Sum-Rate Maximization in Sub-28-GHz Millimeter-Wave MIMO Interfering Networks

Hadi Ghauch, *Student Member, IEEE*, Taejoon Kim, *Member, IEEE*, Mats Bengtsson, *Senior Member, IEEE*, and Mikael Skoglund, *Senior Member, IEEE*

***Abstract*— MIMO systems in the lower part of the millimetre-wave (mmWave) spectrum band (i.e., below 28 GHz) do not exhibit enough directivity and selectively, as compared to their counterparts in higher bands of the spectrum (i.e., above 60 GHz), and thus still suffer from the detrimental effect of interference, on the system sum rate. As such systems exhibit large numbers of antennas and short coherence times for the channel, traditional methods of distributed coordination are ill-suited, and the resulting communication overhead would offset the gains of coordination. In this paper, we propose algorithms for tackling the sum-rate maximization problem that are designed to address the above-mentioned limitations. We derive a lower bound on the sum rate, a so-called difference of log and trace (DLT) bound, shed light on its tightness, and highlight its decoupled nature at both the transmitters and receivers. Moreover, we derive the solution to each of the subproblems that we dub non-homogeneous waterfilling (a variation on the MIMO waterfilling solution), and underline an inherent desirable feature: its ability to turn-OFF streams exhibiting low SINR, and contribute to greatly speeding up the convergence of the proposed algorithm. We then show the convergence of the resulting algorithm, max-DLT, to a stationary point of the DLT bound. Finally, we rely on extensive simulations of various network configurations, to establish the fast-converging nature of our proposed schemes, and thus their suitability for addressing the short coherence interval, as well as the increased system dimensions, arising when managing interference in lower bands of the mmWave spectrum. Moreover, our results suggest that interference management still brings about significant performance gains, especially in dense deployments.**

***Index Terms*— Sub-28 GHz millimeter-wave, interference management, fast-converging algorithms, distributed optimization, difference of log and trace (DLT), non-homogeneous waterfilling, max-DLT, alternating iterative maximal separation (AIMS).**

I. Introduction

COMMUNICATION systems in the millimeter-wave (mmWave) band are one of the most promising candidate technologies for 5G systems, to address the ever-increasing demand for data-rates in cellular systems [1], [2]. The systems we consider in this work operate at the lower bands of the mmWave frequency spectrum, sub-28 GHz systems, e.g., X-band (8-12) GHz, Ku-band (12-18) GHz, and 28 GHz in the Ka band. The antenna spacing is not small enough to allow for hundreds of antennas, but rather a few tens (at most) at each transmitter/receiver. Thus, fully digital precoding is feasible, as the analog-to-digital converter power consumption is not a limiting factor. While the available bandwidth is narrower than higher mmWave frequency spectrum, sub-28 GHz systems offer several advantages over the latter: classical *narrow-band* transmission/signal processing is feasible [3], channels follow Rayleigh/Rician fading in non line-of-sight environments [3], and pilot-based channel estimation is more suitable than beam alignment and channel sounding [4], [5]. Investigations in the Ku band show that the narrow-band model is substantiated [6]. In that sense, they are *transitional* architectures, between conventional LTE architectures (where interference management is critical), and future mmWave systems believed to be in the higher end of the spectrum (that are virtually interference-free).

Interference management is less critical to mmWave communication, at 60 GHz and beyond. Indeed outdoor links operating at 60 GHz are shown to behave as *pseudo-wired*, due to their highly directional nature [7]. This results in channels whose sparsity (in terms of eigenmodes) is usually exploited for channel estimation [5], [8]. In the system under consideration however, channels and beamforming are not highly directional (for a fixed array aperture), as compared to higher mmWave bands. Moreover, the channels exhibit less *sparsity* in non line-of-sight scenarios, than their counterparts that operate at higher frequencies: narrowband/wideband channel measurements over the 9.6 GHz, 11.4 GHz, and 28.8 GHz bands, reveal that multipath components form a significant part of the received signal, in urban environments [9].

Thus, in such systems, when considering a multi-user multi-cell setup, interference is still a potentially limiting factor, and effective means of *interference management* are still be needed. While several works have focused on coordination at the MAC layer (an exhaustive survey was done in [10]), little-to-no work addresses the problem from a physical layer perspective. Performance evaluations of coordinated transmission at 28 GHz, in a realistic propagation environment, reported gains in spectral efficiency - albeit moderate [11]. Moreover, while earlier works such as [12] suggest that coordination and interference management bring about modest/little gains, for 28 GHz systems, one has to also consider additional interference inherent to (ultra) dense deployments - a key feature of 5G systems [13]. In addition, interference-limited

Manuscript received November 4, 2016; revised March 15, 2017; accepted March 20, 2017. Date of publication April 27, 2017; date of current version June 19, 2017. The work of T. Kim was supported by the Research Grant Council, Hong Kong, under Project CityU 11272516 and Project 21200714. *(Corresponding author: Hadi Ghauch.)*

H. Ghauch, M. Bengtsson, and M. Skoglund are with the ACCESS Linnaeus Center, School of Electrical Engineering, KTH Royal Institute of Technology, 10044 Stockholm, Sweden (e-mail: ghauch@kth.se; mats.bengtsson@ee.kth.se; skoglund@kth.se).

T. Kim is with the State Key Laboratory of Millimeter Wave, Department of Electronic Engineering, City University of Hong Kong, Hong Kong (e-mail: taejokim@cityu.edu.hk).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/JSAC.2017.2698779

scenarios arise due to intra-cell interference (as it is more stringent than inter-cell interference), in the case of cell edge users, and/or when employing spatial multiplexing. In such cases, neglecting interference might be suboptimal. Shedding light on the above questions is central to this work.

Multi-user multi-cell coordination is often accomplished in an iterative distributed manner, where only local Channel State Information (CSI) is needed at each Base Station (BS) and Mobile Station (MS). Such schemes employ Forward-Backward (F-B) iterations (also known as ping-pong iterations), to iteratively optimize the transmit and receive filters (Definition 1 in Sec. II). Over the last decade, there has been a huge body of distributed coordination algorithms, for traditional multi-user multi-cell networks. Moreover, they can be categorized based on the metric that is optimized: interference leakage minimization [14], [15] and max-SINR [14], minimum mean-squared error [16], [17], weighted minimum mean-squared error [18] and (weighted) sum-rate maximization [18]–[20].

Despite the abundance of such schemes, they are *ill-suited* for the problem at hand, as they require hundreds/thousands of iterations for convergence [21]. Furthermore, the number of required F-B iterations increases with the dimensions of the problem [21]. They are thus only applicable to low-mobility scenarios, because the number of F-B iterations is limited by the coherence time of the channel. Note that the above limitations become stricter in the case of mmWave systems: more antennas at the transmitter and receiver are envisioned (and thus more F-B iterations until convergence), as well as lower coherence times compared to conventional sub-6 GHz systems (and thus a lower number of allowed F-B iterations) when the same mobility is assumed.

Thus, applying such schemes to the sub-28 GHz systems under consideration, generates *communication overhead* that offsets the resulting performance gains (as the communication overhead is dominated by the number of F-B iterations). Though this limitation is critical to coordination algorithms, just a handful of works have explored it, even in conventional sub-6 GHz systems. In line with recent work, [22]–[25], investigating algorithms that operate in the low-overhead regime (where just a few F-B iterations are performed), is one of the main aims of this work. Moreover, our schemes are specifically designed to address the aforementioned limitations, by delivering superior performance under the low-overhead requirement, and increased system dimensions. With that in mind, while such schemes could equally well be applied to traditional cellular systems, their application has higher impact/relevance on the system at hand.

We address the problem of sum-rate maximization in MIMO Interfering Multiple-Access Channels (MIMO IMAC), by formulating lower bounds on the problem. In a first part, we establish that maximizing the *separability* between the signal subspace and the interference-plus-noise subspace (I+N), results in optimizing a bound on the sum-rate. In addition, we advocate the use of another separability metric, a lower bound on the sum-rate, that we refer to as a difference of log and trace (DLT) expression. We highlight the main advantages of using such an expression, namely that it yields optimization problems that decouple in both the transmit and receive filters. We derive the solution to each of the subproblems - that we dub *non-homogeneous waterfilling*, and underline its ability for stream control (by turning off streams that have low-SINR). We then propose a corresponding distributed algorithm, max-DLT, and establish its convergence to a stationary point of the DLT expression. Finally, we gear our numerical results to show the suitability of such schemes to the mmWave systems in question, by highlighting their fast-convergence and superior performance, in the larger antenna regime. We also benchmark against several well-known schemes such as max-SINR [14], (Weighted) MMSE [17], [18], and recent fast-converging approaches, such as CCP-WMMSE [23] and IWU [25].

Though the work addresses the problem at hand for a MIMO IMAC, it is equally applicable to the network-dual problem, the MIMO Interfering Broadcast Channel (IBC), and consequently to all the ensuing special cases, such as the MIMO IFC. This is further explored in the numerical results section.

Notation: we use bold upper-case letters to denote matrices, and bold lower-case denote vectors. For a given matrix $\boldsymbol{A}$, we define $\mathrm{tr}(\boldsymbol{A})$ as its trace, $\|\boldsymbol{A}\|_F^2$ as its Frobenius norm, $|\boldsymbol{A}|$ as its determinant, $\boldsymbol{A}^\dagger$ as its conjugate transpose, and $\boldsymbol{A}^{-\dagger}$ as $(\boldsymbol{A}^\dagger)^{-1}$. In addition, $\boldsymbol{A}_{(i)}$ denotes its ith column, $\boldsymbol{A}_{i:j}$ columns i to j, $\boldsymbol{A}_{(i,j)}$ element (i,j) in $\boldsymbol{A}$, $\lambda_i[\boldsymbol{A}]$ the i^{th} eigenvalue of a Hermitian matrix $\boldsymbol{A}$ (assuming the eigenvalues are sorted in decreasing order), and $v_{1:d}[\boldsymbol{A}]$ denotes the d dominant eigenvectors of $\boldsymbol{A}$. $\mathbb{S}_+^{n,n}$ (resp. $\mathbb{S}_{++}^{n,n}$) is the set of complex $n \times n$ positive semi-definite (resp. positive definite) matrices. Furthermore, $\boldsymbol{A} \succ \boldsymbol{0}$ (resp. $\boldsymbol{A} \succeq \boldsymbol{0}$) implies that $\boldsymbol{A}$ is positive definite (resp. positive semi-definite), and $\boldsymbol{A} \succ \boldsymbol{B}$ (resp. $\boldsymbol{A} \succeq \boldsymbol{B}$) implies that $\boldsymbol{A} - \boldsymbol{B} \succ \boldsymbol{0}$ (resp. $\boldsymbol{A} - \boldsymbol{B} \succeq \boldsymbol{0}$). Finally, $\boldsymbol{I}_n$ denotes the $n \times n$ identity matrix, $\{n\} = \{1, \cdots, n\}$, and $x^+ \triangleq \max\{0, x\}$.

II. System Model and Problem Formulation

Consider a system with L cells (each having one BS), where each cell is serving K MSs (Fig. 1). Each receiver (transmitter, resp.) is equipped with N (M, resp.) antennas, and decodes d data streams from each of its users ($d \leq \min(M, N)$). In the considered MIMO IMAC, transmitters (receivers, resp.) are MSs (BSs, resp.). Note that transmitters (receivers, resp.) become BSs (MSs, resp.), in the MIMO IBC scenario. Let $\mathcal{L}$ be the set of BSs, $\mathcal{K}$ the set of users served by BS $l \in \mathcal{L}$, and l_j denote the index of user $j \in \mathcal{K}$, at BS $l \in \mathcal{L}$. We denote by $\mathcal{I}$ the total set of users, i.e., $\mathcal{I} = \{l_j \mid l \in \mathcal{L},\ j \in \mathcal{K}\}$. The received signal at BS $l \in \mathcal{L}$ is given by,

$$\boldsymbol{y}_l = \sum_{i \in \mathcal{L}} \sum_{k \in \mathcal{K}} \boldsymbol{H}_{l,i_k} \boldsymbol{V}_{i_k} \boldsymbol{s}_{i_k} + \boldsymbol{n}_l, \quad l \in \mathcal{L}, \tag{1}$$

To recover the signal of user $j \in \mathcal{K}$, in cell $l \in \mathcal{L}$ (henceforth referred to as user $l_j \in \mathcal{I}$), $\boldsymbol{y}_l$ is processed with a linear filter, $\boldsymbol{U}_{l_j} \in \mathbb{C}^{N \times d}$, i.e.,

$$\tilde{\boldsymbol{s}}_{l_j} = \boldsymbol{U}_{l_j}^\dagger \boldsymbol{H}_{l,l_j} \boldsymbol{V}_{l_j} \boldsymbol{s}_{l_j} + \sum_{\substack{i \in \mathcal{L} \\ i \neq l}} \sum_{k \in \mathcal{K}} \boldsymbol{U}_{l_j}^\dagger \boldsymbol{H}_{l,i_k} \boldsymbol{V}_{i_k} \boldsymbol{s}_{i_k} + \boldsymbol{U}_{l_j}^\dagger \boldsymbol{n}_l, \quad \forall\, l_j \in \mathcal{I} \tag{2}$$

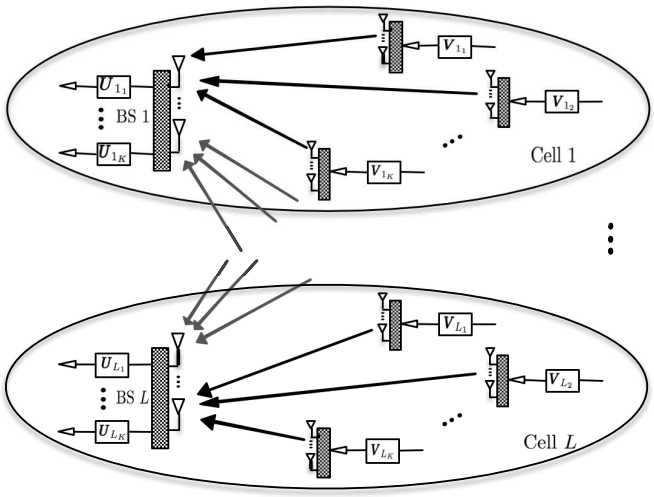


Fig. 1. L-cell MIMO interfering multiple-access channel.

where the first term represents the desired signal, the second both the intra and inter-cell interference. In the above, $\boldsymbol{s}_{i_k}$ represents the d-dimensional vector of *independently encoded symbols* for user $i_k \in \mathcal{I}$, with covariance matrix $\mathbb{E}[\boldsymbol{s}_{i_k}\boldsymbol{s}_{i_k}^\dagger] = \boldsymbol{I}_d$. In addition, $\boldsymbol{V}_{i_k}$ denotes the $M \times d$ transmit filter of user $i_k \in \mathcal{I}$, and $\boldsymbol{H}_{l,i_k}$ the $N \times M$ MIMO channel from user $i_k \in \mathcal{I}$, to BS l (assumed to be block-fading with i.i.d. entries). $\boldsymbol{n}_l$ represents the N-dimensional AWGN noise at BS $l \in \mathcal{L}$, such that $\mathbb{E}[\boldsymbol{n}_l\boldsymbol{n}_l^\dagger] = \sigma_l^2\boldsymbol{I}_N$. Note that our model (and the results presented thereafter) can easily be extended to cases where M, N and d are different across users and BSs.

If we assume that joint encoding/decoding of each user's streams is performed at the users and BSs, and treating interference as noise, the achievable rate of user $l_j \in \mathcal{I}$ is given by,

$$r_{l_j} = \log_2 |\boldsymbol{I}_d + (\boldsymbol{U}_{l_j}^\dagger \boldsymbol{R}_{l_j}\boldsymbol{U}_{l_j})(\boldsymbol{U}_{l_j}^\dagger \boldsymbol{Q}_{l_j}\boldsymbol{U}_{l_j})^{-1}|, \quad l_j \in \mathcal{I} \tag{3}$$

where $\boldsymbol{R}_{l_j}$ and $\boldsymbol{Q}_{l_j}$ are the desired signal and interference-plus-noise (I+N) covariance matrices for user j, at BS l, respectively, and are given by,

$$\boldsymbol{R}_{l_j} = \boldsymbol{H}_{l,l_j}\boldsymbol{V}_{l_j}\boldsymbol{V}_{l_j}^\dagger \boldsymbol{H}_{l,l_j}^\dagger, \quad l_j \in \mathcal{I}$$

$$\boldsymbol{Q}_{l_j} = \sum_{i=1}^{L}\sum_{k=1}^{K} \boldsymbol{H}_{l,i_k}\boldsymbol{V}_{i_k}\boldsymbol{V}_{i_k}^\dagger \boldsymbol{H}_{l,i_k}^\dagger + \sigma_l^2\boldsymbol{I}_N - \boldsymbol{R}_{l_j}, \quad l_j \in \mathcal{I}.$$

Moreover, we define

$$\bar{\boldsymbol{R}}_{i_k} = \boldsymbol{H}_{i,i_k}^\dagger \boldsymbol{U}_{i_k}\boldsymbol{U}_{i_k}^\dagger \boldsymbol{H}_{i,i_k}, \quad i_k \in \mathcal{I}$$

$$\bar{\boldsymbol{Q}}_{i_k} = \sum_{l=1}^{L}\sum_{j=1}^{K} \boldsymbol{H}_{l,i_k}^\dagger \boldsymbol{U}_{l_j}\boldsymbol{U}_{l_j}^\dagger \boldsymbol{H}_{l,i_k} + \bar{\sigma}_{i_k}^2\boldsymbol{I}_M - \bar{\boldsymbol{R}}_{i_k}, \quad i_k \in \mathcal{I}$$

as the signal and I+N covariance matrices of user i_k, in the reverse network (where $\bar{\sigma}_{i_k}^2$ is the noise variance at user i_k). Finally, we henceforth denote $\boldsymbol{L}_{l_j}\boldsymbol{L}_{l_j}^\dagger$ as the Cholesky Decomposition of $\boldsymbol{Q}_{l_j}$, and $\boldsymbol{K}_{i_k}\boldsymbol{K}_{i_k}^\dagger$ as that of $\bar{\boldsymbol{Q}}_{i_k}$. We formulate the (unweighted) sum-rate maximization problem as follows,

$$\max R_\Sigma(\{\boldsymbol{U}_{l_j}\}, \{\boldsymbol{V}_{l_j}\}) \triangleq \sum_{l\in\mathcal{L}}\sum_{j\in\mathcal{K}} r_{l_j}. \tag{4}$$

In the next section we generalize the well-known max-SINR algorithm from a stream-by-stream optimization algorithm, to an algorithm that optimizes the whole transmit/receive filter. For that purpose, we show that this generalized form can be formulated using *separability* metrics, namely, the Generalized Multi-dimensional Rayleigh Quotient (GMRQ), defined next. We next highlight the central assumptions/definitions of this work.

A. Preliminaries

The schemes we consider in the present work fall under the category of Forward-Backward training, recapped in the definition below.

Definition 1 *(F-B Training):* Schemes employing Forward-Backward (F-B) iterations (also known as ping-pong iterations, or bi-directional training), consist of optimizing the receive filters (at the BSs) in the forward training phase, then the transmit filters (at the MSs) in the reverse training phase. They exploit channel reciprocity in Time-Division Duplex (TDD) systems, and result in fully distributed algorithms. The basic iteration structure is shown in Fig. 2.

Definition 2 *(Separability):* Given two sets of points with covariance matrices $\boldsymbol{R}$ (red points) and $\boldsymbol{Q}$ (blue points), separability is a measure of the distance between the sets, after projecting on a subspace $\boldsymbol{U}$ (green plane) (shown in Fig. 3). Separability metrics - the building blocks of areas such as linear discriminant analysis [26, Ch. 4.1], include the *Generalized Multi-dimensional Rayleigh Quotient*. In the context of this work, $\boldsymbol{R}$ and $\boldsymbol{Q}$ represent the signal and I+N covariance matrices, respectively, and $\boldsymbol{U}$ the linear filter at the receiver.

Assumption 1 *(Local CSI):* We assume that each MS/BS has local CSI, i.e., each MS (resp. BS) knows the channels to its desired and interfering BSs (resp. users). Although we underline that methods in [27] are applicable for acquiring such quantities (discussion in Sect. V-C), investigating the CSI acquisition mechanism is not part of this work. Moreover, local CSI at each MS/BS is assumed to be perfectly known.

Assumption 2 *(Distributed Operation):* All schemes are required to use local CSI only, using the framework of F-B training.

Assumption 3 *(Low-Overhead Regime):* We restrict our proposed schemes to operate in the low-overhead regime, where only a small number of F-B iterations is used (in line with recent work such as [22]–[25]).

III. Separability and Sum-Rate Maximization

A. Problem Formulation

In this part, we shed light on the intimate relation between sum-rate maximization and maximization of the GMRQ separability metric. We make use of the fact that $\log|\boldsymbol{X}|$ is monotonically increasing on the positive-definite cone, i.e., $\log|\boldsymbol{X}_2| \geq \log|\boldsymbol{X}_1|$, for $\boldsymbol{X}_2 \succeq \boldsymbol{X}_1 \succ \boldsymbol{0}$. Applying the above property, we lower bound r_{l_j} in (3) as,

$$r_{l_j} > \log_2 |(\boldsymbol{U}_{l_j}^\dagger \boldsymbol{R}_{l_j}\boldsymbol{U}_{l_j})(\boldsymbol{U}_{l_j}^\dagger \boldsymbol{Q}_{l_j}\boldsymbol{U}_{l_j})^{-1}|$$

$$= \log_2 \frac{|\boldsymbol{U}_{l_j}^\dagger \boldsymbol{R}_{l_j}\boldsymbol{U}_{l_j}|}{|\boldsymbol{U}_{l_j}^\dagger \boldsymbol{Q}_{l_j}\boldsymbol{U}_{l_j}|} \triangleq \tilde{r}_{l_j}, \quad \forall l_j \in \mathcal{I} \tag{5}$$

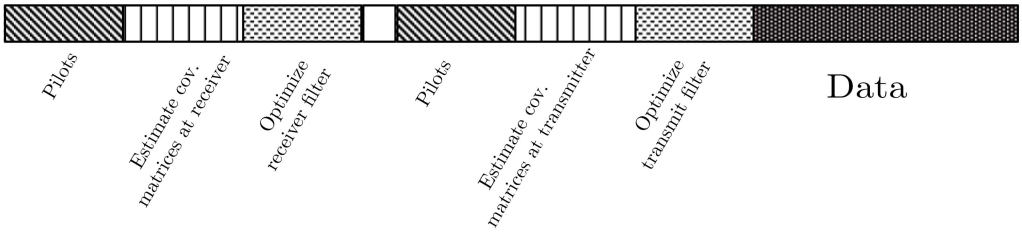
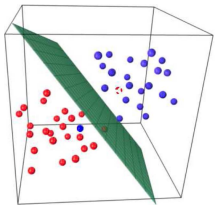


Fig. 2. Basic structure of forward-backward iteration.



$$\text{GMRQ} = \frac{|\boldsymbol{U}^\dagger \boldsymbol{R} \boldsymbol{U}|}{|\boldsymbol{U}^\dagger \boldsymbol{Q} \boldsymbol{U}|}$$

Fig. 3. Example of separability for Definition 2.

Note that $\tilde{r}_{l_j}$ is a high-SNR approximation of the actual user rate r_{l_j}, where the approximation error is $O(\text{tr}[(\boldsymbol{U}_{l_j}^\dagger \boldsymbol{Q}_{l_j} \boldsymbol{U}_{l_j})(\boldsymbol{U}_{l_j}^\dagger \boldsymbol{R}_{l_j} \boldsymbol{U}_{l_j})^{-1}])$ (refer to Appendix B). Moreover, bounds such as (5) are already prevalent in the MIMO literature (e.g., [28, Proposition 2]). Thus, the sum-rate R_Σ can be bounded below, as follows,

$$R_\Sigma > \sum_{l_j \in I} \tilde{r}_{l_j} = \log_2(\prod_{l_j} q_{l_j}), \quad \text{where } q_{l_j} \triangleq \frac{|\boldsymbol{U}_{l_j}^\dagger \boldsymbol{R}_{l_j} \boldsymbol{U}_{l_j}|}{|\boldsymbol{U}_{l_j}^\dagger \boldsymbol{Q}_{l_j} \boldsymbol{U}_{l_j}|}$$

Since $\log(-)$ is monotonic, the sum-rate maximization problem, (4), is lower bounded by,

$$(SRM) \begin{cases} \max\limits_{\{\boldsymbol{U}_{l_j}, \boldsymbol{V}_{l_j}\}} \prod\limits_{l_j \in I} q_{l_j} \\ \text{s. t. } \|\boldsymbol{U}_{l_j}\|_F^2 = P_r, \|\boldsymbol{V}_{l_j}\|_F^2 = P_t, \quad \forall\, l_j \in I \end{cases} \tag{6}$$

Remark 1 (Power Constraint): In distributed optimization schemes employing Forward-Backward (F-B) training, receivers are active in one of the phases (i.e., by sending data / pilots). Thus, generally, in the forward training phase (Definition 1), one needs a maximum transmit power constraint for the receiver filter (at each BS). This is in addition to the maximum transmit power constraint of the transmitter (at each MS), widely used in cellular systems. In addition, in scenarios involving a multi-cellular downlink communication, each BS employs a sum-power constraint for its users, e.g., [18]. However, the same does not hold in the considered setup (multi-cellular uplink), since it would lead to a sum-power constraint, across all MSs: clearly this is not applicable in practice. We thus adopt an individual per-user power constraint. We also assume equal power allocation among the users in a cell, to avoid the need for power allocation (outside the scope of this work). From a mathematical perspective however, the cost function in (SRM) renders the presence of a receive power constraint, irrelevant.

Referring to (SRM), q_{l_j} is nothing but the GMRQ separability metric in Definition 2. Consequently, given the signal and I+N covariance matrices, $\boldsymbol{R}_{l_j}$ and $\boldsymbol{Q}_{l_j}$, each receiver chooses its filter such as to maximize the separation between signal and I+N subspaces.

B. Maximization of GMRQ

The main limitation of solving problems such (SRM) is the fact it is not jointly convex in all the optimization variables. Though Block Coordinate Decent (BCD) stands out as a strong candidate, one major obstacle persists: while the problem decouples in the receive filters (as shown in (SRM)), attempting to write a similar expression by factoring out the transmit filters, leads to a coupled problem. Therefore, we propose an alternative (purely heuristic) method: the receive filters are updated as the solution to maximizing the sum-rate (assuming fixed transmit filters), while the transmit filters are chosen as the solution of the reverse network sum rate maximization (this same structure is implicitly exploited in max-SINR [14]), i.e.,

$$(SRM_F) \begin{cases} \max\limits_{\{\boldsymbol{U}_{l_j}\}} \prod\limits_{l_j \in I} q_{l_j}(\boldsymbol{U}_{l_j}) = \dfrac{|\boldsymbol{U}_{l_j}^\dagger \boldsymbol{R}_{l_j} \boldsymbol{U}_{l_j}|}{|\boldsymbol{U}_{l_j}^\dagger \boldsymbol{Q}_{l_j} \boldsymbol{U}_{l_j}|} \\ \text{s. t. } \|\boldsymbol{U}_{l_j}\|_F^2 = P_r, \quad \forall\, l_j \in I, \end{cases} \quad \text{and,}$$

$$(SRM_B) \begin{cases} \max\limits_{\{\boldsymbol{V}_{i_k}\}} \prod\limits_{i_k \in I} p_{i_k}(\boldsymbol{V}_{i_k}) = \dfrac{|\boldsymbol{V}_{i_k}^\dagger \bar{\boldsymbol{R}}_{i_k} \boldsymbol{V}_{i_k}|}{|\boldsymbol{V}_{i_k}^\dagger \bar{\boldsymbol{Q}}_{i_k} \boldsymbol{V}_{i_k}|} \\ \text{s. t. } \|\boldsymbol{V}_{i_k}\|_F^2 = P_t, \quad \forall\, i_k \in I. \end{cases}$$

In other words, assuming transmit filters as fixed, the receive filters are updated such as to maximize the separability metric in the forward phase. Similarly, the transmit filters are chosen to maximize the separability in the backward training phase. Moreover, as seen from the above problems, the objective in each subproblem is invariant to scaling of the optimal solution. Thus, they can be solved as unconstrained problems, without loss of optimality.

We first require a solution to the GMRQ maximization. The solution was earlier proposed in [29], and is not a contribution of this work. It is restated below for completeness.

Lemma 1: Consider the following maximization of the r-dimensional GMRQ,

$$\boldsymbol{X}^\star \triangleq \underset{\boldsymbol{X} \in \mathbb{C}^{n \times r}}{\text{argmax}}\ q(\boldsymbol{X}) = \frac{|\boldsymbol{X}^\dagger \boldsymbol{R} \boldsymbol{X}|}{|\boldsymbol{X}^\dagger \boldsymbol{Q} \boldsymbol{X}|}, \tag{7}$$

where $\boldsymbol{Q} \in \mathbb{S}_{++}^{n \times n}$, $\boldsymbol{R} \in \mathbb{S}_{+}^{n \times n}$ *and* $r < n$. *The set of optimal solutions to this non-convex problem are given by,*

$$\boldsymbol{X}^\star = \boldsymbol{L}^{-\dagger} \boldsymbol{\Psi} \hat{\boldsymbol{V}}, \tag{8}$$

where $\boldsymbol{L}\boldsymbol{L}^{\dagger} = \boldsymbol{Q}$ ($\boldsymbol{L} \in \mathbb{C}^{n\times n}$), $\boldsymbol{\Psi} = \upsilon_{1:r}[\boldsymbol{L}^{-1}\boldsymbol{R}\boldsymbol{L}^{-\dagger}]$ ($\boldsymbol{\Psi} \in \mathbb{C}^{n\times r}$), *and* $\hat{\boldsymbol{V}} \in \mathbb{C}^{r\times r}$ *is arbitrary and non-singular.*

Proof: It was shown in [29] that a solution to (7) is given by, $\boldsymbol{X}^{\star} = \boldsymbol{L}^{-\dagger}\boldsymbol{\Psi}$. We note that it can verified that this optimal solution is invariant to multiplication by a non-singular matrix $\hat{\boldsymbol{V}}$, i.e., $q(\boldsymbol{X}^{\star}\hat{\boldsymbol{V}}) = q(\boldsymbol{X}^{\star})$. Thus, a generic form of the solution is, $\boldsymbol{X}^{\star} = \boldsymbol{L}^{-\dagger}\boldsymbol{\Psi}\hat{\boldsymbol{V}}$ ∎

Note that the above solution is a generalized formulation of the well-known generalized eigenvalues solution. This equivalence was also established in in [29], and is restated below for convenience.

Corollary 1: Consider a special case of (8) *where* $\hat{\boldsymbol{V}} = \boldsymbol{I}_r$. *Then, this corresponds to the generalized eigenvalues solution,*

$$\boldsymbol{X}^{\star} = \boldsymbol{L}^{-\dagger}\boldsymbol{\Psi} \Leftrightarrow \boldsymbol{R}\boldsymbol{X}^{\star} = \boldsymbol{Q}\boldsymbol{X}^{\star}\boldsymbol{\Lambda}_r \tag{9}$$

where $\boldsymbol{\Lambda}_r \in \mathbb{R}^{r\times r}$ *be the (diagonal) matrix of eigenvalues for* $\boldsymbol{L}^{-1}\boldsymbol{R}\boldsymbol{L}^{-\dagger}$.

Proof: Refer to [29]. ∎

With this in mind, we can write the optimal transmit and receive filter updates, as follows,

$$\begin{aligned} \boldsymbol{U}_{l_j}^{\star} &= \boldsymbol{L}_{l_j}^{-\dagger}\boldsymbol{\Psi}_{l_j}, \quad \boldsymbol{\Psi}_{l_j} \triangleq \upsilon_{1:d}[\boldsymbol{L}_{l_j}^{-1}\boldsymbol{R}_{l_j}\boldsymbol{L}_{l_j}^{-\dagger}], \quad \forall\, l_j, \\ \boldsymbol{V}_{i_k}^{\star} &= \boldsymbol{K}_{i_k}^{-\dagger}\boldsymbol{\Theta}_{i_k}, \quad \boldsymbol{\Theta}_{i_k} \triangleq \upsilon_{1:d}[\boldsymbol{K}_{i_k}^{-1}\bar{\boldsymbol{R}}_{i_k}\boldsymbol{K}_{i_k}^{-\dagger}], \quad \forall i_k, \end{aligned} \tag{10}$$

where we used the fact we can set $\hat{\boldsymbol{V}} = \boldsymbol{I}_d$ in the solution of (8). We note that the optimal filter updates for the transmitter are more heuristic than the receiver ones: While the receive filter updates directly maximize a lower bound on the sum-rate - as seen in (SRM), no such claim can be made about the transmit filter updates. The details of our algorithm, Alternating Iterative Maximal Separation (AIMS), are shown in Algorithm 1 (where T denotes the number of F-B iterations).

Algorithm 1 Alternating Iterative Maximal Separation (AIMS)

for $t = 1, 2, \ldots, T$ **do**
  // *forward network optimization: receive filter update*
    Estimate $\boldsymbol{R}_{l_j}, \boldsymbol{Q}_{l_j}$, and compute $\boldsymbol{L}_{l_j}$, $\forall l_j$
    $\boldsymbol{U}_{l_j} \leftarrow \boldsymbol{L}_{l_j}^{-\dagger}\upsilon_{1:d}[\boldsymbol{L}_{l_j}^{-1}\boldsymbol{R}_{l_j}\boldsymbol{L}_{l_j}^{-\dagger}], \quad \forall l_j$
    $\boldsymbol{U}_{l_j} \leftarrow \sqrt{P_r}\ \boldsymbol{U}_{l_j}/\|\boldsymbol{U}_{l_j}\|_F$
  // *reverse network optimization: transmit filter update*
    Estimate $\bar{\boldsymbol{R}}_{i_k}, \bar{\boldsymbol{Q}}_{i_k}$, and compute $\boldsymbol{K}_{i_k}$, $\forall i_k$
    $\boldsymbol{V}_{i_k} \leftarrow \boldsymbol{K}_{i_k}^{-\dagger}\upsilon_{1:d}[\boldsymbol{K}_{i_k}^{-1}\bar{\boldsymbol{R}}_{i_k}\boldsymbol{K}_{i_k}^{-\dagger}], \quad \forall i_k$
    $\boldsymbol{V}_{i_k} \leftarrow \sqrt{P_t}\ \boldsymbol{V}_{i_k}/\|\boldsymbol{V}_{i_k}\|_F$
end for

Using the above solution, we next establish the result that employing unitary filters is not optimal, from the perspective of separability. We stress that the latter is not central to the main story of this work, but rather an interesting result from the separability perspective, that is obtained 'for free'. We thus restrict our presentation to sketching a proof, in Appendix VII-A.

Proposition 1: Consider the optimal receive filter given in (10), i.e., $\boldsymbol{U}_{l_j}^{\star} = \boldsymbol{L}_{l_j}^{-\dagger}\boldsymbol{\Psi}_{l_j}$, *where* $\boldsymbol{\Psi}_{l_j} = \upsilon_{1:d}[\boldsymbol{L}_{l_j}^{-1}\boldsymbol{R}_{l_j}\boldsymbol{L}_{l_j}^{-\dagger}]$. *Then, assuming MIMO channel coefficients are i.i.d. (as defined in Sec. II),* $\boldsymbol{U}_{l_j}^{\star}$ *is not orthonormal, almost surely.*

A few comments are in order at this stage, regarding the difference between AIMS and max-SINR. Referring to (SRM_F) and (SRM_B), it is clear that our proposed algorithm reduces to max-SINR, in case of single-stream transmission, i.e., setting $d = 1$. Moreover, an inherent property of the max-SINR solution is that it yields equal power allocation across all the streams (since the individual columns of each transmit/receive filter are normalized to unity). However, as evident from (10), our proposed solution does not normalize the individual columns of the receive filter, but rather the whole filter norm (as seen in Algorithm 1). This allows for different power allocation, across columns of the same filter. That being said, the proposed solution is expected to yield better sum-rate performance (w.r.t. max-SINR), especially in the interference-limited regime. This is due to the intuitive fact that much can be gained from allocating low power to streams that suffer from severe interference, and higher power to streams with lesser interference (this will be validated in the numerical results section). We next introduce a rank adaptation mechanism that further enhances the interference suppression capabilities of the algorithm.

C. AIMS With Rank Adaptation

We introduce an additional (heuristic) mechanism to robustify AIMS against severely interference-limited scenarios, by introducing a mechanism of Rank Adaptation (RA): in addition to the transmit / receive filter optimization (Lemma 1), the latter allows the filter rank to be optimized as well. Mathematically speaking, RA addresses the following problem,

$$r^{\star} \triangleq \underset{r}{\operatorname{argmax}} \left[\boldsymbol{X}^{\star} \triangleq \underset{\boldsymbol{X}\in\mathbb{C}^{n\times r}}{\operatorname{argmax}} \frac{|\boldsymbol{X}^{\dagger}\boldsymbol{R}\boldsymbol{X}|}{|\boldsymbol{X}^{\dagger}\boldsymbol{Q}\boldsymbol{X}|} \right], \tag{11}$$

Using the same argument as Lemma 1, one can verify that $\boldsymbol{X}^{\star}$ and $r^{\star}$ are as follows,

$$\begin{aligned} \boldsymbol{X}^{\star} &= [\boldsymbol{L}^{-\dagger}\boldsymbol{\Psi}]_{1:r^{\star}}, \quad \text{where } \boldsymbol{\Psi} = \upsilon_{1:n}[\boldsymbol{L}^{-1}\boldsymbol{R}\boldsymbol{L}^{-\dagger}] \\ r^{\star} &= \underset{r}{\operatorname{argmax}}\ |\boldsymbol{\Lambda}_r| = \left|\{i \mid \lambda_i[\boldsymbol{L}^{-1}\boldsymbol{R}\boldsymbol{L}^{-\dagger} \geq 1\}\right| \end{aligned} \tag{12}$$

where $\boldsymbol{\Lambda}_r \in \mathbb{R}^{r\times r}$ is the (diagonal) matrix consisting of the r-largest eigenvalues of $\boldsymbol{L}^{-1}\boldsymbol{R}\boldsymbol{L}^{-\dagger}$. Simply put, $r^{\star}$ is the number of eigenvalues greater than one.

When RA is incorporated into AIMS, this mechanism will boost the performance of the algorithm (namely in interference-limited settings). However, one still needs to ensure that the filter ranks for each transmit-receive pair are the same, i.e., $\operatorname{rank}(\boldsymbol{U}_{l_j}) = \operatorname{rank}(\boldsymbol{V}_{l_j})\ \forall\, l_j$. One quick (heuristic) solution is as follows. For each transmit-receive filter pair, compute the optimal filter rank for both the transmit and receive filter, and use the minimum.[1] Needless to say, ensuring this condition requires additional signaling overhead. We thus envision RA, as potential 'add-on' for AIMS, when one can afford the resulting overhead increase. The added performance boost from RA in further discussed in the numerical results section.

[1] Alternately, one can apply RA to the receive filters only, in the last iteration of the algorithm, since the transmit filter updates are more heuristic than the receive filter updates.

IV. Maximizing a DLT Bound

In this section we propose another approach to tackle the sum-rate optimization problem. The central idea behind this approach is to use a lower bound on the sum-rate, that results in separable subproblems.

A. Problem Formulation

We focus the derivations to the interference-limited case, where the following holds,

$$\lambda_i[\boldsymbol{U}_{l_j}^\dagger \boldsymbol{Q}_{l_j}\boldsymbol{U}_{l_j}] \to \infty,\ \forall i \in \{d\} \Leftrightarrow \begin{cases} A1)\ \lambda_i[(\boldsymbol{U}_{l_j}^\dagger \boldsymbol{Q}_{l_j}\boldsymbol{U}_{l_j})^{-1}] \to 0, \quad \forall i \in \{d\} \\ A2)\ \boldsymbol{I}_d \succeq (\boldsymbol{U}_{l_j}^\dagger \boldsymbol{Q}_{l_j}\boldsymbol{U}_{l_j})^{-1} \end{cases} \tag{13}$$

Proposition 2: In the interference-limited regime, the user-rate r_{l_j} in (3) is lower bounded by,

$$r_{l_j} \geq \log_2|\boldsymbol{I}_d + \boldsymbol{U}_{l_j}^\dagger \boldsymbol{R}_{l_j}\boldsymbol{U}_{l_j}| - \log_2|\boldsymbol{U}_{l_j}^\dagger \boldsymbol{Q}_{l_j}\boldsymbol{U}_{l_j}|, \tag{14}$$

$$\geq \log_2|\boldsymbol{I}_d + \boldsymbol{U}_{l_j}^\dagger \boldsymbol{R}_{l_j}\boldsymbol{U}_{l_j}| - \mathrm{tr}(\boldsymbol{U}_{l_j}^\dagger \boldsymbol{Q}_{l_j}\boldsymbol{U}_{l_j}) \triangleq r_{l_j}^{(LB)}, \tag{15}$$

where $r_{l_j}^{(LB)}$ is such that,

$$\begin{aligned}\Delta_{l_j} &\triangleq r_{l_j} - r_{l_j}^{(LB)} \\ &= \mathrm{tr}(\boldsymbol{U}_{l_j}^\dagger \boldsymbol{Q}_{l_j}\boldsymbol{U}_{l_j}) - \log_2|\boldsymbol{U}_{l_j}^\dagger \boldsymbol{Q}_{l_j}\boldsymbol{U}_{l_j}| \\ &\quad + \mathcal{O}(\mathrm{tr}[(\boldsymbol{U}_{l_j}^\dagger \boldsymbol{Q}_{l_j}\boldsymbol{U}_{l_j})(\boldsymbol{U}_{l_j}^\dagger \boldsymbol{R}_{l_j}\boldsymbol{U}_{l_j})^{-1}]), \quad \forall l_j \in I \end{aligned} \tag{16}$$

Proof: Refer to Appendix B. ∎

We refer to expressions such as $r_{l_j}^{(LB)}$, as a Difference of Log-Trace (DLT) expressions. They shall be used as basis for the optimization algorithm. With that in mind, the sum-rate R_Σ, can be lower bounded by $R_\Sigma^{(LB)}$,

$$R_\Sigma^{(LB)} = \sum_{l_j \in I} \log_2|\boldsymbol{I}_d + \boldsymbol{U}_{l_j}^\dagger \boldsymbol{R}_{l_j}\boldsymbol{U}_{l_j}| - \mathrm{tr}(\boldsymbol{U}_{l_j}^\dagger \boldsymbol{Q}_{l_j}\boldsymbol{U}_{l_j}) \tag{17}$$

$$= \sum_{i_k \in I} \log_2|\boldsymbol{I}_d + \boldsymbol{V}_{i_k}^\dagger \bar{\boldsymbol{R}}_{i_k}\boldsymbol{V}_{i_k}| - \mathrm{tr}(\boldsymbol{V}_{i_k}^\dagger \bar{\boldsymbol{Q}}_{i_k}\boldsymbol{V}_{i_k}) \tag{18}$$

where the last equality is due to $\log|\boldsymbol{I} + \boldsymbol{AB}| = \log|\boldsymbol{I} + \boldsymbol{BA}|$, and the linearity of tr(.). Then, the sum-rate optimization problem in (4) can be bounded below by solving the following,

$$\begin{cases} \max\limits_{\{\boldsymbol{V}_{l_j}, \boldsymbol{U}_{l_j}\}} R_\Sigma^{(LB)} \\ \text{s. t. } \|\boldsymbol{U}_{l_j}\|_F^2 = P_r,\ \|\boldsymbol{V}_{l_j}\|_F^2 = P_t, \quad \forall l_j \in I \end{cases} \tag{19}$$

Note that the above problem is not jointly convex in all the optimization variables, mainly due to the coupling between the transmit and receive filters. Moreover, the reason for having transmit/receive power constraints with equality, will become clear in Sec. IV-C.

B. Proposed Algorithm

The formulation in (19) is ideal for a Block Coordinate Descent (BCD) approach. We use the superscript $^{(n)}$ to denote the iteration number: at the nth iteration, the transmit filters, $\{\boldsymbol{V}_{l_j}^{(n)}\}$, are fixed, and the update for the receive filters, $\{\boldsymbol{U}_{l_j}^{(n+1)}\}$, is the one that maximizes the objective. The same is done for the transmit filter update. In each of the two stages, BCD decomposes the original coupled problem (19), into a set of parallel subproblems, that can solved in distributed fashion. This is formalized in (20), as shown at the bottom of this page and each of the resulting subproblems are detailed below. When the transmit filters are fixed, the problem decouples in the receive filters $\{\boldsymbol{U}_{l_j}\}$ (as seen from (19)), and the resulting subproblems are given by,

$$(J1)\ \begin{cases} \min\limits_{\boldsymbol{U}_{l_j}} \mathrm{tr}(\boldsymbol{U}_{l_j}^\dagger \boldsymbol{Q}_{l_j}\boldsymbol{U}_{l_j}) - \log_2|\boldsymbol{I}_d + \boldsymbol{U}_{l_j}^\dagger \boldsymbol{R}_{l_j}\boldsymbol{U}_{l_j}| \\ \text{s. t. } \|\boldsymbol{U}_{l_j}\|_F^2 = P_r \end{cases} \tag{21}$$

By recalling that (19) can rewritten as (18), we see that the above objective decouples in the transmit filters, i.e.,

$$(J2)\ \begin{cases} \min\limits_{\boldsymbol{V}_{i_k}} \mathrm{tr}(\boldsymbol{V}_{i_k}^\dagger \bar{\boldsymbol{Q}}_{i_k}\boldsymbol{V}_{i_k}) - \log_2|\boldsymbol{I}_d + \boldsymbol{V}_{i_k}^\dagger \bar{\boldsymbol{R}}_{i_k}\boldsymbol{V}_{i_k}| \\ \text{s. t. } \|\boldsymbol{V}_{i_k}\|_F^2 = P_t \end{cases} \tag{22}$$

Thus, choosing DLT expressions is rather advantageous, since they lead to subproblems that decouple in both $\{\boldsymbol{U}_{l_j}\}$ and $\{\boldsymbol{V}_{l_j}\}$. Note that the equality constraints in $(J1)$ and $(J2)$, do not affect the convexity of the problems, as they are already non-convex. Indeed, expressions such as $-\log_2|\boldsymbol{I}_d + \boldsymbol{U}_{l_j}^\dagger \boldsymbol{R}_{l_j}\boldsymbol{U}_{l_j}|$ are *not convex* in $\boldsymbol{U}_{l_j}$.[2] However, this does not make BCD less applicable, as long as $(J1)$ and $(J2)$ are solved *globally*. The solution to each of the subproblems is given by the following result.

Lemma 2: Non-homogeneous Waterfilling.
Consider the following problem,

$$(P)\ \begin{cases} \min\limits_{\boldsymbol{X} \in \mathbb{C}^{n\times r}} f(\boldsymbol{X}) \triangleq \mathrm{tr}(\boldsymbol{X}^\dagger \boldsymbol{Q}\boldsymbol{X}) - \log_2|\boldsymbol{I}_d + \boldsymbol{X}^\dagger \boldsymbol{R}\boldsymbol{X}| \\ \text{s. t. } \|\boldsymbol{X}\|_F^2 = \zeta. \end{cases} \tag{23}$$

where $\boldsymbol{Q} \succ \boldsymbol{0}$ and $\boldsymbol{R} \succeq \boldsymbol{0}$, $r < n$. Let $\boldsymbol{Q} \triangleq \boldsymbol{L}\boldsymbol{L}^\dagger$ be the Cholesky factorization of $\boldsymbol{Q}$, and $\boldsymbol{M} \triangleq \boldsymbol{L}^{-1}\boldsymbol{R}\boldsymbol{L}^{-\dagger}$, $\boldsymbol{M} \succeq \boldsymbol{0}$, and define the following, $\{\alpha_i \triangleq \lambda_i[\boldsymbol{M}]\}_{i=1}^r$, $\boldsymbol{\Psi} \triangleq v_{1:r}[\boldsymbol{M}]$, $\{\beta_i \triangleq \boldsymbol{\Psi}_{(i)}^\dagger (\boldsymbol{L}^\dagger \boldsymbol{L})^{-1}\boldsymbol{\Psi}_{(i)}\}_{i=1}^r$. Then the optimal solution for (P) is given by,

$$\boldsymbol{X}^\star = \boldsymbol{L}^{-\dagger}\boldsymbol{\Psi}\boldsymbol{\Sigma}^\star, \tag{24}$$

[2] To see this, consider the (degenerate) scalar case. It is easily verified that $-\log_2(1 + ru^2)$, $r > 0$ is concave for $u \ll 1$, and convex for $u \gg 1$.

$$\{\boldsymbol{V}_{l_j}^{(n+1)}\} \triangleq \underbrace{\underset{\{\boldsymbol{V}_{l_j}\}}{\mathrm{argmax}}\ R_\Sigma^{(LB)} \left(\underbrace{\{\boldsymbol{U}_{l_j}^{(n+1)}\} \triangleq \underset{\{\boldsymbol{U}_{l_j}\}}{\mathrm{argmax}}\ R_\Sigma^{(LB)}(\{\boldsymbol{U}_{l_j}\}, \{\boldsymbol{V}_{l_j}^{(n)}\})}_{J1}\ ,\ \{\boldsymbol{V}_{l_j}\} \right)}_{J2},\ n = 1, 2, \ldots \tag{20}$$

where $\boldsymbol{\Sigma}^\star$ (diagonal) is the optimal power allocation,

$$(P5)\quad \begin{cases} \min\limits_{\{x_i\}} \ \sum_{i=1}^{r} \left(x_i - \log_2(1+\alpha_i x_i)\right) \\ \text{s. t. } \sum_{i=1}^{r} \beta_i x_i = \zeta, \quad x_i \geq 0, \quad \forall i \end{cases} \tag{25}$$

Proof: Refer to Appendix C ∎

We underline that a similar problem was obtained in [30], but in the context of covariance optimization. Hence, this result is not applicable to (P). Moreover, $(P5)$ has a closed-form solution, that can be obtained using standard Lagrangian techniques.

Lemma 3: The solution to the optimal power allocation in $(P5)$ is given by,

$$\boldsymbol{\Sigma}^\star_{(i,i)} = \sqrt{\left(1/(1+\mu^\star\beta_i) - 1/\alpha_i\right)^+}, \quad \forall i, \tag{26}$$

where $\mu^\star$ is the unique root to,

$$g(\mu) \triangleq \sum_{i=1}^{r} \beta_i \left(1/(1+\mu\beta_i) - 1/\alpha_i\right)^+ - \zeta,$$

on the interval $]-1/(\max_i \beta_i),\ \infty[$, and $g(\mu)$ is monotonically decreasing on that interval.

Proof: Refer to Appendix C ∎

With this in mind, we can write the optimal transmit and receive filter updates, as follows,

$$\begin{aligned} \boldsymbol{U}^\star_{l_j} &= \boldsymbol{L}^{-\dagger}_{l_j} \boldsymbol{\Psi}_{l_j} \ \boldsymbol{\Sigma}^\star_{l_j}, \quad \boldsymbol{\Psi}_{l_j} \triangleq v_{1:d}[\boldsymbol{L}^{-1}_{l_j} \boldsymbol{R}_{l_j} \boldsymbol{L}^{-\dagger}_{l_j}], \forall\, l_j, \\ \boldsymbol{V}^\star_{i_k} &= \boldsymbol{K}^{-\dagger}_{i_k} \boldsymbol{\Theta}_{i_k} \ \boldsymbol{\Lambda}^\star_{i_k}, \quad \boldsymbol{\Theta}_{i_k} \triangleq v_{1:d}[\boldsymbol{K}^{-1}_{i_k} \bar{\boldsymbol{R}}_{i_k} \boldsymbol{K}^{-\dagger}_{i_k}], \quad \forall\, i_k, \end{aligned} \tag{27}$$

where $\boldsymbol{\Sigma}^\star_{l_j}$ and $\boldsymbol{\Lambda}^\star_{i_k}$ are the optimal power allocation, given in Lemma 3. The resulting algorithm, max-DLT, is detailed in Algorithm 2 (where T is the number of F-B iterations). Moreover, due to the monotone nature of $g(\mu)$, $\mu^\star$ can be found using simple 1D search methods.

Algorithm 2 Maximal DLT (Max-DLT)

for $t = 1, 2, \ldots, T$ **do**
  // forward network optimization: receive filter update
    Estimate $\boldsymbol{R}_{l_j}, \boldsymbol{Q}_{l_j}$, and compute $\boldsymbol{L}_{l_j}, \forall l_j$
    $\boldsymbol{U}_{l_j} \leftarrow \boldsymbol{L}^{-\dagger}_{l_j} v_{1:d}[\boldsymbol{L}^{-1}_{l_j} \boldsymbol{R}_{l_j} \boldsymbol{L}^{-\dagger}_{l_j}] \boldsymbol{\Sigma}_{l_j}, \quad \forall l_j$
  // reverse network optimization: transmit filter update
    Estimate $\bar{\boldsymbol{R}}_{i_k}, \bar{\boldsymbol{Q}}_{i_k}$, and compute $\boldsymbol{K}_{i_k}, \forall i_k$
    $\boldsymbol{V}_{i_k} \leftarrow \boldsymbol{K}^{-\dagger}_{i_k} v_{1:d}[\boldsymbol{K}^{-1}_{i_k} \bar{\boldsymbol{R}}_{i_k} \boldsymbol{K}^{-\dagger}_{i_k}] \boldsymbol{\Lambda}_{i_k}, \quad \forall i_k$
end for

C. Analysis and Discussions

a) Interpretation: We provide an intuitive interpretation of the problem in Lemma 2 and its solution. It can be verified that $\{\alpha_i \triangleq \lambda_i[\boldsymbol{L}^{-1}\boldsymbol{R}\boldsymbol{L}^{-\dagger}]\}_{i=1}^r$ are also the eigenvalues of $\boldsymbol{Q}^{-1}\boldsymbol{R}$ (where $\boldsymbol{R}$ and $\boldsymbol{Q}$ represent the signal and I+N covariance matrix, respectively). Thus, $\{\alpha_i\}$ acts as a (quasi)-SINR measure, for each of the data streams. Moreover, it can be seen that the optimal power allocated to stream i, $\boldsymbol{\Sigma}^\star_{(i,i)}$ in (26), tends to zero as $\alpha_i \to 0$, i.e., no power is allocated to streams that have low-SINR.[3] Moreover, note that $\{\beta_i\}$ represents the cost of allocating power to each of the streams (this can be seen in $(P5)$). Thus, the non-homogeneous waterfilling solution in (24) allocates power to each of the streams, based on the SINR and cost of each (possibly not allocating power to some streams). In addition, we note that this solution reduces to that of the GMRQ problem, (8), when equal power allocation is assumed.

b) Discussion: We now discuss the reason for adopting the equality power constraints for the problem at hand (i.e., $(J1)$ and $(J2)$), by showing the limitation of using an inequality constraint. Note that in the noise-limited regime, $\sigma_l \gg 1,\ \forall l \in \mathcal{L}$, and consequently $\alpha_i \triangleq \lambda_i[\boldsymbol{L}^{-1}\boldsymbol{R}\boldsymbol{L}^{-\dagger}] \to 0,\ \forall i \in \{r\}$. Using the fact that $\log(1+y) \approx y,\ y \ll 1$, the optimal power allocation in $(P5)$ is approximated as,

$$\begin{aligned} \sum_{i=1}^{r} x_i - \log_2(1+\alpha_i x_i) &\approx \sum_{i=1}^{r} x_i - \alpha_i x_i \\ &= \sum_{i=1}^{r} (1-\alpha_i) x_i \overset{\alpha_i \to 0}{\approx} \sum_{i=1}^{r} x_i \end{aligned} \tag{28}$$

When inequality constraints are considered, $(P5)$ takes the following form,

$$\min \sum_i x_i \quad \text{s. t.} \quad \sum_{i=1}^{r} \beta_i x_i \leq \zeta, \quad x_i \geq 0. \tag{29}$$

One can see that the optimal solution is $x_i^\star = 0, \forall i$, and the optimal transmit/receive filter in $(J1)$ and $(J2)$ is zero. Thus, operating with an inequality power constraint leads to degenerate solutions, in the noise-limited regime. Though it might seem that an equality power constraint makes $(J1)$ and $(J2)$ harder to solve, this is not the case as both have non-convex cost functions already: despite their non-convexity, they can be effectively solved using Lemma 2 and Lemma 3. Moreover, the convergence of BCD is unaffected since the globally optimal solution is found for each subproblem (formalized in the following proposition).

c) Convergence of max-DLT: Regarding convergence of the proposed algorithm, max-DLT, it is established using standard BCD convergence results.

Corollary 2: Let $\psi_n \triangleq R_\Sigma^{(LB)}(\{\boldsymbol{U}_{l_j}^{(n)}\}, \{\boldsymbol{V}_{l_j}^{(n)}\})$, $n = 1, 2, \ldots$ be the sequence of iterates for the objective value. Then, $\{\psi_n\}$ is non-decreasing in n, and converges to a stationary point of $R_\Sigma^{(LB)}(\{\boldsymbol{U}_{l_j}\}, \{\boldsymbol{V}_{l_j}\})$.

Proof: ψ_n converges to a stationary point of the objective, since a unique minimizer is found at each step. This follows form BCD convergence results in [31] and [32, Ch. 7.8]. ∎

d) Relation to other methods: The view that the proposed approach seems close to other heuristics such as successive convex programming (SCP) and the convex-concave procedure (CCP), is slightly misleading. Those methods start with expressions such as (14) in Proposition 2, then approximate $\log_2|\boldsymbol{U}^\dagger_{l_j}\boldsymbol{Q}_{l_j}\boldsymbol{U}_{l_j}|$ with a linear function (in the case of CCP [33]), or optimize a quadratic lower bound (in the case of

[3] Although the optimal power allocation to stream i is zero for some streams, i.e., $\boldsymbol{\Sigma}_{(i,i)} = 0$, in the actual implementation of the algorithm, $\boldsymbol{\Sigma}_{(i,i)} = \delta$ where $\delta \ll 1$.

SCP [34]). The approximation is iteratively updated until convergence. Starting with (14), CCP [33] generates a sequence of iterates $\{\boldsymbol{U}_{l_j}^{(n)}\}_n$, where at iteration n, $\log_2 |\boldsymbol{U}_{l_j}^{\dagger} \boldsymbol{Q}_{l_j} \boldsymbol{U}_{l_j}|$ is approximated along its gradient, $\boldsymbol{A}_{l_j}^{(n)}$, around the point $\boldsymbol{U}_{l_j}^{(n)}$, i.e.,

$$\boldsymbol{U}_{l_j}^{(n+1)} = \underset{\boldsymbol{U}_{l_j}}{\text{argmin}} \ \text{tr}\ ((\boldsymbol{A}_{l_j}^{(n)})^{\dagger} \boldsymbol{U}_{l_j}) - \log_2 |\boldsymbol{I}_d + \boldsymbol{U}_{l_j}^{\dagger} \boldsymbol{R}_{l_j} \boldsymbol{U}_{l_j}|$$

Note that a comparison between the CCP updates above, and those resulting from our proposed approach, e.g., $(J1)$ and $(J2)$, indeed reveals that they different. Moreover, such approaches will inevitably lead to significantly larger communication overhead and complexity; this goes against the main motivation of the work (this is further discussed in Sec. V-D). Thus, iteratively updating the DLT bound around the operating point (in a similar fashion to CCP or SCP), is not applicable: this is not of interest in this work, as the resulting bound would not be *separable* and decouple in the transmit/receive filters. That being said, we will benchmark against a CCP scheme, where transmit covariance optimization was considered in the MIMO IMAC setting [23].

V. Practical Aspects

A. Comparison

A few remarks are in order at this stage, regarding the similarities and differences between AIMS and max-DLT. Referring to the optimal update equations for each algorithm, i.e., (10) and (27), we clearly see that both span the same subspace, i.e. the generalized eigenspace between the signal and I+N covariance matrices. In addition, max-DLT computes the optimal power allocation for each stream. Despite this significant similarity among the two solutions, recall that they are derived from two fundamentally different separability metrics. While AIMS is an extension of max-SINR, that greedily maximizes the separability at each BS/MS, the updates of max-DLT maximize a lower bound on the sum-rate capacity (and are shown to converge to a stationary point of the DLT bound). That being said, their performance evaluation is done via numerical results.

B. Benchmarks

As mentioned earlier, we will also investigate the proposed approach in alternate scenarios such as MIMO IBC, and the MIMO Interference Channel (MIMO IFC). We benchmark our algorithms against widely adopted ones,

- o *max-SINR* [14] in the MIMO IMAC, MIMO IFC and MIMO IBC
- o *MMSE and Weighted-MMSE* [17], [18] in the MIMO IFC and MIMO IBC

as well as relevant fast-converging algorithms,

- o *CCP-WMMSE* [23]: an accelerated version of WMMSE algorithm (using CCP), for the MIMO IMAC
- o *IWU* [25]: a fast-convergent leakage minimization algorithm for the MIMO IFC
- o *Uncoordinated*, wherein each transmit-receive pair perform optimal eigen-beamforming using the left/right eigenvectors of the channel (irrespective of all other pairs).

Both IWU and CCP-WMMSE rely on the use of *turbo iterations*: I inner-loop iterations are carried out within each main F-B iteration, to solve a given optimization problem (analogous to *primal-dual decompositions*, where an inner problem is solved to optimality, and the solution plugged into the main problem). While those turbo iterations are performed at the BS/MS in the case of IWU, they are run over-the-air in the case of CCP-WMMSE, thus leading to higher overhead. We note as well that earlier works applied SCA to MIMO IBC settings, e.g., [20], but their algorithms are restricted to single-stream beamforming/combining.

C. Distributed CSI Acquisition

We underline in this section some practical issues that relate to the proposed schemes, such as the mechanism for distributed CSI acquisition, and the resulting communication overhead and computational complexity. Although additional issues such as robustness have to considered as well, such matters are outside the scope of the current paper. We reiterate the fact that CSI acquisition mechanisms are outside the scope of the paper (we refer the reader to [27]). We just summarize the basic operation behind F-B iterations.

Evidently, the operation of such schemes is contingent upon each transmitter / receiver being able to estimate the signal and the I+N covariance matrices, in a fully distributed manner. From the perspective of this work, this is accomplished via the use of *precoded pilots* to estimate the *effective channels*.[4] In the forward phase, the signal covariance matrix, as well as the I+N covariance matrices, can be computed after estimating the effective signal channel, and the effective interfering channels, respectively. The receive filters at the base stations are updated following any of the proposed algorithms (summarized in Fig. 2). Then, in the downlink phase, the same procedure is used to estimate the signal and I+N covariance matrices, and update the filters at the receivers. This constitutes one F-B iteration. Recall that T is the total number of such iterations, that are carried out.

D. Communication Overhead

For such schemes to be fully distributed, the required CSI quantities have to be obtained via uplink-downlink pilots. Each F-B iteration has an associated communication overhead, namely that of bi-directional transmission of pilots. We adopt a simplistic definition of the communication overhead, as the minimal number of orthogonal pilots symbols, required for estimating the required CSI quantities (keeping in mind that the actual overhead will be dominated by this quantity). We note that almost all prior algorithms that have been proposed in the context of cellular systems, focus on a regime with a high enough number of F-B iterations, i.e., $T = 100 \sim 1000$, such as [14] and [16]–[18] to list a few. In addition, the relatively

[4] A full investigation of the total overhead of this decentralized solution, as compared to a centralized implementation, falls outside the scope of the current paper.

larger number of antennas in lower bands of mmWave systems, results in significantly more iterations. On the contrary, and in line with recent attempts such as [22], [24], and [25], we assume that this modus operandi is not feasible in the systems we consider (since F-B iterations are carried out over-the-air, and the associated overhead would be higher than the potential gains). Indeed the lower coherence time of mmWave channels (compared to that of traditional systems) makes the low-overhead requirement even more stringent. We thus focus on a regime where $T = 2 \sim 5$. In addition, we assume that the minimal number of orthogonal pilots is used, i.e., d orthogonal pilot slots for each uplink/downlink effective channel. Moreover, the pilots are orthogonal across users and cells, resulting in a total of KLd orthogonal pilots for each uplink/downlink training phase. Consequently, the total overhead of both AIMS and max-DLT is approximately,

$$\Omega_{\text{prop}} = T(\underbrace{KLd}_{UL} + \underbrace{KLd}_{DL}) = 2TKLd \text{ channel uses.}$$

It can be verified that the overhead is the same for benchmarks such as max-SINR, IWU and MMSE. Moreover, using similar arguments one can approximate the overhead of CCP-WMMSE and WMMSE (in the number of channel uses), as

$$\Omega_{\text{ccp-wmmse}} = T[(\underbrace{KLM}_{\substack{UL\ chann.\\ estim}})\times \underbrace{(L-1)}_{\substack{CSI\\ sharing}} + \underbrace{I}_{turbo} \times (\underbrace{KLN}_{\substack{cov.\ mat\\ upd.}})]$$

$$\Omega_{\text{wwmse}} = T(\underbrace{KLd}_{UL} + \underbrace{KLM}_{weights} + \underbrace{KLd}_{DL})$$

where I denotes the number of turbo iterations. Though a coarse measure, we can see that the overhead associated with WMMSE and its fast-converging variant CCP-WMMSE are significantly higher than that of the proposed schemes. Moreover, CCP-WMMSE exhibits massively larger overhead than the other two, due to the fact that the turbo optimization is carried over-the-air (as described in Sec.V-B), and that the CSI for the uplink channels is shared among the BSs [23]. The overhead of the aforementioned schemes will be included in the numerical results.

E. Complexity

Despite the fact that the communication overhead is the limiting resource in cellular networks, we nonetheless shed light on the complexity of the proposed approaches, for completeness. The computational complexity of both AIMS and max-DLT is dominated by the Cholesky Decomposition of the I+N covariance matrix, and the Eigenvalue Decomposition of $\boldsymbol{M}$, both of which have similar complexity of $O(N^3)$ (keeping in mind that other operations such as matrix multiplication and bisection search are quite negligible in comparison). Thus, the complexity (per F-B iteration) is approximately,

$$C_{\text{prop}} = O(2KL(M^3 + N^3)).$$

Note that the same holds for benchmarks such as max-SINR, IWU, and WMMSE since they are dominated by matrix inversion of the I+N covariance matrix. While the complexity of CCP-WMMSE is also dominated by the above quantity, it also involves running a series of semi-definite programs (using interior point solvers), within each turbo iteration. This renders the algorithm quite costly.

VI. Numerical Results

A. Simulation Methodology

We use the achievable sum-rate in the network as the performance metric, where the achievable user rate is given by (3). Because the approach here is presented in the context of MIMO IMAC, a significant fraction of the results will be under this setup. The model, algorithms and results presented in the paper, are applicable to the MIMO IBC: in the latter setting, the transmit filters ($\{\boldsymbol{V}_{i_k}\}$) are at the BSs, receive filters ($\{\boldsymbol{U}_{l_j}\}$) at the MSs, and M (resp. N) denotes the number of BS (resp. MS) antennas. For example, a system with 32 BS antennas and 4 MS antennas, is modeled by setting $M = 4, N = 32$ in the MIMO IMAC case, and $N = 4, M = 32$ in the MIMO IBC.

We limit the number of F-B iterations, T, to a small number. Following the calculations in Sec.V-D, we include in the results the overhead, $\boldsymbol{\Omega}$, for each of the algorithms, as a function of T (and the number if turbo iterations, I, when applicable). Moreover, the considered sub-28 GHz systems allow for more BS/MS antennas compared to traditional cellular frequencies, for the same transmit/receive array surface (due to the antenna spacing being smaller). This will be reflected in the numerical results. Finally, we note that all curves are averaged over 500 channel realizations. In the first part, we assume a block-fading channel model with static users, having i.i.d. channels, to benchmark against several known schemes, and canonical channel model. In another part, we will build a more realistic simulation setup that is more reflective of the systems under consideration.

B. Part I

Single-user Multi-cell MIMO Uplink: We will first evaluate the performance of our schemes in a MIMO Interference Channel (IFC), where many schemes such as max-SINR [14], MMSE [17], and IWU [25] are applicable. We set $M = N = 4$, $d = 2$ and fix the number of F-B iterations, $T = 4$, for all schemes. We also included Weighted-MMSE with the corresponding number of F-B iterations ($T = 4$), and a large enough number of iterations (as an upper bound). It is clear from Fig. 4 that while max-DLT has similar performance as WMMSE (for $T = 4$) in the low-to-medium SNR regime, the gap increases in the high-SNR region (SNR $\geq$ 20 dB). Moreover, we note that our proposed schemes outperform all other the benchmarks, across all SNR regimes. In particular, the performance gap between max-DLT and the benchmarks, is quite significant in the medium-to-high SNR region. Moreover, max-DLT offers a 35% gain over the fast-converging IWU: Though both are able to turn off some streams in view of reducing interference, max-DLT also optimizes the signal subspace as well. Note that the 'optimal-performance' of WMMSE is achieved for $T = 200$, but the resulting overhead is massive. Although the performance

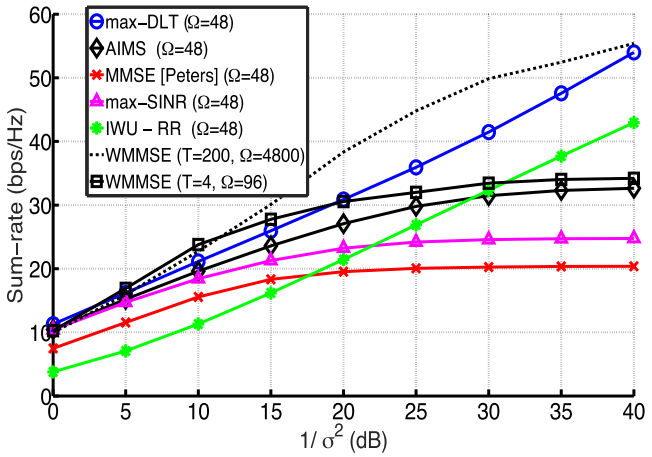


Fig. 4. Ergodic sum-rate vs $1/\sigma^2$, for $L = 3, K = 1, M = N = 4, d = 2$, $T = 4$ (MIMO IFC).

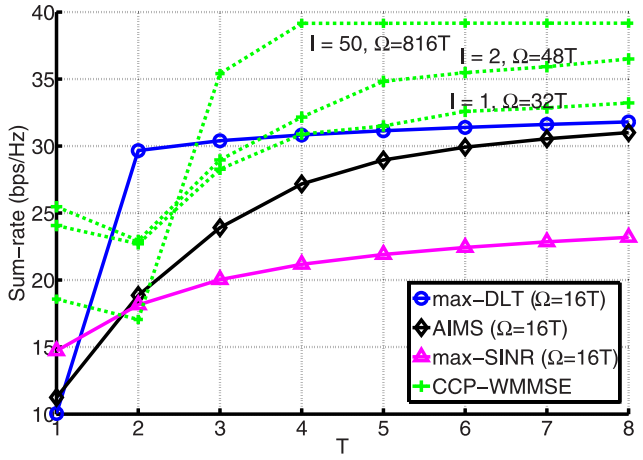


Fig. 5. Ergodic sum-rate vs T, for $L = 2, K = 2, M = 4, N = 4, d = 2$.

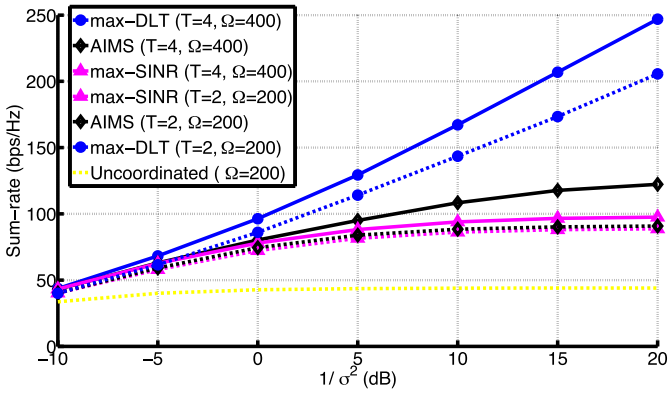


Fig. 6. Ergodic sum-rate vs $1/\sigma^2$, for $L = 5, K = 5, M = 4, N = 32$, $d = 2$ (Uplink).

of max-DLT is similar to WMMSE ($T = 4$) in low-to-medium SNR regime, the overhead is much lower for the former. Moreover, the gap increases in the medium-to-high SNR region. The rest of results will show that the performance gap between our proposed algorithms and several known benchmarks, increases in the regime of interest (low-overhead, large system dimensions).

Multi-user Multi-cell MIMO uplink: We next evaluate a MIMO IMAC setting with $L = 2$, $K = 2$, $M = 4$, $N = 4$, $d = 2$, as a function of the number of F-B iterations, T. We also benchmark against CCP-WMMSE (summarized in Sec. V-B) by varying the number of turbo iterations I, and testing the resulting performance and overhead. Fig. 5 clearly shows the fast-converging features of both algorithms. More specifically, this is apparent in the case of max-DLT, that reaches 95% of its performance in 2 iterations. While the performance of max-DLT is slightly better than CCP-WWMMSE for $I = 1$, the overhead of the latter is twice that of the former (CCP-WMMSE becomes better than max-DLT for $I = 2$, but the resulting overhead is thrice as high). Note that the 'full' performance CCP-WMMSE is achieved for $I = 50$, but the the resulting overhead (and complexity) are orders-of-magnitude larger than the proposed schemes. Its performance is quite sensitive to solving the inner problem to optimality (i.e., until the turbo iteration converges), thus making it ill-suited for larger setups. Indeed, the running time of CCP-WMMSE (using the Mosek solver in CVX) prevented us from testing its performance for larger antenna configurations.

Large-scale Multi-user Multi-cell MIMO uplink: We test the performance of a system where a larger number of antennas is available at the BS (enabled by sub-28 GHz systems). We evaluate a large-scale (in the number of antennas at the BS) multi-user multi-cell uplink with $L = 5, K = 5$, $d = 2, M = 4, N = 32$. Fig. 6 shows the resulting sum-rate of the proposed schemes (and max-SINR), for $T = 2$ and $T = 4$ (we were unable to include CPP-WMMSE as the resulting simulation time was too long to be included). Recall that for each of the simulated values of T, the overhead is the same for all schemes. We observe that both our schemes outperform max-SINR significantly. In particular, max-DLT offer twice the performance of max-SINR at 5dB (this performance gap increases with the SNR). And while both our schemes show significant performance gain by increasing T, the corresponding gain that max-SINR exhibits is negligible in comparison.

Large-scale Multi-user Multi-cell MIMO downlink: We next investigate a dual communication setup of the one just above, a MIMO IBC obtained by setting $M = 32, N = 4$ (all else being the same). We benchmark our results against the well-known WMMSE algorithm [18]. Note that while WMMSE employs a sum-power constraint, our schemes have a per-user power constraint, and thus assume equal power allocation among the users.[5] This implies that a more stringent constraint is placed on our schemes. Despite this unfavorable setup, both our schemes significantly outperform WMMSE, the gap becoming quite large when noise power is $\sigma^2 = 0.01$ (as seen in Fig. 7). Note as well that the overhead of our proposed schemes, is half that of WMMSE as the latter requires feedback of the weights (refer to Sec. V-D for the overhead calculations). Needless to say, the full-performance

[5]If P_t is the per user constraint for our schemes, then KP_t is the per-BS sum-power constraint for WWMMSE.

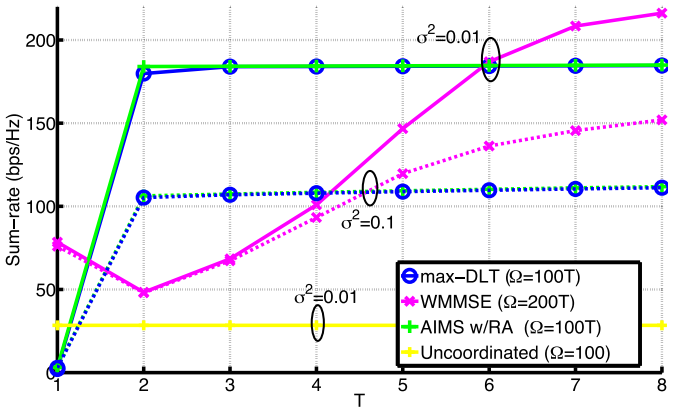


Fig. 7. Ergodic sum-rate vs T, for $L = 5, K = 5, M = 32, N = 4$, $d = 2$ (Downlink). Solid curves correspond to noise power $\sigma^2 = 10^{-2}$, and dashed ones to $\sigma^2 = 10^{-1}$.

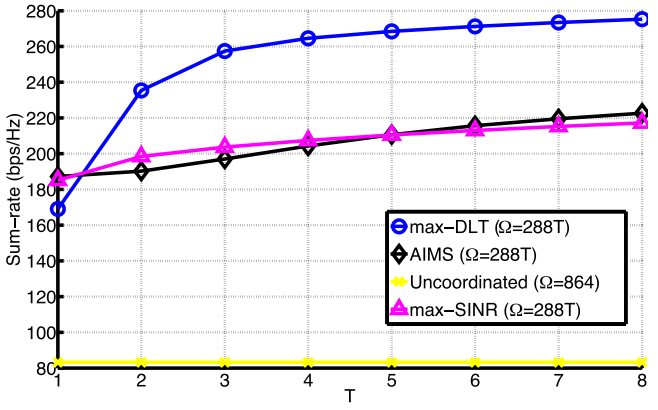


Fig. 8. Ergodic sum-rate vs T, for dense uplink ($N = 8, M = 4, d = 2$, $L = 9, K = 8$).

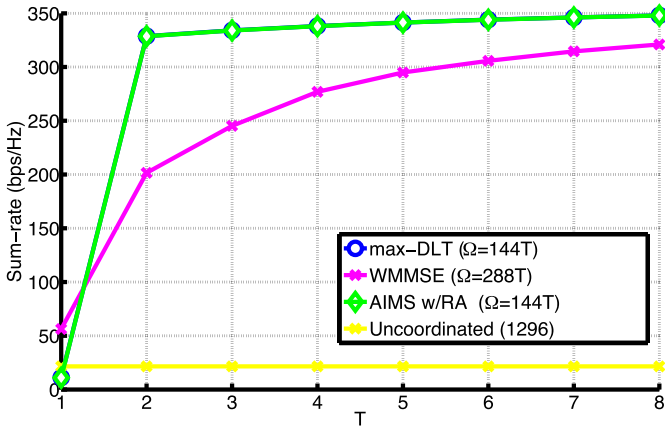


Fig. 9. Ergodic sum-rate vs T for dense downlink ($M = 16$, $N = 4$, $d = 1$, $L = 9$, $K = 8$).

that WMMSE is expected to deliver, is reached after more iterations are performed. The reason behind this behavior is the fast-converging nature of our algorithms, allowing them to reach a good operating point, in just 2 iterations. In the case of the max-DLT, this is due to the stream control feature of the non-homogeneous waterfilling.

While Figs. 6-7 have the same configuration (same number of BS/MS antennas, data streams, users and cells), as far as the algorithms are concerned, M and N are different ($N = 32, M = 4$ in the uplink, and $N = 4, M = 32$ in the downlink). Consequently, comparing the performance of any scheme in both configurations, is not informative. This is further reinforced by the absence of any result linking sum-rate, for the MIMO IMAC and the MIMO IBC (unlike sum-rate in the MIMO MAC and the MIMO BC, related by duality).

C. Part II

In this part we use a more comprehensive simulation setup. For lack of channel measurements in the $6-18$ GHz band, we will use recent results in the 28 GHz band [35], in the non line-of-sight setting. The MIMO channel $\boldsymbol{H}_{l,i_k}$, from user $i_k \in \mathcal{I}$ to BS $l \in \mathcal{L}$, has coefficients

$$[\boldsymbol{H}_{l,i_k}]_{p,q} = \sqrt{\gamma_{l,i_k}} g_{p,q}, \quad \forall (p,q) \in \{N\} \times \{M\}. \tag{30}$$

In the above, $(\gamma_{l,i_k})_{dB} = 20\log_{10}(4\pi/\lambda_c) + 10\ n_p \log_{10}(D_{l,i_k}) + \psi_{l,i_k}$ is the pathloss between user $i_k \in \mathcal{I}$ and BS $l \in \mathcal{L}$, where D_{l,i_k} is the corresponding distance, $n_p = 3.4$ the pathloss exponent, λ_c the carrier wavelength (corresponding to 28 GHz), and ψ_{l,i_k} is log-normal with zero mean and variance of 9dB [3]. Moreover, $g_{p,q}$ follows a Rician distribution with zero mean and unit variance, to model the line-of-sight components. We consider a 'dense' multi-user multi-cell setup with $L = 9$ cells, each with radius 10m and serving $K = 8$ users (dropped uniformly within the cell). We investigate both uplink and downlink communication.

Dense Multi-user Multi-cell uplink: BSs are equipped with $N = 8$ antennas, and MSs with $M = 4$, sending $d = 2$ streams each. The ergodic sum-rage is shown in Fig. 9, as a function of T, and the average effective SNR (including pathloss) is around 19dB. As we can see in Fig. 8, max-DLT provides significant gains over the other schemes. Moreover, the poor performance of the uncoordinated scheme further confirms our motivation for this work: mmWave systems in the lower bands, are not selective enough to bypass the need for interference management; indeed *any* of the coordination schemes considered here doubles the sum-rate performance (Fig.8). Note that the overhead for max-DLT at $T = 3$, is similar to that of the uncoordinated transmission, while providing around three times higher sum-rate.

Dense Multi-user Multi-cell downlink: We also investigate another 'dense' setting ($M = 16, N = 4, d = 1$, $L = 9, K = 8$), with downlink communication (using the same simulation setup). The power constraint on WMMSE follows the same convention as the large-scale multi-cell downlink above, and the observed trends are still the same (as seen in Fig. 9). Despite the unfavourable setup, our schemes offer a massive performance gain over WMMSE, with half the resulting overhead (the curves for max-DLT and AIMS w/RA are overlapping). We reiterate the fact that this is due to their ability to shut down streams with low-SINR, and thus converging to good sum-rate points, in less than 2 iterations.

While max-DLT has the distinct feature of quickly converging (due to the built-in stream-control), WMMSE has

a slower convergence converge to a stationary point of the sum-rate. In this work, we leverage this feature of max-DLT to significantly reduce the overhead (refer to corresponding Ω in Figs. 7,9). However, in general, the performance of WMMSE will exceed that of max-DLT, as T increases: this occurs at $T = 6$ iterations in the case of Fig. 7, and $T > 8$ in Fig. 9).

D. Discussions

As mentioned earlier, the non-homogeneous waterfilling solution clearly shows that streams that have low SINR are turned-off, and power is only allocated to the ones that exhibit relatively high SINR. This greatly speeds up the convergence of max-DLT, and allows it to achieve its required performance, with that limited number of F-B iterations (e.g., 2). On the other hand, due to the large dimensions inherent to lower band mmWave systems (i.e., more antennas, cells, users) other benchmarks will require more iterations to reach a similar performance. In addition, the rank-adaptation feature in AIMS offers a trade-off between reducing interference and diversity of the signal: however, in interference-limited scenarios, the former is more critical than the latter. The improved performance from RA, shown in the numerical results section, validates this proposition. As for the overhead, our schemes are based on the framework of F-B iterations and result in minimal overhead (the overhead consisting of uplink/downlink pilots only). However, other schemes such as WMMSE and CCP-WMMSE require additional pilots and feedback, and result in significantly higher overhead (as detailed in Sec. V-D). Moreover, the evaluations from the realistic dense uplink and downlink setup both conclude that interference management is a vital component in sub-28 GHz mmWave systems.

VII. Conclusion

In this work, we shed light on the need for interference management in lower bands of the mmWave spectrum, while highlighting the inapplicability of conventional approaches for distributed coordination. We thus proposed AIMS, a distributed algorithm that alternately maximizes the separability metric, for both the uplink and downlink, and established the fact that this is a generalization of the well-known max-SINR. Moreover, we advocated the use of DLT bounds, and highlighted their significant advantage in yielding optimization problems that decouple at both the transmitters and receivers. We provided a generic solution to this problem, the so-called non-homogeneous waterfilling (underlining its built-in stream-control feature), and proposed another distributed algorithm, max-DLT, that solves the problem in a distributed manner. Convergence to a stationary point of the DLT bound was also established. We later verified through extensive simulations that our proposed algorithms significantly outperform other well-known schemes, in the desired low-overhead regime (while still requiring less overhead). Moreover, the results also confirmed the need for interference management, and that the proposed approaches are a good fit for the this task.

Appendix

A. Sketch of Proof for Proposition 1

We need to show that $(\boldsymbol{U}_{l_j}^{\star})^{\dagger}\boldsymbol{U}_{l_j}^{\star} \approx \alpha \boldsymbol{I}_d$ happens with probability zero. Note that due to the i.i.d. nature, of MIMO channel coefficients, the eigenvalues of $\boldsymbol{Q}_{l_j}$ can be assumed to be distinct (and $\boldsymbol{Q}_{l_j}$ is full rank), almost surely. Then, it can be verified that the same holds for $\boldsymbol{L}_{l_j}, \boldsymbol{L}_{l_j}^{\dagger}$ and $(\boldsymbol{L}_{l_j}^{\dagger}\boldsymbol{L}_{l_j})^{-1}$. Then, $(\boldsymbol{L}_{l_j}^{\dagger}\boldsymbol{L}_{l_j})^{-1} \approx \alpha \boldsymbol{I}_N$ happens with probability zero. Recalling that $\boldsymbol{\Psi}_{l_j}^{\dagger}\boldsymbol{\Psi}_{l_j} = \boldsymbol{I}_d$, then the following equivalent statements happen with probability zero,

$$\begin{aligned}
&\boldsymbol{\Psi}_{l_j}^{\dagger}(\boldsymbol{L}_{l_j}^{\dagger}\boldsymbol{L}_{l_j})^{-1}\boldsymbol{\Psi}_{l_j} \\
&\quad\approx \boldsymbol{\Psi}_{l_j}^{\dagger}(\alpha \boldsymbol{I}_N)\boldsymbol{\Psi}_{l_j} \\
&\quad\Leftrightarrow (\boldsymbol{\Psi}_{l_j}^{\dagger}\boldsymbol{L}_{l_j}^{-1})(\boldsymbol{L}_{l_j}^{-\dagger}\boldsymbol{\Psi}_{l_j}) \approx \alpha \boldsymbol{I}_d \Leftrightarrow \ (\boldsymbol{U}_{l_j}^{\star})^{\dagger}\boldsymbol{U}_{l_j}^{\star} \approx \alpha \boldsymbol{I}_d
\end{aligned}$$

B. Proof of Proposition 2

We start by lower bounding the user rate in (3), as

$$\begin{aligned}
r_{l_j} &\geq \log_2 |(\boldsymbol{U}_{l_j}^{\dagger}\boldsymbol{Q}_{l_j}\boldsymbol{U}_{l_j})^{-1} + (\boldsymbol{U}_{l_j}^{\dagger}\boldsymbol{R}_{l_j}\boldsymbol{U}_{l_j})(\boldsymbol{U}_{l_j}^{\dagger}\boldsymbol{Q}_{l_j}\boldsymbol{U}_{l_j})^{-1}| \\
&= \log_2 |(\boldsymbol{I}_d + \boldsymbol{U}_{l_j}^{\dagger}\boldsymbol{R}_{l_j}\boldsymbol{U}_{l_j})(\boldsymbol{U}_{l_j}^{\dagger}\boldsymbol{Q}_{l_j}\boldsymbol{U}_{l_j})^{-1}| \\
&= \log_2 |\boldsymbol{I}_d + \boldsymbol{U}_{l_j}^{\dagger}\boldsymbol{R}_{l_j}\boldsymbol{U}_{l_j}| - \log_2 |\boldsymbol{U}_{l_j}^{\dagger}\boldsymbol{Q}_{l_j}\boldsymbol{U}_{l_j}| \\
&\geq \log_2 |\boldsymbol{I}_d + \boldsymbol{U}_{l_j}^{\dagger}\boldsymbol{R}_{l_j}\boldsymbol{U}_{l_j}| - \mathrm{tr}(\boldsymbol{U}_{l_j}^{\dagger}\boldsymbol{Q}_{l_j}\boldsymbol{U}_{l_j}) \triangleq r_{l_j}^{(LB)} \qquad (31)
\end{aligned}$$

where the first inequality follows from combining A2) in (13), and the monotonically increasing nature of $\log|\boldsymbol{X}|$. Moreover the last one follows from using $\log|\boldsymbol{A}| \leq \mathrm{tr}(\boldsymbol{A})$ for $\boldsymbol{A} \succeq \boldsymbol{0}$. We rewrite r_{l_j} in (3) as,

$$\begin{aligned}
r_{l_j} &= \log_2 |(\boldsymbol{U}_{l_j}^{\dagger}\boldsymbol{R}_{l_j}\boldsymbol{U}_{l_j})(\boldsymbol{U}_{l_j}^{\dagger}\boldsymbol{Q}_{l_j}\boldsymbol{U}_{l_j})^{-1}[\boldsymbol{I}_d \\
&\quad + (\boldsymbol{U}_{l_j}^{\dagger}\boldsymbol{Q}_{l_j}\boldsymbol{U}_{l_j})(\boldsymbol{U}_{l_j}^{\dagger}\boldsymbol{R}_{l_j}\boldsymbol{U}_{l_j})^{-1}]| \\
&= \log_2 |(\boldsymbol{U}_{l_j}^{\dagger}\boldsymbol{R}_{l_j}\boldsymbol{U}_{l_j})(\boldsymbol{U}_{l_j}^{\dagger}\boldsymbol{Q}_{l_j}\boldsymbol{U}_{l_j})^{-1}| \\
&\quad + \log_2 |\boldsymbol{I}_d + (\boldsymbol{U}_{l_j}^{\dagger}\boldsymbol{Q}_{l_j}\boldsymbol{U}_{l_j})(\boldsymbol{U}_{l_j}^{\dagger}\boldsymbol{R}_{l_j}\boldsymbol{U}_{l_j})^{-1}| \\
&= \log_2 |\boldsymbol{U}_{l_j}^{\dagger}\boldsymbol{R}_{l_j}\boldsymbol{U}_{l_j}| - \log_2 |\boldsymbol{U}_{l_j}^{\dagger}\boldsymbol{Q}_{l_j}\boldsymbol{U}_{l_j}| \\
&\quad + O(\mathrm{tr}[(\boldsymbol{U}_{l_j}^{\dagger}\boldsymbol{Q}_{l_j}\boldsymbol{U}_{l_j})(\boldsymbol{U}_{l_j}^{\dagger}\boldsymbol{R}_{l_j}\boldsymbol{U}_{l_j})^{-1}])
\end{aligned}$$

Thus, r_{l_j} is approximated by $\log_2 |\boldsymbol{U}_{l_j}^{\dagger}\boldsymbol{R}_{l_j}\boldsymbol{U}_{l_j}| - \log_2 |\boldsymbol{U}_{l_j}^{\dagger}\boldsymbol{Q}_{l_j}\boldsymbol{U}_{l_j}|$ (where the error is given in the above equation). Plugging this result in Δ_{l_j} yields,

$$\begin{aligned}
\Delta_{l_j} &= \log_2 |\boldsymbol{U}_{l_j}^{\dagger}\boldsymbol{R}_{l_j}\boldsymbol{U}_{l_j}| - \log_2 |\boldsymbol{U}_{l_j}^{\dagger}\boldsymbol{Q}_{l_j}\boldsymbol{U}_{l_j}| \\
&\quad - [\log_2 |\boldsymbol{I}_d + \boldsymbol{U}_{l_j}^{\dagger}\boldsymbol{R}_{l_j}\boldsymbol{U}_{l_j}| - \mathrm{tr}(\boldsymbol{U}_{l_j}^{\dagger}\boldsymbol{Q}_{l_j}\boldsymbol{U}_{l_j})] \\
&\quad + O(\mathrm{tr}[(\boldsymbol{U}_{l_j}^{\dagger}\boldsymbol{Q}_{l_j}\boldsymbol{U}_{l_j})(\boldsymbol{U}_{l_j}^{\dagger}\boldsymbol{R}_{l_j}\boldsymbol{U}_{l_j})^{-1}])
\end{aligned}$$

Referring to the above, in the interference-limited regime, i.e., A1) in (13), the first and third terms become negligible w.r.t. the second and fourth. Consequently,

$$\begin{aligned}
\Delta_{l_j} &= \mathrm{tr}(\boldsymbol{U}_{l_j}^{\dagger}\boldsymbol{Q}_{l_j}\boldsymbol{U}_{l_j}) - \log_2 |\boldsymbol{U}_{l_j}^{\dagger}\boldsymbol{Q}_{l_j}\boldsymbol{U}_{l_j}| \\
&\quad + O(\mathrm{tr}[(\boldsymbol{U}_{l_j}^{\dagger}\boldsymbol{Q}_{l_j}\boldsymbol{U}_{l_j})(\boldsymbol{U}_{l_j}^{\dagger}\boldsymbol{R}_{l_j}\boldsymbol{U}_{l_j})^{-1}])
\end{aligned}$$

C. Proof of Lemma 2

We rewrite the problem into a series of equivalent forms. Letting $\boldsymbol{Z} = \boldsymbol{L}^\dagger \boldsymbol{X} \Leftrightarrow \boldsymbol{X} = \boldsymbol{L}^{-\dagger}\boldsymbol{Z}$, then (P) in (23) is equivalent to,

$$(P2)\ \begin{cases} \min\limits_{\boldsymbol{Z}}\ f(\boldsymbol{Z}) \triangleq \mathrm{tr}(\boldsymbol{Z}^\dagger \boldsymbol{Z}) - \log_2 |\boldsymbol{I}_d + \boldsymbol{Z}^\dagger \boldsymbol{M} \boldsymbol{Z}| \\ \text{s. t. } \mathrm{tr}(\boldsymbol{Z}^\dagger \boldsymbol{A} \boldsymbol{Z}) = \zeta \end{cases}$$

where $\boldsymbol{A} = (\boldsymbol{L}^\dagger \boldsymbol{L})^{-1}$. Letting $\boldsymbol{Z} = \boldsymbol{T}\boldsymbol{\Sigma}\boldsymbol{V}^\dagger$ be the SVD of $\boldsymbol{Z}$ ($\boldsymbol{T} \in \mathbb{C}^{n\times r}$, $\boldsymbol{\Sigma} \in \mathbb{R}^{r\times r}$) we rewrite $(P2)$ into an equivalent form,

$$(P3)\ \begin{cases} \min\limits_{\boldsymbol{T},\boldsymbol{\Sigma}}\ \mathrm{tr}(\boldsymbol{\Sigma}^2) - \log_2 |\boldsymbol{I}_d + \boldsymbol{\Sigma}^2 \boldsymbol{T}^\dagger \boldsymbol{M} \boldsymbol{T}| \\ \text{s. t. } \mathrm{tr}(\boldsymbol{\Sigma}^2 \boldsymbol{T}^\dagger \boldsymbol{A} \boldsymbol{T}) = \zeta \end{cases}$$

Let us first look only at the objective in $(P3)$, to illustrate the argument. Note that for any given $\boldsymbol{\Sigma}$, the optimal $\boldsymbol{T}$ is given by $\boldsymbol{T}^\star \triangleq v_{1:r}[\boldsymbol{M}] = \boldsymbol{\Psi}$. [6] Moreover, $\boldsymbol{T}^\star$ does not depends on $\boldsymbol{\Sigma}$ (only on $\boldsymbol{M}$): thus, $\boldsymbol{T}^\star$ can be plugged into the objective, and one can solve for $\boldsymbol{\Sigma}$. Though the presence of a constraint makes this approach not suitable in general, note that this particular constraint allows scaling of the optimal solution, to always satisfy the constraint (this becomes clear when we express the problem as a function of the columns of $\boldsymbol{T}$, and the diagonal entries in $\boldsymbol{\Sigma}$). This can also be checked by considering solutions of the form $\boldsymbol{T} \neq v_{1:r}[\boldsymbol{M}]$, and showing that they cannot be optimal. With that in mind, the feasible set of $(P3)$ becomes $\mathrm{tr}(\boldsymbol{\Sigma}^2 \boldsymbol{\Psi}^\dagger \boldsymbol{A} \boldsymbol{\Psi}) = \sum_i \sigma_i^2 \beta_i$, where $\{\sigma_i\}$ are the diagonal elements of $\boldsymbol{\Sigma}$. Thus, $(P3)$ becomes,

$$(P4)\ \begin{cases} \min\limits_{\{\sigma_i\}}\ \sum_{i=1}^r \left(\sigma_i^2 - \log_2(1+\alpha_i\sigma_i^2)\right) \\ \text{s. t. } \sum_{i=1}^r \beta_i \sigma_i^2 = \zeta \end{cases}$$

Letting $x_i = \sigma_i^2$, we can rewrite the problem as,

$$(P5)\ \begin{cases} \min\limits_{\{x_i\}}\ \sum_{i=1}^r \left(x_i - \log_2(x_i + \dfrac{1}{\alpha_i})\right) \\ \text{s. t. } \sum_{i=1}^r \beta_i x_i = \zeta, \quad x_i \geq 0, \quad \forall i \end{cases}$$

$(P5)$ is a generalization of the well-known waterfilling problem: in fact, $(P5)$ reduces to the waterfilling problem, if $\beta_i = 1, \forall i$, and by dropping the first term in the objective. We start by writing the associated KKT conditions.

$$\begin{cases} 1 - (x_i + \alpha_i^{-1})^{-1} + \mu\beta_i - \lambda_i = 0, \quad \forall i \\ \sum_i \beta_i x_i = \zeta, \quad x_i \geq 0 \\ \lambda_i x_i = 0, \quad \lambda_i \geq 0, \quad \mu \neq 0, \quad \forall i \end{cases}$$

Firstly, note that λ_i act as slack variables and can thus easily be eliminated. Considering two cases, $\lambda_i = 0, \forall i$ or $\lambda_i > 0, \forall i$, the optimal solution can be easily found as,

$$\begin{aligned} x_i^\star &= \begin{cases} (1+\mu\beta_i)^{-1} - \alpha_i^{-1}, \text{ if } \mu < (\alpha_i - 1)/\beta_i \\ 0, \text{ if } \mu > (\alpha_i - 1)/\beta_i \end{cases} \\ &= \left(1/(1+\mu^\star\beta_i) - 1/\alpha_i\right)^+, \forall i \end{aligned}$$

[6] This follows from maximizing $\log_2 |\boldsymbol{I}_d + \boldsymbol{T}^\dagger \boldsymbol{M} \boldsymbol{T}|$, over the set of unitary matrices $\boldsymbol{T}$. Recall that $\boldsymbol{U}^\star = \mathrm{argmax}_{\boldsymbol{U}^\dagger\boldsymbol{U}=\boldsymbol{I}} |\boldsymbol{I} + \boldsymbol{U}^\dagger \boldsymbol{S} \boldsymbol{U}| = \mathrm{argmax}_{\boldsymbol{U}^\dagger\boldsymbol{U}=\boldsymbol{I}} |\boldsymbol{U}^\dagger \boldsymbol{S} \boldsymbol{U}|$, $\boldsymbol{S} \succeq \boldsymbol{0}$. Then, it is well known that $\boldsymbol{U}^\star \triangleq v_{1:r}[\boldsymbol{S}]$ [36].

where $\mu^\star$ is the unique root to

$$g(\mu) \triangleq \sum_{i=1}^r \beta_i \left(1/(1+\mu\beta_i) - 1/\alpha_i\right)^+ - \zeta$$

Note that $g(\mu)$ is monotonically decreasing, for $\mu > -1/(\max_i \beta_i)$, and $\mu^\star$ can be found using standard 1D search methods, such as bisection. Thus, the optimal solution for $(J1)$ is $\boldsymbol{Z}^\star = \boldsymbol{\Psi}\boldsymbol{\Sigma}^\star$(where $\boldsymbol{\Sigma}^\star_{(i,i)} = \sqrt{x_i}, \forall i$), and that of (23) is $\boldsymbol{X}^\star = \boldsymbol{L}^{-\dagger}\boldsymbol{\Psi}\boldsymbol{\Sigma}^\star$

REFERENCES

[1] J. G. Andrews *et al.*, "What will 5G be?" *IEEE J. Sel. Areas Commun.*, vol. 32, no. 6, pp. 1065–1082, Jun. 2014.

[2] T. S. Rappaport *et al.*, "Millimeter wave mobile communications for 5G cellular: It will work!" *IEEE Access*, vol. 1, pp. 335–349, May 2013.

[3] M. K. Samimi and T. S. Rappaport, "Characterization of the 28 GHz millimeter-wave dense urban channel for future 5G mobile cellular," Tech. Rep. TR 2014-04, Mar. 2014

[4] S. Hur, T. Kim, D. J. Love, J. V. Krogmeier, T. A. Thomas, and A. Ghosh, "Millimeter wave beamforming for wireless backhaul and access in small cell networks," *IEEE Trans. Commun.*, vol. 61, no. 10, pp. 4391–4403, Oct. 2013.

[5] A. Alkhateeb, O. El Ayach, G. Leus, and R. W. Heath, Jr., "Channel estimation and hybrid precoding for millimeter wave cellular systems," *IEEE J. Sel. Topics Signal Process.*, vol. 8, no. 5, pp. 831–846, Oct. 2014.

[6] S. Scalise, H. Ernst, and G. Harles, "Measurement and modeling of the land mobile satellite channel at ku-band," *IEEE Trans. Veh. Technol.*, vol. 57, no. 2, pp. 693–703, Mar. 2008.

[7] I. K. Son, S. Mao, M. X. Gong, and Y. Li, "On frame-based scheduling for directional mmWave WPANs," in *Proc. IEEE INFOCOM*, Mar. 2012, pp. 2149–2157.

[8] H. Ghauch, T. Kim, M. Bengtsson, and M. Skoglund, "Subspace estimation and decomposition for large millimeter-wave MIMO systems," *IEEE J. Sel. Topics Signal Process.*, vol. 10, no. 3, pp. 528–542, Apr. 2016.

[9] E. J. Violette, R. H. Espeland, R. O. DeBolt, and F. K. Schwering, "Millimeter-wave propagation at street level in an urban environment," *IEEE Trans. Geosci. Remote Sens.*, vol. 26, no. 3, pp. 368–380, May 1988.

[10] Y. Niu, Y. Li, D. Jin, L. Su, and A. V. Vasilakos, "A survey of millimeter wave communications (mmWave) for 5G: opportunities and challenges," *Wireless Netw.*, vol. 21, no. 8, pp. 2657–2676, 2015.

[11] Ritayan Biswas, "Performance evaluation of coordinated multipoint techniques at millimeter wave frequencies," M.S. thesis, Faculty Comput. Elect. Eng., Univ. Tampere, Tampere, Finland, Mar. 2016.

[12] S. Rangan, T. S. Rappaport, and E. Erkip, "Millimeter-wave cellular wireless networks: Potentials and challenges," *Proc. IEEE*, vol. 102, no. 3, pp. 366–385, Mar. 2014.

[13] *Initial Report on Horizontal Topics, First Results and 5G System Concept*, document METIS D6.2, Mar. 2014.

[14] K. Gomadam, V. R. Cadambe, and S. A. Jafar, "A distributed numerical approach to interference alignment and applications to wireless interference networks," *IEEE Trans. Inf. Theory*, vol. 57, no. 6, pp. 3309–3322, Jun. 2011.

[15] H. Ghauch and C. Papadias, "Interference alignment: A one-sided approach," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, Dec. 2011, pp. 1–5.

[16] D. Schmidt, C. Shi, R. Berry, M. Honig, and W. Utschick, "Minimum mean squared error interference alignment," in *Proc. Conf. Rec. 43rd Asilomar Conf. Signals, Syst. Comput.*, Nov. 2009, pp. 1106–1110.

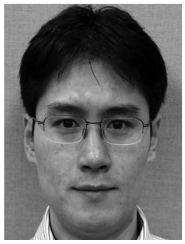
[17] S. W. Peters and R. W. Heath, Jr., "Cooperative algorithms for MIMO interference channels," *IEEE Trans. Veh. Technol.*, vol. 60, no. 1, pp. 206–218, Jan. 2011.

[18] Q. Shi, M. Razaviyayn, Z.-Q. Luo, and C. He, "An iteratively weighted MMSE approach to distributed sum-utility maximization for a MIMO interfering broadcast channel," *IEEE Trans. Signal Process.*, vol. 59, no. 9, pp. 4331–4340, Sep. 2011.

[19] I. Santamaria, O. Gonzalez, R. W. Heath, Jr., and S. W. Peters, "Maximum sum-rate interference alignment algorithms for MIMO channels," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, Dec. 2010, pp. 1–6.

[20] J. Kaleva, A. Tölli, and M. Juntti, "Weighted sum rate maximization for interfering broadcast channel via successive convex approximation," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, Dec. 2012, pp. 3838–3843.
[21] D. A. Schmidt, C. Shi, R. A. Berry, M. L. Honig, and W. Utschick, "Comparison of distributed beamforming algorithms for MIMO interference networks," *IEEE Trans. Signal Process.*, vol. 61, no. 13, pp. 3476–3489, Jul. 2013.
[22] P. Komulainen, A. Tölli, and M. Juntti, "Effective CSI signaling and decentralized beam coordination in TDD multi-cell MIMO systems," *IEEE Trans. Signal Process.*, vol. 61, no. 9, pp. 2204–2218, May 2013.
[23] D. H. N. Nguyen and T. Le-Ngoc, "Sum-rate maximization in the multicell MIMO multiple-access channel with interference coordination," *IEEE Trans. Wireless Commun.*, vol. 13, no. 1, pp. 36–48, Jan. 2014.
[24] R. Brandt and M. Bengtsson, "Fast-convergent distributed coordinated precoding for TDD multicell MIMO systems," in *Proc. IEEE 6th Int. Workshop Comput. Adv. Multi-Sensor Adapt. Process. (CAMSAP)*, Dec. 2015, pp. 457–460.
[25] H. Ghauch, T. Kim, M. Bengtsson, and M. Skoglund, "Distributed low-overhead schemes for multi-stream MIMO interference channels," *IEEE Trans. Signal Process.*, vol. 63, no. 7, pp. 1737–1749, Apr. 2015.
[26] C. M. Bishop, "Pattern recognition and machine learning," in *Information Science and Statistics*. Secaucus, NJ, USA: Springer-Verlag, 2006.
[27] R. Brandt and M. Bengtsson, "Distributed CSI acquisition and coordinated precoding for TDD multicell MIMO systems," *IEEE Trans. Veh. Technol.*, vol. 65, no. 5, pp. 2890–2906, May 2015.
[28] O. Oyman, R. U. Nabar, H. Bolcskei, and A. J. Paulraj, "Characterizing the statistical properties of mutual information in MIMO channels," *IEEE Trans. Signal Process.*, vol. 51, no. 11, pp. 2784–2795, Nov. 2003.
[29] R. Prieto, "A general solution to the maximization of the multidimensional generalized Rayleigh quotient used in linear discriminant analysis for signal classification," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, vol. 6. Apr. 2003, p. VI-157-60.
[30] S.-J. Kim and G. B. Giannakis, "Optimal resource allocation for MIMO ad hoc cognitive radio networks," *IEEE Trans. Inf. Theory*, vol. 57, no. 5, pp. 3117–3131, May 2011.
[31] P. Tseng, "Convergence of a block coordinate descent method for nondifferentiable minimization," *J. Optim. Theory Appl.*, vol. 109, no. 3, pp. 475–494, 2001.
[32] D. G. Luenberger and Y. Ye, *Linear and Nonlinear Programming*. New York, NY, USA: Springer, 1973.
[33] T. Lipp and S. Boyd, "Variations and extension of the convex–concave procedure," *Optim. Eng.*, vol. 17, no. 2, pp. 263–287, 2016.
[34] K. Schittkowski and C. Zillober, "Stochastic programming: Numerical techniques and engineering applications," in *Sequential Convex Programming Methods*. Berlin, Germany: Springer, 1995, pp. 123–141.
[35] G. R. MacCartney, Jr., M. K. Samimi, and T. S. Rappaport, "Omnidirectional path loss models in New York City at 28 GHz and 73 GHz," in *Proc. IEEE 25th Int. Symp. Pers. Indoor Mobile Radio Commun. (PIMRC)*, Sep. 2014, pp. 227–231.
[36] G. H. Golub and C. F. Van Loan, *Matrix Computations*, 3rd ed. Baltimore, MD, USA: Johns Hopkins Univ. Press, 1996.

Hadi Ghauch (S'06) received the Ph.D. degree in electrical engineering from the KTH Royal Institute of Technology, Stockholm, Sweden, in 2016, and the M.S. degree (Hons.) in information networking from Carnegie Mellon University, USA, in 2011. He was a Visiting Graduate Research Assistant with the City University of Hong Kong in 2014. He is currently a Post-Doctoral Researcher with the Department of Information Science and Engineering, KTH Royal Institute of Technology. His focus is on optimization techniques for signal processing/communication and learning. He was a recipient of several awards, namely, the Fulbright Scholarship and the KTH-EE Excellence Scholarship. He serves as a Reviewer for several IEEE journals, such as the IEEE TRANSACTIONS ON SIGNAL PROCESSING, the IEEE TRANSACTIONS ON COMMUNICATIONS, and the IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS.



Taejoon Kim (S'08–M'11) received the B.S. degree from Sogang University (with highest Hons.) in 2002, M.S. degree from KAIST, Korea, in 2004, and Ph.D. degree from Purdue University, IN, USA in 2011, all in electrical engineering. From 2011 to 2012, he was with the Nokia Research Center (NRC), CA, USA. Before joining City University of Hong Kong as an assistant professor in 2013, he was a postdoctoral researcher at KTH, Sweden. He was the recipient of The President's Award, City University of Hong Kong in 2017, the IEEE TRANSACTIONS ON COMMUNICATIONS Best Paper Award (Stephen O. Rice Prize) in 2016, and IEEE PIMRC Best Paper Award in 2012. He has served as an Associate Editor for the IEEE TRANSACTIONS ON COMMUNICATIONS since 2016.

Mats Bengtsson (M'08–SM'06) received the M.S. degree in computer science from Linkping University, Linkping, Sweden, in 1991, and the Tech.Lic. and Ph.D. degrees in electrical engineering from the KTH Royal Institute of Technology, Stockholm, Sweden, in 1997 and 2000, respectively. From 1991 to 1995, he was with Ericsson Telecom AB, Karlstad. He is currently a Professor with the Information Science and Engineering Department, School of Electrical Engineering, KTH Royal Institute of Technology. His research interests include statistical signal processing and optimization theory and its applications to communications, multi-antenna processing, cooperative communication, radio resource management, and propagation channel modeling. He was the Technical Chair of the IEEE SPAWC 2015 and a member of the IEEE SPCOM Technical Committee 2007–2012. He serves as an Associate Editor of the IEEE TRANSACTIONS ON SIGNAL PROCESSING.

Mikael Skoglund (S'93–M'97–SM'04) received the Ph.D. degree from the Chalmers University of Technology, Sweden, in 1997. In 1997, he joined the KTH Royal Institute of Technology, Stockholm, Sweden, where he was the Chair in communication theory in 2003. He is currently the Head of the Department of Information Science and Engineering, KTH Royal Institute of Technology, and the Assistant Dean of the School of Electrical Engineering. He is also a Founding Faculty Member of the ACCESS Linnaeus Center and the Director of the the Center Graduate School. He has been involved in problems in source–channel coding, coding and transmission for wireless communications, Shannon theory, information and control, and statistical signal processing. He has authored or co-authored over 130 journal and 300 conference papers. He holds six patents. He has served on numerous technical program committees for IEEE sponsored conferences. From 2003 to 2008, he was an Associate Editor of the IEEE TRANSACTIONS ON COMMUNICATIONS. From 2008 to 2012, he was on the Editorial Board of the IEEE TRANSACTIONS ON INFORMATION THEORY.