



<http://www.diva-portal.org>

Postprint

This is the accepted version of a paper published in *IEEE Transactions on Vehicular Technology*. This paper has been peer-reviewed but does not include the final publisher proof-corrections or journal pagination.

Citation for the original published paper (version of record):

Brandt, R., Bengtsson, M. (2015)

Distributed CSI Acquisition and Coordinated Precoding for TDD Multicell MIMO Systems.

IEEE Transactions on Vehicular Technology

<http://dx.doi.org/10.1109/TVT.2015.2432051>

Access to the published version may require subscription.

N.B. When citing this work, cite the original published paper.

Permanent link to this version:

<http://urn.kb.se/resolve?urn=urn:nbn:se:kth:diva-166426>

Distributed CSI Acquisition and Coordinated Precoding for TDD Multicell MIMO Systems

Rasmus Brandt*, *Student Member, IEEE*, and Mats Bengtsson, *Senior Member, IEEE*

Abstract—Several distributed coordinated precoding methods exist in the downlink multicell MIMO literature, many of which assume perfect knowledge of received signal covariance and local effective channels. In this work, we let the notion of channel state information (CSI) encompass this knowledge of covariances and effective channels. We analyze what local CSI is required in the WMMSE algorithm for distributed coordinated precoding, and study how this required CSI can be obtained in a distributed fashion. Based on pilot-assisted channel estimation, we propose three CSI acquisition methods with different tradeoffs between feedback and signaling, backhaul use, and computational complexity. One of the proposed methods is fully distributed, meaning that it only depends on over-the-air signaling but requires no backhaul, and results in a fully distributed joint system when coupled with the WMMSE algorithm. Naïvely applying the WMMSE algorithm together with the fully distributed CSI acquisition results in catastrophic performance however, and therefore we propose a robustified WMMSE algorithm based on the well known diagonal loading framework. By enforcing properties of the WMMSE solutions with perfect CSI onto the problem with imperfect CSI, the resulting diagonally loaded spatial filters are shown to perform significantly better than the naïve filters. The proposed robust and distributed system is evaluated using numerical simulations, and shown to perform well compared with benchmarks. Under centralized CSI acquisition, the proposed algorithm performs on par with other existing centralized robust WMMSE algorithms. When evaluated in a large scale fading environment, the performance of the proposed system is promising.

I. INTRODUCTION

MULTIPLE-ANTENNA coordinated precoding is a promising technique for improving spectral efficiency in multicell multiple-input multiple-output (MIMO) networks, by serving several spatially separated users simultaneously in the same time/frequency resource block [1], [2]. The cascade of physical channels and precoders are the *effective channels* experienced by the receivers. By suitably selecting the precoders, the downlink weighted sum rate of the network can be maximized. The requirements for practical implementation of coordinated precoding include channel estimation [3]–[5], robustness against channel estimation errors [6]–[10], and sufficiently low complexity; preferably achieved using distributed methods [11]–[14].

In the multicell MIMO literature, there are several examples of distributed coordinated precoding methods; see e.g. [11] and

references therein. These methods typically require information about the received signal covariance and local effective channels at the involved nodes, and it is often assumed that this information is perfectly known. In this work, we denote the information about the received signal covariance and effective channels as *channel state information (CSI)*. We take a systems perspective and propose methods for estimating and acquiring the necessary CSI at the involved nodes in a distributed fashion. The resource allocation is based on the *WMMSE algorithm*¹ [12] for distributed weighted sum rate optimization, because of its low per-iteration complexity and tractable form. Due to poor performance when naïvely applying the WMMSE algorithm, we also propose some robustifying procedures, leading to a robust and fully distributed joint coordinated precoding and CSI acquisition system.

As the first step in the system design, we succinctly describe what information, in terms of weights and CSI, that is needed for the nodes of the network to perform their part in the WMMSE algorithm. There is a multitude of conceivable methods to obtain the necessary information at the nodes, e.g. using various combinations of channel estimation, feedback, signaling, backhaul, etc. We propose three methods for acquiring the necessary CSI. Based on channel estimation through pilot transmissions, feedback, signaling, and backhaul use, the proposed CSI acquisition methods correspond to different tradeoffs between these techniques. In particular, one of the proposed CSI acquisition methods is *fully distributed*, in the sense that the nodes of the network solely cooperate by means of over-the-air signaling, thus requiring no backhaul. A key component of the proposed CSI acquisition methods is the estimation of the effective channels. It is based on synchronous pilot transmission in the downlink, enabling the receiving user equipments (UEs) to estimate both desired and interfering effective channels [4]. Assuming time-division duplex (TDD) operation and perfectly calibrated transceivers [15], [16], similar channel estimation can be performed in the uplink at the receiving base stations (BSs). This is contrary to frequency-division duplex operation, where the BSs obtain their required information by feedback and backhaul signaling. When combining the fully distributed CSI acquisition with the WMMSE algorithm, the joint system is fully distributed.

Naïvely applying the original WMMSE algorithm together with the proposed fully distributed CSI acquisition method leads to catastrophic performance however. This is because the original algorithm was not developed to be robust against

Copyright (c) 2015 IEEE. Personal use of this material is permitted. However, permission to use this material for any other purposes must be obtained from the IEEE by sending a request to pubs-permissions@ieee.org.

The authors are with the Department of Signal Processing, School of Electrical Engineering, KTH Royal Institute of Technology, SE-100 44 Stockholm, Sweden. E-mails: rabr5411@kth.se, mats.bengtsson@ee.kth.se.

¹The algorithm takes this name since it is a Weighted Minimization of the Mean Squared Error (WMMSE).

imperfect CSI. We therefore propose a robustified WMMSE algorithm which *retains the distributedness* of the original algorithm, contrary to state of the art [7]–[10]. We formulate a worst-case WMMSE problem, and solve an upper bounded version of the problem. The resulting precoders are diagonally loaded, a technique which is well known for its robustifying effect on beamformers [17]–[23]. The optimal amount of diagonal loading is determined by the worst-case channel estimation errors, whose statistics are unfortunately unavailable in the proposed CSI acquisition setup. Instead, we propose a practical method for implicitly selecting the amount of diagonal loading for the precoders. At the UEs, we show an inherent property of the (spatial) receive filters and mean squared error (MSE) weights obtained from the WMMSE algorithm with perfect CSI. When this property is enforced onto the filters with imperfect CSI, the resulting receive filters are also diagonally loaded. The robust MSE weights have smaller eigenvalues than the non-robust MSE weights. This can be interpreted as the receivers requesting lower data rates when there are large discrepancies in their estimated CSI.

A. Related Work

In [4], a reciprocal channel was exploited to directly estimate the filters maximizing the signal-to-interference-and-noise ratio, requiring no other signaling. Similar work was performed in [24], where non-linear filters also were studied. Focusing on the reciprocity, and using the receive filters as transmit filters in the uplink, [25] performed extensive simulations for a beam selection approach. Our proposed effective channel estimation is similar to their ‘busy burst’ methodology. For the single-stream multiple-input single-output (MISO) interference channel, an analytical method for finding the rate-maximizing zero-forcing beamformers was derived in [13]. Saliiently, the method does not require cross-link CSI, and was consequently shown to be highly robust against CSI imperfections. In [14], decentralized algorithms based on WMMSE ideas were proposed, achieving faster convergence than the original WMMSE algorithm in [12], in addition to signaling strategies for obtaining the necessary CSI. TDD reciprocity was assumed, and the UEs used combinations of inter-cell and intra-cell effective channel pilot transmissions. Contrary to our work, perfect channel estimation was assumed, and their decentralized algorithms still require some BS backhaul.

Weighted sum rate maximization by means of weighted MMSE minimization was originally proposed for multiuser MIMO systems in [26], where the MSE weights were used to equate the Karush-Kuhn-Tucker (KKT) conditions of the weighted MMSE problem to the KKT conditions of the weighted sum rate problem. This same method was directly applied to multicell MIMO systems in [7], [27], but it was not until [12] that a rigorous connection to the multicell weighted sum rate problem was presented. An earlier work is [6], where the weighted MMSE optimization problem was solved using the same technique, but without explicitly providing the rigorous connection to the weighted sum rate problem. In [6], a robust WMMSE algorithm was also suggested for the case of norm bounded channel uncertainty arising from limited

quantized feedback. Other robustified versions of the WMMSE algorithm, where the contribution of the downlink channel estimation errors in the involved covariance matrices was averaged out, were proposed in [7], [8]. The same approach was taken in [9], where it was mentioned that this corresponds to optimizing a lower bound on the achieved performance, and in [10] where the lower bound was explicitly derived. The filters were in effect robustified by diagonal loading, where the diagonal loading factors were determined by the downlink channel estimation performance. The work in [7]–[10] was mainly focused on proposing robust WMMSE methods and thus the actual CSI acquisition was not conclusively studied, contrary to this paper. The major assumption in the system model of [7]–[10] is that downlink channel estimation is performed at the UEs, and that the downlink channel estimates are fed back to the BSs. In this work we are interested in TDD channel estimation and although the algorithms in [7]–[10] could be applied in such a setting, doing so leads to some idiosyncrasies that will be detailed in Sec. IV-E. Due to the system model in [7]–[10], the nodes of the network require feedback of all filters in all iterations of the algorithm, leading to a large amount of feedback which would typically be implemented using a centralized CSI acquisition infrastructure. In this paper, contrary to [7]–[10], we incorporate a detailed analysis of the CSI acquisition component of the system, leading up to a *robust* and *distributed* coordinated precoding system.

B. Contributions

The major contributions of this work are as follows.

- We succinctly describe the required information for the network nodes to perform one WMMSE iteration. We propose three CSI acquisition methods which provide the necessary information. The methods have varying levels of distributedness and signaling needs. One of the proposed methods is fully distributed, meaning that it can be implemented entirely by over-the-air signaling.
- For resilient performance against channel estimation errors, we propose a robustified, but still distributed, WMMSE algorithm to be applied together with the proposed CSI acquisition schemes. The robustness is due to diagonal loading, and the level of diagonal loading for the precoders is determined implicitly by a practical procedure.
- We identify and explore new inherent properties of the WMMSE algorithm. When the properties are explicitly enforced onto solutions with imperfect CSI, the resulting receive filters are diagonally loaded.
- Performance is evaluated numerically, and it is shown that the proposed fully distributed system performs excellently compared with the naïve WMMSE algorithm with fully distributed CSI acquisition. With centralized CSI acquisition, the proposed robust WMMSE algorithm performs on par with existing robust WMMSE algorithms, which however require centralized CSI acquisition.

C. Notation

The operations $(\cdot)^*$, $(\cdot)^H$, $(\cdot)^T$ are complex conjugate, Hermitian transpose, and regular transpose, respectively. The operators $\text{Tr}(\cdot)$, $\|\cdot\|_2$, $\|\cdot\|_F$ are the matrix trace, Euclidean norm and Frobenius matrix norm, respectively. We denote the partial ordering of positive (negative) semidefinite matrices as \succeq (\preceq). The m th largest eigenvalue (singular value) of \mathbf{Q} is denoted $\lambda_m(\mathbf{Q})$ ($s_m(\mathbf{Q})$). The zero-mean and covariance \mathbf{Q} complex symmetric Gaussian distribution is $\mathcal{CN}(\mathbf{0}, \mathbf{Q})$, and $\mathbb{E}(\cdot)$ denotes expectation. Estimated quantities are denoted with a hat \hat{a} and uplink quantities with an arrow \overleftarrow{a} . The Kronecker delta is $\delta_{i,j}$.

II. DISTRIBUTED WEIGHTED SUM RATE OPTIMIZATION FOR THE MULTICELL MIMO DOWNLINK

Our system model is a multicell system with K_t BSs, each serving K_c UEs, for a total of $K_r = K_t K_c$ UEs. We index the BSs as $i \in \{1, \dots, K_t\}$. The k th served UE of BS i is indexed by the pair of indices (i, k) . For compactness, we will often write this pair of indices as i_k . The system is operating using *coordinated precoding*, i.e. each UE is only served data from one BS and the signals from the other BSs constitute inter-cell interference². When $K_c \geq 2$, intra-cell interference is also observed. The BSs are equipped with M_t antennas each, the UEs have M_r antennas and are served N_d data streams each³.

Communication takes place both in the downlink and in the uplink. We focus on optimizing performance in the downlink, since that typically experiences heavier traffic loads than the uplink. The presented method could equally well be applied in the uplink however. In the downlink, the multiuser interaction is described by the interfering broadcast channel. Denote a realization of the flat-fading MIMO channel between BS j and UE i_k as $\mathbf{H}_{i_k j}$ and let each user's data signal $\mathbf{x}_{i_k} \sim \mathcal{CN}(\mathbf{0}, \mathbf{I}_{N_d})$ be linearly precoded by $\mathbf{V}_{i_k} \in \mathbb{C}^{M_t \times N_d}$. The received signal at UE i_k is then

$$\mathbf{y}_{i_k} = \mathbf{H}_{i_k i} \mathbf{V}_{i_k} \mathbf{x}_{i_k} + \sum_{(j,l) \neq (i,k)} \mathbf{H}_{i_k j} \mathbf{V}_{j l} \mathbf{x}_{j l} + \mathbf{z}_{i_k}, \quad (1)$$

where the last term is a white Gaussian noise term $\mathbf{z}_{i_k} \sim \mathcal{CN}(\mathbf{0}, \sigma_r^2 \mathbf{I}_{M_r})$. The signals $\{\mathbf{x}_{i_k}\}$ and $\{\mathbf{z}_{i_k}\}$ are i.i.d. over users. Given these assumptions, the received interference plus noise covariance matrix for UE i_k is $\Phi_{i_k}^{\text{in}} = \sum_{(j,l) \neq (i,k)} \mathbf{H}_{i_k j} \mathbf{V}_{j l} \mathbf{V}_{j l}^H \mathbf{H}_{i_k j}^H + \sigma_r^2 \mathbf{I}$.

Assuming that the decoders in the UE terminals treat interference as additive noise, the achievable downlink data rate for UE i_k is

$$R_{i_k} = \log \det \left(\mathbf{I} + \mathbf{V}_{i_k}^H \mathbf{H}_{i_k i}^H (\Phi_{i_k}^{\text{in}})^{-1} \mathbf{H}_{i_k i} \mathbf{V}_{i_k} \right). \quad (2)$$

Note that (2) is non-convex in $\{\mathbf{V}_{i_k}\}$, since the precoders appear inside $\Phi_{i_k}^{\text{in}}$. This non-convex dependence on the precoders describes the coupling between users, and will be the key challenge in the optimization to come.

²Since the focus of this paper is the distributed implementation of multicell processing, we do not investigate *joint transmission*, where several BSs jointly serve the UEs with data. Such joint transmission requires significant backhaul between BSs, and is not amenable to fully distributed implementation.

³The system model can easily be extended to scenarios where the BSs serve different number of UEs each and scenarios where the nodes have different number of antennas.

One main assumption in this work is that there is a perfectly reciprocal uplink channel available. That is, the channel in the uplink from UE j_l to BS i is $\overleftarrow{\mathbf{H}}_{j_l i} = \mathbf{H}_{j_l i}^T$. Let $\overleftarrow{\mathbf{x}}_{i_k}^* \sim \mathcal{CN}(\mathbf{0}, \overleftarrow{\Sigma}_{i_k}^*)$ be the transmitted signal from UE i_k in the uplink. The uplink is described by the interfering multiple access channel, and the received signal for BS i is then

$$\overleftarrow{\mathbf{y}}_i^* = \sum_{k=1}^{K_c} \mathbf{H}_{i_k i}^T \overleftarrow{\mathbf{x}}_{i_k}^* + \sum_{j \neq i} \sum_{l=1}^{K_c} \mathbf{H}_{j l i}^T \overleftarrow{\mathbf{x}}_{j l}^* + \overleftarrow{\mathbf{z}}_i^*, \quad (3)$$

where $\overleftarrow{\mathbf{z}}_i^* \sim \mathcal{CN}(\mathbf{0}, \sigma_t^2 \mathbf{I}_{M_t})$. For convenience, we work with the complex conjugate version of the received signal. That is, the model we will use for the uplink is:

$$\overleftarrow{\mathbf{y}}_i = (\overleftarrow{\mathbf{y}}_i^*)^* = \sum_{k=1}^{K_c} \mathbf{H}_{i_k i}^H \overleftarrow{\mathbf{x}}_{i_k} + \sum_{j \neq i} \sum_{l=1}^{K_c} \mathbf{H}_{j l i}^H \overleftarrow{\mathbf{x}}_{j l} + \overleftarrow{\mathbf{z}}_i. \quad (4)$$

With the uplink model in (4), the channel estimation in Sec. III can be tailored to the needs of the weighted sum rate optimization, which we detail in the next section.

A. Weighted Sum Rate Optimization

Since the CSI acquisition to be proposed is tailored for the WMMSE algorithm [12], we now briefly summarize the algorithm, as well as introduce some necessary notation.

By assigning the UEs data rate weights $\alpha_{i_k} \in [0, 1]$, $\forall i_k$, the weighted sum rate is formulated as $\sum_{(i,k)} \alpha_{i_k} R_{i_k}$. This formulation describes the ultimate performance of the system, but is just one way of forming a system-level utility from the user rates [2]. The data rate weights α_{i_k} can be selected corresponding to user priority, e.g. to achieve a proportionally fair solution [28]. In the following, we will assume that the weights are selected at the BSs.

Let P_i be the sum power constraint for BS i . With the precoders $\{\mathbf{V}_{i_k}\}$ as optimization variables, the *weighted sum rate optimization problem* is:

$$\begin{aligned} & \underset{\{\mathbf{V}_{i_k}\}}{\text{maximize}} && \sum_{(i,k)} \alpha_{i_k} R_{i_k} \\ & \text{subject to} && \sum_{k=1}^{K_c} \text{Tr}(\mathbf{V}_{i_k} \mathbf{V}_{i_k}^H) \leq P_i, \quad i = 1, \dots, K_t. \end{aligned} \quad (5)$$

Due to the non-convexity of (2), this is a non-convex optimization problem. At least when $M_r = 1$, the problem is also NP-hard [29]. We can therefore only reasonably strive to find a *locally* optimal solution.

By introducing additional optimization variables $\{\mathbf{W}_{i_k}\}$ (acting like MSE weights), it was shown in [12] that (5) has the same global solutions as the following *weighted MMSE optimization problem*:

$$\begin{aligned} & \underset{\{\mathbf{A}_{i_k}\}, \{\mathbf{V}_{i_k}\}}{\text{minimize}} && \sum_{(i,k)} \alpha_{i_k} (\text{Tr}(\mathbf{W}_{i_k} \mathbf{E}_{i_k}) - \log \det(\mathbf{W}_{i_k})) \\ & \text{subject to} && \sum_{k=1}^{K_c} \text{Tr}(\mathbf{V}_{i_k} \mathbf{V}_{i_k}^H) \leq P_i, \quad i = 1, \dots, K_t. \end{aligned} \quad (6)$$

TABLE I
SUMMARY OF CSI QUANTITY SHORTHANDS

Downlink	$\mathbf{F}_{i_k} = \mathbf{H}_{i_k i} \mathbf{V}_{i_k}$ $\Phi_{i_k} = \mathbf{F}_{i_k} \mathbf{F}_{i_k}^H + \Phi_{i_k}^{i+n}$ $\Phi_{i_k}^{i+n} = \sum_{(j,l) \neq (i,k)} \mathbf{H}_{i_k j} \mathbf{V}_{j l} \mathbf{V}_{j l}^H \mathbf{H}_{i_k j}^H + \sigma_r^2 \mathbf{I}$
Uplink	$\mathbf{G}_{i_k} = \mathbf{H}_{i_k i}^H \mathbf{U}_{i_k}$ $\Gamma_i = \Gamma_i^{s+i} = \sum_{(j,l)} \mathbf{H}_{j l i}^H \mathbf{U}_{j l} \mathbf{U}_{j l}^H \mathbf{H}_{j l i}$

TABLE II
QUANTITIES THAT MUST BE SIGNALLED IN ORDER FOR EACH NODE TO PERFORM ONE ITERATION OF THE WMMSE ALGORITHM

	Covariance matrix	Effective channel(s)	Weight(s) [†]
UE i_k	Φ_{i_k}	\mathbf{F}_{i_k}	α_{i_k}
BS i	Γ_i	$\{\mathbf{G}_{i_k}\}_{k=1}^{K_c}$	$\{\mathbf{W}_{i_k}^{1/2}\}_{k=1}^{K_c}$

[†] Note that the user priorities $\{\alpha_{i_k}\}_{k=1}^{K_c}$ are selected, and thus fully known, at the serving BSs.

The $\{\mathbf{A}_{i_k}\}$ are linear receive filters, and

$$\mathbf{E}_{i_k} = \mathbb{E} \left((\mathbf{x}_{i_k} - \mathbf{A}_{i_k}^H \mathbf{y}_{i_k}) (\mathbf{x}_{i_k} - \mathbf{A}_{i_k}^H \mathbf{y}_{i_k})^H \right) \quad (7)$$

$$= \mathbf{I} - \mathbf{A}_{i_k}^H \mathbf{H}_{i_k i} \mathbf{V}_{i_k} - \mathbf{V}_{i_k}^H \mathbf{H}_{i_k i}^H \mathbf{A}_{i_k} + \mathbf{A}_{i_k}^H \Phi_{i_k} \mathbf{A}_{i_k}$$

is the MSE matrix for UE i_k . Further, $\Phi_{i_k} = \mathbf{H}_{i_k i} \mathbf{V}_{i_k} \mathbf{V}_{i_k}^H \mathbf{H}_{i_k i}^H + \Phi_{i_k}^{i+n}$ is the received signal and interference plus noise covariance matrix for UE i_k .

The optimization problem in (6) is still non-convex over the joint set $\{\mathbf{A}_{i_k}, \mathbf{W}_{i_k}, \mathbf{V}_{i_k}\}$, but the key benefit of (6) over (5) is that it is independently convex in the blocks of variables $\{\mathbf{A}_{i_k}\}$, $\{\mathbf{W}_{i_k}\}$, and $\{\mathbf{V}_{i_k}\}$, when the remaining blocks are kept fixed. Further, a stationary point can be found through alternating minimization⁴ [30, Ch. 2.7] over the blocks [12]. There is a one-to-one correspondence between the stationary points of (5) and the stationary points of (6) [12], and since (6) optimizes a locally tight lower bound of (5), alternating minimization of (6) will also converge to a stationary point of (5) [31].

1) *WMMSE Algorithm for Distributed Weighted Sum Rate Optimization*: First, by fixing $\{\mathbf{W}_{i_k}, \mathbf{V}_{i_k}\}$ in (6), it can easily be shown that the problem decouples over the UEs. The solution is the well known MMSE receiver

$$\mathbf{A}_{i_k} = \Phi_{i_k}^{-1} \mathbf{H}_{i_k i} \mathbf{V}_{i_k}, \quad \forall i_k. \quad (8)$$

Next, by fixing $\{\mathbf{A}_{i_k}, \mathbf{V}_{i_k}\}$, the problem again decouples over the UEs. UE i_k should therefore solve $\min_{\mathbf{W}_{i_k}} \text{Tr}(\mathbf{W}_{i_k} \mathbf{E}_{i_k}) - \log \det(\mathbf{W}_{i_k})$, and the solutions are

$$\mathbf{W}_{i_k} = \mathbf{E}_{i_k}^{-1} = (\mathbf{I} - \mathbf{V}_{i_k}^H \mathbf{H}_{i_k i}^H \Phi_{i_k}^{-1} \mathbf{H}_{i_k i} \mathbf{V}_{i_k})^{-1}, \quad \forall i_k, \quad (9)$$

where the last equality comes from plugging in \mathbf{A}_{i_k} from (8).

Finally, it remains to solve (6) for $\{\mathbf{V}_{i_k}\}$, while keeping the UE variables $\{\mathbf{A}_{i_k}, \mathbf{W}_{i_k}\}$ fixed. The problem decouples over the BSs, and it can be shown that the remaining problem

⁴This technique is also known as *block coordinate descent* or *block nonlinear Gauss-Seidel* in the literature.

Algorithm 1 WMMSE Algorithm [12] (Perfect CSI)

- 1: **repeat**
- At UEs:
- 2: $\mathbf{W}_{i_k} = (\mathbf{I} - \mathbf{F}_{i_k}^H \Phi_{i_k}^{-1} \mathbf{F}_{i_k})^{-1}$
- 3: $\mathbf{A}_{i_k} = \Phi_{i_k}^{-1} \mathbf{F}_{i_k}$, $\mathbf{U}_{i_k} = \sqrt{\alpha_{i_k}} \mathbf{A}_{i_k} \mathbf{W}_{i_k}^{1/2}$
- At BSs:
- 4: Find μ_i which satisfies $\sum_{k=1}^{K_c} \text{Tr}(\mathbf{V}_{i_k} \mathbf{V}_{i_k}^H) \leq P_i$
- 5: $\mathbf{B}_{i_k} = (\Gamma_i + \mu_i \mathbf{I})^{-1} \mathbf{G}_{i_k}$, $\mathbf{V}_{i_k} = \sqrt{\alpha_{i_k}} \mathbf{B}_{i_k} \mathbf{W}_{i_k}^{1/2}$
- 6: **until** convergence criterion met, or fixed number of iters.

for BS i is a quadratically constrained quadratic program with optimization variables $\{\mathbf{V}_{i_k}\}_{k=1}^{K_c}$. The solution is [12]

$$\mathbf{V}_{i_k} = \alpha_{i_k} (\Gamma_i + \mu_i \mathbf{I})^{-1} \mathbf{H}_{i_k i}^H \mathbf{A}_{i_k} \mathbf{W}_{i_k}, \quad \forall i_k, \quad (10)$$

where $\Gamma_i = \Gamma_i^{s+i} = \sum_{(j,l)} \alpha_{j l} \mathbf{H}_{j l i}^H \mathbf{A}_{j l} \mathbf{W}_{j l} \mathbf{A}_{j l}^H \mathbf{H}_{j l i}$ is a signal plus interference covariance matrix for BS i in the uplink. If $\sum_{k=1}^{K_c} \text{Tr}(\mathbf{V}_{i_k} \mathbf{V}_{i_k}^H) \leq P_i$ is satisfied for $\mu_i = 0$, the sum power constraint for BS i is inactive and the problem is solved. Otherwise, $\mu_i > 0$ is found such that $\sum_{k=1}^{K_c} \text{Tr}(\mathbf{V}_{i_k} \mathbf{V}_{i_k}^H) = P_i$ holds. This can be done efficiently using e.g. bisection [12]. When the precoders have been found, a new iteration is commenced by again optimizing over $\{\mathbf{A}_{i_k}\}$. With each update of $\{\mathbf{A}_{i_k}\}$, $\{\mathbf{W}_{i_k}\}$ or $\{\mathbf{V}_{i_k}\}$, the objective value in (6) cannot increase. The iterations thus continue until convergence, or for a fixed number of iterations.

2) *Required Local Information for the WMMSE Iterations*: In order to clarify what information the CSI acquisition schemes should provide, we introduce some shorthands for the quantities involved in the WMMSE algorithm. For UE i_k , we define a *weighted receive filter* as $\mathbf{U}_{i_k} = \sqrt{\alpha_{i_k}} \mathbf{A}_{i_k} \mathbf{W}_{i_k}^{1/2}$ and denote the *effective downlink channel* as $\mathbf{F}_{i_k} = \mathbf{H}_{i_k i} \mathbf{V}_{i_k}$. The *receive filter* can then be written as $\mathbf{A}_{i_k} = \Phi_{i_k}^{-1} \mathbf{F}_{i_k}$. Symmetrically, in the uplink for UE i_k , the *precoder* is $\mathbf{V}_{i_k} = \sqrt{\alpha_{i_k}} \mathbf{B}_{i_k} \mathbf{W}_{i_k}^{1/2}$ and the *effective uplink channel* is $\mathbf{G}_{i_k} = \mathbf{H}_{i_k i}^H \mathbf{U}_{i_k}$. Finally, the *component precoder* is $\mathbf{B}_{i_k} = (\Gamma_i + \mu_i \mathbf{I})^{-1} \mathbf{G}_{i_k}$. We summarize the shorthands in Table I, and the WMMSE algorithm written using these shorthands in Algorithm 1.

The WMMSE algorithm operates in two phases: one in which the UEs form their receive filters and weights, and one in which the BSs form the precoders for their served UEs. The optimization steps at the UEs and BSs are completely decoupled, and as summarized in Table II, the nodes only require local CSI and local weights. Hence, the WMMSE algorithm is an example of distributed resource allocation. In Sec. III, we will describe how the nodes can exploit the channel reciprocity to obtain local CSI in a distributed fashion.

III. DISTRIBUTED CSI ACQUISITION

According to Table II, the UEs require knowledge about the effective channel \mathbf{F}_{i_k} from their serving BSs, as well as the signal and interference plus noise covariance matrix Φ_{i_k} . The BSs need to know the effective uplink channels $\{\mathbf{G}_{i_k}\}_{k=1}^{K_c}$ to the UEs they serve, the corresponding MSE weights $\{\mathbf{W}_{i_k}^{1/2}\}_{k=1}^{K_c}$, and the uplink signal plus interference covariance matrix Γ_i . Several methods for obtaining the

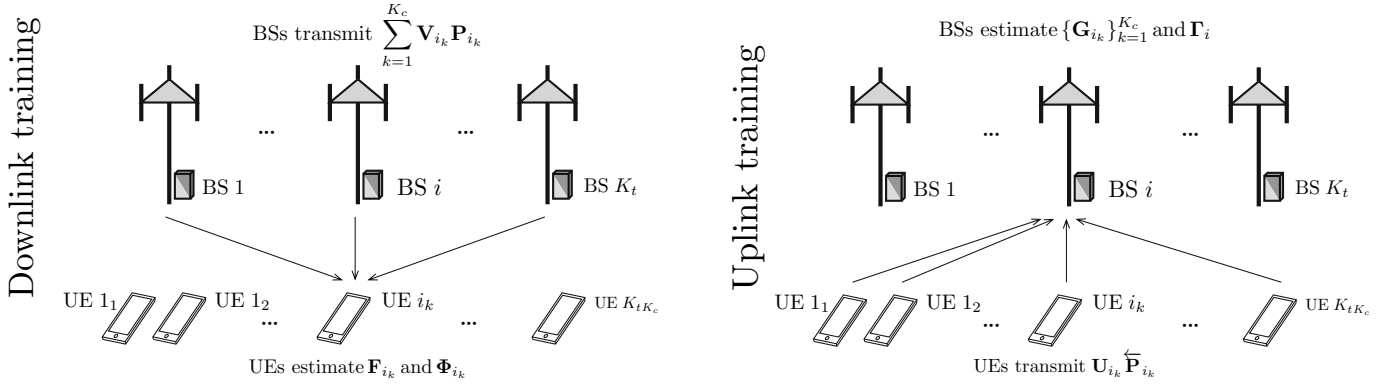


Fig. 1. CSI estimation in one subframe (cf. Fig. 2). In each subframe, the downlink channels are estimated using pilots from the BSs. Later, the uplink pilots are estimated using pilots from the UEs. Additionally, the UEs feed back \mathbf{W}_{i_k} to their serving BS using an out-of-band feedback link.

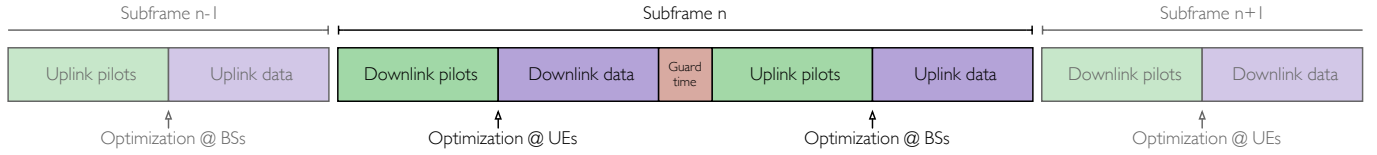


Fig. 2. Schematic drawing of subframes.

required CSI⁵ at the nodes can be imagined, using various combinations of channel estimation, feedback, signaling and backhaul. In this section, we will propose three CSI acquisition methods, with different tradeoffs between these aspects.

The channel estimation in the proposed methods exploits the reciprocity of the network, and uses pilot transmissions in both uplink and downlink. As the effective channels change between iterations in the WMMSE algorithm, we propose to perform a training phase between one iteration and the next. A schematic drawing of the subframe structure that we envision can be seen in Fig. 2. The subframe is split between pilot transmission and data transmission, in both the uplink and downlink. Data transmission thus takes place *between* the filter updates of the algorithm. The ratio between uplink and downlink data transmission lengths could be flexibly allocated [32]. Before the iterative algorithm has converged, the data rates that are achievable in the downlink data transmission phase may be low, but not negligible, as shown by the numerical results in Sec. V-A1. An illustration of the channel estimation in one subframe is shown in Fig. 1.

In block fading channels, the coherence interval should be sufficiently long such that the iterative algorithm can perform enough iterations to reach good performance. The deployment scenario will determine the coherence time of the channel, and the details of the frame structure will determine the number of subframes that can be transmitted within one coherence interval. As a brief example, under a block fading channel with carrier frequency $f_c = 2$ GHz and UE speed $v = 3$ km/h, the coherence time can be modeled as $T_c = \frac{1}{2f_c} \frac{c}{v} = 90$ ms [33]. For future 5G systems, the TDD switching periodicity is planned to be 1 ms or less [32], [34], leading to at least

90 uplink-downlink iterations in one coherence interval when the UEs are slowly moving. In continuous fading channels, the proposed algorithm would possibly instead be able to track the channel variations, assuming that they are slow enough. In the rest of the paper, we make the assumption that the channel is changing slowly enough for the iterative algorithm to reach adequate performance.

We now detail the different CSI acquisition methods, which all rely on pilot-assisted channel estimation. When a statistical characterization of the channel is available, the MMSE channel estimator [5] is typically used. Here we estimate the effective channels, which are updated in each WMMSE iteration based on the current channel conditions. Obtaining a statistical characterization of the effective channel is thus complicated. In the estimation, we therefore regard the effective channels as deterministic but unknown. Under this perspective from classical estimation theory, it is easy to find the minimum variance unbiased (MVU) estimator.

A. Fully Distributed CSI Acquisition

First, we seek to estimate the effective downlink channel $\mathbf{F}_{i_k} = \mathbf{H}_{i_k} \mathbf{V}_{i_k}$ using synchronous pilot transmissions. In the downlink training phase, the BSs transmit orthogonal pilot sequences⁶ $\mathbf{P}_{i_k} \in \mathbb{C}^{N_d \times N_{p,d}}$ per user, such that $\mathbf{P}_{i_k} \mathbf{P}_{j_l}^H = N_{p,d} \mathbf{I}_{N_d} \delta_{i_k, j_l}$. In order to fulfill the orthogonality requirement, $N_{p,d} \geq K_r N_d$. The received signal $\mathbf{Y}_{i_k} \in \mathbb{C}^{M_r \times N_{p,d}}$ at UE i_k is then

$$\mathbf{Y}_{i_k} = \mathbf{H}_{i_k} \mathbf{V}_{i_k} \mathbf{P}_{i_k} + \sum_{(j,l) \neq (i,k)} \mathbf{H}_{i_k, j} \mathbf{V}_{j_l} \mathbf{P}_{j_l} + \mathbf{Z}_{i_k}. \quad (11)$$

⁶The framework can be extended to allow for non-orthogonal pilots, but then a pilot allocation scheme must be set up to minimize the problem of pilot contamination [35]. Furthermore, the resource allocation step should take the pilot contamination into account.

⁵We remind the reader that our notion of ‘CSI’ encompasses knowledge of the effective channels and the covariance matrices; see Tables I and II.

Notice that the power allocated to the pilots is the same as the power allocated to the data symbols in (1). This will enable distributed and unbiased estimation of Φ_{i_k} . This type of pilot transmissions, intended to estimate the effective channels, are called ‘UE-specific reference signals’ in the LTE standard [36].

Assuming that UE i_k knows its designated pilot \mathbf{P}_{i_k} , this is a deterministic parameter estimation problem in Gaussian noise. The MVU estimator of the effective channel $\mathbf{F}_{i_k} \in \mathbb{C}^{M_r \times N_d}$ is then [5]:

$$\hat{\mathbf{F}}_{i_k} = \frac{1}{N_{p,d}} \mathbf{Y}_{i_k} \mathbf{P}_{i_k}^H = \mathbf{H}_{i_k} \mathbf{V}_{i_k} + \frac{1}{N_{p,d}} \mathbf{Z}_{i_k} \mathbf{P}_{i_k}^H. \quad (12)$$

The MVU estimator is an unbiased, efficient and asymptotically consistent (in $N_{p,d}$) estimator of \mathbf{F}_{i_k} . In addition to knowing \mathbf{F}_{i_k} , UE i_k also needs knowledge of $\Phi_{i_k} \in \mathbb{C}^{M_r \times M_r}$. This can be achieved by applying the sample covariance estimator:

$$\begin{aligned} \hat{\Phi}_{i_k} &= \frac{1}{N_{p,d}} \mathbf{Y}_{i_k} \mathbf{Y}_{i_k}^H \\ &= \sum_{(j,l)} (\mathbf{H}_{i_k j} \mathbf{V}_{j l} \mathbf{V}_{j l}^H \mathbf{H}_{i_k j}^H) + \frac{1}{N_{p,d}} \mathbf{Z}_{i_k} \mathbf{Z}_{i_k}^H \\ &\quad + \frac{1}{N_{p,d}} \sum_{(j,l)} (\mathbf{H}_{i_k j} \mathbf{V}_{j l} \mathbf{P}_{j l} \mathbf{Z}_{i_k}^H + \mathbf{Z}_{i_k} \mathbf{P}_{j l}^H \mathbf{V}_{j l}^H \mathbf{H}_{i_k j}^H). \end{aligned} \quad (13)$$

Since the only stochastic component of \mathbf{Y}_{i_k} is \mathbf{Z}_{i_k} , the estimator in (13) is unbiased.

The uplink estimation is performed in a similar manner as the downlink estimation. Now the UEs each transmit a signal $\tilde{\mathbf{X}}_{i_k} = \gamma \mathbf{U}_{i_k} \tilde{\mathbf{P}}_{i_k}$, where $\tilde{\mathbf{P}}_{i_k} \in \mathbb{C}^{N_d \times N_{p,u}}$ are orthogonal pilots, such that $\tilde{\mathbf{P}}_{i_k} \tilde{\mathbf{P}}_{j_l}^H = N_{p,u} \mathbf{I}_{N_d} \delta_{i_k, j_l}$. As will be shown by Proposition 1 in Sec. IV-D, $\|\mathbf{U}_{i_k}\|_F^2 = \alpha_{i_k} \|\mathbf{A}_{i_k} \mathbf{W}_{i_k}^{1/2}\|_F^2 \leq \alpha_{i_k} N_d / \sigma_r^2$. In order to maximize the uplink estimation SNR, the scaling factor γ is set as⁷ $\gamma = \sqrt{P_r \sigma_r^2 / N_d}$, where P_r is the maximum transmit power of the UEs. The UE quantities P_r and σ_r^2 are assumed to be known at the BSs, such that they have perfect a priori knowledge of γ . For this setup, assuming synchronized pilot transmissions from the UEs, the received signal $\tilde{\mathbf{Y}}_i \in \mathbb{C}^{M_t \times N_{p,u}}$ at BS i during the uplink training phase is

$$\tilde{\mathbf{Y}}_i = \gamma \sum_{k=1}^{K_c} \mathbf{H}_{i_k}^H \mathbf{U}_{i_k} \tilde{\mathbf{P}}_{i_k} + \gamma \sum_{j \neq i} \sum_{l=1}^{K_c} \mathbf{H}_{j l}^H \mathbf{U}_{j l} \tilde{\mathbf{P}}_{j l} + \tilde{\mathbf{Z}}_i. \quad (14)$$

The MVU estimator of the uplink effective channel $\mathbf{G}_{i_k} \in \mathbb{C}^{M_t \times N_d}$ is

$$\hat{\mathbf{G}}_{i_k} = \frac{1}{\gamma N_{p,u}} \tilde{\mathbf{Y}}_i \tilde{\mathbf{P}}_{i_k}^H = \mathbf{H}_{i_k}^H \mathbf{U}_{i_k} + \frac{1}{\gamma N_{p,u}} \tilde{\mathbf{Z}}_i \tilde{\mathbf{P}}_{i_k}^H. \quad (15)$$

Furthermore, the signal and interference *plus scaled noise* covariance matrix $\hat{\Gamma}_i^{s+i+n} \in \mathbb{C}^{M_t \times M_t}$ is estimated using the sample covariance:

$$\hat{\Gamma}_i^{s+i+n} = \frac{1}{\gamma^2} \frac{1}{N_{p,u}} \tilde{\mathbf{Y}}_i \tilde{\mathbf{Y}}_i^H. \quad (16)$$

⁷Note that the UE dependent factor α_{i_k} in $\|\mathbf{U}_{i_k}\|_F^2$ should not be removed by the scaling, since then Γ_i cannot be estimated in a fully distributed fashion. If $\alpha_{i_k} < 1$, the full transmit power of UE i_k cannot be used.

The WMMSE algorithm however needs an estimate of $\Gamma_i = \Gamma_i^{s+i}$, without the noise covariance component of Γ_i^{s+i+n} . In Sec. IV-C, we resolve this issue by modifying the WMMSE algorithm.

When forming the precoder in (10), the product $\sqrt{\alpha_{i_k}} \mathbf{H}_{i_k}^H \mathbf{A}_{i_k} \mathbf{W}_{i_k} = \mathbf{G}_{i_k} \mathbf{W}_{i_k}^{1/2}$ is needed. Instead of independently estimating this quantity in a second uplink estimation phase, we let UE i_k feed back \mathbf{W}_{i_k} to its serving BS i . Together with (15), BS i can then form $\hat{\mathbf{G}}_{i_k} \mathbf{W}_{i_k}^{1/2}$ and use that in (10). The point of this procedure is to avoid signal cancellation [37], where a small mismatch between the estimate of $\mathbf{G}_{i_k} \mathbf{W}_{i_k}^{1/2}$ and the estimate of Γ_i can have a large detrimental impact on performance. If \mathbf{G}_{i_k} and Γ_i are estimated using the same pilot transmissions, as in (15) and (16), the covariance matrix can be decomposed as $\hat{\Gamma}_i^{s+i+n} = \hat{\Gamma}_i^{i+n} + \hat{\mathbf{G}}_{i_k} \hat{\mathbf{G}}_{i_k}^H$. Because of this structure, there is no mismatch between $\hat{\mathbf{G}}_{i_k} \mathbf{W}_{i_k}^{1/2}$ and $\hat{\Gamma}_i$, and signal cancellation does not occur [37].

It can be shown that $R_{i_k} = \log \det(\mathbf{W}_{i_k})$. Feedback of the eigenvalues of \mathbf{W}_{i_k} therefore constitutes a *rate request* for each data stream of UE i_k , describing what rate that stream can handle under the current network conditions. This information is already fed back to the serving BS in a practical system. Recall that α_{i_k} is fixed and known at BS i , and does therefore not need to be fed back.

Remark 1. *The CSI acquisition proposed in this section is fully distributed over BSs and UEs, in the sense that only over-the-air signaling is required. UE i_k feeds back \mathbf{W}_{i_k} to its serving BS, but the BSs do not need to share any information over a BS backhaul.*

B. CSI Acquisition with Global Sharing of Individual Scaling Parameters

As noted in the previous section, and proved in Sec. IV-D, $\|\mathbf{U}_{i_k}\|_F^2 \leq \alpha_{i_k} N_d / \sigma_r^2$. The scaling factor γ was set based on this to maximize the uplink transmit power. However, unless the inequality is met with equality and $\alpha_{i_k} = 1$, the transmit power constraint of that particular UE is not met. Correspondingly, the uplink estimation SNR suffers for that UE. If the requirement of fully distributed estimation of the uplink covariance matrix is dropped, and by introducing individual scaling factors for the UEs, the maximum uplink transmit power can always be used.

In this section, we keep the downlink estimation the same as in Sec. III-A, but modify the uplink estimation to maximize the transmit power used. The BSs will then need access to a backhaul network, where information about the individual scaling parameters can be shared.

Letting $\tilde{\mathbf{X}}_{i_k} = \frac{\sqrt{P_r}}{\|\mathbf{U}_{i_k}\|_F} \mathbf{U}_{i_k} \tilde{\mathbf{P}}_{i_k}$, the effective uplink transmit power is maximized for UE i_k . The received signal at BS i is then

$$\tilde{\mathbf{Y}}_i = \sqrt{P_r} \sum_{(j,l)} \frac{1}{\|\mathbf{U}_{j l}\|_F} \mathbf{H}_{j l}^H \mathbf{U}_{j l} \tilde{\mathbf{P}}_{j l} + \tilde{\mathbf{Z}}_i. \quad (17)$$

We now assume that the individual scaling factors $\|\mathbf{U}_{i_k}\|_F$ are fed back from the UEs to their serving BSs, and then globally

TABLE III
FEEDBACK AND ESTIMATION NEEDED FOR THE DIFFERENT CSI ACQUISITION METHODS.

Method	Estimated at UE i_k	BS i feedback to served UE i_k	Estimated at BS i	UE i_k feedback to serving BS i	Shared information over BS backhaul
Fully distributed (Sec. III-A)	$\Phi_{i_k}, \mathbf{F}_{i_k}$	$\alpha_{i_k}^\diamond$	$\Gamma_i, \{\mathbf{G}_{i_k}\}$	\mathbf{W}_{i_k}	—
Globally shared individual scale factors (Sec. III-B)	$\Phi_{i_k}, \mathbf{F}_{i_k}$	$\alpha_{i_k}^\diamond$	$\{\mathbf{G}_{j_l i}\}$	$\ \mathbf{U}_{i_k}\ _F, \mathbf{W}_{i_k}$	$\{\ \mathbf{U}_{j_l}\ _F\}$
Globally shared filters (Sec. III-C)	$\{\mathbf{H}_{i_k j}\}^\dagger$	$\alpha_{i_k}^\diamond, \{\mathbf{V}_{j_l}\}$	$\{\mathbf{H}_{j_l i}\}^\dagger$	$\mathbf{U}_{i_k}, \mathbf{W}_{i_k}$	$\{\mathbf{U}_{j_l}\}, \{\mathbf{W}_{j_l}\}, \{\mathbf{V}_{j_l}\}$

\diamond The user priorities only need to be fed back if/when they are changed.

\dagger The estimated quantities for the methods in Sec. III-A and Sec. III-B depend on the transmit and receive filters, and must therefore be re-estimated in every subframe. The estimated quantities for the method in Sec. III-C do not change within one coherence block, and can therefore be improved upon in every subframe.

shared over a BS backhaul. BS i can then estimate the effective channels from UE j_l as

$$\hat{\mathbf{G}}_{j_l i} = \frac{\|\mathbf{U}_{j_l}\|_F}{\sqrt{P_r N_{p,u}}} \hat{\mathbf{Y}}_i \hat{\mathbf{P}}_{j_l}^H = \mathbf{H}_{i_k j}^H \mathbf{U}_{j_l} + \frac{\|\mathbf{U}_{j_l}\|_F}{\sqrt{P_r N_{p,u}}} \hat{\mathbf{Z}}_i \hat{\mathbf{P}}_{j_l}^H. \quad (18)$$

Since the scaled pilots effectively all have the same weight, the sample covariance estimator of Γ_i in (16) cannot be used. Instead, we rely on the biased estimator $\hat{\Gamma}_i^{s+i+n} = \sum_{(j,l)} \hat{\mathbf{G}}_{j_l i} \hat{\mathbf{G}}_{j_l i}^H$. The bias is determined by the factors $\|\mathbf{U}_{i_k}\|_F / \sqrt{P_r}$ and the pilots $\{\hat{\mathbf{P}}_{j_l}\}$. Since $\|\mathbf{U}_{i_k}\|_F^2 \leq \alpha_{i_k} N_d / \sigma_r^2$, the scaling factors could be quantized over $[0, \sqrt{\alpha_{i_k} N_d / \sigma_r^2}]$, $\forall i_k$.

This estimation scheme is similar to one proposed in [14], where a scaled version of \mathbf{A}_{i_k} was used as the uplink precoder. The MSE weights \mathbf{W}_{i_k} can then be directly estimated at the serving BSs, and do not need to be fed back. However, in order for the BSs to estimate $\hat{\Gamma}_i^{s+i+n}$ in that estimation scheme, they must exchange the MSE weights for their corresponding UEs over the backhaul. In essence, reduced over-the-air feedback has been traded for more backhaul use.

Remark 2. *The CSI acquisition proposed in this section is fully distributed over the UEs, but not over the BSs. Each BS needs knowledge of the individual scaling factors for all UEs, information which is shared over a BS backhaul.*

C. CSI Acquisition with Global Sharing of Precoders, Receive Filters and MSE Weights

Lastly, we present an CSI acquisition scheme which relies even further on feedback, signaling and backhaul. We present this method since the state-of-the-art robust WMMSE algorithms in [7]–[10] require this type of CSI acquisition. In this scheme, only the underlying channels are estimated exploiting the reciprocity, and the filters and MSE weights must be signaled between all nodes.

In the downlink training, $\mathbf{P}_j \in \mathbb{C}^{M_t \times N_{p,d}}$ are orthogonal pilots sent from BS j such that $\mathbf{P}_i \mathbf{P}_j^H = N_{p,d} \mathbf{I}_{M_t} \delta_{i,j}$. The received signal at UE i_k is then

$$\mathbf{Y}_{i_k} = \sqrt{\frac{P_i}{K_c M_t}} \mathbf{H}_{i_k i} \mathbf{P}_i + \sum_{(j,l) \neq (i,k)} \sqrt{\frac{P_j}{K_c M_t}} \mathbf{H}_{i_k j} \mathbf{P}_j + \mathbf{Z}_{i_k}. \quad (19)$$

This type of pilot transmissions are called ‘cell-specific reference signals’ in the LTE standard [36]. For the case of

TABLE IV
TOTAL ESTIMATION COMPLEXITY, PER ITERATION AND UE.

Method	Approximate number of flops
Sec. III-A	$(M_r N_d + M_r^2) N_{p,d} + M_t N_d N_{p,u} + M_t^2 N_{p,u} / K_c$
Sec. III-B	$(M_r N_d + M_r^2) N_{p,d} + (M_t N_d N_{p,u} + M_t^2 N_d + M_t^2) K_t$
Sec. III-C	$M_r M_t K_t (N_{p,d} + N_{p,u}) + (M_r M_t N_d + M_r^2 N_d + M_r^2) K_r + (M_r M_t N_d + M_t^2 N_d + M_t^2) K_t + M_r$

Rayleigh fading, $\text{vec}(\mathbf{H}_{i_k j}) \sim \mathcal{CN}(0, \mathbf{I})$, the MMSE estimator [5] is

$$\hat{\mathbf{H}}_{i_k j} = \frac{\sqrt{P_j / (K_c M_t)}}{N_{p,d} \frac{P_j}{K_c M_t} + \sigma_r^2} \mathbf{Y}_{i_k} \mathbf{P}_j^H. \quad (20)$$

Assuming that the noise variance σ_r^2 is known, and that all precoders $\{\mathbf{V}_{j_l}\}$ have been fed back to UE i_k , it can form

$$\hat{\mathbf{F}}_{i_k} = \hat{\mathbf{H}}_{i_k i} \mathbf{V}_{i_k}, \quad \hat{\Phi}_{i_k} = \sum_{(j,l)} \hat{\mathbf{H}}_{i_k j} \mathbf{V}_{j_l} \mathbf{V}_{j_l}^H \hat{\mathbf{H}}_{i_k j}^H + \sigma_r^2 \mathbf{I}. \quad (21)$$

With the subframe structure in Fig. 2, consecutive training phases can be used to monotonically improve the channel estimates in one coherence block of the channel. This can be done using iterative techniques, see e.g. [38, Ch. 12.6].

In the uplink, assuming feedback of receive filters and MSE weights, which are shared among all BSs, $\hat{\mathbf{G}}_{i_k}$ and $\hat{\Gamma}_i$ are formed in a similar fashion as in (21).

Remark 3. *The CSI acquisition proposed in this section is centralized. It requires significant signaling of filters among BSs and UEs in every subframe. In terms of estimating the underlying channels $\{\mathbf{H}_{i_k j}\}$, it is however distributed over the BSs and UEs.*

D. Feedback Requirements, Computational Complexity and Quantized MSE Weight Feedback

We compare the feedback requirements of the proposed estimation schemes in Table III. The estimation matrix operation complexities [39] are shown in Table IV. The $M_r M_t K_t (N_{p,d} + N_{p,u})$ term in the Sec. III-C estimation method flop count dominates all other terms when the number of pilots is large. This effect is illustrated in Fig. 8 of Sec. V-A3, where the complexities of the estimators are visually compared with the correspondingly achieved sum rates.

In the CSI acquisition proposed in Sec. III-A and Sec. III-B, feedback of the MSE weights to the serving BS is required. In order to be practical, the MSE weights should be *quantized* and fed back. Since the MSE matrix \mathbf{W}_{i_k} is Hermitian, it can be quantized by quantizing its eigenvalue decomposition. The eigenvectors can be quantized using e.g. Grassmannian subspace packing [40]. Recalling that $s_1(\cdot)$ denotes the largest singular value, we have the following helpful lemma for the quantization of the eigenvalues:

Lemma 1. *The eigenvalues of the MSE weights are bounded as $1 \leq \lambda_n(\mathbf{W}_{i_k}) \leq 1 + \frac{P_i s_1^2(\mathbf{H}_{i_k i})}{\sigma_i^2}$, $\forall i_k, n$.*

Proof: The proof is given in Appendix A. ■

The square roots of the eigenvalues of \mathbf{W}_{i_k} can therefore be quantized by finding⁸ a suitable scalar quantizer over the interval in Lemma 1. After UE i_k has fed back the quantized eigenvectors and the quantized square roots of the eigenvalues, the serving BS can then form the reconstructed square root of the MSE weight $\widehat{\mathbf{W}}_{i_k}^{1/2}$.

As mentioned in Sec. III-A, $R_{i_k} = \log \det(\mathbf{W}_{i_k}) = \sum_n \log(\lambda_n(\mathbf{W}_{i_k}))$ can be seen as the data rate (summed over data streams) for UE i_k . Quantizing $\lambda_n(\mathbf{W}_{i_k})$ therefore corresponds to making a set of discrete rates available to the UE, corresponding to e.g. a set of different modulation and coding schemes.

IV. INHERENT AND ENFORCED ROBUSTNESS OF WMMSE SOLUTIONS

In this section, we propose some modifications to the original WMMSE algorithm that lead to an algorithm which is robustified against CSI estimation errors.

A. Naïve WMMSE Algorithm with Estimated CSI

It is straightforward to naïvely feed the WMMSE algorithm the estimated CSI from one of the presented CSI acquisition methods. An example of the resulting performance can be seen in Fig. 3. The simulation settings are described in detail in Sec. V-A. It is clear that the naïve application of the WMMSE algorithm works moderately well for the centralized CSI acquisition schemes, but the performance for the fully distributed CSI acquisition scheme catastrophically deteriorates at high SNR. Thus, some form of robustification against CSI estimation errors is necessary.

B. General Worst-Case Robustness WMMSE Problem

One approach to robustifying the optimization problem in (6) is to minimize the objective function under the worst-case error conditions:

$$\begin{aligned} & \min_{\{\mathbf{A}_{i_k}\}} \max_{\{\text{uncertainty}\}} \sum_{(i,k)} \alpha_{i_k} (\text{Tr}(\mathbf{W}_{i_k} \mathbf{E}_{i_k}) - \log \det(\mathbf{W}_{i_k})) \\ & \quad \{\mathbf{W}_{i_k} \succ \mathbf{0}\} \\ & \quad \{\mathbf{V}_{i_k}\} \\ \text{subject to} & \quad \sum_{k=1}^{K_c} \text{Tr}(\mathbf{V}_{i_k} \mathbf{V}_{i_k}^H) \leq P_i, \quad i = 1, \dots, K_t. \end{aligned} \quad (22)$$

⁸The details of designing such a quantizer is outside the scope of this paper.

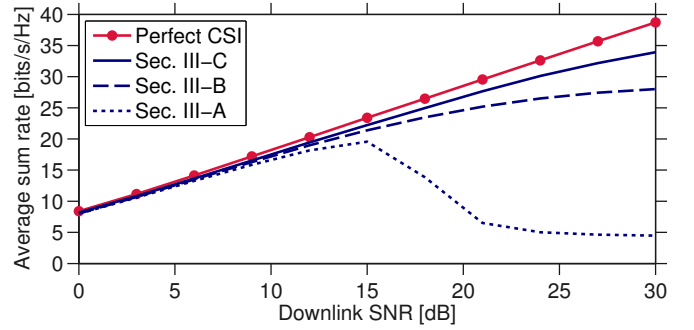


Fig. 3. Sum rate performance when naively applying the WMMSE algorithm together with the CSI acquisition schemes. The scenario was a $K_t = 3$, $K_c = 2$, $M_t = 4$, $M_r = 2$ interfering broadcast channel with $N_d = 1$. The channels were i.i.d. Rayleigh fading, and the uplink SNR was set as $\text{SNR}_u = P_r/\sigma_t^2 = 10$ dB for all links. Note that $\text{SNR}_d = P_t/\sigma_r^2$ affects both the power constraint in the WMMSE algorithm, as well as the estimation performance in the downlink estimation, since the downlink pilots are precoded with the same precoders as used in the data transmission.

The proposed CSI acquisition methods provide estimates both of the downlink effective channels, as well as of the uplink effective channels. Due to the definition of these effective channels, the general uncertainty set in (22) cannot be explicitly defined in terms of the uplink and downlink estimation errors simultaneously. For example, one of the terms in the objective function of (22) is $\mathbf{W}_{i_k} \mathbf{A}_{i_k}^H \mathbf{H}_{i_k i} \mathbf{V}_{i_k} = \mathbf{W}_{i_k}^{1/2} \mathbf{G}_{i_k}^H \mathbf{V}_{i_k} = \mathbf{W}_{i_k} \mathbf{A}_{i_k}^H \mathbf{F}_{i_k}$, which cannot be written in terms of \mathbf{G}_{i_k} and \mathbf{F}_{i_k} simultaneously. In the forthcoming alternating minimization, we will therefore solve (22) with the CSI uncertainty relating to the particular block of variables for which (22) is solved for. That is, for the precoders the CSI uncertainty at the BSs will be considered, whereas for the receive filters and MSE weights, the CSI uncertainty at the UEs will be considered. We now detail the alternating minimization solutions for the three blocks of (22).

C. Precoder Robustness

First we fix $\{\mathbf{A}_{i_k}, \mathbf{W}_{i_k}\}$ and solve (22) with respect to the precoders $\{\mathbf{V}_{i_k}\}$. The optimization problem can then be interpreted as a local optimization problem at each BS, given that the CSI uncertainty in (22) comes from the uplink channel estimation phase. We let the estimation errors for BS i be $\tilde{\Gamma}_i = \Gamma_i - \hat{\Gamma}_i^{s+i}$ and $\tilde{\mathbf{G}}_{i_k} = \mathbf{G}_{i_k} - \hat{\mathbf{G}}_{i_k}$, $k = 1, \dots, K_c$. For the local uncertainty set, we assume that the errors are norm bounded as $\|\tilde{\Gamma}_i\|_F \leq \varepsilon_i^{(\text{BS})}$ and $\|\tilde{\mathbf{G}}_{i_k} \mathbf{W}_{i_k}^{1/2}\|_F \leq \xi_{i_k}^{(\text{BS})}$, $k = 1, \dots, K_c$. Note that $\xi_{i_k}^{(\text{BS})}$ depends on \mathbf{W}_{i_k} , which is fixed. The local worst-case optimization problem for BS i is then:

$$\begin{aligned} & \text{minimize}_{\{\mathbf{V}_{i_k}\}} \max_{\|\tilde{\Gamma}_i\|_F \leq \varepsilon_i^{(\text{BS})}} \sum_{k=1}^{K_c} \text{Tr}(\mathbf{V}_{i_k}^H (\hat{\Gamma}_i^{s+i} + \tilde{\Gamma}_i) \mathbf{V}_{i_k}) \\ & \quad \|\tilde{\mathbf{G}}_{i_k} \mathbf{W}_{i_k}^{1/2}\|_F \leq \xi_{i_k}^{(\text{BS})} \\ & \quad - 2\sqrt{\alpha_{i_k}} \text{Re} \left(\text{Tr} \left(\mathbf{W}_{i_k}^{1/2} (\hat{\mathbf{G}}_{i_k} + \tilde{\mathbf{G}}_{i_k})^H \mathbf{V}_{i_k} \right) \right) \\ \text{subject to} & \quad \sum_{k=1}^{K_c} \text{Tr}(\mathbf{V}_{i_k} \mathbf{V}_{i_k}^H) \leq P_i, \quad i = 1, \dots, K_t. \end{aligned} \quad (23)$$

The solution to the inner optimization problem of (23) can be found by extending the results of [22], [23] to the multiuser matrix case. By upper bounding the optimal value of the inner optimization problem using the triangle inequality⁹ and the submultiplicativity of the Frobenius norm, the (pessimistic) robust optimal precoder for UE i_k is

$$\mathbf{V}_{i_k}^{\text{rob}} = \sqrt{\alpha_{i_k}} \left(\hat{\Gamma}_i^{\text{S+I}} + \left(\varepsilon_i^{(\text{BS})} + \frac{\xi_{i_k}^{(\text{BS})}}{\|\mathbf{V}_{i_k}^{\text{rob}}\|_{\text{F}}} + \mu_i \right) \mathbf{I} \right)^{-1} \hat{\mathbf{G}}_{i_k} \mathbf{W}_{i_k}^{1/2}. \quad (24)$$

As before, μ_i is the Lagrange multiplier for the sum power constraint. Note that the robust precoder in (24) is *diagonally loaded* by a constant factor $\varepsilon_i^{(\text{BS})}$, a data dependent factor $\xi_{i_k}^{(\text{BS})} / \|\mathbf{V}_{i_k}^{\text{rob}}\|_{\text{F}}$, and the Lagrange multiplier μ_i . Diagonal loading is well known to robustify beamformers in various settings, and a large body of literature has studied its robustifying effects; see e.g. [17]–[23].

In order to construct the robust precoder in (24), the parameters $\varepsilon_i^{(\text{BS})}$ and $\xi_{i_k}^{(\text{BS})}$ must be known. For the fully distributed CSI estimation in Sec. III-A, the effective channel error $\hat{\mathbf{G}}_{i_k}$ follows a zero-mean Gaussian distribution with known covariance, and $\xi_{i_k}^{(\text{BS})}$ can thus be selected such that $\|\hat{\mathbf{G}}_{i_k} \mathbf{W}_{i_k}^{1/2}\|_{\text{F}} \leq \xi_{i_k}^{(\text{BS})}$ holds with some probability. The statistics of the covariance error $\tilde{\Gamma}_i$ however depend on the filters $\{\mathbf{U}_{i_k}\}$, which are unknown at BS i . Since the optimal amount of diagonal loading is unknown, we therefore propose to disregard $\varepsilon_i^{(\text{BS})}$ and $\xi_{i_k}^{(\text{BS})}$, and let the factor μ_i handle all the diagonal loading. To compensate for the missing $\varepsilon_i^{(\text{BS})}$ and $\xi_{i_k}^{(\text{BS})}$, we implicitly amplify μ_i using a scaling procedure.

1) Implicitly Selecting the Diagonal Loading Parameter:

When applying diagonal loading for robustness, a heuristic often used in the literature [23] is to select a fixed loading level around 10 dB over the noise level. Instead, we propose a data dependent method for selecting the diagonal loading parameter implicitly. We note that, given estimates $\hat{\Gamma}_i^{\text{S+I+n}}$, $\hat{\mathbf{G}}_{i_k}$ and fed back \mathbf{W}_{i_k} , the precoders in the WMMSE algorithm are formed like

$$\mathbf{V}_{i_k} = \sqrt{\alpha_{i_k}} \left(\hat{\Gamma}_i^{\text{S+I+n}} + \mu_i \mathbf{I} \right)^{-1} \hat{\mathbf{G}}_{i_k} \mathbf{W}_{i_k}^{1/2}. \quad (25)$$

The form of (25) and (24) are similar, and it can therefore be concluded that μ_i alone acts as the diagonal loading for the naïve WMMSE precoder. The factor μ_i therefore robustifies the solution, and the amount of diagonal loading is determined by $\hat{\Gamma}_i^{\text{S+I+n}}$, $\hat{\mathbf{G}}_{i_k} \mathbf{W}_{i_k}^{1/2}$ and P_i .

In order to artificially amplify the factor μ_i , we now introduce a scaling procedure. We let $0 \leq \rho \leq 1$ be a scaling factor, and modify the WMMSE algorithm as follows:

- 1) In the precoder optimization at BS i (step 4 in Algorithm 1), let the sum power constraint be ρP_i . The resulting precoders from (25) are denoted $\{\mathbf{V}_{i_k}^{(\rho)}\}$, and will have equal or higher diagonal loading level than the original precoder in (25), since μ_i is nonincreasing in the sum power constraint value.

⁹This relaxes the problem such that $\tilde{\Gamma}_i$ is the worst for each UE simultaneously. This is equivalent to replacing the existing covariance constraint with $\|\hat{\Gamma}_{i_k}\|_{\text{F}} \leq \varepsilon_i^{(\text{BS})}$, $k = 1, \dots, K_c$, and changing the objective accordingly.

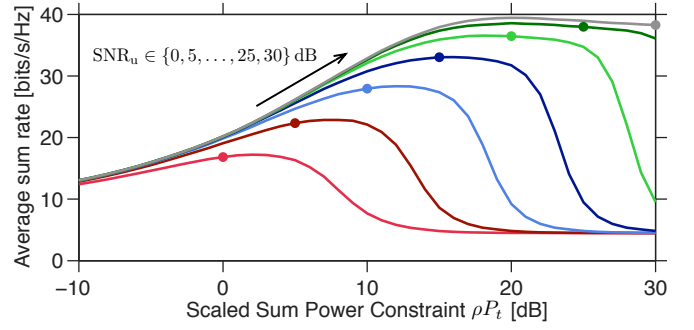


Fig. 4. Sum rate performance when varying ρ , for $P_t = 1000$ and $\sigma_\gamma^2 = 1$ and varying SNR_u . The solid markers represent the performance for $\rho = \min\left(\frac{P_t/\sigma_\gamma^2}{P_t/\sigma_\gamma^2}, 1\right)$. The scenario is the same as in Fig. 3.

- 2) Form scaled precoders $\mathbf{V}_{i_k}^{\text{opt}} = \frac{1}{\sqrt{\rho}} \mathbf{V}_{i_k}^{(\rho)}$, and use these for downlink pilot and data transmission. This scaling ensures that the correct transmit power is used.
- 3) At the UEs, perform the estimation given the precoders $\{\mathbf{V}_{i_k}\}$, giving $\{\hat{\mathbf{F}}_{i_k}^{(\rho)}\}$ and $\{\hat{\Phi}_{i_k}^{(\rho)}\}$.
- 4) Scale the estimates as $\hat{\Phi}_{i_k} = \rho \hat{\Phi}_{i_k}^{(\rho)}$, $\hat{\mathbf{F}}_{i_k} = \sqrt{\rho} \hat{\mathbf{F}}_{i_k}^{(\rho)}$, and use $\{\hat{\Phi}_{i_k}\}$ and $\{\hat{\mathbf{F}}_{i_k}\}$ to form receive filters and MSE weights. This scaling is necessary in order for the WMMSE algorithm at the UEs to be aware of what the original precoders $\mathbf{V}_{i_k}^{(\rho)}$ were.

The same scaling ρ is used at all BSs, and therefore the signal-to-interference ratios of the cross-links are not affected. In Fig. 4, we plot the impact of selecting different ρ . A simple selection that appears to work well is $\rho = \min\left(\frac{P_t/\sigma_\gamma^2}{P_t/\sigma_\gamma^2}, 1\right)$.

2) *Removing the Noise Component of $\hat{\Gamma}_i^{\text{S+I+n}}$* : Comparing (25) with (10), it can be noted that the covariance matrix should be $\hat{\Gamma}_i^{\text{S+I}}$, and not $\hat{\Gamma}_i^{\text{S+I+n}}$. The noise portion of $\hat{\Gamma}_i^{\text{S+I+n}}$ will on average be $\frac{\sigma_i^2}{\gamma^2} \mathbf{I}$, but simply subtracting that might make the resulting matrix indefinite. Instead, we modify μ_i to allow for *negative* values; this is the same as seeing μ_i as the difference of a non-negative Lagrange multiplier with an estimate of the noise power. Specifically, we allow $\mu_i \geq -\min\left(\frac{\sigma_i^2}{\gamma^2}, \lambda_{M_t}\left(\hat{\Gamma}_i^{\text{S+I+n}}\right) - \zeta\right)$ where ζ is some constant value determining how close to singular $\hat{\Gamma}_i^{\text{S+I+n}} + \mu_i \mathbf{I}$ can be.

D. Receive Filter and MSE Weight Robustness

With similar notation and assumptions as in (23), the local worst-case optimization problem for the receive filter at UE i_k is

$$\min_{\{\mathbf{A}_{i_k}\}} \max_{\substack{\|\tilde{\Phi}_{i_k}\|_{\text{F}} \leq \varepsilon_i^{(\text{UE})} \\ \|\tilde{\mathbf{F}}_{i_k} \mathbf{W}_{i_k}^{1/2}\|_{\text{F}} \leq \xi_{i_k}^{(\text{UE})}}} \text{Tr} \left(\mathbf{W}_{i_k} \left(\mathbf{I} + \mathbf{A}_{i_k}^H \left(\hat{\Phi}_{i_k} + \tilde{\Phi}_{i_k} \right) \mathbf{A}_{i_k} \right) \right) - 2\text{Re} \left(\text{Tr} \left(\mathbf{W}_{i_k} \left(\hat{\mathbf{F}}_{i_k} + \tilde{\mathbf{F}}_{i_k} \right)^H \mathbf{A}_{i_k} \right) \right), \quad (26)$$

whose (pessimistic) solution [23] is

$$\mathbf{A}_{i_k}^{\text{rob}} = \left(\hat{\Phi}_{i_k} + \left(\varepsilon_i^{(\text{UE})} + \frac{\xi_{i_k}^{(\text{UE})}}{\|\mathbf{A}_{i_k}^{\text{rob}} \mathbf{W}_{i_k}^{1/2}\|_{\text{F}}} \right) \mathbf{I} \right)^{-1} \hat{\mathbf{F}}_{i_k}. \quad (27)$$

Finally, the corresponding (pessimistic) robust MSE weight is

$$\mathbf{W}_{i_k}^{\text{rob}} = \left(\mathbf{I} - \widehat{\mathbf{F}}_{i_k}^H \left(\widehat{\Phi}_{i_k} + \left(\varepsilon_i^{(\text{UE})} + \frac{\xi_{i_k}^{(\text{UE})}}{\|\mathbf{A}_{i_k}^{\text{rob}} \mathbf{W}_{i_k}^{1/2}\|_{\text{F}}} \right) \mathbf{I} \right)^{-1} \widehat{\mathbf{F}}_{i_k} \right)^{-1} \widehat{\mathbf{F}}_{i_k} \quad (28)$$

Again, the optimal level of diagonal loading is unknown. This is because $\varepsilon_i^{(\text{UE})}$ depends on the statistics of the covariance error $\tilde{\Phi}_{i_k} = \Phi_{i_k} - \widehat{\Phi}_{i_k}$, which in turn depend on the unknown precoders $\{\mathbf{V}_{i_k}\}$. We therefore propose to indirectly apply diagonal loading at the UEs instead, based on the following observation:

Proposition 1. *The receive filter \mathbf{A}_{i_k} and MSE weight \mathbf{W}_{i_k} obtained in the UE side optimization of the WMMSE algorithm with perfect CSI satisfies $\|\mathbf{A}_{i_k} \mathbf{W}_{i_k}^{1/2}\|_{\text{F}}^2 = \text{Tr} \left((\Phi_{i_k}^{\text{in}})^{-1} - \Phi_{i_k}^{-1} \right) \leq N_d / \sigma_r^2$. If the effective channel is fully contained in an interference-free subspace of dimension $N_s \leq N_d$, then asymptotically $\|\mathbf{A}_{i_k} \mathbf{W}_{i_k}^{1/2}\|_{\text{F}}^2 \rightarrow N_s / \sigma_r^2$ as the SNR grows large.*

Proof: The proof is given in Appendix B. ■

The first part of this proposition has an important connection to the uplink training stage in the fully distributed estimation scheme (Sec. III-A). Since $\gamma \mathbf{U}_{i_k} = \sqrt{P_r \sigma_r^2 / N_d} \sqrt{\alpha_{i_k}} \mathbf{A}_{i_k} \mathbf{W}_{i_k}^{1/2}$ is acting as the uplink training stage precoder, $\|\mathbf{A}_{i_k} \mathbf{W}_{i_k}^{1/2}\|_{\text{F}}^2$ determines the effective UE transmit power, and hence the uplink estimation SNR. The second part shows that $\|\mathbf{A}_{i_k} \mathbf{W}_{i_k}^{1/2}\|_{\text{F}}^2$ also indicates whether *perfect interference alignment* is achieved for UE i_k .

1) *Enforcing Proposition 1 onto WMMSE Solutions with Imperfect CSI:* Proposition 1 relates to perfect CSI, but the inequality may not hold for the naïve solutions in (8) and (9) with imperfect CSI. In order to robustify the algorithm, we therefore explicitly impose the constraint on the UE side optimization problem with imperfect CSI. The problems still decouple over users, and the problem each UE should solve is

$$\begin{aligned} \underset{\mathbf{A}_{i_k}, \mathbf{W}_{i_k} \succ \mathbf{0}}{\text{minimize}} \quad & \text{Tr} \left(\mathbf{W}_{i_k} \left(\mathbf{I} + \mathbf{A}_{i_k}^H \widehat{\Phi}_{i_k} \mathbf{A}_{i_k} \right) \right) \\ & - 2 \text{Re} \left(\mathbf{W}_{i_k} \widehat{\mathbf{F}}_{i_k}^H \mathbf{A}_{i_k} \right) - \log \det \left(\mathbf{W}_{i_k} \right) \quad (29) \end{aligned}$$

$$\text{subject to} \quad \|\mathbf{A}_{i_k} \mathbf{W}_{i_k}^{1/2}\|_{\text{F}}^2 \leq N_d / \sigma_r^2.$$

Proposition 2. *The solution to (29) is*

$$\begin{aligned} \mathbf{A}_{i_k}^{\text{opt}} &= \left(\widehat{\Phi}_{i_k} + \nu_{i_k}^{\text{opt}} \mathbf{I} \right)^{-1} \widehat{\mathbf{F}}_{i_k} \\ \mathbf{W}_{i_k}^{\text{opt}} &= \left(\mathbf{I} - \widehat{\mathbf{F}}_{i_k}^H \left(\widehat{\Phi}_{i_k} + \nu_{i_k}^{\text{opt}} \mathbf{I} \right)^{-1} \widehat{\mathbf{F}}_{i_k} \right)^{-1}. \end{aligned}$$

If $\|\mathbf{A}_{i_k}^{\text{opt}} \left(\mathbf{W}_{i_k}^{\text{opt}} \right)^{1/2}\|_{\text{F}}^2 \leq N_d / \sigma_r^2$ holds for $\nu_{i_k}^{\text{opt}} = 0$, the constraint is not active and the solution has the same form as the original solution in Sec. II-A1. Otherwise, $\nu_{i_k}^{\text{opt}}$ can be found by bisection over $(0, \sigma_r^2]$ such that $\|\mathbf{A}_{i_k}^{\text{opt}} \left(\mathbf{W}_{i_k}^{\text{opt}} \right)^{1/2}\|_{\text{F}}^2 = N_d / \sigma_r^2$.

Proof: The proof is given in Appendix C. ■

Interestingly, explicitly imposing Proposition 1 as a constraint in (29) corresponds to diagonal loading of the receive filter $\mathbf{A}_{i_k}^{\text{opt}}$, giving it the same form as $\mathbf{A}_{i_k}^{\text{rob}}$. Likewise, $\mathbf{W}_{i_k}^{\text{opt}}$ has

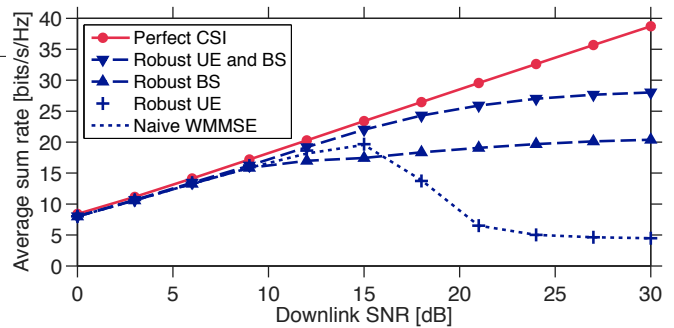


Fig. 5. Sum rate performance when selectively applying the robustifying measures in Sec. IV, together with the fully distributed CSI acquisition in Sec. III-A. For comparison purposes, the scenario is the same as in Fig. 3.

the same form as $\mathbf{W}_{i_k}^{\text{rob}}$. The important difference is that $\mathbf{A}_{i_k}^{\text{opt}}$ and $\mathbf{W}_{i_k}^{\text{opt}}$ do not depend on unknown parameters, whereas $\mathbf{A}_{i_k}^{\text{rob}}$ and $\mathbf{W}_{i_k}^{\text{rob}}$ do. Thus, $\mathbf{A}_{i_k}^{\text{opt}}$ and $\mathbf{W}_{i_k}^{\text{opt}}$ can be applied as realizable proxies for the unrealizable $\mathbf{A}_{i_k}^{\text{rob}}$ and $\mathbf{W}_{i_k}^{\text{rob}}$ in the robust WMMSE algorithm to be proposed.

By increasing ν_{i_k} , the requested rate $\log \det \left(\mathbf{W}_{i_k}^{\text{opt}} \right)$ is decreased. A large ν_{i_k} would occur when there are obvious discrepancies in the estimated CSI, such that $\|\mathbf{A}_{i_k} \mathbf{W}_{i_k}^{1/2}\|_{\text{F}}^2 \leq N_d / \sigma_r^2$ is far from being fulfilled without the diagonal loading.

We visualize the robustifying effects in Fig. 5 for the same simulation settings as in Fig. 3. The robustifying measures are effective, and result in up to a factor 5 sum rate gain over the naïve WMMSE algorithm.

E. Robustified WMMSE Algorithm

We now combine the diagonal loading robustifications in Sec. IV-C and Sec. IV-D (i.e. $\mathbf{V}_{i_k}^{\text{opt}}$, $\mathbf{A}_{i_k}^{\text{opt}}$, and $\mathbf{W}_{i_k}^{\text{opt}}$) to form a *Robustified WMMSE algorithm* (RB-WMMSE); see Algorithm 2. This algorithm can be combined with any of the channel estimation procedures outlined in Sec. III, and the joint system is fully distributed if the CSI acquisition is distributed.

The existing robust WMMSE algorithms in [7]–[10] also gain their robustness from diagonal loading, obtained by optimizing a lower bound on performance. Although not being directly tailored for TDD channel estimation, these algorithms can be applied together with the centralized CSI acquisition method¹⁰ proposed in Sec. III-C. In doing so, an implicit assumption on the channel estimation errors in the uplink and downlink is made however. Since these algorithms only have a notion of *downlink* channel estimation errors, they are unaware of the *uplink* channel estimation errors in the TDD channel estimation. Thus, the implicit assumption that the channel estimation errors in the downlink and uplink are *identical* is made. The performance of this approach is studied in Sec. V-A2.

¹⁰The algorithms in [7]–[10] need the statistics of the CSI uncertainty, which is very complicated to derive for the CSI acquisition methods in Sec. III-A and Sec. III-B since those methods estimate the *effective channels*.

Algorithm 2 RB-WMMSE Algorithm (Estimated CSI)

-
- 1: **repeat**
- At UEs:
- 2: Pilot transmission from BSs: estimate $\widehat{\mathbf{\Phi}}_{i_k}^{(\rho)}$ and $\widehat{\mathbf{F}}_{i_k}^{(\rho)}$ using one of the methods in Sec. III.
 - 3: Rescale $\widehat{\mathbf{\Phi}}_{i_k}^{(\rho)} = \rho \widehat{\mathbf{\Phi}}_{i_k}^{(\rho)}$, $\widehat{\mathbf{F}}_{i_k}^{(\rho)} = \sqrt{\rho} \widehat{\mathbf{F}}_{i_k}^{(\rho)}$
 - 4: Find ν_{i_k} to satisfy $\left\| \mathbf{A}_{i_k} \mathbf{W}_{i_k}^{1/2} \right\|_F^2 \leq N_d / \sigma_r^2$
 - 5: $\mathbf{W}_{i_k} = \left(\mathbf{I} - \widehat{\mathbf{F}}_{i_k}^H \left(\widehat{\mathbf{\Phi}}_{i_k}^{(\rho)} + \nu_{i_k} \mathbf{I} \right)^{-1} \widehat{\mathbf{F}}_{i_k} \right)^{-1}$
 - 6: $\mathbf{A}_{i_k} = \left(\widehat{\mathbf{\Phi}}_{i_k}^{(\rho)} + \nu_{i_k} \mathbf{I} \right)^{-1} \widehat{\mathbf{F}}_{i_k}$, $\mathbf{U}_{i_k} = \sqrt{\alpha_{i_k}} \mathbf{A}_{i_k} \mathbf{W}_{i_k}^{1/2}$
- At BSs:
- 7: Pilot transmission from UEs: estimate $\widehat{\mathbf{\Gamma}}_i^{s+i+n}$ and $\widehat{\mathbf{G}}_{i_k}$ using one of the methods in Sec. III.
 - 8: Obtain $\mathbf{W}_{i_k}^{1/2}$ through feedback.
 - 9: Find $\mu_i \geq -\min \left(\frac{\sigma_t^2}{\gamma^2}, \lambda_{M_t} \left(\widehat{\mathbf{\Gamma}}_i^{s+i+n} \right) - \zeta \right)$ to satisfy $\sum_{k=1}^{K_c} \text{Tr} \left(\mathbf{V}_{i_k}^{(\rho)} \mathbf{V}_{i_k}^{(\rho),H} \right) \leq \rho P_i$
 - 10: $\mathbf{B}_{i_k}^{(\rho)} = \left(\widehat{\mathbf{\Gamma}}_i^{s+i+n} + \mu_i \mathbf{I} \right)^{-1} \widehat{\mathbf{G}}_{i_k}$, $\mathbf{V}_{i_k}^{(\rho)} = \sqrt{\alpha_{i_k}} \mathbf{B}_{i_k}^{(\rho)} \mathbf{W}_{i_k}^{1/2}$
 - 11: Scale $\mathbf{V}_{i_k} = \frac{1}{\sqrt{\rho}} \mathbf{V}_{i_k}^{(\rho)}$
 - 12: **until** fixed number of iterations
-

V. PERFORMANCE EVALUATION

Performance of the proposed system is evaluated by means of numerical simulations¹¹. Two scenarios are studied:

- 1) A *canonical interfering broadcast channel*, without large scale fading. This model is relevant in local environments where the inter-cell interference power levels are on par with the desired power levels.
- 2) A *large scale 3-cell network*, with path loss, shadow fading, and small scale fading. This models a possible large scale deployment scenario, where only cell-edge users are significantly affected by inter-cell interference.

In both scenarios, we study a case with $K_t = 3$ BSs, each serving K_c users with $N_d = 1$ data streams each. The number of antennas were $M_t = 4$ and $M_r = 2$. The BSs transmit power was $P_i = P_t$ for all BSs, and the UEs transmit power was P_r for all UEs. Unless otherwise stated, the RB-WMMSE BS power scaling was set as $\rho = \min \left(\frac{P_r / \sigma_r^2}{P_t / \sigma_t^2}, 1 \right)$, based on the findings in Fig. 4. For numerical stability, we let the constant $\zeta = 10^{-10}$ such that $\lambda_{M_t} \left(\widehat{\mathbf{\Gamma}}_i^{s+i+n} + \mu_i \mathbf{I} \right) \geq 10^{-10}$, $\forall i$, in the RB-WMMSE algorithm. The UE data rate weights were $\alpha_{i_k} = 1$ for all UEs. Truncated discrete Fourier transform (DFT) matrices of appropriate dimensions were used for the pilot matrices \mathbf{P}_{i_k} and \mathbf{P}_i , as well as for the initial precoders.

As a baseline performance measure, we used single-user eigenprecoding and waterfilling with channels estimated by the MMSE estimator in Sec. III-C. With the single-user processing, we show the performance under time-division multiple access (TDMA), as well as under nonorthogonal concurrent transmissions from all BSs simultaneously ('uncoordinated transmission').

¹¹In the spirit of reproducibility, we provide the full Matlab simulation package as open source. It is available for download at [41].

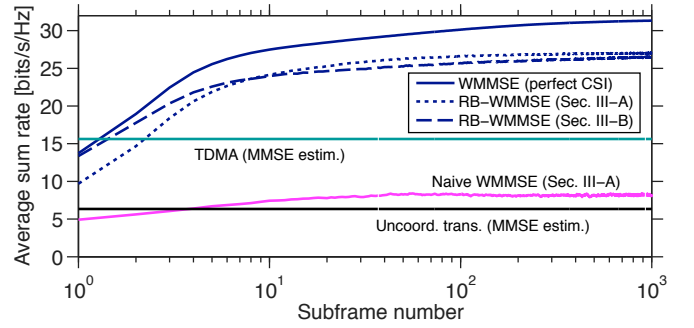


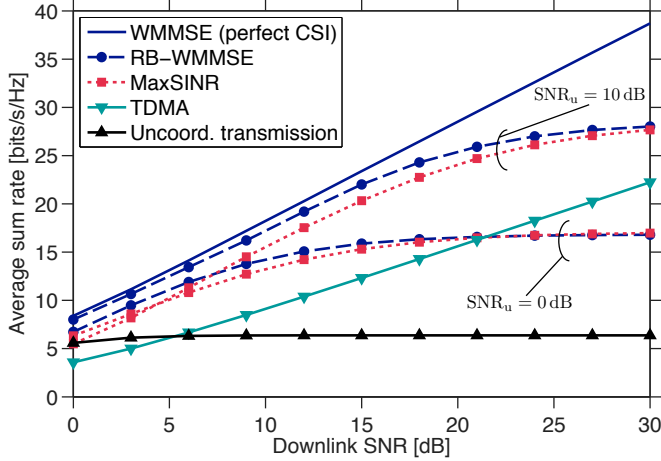
Fig. 6. Convergence comparison of the different methods for $K_t = 3$, $K_c = 2$, $M_t = 4$, $M_r = 2$, $N_d = 1$, $\text{SNR}_d = 20$ dB and $\text{SNR}_u = 10$ dB.

A. Canonical interfering broadcast channel

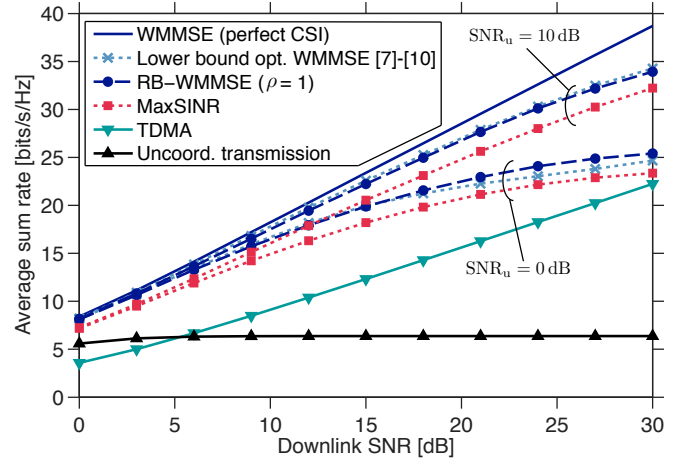
For the simulations with the canonical channel the channel model was i.i.d. Rayleigh fading on all antenna-pairs in the system such that $[\mathbf{H}_{i_k,j}]_{mn} \sim \mathcal{CN}(0, 1)$. This models a setting where each interfering link on average is equally strong as the desired channel. We assume a sufficiently long coherence interval, such that the channels do not change between iterations. We let each BS serve $K_c = 2$ UEs, a setting which is feasible for interference alignment [42]. For fairness when comparing estimation schemes, we let $N_{p,d} = K_t M_t$ and $N_{p,u} = K_r M_r$. The results were averaged over 1000 independent Monte Carlo realizations.

1) *Convergence*: First, we investigate the average convergence behaviour of the RB-WMMSE algorithm with $\text{SNR}_d = P_t / \sigma_r^2 = 20$ dB and $\text{SNR}_u = P_r / \sigma_t^2 = 10$ dB. The results in Fig. 6, indicate that the RB-WMMSE algorithm needs on the order of 1000 iterations to converge, which is consistent with the findings of [11]. We do however note that a significant fraction of the final performance is achieved after just around 10 to 20 iterations. In the following, we therefore let the algorithms iterate for 20 iterations.

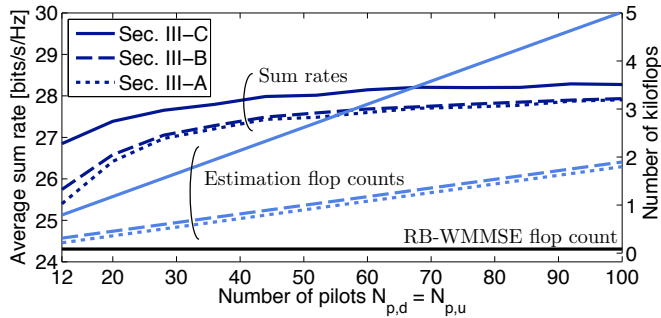
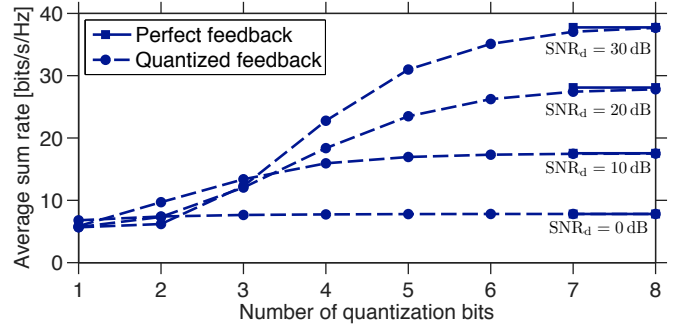
2) *Sum Rate vs. Signal-to-Noise Ratio*: Next, we study the sum rate when the downlink and uplink SNRs are varied. Recall that the downlink SNR affects both the downlink data transmission, as well as the downlink estimation performance (cf. Sec. III-A). The uplink SNR only affects the uplink estimation performance. We compare with MaxSINR [43], for which we actively turn off two users in order not to overload the algorithm. The results for the fully distributed CSI acquisition (Sec. III-A) are shown in Fig. 7a. The RB-WMMSE algorithm consistently performs better than MaxSINR, and better than TDMA for sufficiently high uplink SNR. The results for the CSI acquisition with global sharing of filters and MSE weights (Sec. III-C) are shown in Fig. 7b. Here we also compare with the lower bound optimization method of [7]–[10], which *requires* this form of centralized CSI estimation. We relax their requirement of downlink and uplink estimation errors being identical (cf. Sec. IV-E). In Fig. 7b, it can be seen that the RB-WMMSE algorithm exhibits similar performance as the lower bound optimization method of [7]–[10]. The sum rates in Fig. 7b are higher than the corresponding sum rates in Fig. 7a. This is because the improved channel estimation performance, due to the perfect feedback of filters, and that



(a) Fully distributed CSI acquisition (Sec. III-A)



(b) CSI acquisition with globally shared filters and MSE weights (Sec. III-C)

Fig. 7. Sum rate after the 20th iteration for $K_t = 3, K_c = 2, M_t = 4, M_r = 2$ canonical interfering broadcast channel with $N_d = 1$.Fig. 8. Comparison of complexity and sum rate performance after the 20th iteration for $K_t = 3, K_c = 2, M_t = 4, M_r = 2$ canonical interfering broadcast channel with $N_d = 1, \text{SNR}_d = 20 \text{ dB}$, and $\text{SNR}_u = 10 \text{ dB}$.Fig. 9. Sum rate as a function of quantization accuracy for $K_t = 3, K_c = 2, M_t = 4, M_r = 2$ canonical interfering broadcast channel with $N_d = 1$ and $\text{SNR}_u = 30 \text{ dB}$.

the estimates of the channels are improved in every iteration, as described in Sec. III-C.

3) *Sum Rate and Complexity vs. Flop Count*: For the case with $\text{SNR}_d = 20 \text{ dB}$, and $\text{SNR}_u = 10 \text{ dB}$, we vary the number of pilots $N_{p,d} = N_{p,u}$ and study the resulting performance and complexity of the system. The results can be seen in Fig. 8. The more complex CSI acquisition methods perform slightly better in the sum rate sense. The centralized CSI acquisition from Sec. III-C requires particularly many flops, since it estimates all interfering channels.

4) *Quantized MSE Weight Feedback*: So far, the feedback of the MSE weights was assumed to be perfect. We now study performance of the system, using quantized MSE weights, while varying the number of feedback bits used. For the case with fixed uplink $\text{SNR}_u = 30 \text{ dB}$, we vary the SNR_d and the number of quantization bits. Each UE had an individual codebook with weights uniformly quantized on $\left[0, 10 \log_{10} \left(1 + \frac{P_i \sigma_1^2 (\mathbf{H}_{i,k,i})}{\sigma_2^2}\right)\right]$ dB. The performance is shown in Fig. 9. For higher downlink SNR, more bits are needed for good performance. For high resolution quantization, the performance is equal to that of perfect feedback.

B. Large scale 3-cell network

The results presented so far describe performance in a setting where the desired signal and interfering signals had equal average power levels. In realistic deployments, e.g. macrocell setups, large scale fading such as path loss and shadow fading are present however, leading to a more heterogeneous setting. In order to investigate the performance for such a setting, we study a scenario where the BSs are located at the vertices of an equilateral triangle, and the antenna bore sights are aimed towards the centre of the triangle (see Fig. 10). This scenario models three interfering sectors in a larger hexagonal macrocell deployment. In particular, we assume a setup where fractional frequency reuse is combined with coordinated precoding, such that cell centre and cell edge users are served on orthogonal subbands [44]. Since the cell centre users typically have very high signal-to-interference ratios (SIRs), they can be served well using single cell techniques. The cell edge users experience low SIRs however, and thus multicell coordinated precoding is a fruitful transmission strategy for these users. Since our focus is on coordinated precoding, our simulations only study the performance of the cell edge users. The simulation parameters (see Table V) can be described as a simplified version of the 3GPP Case 1 [45], [46], where the

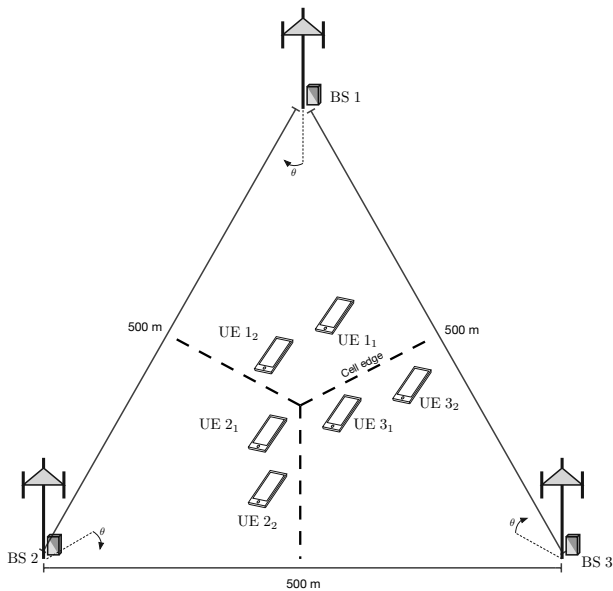


Fig. 10. Cell layout for large scale 3-cell network, here displayed with $K_c = 2$ UEs per cell.

small scale fading is i.i.d. Rayleigh fading, and where we only study one subcarrier.

The purpose of the simulation study is to investigate the impact on sum rate performance, when the number of cell edge users per cell K_c is varied. For each of the two simulations to be described, we generated 500 independent user drops (including shadow fading), where the UEs were dropped uniformly at random in the cell, but never farther than 50 m from the cell edge. For each user drop, 5 independent small scale fading realizations were generated. We used the fully distributed estimation method in Sec. III-A for channel estimation in the RB-WMMSE algorithm and MaxSINR, and we assumed perfect feedback for the MSE weights. In order to have the same estimation performance regardless of K_c , we fixed $N_{p,d} = N_{p,u} = 3 \cdot 10 \cdot 1 = 30$.

We used the same baseline methods as described in Sec. V-A. For large K_c , all baseline methods would perform poorly due to the high interference levels experienced at the cell edge. We therefore coupled the baseline methods with a user selection procedure that determined which UEs to serve. Before describing the main results of the simulation study, we first detail the user selection procedure performance.

1) *Sum Rate vs. Number of Users Selected for Transmission:* In order to study how many users to select for transmission for the baseline methods, we performed simulations where $K_c = 10$ users were dropped per cell, and the number of users selected for transmission was varied. The user selection was based solely on the channel strength to the serving BS. The results are plotted in Fig. 11a. It can be seen that performance for MaxSINR is maximized when 2 users are selected for transmission. For TDMA and uncoordinated transmission, performance is maximized when only a single user is selected for transmission in each cell. The performance of the RB-WMMSE algorithm is the highest when 3 users are selected for transmission in each cell, but performance only

TABLE V
SIMULATION PARAMETERS FOR LARGE SCALE 3-CELL NETWORK
(ADAPTED FROM 3GPP CASE 1 [45], [46])

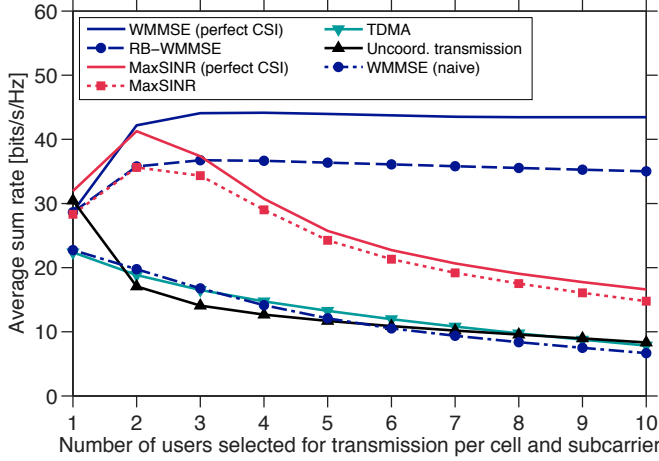
Inter site distance	500 m
Max. distance, UE to cell edge	50 m
Path loss	$15.3 + 37.6 \log_{10}(\text{distance [m]})$ dB
Penetration loss	20 dB
BS antenna gain [47]	$-\min\left(12\left(\frac{\theta}{35^\circ}\right)^2, 23\right)$ dB
UE antenna gain	0 dB
Shadow fading	Lognormal with std. dev. 8 dB
Small scale fading	i.i.d. $\mathcal{CN}(0, 1)$
Bandwidth*	15 kHz
BS transmit power*	$P_t = 18.2$ dBm
UE receiver noise power*	$\sigma_r^2 = -123.2$ dBm (9 dB NF)
UE transmit power*	$P_r = -4.8$ dBm
BS receiver noise power*	$\sigma_t^2 = -127.2$ dBm (5 dB NF)

* We only study one subcarrier in a 10 MHz system with 600 subcarriers. The total transmit powers are thus $P_t^{\text{tot}} = 46$ dBm and $P_r^{\text{tot}} = 23$ dBm.

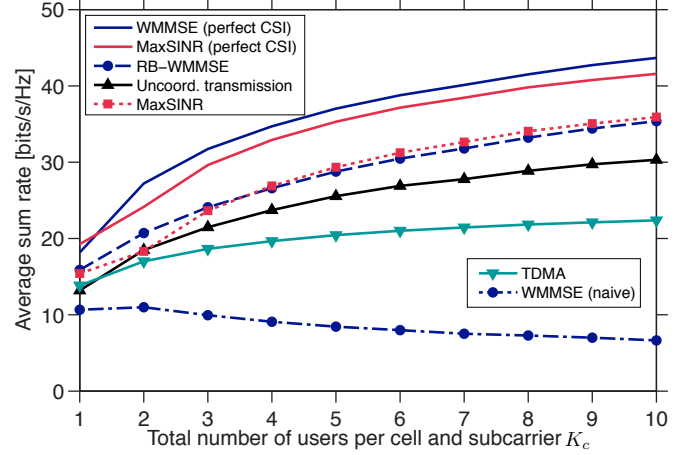
slightly drops as more users are selected. This is because the RB-WMMSE algorithm, just like the WMMSE algorithm, is able to implicitly perform user selection in the iterations. The fact that the performance is almost constant when more than 3 users are selected for transmission in each cell suggests that the RB-WMMSE algorithm is able to find a good local solution to the weighted sum rate problem.

2) *Sum Rate vs. Total Number of Users per Cell:* We now study system performance as a function of the total number of users per cell K_c . Given the results in the previous section, we select at most 2 users for transmission per cell for MaxSINR. Similarly for TDMA and uncoordinated transmission, disregarding the obvious unfairness of such a strategy, we only select 1 user per cell. For the RB-WMMSE and WMMSE algorithms, we do not explicitly perform any user selection, but rather let the algorithms perform their own implicit user selection in the iterations. The results of the simulation can be seen in Fig. 11b. For the baselines with user selection, the improved performance with K_c is due to the increased multiuser diversity. The RB-WMMSE algorithm is also able to harness this increase in multiuser diversity. The large gap between the perfect CSI case, and the RB-WMMSE algorithm with fully distributed CSI acquisition (Sec. III-A) is due to the low uplink SNR in this scenario¹². The RB-WMMSE algorithm performs slightly worse than the MaxSINR with user selection, for large K_c . It has however been verified that this gap can be closed by combining the RB-WMMSE algorithm together with explicit user selection, serving at most 3 users per cell (cf. Fig. 11a). Here we however show the results without explicit user selection, in order to display the self-reliant performance of the algorithm. The uncoordinated transmission strategy works fairly well in terms of sum rate,

¹²Along the cell edge, the average uplink SNR ranges between -4.9 dB and -11.8 dB. The corresponding average downlink SNR ranges between 7.2 dB and 14.1 dB. Note that the average SNRs are determined both by the distance to the BS, as well as the angle to the antenna bore sight.



(a) Varying the number of users selected for transmission, given $K_c = 10$ users per cell.



(b) Varying the total number of users per cell K_c . At most 2 users per cell were served by MaxSINR. For TDMA and uncoordinated transmission, only 1 user per cell was served.

Fig. 11. Sum rate after the 20th iteration for the large scale scenario with fully distributed CSI acquisition (Sec. III-A).

but as noted earlier, corresponds to a highly unfair situation where only one user is served per cell.

VI. CONCLUSIONS

Many distributed coordinated precoding algorithms have been proposed in the literature, but few of these works study the issue of robustness against imperfect CSI. The works that do study this important issue can however not be coupled with distributed CSI acquisition, thus leading to centralized CSI acquisition requiring large amounts of BS backhaul usage. To our knowledge, the present paper is the first that proposes a robust, and yet still fully distributed, coordinated precoding algorithm. In doing so, three CSI acquisition methods have been proposed, and the corresponding requirements in terms of channel estimation, feedback, and signaling have been illuminated. The robustification of the algorithm comes from using inherent properties of the WMMSE solutions and applying a data dependent scaling procedure, leading to a realizable algorithm which does not depend on unknown parameters of the CSI uncertainty. Compared with the centralized state-of-the-art methods, our system performs similarly, with the major difference that our system can be implemented in a fully distributed manner. When evaluated in a macrocell setup, the proposed system also performs well.

APPENDIX

A. Proof of Lemma 1

For UE i_k it holds that $\Phi_{i_k}^{i+n} \succeq \sigma_r^2 \mathbf{I}$, with equality if the UE does not experience any interference. Thus, the MSE weight for that UE satisfies

$$\mathbf{W}_{i_k} = \mathbf{I} + \mathbf{V}_{i_k}^H \mathbf{H}_{i_k}^H (\Phi_{i_k}^{i+n})^{-1} \mathbf{H}_{i_k} \mathbf{V}_{i_k} \quad (30)$$

$$\leq \mathbf{I} + \frac{1}{\sigma_r^2} \mathbf{V}_{i_k}^H \mathbf{H}_{i_k}^H \mathbf{H}_{i_k} \mathbf{V}_{i_k}. \quad (31)$$

Now introduce the spectral norm $\|\mathbf{D}\|_2 = \max_c \frac{\|\mathbf{D}\mathbf{c}\|_2}{\|\mathbf{c}\|_2} = s_1(\mathbf{D})$. Then, for all \mathbf{c}_{i_k} such that $\|\mathbf{c}_{i_k}\|_2 = 1$, we have that

$$\begin{aligned} \mathbf{c}_{i_k}^H \mathbf{V}_{i_k}^H \mathbf{H}_{i_k}^H \mathbf{H}_{i_k} \mathbf{V}_{i_k} \mathbf{c}_{i_k} &\leq \|\mathbf{V}_{i_k} \mathbf{c}_{i_k}\|_2^2 \cdot \lambda_1(\mathbf{H}_{i_k}^H \mathbf{H}_{i_k}) \\ &= \|\mathbf{V}_{i_k} \mathbf{c}_{i_k}\|_2^2 \cdot s_1^2(\mathbf{H}_{i_k}) \leq \|\mathbf{V}_{i_k}\|_2^2 \cdot \|\mathbf{c}_{i_k}\|_2^2 \cdot s_1^2(\mathbf{H}_{i_k}) \\ &= \|\mathbf{V}_{i_k}\|_2^2 \cdot s_1^2(\mathbf{H}_{i_k}) \leq \|\mathbf{V}_{i_k}\|_F^2 \cdot s_1^2(\mathbf{H}_{i_k}) \leq P_i s_1^2(\mathbf{H}_{i_k}). \end{aligned} \quad (32)$$

Thus, $\lambda_1(\mathbf{V}_{i_k}^H \mathbf{H}_{i_k}^H \mathbf{H}_{i_k} \mathbf{V}_{i_k}) \leq P_i s_1^2(\mathbf{H}_{i_k})$, and the upper bound then directly follows from (31). For the lower bound, note that $\mathbf{I} \preceq \mathbf{I} + \mathbf{V}_{i_k}^H \mathbf{H}_{i_k}^H (\Phi_{i_k}^{i+n})^{-1} \mathbf{H}_{i_k} \mathbf{V}_{i_k} = \mathbf{W}_{i_k}$. ■

B. Proof of Proposition 1

Decompose $\Phi_{i_k} = \mathbf{F}_{i_k} \mathbf{F}_{i_k}^H + \Phi_{i_k}^{i+n}$ and let $\mathbf{C}_{i_k} = \mathbf{F}_{i_k}^H (\Phi_{i_k}^{i+n})^{-1} \mathbf{F}_{i_k}$ and $\mathbf{D}_{i_k} = \mathbf{F}_{i_k}^H (\Phi_{i_k}^{i+n})^{-2} \mathbf{F}_{i_k}$. We have that $\mathbf{A}_{i_k} = \Phi_{i_k}^{-1} \mathbf{F}_{i_k}$ and $\mathbf{W}_{i_k} = \mathbf{I} + \mathbf{C}_{i_k}$. Plugging in, $\|\mathbf{A}_{i_k} \mathbf{W}_{i_k}^{1/2}\|_F^2 = \text{Tr}(\mathbf{A}_{i_k} \mathbf{W}_{i_k} \mathbf{A}_{i_k}^H) = \text{Tr}(\mathbf{A}_{i_k}^H \mathbf{A}_{i_k} \mathbf{W}_{i_k}) = \text{Tr}(\mathbf{F}_{i_k}^H \Phi_{i_k}^{-2} \mathbf{F}_{i_k} (\mathbf{I} + \mathbf{C}_{i_k}))$. Applying the matrix inversion lemma to $\Phi_{i_k}^{-1}$, it can be shown that $\mathbf{F}_{i_k}^H \Phi_{i_k}^{-2} \mathbf{F}_{i_k} = (\mathbf{I} + \mathbf{C}_{i_k})^{-1} \mathbf{D}_{i_k} (\mathbf{I} + \mathbf{C}_{i_k})^{-1}$ after simplifications. Thus, $\|\mathbf{A}_{i_k} \mathbf{W}_{i_k}^{1/2}\|_F^2 = \text{Tr}((\mathbf{I} + \mathbf{C}_{i_k})^{-1} \mathbf{D}_{i_k}) = \text{Tr}((\Phi_{i_k}^{i+n})^{-1} \mathbf{F}_{i_k} (\mathbf{I} + \mathbf{F}_{i_k}^H (\Phi_{i_k}^{i+n})^{-1} \mathbf{F}_{i_k})^{-1} \mathbf{F}_{i_k}^H (\Phi_{i_k}^{i+n})^{-1})$. Applying the matrix inversion lemma backwards, we then get $\|\mathbf{A}_{i_k} \mathbf{W}_{i_k}^{1/2}\|_F^2 = \text{Tr}((\Phi_{i_k}^{i+n})^{-1} - \Phi_{i_k}^{-1})$. Further,

$$\begin{aligned} \|\mathbf{A}_{i_k} \mathbf{W}_{i_k}^{1/2}\|_F^2 &= \text{Tr}((\mathbf{I} + \mathbf{C}_{i_k})^{-1} \mathbf{D}_{i_k}) \stackrel{(a)}{\leq} \text{Tr}(\mathbf{C}_{i_k}^{-1} \mathbf{D}_{i_k}) \\ &= \text{Tr}\left(\left(\mathbf{F}_{i_k}^H (\Phi_{i_k}^{i+n})^{-1} \mathbf{F}_{i_k}\right)^{-1} \mathbf{F}_{i_k}^H (\Phi_{i_k}^{i+n})^{-2} \mathbf{F}_{i_k}\right) \\ &= \left[\text{Let } \tilde{\mathbf{F}}_{i_k} = (\Phi_{i_k}^{i+n})^{-1/2} \mathbf{F}_{i_k}\right] \\ &= \text{Tr}\left(\left(\tilde{\mathbf{F}}_{i_k}^H \tilde{\mathbf{F}}_{i_k}\right)^{-1} \tilde{\mathbf{F}}_{i_k}^H (\Phi_{i_k}^{i+n})^{-1} \tilde{\mathbf{F}}_{i_k}\right) \\ &= \text{Tr}\left(\left(\Phi_{i_k}^{i+n}\right)^{-1} \tilde{\mathbf{F}}_{i_k} \left(\tilde{\mathbf{F}}_{i_k}^H \tilde{\mathbf{F}}_{i_k}\right)^{-1} \tilde{\mathbf{F}}_{i_k}^H\right) \end{aligned}$$

$$\begin{aligned} &\stackrel{(b)}{\leq} \max_{\text{rank}(\mathbf{\Pi}_{i_k})=N_d} \text{Tr} \left((\mathbf{\Phi}_{i_k}^{i+n})^{-1} \mathbf{\Pi}_{i_k} \right) \\ &\leq N_d \lambda_1 \left((\mathbf{\Phi}_{i_k}^{i+n})^{-1} \right) = \frac{N_d}{\lambda_{M_r}(\mathbf{\Phi}_{i_k}^{i+n})} \leq \frac{N_d}{\sigma_r^2} \end{aligned}$$

where $\mathbf{\Pi}_{i_k}$ is a rank- N_d projection matrix. The inequality (a) is due to the trace being an increasing function on the cone of positive definite matrices and the fact that $\mathbf{D}_{i_k}^{1/2} (\mathbf{I} + \mathbf{C}_{i_k})^{-1} \mathbf{D}_{i_k}^{1/2} \preceq \mathbf{D}_{i_k}^{1/2} \mathbf{C}_{i_k}^{-1} \mathbf{D}_{i_k}^{1/2}$. The inequality (b) holds since $\tilde{\mathbf{F}}_{i_k} (\tilde{\mathbf{F}}_{i_k}^H \tilde{\mathbf{F}}_{i_k})^{-1} \tilde{\mathbf{F}}_{i_k}^H$ is a rank- N_d projection matrix.

Now assume there are $N_s \leq N_d$ interference-free dimensions, and that the effective channel is fully contained in those. Let the eigenvalues of $\mathbf{\Phi}_{i_k}$ be $\{\kappa_m^{s+i+n}\}$ and the eigenvalues of $\mathbf{\Phi}_{i_k}^{i+n}$ be $\{\kappa_m^{i+n}\}$. Let the eigenvalues be ordered such that $\kappa_m^{s+i+n} = \kappa_m^{i+n}$ for all $m \in \{1, \dots, M_r - N_s\}$. Consequently, $\kappa_m^{s+i+n} = \kappa_m^s + \sigma_r^2$ and $\kappa_m^{i+n} = \sigma_r^2$ for all $m \in \{M_r - N_s + 1, \dots, M_r\}$. Here, $\{\kappa_m^s\}$ are the received signal powers in the interference-free subspace. Then,

$$\begin{aligned} \text{Tr} \left((\mathbf{\Phi}_{i_k}^{i+n})^{-1} - \mathbf{\Phi}_{i_k}^{-1} \right) &= \sum_{m=1}^{M_r} \left(\frac{1}{\kappa_m^{i+n}} - \frac{1}{\kappa_m^{s+i+n}} \right) \\ &= \underbrace{\sum_{m=1}^{M_r - N_s} \left(\frac{1}{\kappa_m^{i+n}} - \frac{1}{\kappa_m^{i+n}} \right)}_{\text{Interference dimensions}} + \underbrace{\sum_{m=M_r - N_s + 1}^{M_r} \left(\frac{1}{\sigma_r^2} - \frac{1}{\kappa_m^s + \sigma_r^2} \right)}_{\text{Interference-free dimensions}} \\ &= \sum_{m=M_r - N_s + 1}^{M_r} \frac{\kappa_m^s}{\sigma_r^2 (\kappa_m^s + \sigma_r^2)} \rightarrow \sum_{m=M_r - N_s + 1}^{M_r} \frac{1}{\sigma_r^2} = \frac{N_s}{\sigma_r^2} \end{aligned}$$

as the $\{\kappa_m^s\}$ grow large with respect to σ_r^2 . ■

C. Proof of Proposition 2

It can be shown that all feasible points are regular, and thus any minimizer of this non-convex problem satisfies the KKT conditions [30, Ch. 3.3.1]. The expressions are obtained by forming the Lagrangian $\mathcal{L}_{i_k}(\mathbf{A}_{i_k}, \mathbf{W}_{i_k}, \nu_{i_k}) = \text{Tr} \left(\mathbf{W}_{i_k} (\mathbf{I} + \mathbf{A}_{i_k}^H \hat{\mathbf{\Phi}}_{i_k} \mathbf{A}_{i_k}) \right) - 2\text{Re} \left(\mathbf{W}_{i_k} \hat{\mathbf{F}}_{i_k}^H \mathbf{A}_{i_k} \right) - \log \det (\mathbf{W}_{i_k}) + \nu_{i_k} \left(\text{Tr} (\mathbf{A}_{i_k} \mathbf{W}_{i_k} \mathbf{A}_{i_k}^H) - \frac{N_d}{\sigma_r^2} \right)$, setting the complex partial derivatives to zero, and applying the matrix inversion lemma. It now remains to find the optimal $\nu_{i_k}^{\text{opt}} \geq 0$. If the constraint is satisfied for $\nu_{i_k}^{\text{opt}} = 0$, the problem is solved and the form is identical to the solution in Sec. II-A1. Otherwise, let $\hat{\mathbf{\Phi}}_{i_k} = \mathbf{L}_{i_k} \mathbf{\Lambda}_{i_k} \mathbf{L}_{i_k}^H$ and $\hat{\mathbf{\Phi}}_{i_k}^{i+n} = \mathbf{L}_{i_k}^{i+n} \mathbf{\Lambda}_{i_k}^{i+n} \mathbf{L}_{i_k}^{i+n,H}$ be eigenvalue decompositions. Then as can be seen in (33), $\|\mathbf{A}_{i_k} \mathbf{W}_{i_k}^{1/2}\|_F^2$ is decreasing in ν_{i_k} , and the $\nu_{i_k}^{\text{opt}}$ which satisfies the inequality constraint with equality can be found by bisection. A natural starting point for the lower value in the bisection is $\nu_{i_k}^{\text{lower}} = 0$. Using the same argument as in the proof for Proposition 1, we have that $\|\mathbf{A}_{i_k} \mathbf{W}_{i_k}^{1/2}\|_F^2 \leq \frac{N_d}{\lambda_{M_r}(\hat{\mathbf{\Phi}}_{i_k}^{i+n} + \nu_{i_k} \mathbf{I})}$ and we can enforce $\|\mathbf{A}_{i_k} \mathbf{W}_{i_k}^{1/2}\|_F^2 \Big|_{\nu_{i_k} = \nu_{i_k}^{\text{upper}}} \leq \frac{N_d}{\sigma_r^2}$ with $\nu_{i_k}^{\text{upper}} = \sigma_r^2$. The optimal $\nu_{i_k}^{\text{opt}}$ can now be found using bisection on (33), given the bounds. With $\nu_{i_k}^{\text{opt}}$ found, the minimizer can be identified as

unique. The fact that $(\mathbf{A}_{i_k}^{\text{opt}}, \mathbf{W}_{i_k}^{\text{opt}})$ is indeed a minimizer is clear, since each variable minimizes the objective function when the other variable is kept fixed. ■

REFERENCES

- [1] D. Gesbert, S. Hanly, H. Huang, S. Shamai Shitz, O. Simeone, and W. Yu, "Multi-cell MIMO cooperative networks: A new look at interference," *IEEE J. Sel. Areas Commun.*, vol. 28, no. 9, pp. 1380–1408, 2010.
- [2] E. Björnson and E. Jorswieck, "Optimal resource allocation in coordinated multi-cell systems," *Foundations and Trends in Communications and Information Theory*, vol. 9, no. 2–3, pp. 113–381, 2013.
- [3] J. Jose, A. Ashikhmin, P. Whiting, and S. Vishwanath, "Channel estimation and linear precoding in multiuser multiple-antenna TDD systems," *IEEE Trans. Veh. Technol.*, vol. 60, no. 5, pp. 2102–2116, 2011.
- [4] C. Shi, R. Berry, and M. Honig, "Bi-directional training for adaptive beamforming and power control in interference networks," *IEEE Trans. Signal Process.*, vol. 62, no. 3, pp. 607–618, Feb. 2014.
- [5] M. Biguesh and A. B. Gershman, "Training-based MIMO channel estimation: a study of estimator tradeoffs and optimal training signals," *IEEE Trans. Signal Process.*, vol. 54, no. 3, 2006.
- [6] J. Jose, N. Prasad, M. Khojastepour, and S. Rangarajan, "On robust weighted-sum rate maximization in MIMO interference networks," in *Proc. IEEE Int. Conf. Commun. (ICC'11)*, Jun. 2011, pp. 1–6.
- [7] J. Shin and J. Moon, "Weighted-sum-rate-maximizing linear transceiver filters for the K-user MIMO interference channel," *IEEE Trans. Commun.*, vol. 60, no. 10, pp. 2776–2783, Oct. 2012.
- [8] H.-H. Lee, Y.-C. Ko, and H.-C. Yang, "On robust weighted sum rate maximization for MIMO interfering broadcast channels with imperfect channel knowledge," *IEEE Commun. Lett.*, vol. 17, no. 6, pp. 1156–1159, 2013.
- [9] M. Razaviyayn, M. Baligh, A. Callard, and Z.-Q. Luo, "Robust transceiver design," WO Patent Application WO 2013/044 824 A1, 04, 2013. [Online]. Available: <http://patentscope.wipo.int/search/en/detail.jsf?docId=WO2013044824>
- [10] F. Negro, I. Ghauri, and D. T. M. Stock, "Sum rate maximization in the noisy MIMO interfering broadcast channel with partial CSIT via the expected weighted MSE," in *Proc. Int. Symp. Wireless Commun. Systems (ISWCS'12)*, 2012, pp. 576–580.
- [11] D. Schmidt, C. Shi, R. Berry, M. Honig, and W. Utschick, "Comparison of distributed beamforming algorithms for MIMO interference networks," *IEEE Trans. Signal Process.*, vol. 61, no. 13, pp. 3476–3489, 2013.
- [12] Q. Shi, M. Razaviyayn, Z. Luo, and C. He, "An iteratively weighted MMSE approach to distributed sum-utility maximization for a MIMO interfering broadcast channel," *IEEE Trans. Signal Process.*, vol. 59, no. 9, pp. 4331–4340, 2011.
- [13] J. Ren, K.-K. Wong, and J. Hou, "Robust analytical ZF optimization for MISO IFC," *IEEE Wireless Commun. Lett.*, vol. 2, no. 4, pp. 451–454, 2013.
- [14] P. Komulainen, A. Tölli, and M. Juntti, "Effective CSI signaling and decentralized beam coordination in TDD multi-cell MIMO systems," *IEEE Trans. Signal Process.*, vol. 61, no. 9, pp. 2204–2218, May 2013.
- [15] M. Guillaud, D. T. M. Stock, and R. Knopp, "A practical method for wireless channel reciprocity exploitation through relative calibration," in *Proc. 8th Int. Symp. Signal Process. Applicat. (ISSPA'05)*, 2005, pp. 403–406.
- [16] R. Rogalin, O. Bursalioglu, H. Papadopoulos, G. Caire, A. Molisch, A. Michaloliakos, V. Balan, and K. Psounis, "Scalable synchronization and reciprocity calibration for distributed multiuser MIMO," *IEEE Trans. Wireless Commun.*, vol. 13, no. 4, pp. 1815–1831, Apr. 2014.
- [17] H. Cox, R. M. Zeskind, and M. M. Owen, "Robust adaptive beamforming," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 35, no. 10, pp. 1365–1376, 1987.
- [18] B. D. Carlson, "Covariance matrix estimation errors and diagonal loading in adaptive arrays," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 24, no. 4, pp. 397–401, Jul. 1988.
- [19] R. Wu, Z. Bao, and Y. Ma, "Control of peak sidelobe level in adaptive arrays," *IEEE Trans. Antennas Propag.*, vol. 44, no. 10, pp. 1341–1347, 1996.
- [20] J. Li, P. Stoica, and Z. Wang, "On robust Capon beamforming and diagonal loading," *IEEE Trans. Signal Process.*, vol. 51, no. 7, pp. 1702–1715, 2003.

$$\begin{aligned}
& \left\| \mathbf{A}_{i_k} \mathbf{W}_{i_k}^{1/2} \right\|_F^2 = \text{Tr} \left(\mathbf{A}_{i_k} \mathbf{W}_{i_k} \mathbf{A}_{i_k}^H \right) = \text{Tr} \left(\widehat{\mathbf{F}}_{i_k}^H \left(\widehat{\boldsymbol{\Phi}}_{i_k} + \nu_{i_k} \mathbf{I} \right)^{-2} \widehat{\mathbf{F}}_{i_k} \left(\mathbf{I} + \widehat{\mathbf{F}}_{i_k}^H \left(\widehat{\boldsymbol{\Phi}}_{i_k} + \nu_{i_k} \mathbf{I} \right)^{-1} \widehat{\mathbf{F}}_{i_k} \right) \right) = \\
& \text{Tr} \left(\mathbf{L}_{i_k}^H \widehat{\mathbf{F}}_{i_k} \widehat{\mathbf{F}}_{i_k}^H \mathbf{L}_{i_k} \left(\boldsymbol{\Lambda}_{i_k} + \nu_{i_k} \mathbf{I} \right)^{-2} + \left(\boldsymbol{\Lambda}_{i_k} + \nu_{i_k} \mathbf{I} \right)^{-1/2} \mathbf{L}_{i_k}^{i+n, H} \widehat{\mathbf{F}}_{i_k} \widehat{\mathbf{F}}_{i_k}^H \mathbf{L}_{i_k} \left(\boldsymbol{\Lambda}_{i_k} + \nu_{i_k} \mathbf{I} \right)^{-2} \mathbf{L}_{i_k}^H \widehat{\mathbf{F}}_{i_k} \widehat{\mathbf{F}}_{i_k}^H \mathbf{L}_{i_k}^{i+n} \left(\boldsymbol{\Lambda}_{i_k} + \nu_{i_k} \mathbf{I} \right)^{-1/2} \right) \\
& = \sum_{m=1}^{M_r} \frac{\left[\mathbf{L}_{i_k}^H \widehat{\mathbf{F}}_{i_k} \widehat{\mathbf{F}}_{i_k}^H \mathbf{L}_{i_k} \right]_{mm}}{\left(\left[\boldsymbol{\Lambda}_{i_k} \right]_{mm} + \nu_{i_k} \right)^2} + \left\| \left(\boldsymbol{\Lambda}_{i_k} + \nu_{i_k} \mathbf{I} \right)^{-1} \mathbf{L}_{i_k}^H \widehat{\mathbf{F}}_{i_k} \widehat{\mathbf{F}}_{i_k}^H \mathbf{L}_{i_k}^{i+n} \left(\boldsymbol{\Lambda}_{i_k} + \nu_{i_k} \mathbf{I} \right)^{-1/2} \right\|_F^2 \\
& = \sum_{m=1}^{M_r} \frac{\left[\mathbf{L}_{i_k}^H \widehat{\mathbf{F}}_{i_k} \widehat{\mathbf{F}}_{i_k}^H \mathbf{L}_{i_k} \right]_{mm}}{\left(\left[\boldsymbol{\Lambda}_{i_k} \right]_{mm} + \nu_{i_k} \right)^2} + \sum_{m=1}^{M_r} \sum_{p=1}^{M_r} \frac{1}{\left(\left[\boldsymbol{\Lambda}_{i_k} \right]_{mm} + \nu_{i_k} \right)^2} \frac{1}{\left(\left[\boldsymbol{\Lambda}_{i_k}^{i+n} \right]_{pp} + \nu_{i_k} \right)} \left| \left[\mathbf{L}_{i_k}^H \widehat{\mathbf{F}}_{i_k} \widehat{\mathbf{F}}_{i_k}^H \mathbf{L}_{i_k}^{i+n} \right]_{mp} \right|^2
\end{aligned} \tag{33}$$

- [21] S. Vorobyov, A. Gershman, and Z.-Q. Luo, "Robust adaptive beamforming using worst-case performance optimization: a solution to the signal mismatch problem," *IEEE Trans. Signal Process.*, vol. 51, no. 2, pp. 313–324, Feb. 2003.
- [22] S. Shahbazpanahi, A. Gershman, Z.-Q. Luo, and K. M. Wong, "Robust adaptive beamforming for general-rank signal models," *IEEE Trans. Signal Process.*, vol. 51, no. 9, pp. 2257–2269, Sep. 2003.
- [23] K. Zarifi, S. Shahbazpanahi, A. Gershman, and Z.-Q. Luo, "Robust blind multiuser detection based on the worst-case performance optimization of the MMSE receiver," *IEEE Trans. Signal Process.*, vol. 53, no. 1, pp. 295–305, Jan. 2005.
- [24] M. Xu, D. Guo, and M. L. Honig, "Joint bi-directional training of nonlinear precoders and receivers in cellular networks," *arXiv:1408.6924 [cs.IT]*, Aug. 2014.
- [25] B. Ghimire, G. Auer, and H. Haas, "Busy burst enabled coordinated multipoint network with decentralized control," *IEEE Trans. Wireless Commun.*, vol. 10, no. 10, pp. 3310–3320, 2011.
- [26] S. Christensen, R. Agarwal, E. Carvalho, and J. 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.
- [27] F. Negro, S. P. Shenoy, I. Ghauri, and D. T. M. Slock, "Weighted sum rate maximization in the MIMO interference channel," *Proc. IEEE Int. Symp. Personal, Indoor, Mobile Radio Commun. (PIMRC'10)*, pp. 684–689, 2010.
- [28] F. P. Kelly, A. K. Maulloo, and D. K. H. Tan, "Rate control for communication networks: Shadow prices, proportional fairness and stability," *J. Operational Research Soc.*, vol. 49, no. 3, pp. 237–252, 1998.
- [29] Y.-F. Liu, Y.-H. Dai, and Z.-Q. Luo, "Coordinated beamforming for MISO interference channel: Complexity analysis and efficient algorithms," *IEEE Trans. Signal Process.*, vol. 59, no. 3, pp. 1142–1157, Mar. 2011.
- [30] D. Bertsekas, *Nonlinear programming*. Athena Scientific, 2006.
- [31] M. Razaviyayn, M. Hong, and Z. Luo, "A unified convergence analysis of block successive minimization methods for nonsmooth optimization," *SIAM J. Optimization*, vol. 23, no. 2, pp. 1126–1153, 2013.
- [32] T. A. Levanen, J. Pirskanen, T. Koskela, J. Talvitie, and M. Valkama, "Radio interface evolution towards 5G and enhanced local area communications," *IEEE Access*, vol. 2, pp. 1005–1029, 2014.
- [33] N. Jindal and A. Lozano, "A unified treatment of optimum pilot overhead in multipath fading channels," *IEEE Trans. Commun.*, vol. 58, no. 10, pp. 2939–2948, Oct. 2010.
- [34] E. Lähäkangas, K. Pajukoski, E. Tirola, G. Berardinelli, I. Harjula, and J. Vihriälä, "On the TDD subframe structure for beyond 4G radio access network," in *Future Network and Mobile Summit*, Jul. 2013, pp. 1–10.
- [35] J. Jose, A. Ashikhmin, T. L. Marzetta, and S. Vishwanath, "Pilot contamination and precoding in multi-cell TDD systems," *IEEE Trans. Wireless Commun.*, vol. 10, no. 8, pp. 2640–2651, 2011.
- [36] E. Dahlman, S. Parkvall, and J. Sköld, *4G LTE/LTE-Advanced for Mobile Broadband*. Academic Press, 2011.
- [37] H. Cox, "Resolving power and sensitivity to mismatch of optimum array processors," *J. Acoustical Soc. America*, vol. 54, no. 3, pp. 771–785, 1973.
- [38] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*. Prentice Hall, 1993.
- [39] L. N. Trefethen and D. Bau III, *Numerical Linear Algebra*. SIAM, 1997.
- [40] D. Love and R. Heath, "Limited feedback unitary precoding for spatial multiplexing systems," *IEEE Trans. Inf. Theory*, vol. 51, no. 8, pp. 2967–2976, 2005.
- [41] R. Brandt, "Simulation Matlab code." [Online]. Available: http://github.com/rasmusbrandt/tvt2015_distributed_CSI_acquisition
- [42] T. Liu and C. Yang, "On the feasibility of linear interference alignment for MIMO interference broadcast channels with constant coefficients," *IEEE Trans. Signal Process.*, vol. 61, no. 9, pp. 2178–2191, 5 2013.
- [43] K. Gomadam, V. R. Cadambe, and S. Jafar, "A distributed numerical approach to interference alignment and applications to wireless interference networks," *IEEE Trans. Inf. Theory*, vol. 57, no. 6, pp. 3309–3322, 2011.
- [44] L.-C. Wang and C.-J. Yeh, "3-cell network MIMO architectures with sectorization and fractional frequency reuse," *IEEE J. Sel. Areas Commun.*, vol. 29, no. 6, pp. 1185–1199, Jun. 2011.
- [45] 3GPP, "TR 25.814, Physical layer aspects for evolved universal terrestrial radio access (Release 7)," 3GPP, Tech. Rep., 2006.
- [46] —, "TR 36.814, Further advancements for E-UTRA physical layer aspects (Release 9)," 3GPP, Tech. Rep., 2010.
- [47] —, "TR 25.996, Spatial channel model for multiple input multiple output (MIMO) simulations (Release 11)," 3GPP, Tech. Rep., 2012.



Rasmus Brandt (S'10) was born in Uppsala, Sweden, in 1985. He received the Tech.Lic. degree in Electrical Engineering from KTH Royal Institute of Technology, Stockholm, Sweden, in 2014 and the M.Sc. degree in Engineering Physics from Uppsala University, Uppsala, Sweden, in 2010. As part of his M.Sc. program, he spent the academic year 2007/2008 at Queen's University, Kingston, Canada and the summer of 2009 as an IAESTE Intern at CMC Microsystems, Kingston, Canada. He is currently working towards the Ph.D. degree in Electrical

Engineering at the KTH Royal Institute of Technology. His research interests include statistical signal processing for wireless communications as well as resource allocation and interference alignment for interference management. Mr. Brandt is a Publicity Chair of the 16th IEEE International Workshop on Signal Processing Advances in Wireless Communications (SPAWC'15).



Mats Bengtsson (M'00–SM'06) received the M.S. degree in computer science from Linköping University, Linköping, Sweden, in 1991 and the Tech. Lic. and Ph.D. degrees in electrical engineering from the KTH Royal Institute of Technology, Stockholm, Sweden, in 1997 and 2000, respectively.

From 1991 to 1995, he was with Ericsson Telecom AB Karlstad. He currently holds a position as Associate Professor at the Signal Processing department, School of Electrical Engineering, KTH. His research interests include statistical signal processing and its

applications to communications, multi-antenna processing, cooperative communication, radio resource management, and propagation channel modelling. Dr. Bengtsson served as Associate Editor for the IEEE Transactions on Signal Processing 2007-2009 and was a member of the IEEE SPCOM Technical Committee 2007-2012.