Precoding for Scalable Cell-free Massive MIMO with Radio Stripes

Lorenzo Miretti[1], Emil Björnson[2,3], and David Gesbert[1]

[1]Dept. of Communications Systems, EURECOM, Sophia Antipolis, France
[2]Dept. of Computer Science, KTH Royal Institute of Technology, Stockholm, Sweden
[3]Dept. of Electrical Engineering, Linköping University, Linköping, Sweden
{miretti,gesbert}@eurecom.fr, emilbjo@kth.se

***Abstract*—This paper studies a novel distributed precoding scheme for cell-free massive MIMO networks. Our scheme, coined *team minimum mean-square error* (TMMSE) precoding, generalizes classical centralized MMSE precoding to arbitrary patterns of channel state information (CSIT) sharing among the transmitters. Building on the so-called *theory of teams*, we show that designing the optimal TMMSE precoders is equivalent to solving an infinite dimensional linear system of equations. We solve the problem explicitly for two important CSIT sharing patterns, i.e., the classical case of purely local CSIT and the case of unidirectional CSIT sharing along a serial fronthaul. The latter scenario is relevant, e.g., for the recently proposed *radio stripes* concept. In both cases, our optimal design outperforms the heuristic methods that are known from the previous literature. Duality arguments and numerical simulations validate the effectiveness of the proposed schemes in terms of ergodic achievable rates under a sum power constraint.**

I. Introduction

One of the major barriers currently preventing the practical deployment of wireless communication systems capitalizing on transmit (TX) cooperation is the severe scalability issue arising from network-wide information sharing [1]. By advocating simple distributed precoding strategies which do not require the sharing of channel state information (CSIT) among the TXs, the cell-free massive MIMO paradigm emerged as a promising solution for making cooperative wireless networks feasible in practice [2]. From this original idea, several extensions have been proposed considering more involved CSIT sharing patterns and network clustering techniques to better shape the cooperation regime [3]–[7]. However, most of the available distributed precoding designs are essentially heuristic adaptations of known centralized schemes such as maximum-ratio transmission (MRT), zero-forcing (ZF), or minimum mean-square error (MMSE) precoding [8].

In contrast to the previous literature, in this work, we cast the distributed precoding design problem for cell-free massive MIMO in a general form which explicitly considers limited CSIT sharing. Specifically, we propose a novel scheme called *team* MMSE (TMMSE) precoding, which rigorously generalizes centralized MMSE precoding to arbitrary CSIT sharing patterns. Its optimality in terms of ergodic achievable rates estimated by the popular *hardening* bound [8] is motivated by means of the uplink-downlink (UL-DL) duality principle, assuming a sum power constraint. The main contribution is the derivation of a useful set of necessary and sufficient conditions for optimal TMMSE precoding design in the form of an infinite dimensional linear system of equations. The key novelty lies in the exploitation of previously unexplored elements from the *theory of teams*, a mathematical framework for multi-agent coordinated decision making in presence of asymmetry of information, popularized by the economic and control theoretical literature [9], [10]. As a first non-trivial application, we derive the optimal TMMSE precoders based on local CSIT only, improving upon known local precoding schemes studied, e.g., in [2], [6]. We then derive the optimal TMMSE precoders by assuming that CSIT is shared unidirectionally along a serial fronthaul, an architecture also known as a *radio stripe* [11], [12]. The proposed scheme can be efficiently implemented in a sequential fashion, an idea that has been explored already in [11], [12] for uplink processing, and in [13] under a different cellular context. As a byproduct, we also obtain a novel distributed implementation of centralized MMSE precoding.

Notation: We reserve italic letters (e.g., a, A) for scalars and functions, boldface letters (e.g., $\mathbf{a}$, $\mathbf{A}$) for vectors and matrices, and calligraphic letters (e.g., $\mathcal{A}$) for sets. The operators $(\cdot)^{\mathsf{T}}$, $(\cdot)^{\mathsf{H}}$ denote respectively the transpose and Hermitian transpose, and $\|\cdot\|$ is the Euclidean norm. We denote by $\mathbf{I}_n$ the identity matrix of dimension n, and by $\mathbf{0}_{n\times m}$ an $n\times m$ matrix of zeros, omitting the subscripts when no confusion arises. We use $\mathrm{diag}(\mathbf{A}_1,\dots,\mathbf{A}_n)$ to denote a block-diagonal matrix with the matrices $\mathbf{A}_1,\dots,\mathbf{A}_n$ on its diagonal. We use $\prod_{i=l'}^{l}\mathbf{A}_i := \mathbf{A}_l\mathbf{A}_{l-1}\dots\mathbf{A}_{l'}$ for integers $l\geq l'\geq 1$ to denote the *left* product chain of $l-l'+1$ matrices of compatible dimension, with the convention $\prod_{i=l'}^{l}\mathbf{A}_i=\mathbf{I}$ for $l<l'$.

II. System model and preliminaries

A. Channel model

Consider a network of L TXs indexed by $\mathcal{L}:=\{1,\dots,L\}$, each of them equipped with N antennas, and K single-antenna receivers (RXs) indexed by $\mathcal{K}:=\{1,\dots,K\}$. Let an arbitrary channel use be governed by the MIMO channel law

$$\mathbf{y}=\sum_{l\in\mathcal{L}}\mathbf{H}_l\mathbf{x}_l+\mathbf{n} \tag{1}$$

where the k-th element of $\mathbf{y}\in\mathbb{C}^K$ is the received signal at RX k, $\mathbf{H}_l\in\mathbb{C}^{K\times N}$ is a sample of a stationary ergodic random process modelling the fading between TX l and all RXs,

This work received partial support from the Huawei funded Chair on Future Wireless Networks at EURECOM. E. Björnson was supported by the Grant 2019-05068 from the Swedish Research Council.

$\mathbf{x}_l \in \mathbb{C}^N$ is the transmitted signal at TX l, and $\mathbf{n} \sim \mathcal{CN}(\mathbf{0}, \mathbf{I})$ is a sample of a white noise process. This channel model is relevant, e.g, for narrowband or wideband OFDM systems [8]. For most parts of this work, we do not specify the distribution of the channel matrix $\mathbf{H} := \begin{bmatrix}\mathbf{H}_1, \ldots, \mathbf{H}_L\end{bmatrix}$. However, we reasonably assume the channel submatrices corresponding to different TX-RX pairs to be mutually independent, and that each channel coefficient is governed by a physically consistent fading distribution with bounded support. Furthermore, we focus on $N < K$, that is, on the regime where cooperation is necessary for effective spatial interference management [1].

B. Distributed precoding with limited CSIT sharing

In the scope of this study, the main feature of the cell-free massive MIMO paradigm is the use of TDD operations to efficiently acquire the *local* channel $\mathbf{H}_l$ at each TX l via over-the-uplink training [2]–[6]. These local measurements may be subsequently shared through the fronthaul according to some predefined CSIT sharing mechanism, forming at each TX l (or at the network equipment virtualizing its physical layer operations) some *individual* knowledge $\hat{\mathbf{H}}_l$ of the *global* channel matrix $\mathbf{H}$. Consistently, in this work, we let

$$\hat{\mathbf{H}}_l := \mathbf{H}\mathbf{\Omega}_l, \quad \mathbf{\Omega} := \mathrm{diag}(\mathbf{\Omega}_{l,1}, \ldots, \mathbf{\Omega}_{l,L}), \tag{2}$$

where $\mathbf{\Omega}_{l,j} = \mathbf{I}_N$ if TX j shares its local channel $\mathbf{H}_j$ to TX l, and $\mathbf{\Omega}_{l,j} = \mathbf{0}_{N\times N}$ otherwise. Clearly, $\mathbf{\Omega}_{l,l} = \mathbf{I}_N$ always holds. This model captures limited CSIT sharing, but does not consider for simplicity local CSIT estimation errors, nor imperfect fronthaul signalling. The extension of the results presented in this work to more general CSIT models requires significant mathematical sophistication, and is given in the journal version of this work in preparation [14].

We then let each TX l form its transmit signal according to the following distributed linear precoding scheme:

$$\mathbf{x}_l = \sum_{k\in\mathcal{K}} \mathbf{t}_{l,k} s_k, \quad \mathbf{t}_{l,k} = f_{l,k}(\hat{\mathbf{H}}_l), \tag{3}$$

where $s_k \sim \mathcal{CN}(0, P_k)$ is the independently encoded message for RX k, shared by all TXs, and where $\mathbf{t}_{l,k} \in \mathbb{C}^N$ is a linear precoder applied at TX l to message s_k based only on $\hat{\mathbf{H}}_l$ according to a function $f_{l,k}$ belonging to the space $\mathcal{F}_l$ of square-integrable $\hat{\mathbf{H}}_l$-measurable functions, i.e., such that $\mathbb{E}[\|f_{l,k}(\hat{\mathbf{H}}_l)\|^2] < \infty$. This last assumption corresponds to a reasonable constraint $\mathbb{E}[\|\mathbf{t}_{l,k}\|^2] < \infty$ on precoders' power. We then denote the full precoding vector for message s_k by $\mathbf{t}_k^\mathsf{T} := \begin{bmatrix}\mathbf{t}_{1,k}^\mathsf{T} & \ldots & \mathbf{t}_{L,k}^\mathsf{T}\end{bmatrix}^\mathsf{T}$, and the corresponding tuple of functions by $f_k := (f_{1,k}, \ldots, f_{L,k}) \in \mathcal{F} := \prod_{l=1}^L \mathcal{F}_l$. Finally, we assume the first order distribution of $\mathbf{H}$ to be known at all TXs, which is reasonable if the fading process is indeed approximately stationary for a sufficiently long time span.

C. Performance metric

We measure the network performance under the specified transmission scheme by using achievable ergodic rates estimated by the popular *hardening* bound [8]

$$R_k^{\mathrm{hard}} := \log\left(1 + \frac{P_k|\mathbb{E}[\mathbf{g}_k^\mathsf{H}\mathbf{t}_k]|^2}{P_k\mathrm{Var}[\mathbf{g}_k^\mathsf{H}\mathbf{t}_k] + \sum_{j\neq k} P_j \mathbb{E}[|\mathbf{g}_k^\mathsf{H}\mathbf{t}_j|^2] + 1}\right), \tag{4}$$

where $\begin{bmatrix}\mathbf{g}_1 & \ldots & \mathbf{g}_K\end{bmatrix} := \mathbf{H}^\mathsf{H}$. The rates in (4) are achievable by treating interference as noise (TIN), without channel state information at the RX (CSIR), and without exploiting any memory across the realizations of $\mathbf{H}$ [15].

We then let $\mathcal{R}^{\mathrm{hard}}$ be the union of all rate tuples $(R_1, \ldots, R_K) \in \mathbb{R}^K$ such that $R_k \leq R_k^{\mathrm{hard}}\ \forall k \in \mathcal{K}$ for some set of distributed precoders $\{f_k\}_{k=1}^K$ and power allocation policy $\{P_k\}_{k=1}^K$ satisfying a long-term sum power constraint $\sum_{l=1}^L \mathbb{E}[\|\mathbf{x}_l\|^2] \leq P_{\mathrm{sum}} < \infty$. Due to its importance in system design and resource allocation, we consider the notion of (weak) Pareto optimality on $\mathcal{R}^{\mathrm{hard}}$ and we mostly focus on the (weak) Pareto boundary $\partial\mathcal{R}^{\mathrm{hard}}$ of $\mathcal{R}^{\mathrm{hard}}$. The sum power constraint is chosen because it allows for strong analytical results and simplifies system design. This constraint may be directly relevant for systems such as the radio stripes, where all the TXs share the same power supply [11]. However, note that many simple heuristic methods (such as power scaling factors) can be applied to adapt systems designed under a sum power constraint to a more restrictive per-TX power constraint.

III. Team MMSE precoding

In this work, we study the following novel *team* MMSE precoding design criterion: given a vector of nonnegative weights $\mathbf{w} := [w_1, \ldots, w_K]$ belonging to $\mathcal{W} := \{\mathbf{w} \in \mathbb{R}_+^K \mid \sum_{k=1}^K w_k = K\}$, we consider the problem

$$\underset{f_k\in\mathcal{F}}{\mathrm{minimize}}\ \mathrm{MSE}_k(f_k) := \mathbb{E}\left[\left\|\mathbf{W}^{\frac{1}{2}}\mathbf{H}\mathbf{t}_k - \mathbf{e}_k\right\|^2 + \frac{\|\mathbf{t}_k\|^2}{P}\right], \tag{5}$$

where $\mathbf{W} := \mathrm{diag}(w_1, \ldots, w_K)$, $\mathbf{e}_k$ is the k-th column of $\mathbf{I}_K$, and $P := P_{\mathrm{sum}}/K$. A solution to the above problem can be recognized as a distributed version of the classical centralized MMSE precoding design [8]. By means of team theoretical arguments [9], [10], this section provides fundamental properties of the optimal solution to Problem (5). Before providing the main results of this section, we also motivate the effectiveness of the MSE criterion in (5) in terms of network performance, which is well-known for centralized precoding.

A. Achievable rates via uplink-downlink duality

In the following, we discuss the formal connection between the objective of Problem (5) and network performance by revisiting classical UL-DL duality arguments behind centralized MMSE precoding [8], [16] under arbitrary CSIT sharing.

Theorem 1. *Consider an arbitrary set of distributed precoders $\{f_k\}_{k=1}^K$ and weights $\mathbf{w} \in \mathcal{W}$. Then, any rate tuple $(R_1, \ldots, R_K) \in \mathbb{R}^K$ such that*

$$R_k \leq \log(\mathrm{MSE}_k(f_k))^{-1} \tag{6}$$

belongs to $\mathcal{R}^{\mathrm{hard}}$. Furthermore, if f_k solves Problem (5) $\forall k \in \mathcal{K}$, then $(R_1, \ldots, R_K)$ with $R_k = \log(\mathrm{MSE}_k(f_k))^{-1}$ is Pareto optimal, and $\partial\mathcal{R}^{\mathrm{hard}}$ is fully parametrized by $\mathcal{W}$.

Proof. The proof is based on observing that the solution to Problem (5) gives the rate-optimal *distributed* linear combiners in a dual UL channel, where $\mathbf{w}$ is a dual UL power allocation vector, and where achievable rates are measured by using the so-called *use-and-then-forget* (UatF) bound [8, Th. 4.4]. The proof follows by the duality principle between the hardening bound (4) and the UatF bound [8, Th. 4.8], which also provides the method for computing the optimal $\{P_k\}_{k=1}^K$. We omit the details due to space limitations. The detailed proof is available in the journal version of this work [14]. □

Theorem 1 states that, similarly to known results for deterministic channels (reviewed, e.g., in [16]), the Pareto boundary of $\mathcal{R}^{\text{hard}}$ is achieved by TMMSE precoding and can be parametrized by $K-1$ nonnegative real parameters, i.e., by the weights $\mathbf{w} \in \mathcal{W}$. In practice, $\mathbf{w}$ is often fixed heuristically as in classical MMSE design, while the network utility is optimized a posteriori by varying the power allocation policy $\{P_k\}_{k=1}^K$.

Remark 1. *Hereafter, we derive our main results by considering w.l.o.g. the operating point* $\mathbf{W} = \mathbf{I}$. *The general case will readily follow by replacing* $\mathbf{H}_l$ *with* $\mathbf{W}^{\frac{1}{2}}\mathbf{H}_l$ *everywhere.*

B. Quadratic teams for distributed precoding design

Problem (5) belongs to the known category of *team decision* problems [9], [10], which are generally difficult to handle for nontrivial information constraints $f_k \in \mathcal{F}$. However, we recognize that Problem (5) is an instance of a *quadratic team*, a particular type of team decision problem pioneered by Radner in [9] for which solid solution approches are available (see, e.g., [10]). In particular, we obtain the following result:

Theorem 2. *Problem* (5) *admits a unique optimal solution, which is given by the unique* $f_k^\star = (f_{1,k}^\star, \dots, f_{L,k}^\star) \in \mathcal{F}$ *satisfying* ($\forall l \in \mathcal{L}$)

$$f_{l,k}^\star(\hat{\mathbf{H}}_l) = \mathbf{F}_l \left(\mathbf{e}_k - \sum_{j \neq l} \mathbb{E}\left[\mathbf{H}_j f_{j,k}^\star(\hat{\mathbf{H}}_j) \middle| \hat{\mathbf{H}}_l \right] \right) \quad \text{a.s.,} \quad (7)$$

where $\mathbf{F}_l := \left(\mathbf{H}_l^{\mathsf{H}}\mathbf{H}_l + P^{-1}\mathbf{I}\right)^{-1}\mathbf{H}_l^{\mathsf{H}}$.

Proof. Problem (5) is a quadratic team as defined in [9, Sect. 4]. Furthermore, the bounded fading assumption and the square-integrability of $f_{l,k}$ satisfy the assumptions of [10, Th. 2.6.6], which provides a set of optimality conditions for quadratic teams here specialized to (7). □

Remark 2. *Theorem 2 does not cover the physically inconsistent yet useful Gaussian fading model, nor CSIT acquisition errors. The extension to these cases is not trivial and is provided in the journal version of this work [14].*

Theorem 2 establishes a set of necessary and sufficient optimality conditions in the form of an infinite dimensional linear system of equalities. Interestingly, $\mathbf{F}_l$ can be recognized as a *local* MMSE precoding matrix [6], while the remaining part can be interpreted as a 'corrective' stage which takes into account the other precoders based on $\hat{\mathbf{H}}_l$ and statistical information. It is generally difficult to solve (7) in closed form. However, the optimal TMMSE precoders may be approached via one of the many approximation methods available in the literature (see, e.g., [10]). Further discussions on approximate solution methods are left for future work, and hereafter we focus on special cases where (7) can be solved explicitly.

IV. Applications to radio stripes

In this section, we consider the cell-free massive MIMO network in Figure 1, where CSIT, messages, and power are distributed along a serial fronthaul from and/or towards a CPU located at one edge, an architecture also known as a *radio stripe* [11], [12]. The main result of this section is the solution to the optimality conditions in (7) for the cases of no CSIT sharing and unidirectional CSIT sharing depicted in Figure 1.

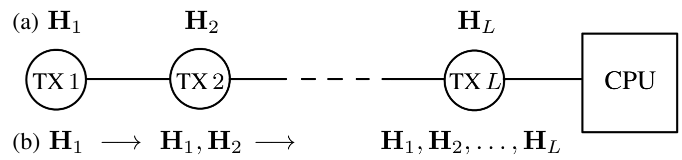

Fig. 1. Pictorial representation of a radio stripe with (a) no CSIT sharing, and (b) unidirectional CSIT sharing along the serial fronthaul.

A. No CSIT sharing

We assume that no local CSIT is shared along the fronthaul. Specifically, this can be modelled by the limited CSIT sharing pattern $\mathbf{\Omega}_{l,j} = \mathbf{0}_{N\times N}$, $\forall j \neq l$, $\forall l \in \mathcal{L}$. Note that this scenario is relevant for any fronthaul architecture.

Theorem 3. *The TMMSE precoders under no CSIT sharing are given by*

$$f_{l,k}^\star(\hat{\mathbf{H}}_l) = \mathbf{F}_l\mathbf{C}_l\mathbf{e}_k, \quad \forall l \in \mathcal{L}, \quad (8)$$

for some matrices of coefficients $\mathbf{C}_l \in \mathbb{C}^{K\times K}$. *Furthermore, the optimal* $\mathbf{C}_l$ *are given by the unique solution of the linear system* $\mathbf{C}_l + \sum_{j\neq l}\mathbf{\Pi}_j\mathbf{C}_j = \mathbf{I}$, $\forall l \in \mathcal{L}$, *where* $\mathbf{\Pi}_l := \mathbb{E}\left[\mathbf{H}_l\mathbf{F}_l\right]$.

Proof. By Theorem 2, (8) is the unique TMMSE solution if and only if it satisfies the optimality conditions (7), which are readily rewritten as ($\forall l \in \mathcal{L}$)

$$\mathbf{H}_l^{\mathsf{H}} \left(\mathbf{C}_l + \sum_{j\neq l} \mathbb{E}\left[\mathbf{H}_j\mathbf{F}_j\mathbf{C}_j \middle| \hat{\mathbf{H}}_l \right] - \mathbf{I} \right) \mathbf{e}_k = \mathbf{0} \quad \text{a.s.}$$

Since $\mathbf{H}_l$ and $\mathbf{H}_j$ are independent, we can drop the conditioning on $\hat{\mathbf{H}}_l$ and obtain $\mathbf{H}_l^{\mathsf{H}}\left(\mathbf{C}_l + \sum_{j\neq l}\mathbf{\Pi}_j\mathbf{C}_j - \mathbf{I}\right)\mathbf{e}_k = \mathbf{0}$ a.s., $\forall l \in \mathcal{L}$. It can be finally shown that the linear system $\mathbf{C}_l + \sum_{j\neq l}\mathbf{\Pi}_j\mathbf{C}_j = \mathbf{I}$, $\forall l \in \mathcal{L}$ always has a unique solution. The details are here omitted due to space limitations, and are provided in [14]. □

By assuming a zero-mean symmetric distribution for all channel coefficients, corresponding to a non-line-of-sight (NLoS) channel model, it can be shown that the matrices $\mathbf{\Pi}_l$ are diagonal and (8) takes the same form of a local MMSE precoding scheme (studied, e.g., in [6])

$$f_{l,k}^\star(\hat{\mathbf{H}}_l) = c_{l,k}\mathbf{F}_l\mathbf{e}_k, \quad (9)$$

up to an optimized scaling factor $c_{l,k}$. However, if the channels have non-zero mean, such as in line-of-sight (LoS) scenarios, (8) may provide significantly higher rates than (9). To see this, suppose that all channels are dominated by their LoS component, i.e., let $\mathbf{H}_l \approx \bar{\mathbf{H}}_l$ for some deterministic matrix $\bar{\mathbf{H}}_l$, $\forall l \in \mathcal{L}$. Then, since $\bar{\mathbf{H}}_l$ is statistical information known to all TXs, the optimal TMMSE precoders should take a form similar to centralized MMSE precoding.

B. Unidirectional CSIT sharing

We now let the local CSIT be shared unidirectionally along the serial fronthaul. Specifically, this can be modeled by the limited CSIT sharing pattern $\mathbf{\Omega}_{l,j} = \mathbf{I}_N$ if $j \leq l$, and $\mathbf{\Omega}_{l,j} = \mathbf{0}_{N\times N}$ otherwise, $\forall l \in \mathcal{L}$. This particular information structure can be interpreted as the CSIT which is accumulated at every TX during the first phase of a centralized MMSE precoding scheme for radio stripes, where the CPU collects the $K \times LN$ channel matrix $\mathbf{H}$ through the serial fronthaul.

Theorem 4. *The TMMSE precoders under unidirectional CSIT sharing are given by*

$$f^{\star}_{l,k}(\hat{\mathbf{H}}_l) = \mathbf{F}_l \mathbf{S}_l \left[\prod_{i=1}^{l-1} \bar{\mathbf{S}}_i\right] \mathbf{e}_k, \quad \forall l \in \mathcal{L}, \tag{10}$$

where we use the following short-hands:

- $\mathbf{S}_l := (\mathbf{I} - \mathbf{\Pi}_l \mathbf{P}_l)^{-1} (\mathbf{I} - \mathbf{\Pi}_l)$;
- $\bar{\mathbf{S}}_l := \mathbf{I} - \mathbf{P}_l \mathbf{S}_l$;
- $\mathbf{P}_l := \mathbf{H}_l \mathbf{F}_l$;
- $\mathbf{\Pi}_l := \mathbb{E}[\mathbf{P}_{l+1}\mathbf{S}_{l+1}] + \mathbf{\Pi}_{l+1}\mathbb{E}[\bar{\mathbf{S}}_{l+1}]$, $\mathbf{\Pi}_L := \mathbf{0}$.

Proof. We assume that all the matrix inverses involved exist. Substituting (10) into (7), we need to show that ($\forall l \in \mathcal{L}$)

$$\mathbf{H}_l^{\mathsf{H}} \left(\mathbf{S}_l \prod_{i=1}^{l-1} \bar{\mathbf{S}}_i + \sum_{j \neq l} \mathbb{E}\left[\mathbf{P}_j \mathbf{S}_j \prod_{i=1}^{j-1} \bar{\mathbf{S}}_i \middle| \hat{\mathbf{H}}_l \right] - \mathbf{I} \right) \mathbf{e}_k = \mathbf{0} \text{ a.s.}$$

To verify the above statement, we rewrite the first two terms inside the outer brackets as:

$$\left(\mathbf{S}_l + \sum_{j>l} \mathbb{E}\left[\mathbf{P}_j \mathbf{S}_j \prod_{i=l+1}^{j-1} \bar{\mathbf{S}}_i \right] \bar{\mathbf{S}}_l \right) \prod_{i=1}^{l-1} \bar{\mathbf{S}}_i + \sum_{j<l} \mathbf{P}_j \mathbf{S}_j \prod_{i=1}^{j-1} \bar{\mathbf{S}}_i, \tag{11}$$

where we use the fact that $\mathbf{P}_j$, $\mathbf{S}_j$, and $\bar{\mathbf{S}}_j$ are deterministic functions of $\mathbf{H}_j$ only, hence they are independent from $\hat{\mathbf{H}}_l$ for $j > l$, while they are deterministic functions of $\hat{\mathbf{H}}_l$ otherwise. Furthermore, since $\mathbf{P}_j$, $\mathbf{S}_j$, and $\bar{\mathbf{S}}_j$ are independent from $\mathbf{P}_i$, $\mathbf{S}_i$, and $\bar{\mathbf{S}}_i$ $\forall i \neq j$, we have

$$\begin{aligned}
\sum_{j>l} \mathbb{E}\left[\mathbf{P}_j \mathbf{S}_j \prod_{i=l+1}^{j-1} \bar{\mathbf{S}}_i \right] &= \sum_{j>l} \mathbb{E}\left[\mathbf{P}_j \mathbf{S}_j\right] \prod_{i=l+1}^{j-1} \mathbb{E}\left[\bar{\mathbf{S}}_i\right] \\
&= \mathbb{E}\left[\mathbf{P}_{l+1}\mathbf{S}_{l+1}\right] + \sum_{j>l+1} \mathbb{E}\left[\mathbf{P}_j \mathbf{S}_j\right] \prod_{i=l+1}^{j-1} \mathbb{E}\left[\bar{\mathbf{S}}_i\right] \\
&= \mathbb{E}\left[\mathbf{P}_{l+1}\mathbf{S}_{l+1}\right] + \left(\sum_{j>l+1} \mathbb{E}\left[\mathbf{P}_j \mathbf{S}_j\right] \prod_{i=l+2}^{j-1} \mathbb{E}\left[\bar{\mathbf{S}}_i\right] \right) \mathbb{E}\left[\bar{\mathbf{S}}_{l+1}\right].
\end{aligned}$$

The second and last term of the above chain of equalities define a recursion terminating with $\mathbb{E}\left[\mathbf{P}_L \mathbf{S}_L\right] + \mathbf{0}\mathbb{E}\left[\bar{\mathbf{S}}_L\right] = \mathbf{\Pi}_{L-1}$. This recursion gives precisely $\mathbf{\Pi}_l = \sum_{j>l} \mathbb{E}\left[\mathbf{P}_j \mathbf{S}_j \prod_{i=1}^{j-1} \bar{\mathbf{S}}_i\right]$. Together with the property $\mathbf{S}_l + \mathbf{\Pi}_l \bar{\mathbf{S}}_l = \mathbf{I}$, (11) simplifies to

$$\begin{aligned}
\prod_{i=1}^{l-1} \bar{\mathbf{S}}_i + \sum_{j<l} \mathbf{P}_j \mathbf{S}_j \prod_{i=1}^{j-1} \bar{\mathbf{S}}_i &= \left(\bar{\mathbf{S}}_{l-1} + \mathbf{P}_{l-1}\mathbf{S}_{l-1}\right) \prod_{i=1}^{l-2} \bar{\mathbf{S}}_i \\
&+ \sum_{j<l-1} \mathbf{P}_j \mathbf{S}_j \prod_{i=1}^{j-1} \bar{\mathbf{S}}_i = \prod_{i=1}^{l-2} \bar{\mathbf{S}}_i + \sum_{j<l-1} \mathbf{P}_j \mathbf{S}_j \prod_{i=1}^{j-1} \bar{\mathbf{S}}_i,
\end{aligned}$$

where the last equation follows from the definition of $\bar{\mathbf{S}}_l$ and where we identify another recursive structure among the remaining terms. By continuing until termination, we finally obtain $\prod_{i=1}^{l-1} \bar{\mathbf{S}}_i + \sum_{j<l} \mathbf{P}_j \mathbf{S}_j \prod_{i=1}^{j-1} \bar{\mathbf{S}}_i = \mathbf{I}$, which proves the main statement under the assumption that all the matrix inverses involved exist. This assumption is always satisfied. Due to space limitations, the details are given in [14]. □

Corollary 1. *An alternative expression for centralized MMSE precoding is given by* (10) *with* $\mathbf{\Pi}_l$ *replaced by*

$$\bar{\mathbf{P}}_l := \mathbf{P}_{l+1}\mathbf{S}_{l+1} + \bar{\mathbf{P}}_{l+1}\bar{\mathbf{S}}_{l+1}, \quad \bar{\mathbf{P}}_L := \mathbf{0}. \tag{12}$$

Proof. Centralized MMSE precoding is equivalent to TMMSE precoding for $\hat{\mathbf{H}}_l = \mathbf{H}$ $\forall l \in \mathcal{L}$. Since all random quantities become deterministic after conditioning on $\hat{\mathbf{H}}_l$, the proof of Theorem 4 can be repeated by removing $\mathbb{E}[\cdot]$ everywhere. □

By locally computing precoders based on $\hat{\mathbf{H}}_l$ only, and at the expense of some performance loss, (10) eliminates the additional overhead required by centralized MMSE precoding to share back the $LN \times K$ precoding matrix from the CPU.

Remark 3. *The scheme in* (10) *can be alternatively implemented via a recursive algorithm involving a* $K \times K$ *aggregate information matrix which is sequentially processed and forwarded in the direction from TX 1 to TX L, thus further reducing the overhead. This idea was already explored in [13]. Furthermore, Corollary 1 provides a novel distributed implementation of centralized MMSE precoding. The main difference is that the computation of* $\bar{\mathbf{P}}_l$ *entails an additional sequential procedure in the reverse direction, thus doubling the overhead. In contrast,* $\mathbf{\Pi}_l$ *can be computed offline.*

V. Performance evaluation

A. Comparison among CSIT sharing patterns

We simulate a network with a radio stripe of $L = 30$ equally spaced TXs with $N = 2$ antennas each wrapped around a circular area of radius $r_1 = 60$ m, and $K = 7$ RXs independently and uniformly drawn within a concentric circular area of radius $r_2 = 50$ m. We let the channel coefficient $h_{l,k,n}$ between the n-th antenna of TX l and RX k be independently distributed as $h_{l,k,n} \sim \mathcal{CN}(0, \rho^2_{l,k})$ approximated using a finite precision random numbers generator, hence inducing a bounded fading distribution, where $\rho^2_{l,k}$ denotes the channel gain between TX l and RX k. We follow the 3GPP NLoS Urban Microcell path-loss model [17, Table B.1.2.1-1]

$$\mathrm{PL}_{l,k} = 36.7 \log_{10}\left(\frac{d_{l,k}}{1 \text{ m}}\right) + 22.7 + 26 \log_{10}\left(\frac{f_c}{1 \text{ GHz}}\right) \text{ [dB]}, \tag{13}$$

where $f_c = 2$ GHz is the carrier frequency, and $d_{l,k}$ is the distance between TX l and RX k including a difference in height of 10 m. We let the noise power at all RXs be given by $P_{\text{noise}} = -174 + 10\log_{10}(B) + F$ dBm, where $B = 20$ MHz is the system bandwidth, and $F = 7$ dB is the noise figure. Finally, we let $\rho^2_{l,k} := 10^{-\frac{\mathrm{PL}_{l,k} + P_{\text{noise}}}{10}}$ mW^{-1}, and assume a relatively low total radiated power $P_{\text{sum}} = KP = K \times 1$ mW.

Figure 2a reports the empirical cumulative distribution function (CDF) of the Pareto optimal achievable rates $R_k = -\log(\mathrm{MSE}_k(f^{\star}_k))$, where $f^{\star}_k$ denotes the optimal solution of Problem (5) for: (i) no CSIT sharing (local TMMSE), (ii) unidirectional CSIT sharing (unidirectional TMMSE), and (iii) full CSIT sharing (centralized MMSE). As expected, adding stricter information constraints leads to performance degradation. However, the degradation is less pronounced from centralized to unidirectional MMSE precoding. Hence, the unidirectional MMSE scheme appears as a promising intermediate solution for supporting network-wide interference

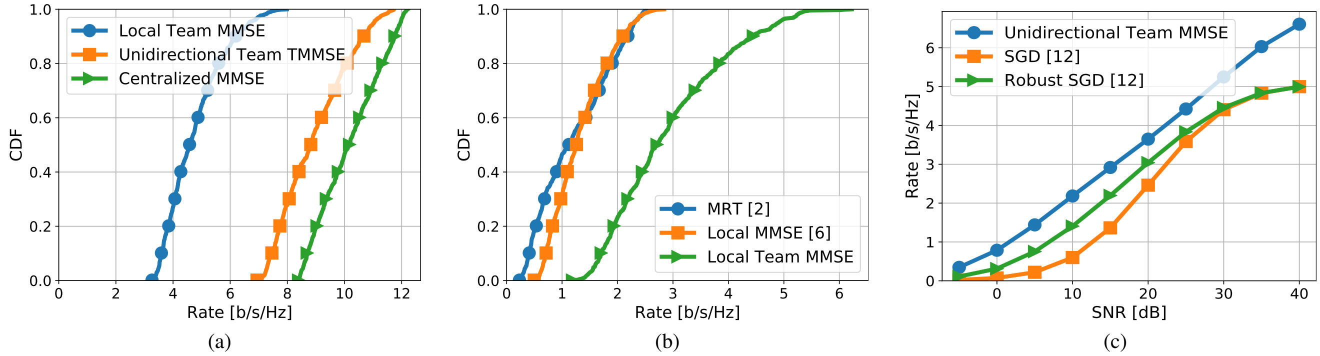

Fig. 2. Simulation results: (a) CDF of the optimal achievable rates for different CSIT sharing patterns; (b) CDF of the achievable rates for different local precoding schemes in a relatively weak LoS setup ($\kappa = 1$); (c) Rate vs the SNR of RX 1 for different precoding schemes using unidirectional CSIT sharing. In contrast to previously known heuristics, the team theoretical approach optimally exploits the available information, and hence exhibits superior performance.

management in a wider range of network setups, for instance, in those cases where significant RXs mobility could make the cost of timely CSIT sharing too high for centralized precoding.

B. Comparison among local precoding schemes

We compare the local TMMSE solution against classical MRT and local MMSE precoding [2], [6]. Since the bound in (6) may be overly pessimistic with suboptimal schemes, for a fair comparison we compute the DL rates $R_k = R_k^{\mathrm{hard}}$ by means of their dual UL rates R_k^{UatF} as defined in [8] using the same dual UL power allocation $w_k = 1 \ \forall k \in \mathcal{K}$. One of the major weaknesses of the above schemes is that they do not exploit channel mean information. To study this effect, we modify our setup by letting $N = 1$ and by considering a Ricean fading model $h_{l,k,1} \sim \mathcal{CN}\left(\sqrt{\frac{\kappa}{\kappa+1}\rho_{l,k}^2}, \frac{1}{\kappa+1}\rho_{l,k}^2\right)$ with $\kappa = 1$. Figure 2b confirms the above observation: while local MMSE precoding is shown in Sect. IV-A to be close to optimal for a NLoS setup ($\kappa = 0$), it incurs significant performance loss even in case of relatively weak LoS components ($\kappa = 1$).

C. Comparison with the SGD scheme [13]

We finally compare the unidirectional TMMSE solution against the suboptimal SGD scheme given by [13] using the suggested parameter choice $\mu = 1$, and against its robust version with μ optimized using line search. Despite being restricted to $N = 1$, this scheme exploits unidirectional CSIT sharing and possesses a similar sequential implementation as unidirectional TMMSE precoding. Figure 2c plots the rate $R_1 = R_1^{\mathrm{hard}}$ of the first RX (measured via its dual UL rate as in the previous section) versus the $\mathrm{SNR} := P\sum_l \rho_{l,1}^2$ for a single realization of the simulation setup. Although optimizing μ improves robustness against low SNR as observed in [13], this procedure seems not sufficient to recover the loss w.r.t. optimal TMMSE precoding in the considered scenario (in contrast, we report that this seems sufficient for the rather unrealistic equal path-loss case $\rho_{l,1}^2 = 1 \ \forall l$ studied in [13]).

VI. Conclusion

Our findings demonstrate the gains and feasibility of optimal precoding design in cell-free networks with limited CSIT sharing. Extensions may cover limited message sharing, per-TX power constraints, and additional examples of application.

References

[1] D. Gesbert, S. Hanly, H. Huang, S. Shamai Shitz, O. Simeone, and W. Yu, "Multi-cell MIMO cooperative networks: A new look at interference," *IEEE J. Sel. Areas Commun.*, vol. 28, no. 9, pp. 1380–1408, Sept. 2010.

[2] H. Q. Ngo, A. Ashikhmin, H. Yang, E. G. Larsson, and T. L. Marzetta, "Cell-free massive MIMO versus small cells," *IEEE Trans. Wireless Commun.*, vol. 16, no. 3, pp. 1834–1850, Mar. 2017.

[3] E. Nayebi, A. Ashikhmin, T. L. Marzetta, H. Yang, and B. D. Rao, "Precoding and power optimization in cell-free massive MIMO systems," *IEEE Trans. Wireless Commun.*, vol. 16, no. 7, pp. 4445–4459, July 2017.

[4] L. Du, L. Li, H. Q. Ngo, T. C. Mai, and M. Matthaiou, "Cell-free massive MIMO: Joint maximum-ratio and zero-forcing precoder with power control," *IEEE Trans. Commun.*, vol. 69, no. 6, pp. 3741–3756, June 2021.

[5] S. Buzzi, C. D'Andrea, A. Zappone, and C. D'Elia, "User-centric 5G cellular networks: Resource allocation and comparison with the cell-free massive MIMO approach," *IEEE Trans. Wireless Commun.*, vol. 19, no. 2, pp. 1250–1264, Feb. 2020.

[6] E. Björnson and L. Sanguinetti, "Scalable cell-free massive MIMO systems," *IEEE Trans. Commun.*, vol. 68, no. 7, pp. 4247–4261, July 2020.

[7] M. Attarifar, A. Abbasfar, and A. Lozano, "Subset MMSE receivers for cell-free networks," *IEEE Trans. Wireless Commun.*, vol. 19, no. 6, pp. 4183–4194, June 2020.

[8] E. Björnson, J. Hoydis, and L. Sanguinetti, "Massive MIMO networks: Spectral, energy, and hardware efficiency," *Foundations and Trends® in Signal Processing*, vol. 11, no. 3-4, pp. 154–655, 2017.

[9] R. Radner, "Team decision problems," *The Annals of Mathematical Statistics*, vol. 33, no. 3, pp. 857–881, 1962.

[10] S. Yüksel and T. Başar, *Stochastic networked control systems: Stabilization and optimization under information constraints*, Springer Science & Business Media, 2013.

[11] G. Interdonato, E. Björnson, H. Q. Ngo, P. Frenger, and E. G. Larsson, "Ubiquitous cell-free massive MIMO communications," *EURASIP J. Wireless Commun. and Netw.*, vol. 2019, no. 1, pp. 197, 2019.

[12] Z. H. Shaik, E. Björnson, and E. G. Larsson, "MMSE-optimal sequential processing for cell-free massive MIMO with radio stripes," *arXiv preprint arXiv:2012.13928*, 2020.

[13] J. Rodríguez Sánchez, F. Rusek, O. Edfors, M. Sarajlić, and L. Liu, "Decentralized massive MIMO processing exploring daisy-chain architecture and recursive algorithms," *IEEE Trans. Signal Process.*, vol. 68, no. Jan., pp. 687–700, 2020.

[14] L. Miretti, E. Björnson, and D. Gesbert, "Team MMSE precoding with applications to cell-free massive MIMO," *arXiv preprint arXiv:2104.15027v1*, 2021.

[15] G. Caire, "On the ergodic rate lower bounds with applications to massive MIMO," *IEEE Trans. Wireless Commun.*, vol. 17, no. 5, pp. 3258–3268, May 2018.

[16] E. Björnson and E. Jorswieck, *Optimal resource allocation in coordinated multi-cell systems*, Now Publishers Inc, 2013.

[17] 3GPP, "Further advancements for E-UTRA physical layer aspects (release 9)," 3GPP TS 36.814, Mar. 2017.