Team MMSE Precoding With Applications to Cell-Free Massive MIMO

Lorenzo Miretti, *Member, IEEE*, Emil Björnson, *Fellow, IEEE*, and David Gesbert, *Fellow, IEEE*

Abstract—This article studies a novel distributed precoding design, coined *team minimum mean-square error* (TMMSE) precoding, which rigorously generalizes classical centralized MMSE precoding to distributed operations based on transmitter-specific channel state information (CSIT). Building on the so-called *theory of teams*, we derive a set of necessary and sufficient conditions for optimal TMMSE precoding, in the form of an infinite dimensional linear system of equations. These optimality conditions are further specialized to cell-free massive MIMO networks, and explicitly solved for two important examples, i.e., the classical case of local CSIT and the case of unidirectional CSIT sharing along a serial fronthaul. The latter case is relevant, e.g., for the recently proposed *radio stripe* concept and the related advances on sequential processing exploiting serial connections. 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 team theoretical approach in terms of ergodic achievable rates under a sum-power constraint.

Index Terms—Cell-free massive MIMO, radio stripes, distributed processing, precoding, team minimum mean-square error (TMMSE), team theory.

I. INTRODUCTION

INTER-CELL interference is a major limiting factor of wireless communication systems capitalizing on aggressive spectrum reuse and network densification to increase capacity. To mitigate this effect, future generation systems are expected to implement advanced cooperative communication techniques, in particular by letting geographically distributed base stations jointly serve their users. However, the practical deployment of cooperative wireless networks is currently prevented by the severe scalability issue arising from network-wide processing [2]. Specifically, the excessive amount of

data and channel state information (CSI) to be timely shared for implementing fully cooperative regimes such as in the original *network MIMO* or *cloud radio access network* (C-RAN) concepts [3]–[5] often becomes the main bottleneck when practical fronthaul capacity constraints are introduced. Studying more realistic cooperation regimes entailing limited data and CSI sharing is hence of fundamental importance for making network cooperation an attractive technology for next generation systems [2], [6].

A. Cooperative Transmission With Distributed CSIT

In this work we explore a downlink (DL) cooperation regime with full data sharing and general *distributed* CSI at the TXs (CSIT) [7]–[9], that is, we let each TX operate on the basis of possibly different estimates of the *global* channel state obtained through some arbitrary CSIT acquisition and sharing mechanism. This assumption is relevant, e.g., for all service situations where, compared to data sharing, CSIT sharing needs to be performed within much tighter time constraints, and hence may dominate the fronthaul overhead. For instance, it is suitable in case of rapidly varying channels due to user mobility, where full CSIT sharing may result in outdated information or occupy an excessive portion of the coherence time, or when delay-tolerant data is proactively made available at the TXs using caching techniques (see [9] and reference therein for a detailed discussion). As an extreme example, a cooperation regime with full data sharing and no CSIT sharing (a configuration here referred to as *local* CSIT) is perhaps the leading motivation behind the early development of the now popular *cell-free massive MIMO* paradigm [10]. This paradigm combines the benefits of ultra-dense networks with simple yet effective TX cooperation schemes, and emerged as a promising evolution of the network MIMO and C-RAN concepts. The distributed CSIT assumption also covers extensions of [10] to more complex setups ranging from partial to full CSIT sharing (see, e.g., [11], [12]). Clearly, these cooperation regimes are still far from being scalable, since they all assume network-wide data sharing. Splitting the network into clusters of cooperating TXs [12]–[14], possibly dynamically and with a user-centric approach [15]–[18], and applying similar transmission techniques assuming full data sharing within each cluster, emerged as a viable solution for implementing scalable cooperation regimes in practical systems. However, due to space limitations and to better focus on limited CSIT sharing, in this work, we do not consider network clusterization. In particular, we do not cover complementary service situations where CSIT can be more easily shared than

data, hence entering the realm of interference *coordination* or *alignment* [2], [19]. Nevertheless, if combined with network clustering techniques, the results presented in this study can be seen as a first step towards a more general theory jointly covering limited data and CSIT sharing.

B. Summary of Contributions and Related Works

Although the importance of cooperative transmission schemes based on limited CSIT sharing has been acknowledged in the literature, a satisfactory understanding of systems with distributed CSIT is still missing. Most of the available information theoretical results rely on asymptotic signal-to-noise ratio (SNR) tools [7], [9], or focus on simple settings with a single receiver (RX) [20]. However, [7], [9], [20] do not lead to practical schemes for complex settings such as cell-free massive MIMO networks. On the other hand, the available practical schemes are essentially based on heuristic adaptations of known centralized precoding designs such as maximum-ratio transmission (MRT), zero-forcing (ZF), or minimum mean-square error (MMSE) precoding [10], [11], [18]. Hence, there is a need to develop a mathematical framework that allows the sound derivation of schemes that cater for distributed CSIT setups.

This work provides considerable progress in this direction. By using simplified yet practical point-to-point information theoretical tools, namely by using standard linearly precoded Gaussian codes, treating interference as noise, and the non-coherent ergodic rate bounds popularized by the massive MIMO literature [21], [22], we propose a novel distributed precoding design, coined *team* MMSE (TMMSE) precoding, generalizing classical centralized MMSE precoding [22] to systems with distributed CSIT. Its optimality in terms of achievable ergodic rates under a sum-power constraint is formally established by revisiting the uplink-downlink (UL-DL) duality principle [22] in light of the distributed CSIT assumption. Our first main result[1] is showing that the problem of optimal TMMSE precoding design can be solved by means of a useful set of optimality conditions in the form of an infinite dimensional linear system of equations, for which many standard solution tools exist. The key novelty lies in the introduction of previously unexplored elements from the *theory of teams*, a mathematical framework for multi-agent coordinated decision making in presence of asymmetry of information. This framework was pioneered in theoretical economics by Marschak and Radner [23], [24], and then further developed in the control theoretical literature (see the excellent survey in [25]). Early applications of team theory to wireless communication, including the problem of distributed precoding design, are reported in [2], [8]. However, compared to previous attempts for distributed precoding design, this work is the first exploiting (and partially extending) known results for the class of *quadratic teams* [24], [25], which is one of the few cases where solid globally optimal solution approaches are available.

[1]The preliminary version of this work [1] focuses on a simplified cell-free massive MIMO setup. This work extends [1] to more general networks, Gaussian fading, and channel estimation errors; provides complete theoretical derivations; improves the comparison with the previous literature.

In the second part of this work, the aforementioned optimality conditions are specialized to cell-free massive MIMO networks. To the best of the authors' knowledge, this is the first work connecting cell-free massive MIMO to the theory of teams. The first non-trivial application is the derivation of the optimal TMMSE precoders based on local CSIT only, improving upon previous local precoding strategies studied, e.g., in [10], [18], [26]. We then consider a cell-free massive MIMO network with serial fronthaul, an efficient architecture also known as a *radio stripe* [17], [27]. We derive optimal TMMSE precoders by assuming that CSIT is shared unidirectionally along the stripe. The proposed scheme can be efficiently implemented in a sequential fashion, an idea that has been explored already in [17], [27], [28] for UL processing, and in [29] under a different cellular context. As a byproduct, we also obtain a novel distributed implementation of classical centralized MMSE precoding tailored to radio stripes. Finally, we present extensive numerical results comparing the effects of different CSIT sharing patterns in a radio stripe system and evaluating the suboptimality of the competing schemes. Interestingly, our numerical results suggest that unidirectional information sharing is a promising candidate for enlarging the domain of applications of radio stripes beyond the regimes supported by centralized or local precoding; for instance, it may allow effective interference management for a wider range of mobility patterns. Moreover, we show that the known local MMSE precoding scheme studied, e.g., in [18], [28], is optimal in a non line-of-sight (NLoS) scenario, while it may be significantly outperformed by the TMMSE solution for local CSIT in the presence of line-of-sight (LoS) components.

C. Outline and Notation

The present study is structured as follows: Section II presents the system model and other necessary preliminaries. The main results on team MMSE precoding and their specialization to cell-free massive MIMO networks are given respectively in Section III and Section IV. The numerical results are given in Section V (*Reproducible Research*: simulation code available at https://github.com/emilbjornson/team-MMSE).

Hereafter, we use the following notation. We reserve italic letters (e.g., 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. Random quantities are distinguished from their realizations as follows: $\mathbb{a}$, $\mathbb{A}$ denote random vectors and matrices; A denotes a random scalar, or a generic random variable taking values in some unspecified set $\mathcal{A}$. We use := for definitions, and $\preceq$, $\succeq$ ($\prec$, $\succ$) for (strict) generalized inequalities w.r.t. the cone of nonnegative Hermitian matrices. The operators $(\cdot)^{\mathsf{T}}$, $(\cdot)^{\mathsf{H}}$ denote respectively the transpose and Hermitian transpose of matrices and vectors, and $\Re(\cdot)$ is the real part. We denote the Euclidean norm by $\|\cdot\|$, the Frobenius norm by $\|\cdot\|_{\mathrm{F}}$, and the trace operator by $\mathrm{tr}(\cdot)$. The (conditional) expectation of $\mathbb{A}$ (given $\mathbb{B}$) is denoted by $\mathsf{E}[\mathbb{A}]$ ($\mathsf{E}[\mathbb{A}|\mathbb{B}]$), and $\mathsf{Var}[A]$ is the variance of A. Given $n > 2$ random matrices $\mathbb{A}_1, \ldots, \mathbb{A}_n$ with joint distribution $p(\mathbf{A}_1, \ldots, \mathbf{A}_n)$, we say that $\mathbb{A}_1 \rightarrow \mathbb{A}_2 \rightarrow \ldots \rightarrow \mathbb{A}_n$ forms a Markov chain if $p(\mathbf{A}_i|\mathbf{A}_{i-1}, \ldots, \mathbf{A}_1) = p(\mathbf{A}_i|\mathbf{A}_{i-1})$

$\forall i \geq 2$. We use $\mathrm{diag}(\mathbf{A}_1, \ldots, \mathbf{A}_n)$ to denote a block-diagonal matrix with the matrices $\mathbf{A}_1, \ldots, \mathbf{A}_n$ on its diagonal, and $\mathrm{vec}(\mathbf{A})$ to denote a vector obtained by stacking column-wise the elements of $\mathbf{A}$. We denote by $\mathbf{e}_n$ the n-th column of the identity matrix $\mathbf{I}$, with dimension extrapolated from the context. We use $\prod_{i=l'}^{l} \mathbf{A}_i := \mathbf{A}_l \mathbf{A}_{l-1} \ldots \mathbf{A}_{l'}$ for integers $l \geq l' \geq 1$ to denote the *left* product chain of $l - l' + 1$ ordered matrices of compatible dimension, and we adopt the convention $\prod_{i=l'}^{l} \mathbf{A}_i = \mathbf{I}$ for $l < l'$. Finally, $h(A)$ $(h(A|B))$ denotes the (conditional) entropy, $I(A;B)$ is the mutual information, and all logarithms are expressed in base 2 unless differently specified.

II. SYSTEM MODEL AND PRELIMINARIES

A. Channel Model

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

$$\mathbb{y} = \sum_{l=1}^{L} \mathbb{H}_l \mathbb{x}_l + \mathbb{n}$$

where the k-th element Y_k of $\mathbb{y} \in \mathbb{C}^K$ is the received signal at RX k, $\mathbb{H}_l \in \mathbb{C}^{K \times N}$ is a sample of a stationary ergodic random process modelling the fading between TX l and all RXs, $\mathbb{x}_l \in \mathbb{C}^N$ is the transmitted signal at TX l, and $\mathbb{n} \sim \mathcal{CN}(\mathbf{0}, \mathbf{I})$ is a sample of a white noise process. This channel model is relevant, e.g, for narrowband or wideband OFDM systems [30] where transmission spans several realizations of the fading process. For most parts of this work, we do not specify the distribution of $\mathbb{H} := [\mathbb{H}_1, \ldots, \mathbb{H}_L]$. However, we reasonably assume the channel submatrices corresponding to different TX-RX pairs to be mutually independent, and finite fading power $\mathsf{E}[\|\mathbb{H}\|_{\mathrm{F}}^2] < \infty$. Furthermore, we focus on $N < K$, that is, on the regime where cooperation is crucial for interference management [2].

B. Distributed Linear Precoding

Consider a *distributed* CSIT configuration [8], i.e., where each TX has some potentially different side information S_l about the *global* channel matrix $\mathbb{H}$. For instance, this could model frequency division duplex (FDD) systems where each S_l is composed by different feedback signals from the RXs, or time-division duplex (TDD) systems where over-the-uplink *local* estimates $\hat{\mathbb{H}}_l$ of the *local* channel $\mathbb{H}_l$ are not perfectly shared across the network. Importantly, this is more general than the typical assumption in the cell-free massive MIMO literature covering distributed operations, which limits S_l to $\hat{\mathbb{H}}_l$ only [10], [18]. We assume $(\mathbb{H}, S_1, \ldots, S_L)$ to be a sample of an ergodic stationary process with first order joint distribution fixed by nature/design, and known by all TXs.

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

$$\mathbb{x}_l = \sum_{k=1}^{K} \mathbb{t}_{l,k} U_k, \quad \mathbb{t}_{l,k} = \mathbb{t}_{l,k}(S_l), \tag{1}$$

where $U_k \sim \mathcal{CN}(0, p_k)$ is the independently encoded message for RX k, shared by all TXs, and where $\mathbb{t}_{l,k} \in \mathbb{C}^N$ is a linear precoder applied at TX l to message U_k based only on the side information S_l. More formally, by letting $(\Omega, \Sigma, \mathsf{P})$ be the underlying probability space over which all random quantities are defined, we constrain $\mathbb{t}_{l,k}$ within the vector space $\mathcal{T}_l$ of square-integrable Σ_l-measurable functions $\Omega \to \mathbb{C}^N$, where $\Sigma_l \subseteq \Sigma$ denotes the sub-σ-algebra generated by S_l on Ω, called the *information subfield* of TX l [24], [25]. This assumption[2] rigorously describes the functional dependency of $\mathbb{t}_{l,k}$ on the portion S_l of the overall system randomness, and includes a reasonable finiteness constraint $\mathsf{E}[\|\mathbb{t}_{l,k}\|^2] < \infty$ on precoders power. We finally denote the full precoding vector for message U_k by $\mathbb{t}_k^{\mathsf{T}} := [\mathbb{t}_{1,k}^{\mathsf{T}} \ldots \mathbb{t}_{L,k}^{\mathsf{T}}]^{\mathsf{T}}$, and let $\mathbb{t}_k \in \mathcal{T} := \prod_{l=1}^{L} \mathcal{T}_l$.

C. Performance Metric

We measure the network performance under the specified transmission scheme by using Shannon (ergodic) rates $R_k^{\mathrm{DL}} := I(U_k; Y_k)$, which are achievable without channel state information at the RX (CSIR) by treating interference as noise (TIN) and by neglecting any memory across the realizations $(\mathbb{H}, S_1, \ldots, S_L)$ of the state and CSIT process [31], [32]. Because of the difficulties in evaluating the mutual information, we consider the following lower bound, known as the *hardening* bound [21], [22],

$$R_k^{\mathrm{hard}} := \log\left(1 + \frac{p_k |\mathsf{E}[\mathbb{g}_k^{\mathsf{H}} \mathbb{t}_k]|^2}{p_k \mathsf{Var}[\mathbb{g}_k^{\mathsf{H}} \mathbb{t}_k] + \sum_{j \neq k} p_j \mathsf{E}[|\mathbb{g}_k^{\mathsf{H}} \mathbb{t}_j|^2] + 1}\right)$$
$$\leq I(U_k; Y_k), \tag{2}$$

where $[\mathbb{g}_1 \ldots \mathbb{g}_K] := \mathbb{H}^{\mathsf{H}}$. An alternative classical performance metric would be given by the following upper bound

$$I(U_k; Y_k) \leq \mathsf{E}\left[\log\left(1 + \frac{p_k |\mathbb{g}_k^{\mathsf{H}} \mathbb{t}_k|^2}{\sum_{j \neq k} p_j |\mathbb{g}_k^{\mathsf{H}} \mathbb{t}_j|^2 + 1}\right)\right] =: R_k^{\mathrm{ub}}, \tag{3}$$

which is in fact achievable with perfect local CSIR $\{\mathbb{g}_k^{\mathsf{H}} \mathbb{t}_j\}_{j=1}^{K}$ by TIN and by taking into account channel memory [31], [32]. We consider R_k^{hard} instead of R_k^{ub} because of the less stringent CSIR requirements and, perhaps most importantly, for treatability reasons. Due to its name and historical use, it is sometimes believed that the ability of R_k^{hard} to produce good approximations of R_k^{DL} relies on the *channel hardening* effect arising in massive MIMO systems [22]. Although this is correct for some precoding design such as MRT [32], we remark that (2) may perform well also in absence of channel hardening. For instance, R_k^{hard} and R_k^{up} coincide under a ZF scheme with perfect CSIT putting $\mathbb{g}_k^{\mathsf{H}} \mathbb{t}_j = 0$ for $j \neq k$ and $\mathbb{g}_k^{\mathsf{H}} \mathbb{t}_k = 1$, for any long-term power allocation policy $\{p_k\}_{k=1}^{K}$ and feasible antenna regime.

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 $\{\mathbb{t}_k\}_{k=1}^{K}$ and power allocation

---

[2]The measure theoretical formulation presented above is necessary for establishing Theorem 3. However, the rest of this study does not require any particular measure theoretical background.

policy $\{p_k\}_{k=1}^K$ satisfying $\sum_{l=1}^L \mathsf{E}[\|\mathbb{x}_l\|^2] \leq P_{\text{sum}} < \infty$. The set $\mathcal{R}^{\text{hard}}$ is an inner bound for the capacity region of the considered network with distributed CSIT and subject to a long-term sum power constraint P_{sum}. Due to its importance in system design and resource allocation, we consider the notion of (weak) Pareto optimality on $\mathcal{R}^{\text{hard}}$ and we mostly focus on the (weak) Pareto boundary of $\mathcal{R}^{\text{hard}}$, denoted by $\partial\mathcal{R}^{\text{hard}}$ [15]. Note that, by the Cauchy–Schwarz inequality and the mild assumptions given in the previous sections, we have $|\mathsf{E}[\mathbb{g}_k^{\mathsf{H}}\mathbb{t}_k]|^2 < \infty \ \forall \mathbb{t}_k \in \mathcal{T}, \ \forall k \in \mathcal{K}$, hence $\partial\mathcal{R}^{\text{hard}}$ is finite.

The long-term 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, treated in Section IV-B, where all the TXs share the same power supply [17]. However, note that many simple heuristic methods (such as power scaling factors) can be applied to adapt systems designed under a long-term sum power constraint to more restrictive cases such as per-TX power constraints. Further analyses on different power constraints are left for future work.

III. TEAM MMSE PRECODING

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

$$\underset{\mathbb{t}_k \in \mathcal{T}}{\text{minimize}} \ \text{MSE}_k(\mathbb{t}_k) := \mathsf{E}\left[\left\|\mathbf{W}^{\frac{1}{2}}\mathbb{H}\mathbb{t}_k - \mathbf{e}_k\right\|^2 + \frac{\|\mathbb{t}_k\|^2}{P}\right], \tag{4}$$

where $\mathbf{W} := \text{diag}(w_1, \ldots, w_K)$, $\mathbf{e}_k$ is the k-th column of $\mathbf{I}_K$, and $P := P_{\text{sum}}/K$. A solution to the above problem can be recognized as a distributed version of the classical centralized MMSE precoding design [22]. For $P \to \infty$, it can be interpreted as the 'closest' distributed approximation of the ZF solution. By means of team theoretical arguments [24], [25], this section provides rigorous yet practical guidelines for optimally solving Problem (4). Before providing the main results of this section, we also revisit the effectiveness of the MSE criterion in terms of network performance, which is well-known for centralized precoding.

Remark 1: Hereafter, with the exception of Section III-A, we consider w.l.o.g. $\mathbf{W} = \mathbf{I}$. The general case will readily follow by replacing $\mathbb{H}_l$ with $\mathbf{W}^{\frac{1}{2}}\mathbb{H}_l$ everywhere.

A. Achievable Rates via Uplink-Downlink Duality

This section discusses the formal connection between the objective of Problem (4) and $\mathcal{R}^{\text{hard}}$ by revisiting UL-DL duality under a general distributed CSIT assumption.

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

$$R_k \leq \log(\text{MSE}_k(\mathbb{t}_k))^{-1} \tag{5}$$

belongs to $\mathcal{R}^{\text{hard}}$. The power allocation policy $\{p_k\}_{k=1}^K$ achieving the above inner bound is given in Appendix A,

and satisfies $\sum_l \mathsf{E}[\|\mathbb{x}_l\|^2] = P_{\text{sum}}$. Furthermore, if $\mathbb{t}_k$ solves Problem (4) $\forall k \in \mathcal{K}$, then $(R_1, \ldots, R_K)$ with $R_k = \log(\text{MSE}_k(\mathbb{t}_k))^{-1}$ is Pareto optimal, and every rate tuple in $\partial\mathcal{R}^{\text{hard}}$ is obtained for some $\mathbf{w} \in \mathcal{W}$.

Proof: The proof is based on connecting Problem (4) to the problem of ergodic rate maximization in a dual UL channel, where $\mathbf{w}$ is an UL power allocation vector, $\mathbb{t}_k$ is a distributed UL combiner, and where achievable rates are measured by using the so-called *use-and-then-forget* (UatF) bound [22, Theorem 4.4]. The details are given in Appendix A. $\square$

Theorem 1 states that the Pareto boundary of $\mathcal{R}^{\text{hard}}$ can be parametrized by $K-1$ nonnegative real parameters, i.e., by the weights $\mathbf{w} \in \mathcal{W}$. A similar parametrization was already known for deterministic channels (see, e.g., [15]), or, equivalently, for fading channels with perfect CSIT and CSIR. This work extends the aforementioned results to imperfect and possibly distributed CSIT, and no CSIR. In theory, $\mathbf{w}$ should be selected according to some network utility (e.g., the sum-rate or the max-min rate). In practice, $\mathbf{w}$ is often fixed heuristically (e.g., from the real UL powers), while the network utility is optimized a posteriori by varying the DL power allocation policy $\{p_k\}_{k=1}^K$.

From a precoding design point of view, Theorem 1 generalizes the duality-based argument behind classical MMSE precoding given by [22]. While [22] motivates the MMSE solution as the optimal combiner maximizing a dual UL ergodic rate bound based on coherent decoding, the proof of Theorem 1 directly relates the MSE criterion to the more conservative UatF bound. This last point is particularly relevant under distributed CSIT, where an optimal solution to the coherent ergodic rate maximization problem is not known in general. We recall that, in turn, [22] generalizes classical duality-based arguments for deterministic channels to fading channels. As a concluding remark, we stress that the inner bound (5) should not be confused with the well-known inner bound based on the notion of (weighted) MSE on the DL channel [33], where the precoders for *all* messages contribute to each rate bound.

B. Quadratic Teams for Distributed Precoding Design

Problem (4) belongs to the known family of *team decision* problems [24], [25], which are generally difficult to solve for general information constraints $\mathbb{t}_k \in \mathcal{T}$. However, by rewriting the objective as $\text{MSE}_k(\mathbb{t}_k) = \mathsf{E}[c_k(\mathbb{H}, \mathbb{t}_{1,k}, \ldots, \mathbb{t}_{L,k})]$,

$$c_k(\mathbf{H}, \mathbf{t}_{1,k}, \ldots, \mathbf{t}_{L,k}) := \mathbf{t}_k^{\mathsf{H}}\mathbb{Q}\mathbf{t}_k - 2\Re\left(\mathbb{g}_k^{\mathsf{H}}\mathbf{t}_k\right) + 1, \tag{6}$$

where $\mathbb{Q} := \mathbb{H}^{\mathsf{H}}\mathbb{H} + \frac{1}{P}\mathbf{I}$, $\mathbb{g}_k = \mathbb{H}^{\mathsf{H}}\mathbf{e}_k$, and by noticing that $\mathbb{Q} \succ \mathbf{0}$ a.s., we recognize that Problem (4) belongs to the class of *quadratic teams* as defined in [24, Sect. 4]. This class exhibits strong structural properties, in particular related to the following solution concept:

Definition 1 (Stationary Solution [25]): A solution $\mathbb{t}_k^{\star} \in \mathcal{T}$ is a stationary solution for Problem (4) if $\text{MSE}_k(\mathbb{t}_k^{\star}) < \infty$ and if the following set of equalities hold ($\forall l \in \mathcal{L}$)

$$\nabla_{\mathbf{t}_{l,k}} \mathsf{E}\left[c_k(\mathbb{H}, \mathbb{t}_{-l,k}^{\star}, \mathbf{t}_{l,k}) \big| S_l\right]\bigg|_{\mathbf{t}_{l,k}=\mathbb{t}_{l,k}^{\star}(S_l)} = \mathbf{0} \text{ a.s.,} \tag{7}$$

where $(\mathbb{t}_{-l,k}, \mathbf{t}_{l,k}) := (\mathbb{t}_{1,k}, \ldots, \mathbb{t}_{l-1,k}, \mathbf{t}_{l,k}, \mathbb{t}_{l+1,k}, \ldots, \mathbb{t}_{L,k})$. By evaluating the stationary conditions (7) using standard results on differentiation of real-valued quadratic forms over a complex domain, we obtain that a stationary solution may be given by any solution to the following feasibility problem:

$$\text{find } \mathbb{t}_k \in \mathcal{T}$$
$$\text{s.t. } (\forall l \in \mathcal{L})$$
$$\mathsf{E}[\mathbb{Q}_{l,l}|S_l]\mathbb{t}_{l,k}^\star(S_l) + \sum_{j \neq l} \mathsf{E}[\mathbb{Q}_{l,j}\mathbb{t}_{j,k}^\star|S_l] - \mathsf{E}[\mathbb{g}_{l,k}|S_l] = \mathbf{0} \text{ a.s., (8)}$$

where $\mathbb{Q}_{l,l} := \mathbb{H}_l^{\mathsf{H}}\mathbb{H}_l + \frac{1}{P}\mathbf{I}$, $\mathbb{Q}_{l,j} := \mathbb{H}_l^{\mathsf{H}}\mathbb{H}_j$ for $j \neq l$, and $\mathbb{g}_{l,k} := \mathbb{H}_l^{\mathsf{H}}\mathbf{e}_k$, provided that all expectations are finite. Since the considered quadratic cost c_k is convex and differentiable a.s. in each of the $\mathbf{t}_{l,k}$, under some mild technical assumptions the notion of stationarity can be interpreted as enforcing each function $\mathbb{t}_{l,k}^\star(S_l)$ to be optimal while keeping the functions $\mathbb{t}_{j,k}^\star(S_j)$ of all the other TXs $j \neq l$ fixed. This is reminiscent of the game theoretical notion of *Nash equilibrium*, with the difference that here all the TXs share the same objective, and hence they act as a team. Similarly to Nash equilibria, stationary solutions may be in general inefficient, i.e., lead to a local optimum. However, a stronger result holds for quadratic teams:

Theorem 2: If $\mathbb{Q}$ is uniformly bounded above, i.e., there exists a positive scalar $B < \infty$ such that $\mathbb{Q} \prec B\mathbf{I}$ a.s., then Problem (4) admits a unique optimal solution, which is also the unique stationary solution solving Problem (8).

Proof: Theorem 2.6.6 of [25]. □

Theorem 2 and Problem (8) are of fundamental theoretical importance since they concisely identify the two key ingredients for optimal distributed precoding design:

1) Robustness against local channel estimation errors, captured by $\mathsf{E}[\mathbb{Q}_{l,l}|S_l]$ and $\mathsf{E}[\mathbb{g}_{l,k}|S_l]$;
2) Robustness against the effect of the "decisions" taken at the other TXs, captured by $\sum_{j \neq l} \mathsf{E}[\mathbb{Q}_{l,j}\mathbb{t}_{j,k}^\star|S_l]$. This is the main new difficulty which is introduced while moving from centralized to distributed precoding design.

From a practical perspective, the above results also provide a very powerful tool to solve the difficult distributed precoding design problem. Specifically, they provide a set of *optimality conditions* for Problem (4) in the form of a standard infinite dimensional *linear* feasibility problem, for which many approximate solution methods are available. For instance, the optimal TMMSE precoders may be approached via one of the iterative methods surveyed in [25] based on interpreting the solution to (8) as the unique fixed point of a linear map. Other promising methods may also include finite dimensional approximations of (8) obtained, e.g., by sampling the CSI process $(\mathbb{H}, S_1, \ldots, S_L)$ and by interpreting the sampled version of (8) as a classical function interpolation problem from a finite set of linear measurements [34]. Further discussions on approximate solution methods are left for future work, and most parts of this study will focus on cell-free massive MIMO networks and in particular on special cases where (8) can be solved explicitly. However, we remark that the content of this section can be readily applied to study general networks with distributed CSIT as described in Section II, not necessarily restricted to cell-free massive MIMO networks.

Before moving to the aforementioned results, we focus on a rather technical yet important weakness of Theorem 2. The assumption of Theorem 2 is essentially used to ensure the existence of all the expectations in the steps of the proof, and is satisfied for any fading distribution with bounded support. However, it is not satisfied for the classical Gaussian fading model. Despite being unrealistic, since physically consistent fading distributions cannot have unbounded support, Gaussian fading is a very common model in the literature due to its analytical treatability, for example in deriving simple channel estimation error models [22], [30]. Furthermore, except for the tails of the distribution, it usually fits measurements well. Therefore, in the following we derive more general optimality results covering this case.

Theorem 3: If $\mathsf{E}[\|\mathbb{Q}\|_{\mathrm{F}}^2] < \infty$, then Problem (4) admits a unique optimal solution, which is also the unique stationary solution solving Problem (8).

Proof: The proof is given in Appendix B. □

Finally, we conclude this section by observing that the optimality conditions (8) are not only useful to characterize the optimal TMMSE solution, but also to evaluate the suboptimality of its approximations. This can be done via an appropriate measure of violation of the optimality conditions. Specifically, we have the following result:

Lemma 1: Suppose that $\mathsf{E}[\|\mathbb{Q}\|_{\mathrm{F}}^2] < \infty$ holds, and let $\mathbb{t}_k^\star \in \mathcal{T}$ be the unique solution to Problem (4). Furthermore, define $\mathbb{z}_k := [\mathbb{z}_{1,k}^{\mathsf{T}}, \ldots, \mathbb{z}_{L,k}^{\mathsf{T}}]^{\mathsf{T}}$, where $\mathbb{z}_{l,k} = \mathbb{z}_{l,k}(S_l)$ is given by the left-hand side of the stationary conditions in (8) with $\mathbb{t}_k^\star$ replaced by an arbitrary $\mathbb{t}_k \in \mathcal{T}$. If $\mathbb{z}_k \in \mathcal{T}$, i.e., if $\mathsf{E}[\|\mathbb{z}_{l,k}(S_l)\|^2] < \infty$, the following optimality bounds hold:

$$\mathrm{MSE}_k(\mathbb{t}_k) - \mathrm{MSE}_k(\mathbb{t}_k^\star) \leq \mathsf{E}\left[\|\mathbb{Q}^{-\frac{1}{2}}\mathbb{z}_k\|^2\right] \quad (9)$$
$$\leq P\mathsf{E}\left[\|\mathbb{z}_k\|^2\right], \quad (10)$$

Proof: The proof is given in Appendix C. □

Clearly, $\mathbb{z}_k(S_1, \ldots, S_L) = \mathbf{0}$ a.s. gives the optimality conditions in (8), and in fact it corresponds to a zero optimality gap in (10). Intuitively, the bounds in (10) can be quite tight if $\mathbb{z}_k(S_1, \ldots, S_L) \approx \mathbf{0}$ with high probability. However, if this is not satisfied, we remark that both bounds can be looser than other trivial bounds obtained, e.g., by assuming a centralized information structure, or even output negative estimates of $\mathrm{MSE}_k(\mathbb{t}_k^\star)$. As already mentioned, we leave further studies on suboptimal solutions for future work, and we use (10) only in Section IV-C for getting analytical insights into a particular setup.

IV. Applications to Cell-Free Massive MIMO

In this section we specialize the theory of Section III to cell-free massive MIMO networks, and explicitly derive optimal TMMSE precoders for two practical examples. In the scope of this study, the important feature of the cell-free massive MIMO paradigm is the exploitation of time division duplex operations and channel reciprocity to efficiently acquire estimates $\hat{\mathbb{H}}_l$ of the local channel $\mathbb{H}_l$ at each TX l via over-the-uplink training [10]. These estimates may be

subsequently shared through the fronthaul according to some predefined CSIT sharing mechanism, forming at each TX l a side information about the global channel $\mathbb{H}$ of the type $S_l := \left(\hat{\mathbb{H}}_l, \bar{S}_l \right)$, where $\bar{S}_l$ denotes the side information about the other channels $\{\mathbb{H}_j\}_{j \neq l}$ collected at TX l. Depending on the CSIT sharing mechanism, $\bar{S}_l$ may be a function of the other local channel estimates $\{\hat{\mathbb{H}}_j\}_{j \neq l}$ (e.g., in case of error-free digital signalling), or include additional noise (e.g., in case of random events such as protocol delays). Consistently with the above discussion, we consider the following assumptions:

Assumption 1 (Local Channel Estimation): For every $l \in \mathcal{L}$, let $\mathbb{E}_l := \mathbb{H}_l - \hat{\mathbb{H}}_l$ be the local channel estimation error for the local channel. Assume that $\hat{\mathbb{H}}_l$ and $\mathbb{E}_l$ are independent. Furthermore, assume $\mathsf{E}[\mathbb{E}_l] = \mathbf{0}$, and that $\mathsf{E}[\mathbb{E}_l^{\mathsf{H}} \mathbb{E}_l] =: \mathbf{\Sigma}_l$ has finite elements. Finally, assume that $(\hat{\mathbb{H}}_l, \mathbb{E}_l)$ and $(\hat{\mathbb{H}}_j, \mathbb{E}_j)$ are independent for $l \neq j$.

Assumption 2 (CSIT Sharing Mechanism): For every $(l, j) \in \mathcal{L}^2$ s.t. $l \neq j$, assume the following Markov chain:

$$\mathbb{H}_l \to \hat{\mathbb{H}}_l \to S_l \to S_j \to \hat{\mathbb{H}}_j \to \mathbb{H}_j.$$

Assumption 1 is widely used in the wireless communication literature and it holds, e.g., for pilot-based MMSE estimates of Gaussian channels [10], [18]. Assumption 2 essentially states that all the available information about $\mathbb{H}_l$ is fully contained in $\hat{\mathbb{H}}_l$ at TX l, and that TX j can only obtain a degraded version of it. We now rewrite the optimality conditions given by (8) in light of the considered model:

Lemma 2: Suppose that Assumption 1, Assumption 2, and the assumption of Theorem 3 hold. Then, the unique TMMSE solution to Problem (4) is given by the unique $\mathbb{t}_k^\star \in \mathcal{T}$ satisfying $(\forall l \in \mathcal{L})$

$$\mathbb{t}_{l,k}^\star(S_l) = \mathbb{F}_l \left(\mathbf{e}_k - \sum_{j \neq l} \mathsf{E} \left[\hat{\mathbb{H}}_j \mathbb{t}_{j,k}^\star \Big| S_l \right] \right) \text{ a.s.,} \quad (11)$$

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

Proof: The first term of the stationarity conditions (8) is evaluated by letting

$$\begin{aligned}
\mathsf{E}[\mathbb{Q}_{l,l} | S_l] &= \mathsf{E}[\mathbb{H}_l^{\mathsf{H}} \mathbb{H}_l | S_l] + P^{-1} \mathbf{I} \\
&= \mathsf{E}[(\hat{\mathbb{H}}_l + \mathbb{E}_l)^{\mathsf{H}} (\hat{\mathbb{H}}_l + \mathbb{E}_l) | \hat{\mathbb{H}}_l] + P^{-1} \mathbf{I} \\
&= \hat{\mathbb{H}}_l^{\mathsf{H}} \hat{\mathbb{H}}_l + \mathbf{\Sigma}_l + P^{-1} \mathbf{I},
\end{aligned}$$

where we used the Markov chain $\mathbb{H}_l \to \hat{\mathbb{H}}_l \to S_l$ and Assumption 1. Then, for $j \neq l$:

$$\begin{aligned}
\mathsf{E}[\mathbb{Q}_{l,j} \mathbb{t}_j^\star | S_l] &= \mathsf{E}[\mathbb{H}_l^{\mathsf{H}} \mathbb{H}_j \mathbb{t}_{j,k}^\star | S_l] \\
&\stackrel{(a)}{=} \mathsf{E} \left[\mathsf{E}[\mathbb{H}_l^{\mathsf{H}} \mathbb{H}_j | S_l, S_j] \mathbb{t}_{j,k}^\star \big| S_l \right] \\
&\stackrel{(b)}{=} \mathsf{E} \left[\mathsf{E}[\mathbb{H}_l^{\mathsf{H}} | S_l, S_j] \mathsf{E}[\mathbb{H}_j | S_l, S_j] \mathbb{t}_{j,k}^\star \big| S_l \right] \\
&\stackrel{(c)}{=} \mathsf{E} \left[\mathsf{E}[\mathbb{H}_l^{\mathsf{H}} | \hat{\mathbb{H}}_l] \mathsf{E}[\mathbb{H}_j | \hat{\mathbb{H}}_j] \mathbb{t}_{j,k}^\star \big| S_l \right] \\
&\stackrel{(d)}{=} \hat{\mathbb{H}}_l^{\mathsf{H}} \mathsf{E}[\hat{\mathbb{H}}_j \mathbb{t}_{j,k}^\star | S_l],
\end{aligned}$$

where (a) follows from the law of total expectation and $\mathbb{t}_{j,k} = \mathbb{t}_{j,k}(S_j)$, (b) from the Markov chain $\mathbb{H}_l \to (S_l, S_j) \to \mathbb{H}_j$, (c) from the Markov chain $\mathbb{H}_l \to \hat{\mathbb{H}}_l \to (S_l, S_j)$, and (d)

from Assumption 1. Note that all the aforementioned Markov chains are implied by Assumption 2. The proof is concluded by using $\mathsf{E}[\mathbb{g}_{l,k} | S_l] = \hat{\mathbb{H}}_l^{\mathsf{H}} \mathbf{e}_k$ and by rearranging the terms. $\square$

The above lemma reveals the following structure of the optimal TMMSE solution: the matrix $\mathbb{F}_l$ can be recognized as a *local* MMSE precoding stage (studied, e.g., in [18]), that is, a centralized MMSE solution [22] assuming that there are no other TXs than TX l; the remaining part can be then interpreted as a 'corrective' stage which takes into account the effect of the other TXs based on the available CSIT and long-term statistical information.

A. No CSIT Sharing

As an important example, we assume that no local channel estimate is shared along the fronthaul. This corresponds to the original cell-free massive MIMO setup studied in [10]. Specifically, we let $\hat{\mathbb{H}}_l$ as in Assumption 1 and

$$(\forall l \in \mathcal{L}) \quad S_l = \hat{\mathbb{H}}_l. \quad (12)$$

Theorem 4: The TMMSE precoders solving (11) under no CSIT sharing (12) are given by

$$\mathbb{t}_{l,k}^\star(S_l) = \mathbb{F}_l \mathbf{C}_l \mathbf{e}_k, \quad (13)$$

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 $\mathbf{\Pi}_l := \mathsf{E} \left[\hat{\mathbb{H}}_l \mathbb{F}_l \right]$.

Proof: Substituting (13) into the optimality conditions (11), we need to show that

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

By the independence between $\hat{\mathbb{H}}_l$ and $\hat{\mathbb{H}}_j$, we can drop the conditioning on S_l and obtain $\hat{\mathbb{H}}_l^{\mathsf{H}} \left(\mathbf{C}_l + \sum_{j \neq l} \mathbf{\Pi}_j \mathbf{C}_j - \mathbf{I} \right) \mathbf{e}_k = \mathbf{0}$ a.s., $\forall l \in \mathcal{L}$. The proof is concluded by observing that $\mathbf{C}_l + \sum_{j \neq l} \mathbf{\Pi}_j \mathbf{C}_j = \mathbf{I}$, $\forall l \in \mathcal{L}$, always has a unique solution, as shown in Appendix D. $\square$

The optimal solution (13) corresponds to a two-stage precoding scheme composed by a local MMSE precoding stage $\mathbb{F}_l$ preceded by a statistical precoding stage $\mathbf{C}_l$. By letting the rows $\hat{\mathbb{g}}_{l,k}^{\mathsf{H}}$ of $\hat{\mathbb{H}}_l$ to be independent and distributed as $\mathcal{CN}(\mathbf{0}, \mathbf{K}_{l,k})$, corresponding for instance to a non-line-of-sight (NLoS) scenario with no pilot contamination [18], it can be shown that the matrices $\mathbf{\Pi}_l$ are diagonal. Hence, (13) takes the simpler form

$$(\forall l \in \mathcal{L}) \quad \mathbb{t}_{l,k}^\star(S_l) = c_{l,k} \mathbb{F}_l \mathbf{e}_k, \quad (14)$$

which, by mapping the optimal $c_{l,k}$ to the optimal large-scale fading decoding coefficients in a dual UL channel, was already studied in [28]. However, if the channels have non-zero mean, such as in line-of-sight (LoS) models, (13) may provide significantly higher rates than (14). To see this, let $\hat{\mathbb{H}}_l \approx \bar{\mathbf{H}}_l$ for some fixed matrix $\bar{\mathbf{H}}_l$, $\forall l \in \mathcal{L}$. Then, since $\bar{\mathbf{H}}_l$ is statistical information known to all TXs, the TMMSE precoders should take a form similar to a 'long-term' centralized

(a) $\hat{\mathbb{H}}_1$ $\hat{\mathbb{H}}_2$ $\hat{\mathbb{H}}_L$

(TX 1) — (TX 2) — - - - — (TX L) — CPU

(b) $\hat{\mathbb{H}}_1 \longrightarrow \quad \hat{\mathbb{H}}_1, \hat{\mathbb{H}}_2 \longrightarrow \quad\quad \hat{\mathbb{H}}_1, \hat{\mathbb{H}}_2, \ldots, \hat{\mathbb{H}}_L$
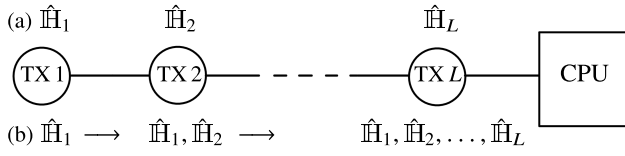
Fig. 1. Pictorial representation of a radio stripe with (a) no CSIT sharing, and (b) unidirectional CSIT sharing.

MMSE solution, which cannot be implemented using (14). Finally, we point out that a suboptimal variation of (13) called *optimal bilinear equalizer* (OBE), with $\mathbb{F}_l$ replaced by $\hat{\mathbb{H}}_l^{\mathsf{H}}$, was already proposed in [26] as a low-complexity alternative to centralized MMSE precoding which maintains robustness against pilot contamination.

B. Unidirectional CSIT Sharing

We now consider a more involved example and let the local channel measurements be shared unidirectionally along a serial fronthaul. This setup is relevant, e.g., for 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 central processing unit (CPU) located at one edge, an architecture also known as a *radio stripe* [17], [27].

Specifically, $\forall l \in \mathcal{L}$, we let $\hat{\mathbb{H}}_l$ as in Assumption 1 and

$$(\forall l \in \mathcal{L}) \ S_l = (\hat{\mathbb{H}}_1, \ldots, \hat{\mathbb{H}}_l). \tag{15}$$

This particular information structure can be interpreted as the CSIT which is accumulated at every TX during the first phase of a centralized precoding scheme for radio stripes, where the CPU collects measurements of the $K \times LN$ channel matrix $\mathbb{H}$ through the serial fronthaul.

Theorem 5: The TMMSE precoders solving (11) under unidirectional CSIT sharing (15) are given by

$$(\forall l \in \mathcal{L}) \ \mathtt{t}_{l,k}^{\star}(S_l) = \mathbb{F}_l \mathbb{V}_l \left[\prod_{i=1}^{l-1} \bar{\mathbb{V}}_i \right] \mathbf{e}_k \tag{16}$$

where we use the following short-hands:

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

Proof: We first assume that all the matrix inverses involved in the following steps exist. Substituting (16) into (11), we need to show that $(\forall l \in \mathcal{L})$

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

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

$$\left(\mathbb{V}_l + \sum_{j>l} \mathsf{E}\left[\mathbb{P}_j \mathbb{V}_j \prod_{i=l+1}^{j-1} \bar{\mathbb{V}}_i \right] \bar{\mathbb{V}}_l \right) \prod_{i=1}^{l-1} \bar{\mathbb{V}}_i$$

$$+ \sum_{j<l} \mathbb{P}_j \mathbb{V}_j \prod_{i=1}^{j-1} \bar{\mathbb{V}}_i, \tag{17}$$

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

$$\sum_{j>l} \mathsf{E}\left[\mathbb{P}_j \mathbb{V}_j \prod_{i=l+1}^{j-1} \bar{\mathbb{V}}_i \right]$$

$$= \sum_{j>l} \mathsf{E}[\mathbb{P}_j \mathbb{V}_j] \prod_{i=l+1}^{j-1} \mathsf{E}[\bar{\mathbb{V}}_i]$$

$$= \mathsf{E}[\mathbb{P}_{l+1} \mathbb{V}_{l+1}] + \sum_{j>l+1} \mathsf{E}[\mathbb{P}_j \mathbb{V}_j] \prod_{i=l+1}^{j-1} \mathsf{E}[\bar{\mathbb{V}}_i]$$

$$= \mathsf{E}[\mathbb{P}_{l+1} \mathbb{V}_{l+1}] + \left(\sum_{j>l+1} \mathsf{E}[\mathbb{P}_j \mathbb{V}_j] \prod_{i=l+2}^{j-1} \mathsf{E}[\bar{\mathbb{V}}_i] \right) \mathsf{E}[\bar{\mathbb{V}}_{l+1}].$$

The second and last term of the above chain of equalities define a recursion terminating with $\mathsf{E}[\mathbb{P}_L \mathbb{V}_L] + \mathbf{0}\mathsf{E}[\bar{\mathbb{V}}_L] = \mathbf{\Pi}_{L-1}$. This recursion gives precisely $\sum_{j>l} \mathsf{E}\left[\mathbb{P}_j \mathbb{V}_j \prod_{i=1}^{j-1} \bar{\mathbb{V}}_i \right] = \mathbf{\Pi}_l$. Together with the property $\mathbb{V}_l + \mathbf{\Pi}_l \bar{\mathbb{V}}_l = \mathbf{I}$, (17) simplifies to

$$\prod_{i=1}^{l-1} \bar{\mathbb{V}}_i + \sum_{j<l} \mathbb{P}_j \mathbb{V}_j \prod_{i=1}^{j-1} \bar{\mathbb{V}}_i$$

$$= \left(\bar{\mathbb{V}}_{l-1} + \mathbb{P}_{l-1} \mathbb{V}_{l-1} \right) \prod_{i=1}^{l-2} \bar{\mathbb{V}}_i + \sum_{j<l-1} \mathbb{P}_j \mathbb{V}_j \prod_{i=1}^{j-1} \bar{\mathbb{V}}_i$$

$$= \prod_{i=1}^{l-2} \bar{\mathbb{V}}_i + \sum_{j<l-1} \mathbb{P}_j \mathbb{V}_j \prod_{i=1}^{j-1} \bar{\mathbb{V}}_i,$$

where the last equation follows from the definition of $\bar{\mathbb{V}}_l$, and where we identify another recursive structure among the remaining terms. By continuing until termination, we finally obtain $\prod_{i=1}^{l-1} \bar{\mathbb{V}}_i + \sum_{j<l} \mathbb{P}_j \mathbb{V}_j \prod_{i=1}^{j-1} \bar{\mathbb{V}}_i = \mathbf{I}$, which proves the main statement under the assumption that all the matrix inverses involved exist. This assumption is indeed always satisfied, as shown in Appendix E. $\square$

By locally computing precoders based on S_l only, and at the expense of some performance loss, the scheme in (16) eliminates the additional overhead required by centralized precoding to share back the computed $K \times LN$ precoding matrix from the CPU to the TXs. Furthermore, inspired by the schemes proposed in [17], [27], [28] for UL processing exploiting the peculiarity of a serial fronthaul, the CSIT sharing overhead can be further reduced as follows:

Remark 2: The scheme in (16) can be alternatively implemented via a recursive algorithm involving a $K \times K$ aggregate information matrix $\prod_{i=1}^{l-1} \bar{\mathbb{V}}_i$ which is sequentially processed and forwarded in the direction from TX 1 to TX L. Therefore, the capacity of the serial fronthaul can be made independent from L, which is typically larger than K.

Furthermore, if data sharing is implemented through the sequential forwarding of a vector $\mathtt{u} := [U_1, \ldots, U_K]^{\mathsf{T}} \in \mathbb{C}^K$ of coded and modulated I/Q symbols originating from a CPU placed next to TX 1, then this can be replaced by the

forwarding of a sequentially precoded K-dimensional vector $\prod_{i=1}^{l-1} \bar{\mathbb{V}}_i \mathbb{u}$, thus eliminating the CSIT sharing overhead.

We conclude this section by providing the following corollary to Theorem 5.

Corollary 1: An alternative expression for centralized MMSE precoding [18], [22], or equivalently, for the TMMSE solution under full CSIT sharing $S_l = (\hat{\mathbb{H}}_1, \ldots, \hat{\mathbb{H}}_L) \; \forall l \in \mathcal{L}$, is given by (16) *with $\mathbf{\Pi}_l$ replaced by $\bar{\mathbb{P}}_l := \mathbb{P}_{l+1}\mathbb{V}_{l+1} + \bar{\mathbb{P}}_{l+1}\bar{\mathbb{V}}_{l+1}$, $\bar{\mathbb{P}}_L := \mathbf{0}$.*

Proof: Since all random quantities become deterministic after conditioning on S_l, the proof of Theorem 5 can be repeated by removing $\mathsf{E}[\cdot]$ everywhere. □

The expression in Corollary 1 can be alternatively derived by applying recursively known block-matrix inversion lemmas to the original centralized MMSE precoding expression [22]. The details are omitted due to space limitations. Similarly to the implementation of (16) described in Remark 2, Corollary 1 provides a novel distributed and recursive implementation of centralized MMSE precoding. The main difference is that, in contrast to $\mathbf{\Pi}_l$ which can be computed offline, the computation of $\bar{\mathbb{P}}_l$ entails an additional sequential procedure in the reverse direction, thus increasing the overhead.

C. Asymptotic Results and Relation With the SGD Scheme [29]

The idea of designing recursive precoding schemes exploiting the opportunities of a serial connection between antenna elements has been also explored by [29]. Motivated by the need of reducing hardware complexity of a massive MIMO cellular base station, and by focusing on $N = 1$ and no channel estimation error, the authors of [29] propose the following so-called *SGD* precoding scheme:

$$(\forall l \in \mathcal{L}) \; T_{l,k}(S_l) = \mu_{l,k} \mathbb{h}_l^{\mathsf{H}} \left(\mathbf{e}_k - \sum_{j=1}^{l-1} \mathbb{h}_j T_{j,k}(S_j) \right), \quad (18)$$

where $\mathbb{h}_l := \hat{\mathbb{H}}_l = \mathbb{H}_l$, S_l is given by (15) assuming unidirectional CSIT sharing, and $\mu_{l,k} \in \mathbb{R}$ are tunable step-sizes of a stochastic gradient descent algorithm. The choice $\mu_{l,k} = \|\mathbb{h}_l\|^{-2}$ is motivated by [29] as a good solution for i.i.d. Rayleigh fading and high SNR. Furthermore, to cope with finite SNR, [29] suggests to take $\mu_{l,k} = \mu_k \|\mathbb{h}_l\|^{-2}$ for a single deterministic scalar $\mu_k \in \mathbb{R}$ per RX to be optimized, e.g., using line search. Interestingly, the SGD scheme with $\mu_{l,k} = \|\mathbb{h}_l\|^{-2}$ can be also derived from team theoretical arguments, as a particular case of the following asymptotic result for $N \geq 1$:

Lemma 3: Assume $\mathrm{vec}(\mathbb{H}) \sim \mathcal{CN}(\mathbf{0}, \mathbf{I})$, $\hat{\mathbb{H}}_l = \mathbb{H}_l \; \forall l \in \mathcal{L}$, *unidirectional CSIT sharing* (15)*, and let $\mathbb{t}_k^\star$ be the optimal TMMSE solution of Problem* (4)*. Then,*

$$R_k = \log(\mathrm{MSE}_k(\mathbb{t}_k^\star))^{-1} \leq L \log\left(\frac{K}{K-N}\right),$$

with equality attained as $P \to \infty$ by $(\forall l \in \mathcal{L})$

$$\mathbb{t}_{l,k}(S_l) = (\mathbb{H}_l^{\mathsf{H}}\mathbb{H}_l)^{-1}\mathbb{H}_l^{\mathsf{H}}\left(\mathbf{e}_k - \sum_{j=1}^{l-1}\mathbb{H}_j \mathbb{t}_{j,k}(S_j)\right). \quad (19)$$
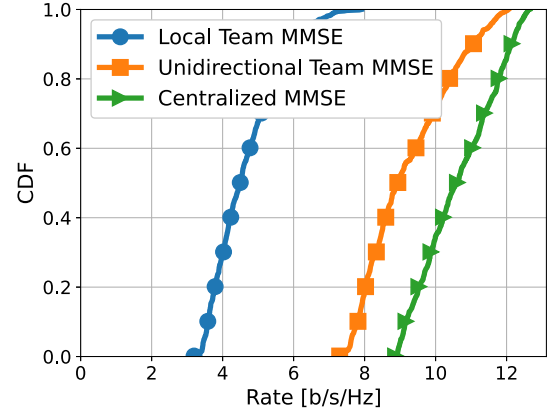
Proof: The proof is given in Appendix F. □

Fig. 2. Comparison among different CSIT configuration, empirical CDF of the optimal per-RX achievable rates. Unidirectional TMMSE is a promising intermediate solution for supporting network-wide interference management when centralized precoding becomes too costly.

V. PERFORMANCE EVALUATION

A. Simulation Setup

Inspired by the "football arena" [27] or "outdoor piazza" [17] scenarios, 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 $H_{l,k,n} \sim \mathcal{CN}(0, \rho_{l,k}^2)$, where $\rho_{l,k}^2$ denotes the channel gain between TX l and RX k. We follow the 3GPP NLoS Urban Microcell path-loss model [35, Table B.1.2.1-1]

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

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 $P_{\mathrm{noise}} = -174 + 10\log_{10}(B/1 \text{ Hz}) + F$ dBm, where $B = 20$ MHz is the system bandwidth, and $F = 7$ dB is the noise figure. Finally, we let $\rho_{l,k}^2 := 10^{-\frac{\mathrm{PL}_{l,k}+P_{\mathrm{noise}}}{10}}$ mW^{-1}, and, leveraging the short distances, we consider a relatively low total radiated power $P_{\mathrm{sum}} = 100$ mW.

B. Comparison Among Different CSIT Configurations

We numerically evaluate the Pareto optimal achievable rates $R_k = -\log(\mathrm{MSE}_k(\mathbb{t}_k^\star))$, where $\mathbb{t}_k^\star$ denotes the optimal solution of Problem (4), under the following CSIT configurations: (i) no CSIT sharing (12), (ii) unidirectional CSIT sharing (15), and (iii) full CSIT sharing as in Corollary 1. The resulting optimal precoding schemes are respectively denoted by (i) local TMMSE, (ii) unidirectional TMMSE, and (iii) centralized MMSE. We assume for simplicity $\hat{\mathbb{H}}_l = \mathbb{H}_l$ to study the impact of the different CSIT configurations in absence of measurement noise, and focus on the Pareto optimal point parametrized by $w_k = 1 \; \forall k \in \mathcal{K}$.

Figure 2 reports the empirical cumulative distribution function (CDF) of R_k for multiple i.i.d. realizations of the RX locations. As expected, adding information constraints on the

CSIT configuration leads to performance degradation. However, the degradation is less pronounced from centralized to unidirectional MMSE precoding, showing that unidirectional CSIT sharing does not prevent effective forms of network-wide interference management. Therefore, the unidirectional team MMSE scheme appears as a promising intermediate solution whenever centralized MMSE precoding becomes too costly, e.g., when the CSIT sharing overhead becomes problematic due to high RXs mobility. Quantifying the savings in terms of CSIT sharing overhead is an interesting open problem which depends on many implementation details. For instance, if computational complexity is not an issue and the message sharing is implemented through the forwarding of high-precision I/Q symbols, unidirectional TMMSE precoding may have the same overhead as local TMMSE precoding, owing to the sequential implementation outlined in Remark 2. If this is not possible, for instance because operations such as matrix inversions at each symbol time are not allowed, then the savings may become less prominent, e.g., down to a factor 2.

C. Comparison Among Local Precoding Schemes

In this section, we compare the optimal local TMMSE solution against classical MRT, the OBE method [26], and local MMSE precoding (14) with optimal large-scale fading coefficients $c_{l,k}$ computed using the method in [28]. Since the bound in (5) may be overly pessimistic for suboptimal schemes, for a fair comparison we compute the DL rates $R_k = R_k^{\text{hard}}$ by means of their dual UL rates R_k^{UatF} as defined in the proof of Theorem 1, using the same dual UL power allocation $w_k = 1 \ \forall k \in \mathcal{K}$. One of the major weaknesses of MRT and local MMSE precoding is that they do not exploit channel mean information, typically arising from LoS components. To study this effect, we modify our simulation setup by letting $N = 1$ and by considering a simple Ricean fading model $H_{l,k,1} \sim \mathcal{CN}\left(\sqrt{\frac{\kappa}{\kappa+1}\rho_{l,k}^2}, \frac{1}{\kappa+1}\rho_{l,k}^2\right)$ for some $\kappa \geq 0$, and consider again no measurement noise $\hat{\mathbb{H}}_l = \mathbb{H}_l$. Figure 3 confirms the above observation: while, as expected, local MMSE precoding is optimal for a NLoS setup ($\kappa = 0$), it may incur significant performance loss w.r.t. local TMMSE precoding and the OBE method even in case of relatively weak LoS components ($\kappa = 1$).

D. Comparison Between Unidirectional TMMSE Precoding and the SGD Scheme [29]

In this section, we compare the unidirectional TMMSE solution (15) for $N = 1$ against the suboptimal SGD scheme (18) proposed in [29] for $\mu_k = 1$ and its robust version obtained by optimizing μ_k statistically via line search. Figure 4 plots the rate $R_1 = R_1^{\text{hard}}$ of the first RX (measured via its dual UL rate as in Section V-C) versus the $\text{SNR} := P \sum_l \rho_{l,1}^2$ for a single realization of the simulation setup, and by focusing on the following aspects:

(a) Equal path-loss, i.e., $r_2 = 0$ (all RXs colocated at the center of the circular service area), and no channel estimation errors, i.e., $\mathbb{E}_l = \mathbf{0}_{K \times 1} \ \forall l \in \mathcal{L}$;
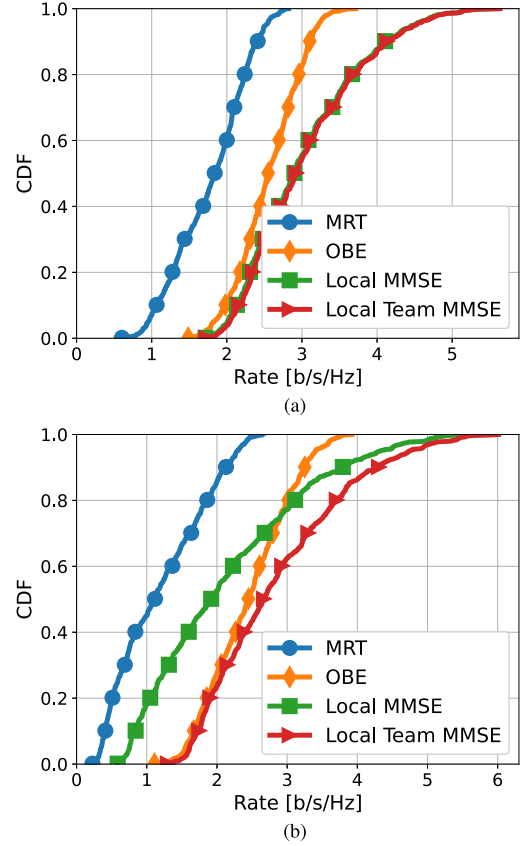

Fig. 3.   Empirical CDF of the per-RX achievable rates for different local precoding schemes, under Ricean factor (a) $\kappa = 0$, and (b) $\kappa = 1$. In contrast to previously known heuristics, the team MMSE approach optimally exploits statistical information such as the channel mean and always exhibits superior performance. Furthermore, consistently with our theoretical results, local MMSE precoding [28] is optimal in case (a). However, as expected, and in contrast to the team MMSE approach and the OBE method [26], it may not handle well the interference originating from the channel mean, as shown by the poor performance of the weaker RXs in case (b).

(b) Equal path-loss, and channel estimation errors, i.e., we let $\mathbb{E}_l \sim \mathcal{CN}(\mathbf{0}_{K \times 1}, \epsilon \mathbf{K}_l) \ \forall l \in \mathcal{L}$ and $\hat{\mathbb{H}}_l \sim \mathcal{CN}(\mathbf{0}_{K \times 1}, (1 - \epsilon)\mathbf{K}_l)$, where $\mathbf{K}_l = \text{diag}(\rho_{l,1}, \dots, \rho_{l,K})$ and $\epsilon = 0.2$;

(c) Realistic path-loss, i.e., $r_2 = 50$ m (single realization), and no channel estimation errors.

Although the SGD scheme assumes no channel estimation errors, in the above experiments we adapt (18) to case (b) by replacing $\mathbb{h}_l$ with $\hat{\mathbb{H}}_l$ everywhere. As expected, from Figure 4a we observe that the SGD scheme is asymptotically optimal in case (a), but its performance degrades for low SNR, or in the presence of channel estimation error noise and/or realistic path-loss as shown in Figure 4b and 4c. In contrast, its robust version seems sufficient to recover most of the loss due to finite SNR and channel estimation errors. However, Figure 4c shows that the (robust) SGD scheme may not handle more realistic path-loss configurations.

The main advantage of the (robust) SGD scheme over optimal unidirectional TMMSE precoding is that the former does not perform any $K \times K$ matrix inversion. Therefore, it may be considered as a low-complexity alternative to unidirectional TMMSE precoding. However, further research is needed in particular regarding the choice of the parameters $\mu_{l,k}$ and the support for $N > 1$ TX antennas.
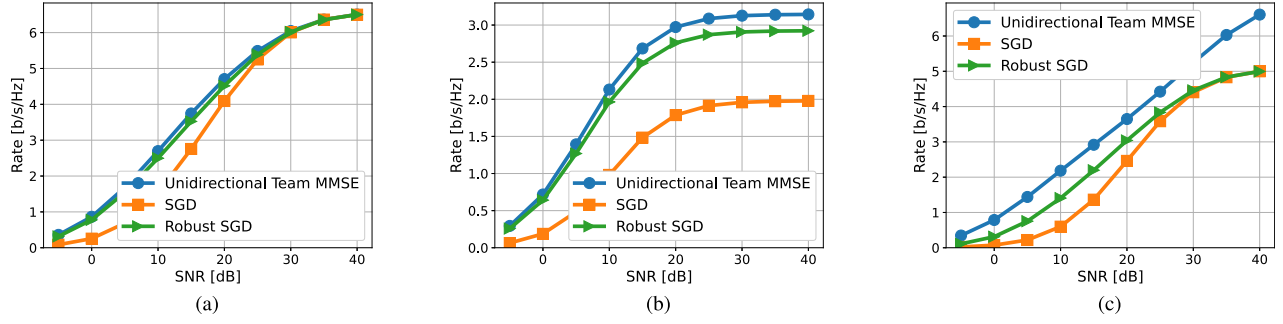
Fig. 4. Rate vs SNR for RX 1 under: (a) equal path-loss and no channel estimation errors; (b) equal path-loss and channel estimation errors; (c) realistic path-loss and no channel estimation errors. In contrast to the (robust) SGD scheme [29], the team MMSE approach optimally exploits the path loss information and hence exhibits superior performance in case (c).

VI. CONCLUDING REMARKS

This work provides novel guidelines for distributed precoding design in systems with distributed CSIT such as cell-free massive MIMO networks. By assuming full data sharing and a sum-power constraint, the proposed optimal approach outperforms previously known heuristic methods in several setups of interest. Going beyond the two chosen examples where the optimal precoders can be derived explicitly, the proposed approach can be potentially applied, exactly or using standard numerical approximations, to a wide range of practical setups. Furthermore, since it exploits the UL-DL duality principle, we remark that the proposed approach also provides optimal distributed combiners for UL operations. Although not covered for simplicity, the proposed approach may be also extended to limited data sharing, for example using network clustering techniques. Other promising lines of research include the extension to different power constraints, the revisitation of cell-free massive MIMO performance analysis involving LoS models or pilot contamination, and the development of new efficient algorithms exploiting fronthaul architectures such as tree or star topologies.

APPENDIX

A. Proof of Theorem 1

Consider a dual UL network with K single-antenna TXs and L cooperating RXs equipped with N antennas each, governed by the MIMO channel law $\mathbb{y}^{\text{UL}} = \sum_{k=1}^{K} \sqrt{P w_k} \mathbb{g}_k V_k + \mathbb{n}^{\text{UL}}$, where $\mathbb{y}^{\text{UL}} \in \mathbb{C}^{LN}$ is the received signal at all RXs, $[\mathbb{g}_1, \ldots, \mathbb{g}_K] = \mathbb{H}^{\mathsf{H}}$ is the dual channel matrix, $V_k \sim \mathcal{CN}(0,1)$ is the independent message of TX k, and $\mathbb{n}^{\text{UL}} \sim \mathcal{CN}(\mathbf{0}, \mathbf{I})$. Then, we consider the processed channel $\hat{V}_k = \frac{1}{\sqrt{P}} \mathbb{t}_k^{\mathsf{H}} \mathbb{y}^{\text{UL}}$, where $\mathbb{t}_k = [\mathbb{t}_{1,k}^{\mathsf{T}} \ldots \mathbb{t}_{L,k}^{\mathsf{T}}]^{\mathsf{T}}$ is a distributed linear combiner satisfying the information constraint $\mathbb{t}_k \in \mathcal{T}$. Let $R_k^{\text{UL}} := I(V_k; \hat{V}_k)$ be achievable rates on this channel. By standard information inequalities [36], we obtain

$$
\begin{aligned}
I(V_k; \hat{V}_k) &= h(V_k) - h(V_k | \hat{V}_k) \\
&\geq \log(\pi e) - h(V_k - \alpha \hat{V}_k) \\
&\geq \log(\pi e) - \log(\pi e \mathsf{E}[|V_k - \alpha \hat{V}_k|^2]).
\end{aligned}
$$

Optimizing α according to channel statistics, i.e., choosing $\alpha = \alpha^\star$ with $\alpha^\star := \mathsf{E}[V_k \hat{V}_k^*]/\mathsf{E}[|\hat{V}_k|^2]$ being the solution of $\min_{\alpha \in \mathbb{C}} \mathsf{E}[|V_k - \alpha \hat{V}_k|^2]$, leads to the well-known UatF

bound [22]

$$
R_k^{\text{UL}} \geq R_k^{\text{UatF}} := \log(1 + \text{SINR}_k),
$$

$$
\text{SINR}_k := \frac{w_k |\mathsf{E}[\mathbb{t}_k^{\mathsf{H}} \mathbb{g}_k]|^2}{w_k \mathsf{Var}[\mathbb{t}_k^{\mathsf{H}} \mathbb{g}_k] + \sum_{j \neq k} w_j \mathsf{E}[|\mathbb{t}_k^{\mathsf{H}} \mathbb{g}_j|^2] + \frac{\mathsf{E}[\|\mathbb{t}_k\|^2]}{P}}.
$$

Alternatively, we can keep $\alpha \in \mathbb{C}$ unoptimized and obtain the bound $R_k^{\text{UL}} \geq R_k^{\text{UatF}} \geq \log(\mathsf{E}[|V_k - \alpha \hat{V}_k|^2])^{-1}$, where after simple manipulations we recognize $\mathsf{E}[|V_k - \alpha \hat{V}_k|^2] = \mathsf{E}\left[\left\|\alpha \mathbf{W}^{\frac{1}{2}} \mathbb{H} \mathbb{t}_k - \mathbf{e}_k\right\|^2\right] + \frac{\alpha^2}{P} \mathsf{E}[\|\mathbb{t}_k\|^2] = \text{MSE}_k(\alpha \mathbb{t}_k)$. The above steps also shows that solving $\min_{\alpha \in \mathbb{C}, \mathbb{t}_k \in \mathcal{T}} \text{MSE}_k(\alpha \mathbb{t}_k)$ is equivalent to solving

$$
\underset{\mathbb{t}_k \in \mathcal{T}}{\text{maximize}} \ R_k^{\text{UatF}}. \tag{20}
$$

Furthermore, we observe that $\min_{\alpha \in \mathbb{C}, \mathbb{t}_k \in \mathcal{T}} \text{MSE}_k(\alpha \mathbb{t}_k) = \min_{\mathbb{t}_k \in \mathcal{T}} \text{MSE}_k(\mathbb{t}_k)$. This is because α is a deterministic scalar, hence $\alpha \mathbb{t}_k \in \mathcal{T}$. Therefore, Problem (20) and Problem (4) have the same optimal solution $\mathbb{t}_k^\star$, and the optima are related by $R_k^{\text{UatF}} = \log(\text{MSE}_k(\mathbb{t}_k^\star)^{-1})$.

Let $R_k^\star(\mathbf{w})$ be the optimum of Problem (20) for some dual UL power allocation policy $\mathbf{w} := (w_1, \ldots, w_K) \in \mathbb{R}_+^K$. Let then $\mathcal{R}^{\text{UatF}}$ be the union of all rate tuples $(R_1, \ldots, R_K) \in \mathbb{R}_+^K$ satisfying $R_k \leq R_k^\star(\mathbf{w}) \ \forall k \in \mathcal{K}$, where the union is taken over all $\mathbf{w}$ satisfying $\sum_{k=1}^{K} P w_k \leq P_{\text{sum}}$. By definition of $\mathcal{R}^{\text{UatF}}$, its Pareto boundary $\partial \mathcal{R}^{\text{UatF}}$ is composed by rate tuples of the type $(R_1, \ldots, R_K)$, $R_k = R_k^\star(\mathbf{w})$, achieved by some $\mathbf{w}$ satisfying $\sum_k w_k \leq K$ and by the optimal combiners $\{\mathbb{t}_k\}_{k=1}^{K}$ solving Problem (4) $\forall k \in \mathcal{K}$. It turns out that it is possible to fully characterize $\partial \mathcal{R}^{\text{UatF}}$ by restricting $\mathbf{w} \in \mathcal{W}$, i.e., by using all the available power $KP = P_{\text{sum}}$. This is because R_k^{UatF} is a continuous monotonic increasing functions of P, and so is its (finite) supremum over $\mathbb{t}_k \in \mathcal{T}$. Furthermore, it can be shown that *all* $\mathbf{w} \in \mathcal{W}$ induce Pareto optimal rate tuples. This last statement can be proven by contradiction as follows.

Let $\mathbf{w} \in \mathcal{W}$ and suppose that $(R_1^\star(\mathbf{w}), \ldots, R_K^\star(\mathbf{w})) \notin \partial \mathcal{R}^{\text{UatF}}$, i.e., $\exists \mathbf{w}' \in \mathcal{W}$, $\mathbf{w}' \neq \mathbf{w}$, s.t. $R_k^\star(\mathbf{w}') > R_k^\star(\mathbf{w}) \ \forall k \in \mathcal{K}$. We now build an iterative procedure which moves from $\mathbf{w}$ to $\mathbf{w}'$ and contradicts the previous supposition. Consider the following sequence of updates $\mathbf{w}^{(i)} := (w_1^{(i)}, \ldots, w_K^{(i)})$ for $i = 0, \ldots, K-1$, where $\mathbf{w}^{(0)} := \mathbf{w}$ and

$$
w_k^{(i)} := \begin{cases} w_k' & \text{if } k \leq i \\ \dfrac{\sum_{j>i} w_j'}{\sum_{j>i} w_j^{(i-1)}} w_k^{(i-1)} & \text{if } k > i \end{cases}.
$$

The i-th step of the above procedure changes w_i into the target w_i' and scales all weights w_k with $k > i$ by a common factor s.t. the constraint $\mathbf{w}^{(i)} \in \mathcal{W}$ is not violated. Note that this constraint also implies that $\mathbf{w}^{(K-1)} = \mathbf{w}'$ without the need of a K-th update. In the following, we use properties of $R_k^\star$ inferred by the fact that $\log_2\left(1 + \frac{ax}{bx+cy+d}\right)$ is continuous monotonic increasing in $x \in \mathbb{R}_+$ and continuous monotonic decreasing in $y \in \mathbb{R}_+$ for any $a \geq 0$, $b \geq 0$, $c \geq 0$, and $d > 0$, and so is its supremum over some family of parameters (a, b, c, d). At step $i = 1$, assume w.l.o.g. that $w_1^{(1)} = w_1' \geq w_1$, which also implies $w_k^{(1)} = \frac{\sum_{j>i} w_j'}{\sum_{j>i} w_j^{(0)}} w_k \leq w_k$ for $k > 1$. In fact, we can always reindex the RXs such that this assumption holds. When going from $x = 1$ to $x = \frac{\sum_{j>i} w_j'}{\sum_{j>i} w_j^{(0)}} \leq 1$, and then subsequently from $y = 1$ to $y = \frac{w_1'}{w_1} \geq 1$, the function $\log$

$$\left(1 + \frac{w_k |\mathsf{E}[\mathbb{t}_k^\mathsf{H} \mathbb{g}_k]|^2 x}{(\sum_{j>1} w_j \mathsf{E}[|\mathbb{t}_k^\mathsf{H} \mathbb{g}_j|^2] - w_k |\mathsf{E}[\mathbb{t}_k^\mathsf{H} \mathbb{g}_k]|^2) x + w_1 \mathsf{E}[|\mathbb{t}_k^\mathsf{H} \mathbb{g}_1|^2] y + \frac{\mathsf{E}[\|\mathbb{t}_k\|^2]}{P}}\right)$$

for $k > 1$ is continuous monotonically decreasing, and so is its supremum over $\mathbb{t}_k \in \mathcal{T}$. Hence, we have $R_k^\star(\mathbf{w}^{(1)}) \leq R_k^\star(\mathbf{w}^{(0)}) \; \forall k > 1$. At step $i = 2$, we assume w.l.o.g. that $w_2' \geq w_2^{(1)}$ (otherwise we can just properly reindex all RXs $k > 2$), and similarly obtain $R_k^\star(\mathbf{w}^{(2)}) \leq R_k^\star(\mathbf{w}^{(1)})$ $\forall k > 2$. By continuing until step $K - 1$, we finally obtain $R_K^\star(\mathbf{w}^{(K-1)}) \leq R_K^\star(\mathbf{w}^{(K-2)})$, which can be combined with the previous steps leading to the desired contradiction $R_K^\star(\mathbf{w}') \leq R_K^\star(\mathbf{w})$ up to a possible reindexing, i.e., at least one rate cannot be strictly increased when moving from $\mathbf{w}$ to $\mathbf{w}'$.

The proof is concluded by invoking the duality principle between the UatF bound and the hardening bound [22, Theorem 4.8], which shows that $\mathcal{R}^{\text{UatF}} = \mathcal{R}^{\text{hard}}$ and that for every rate tuple $(R_1^{\text{UatF}}, \ldots, R_K^{\text{UatF}})$ achieved by some $\{\mathbb{t}_k\}_{k=1}^K$ and $\mathbf{w}$, there is a rate tuple $(R_1^{\text{hard}}, \ldots, R_K^{\text{hard}}) = (R_1^{\text{UatF}}, \ldots, R_K^{\text{UatF}})$ achievable by using the same functions $\{\mathbb{t}_k\}_{k=1}^K$, and by choosing $p_k = \frac{\tilde{p}_k P}{\mathsf{E}[\|\mathbb{t}_k\|^2]} \; \forall k \in \mathcal{K}$ with $\tilde{\mathbf{p}} := [\tilde{p}_1, \ldots, \tilde{p}_K]^\mathsf{T}$ being the solution of $(\mathbf{D}^{-1} - \mathbf{B})\tilde{\mathbf{p}} = (\mathbf{D}^{-1} - \mathbf{B}^\mathsf{T})\mathbf{w}$, where $\mathbf{D} := \text{diag}(d_1, \ldots, d_K)$, $d_k := \text{SINR}_k \frac{\mathsf{E}[\|\mathbb{t}_k\|^2]}{|\mathsf{E}[\mathbb{t}_k^\mathsf{H} \mathbb{g}_k]|^2}$, and where the (k', k)-th element of $\mathbf{B} \in \mathbb{C}^{K \times K}$ is given by

$$[\mathbf{B}]_{k',k} = \begin{cases} \dfrac{\mathsf{E}[|\mathbb{t}_k^\mathsf{H} \mathbb{g}_{k'}|^2]}{\mathsf{E}[\|\mathbb{t}_k\|^2]} & \text{if } k' \neq k \\ \dfrac{\mathsf{E}[|\mathbb{t}_k^\mathsf{H} \mathbb{g}_k|^2] - |\mathsf{E}[\mathbb{t}_k^\mathsf{H} \mathbb{g}_k]|^2}{\mathsf{E}[\|\mathbb{t}_k\|^2]} & \text{otherwise.} \end{cases}$$

The above linear system is guaranteed to have a unique solution satisfying $\sum_{k=1}^K \tilde{p}_k = \sum_{k=1}^K w_k$, which implies $\sum_{l=1}^L \mathsf{E}[\|\mathbb{x}_l\|^2] = \sum_{k=1}^K w_k P = P_{\text{sum}}$.

B. Proof of Theorem 3

To avoid cumbersome notation, we omit the subscript k everywhere. The proof is split into three separate lemmas. We start with a minor extension of [24, Theorem 3] obtained by introducing the constraint $\mathsf{E}[\|\mathbb{t}_k\|^2] < \infty$ and specializing to the considered cost function.

Lemma 4 (Existence and Uniqueness): Problem (4) *admits a unique team optimal solution.*

Proof: Let $\mathcal{H}$ be the space of Σ-measurable functions $\mathbb{a} : \Omega \to \mathbb{C}^{LN}$ s.t. $\mathsf{E}[\mathbb{a}^\mathsf{H} \mathbb{Q} \mathbb{a}] < \infty$. We define the inner product $\langle \mathbb{a}, \mathbb{b} \rangle := \mathsf{E}[\mathbb{b}^\mathsf{H} \mathbb{Q} \mathbb{a}]$, $\forall (\mathbb{a}, \mathbb{b}) \in \mathcal{H}^2$, and its induced norm $\|\mathbb{a}\|_\mathcal{H} := \sqrt{\langle \mathbb{a}, \mathbb{a} \rangle}$, $\forall \mathbb{a} \in \mathcal{H}$. Then, the tuple $(\mathcal{H}, \langle \cdot, \cdot \rangle)$ is a Hilbert space[3] [24]. Let us further define $\mathbb{t}_0 := \mathbb{Q}^{-1} \mathbb{g}$, which is the unique minimizer of $c(\mathbf{H}, \mathbf{t})$ for any realization $\mathbf{H}$. Firstly, we observe that $\mathbb{t}_0 \in \mathcal{H}$, since

$$\|\mathbb{t}_0\|_\mathcal{H}^2 = \mathsf{E}\left[\mathbb{e}^\mathsf{H} \mathbb{H} \left(\mathbb{H}^\mathsf{H} \mathbb{H} + P^{-1} \mathbf{I}\right)^{-1} \mathbb{H}^\mathsf{H} \mathbb{e}\right]$$

$$\leq \mathsf{E}\left[\text{tr}\left(\mathbb{H}\left(\mathbb{H}^\mathsf{H} \mathbb{H} + P^{-1} \mathbf{I}\right)^{-1} \mathbb{H}^\mathsf{H}\right)\right]$$

$$= \mathsf{E}\left[\text{tr}\left(\mathbb{H}^\mathsf{H} \mathbb{H} \left(\mathbb{H}^\mathsf{H} \mathbb{H} + P^{-1} \mathbf{I}\right)^{-1}\right)\right]$$

$$\leq \mathsf{E}\left[\text{tr}\left(\left(\mathbb{H}^\mathsf{H} \mathbb{H} + P^{-1} \mathbf{I}\right)\left(\mathbb{H}^\mathsf{H} \mathbb{H} + P^{-1} \mathbf{I}\right)^{-1}\right)\right]$$

$$= NL. \tag{21}$$

Secondly, we observe that $\mathcal{T}$ is a closed linear subspace of $\mathcal{H}$ [25, Theorem 2.6.6]. Finally, simple algebraic manipulations show that the objective of Problem (4) can be equivalently rewritten as $\text{MSE}(\mathbb{t}) = \|\mathbb{t} - \mathbb{t}_0\|_\mathcal{H}^2 - \|\mathbb{t}_0\|_\mathcal{H}^2 + 1$. Therefore, by following [24], [25], we consider the infinite dimensional orthogonal projection problem:

$$\underset{\mathbb{t} \in \mathcal{T}}{\text{minimize}} \; \|\mathbb{t} - \mathbb{t}_0\|_\mathcal{H}^2. \tag{22}$$

The solution to Problem (4) corresponds to the projection of $\mathbb{t}_0 \in \mathcal{H}$ onto the closed linear subspace $\mathcal{T} \subseteq \mathcal{H}$. By the Hilbert projection theorem, this projection is unique and always exists [37]. $\square$

The following result extends Lemma 4 by following similar lines as [25, Theorem 2.6.6].

Lemma 5 (Sufficiency of Stationarity): Suppose that $\mathsf{E}[\|\mathbb{Q}\|_\mathsf{F}^2] < \infty$. *Then, if* $\mathbb{t}^\star \in \mathcal{T}$ *is stationary, it is also the unique optimal solution to Problem* (4).

Proof: Let us consider again the equivalent problem (22). Since $\mathcal{T}$ is a closed linear subspace, the Hilbert projection theorem also states that a solution $\mathbb{t}^\star \in \mathcal{T}$ is the unique optimal solution if and only if the following orthogonality conditions [37] hold: $(\forall \mathbb{t} \in \mathcal{T})$

$$\langle \mathbb{t}^\star - \mathbb{t}_0, \mathbb{t} \rangle$$
$$= 0 \iff$$
$$\mathsf{E}\left[\mathbb{t}^\mathsf{H} \mathbb{Q}\left(\mathbb{t}^\star - \mathbb{Q}^{-1}\mathbb{g}\right)\right]$$
$$= 0 \iff$$
$$\mathsf{E}\left[\sum_{l=1}^L \mathbb{t}_l^\mathsf{H}\left(\mathsf{E}[\mathbb{Q}_{l,l}|S_l]\mathbb{t}_l^\star + \sum_{j \neq l} \mathsf{E}[\mathbb{Q}_{l,j}\mathbb{t}_j^\star|S_l] - \mathsf{E}[\mathbb{g}_l|S_l]\right)\right]$$
$$= 0, \tag{23}$$

where the last equality follows by the law of total expectation, provided that the inner expectations are finite. Finiteness of $\mathsf{E}[\mathbb{g}_l|S_l]$ and $\mathsf{E}[\mathbb{Q}_{l,l}|S_l]$ follows by the assumption

---

[3]In fact, the positive matrix square root $\mathbb{Q}^{\frac{1}{2}}$ induces an isometry between $(\mathcal{H}, \langle \cdot, \cdot \rangle)$ and the perhaps more familiar Hilbert space of measurable functions such that $\mathsf{E}[\|\mathbb{a}\|^2] < \infty$, equipped with the standard inner product $\langle \mathbb{a}, \mathbb{b} \rangle := \mathsf{E}[\mathbb{b}^\mathsf{H} \mathbb{a}]$.

$\mathsf{E}[\|\mathbb{H}\|_{\mathrm{F}}^2] < \infty$. Finiteness of $\mathsf{E}[\mathbb{Q}_{l,j}\mathfrak{t}_j^\star|S_l]$ follows by applying the Cauchy-Schwarz inequality elementwise, and by using $\mathsf{E}[\|\mathfrak{t}_j\|^2] < \infty$ and $\mathsf{E}[\|\mathbb{Q}\|_{\mathrm{F}}^2] < \infty$. The proof is concluded by observing that if the stationary conditions in (8) are satisfied for some $\mathfrak{t}^\star$, then the orthogonality conditions are satisfied and $\mathfrak{t}^\star$ is the unique optimal solution. $\qquad\square$

To conclude the proof, it remains to show the converse statement of Lemma 5. In the following, we depart from [25, Theorem 2.6.6] and use a different argument tailored to the cost function considered in here.

Lemma 6 (Necessity of Stationarity): Suppose that $\mathsf{E}[\|\mathbb{Q}\|_{\mathrm{F}}^2] < \infty$. *Then, if* $\mathfrak{t}^\star \in \mathcal{T}$ *is the unique optimal solution to Problem* (4), *it is also stationary.*

Proof: We start by using the so-called notion of *person-by-person optimality* [24], [25]. Similarly to the notion of Nash equilibrium, it states that a necessary condition for a tuple $\mathfrak{t}^\star$ to be globally optimal is that it must satisfy

$$(\forall l \in \mathcal{L}) \quad \mathrm{MSE}(\mathfrak{t}^\star) = \min_{\mathfrak{t}_l \in \mathcal{T}_l} \mathrm{MSE}(\mathfrak{t}_{-l}^\star, \mathfrak{t}_l).$$

We relax the the above conditions by letting $\mathcal{T}_{l,\mathrm{unc}}$ be the unconstrained version of $\mathcal{T}_l$, i.e., where we remove the constraint $\mathsf{E}[\|\mathfrak{t}_l\|^2] < \infty$. We then have $(\forall l \in \mathcal{L})$

$$\begin{aligned}
\infty > \mathrm{MSE}(\mathfrak{t}^\star) &= \min_{\mathfrak{t}_l \in \mathcal{T}_l} \mathrm{MSE}(\mathfrak{t}_{-l}^\star, \mathfrak{t}_l) \\
&\geq \min_{\mathfrak{t}_l \in \mathcal{T}_{l,\mathrm{unc}}} \mathrm{MSE}(\mathfrak{t}_{-l}^\star, \mathfrak{t}_l) \\
&= \min_{\mathfrak{t}_l \in \mathcal{T}_{l,\mathrm{unc}}} \mathsf{E}\left[\mathsf{E}\left[c\left(\mathbb{H}, \mathfrak{t}_{-l}^\star, \mathfrak{t}_l\right)\big|S_l\right]\right] \\
&\geq \mathsf{E}\left[\min_{\mathfrak{t}_l \in \mathcal{T}_{l,\mathrm{unc}}} \mathsf{E}\left[c\left(\mathbb{H}, \mathfrak{t}_{-l}^\star, \mathfrak{t}_l\right)\big|S_l\right]\right] \\
&= \mathsf{E}\left[\mathsf{E}\left[c\left(\mathbb{H}, \mathfrak{t}_{-l}^\star, \mathfrak{t}_l^{\star\star}\right)\big|S_l\right]\right] \\
&= \mathrm{MSE}(\mathfrak{t}_{-l}^\star, \mathfrak{t}_l^{\star\star}) \qquad\qquad\qquad (24)
\end{aligned}$$

where $\mathfrak{t}_l^{\star\star}$ is given by the first-order optimality condition $\nabla_{\mathbf{t}_l}\phi_l(S_l, \mathbf{t}_l) = \mathbf{0}$ a.s. applied to the convex function $\phi_l(s_l, \mathbf{t}_l) := \mathsf{E}\left[c(\mathbb{H}, \mathfrak{t}_{-l}^\star, \mathbf{t}_l)\big|S_l = s_l\right]$. Note that

$$\mathfrak{t}_l^{\star\star}(S_l) := (\mathsf{E}[\mathbb{Q}_{l,l}|S_l])^{-1}\left(\mathsf{E}[\mathfrak{g}_l|S_l] - \sum_{j \neq l}\mathsf{E}[\mathbb{Q}_{l,j}\mathfrak{t}_j^\star|S_l]\right)$$

satisfies $\mathfrak{t}_l^{\star\star} \in \mathcal{T}_{l,\mathrm{unc}}$, because it is given by sums and products of measurable functions (we recall that $\mathbb{Q}_{l,l} \succ \mathbf{0}$), and that all the expectations are finite as discussed in the proof of Lemma 5. Finally, we observe that $\mathfrak{t}_l^{\star\star} \in \mathcal{T}_l$, i.e., $\mathsf{E}[\|\mathfrak{t}_l^{\star\star}\|^2] < \infty$, because from the original problem formulation (4) we notice that $\mathrm{MSE}(\mathfrak{t}_{-l}^\star, \mathfrak{t}_l^{\star\star})$ is given by a sum of non-negative terms, one of which is precisely $\frac{1}{P}\mathsf{E}[\|\mathfrak{t}_l^{\star\star}\|^2]$, and $\mathrm{MSE}(\mathfrak{t}_{-l}^\star, \mathfrak{t}_l^{\star\star}) \leq \mathrm{MSE}(\mathfrak{t}^\star) < \infty$. Therefore, the inequalities in (24) are equalities, and the optimal solution must satisfy $\mathrm{MSE}(\mathfrak{t}^\star) = \mathrm{MSE}(\mathfrak{t}_{-l}^\star, \mathfrak{t}_l^{\star\star})$, $\forall l \in \mathcal{L}$. This proves that an optimal solution must satisfy the stationarity conditions given by (8). $\qquad\square$

C. Proof of Lemma 1

We use the same notation and definitions as in the proof of Theorem 3 given in Appendix B. The optimality gap can be expressed as follows:

$$\mathrm{MSE}(\mathfrak{t}) - \mathrm{MSE}(\mathfrak{t}^\star) = \|\mathfrak{t} - \mathfrak{t}_0\|_{\mathcal{H}}^2 - \|\mathfrak{t}^\star - \mathfrak{t}_0\|_{\mathcal{H}}^2$$

$$\begin{aligned}
&\overset{(a)}{=} \|\mathfrak{t} - \mathfrak{t}^\star\|_{\mathcal{H}}^2 \\
&\overset{(b)}{=} \langle\mathfrak{t} - \mathfrak{t}^\star, \mathfrak{t} - \mathfrak{t}^\star\rangle + \langle\mathfrak{t}^\star - \mathfrak{t}_0, \mathfrak{t} - \mathfrak{t}^\star\rangle \\
&= \langle\mathfrak{t} - \mathfrak{t}_0, \mathfrak{t} - \mathfrak{t}^\star\rangle \\
&\overset{(c)}{=} \mathsf{E}[(\mathfrak{t} - \mathfrak{t}^\star)^{\mathsf{H}}\mathbb{z}],
\end{aligned}$$

where (a) follows from Pythagoras' theorem, (b) from the orthogonality condition $\langle\mathfrak{t}^\star - \mathfrak{t}_0, \mathbb{a}\rangle = 0, \forall \mathbb{a} \in \mathcal{T}$ and $\mathfrak{t} - \mathfrak{t}^\star \in \mathcal{T}$, and (c) by applying the law of total expectation as in (23). Then, the proof follows from

$$\begin{aligned}
\|\mathfrak{t} - \mathfrak{t}^\star\|_{\mathcal{H}}^2 &= \mathsf{E}[(\mathfrak{t} - \mathfrak{t}^\star)^{\mathsf{H}}\mathbb{z}] \\
&= \langle\mathbb{Q}^{-1}\mathbb{z}, \mathfrak{t} - \mathfrak{t}^\star\rangle \\
&\leq \|\mathbb{Q}^{-1}\mathbb{z}\|_{\mathcal{H}}\|\mathfrak{t} - \mathfrak{t}^\star\|_{\mathcal{H}},
\end{aligned}$$

where the last step is the Cauchy–Schwarz inequality, and where we use $\mathbb{Q}^{-1} \preceq P\mathbf{I}$ which ensures $\|\mathbb{Q}^{-1}\mathbb{z}\|_{\mathcal{H}}^2 = \mathsf{E}[\mathbb{z}^{\mathsf{H}}\mathbb{Q}^{-1}\mathbb{z}] \leq P\mathsf{E}[\mathbb{z}^{\mathsf{H}}\mathbb{z}] < \infty$.

D. Proof of Theorem 4 (Additional Details)

We rearrange the system at hand as $(\mathbf{D} + \mathbf{U}\mathbf{\Pi}^{\mathsf{T}})\mathbf{C} = \mathbf{U}$, where $\mathbf{C} := \begin{bmatrix}\mathbf{C}_1 \ldots \mathbf{C}_L\end{bmatrix}^{\mathsf{T}}$, $\mathbf{\Pi} := \begin{bmatrix}\mathbf{\Pi}_1 \ldots \mathbf{\Pi}_L\end{bmatrix}^{\mathsf{T}}$, $\mathbf{U} := \begin{bmatrix}\mathbf{I}_K \ldots \mathbf{I}_K\end{bmatrix}^{\mathsf{T}}$, $\mathbf{D} := \mathrm{diag}(\mathbf{I} - \mathbf{\Pi}_1, \ldots, \mathbf{I} - \mathbf{\Pi}_L)$. The proof follows if $\mathbf{D} + \mathbf{U}\mathbf{\Pi}^{\mathsf{T}}$ is invertible, giving the optimal coefficients $\mathbf{C} = (\mathbf{D} + \mathbf{U}\mathbf{\Pi}^{\mathsf{T}})^{-1}\mathbf{U}$. By Lemma 7 given in Appendix G, $\mathbf{D} + \mathbf{U}\mathbf{\Pi}^{\mathsf{T}}$ is invertible if both $\mathbf{D}$ and $\mathbf{D}^{-1} + \mathbf{\Pi}^{\mathsf{T}}\mathbf{U} = \mathbf{D}^{-1} + \sum_l \mathbf{\Pi}_l$ are invertible. Standard arguments show that $\mathbf{0} \preceq \mathbf{\Pi}_l \prec \mathbf{I}$. Therefore, $\mathbf{D}$ is Hermitian positive definite (hence invertible), and so is $\mathbf{D}^{-1} + \sum_l \mathbf{\Pi}_l$, concluding the proof.

E. Proof of Theorem 5 (Additional Details)

Assume $\mathbf{0} \preceq \mathbf{\Pi}_l \prec \mathbf{I}$ for a fixed $l \in \mathcal{L}$. Then, let $\mathbb{Q}_l := \hat{\mathbb{H}}_l^{\mathsf{H}}\hat{\mathbb{H}}_l + \mathbf{\Sigma}_l + P^{-1}\mathbf{I}$ and observe that this implies $\mathbb{Q}_l - \hat{\mathbb{H}}_l^{\mathsf{H}}\mathbf{\Pi}_l\hat{\mathbb{H}}_l = \hat{\mathbb{H}}_l^{\mathsf{H}}(\mathbf{I} - \mathbf{\Pi}_l)\hat{\mathbb{H}}_l + \mathbf{\Sigma}_l + P^{-1}\mathbf{I} \succ \mathbf{0}$. Therefore, we obtain

$$\begin{aligned}
&\hat{\mathbb{H}}_l\left(\mathbb{Q}_l - \hat{\mathbb{H}}_l^{\mathsf{H}}\mathbf{\Pi}_l\hat{\mathbb{H}}_l\right)^{-1}\hat{\mathbb{H}}_l^{\mathsf{H}} \\
&= \hat{\mathbb{H}}_l\left[\mathbb{Q}_l\left(\mathbf{I} - \mathbb{F}_l\mathbf{\Pi}_l\hat{\mathbb{H}}_l\right)\right]^{-1}\hat{\mathbb{H}}_l^{\mathsf{H}} \\
&\overset{(a)}{=} \hat{\mathbb{H}}_l\left(\mathbf{I} - \mathbb{F}_l\mathbf{\Pi}_l\hat{\mathbb{H}}_l\right)^{-1}\mathbb{F}_l \\
&\overset{(b)}{=} \mathbb{P}_l\left(\mathbf{I} - \mathbf{\Pi}_l\mathbb{P}_l\right)^{-1}
\end{aligned}$$

where (a) and (b) follow from Lemma 8 and Lemma 9 given in Appendix G, respectively. These lemmas ensure that all the above inverses exist, and in particular $(\mathbf{I} - \mathbf{\Pi}_l\mathbb{P}_l)^{-1}$. Furthermore, the above chain of equalities also show that

$$\begin{aligned}
\mathbf{\Pi}_{l-1} &= \mathsf{E}[\mathbb{P}_l\mathbb{V}_l] + \mathbf{\Pi}_l\mathsf{E}[\bar{\mathbb{V}}_l] \\
&= \mathbf{\Pi}_l + (\mathbf{I} - \mathbf{\Pi}_l)\mathsf{E}[\mathbb{P}_l\left(\mathbf{I} - \mathbf{\Pi}_l\mathbb{P}_l\right)^{-1}](\mathbf{I} - \mathbf{\Pi}_l) \\
&= \mathbf{\Pi}_l + (\mathbf{I} - \mathbf{\Pi}_l)^{\frac{1}{2}}\mathsf{E}[\tilde{\mathbb{P}}_l](\mathbf{I} - \mathbf{\Pi}_l)^{\frac{1}{2}},
\end{aligned}$$

where $\tilde{\mathbb{P}}_l := \tilde{\mathbb{H}}_l\left(\tilde{\mathbb{H}}_l^{\mathsf{H}}\tilde{\mathbb{H}}_l + \mathbf{\Sigma}_l + P^{-1}\mathbf{I}\right)^{-1}\tilde{\mathbb{H}}_l^{\mathsf{H}}$, and $\tilde{\mathbb{H}}_l := (\mathbf{I} - \mathbf{\Pi}_l)^{\frac{1}{2}}\hat{\mathbb{H}}_l$. By standard argument, it can be shown that $\mathbf{0} \preceq \tilde{\mathbb{P}}_l \prec \mathbf{I}$ holds, and hence $\mathbf{0} \preceq \mathbf{\Pi}_{l-1} \prec \mathbf{I}$. Overall, the above discussion proves that $\mathbf{0} \preceq \mathbf{\Pi}_l \prec \mathbf{I}$ implies the existence

of $(\mathbf{I} - \mathbf{\Pi}_l \mathbb{P}_l)^{-1}$ and that $\mathbf{0} \preceq \mathbf{\Pi}_{l-1} \prec \mathbf{I}$. By finally observing that $\mathbf{0} \preceq \mathbb{P}_{L-1} \prec \mathbf{I}$ and hence $\mathbf{0} \preceq \mathbf{\Pi}_{L-1} = \mathsf{E}[\mathbb{P}_{L-1}] \prec \mathbf{I}$, the proof is concluded by repeating the previous argument recursively.

F. Proof of Lemma 3

We recall the following results from random matrix theory, provided without proof: for $\mathrm{vec}(\mathbb{H}_l) \sim \mathcal{CN}(\mathbf{0}, \mathbf{I})$, we have $\mathsf{E}[\mathbb{H}_l (\mathbb{H}_l^{\mathsf{H}} \mathbb{H}_l)^{-1} \mathbb{H}_l^{\mathsf{H}}] = \frac{N}{K} \mathbf{I}$ and $\mathsf{E}[\mathbb{H}_l (\mathbb{H}_l^{\mathsf{H}} \mathbb{H}_l)^{-2} \mathbb{H}_l^{\mathsf{H}}] = \frac{N}{K(K-N)} \mathbf{I}$. We define the projection matrix $\mathbb{P}_l := \mathbb{H}_l (\mathbb{H}_l^{\mathsf{H}} \mathbb{H}_l)^{-1} \mathbb{H}_l^{\mathsf{H}}$ onto $\mathrm{span}(\mathbb{H}_l)$, the projection matrix $\mathbb{P}_l^{\perp} := \mathbf{I} - \mathbb{P}_l$ onto its orthogonal complement, and let $\mathbb{t}_k = (\mathbb{t}_{1,k}, \dots, \mathbb{t}_{L,k})$ as in (19). A simple recursive calculation shows the identity $\mathbf{e}_k - \sum_{j=1}^l \mathbb{H}_j \mathbb{t}_{j,k} = \mathbb{P}_l^{\perp} \mathbb{P}_{l-1}^{\perp} \dots \mathbb{P}_1^{\perp} \mathbf{e}_k$. The first part of the objective in (4) is then given by

$$
\begin{aligned}
&\mathsf{E}\left[\| \mathbb{H} \mathbb{t}_k - \mathbf{e}_k \|^2 \right] \\
&= \mathbf{e}_k^{\mathsf{H}} \mathsf{E}\left[\mathbb{P}_1^{\perp}, \dots, \mathbb{P}_{L-1}^{\perp} \mathbb{P}_L^{\perp} \mathbb{P}_L^{\perp} \mathbb{P}_{L-1}^{\perp} \dots \mathbb{P}_1^{\perp} \right] \mathbf{e}_k \\
&= \left(1 - \frac{N}{K} \right) \mathbf{e}_k^{\mathsf{H}} \mathsf{E}\left[\mathbb{P}_1^{\perp}, \dots, \mathbb{P}_{L-1}^{\perp} \mathbb{P}_{L-1}^{\perp} \dots \mathbb{P}_1^{\perp} \right] \mathbf{e}_k \\
&= \left(1 - \frac{N}{K} \right)^L,
\end{aligned}
$$

where we used the Hermitian symmetry and idempotency of projection matrices, and the independence between $\mathbb{H}_l$ and $\{\mathbb{H}_j\}_{j \neq l}$. We now measure the suboptimality of $\mathbb{t}_k$ by using Lemma 1, specialized to the current setting similarly to Lemma 2. We have:

$$
\begin{aligned}
&\mathbb{z}_{l,k}(S_l) \\
&= \left(\mathbb{H}_l^{\mathsf{H}} \mathbb{H}_l + P^{-1} \mathbf{I} \right) \mathbb{t}_{l,k}(S_l) \\
&\quad + \mathbb{H}_l^{\mathsf{H}} \left(\sum_{j \neq l} \mathsf{E}\left[\mathbb{H}_j \mathbb{t}_{j,k} \Big| S_l \right] - \mathbf{e}_k \right) \\
&= P^{-1} \mathbb{t}_{l,k}(S_l) - \mathbb{H}_l^{\mathsf{H}} \mathsf{E}\left[\mathbb{P}_L^{\perp} \mathbb{P}_{L-1}^{\perp} \dots \mathbb{P}_1^{\perp} \mathbf{e}_k \Big| S_l \right] \\
&= P^{-1} \mathbb{t}_{l,k}(S_l) - \left(1 - \frac{N}{K} \right)^{L-l+1} \mathbb{H}_l^{\mathsf{H}} \mathbb{P}_l^{\perp} \dots \mathbb{P}_1^{\perp} \mathbf{e}_k \\
&= P^{-1} \mathbb{t}_{l,k}(S_l),
\end{aligned}
$$

where the last step follows from the definition of projection matrices, which gives $\mathbb{H}_l^{\mathsf{H}} \mathbb{P}_l^{\perp} \mathbf{b} = \mathbf{0}$ for any $\mathbf{b} \in \mathbb{C}^K$. Furthermore, we have

$$
\begin{aligned}
&\mathsf{E}\left[\| \mathbb{t}_{l,k} \|^2 \right] \\
&= \mathsf{E}\left[\left(\mathbf{e}_k - \sum_{j=1}^{l-1} \mathbb{H}_j \mathbb{t}_{j,k} \right)^{\mathsf{H}} \mathbb{H}_l (\mathbb{H}_l^{\mathsf{H}} \mathbb{H}_l)^{-2} \mathbb{H}_l^{\mathsf{H}} \right. \\
&\quad \left. \times \left(\mathbf{e}_k - \sum_{j=1}^{l-1} \mathbb{H}_j \mathbb{t}_{j,k} \right) \right] \\
&= \frac{N}{K(N-K)} \left(1 - \frac{N}{K} \right)^{l-1} < \infty.
\end{aligned}
$$

Therefore, Lemma 1 applies and, by using the looser bound in (10), we readily obtain $\mathrm{MSE}_k(\mathbb{t}_k) - \mathrm{MSE}_k(\mathbb{t}_k^{\star}) \leq \frac{1}{P} \sum_{l=1}^L \mathsf{E}\left[\| \mathbb{t}_{l,k} \|^2 \right] \xrightarrow[P \to \infty]{} 0$, and $\mathrm{MSE}_k(\mathbb{t}_k^{\star}) \geq \left(1 - \frac{N}{K} \right)^L$.

G. Linear Algebra Background

Lemma 7 (Woodbury Matrix Identity): Let $\mathbf{A} \in \mathbb{C}^{n \times n}$, $\mathbf{B} \in \mathbb{C}^{n \times m}$, $\mathbf{C} \in \mathbb{C}^{m \times n}$, *and* $\mathbf{D} \in \mathbb{C}^{m \times m}$. *If* $\mathbf{A}$, $\mathbf{C}$, *and* $\mathbf{D}^{-1} + \mathbf{C} \mathbf{A}^{-1} \mathbf{B}$ *are invertible, then* $\mathbf{A} + \mathbf{B} \mathbf{D} \mathbf{C}$ *is invertible and* $(\mathbf{A} + \mathbf{B} \mathbf{D} \mathbf{C})^{-1} = \mathbf{A}^{-1} - \mathbf{A}^{-1} \mathbf{B} \left(\mathbf{D}^{-1} + \mathbf{C} \mathbf{A}^{-1} \mathbf{B} \right)^{-1} \mathbf{C} \mathbf{A}^{-1}$.

Lemma 8 (Inverse of Product): Let $\mathbf{A}$ *and* $\mathbf{B}$ *be two square matrices of the same dimension. If* $\mathbf{A} \mathbf{B}$ *is invertible, then* $\mathbf{A}$ *and* $\mathbf{B}$ *are also invertible, and* $(\mathbf{A} \mathbf{B})^{-1} = \mathbf{B}^{-1} \mathbf{A}^{-1}$.

Lemma 9 (Push-Through Identity): Let $\mathbf{A} \in \mathbb{C}^{n \times m}$ *and* $\mathbf{B} \in \mathbb{C}^{m \times n}$ *be two matrices such that* $\mathbf{I} + \mathbf{A} \mathbf{B}$ *is invertible. Then,* $\mathbf{I} + \mathbf{B} \mathbf{A}$ *is also invertible, and* $\mathbf{B} \left(\mathbf{I} + \mathbf{A} \mathbf{B} \right)^{-1} = (\mathbf{I} + \mathbf{B} \mathbf{A})^{-1} \mathbf{B}$.

REFERENCES

[1] L. Miretti, E. Bjornson, and D. Gesbert, "Precoding for scalable cell-free massive MIMO with radio stripes," in *Proc. IEEE 22nd Int. Workshop Signal Process. Adv. Wireless Commun. (SPAWC)*, Sep. 2021, pp. 411–415.

[2] D. Gesbert, S. V. 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, Sep. 2010.

[3] S. Shamai (Shitz) and B. M. Zaidel, "Enhancing the cellular downlink capacity via co-processing at the transmitting end," in *Proc. IEEE VTS 53rd Veh. Technol. Conf.*, vol. 3, May 2001, pp. 1745–1749.

[4] S. Venkatesan, A. Lozano, and R. Valenzuela, "Network MIMO: Overcoming intercell interference in indoor wireless systems," in *Proc. Conf. Rec. 41st Asilomar Conf. Signals, Syst. Comput.*, Nov. 2007, pp. 83–87.

[5] A. Checko, H. L. Christiansen, Y. Yan, L. Scolari, G. Kardaras, M. S. Berger, and L. Dittmann, "Cloud RAN for mobile networks—A technology overview," *IEEE Commun. Surveys Tuts.*, vol. 17, no. 1, pp. 405–426, 1st Quart., 2015.

[6] J. Zhang, E. Björnson, M. Matthaiou, D. W. K. Ng, H. Yang, and D. J. Love, "Prospective multiple antenna technologies for beyond 5G," *IEEE J. Sel. Areas Commun.*, vol. 38, no. 8, pp. 1637–1660, Aug. 2020.

[7] P. D. Kerret and D. Gesbert, "Degrees of freedom of the network MIMO channel with distributed CSI," *IEEE Trans. Inf. Theory*, vol. 58, no. 11, pp. 6806–6824, Nov. 2012.

[8] D. Gesbert and P. de Kerret, "Team methods for device cooperation in wireless networks," in *Cooperative and Graph Signal Processing*. Amsterdam, The Netherlands: Elsevier, 2018, pp. 469–487.

[9] A. Bazco-Nogueras, P. de Kerret, D. Gesbert, and N. Gresset, "On the degrees-of-freedom of the K-user distributed broadcast channel," *IEEE Trans. Inf. Theory*, vol. 66, no. 9, pp. 5642–5659, Sep. 2020.

[10] 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.

[11] 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, Jul. 2017.

[12] 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, Jun. 2021.

[13] H. Huang, M. Trivellato, A. Hottinen, M. Shafi, P. J. Smith, and R. Valenzuela, "Increasing downlink cellular throughput with limited network MIMO coordination," *IEEE Trans. Wireless Commun.*, vol. 8, no. 6, pp. 2983–2989, Jun. 2009.

[14] J. Zhang, R. Chen, J. G. Andrews, A. Ghosh, and R. W. Heath, Jr., "Networked MIMO with clustered linear precoding," *IEEE Trans. Wireless Commun.*, vol. 8, no. 4, pp. 1910–1921, Apr. 2009.

[15] E. Björnson and E. Jorswieck, *Optimal Resource Allocation in Coordinated Multi-Cell Systems*. Delft, The Netherlands: Now Publishers, 2013.

[16] 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.

[17] G. Interdonato, E. Björnson, H. Q. Ngo, P. Frenger, and E. G. Larsson, "Ubiquitous cell-free massive MIMO communications," *EURASIP J. Wireless Commun. Netw.*, vol. 2019, p. 197, Dec. 2019.

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

[19] V. R. Cadambe and S. A. Jafar, "Interference alignment and degrees of freedom of the K-user interference channel," *IEEE Trans. Inf. Theory*, vol. 54, no. 8, pp. 3425–3441, Aug. 2008.

[20] L. Miretti, M. Kobayashi, D. Gesbert, and P. de Kerret, "Cooperative multiple-access channels with distributed state information," *IEEE Trans. Inf. Theory*, vol. 67, no. 8, pp. 5185–5199, Aug. 2021.

[21] T. L. Marzetta, E. G. Larsson, H. Yang, and H. Q. Ngo, *Fundamentals of Massive MIMO*. Cambridge, U.K.: Cambridge Univ. Press, 2016.

[22] E. Björnson, J. Hoydis, and L. Sanguinetti, "Massive MIMO networks: Spectral, energy, and hardware efficiency," *Found. Trends Signal Process.*, vol. 11, nos. 3–4, pp. 154–655, 2017.

[23] J. Marschak and R. Radner, *Economic Theory of Teams*. New Haven, CT, USA: Yale Univ. Press, 1972.

[24] R. Radner, "Team decision problems," *Ann. Math. Statist.*, vol. 33, no. 3, pp. 857–881, 1962.

[25] S. Yüksel and T. Başar, *Stochastic Networked Control Systems: Stabilization and Optimization Under Information Constraints*. New York, NY, USA: Springer, 2013.

[26] D. Neumann, T. Wiese, M. Joham, and W. Utschick, "A bilinear equalizer for massive MIMO systems," *IEEE Trans. Signal Process.*, vol. 66, no. 14, pp. 3740–3751, Jul. 2018.

[27] Z. H. Shaik, E. Bjornson, and E. G. Larsson, "MMSE-optimal sequential processing for cell-free massive MIMO with radio stripes," *IEEE Trans. Commun.*, vol. 69, no. 11, pp. 7775–7789, Nov. 2021.

[28] E. Björnson and L. Sanguinetti, "Making cell-free massive MIMO competitive with MMSE processing and centralized implementation," *IEEE Trans. Wireless Commun.*, vol. 19, no. 1, pp. 77–90, Jan. 2020.

[29] J. Rodriguez Sanchez, F. Rusek, O. Edfors, M. Sarajlic, and L. Liu, "Decentralized massive MIMO processing exploring daisy-chain architecture and recursive algorithms," *IEEE Trans. Signal Process.*, vol. 68, pp. 687–700, 2020.

[30] D. Tse and P. Viswanath, *Fundamentals of Wireless Communication*. Cambridge, U.K.: Cambridge Univ. Press, 2005.

[31] E. Biglieri, J. Proakis, and S. Shamai (Shitz), "Fading channels: Information-theoretic and communications aspects," *IEEE Trans. Inf. Theory*, vol. 44, no. 6, pp. 2619–2692, Oct. 1998.

[32] 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.

[33] S. S. Christensen, R. Agarwal, E. De Carvalho, and J. M. Cioffi, "Weighted sum-rate maximization using weighted MMSE for MIMO-BC beamforming design," *IEEE Trans. Wireless Commun.*, vol. 7, no. 12, pp. 4792–4799, Dec. 2008.

[34] G. Wahba, *Spline Models for Observational Data*. Philadelphia, PA, USA: SIAM, 1990.

[35] *Further Advancements for E-UTRA Physical Layer Aspects (Release 9)* Tech. Rep. TS 36.814, 3GPP, Mar. 2017.

[36] A. El Gamal and Y.-H. Kim, *Network Information Theory*. Cambridge, U.K.: Cambridge Univ. Press, 2011.

[37] D. G. Luenberger, *Optimization by Vector Space Methods*. Hoboken, NJ, USA: Wiley, 1997.

Lorenzo Miretti (Member, IEEE) received the B.Sc. and M.Sc. degrees in telecommunication engineering from the Politecnico di Torino, Turin, Italy, in 2015 and 2018, respectively, and the Ph.D. degree in wireless communications from EURECOM, Sophia Antipolis, France, and Sorbonne Université, Paris, France, in 2021. He is currently a Post-Doctoral Researcher with the Technical University of Berlin, and a Visiting Researcher with the Fraunhofer Heinrich Hertz Institute, Berlin, Germany. He investigates innovative solutions for next generation wireless networks, using information and communication theoretic tools.

Emil Björnson (Fellow, IEEE) received the M.S. degree in engineering mathematics from Lund University, Sweden, in 2007, and the Ph.D. degree in telecommunications from the KTH Royal Institute of Technology, Sweden, in 2011.

From 2012 to 2014, he was a Post-Doctoral Researcher at the Alcatel-Lucent Chair on Flexible Radio, SUPELEC, France. From 2014 to 2021, he held different professor positions at Linköping University, Sweden. From 2020 to 2021, he was a part-time Visiting Full Professor at the KTH. Since 2022, he has been a Tenured Full Professor of wireless communication at the KTH. He has authored the textbooks *Optimal Resource Allocation in Coordinated Multi-Cell Systems* (2013), *Massive MIMO Networks: Spectral, Energy, and Hardware Efficiency* (2017), and *Foundations of User-Centric Cell-Free Massive MIMO* (2021). He is dedicated to reproducible research and has made a large amount of simulation code publicly available. He has performed MIMO research for over 15 years, his articles have received more than 17000 citations, and he has filed more than 20 patent applications. He is a Co-Host of the Podcast Wireless Future and has a popular YouTube channel. He performs research on MIMO communications, radio resource allocation, machine learning for communications, and energy efficiency.

Dr. Björnson has been a member of the Online Editorial Team of the IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS since 2020. He is a Digital Futures Fellow and a Wallenberg Academy Fellow. He has received the 2014 Outstanding Young Researcher Award from IEEE ComSoc EMEA, the 2015 Ingvar Carlsson Award, the 2016 Best Ph.D. Award from EURASIP, the 2018 IEEE Marconi Prize Paper Award in Wireless Communications, the 2019 EURASIP Early Career Award, the 2019 IEEE Communications Society Fred W. Ellersick Prize, the 2019 *IEEE Signal Processing Magazine* Best Column Award, the 2020 Pierre-Simon Laplace Early Career Technical Achievement Award, the 2020 CTTC Early Achievement Award, and the 2021 IEEE ComSoc RCC Early Achievement Award. He also coauthored articles that received Best Paper Awards at the conferences, including WCSP 2009, the IEEE CAMSAP 2011, the IEEE SAM 2014, the IEEE WCNC 2014, the IEEE ICC 2015, and WCSP 2017. He has been on the Editorial Board of the IEEE TRANSACTIONS ON COMMUNICATIONS since 2017. He has been an Area Editor of *IEEE Signal Processing Magazine* since 2021.

David Gesbert (Fellow, IEEE) received the Ph.D. degree from the Ecole Nationale Superieure des Telecommunications, France, in 1997. From 1997 to 1999, he was with the Information Systems Laboratory, Stanford University. He was the Founding Engineer of Iospan Wireless Inc., a Stanford spin off pioneering MIMO-OFDM (currently Intel). Before joining EURECOM in 2004, he was with the Department of Informatics, University of Oslo, as an Adjunct Professor. He is currently working as the Director of EURECOM, France. He has published about 340 articles and 25 patents, seven of them winning IEEE Best paper awards. He is a Board Member for the OpenAirInterface (OAI) Software Alliance. In 2015, he has been awarded an ERC Advanced Grant. In 2020, he was awarded funding by the French Interdisciplinary Institute on Artificial Intelligence for a Chair in the area of AI for the future IoT. In 2021, he received the Grand Prix in Research from IMT-French Academy of Sciences. He has been the Technical Program Co-Chair of ICC2017 and has been named a Thomson-Reuters Highly Cited Researchers in computer science.