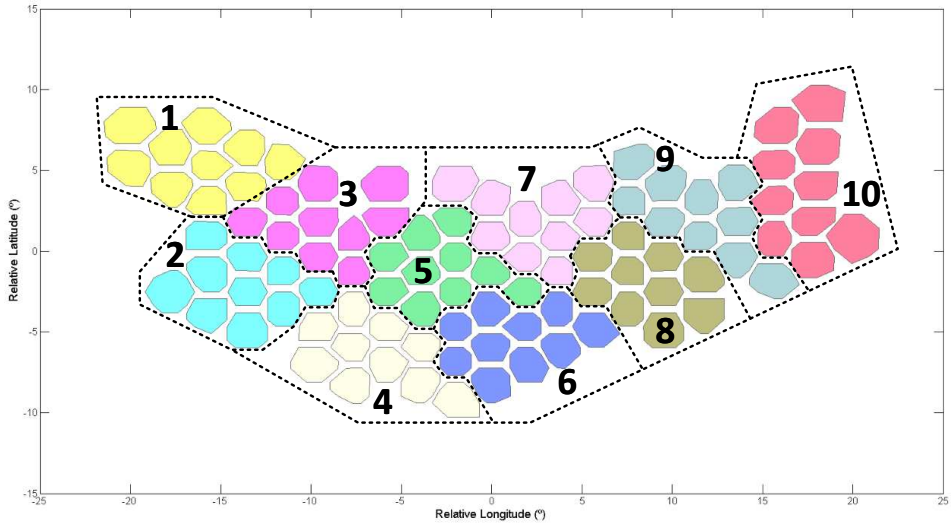


Fig. 4: Spot-beams are grouped into clusters.

$\log_2(1+\text{SINR})$. Results have been obtained for two different number of feeds per cluster, $n = 16$ and $n = 30$. Even though all feeds are available to serve any cluster, the radiation pattern is such that no more than 35 feeds provide significant content to a given cluster. The performance is upper bounded by the OGBF scheme, applicable if the gateways have access to the different feeds without an intermediate BFN and the CSIT is perfect. Even further, the OGBF bound for the single gateway case, i.e., $M = 1$ and different scaling parameters for all users, is also included to illustrate the performance loss due to the lack of data exchange among gateways. All the other curves use an on-board BFN, either adaptive or fixed; the latter uses either the coarse design in Sec. IV-D or the BFN provided by ESA for the four-color reuse scheme. As expected, performance improves if more feeds are assigned to each gateway, keeping in mind that feeds can be shared by different gateways. There is a significant loss from the OBBF-adaptive with respect to the OGBF scheme, which increases with the SNR, due to the separate optimization of precoding and BFN. It can also be noticed that the design of a specific BFN fixed matrix as part of a global multi-gateway interference cancellation scheme provides a gain with respect to a BFN not specifically designed with this in mind, for $n = 30$, whereas for $n = 16$ this gain is barely noticeable. It is left for additional studies whether robustness can be preserved for alternative designs of the BFN able to address the aggregated inter-cluster interference, rather than the noise or the interference to a specific set of users. The empirical pdf of SINR is also shown in Fig. 6 for both OBBF-adaptive and OGBF schemes, with $n = 30$ and SNR = 12dB.

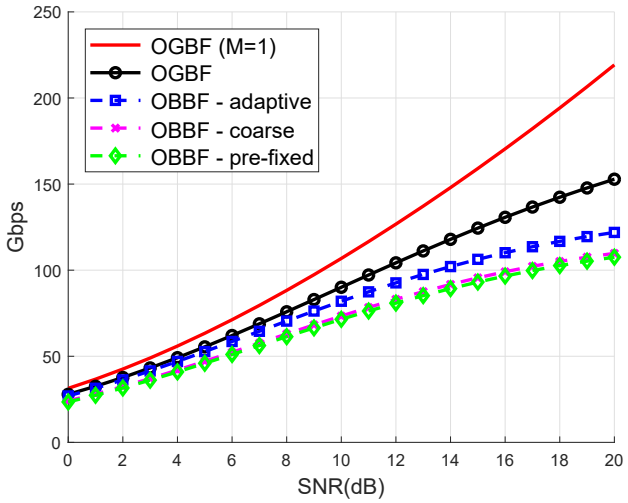
As illustration of the role played by the regularization factor in the precoding process, we have also compared the use of different regularization factors in the computation of the precoder for the adaptive BFN case (OBBF-adaptive):

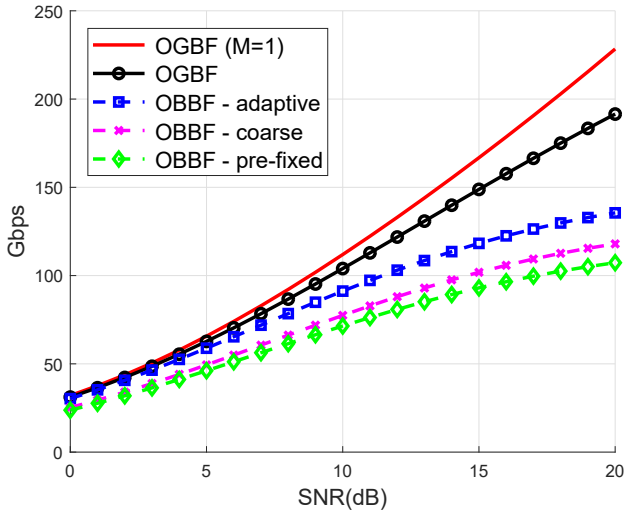
1) $\gamma_m = k/P_m$. This is the regularization factor minimizing the MSE for a single cluster, as it is well established in the literature [9].
2) γ_m the numerical solution of (30), with $\sigma_{ii}^{(m)}$ the ith diagonal entry of $\mathbf{U}_m^H\mathbf{B}_m^H\hat{\mathbf{\Sigma}}_m\mathbf{B}_m\mathbf{U}_m$. This is the regularization factor which has been used to obtain results in Fig. 5, which does not require inter-gateway cooperation, since $\hat{\mathbf{\Sigma}}_m$ in (32) is based on average statistics.
3) Full precoder as in (25).
4) Full precoder with $\mathbf{\Sigma}_m$ approximated as (32).

The expectations have been approximated empirically. The performance of the two following alternatives for γ_m is not included since it is almost indistinguisable from that of the numerical solution of (30): (i) the closed-form expression (33), and (ii) the numerical solution of (30), with $\sigma_{ii}^{(m)}$ the ith diagonal entry of $\mathbf{U}_m^H\mathbf{B}_m^H\mathbf{\Sigma}_m\mathbf{B}_m\mathbf{U}_m$, and $\mathbf{\Sigma}_m$ approximated as (28). As depicted in Fig. 7, which shows the impact of the regularization factor, the inter-cluster solution k/P_m falls short of being effective in the presence of intra-cluster interference. The availability of intercluster CSI does not improve the performance of the regularized precoder, although some gain can be obtained by using the full precoder (25); in such a case, the lack of intercluster real time information has some impact.

Lastly, we have checked the dispersion of the scalars $\{t_m\}$. In order to avoid the interaction among clusters, we have assumed throughout the paper that their values are similar. Otherwise an iterative process to solve the multiple dependencies among $\{\mathbf{T}_m, \gamma_m, \mathbf{\Sigma}_m, t_m\}$ would require the sequential exchange of information among the gateways, and make their autonomous optimization unfeasible. The average ratio of the maximum to minimum values was checked to be lower than 2 for the two settings addressed in the previous simulations, supporting the allocation of similar weights to all the inter-cluster Gramians in (28).

For the purpose of evaluation of the relative merits of the proposed OGBF and OBBF (with regularized precoder)





(b) $n = 30$.

Fig. 5: System capacity, 200 realizations, $M = 10, k = 10$. Upper bound: OGBF, one gateway. Lower bound: fixed BFN provided by ESA, distributed precoders.

techniques, we have computed their empirical averaged spectral efficiency with respect to the previous state-of-the art, in particular:

- Non-precoded scheme. Conventional systems apply a frequency planning such that no neighboring beams share the same frequency channels, so that the co-channel interference is reduced. We have divided the available spectrum into four equal parts (colors), and applied a frequency re-use (FR) pattern used as baseline in references such as [10] and [26]. The BFN matrix is that provided by ESA and labeled as pre-fixed in Fig. 5. The method is labeled as *No-Prec, FR=4*.
- On-ground zero-forcing cancelling precoder for the multiple gateway case, proposed in [26] under a Zero Forcing criterion. The performance of this method can be improved by applying a regularization factor; the corresponding method is labeled as *OGBF-nulling-Regularized*.

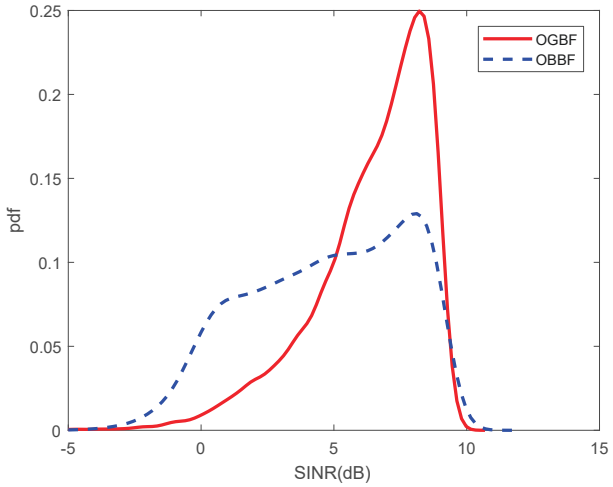
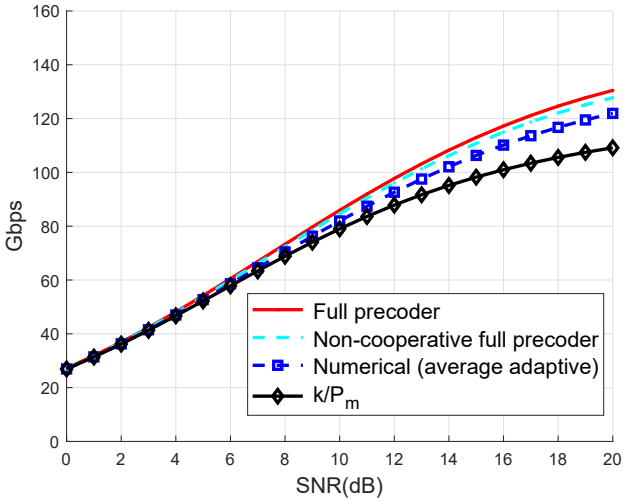
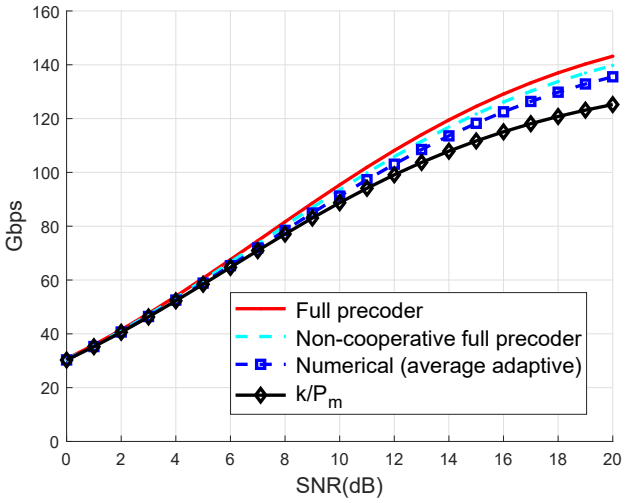


Fig. 6: Empirical SINR pdf's based on 200 realizations. SNR = 12dB, $M = 10, k = 10, n = 30$.





(b) $n = 30$.

Fig. 7: System capacity, 200 realizations, $M = 10, k = 10$. Performance with different regularization factors.

Fig. 8 shows the overall achievable rate, aggregated across all beams and averaged after 200 realizations. The available bandwidth -only one fourth in the case of the non-precoded scheme- multiplies $\log_2(1 + \text{SINR})$. In addition to the non-precoded case with FR = 4 and the nulling method proposed in [26], we have also plotted the performance of the OBBF-nulling method of Sec. IV-C. Results are shown with respect to the satellite transmit power P, and correspond to the same operating points as those in Fig. 5. As noted, all methods improve the performance of the FR = 4 scheme; the cost is the need to acquire the CSI at the gateways and apply the proper precoding. As to the nulling methods, we need to make the following remarks:

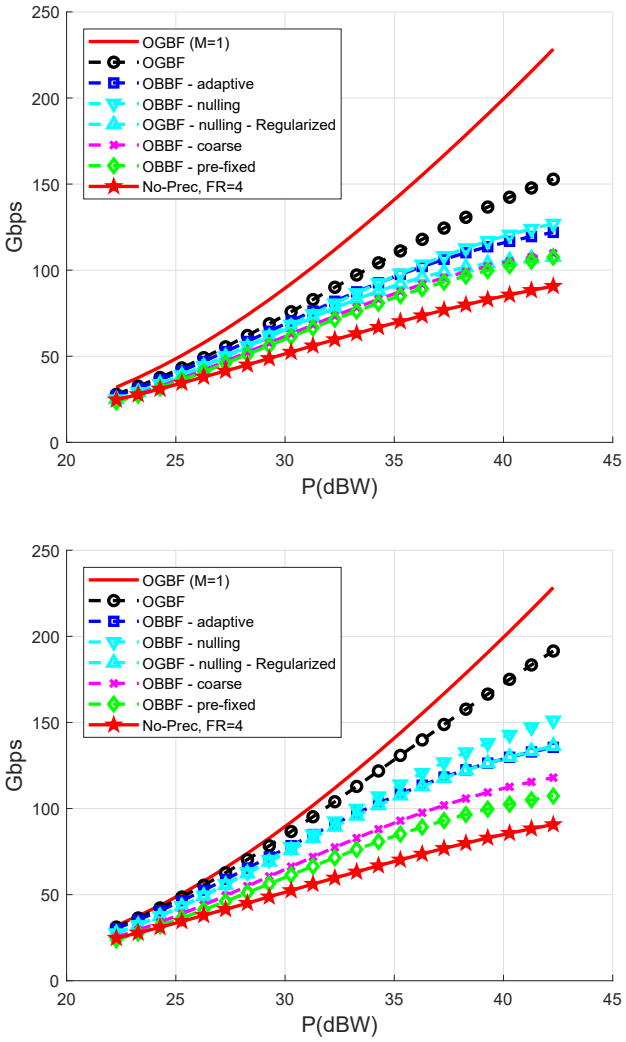
(i) Perfect knowledge of the channel from gateway m to the users in clusters adjacent to cluster m is assumed.
(ii) The number $\bar{k}$ of off-cluster users to protect has been chosen empirically for better performance. As heuristical rule, we have obtained that $\bar{k} = (n - k)/2$.
(iii) Nulls have been created on those $\bar{k}$ users maximizing the Frobenius norm of the matrix $\bar{\mathbf{H}}_{mm}$ introduced in Sec. IV-C for the mth gateway.
(iv) The OGBF nulling scheme in [26] is only included for the modified regularized case, since the original ZF version cannot match the performance of the other schemes for the chosen number of feeds per cluster. The corresponding regularization factor is k/P_m.

Although not shown, results have been also obtained for higher values of n; those methods creating nulls in specific off-cluster users benefit from such an increase, since additional degrees of freedom can be exploited. However, in addition to the required adaptivity of the BFN based on the instantaneous off-cluster channel knowledge, the number of signals to exchange through the feeder link increases accordingly, and all together make these schemes less appealing for practical purposes.

Numerical results have been obtained for a single polarization; the availability in a practical case of an additional polarization would increase the overall throughput for all the candidate techniques, since more bandwidth per beam can be exploited or the inter-beam interference gets reduced. In particular, the inter-play of the two polarizations with the multi-gateway multibeam deserves some attention to exploit in the best possible way the allocation of both polarizations to beams and gateways; this is left for future studies.

VI. Conclusions

The mitigation of co-channel interference in multibeam satellite settings has been addressed in this paper, for the case of several ground stations using the satellite to relay their signals to their respective clusters of beams. Both sources of interference, intra- and inter-cluster, are attenuated by deriving distributed on-ground linear precoders and on-board beamforming weights under a global MSE metric. Practical guidelines have been provided to design the on-ground precoders and on-board beamformer for different degrees of flexibility of the latter, including also a coarse fixed BFN.



(b) $n = 30, \bar{k} = 10$.

Fig. 8: System capacity for different OGBF and OBBF schemes. $M = 10, k = 10$, 200 realizations, $\bar{k}$ is the number of off-cluster users to protect. Non-precoded performance for a four-color deployment is outperformed in all cases.

Under the premise of no cooperation among gateways, the regularization factor of the MMSE precoders was obtained, first numerically and then in an approximated closed form, to account for the presence of inter-cluster leakage, generalizing existing results for a centralized precoder. For the purpose of benchmarking, an On-Groud Beamforming solution has been derived with full flexibility. Exploration of non-linear schemes of the Tomlinson-Harashima type to improve the rate, as proposed in [28] in combination with regularization at the transmit precoder, can be a topic for further improvement of the results exposed in this paper, together with the consideration of CSI errors in the design of the ground precoders. The design of the BFN can also be extended to account for the aggregated inter-cluster leakage; the key performance indicator will be the ability to mitigate the interference with a fixed design, given that a well-performing adaptive solution does not necessarily lead to a valid robust design.

ACKNOWLEDGEMENT

The authors wish to thank the anonymous reviewers for their constructive comments.

APPENDIX A
COMPUTING THE LAGRANGE MULTIPLIER FOR OGBF DESIGN

Consider matrices $\mathbf{F}, \mathbf{X} \in \mathbb{C}^{n\times k}$, and $\mathbf{K} \in \mathbb{C}^{n\times n}$ Hermitian positive (semi)definite. Let $\mathbf{A} = \mathbf{K} + \mathbf{X}\mathbf{X}^H$ with $\operatorname{rank}\mathbf{A} = r \le n$. Let $\mathbf{A} = \mathbf{U}\mathbf{\Gamma}\mathbf{U}^H$ with $\mathbf{U} \in \mathbb{C}^{n\times r}$ semi-unitary, and $\mathbf{\Gamma} = \operatorname{diag}\{\ \gamma_1 \quad \gamma_2 \quad \cdots \quad \gamma_r\}$ with the positive eigenvalues of $\mathbf{A}$. Let $\mathbf{W} \in \mathbb{C}^{n\times(n-r)}$ be semi-unitary with $\mathbf{W}^H\mathbf{U} = \mathbf{0}$. Define $\tilde{\mathbf{F}} = \mathbf{U}^H\mathbf{F}$, $\bar{\mathbf{F}} = \mathbf{W}^H\mathbf{F}$, $\tilde{\mathbf{X}} = \mathbf{U}^H\mathbf{X}$, and $\bar{\mathbf{X}} = \mathbf{W}^H\mathbf{X}$. Note that $\mathbf{F} = \mathbf{U}\tilde{\mathbf{F}} + \mathbf{W}\bar{\mathbf{F}}$ and $\mathbf{X} = \mathbf{U}\tilde{\mathbf{X}} + \mathbf{W}\bar{\mathbf{X}}$.

We claim that $\bar{\mathbf{X}} = \mathbf{0}$. To see this, note that $\mathbf{0} = \mathbf{W}^H\mathbf{A}\mathbf{W} = \bar{\mathbf{X}}\bar{\mathbf{X}}^H + \mathbf{W}^H\mathbf{K}\mathbf{W}$, and therefore $\mathbf{W}^H\mathbf{K}\mathbf{W} = -\bar{\mathbf{X}}\bar{\mathbf{X}}^H$ is negative semidefinite. But clearly $\mathbf{W}^H\mathbf{K}\mathbf{W}$ is also positive semidefinite, so we must have $\mathbf{W}^H\mathbf{K}\mathbf{W} = -\bar{\mathbf{X}}\bar{\mathbf{X}}^H = \mathbf{0}$, and hence $\bar{\mathbf{X}} = \mathbf{0}$.

Consider now the problem

$$\min_{\mathbf{F}} \operatorname{tr}\{\mathbf{F}^H\mathbf{A}\mathbf{F} - \mathbf{F}^H\mathbf{X} - \mathbf{X}^H\mathbf{F}\} \quad \text{subject to} \quad \operatorname{tr}\{\mathbf{F}^H\mathbf{F}\} \le P. \tag{46}$$

Since $\mathbf{F}^H\mathbf{X} = \tilde{\mathbf{F}}^H\tilde{\mathbf{X}} + \bar{\mathbf{F}}^H\bar{\mathbf{X}} = \tilde{\mathbf{F}}^H\tilde{\mathbf{X}}$ (because $\bar{\mathbf{X}} = \mathbf{0}$), and $\mathbf{F}^H\mathbf{F} = \tilde{\mathbf{F}}^H\tilde{\mathbf{F}} + \bar{\mathbf{F}}^H\bar{\mathbf{F}}$, (46) reads as

$$\min_{\tilde{\mathbf{F}},\bar{\mathbf{F}}} \operatorname{tr}\{\tilde{\mathbf{F}}^H\mathbf{\Gamma}\tilde{\mathbf{F}} - \tilde{\mathbf{F}}^H\tilde{\mathbf{X}} - \tilde{\mathbf{X}}^H\tilde{\mathbf{F}}\} \quad \text{subject to} \quad \operatorname{tr}\{\tilde{\mathbf{F}}^H\tilde{\mathbf{F}}\} + \operatorname{tr}\{\bar{\mathbf{F}}^H\bar{\mathbf{F}}\} \le P. \tag{47}$$

The objective in (47) does not depend on $\bar{\mathbf{F}}$, whereas $\bar{\mathbf{F}} = \mathbf{0}$ minimizes the constraint. Thus, the problem reduces to

$$\min_{\tilde{\mathbf{F}}} \operatorname{tr}\{\tilde{\mathbf{F}}^H\mathbf{\Gamma}\tilde{\mathbf{F}} - \tilde{\mathbf{F}}^H\tilde{\mathbf{X}} - \tilde{\mathbf{X}}^H\tilde{\mathbf{F}}\} \quad \text{subject to} \quad \operatorname{tr}\{\tilde{\mathbf{F}}^H\tilde{\mathbf{F}}\} \le P. \tag{48}$$

If the unconstrained solution $\tilde{\mathbf{F}} = \mathbf{\Gamma}^{-1}\tilde{\mathbf{X}}$ is feasible, then it solves (48). Note that for $\{\tilde{\mathbf{F}}, \bar{\mathbf{F}}\} = \{\mathbf{\Gamma}^{-1}\tilde{\mathbf{X}}, \mathbf{0}\}$ one has $\mathbf{F} = \mathbf{U}\mathbf{\Gamma}^{-1}\mathbf{U}^H\mathbf{X} = \mathbf{A}^\dagger\mathbf{X}$.

On the other hand, if $\tilde{\mathbf{F}} = \mathbf{\Gamma}^{-1}\tilde{\mathbf{X}}$ is not feasible, then the constraint must be satisfied with equality. The solution is readily found to be of the form $\tilde{\mathbf{F}} = (\mathbf{\Gamma} + \nu\mathbf{I}_r)^{-1}\tilde{\mathbf{X}}$, where $\nu \ge 0$ is the Lagrange multiplier, and the constraint reads as

$$\operatorname{tr}\{\tilde{\mathbf{F}}^H\tilde{\mathbf{F}}\} = \operatorname{tr}\{\tilde{\mathbf{X}}^H(\mathbf{\Gamma} + \nu\mathbf{I}_r)^{-2}\tilde{\mathbf{X}}\} = P. \tag{49}$$

For $i = 1, \ldots, r$, let $\tilde{\mathbf{x}}_i^H$ be the i-th row of $\tilde{\mathbf{X}}$, i.e., $\tilde{\mathbf{X}} = [\ \tilde{\mathbf{x}}_1 \quad \cdots \quad \tilde{\mathbf{x}}_r\ ]^H$. Then

$$\operatorname{tr}\{\tilde{\mathbf{X}}^H(\mathbf{\Gamma} + \nu\mathbf{I}_r)^{-2}\tilde{\mathbf{X}}\} = \sum_{i=1}^{r} \frac{\tilde{\mathbf{x}}_i^H\tilde{\mathbf{x}}_i}{(\gamma_i + \nu)^2} = \phi(\nu). \tag{50}$$

Thus, the Lagrange multiplier must satisfy $\phi(\nu) = P$. Note that $\phi(0) > P$ (or the unconstrained solution would be feasible) and that $\phi(\nu)$ is monotone decreasing for $\nu > 0$, with $\lim_{\nu\to\infty}\phi(\nu) = 0$. Hence, there is a unique positive solution of $\phi(\nu) = P$, which can be found by bisection, Newton's method, or any root-finding technique. Using the facts that $(\mathbf{A} + \nu\mathbf{I}_n)^{-1} = \mathbf{U}^H(\mathbf{\Gamma} + \nu\mathbf{I}_r)^{-1}\mathbf{U}^H + \nu^{-1}\mathbf{W}\mathbf{W}^H$ and $\mathbf{W}^H\mathbf{X} = \mathbf{0}$, it follows that the solution $\mathbf{F} = \mathbf{U}^H\tilde{\mathbf{F}} + \mathbf{W}^H\bar{\mathbf{F}} = \mathbf{U}^H(\mathbf{\Gamma} + \nu\mathbf{I}_r)^{-1}\mathbf{U}^H\mathbf{X}$ can be written as $\mathbf{F} = (\mathbf{A} + \nu\mathbf{I}_n)^{-1}\mathbf{X}$.

APPENDIX B
DERIVATION OF REGULARIZATION FACTOR FOR ON-GROUND PRECODERS

We start with the decomposition $\mathbf{B}_m^H\mathbf{H}_{mm}^H\mathbf{H}_{mm}\mathbf{B}_m = \mathbf{U}_m\mathbf{S}_m\mathbf{U}_m^H$, which for the adaptive BFN in Sec. IV-B boils down to $\mathbf{S}_{H,m}$ in (58). The error term (7) is expressed as

$$\begin{aligned}\operatorname{tr}\{\mathbf{E}_m\} \;=\; & \operatorname{tr}\{\mathbf{I}_k\} \\ & - 2\operatorname{tr}\{(\mathbf{U}_m\mathbf{S}_m\mathbf{U}_m^H + \gamma_m\mathbf{I}_k)^{-1}\mathbf{U}_m\mathbf{S}_m\mathbf{U}_m^H\} \\ & + \operatorname{tr}\{(\mathbf{U}_m\mathbf{S}_m\mathbf{U}_m^H + \gamma_m\mathbf{I}_k)^{-1}(\mathbf{U}_m\mathbf{S}_m\mathbf{U}_m^H + \mathbf{B}_m\mathbf{\Sigma}_m\mathbf{B}_m) \\ & \qquad \cdot (\mathbf{U}_m\mathbf{S}_m\mathbf{U}_m^H + \gamma_m\mathbf{I}_k)^{-1}\mathbf{U}_m\mathbf{S}_m\mathbf{U}_m^H\} \\ & + \operatorname{tr}\left\{\frac{1}{t_m}\mathbf{I}_k\right\}\end{aligned} \tag{51}$$

which, by using the orthonormality of $\mathbf{U}_m$, can be alternatively expressed as

$$\begin{aligned}\operatorname{tr}\{\mathbf{E}_m\} = k & - 2\operatorname{tr}\left\{(\mathbf{S}_m + \gamma_m\mathbf{I}_k)^{-1}\mathbf{S}_m\right\} \\ & + \operatorname{tr}\left\{(\mathbf{S}_m + \gamma_m\mathbf{I}_k)^{-1}(\mathbf{S}_m + \mathbf{U}_m^H\mathbf{B}_m^H\mathbf{\Sigma}_m\mathbf{B}_m\mathbf{U}_m)\right. \\ & \qquad \left.\cdot(\mathbf{S}_m + \gamma_m\mathbf{I}_k)^{-1}\mathbf{S}_m\right\} + k/t_m.\end{aligned} \tag{52}$$

The scaling t_m in the last term is obtained from the power constraint in (27) and the precoder expression (26):

$$\frac{P_m}{\operatorname{tr}\{(\mathbf{B}_m^H\mathbf{H}_{mm}^H\mathbf{H}_{mm}\mathbf{B}_m + \gamma_m\mathbf{I}_k)^{-1}\mathbf{B}_m^H\mathbf{H}_{mm}^H\mathbf{H}_{mm}\mathbf{B}_m \cdot (\mathbf{B}_m^H\mathbf{H}_{mm}^H\mathbf{H}_{mm}\mathbf{B}_m + \gamma_m\mathbf{I}_k)^{-1}\}} \tag{53}$$

or, equivalently,

$$t_m = \frac{P_m}{\operatorname{tr}\left\{(\mathbf{S}_m + \gamma_m\mathbf{I}_k)^{-1}\mathbf{S}_m(\mathbf{S}_m + \gamma_m\mathbf{I}_k)^{-1}\right\}}. \tag{54}$$

If we insert this into $\operatorname{tr}\{\mathbf{E}_m\}$, then we have the following minimization problem:

$$\begin{aligned}\gamma_m = \arg\min \sum_{i=1}^{k} & \frac{-2\lambda_i^{(m)}}{\lambda_i^{(m)} + \gamma_m} + \frac{(\lambda_i^{(m)})^2}{(\lambda_i^{(m)} + \gamma_m)^2} \\ & + \frac{\sigma_{ii}^{(m)}\lambda_i^{(m)}}{(\lambda_i^{(m)} + \gamma_m)^2} + \frac{k}{P_m}\frac{\lambda_i^{(m)}}{(\lambda_i^{(m)} + \gamma_m)^2}\end{aligned} \tag{55}$$

where $\sigma_{ii}^{(m)}$ is the ith diagonal entry of $\mathbf{U}_m^H\mathbf{B}_m^H\mathbf{\Sigma}_m\mathbf{B}_m\mathbf{U}_m$, and $\lambda_i^{(m)}$ the ith eigenvalue of $\mathbf{B}_m^H\mathbf{H}_{mm}^H\mathbf{H}_{mm}\mathbf{B}_m$. By equating to zero the derivative the relation to satisfy is

$$\sum_{i=1}^{k} \frac{\lambda_i^{(m)}}{(\lambda_i^{(m)} + \gamma_m)^3}\left(\gamma_m - \sigma_{ii}^{(m)} - \frac{k}{P_m}\right) = 0 \tag{56}$$

and the scaling parameter

$$t_m = P_m / \sum_{i=1}^{k} \frac{\lambda_i^{(m)}}{(\lambda_i^{(m)} + \gamma_m)^2}. \tag{57}$$

APPENDIX C
PRECODER DESIGN FOR THE ADAPTIVE BFN CASE

In order to maximize t_m in (36), let us write the eigenvalue decomposition

$$\mathbf{H}_{mm}^H\mathbf{H}_{mm} = \mathbf{U}_{H,m}\begin{pmatrix}\mathbf{S}_{H,m} & \mathbf{0}\\ \mathbf{0} & \mathbf{0}\end{pmatrix}\mathbf{U}_{H,m}^H, \tag{58}$$

with $\mathbf{U}_{H,m} \in \mathbb{C}^{n\times n}$ unitary, $\mathbf{S}_{H,m} = \text{diag}\{\ \lambda_1^{(m)} \ \cdots \ \lambda_k^{(m)}\ \}$, and $\lambda_i^{(m)}$ denoting the k non-zero eigenvalues of $\mathbf{H}_{mm}^H\mathbf{H}_{mm}$ in decreasing order. Note the change of notation of $\mathbf{S}_{H,m}$ with respect to $\mathbf{S}_m$ in (29), since they contain the eigenvalues of different matrices. With this, we have

$$\text{tr}\{(\mathbf{B}_m^H\mathbf{H}_{mm}^H\mathbf{H}_{mm}\mathbf{B}_m)^{-1}\} = \sum_{i=1}^{k}\frac{1}{\lambda_i(\mathbf{B}_m^H\mathbf{U}_{H,m}\left(\begin{smallmatrix}\mathbf{S}_{H,m} & \mathbf{0}\\ \mathbf{0} & \mathbf{0}\end{smallmatrix}\right)\mathbf{U}_{H,m}^H\mathbf{B}_m)} \quad (59)$$

with $\lambda_i(\mathbf{Z}), i = 1,\ldots,k$, denoting the eigenvalues of $\mathbf{Z}$ in decreasing order. $\mathbf{U}_{H,m}^H\mathbf{B}_m$ has orthonormal columns, since $\mathbf{B}_m^H\mathbf{B}_m = \mathbf{I}_k$, so we can apply Poincaré separation theorem [29], which bounds the eigenvalues of $\mathbf{B}_m^H\mathbf{U}_{H,m}\left(\begin{smallmatrix}\mathbf{S}_{H,m} & \mathbf{0}\\ \mathbf{0} & \mathbf{0}\end{smallmatrix}\right)\mathbf{U}_{H,m}^H\mathbf{B}_m$ in terms of those of $\left(\begin{smallmatrix}\mathbf{S}_{H,m} & \mathbf{0}\\ \mathbf{0} & \mathbf{0}\end{smallmatrix}\right)$ in the following way:

$$\lambda_i^{(m)} \geq \lambda_i(\mathbf{B}_m^H\mathbf{U}_{H,m}\left(\begin{smallmatrix}\mathbf{S}_{H,m} & \mathbf{0}\\ \mathbf{0} & \mathbf{0}\end{smallmatrix}\right)\mathbf{U}_{H,m}^H\mathbf{B}_m) \geq \lambda_{n-k+i}^{(m)} \quad (60)$$

so that

$$\sum_{i=1}^{k}\frac{1}{\lambda_i(\mathbf{B}_m^H\mathbf{U}_{H,m}\left(\begin{smallmatrix}\mathbf{S}_{H,m} & \mathbf{0}\\ \mathbf{0} & \mathbf{0}\end{smallmatrix}\right)\mathbf{U}_{H,m}^H\mathbf{B}_m)} \geq \sum_{i=1}^{k}\frac{1}{\lambda_i^{(m)}}. \quad (61)$$

The lower bound is achieved for $\mathbf{B}_m^H\mathbf{U}_{H,m} = \mathbf{Q}[\ \mathbf{I}_k \quad \mathbf{0}\ ]$, with $\mathbf{Q} \in \mathbb{C}^{k\times k}$ an arbitrary unitary matrix; in particular, taking $\mathbf{B}_m$ as the first k columns of $\mathbf{U}_{H,m}$ is optimal. With this solution the $n-k$ degrees of freedom provided by $\mathbf{B}_m$ are exploited to reduce the noise enhancement caused by the intra-cluster interference cancellation.

The design of the BFN is such that $\mathbf{B}_m^H\mathbf{H}_{mm}^H\mathbf{H}_{mm}\mathbf{B}_m = \mathbf{S}_{H,m}$, with $\mathbf{S}_{H,m}$ the diagonal matrix containing the k non-zero eigenvalues of the channel Gramian $\mathbf{H}_{mm}^H\mathbf{H}_{mm}$ in (58), so the precoder matrix reads as

$$\mathbf{T}_m = \sqrt{t_m}\,(\mathbf{S}_{H,m} + \gamma_m\mathbf{I}_k)^{-1}\,\mathbf{B}_m^H\mathbf{H}_{mm}^H. \quad (62)$$

REFERENCES

[1] D. Minoli, *Innovations in Satellite Communications Technology*. John Wiley & Sons, Inc., Hoboken, NJ, USA, 2015.

[2] J. Tronc, P. Angeletti, N. Song, M. Haardt, J. Arendt, and G. Gallinaro, "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.

[3] G. Gallinaro, S. Cioni, F. Vanin, O. Vidal, G. Huggins, M. Gross, S. Andrenacci, S. Chatzinotas, and E. Tirro, "Future ground beamforming," in *8th Advanced Satellite Multimedia Systems Conference and the 14th Signal Processing for Space Communications Workshop (ASMS/SPSC)*. IEEE, 2016.

[4] G. Zheng, S. Chatzinotas, and B. Ottersten, "Multi-gateway cooperation in multibeam satellite systems," in *2012 IEEE 23rd International Symposium on Personal, Indoor and Mobile Radio Communications-(PIMRC)*. IEEE, 2012, pp. 1360–1364.

[5] A. Tolli, H. Pennanen, and P. Komulainen, "Decentralized minimum power multi-cell beamforming with limited backhaul signaling," *IEEE Transactions on Wireless Communications*, vol. 10, no. 2, pp. 570–580, February 2011.

[6] A. Liu and V. Lau, "Hierarchical interference mitigation for massive MIMO cellular networks," *IEEE Transactions on Signal Processing*, vol. 62, no. 18, pp. 4786–4797, Sept 2014.

[7] S. Jin, W. Tan, M. Matthaiou, J. Wang, and K. K. Wong, "Statistical eigenmode transmission for the MU-MIMO downlink in rician fading," *IEEE Transactions on Wireless Communications*, vol. 14, no. 12, pp. 6650–6663, Dec 2015.

[8] Y. Liu, G. Y. Li, and W. Han, "Quantization and feedback of spatial covariance matrix for massive MIMO systems with cascaded precoding," *IEEE Transactions on Communications*, vol. 65, no. 4, pp. 1623–1634, April 2017.

[9] C. B. Peel, B. M. Hochwald, and A. L. Swindlehurst, "A vector-perturbation technique for near-capacity multiantenna multiuser communication-part I: channel inversion and regularization," *IEEE Transactions on Communications*, vol. 53, no. 1, pp. 195–202, Jan 2005.

[10] B. Devillers, A. Pérez-Neira, and C. Mosquera, "Joint linear precoding and beamforming for the forward link of multi-beam broadband satellite systems," in *Global Telecommunications Conference (GLOBECOM 2011), 2011 IEEE*.

[11] S. Chatzinotas, B. Ottersten, and R. De Gaudenzi, Eds., *Cooperative and Cognitive Satellite Systems*. Academic Press, 2015.

[12] T. Ahmed, E. Dubois, J.-B. Dupé, R. Ferrús, P. Gélard, and N. Kuhn, "Software-defined satellite cloud RAN," *International Journal of Satellite Communications and Networking*, 2017.

[13] V. Joroughi, M. A. Vázquez, and A. I. Pérez-Neira, "Precoding in multigateway multibeam satellite systems," *IEEE Transactions on Wireless Communications*, vol. 15, no. 7, pp. 4944–4956, July 2016.

[14] N. Ratkorn, M. Schneider, R. Gehring, and H. Wolf, "MEDUSA-A multiple feeds per beam multi spot beam antenna project," in *30th ESA Antenna Workshop, Noordwijk*, 2008, pp. 59–62.

[15] P. Angeletti and M. Lisi, "Multimode beamforming networks," in *Proceedings of the 32nd ESA Antenna Workshop on Antennas for Space Applications*, 2010, pp. 5–8.

[16] M. Sadek, A. Tarighat, and A. H. Sayed, "A leakage-based precoding scheme for downlink multi-user MIMO channels," *IEEE Transactions on Wireless Communications*, vol. 6, no. 5, pp. 1711–1721, May 2007.

[17] Y. C. Silva and A. Klein, "Linear transmit beamforming techniques for the multigroup multicast scenario," *IEEE Transactions on Vehicular Technology*, vol. 58, no. 8, pp. 4353–4367, 2009.

[18] P. Patcharamaneepakorn, S. Armour, and A. Doufexi, "On the equivalence between SLNR and MMSE precoding schemes with single-antenna receivers," *IEEE Communications Letters*, vol. 16, no. 7, pp. 1034–1037, July 2012.

[19] B. K. Chalise and L. Vandendorpe, "Optimization of MIMO relays for multipoint-to-multipoint communications: Nonrobust and robust designs," *IEEE Transactions on Signal Processing*, vol. 58, no. 12, pp. 6355–6368, 2010.

[20] T. M. Braun, *Satellite Communications Payload and System*. John Wiley & Sons, 2012.

[21] G. Zheng, S. Chatzinotas, and B. Ottersten, "Generic optimization of linear precoding in multibeam satellite systems," *IEEE Transactions on Wireless Communications*, vol. 11, no. 6, pp. 2308–2320, June 2012.

[22] P. Stoica and Y. Selén, "Cyclic minimizers, majorization techniques, and the expectation-maximization algorithm: a refresher," *IEEE Signal Processing Magazine*, vol. 21, no. 1, pp. 112–114, 2004.

[23] *Matrix Computations*, 4th ed. Johns Hopkins University Press, 2012.

[24] C. K. Wen, J. C. Chen, K. K. Wong, and P. Ting, "Message passing algorithm for distributed downlink regularized zero-forcing beamforming with cooperative base stations," *IEEE Transactions on Wireless Communications*, vol. 13, no. 5, pp. 2920–2930, May 2014.

[25] C. Mosquera, R. López-Valcarce, and T. Ramírez, "Distributed precoding systems in multi-gateway multibeam satellites," in *35th AIAA International Communications Satellite Systems Conference*. American Institute of Aeronautics and Astronautics, 2017.

[26] B. Devillers and A. Pérez-Neira, "Advanced interference mitigation techniques for the forward link of multi-beam broadband satellite systems," in *2011 Conference Record of the Forty Fifth Asilomar Conference on Signals, Systems and Computers (ASILOMAR)*. IEEE, 2011, pp. 1810–1814.

[27] Q. H. Spencer, A. L. Swindlehurst, and M. Haardt, "Zero-forcing methods for downlink spatial multiplexing in multiuser MIMO channels," *IEEE Transactions on Signal Processing*, vol. 52, no. 2, pp. 461–471, 2004.

[28] B. M. Hochwald, C. B. Peel, and A. L. Swindlehurst, "A vector-perturbation technique for near-capacity multiantenna multiuser communication-part II: perturbation," *IEEE Transactions on Communications*, vol. 53, no. 3, pp. 537–544, March 2005.

[29] C. R. J. R. A. Horn, *Matrix Analysis*, 2nd ed. Cambridge University Press, 2013.