Van Rompaey and Moonen
*J Wireless Com Network* (2023) 2023:48
https://doi.org/10.1186/s13638-023-02251-6

EURASIP Journal on Wireless Communications and Networking

RESEARCH Open Access

# Distributed MMSE-based uplink receive combining, downlink transmit precoding and optimal power allocation in cell-free massive MIMO systems

Robbe Van Rompaey[1*] and Marc Moonen[1]

*Correspondence:
robbe.vanrompaey@esat.kuleuven.be

[1] Department of Electrical Engineering, KU Leuven, Kasteelpark Arenberg 10 - bus 2446, 3001 Leuven, Belgium

## Abstract

In cell-free massive multiple-input multiple-output systems, a large number of distributed wireless access points (AP) simultaneously serve a number of user equipments (UEs). This setup has recently been introduced as a promising alternative for the current 5G cellular networks. The setup has the ability to offer a good quality of service, especially for the UEs on the cell edges, be it that there is still a need for low-complexity signal processing algorithms. In this paper, the problem of optimal power allocation combined with uplink receive combining and downlink transmit precoding is tackled by providing efficient distributed MMSE-based algorithms. The necessary fronthaul communications to estimate the combining/precoding vectors and the necessary large-scale channel statistics are reduced to a minimum and rely on in-network summation that can be accomplished whenever the APs can be arranged into a tree-topology. Non-weighted max-sum and max–min are used as utilities for the power allocation, but the algorithms are not limited to these cases. Simulations show that the proposed algorithms outperform heuristic power allocation methods, both in uplink and downlink.

**Keywords:** Cell-free massive MIMO (CFmMIMO), Uplink receive combining, downlink transmit precoding, Optimal power allocation, distributed processing, Minimum-mean-squared-error (MMSE)

## 1 Introduction

'multiple-input multiple-output' (MIMO) technology plays an important role in present-day wireless communication standards and networks. It uses multiple antennas to transmit and receive multiple data signals simultaneously over one and the same radio channel and hence increase spectral efficiency. Massive MIMO (mMIMO) systems use a large($r$) number of antennas, to reap additional benefits in rich scattering environments. While mMIMO was initially viewed as a promising concept for future cellular wireless communication networks, it already became a reality in 2018 with commercial deployment in several countries [1]. The signal processing methods

Springer Open

© The Author(s) 2023. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

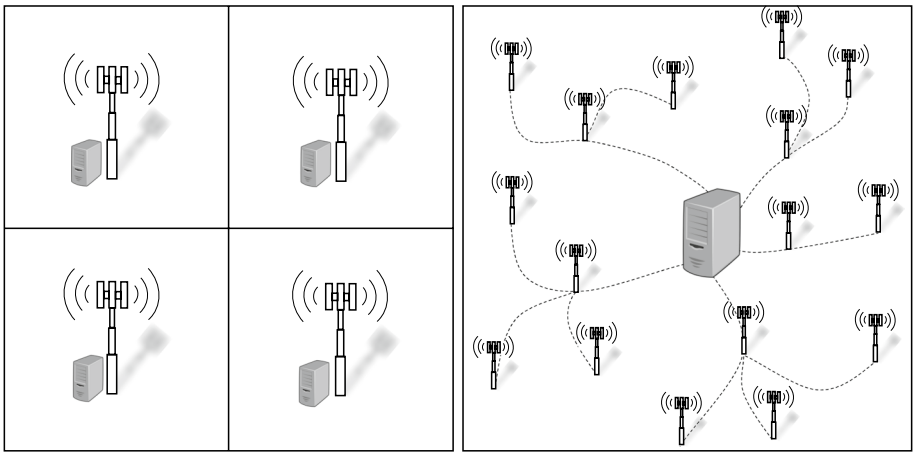


**Fig. 1** Comparison of a cellular massive MIMO system (left) with a CFmMIMO system (right). In a CFmMIMO system, a large number of APs with multiple antennas are connected to a NC, serving a large number of UEs in the coverage area

required to achieve large spectral efficiency gains compared to conventional multi-user MIMO systems [2] are available and well understood [3]. The spectral efficiency gains associated with mMIMO emerge from the higher spatial resolution, its robustness against small-scale fading due to the so-called 'channel hardening' effect, and its ability to suppress interference even with imperfect channel state information (CSI).

One of the major shortcomings of current cellular wireless communication networks, where a user equipment (UE) is served only by the access point (AP) of the cell it resides in, is that UEs on the cell edges experience a low channel gain to their serving AP and a high interference power from nearby cells. Current uplink (UL) receive combining and downlink (DL) transmit precoding algorithms do not effectively mitigate such interference. To address this issue, the concept of cell-free massive MIMO (CFmMIMO) [4–6] has been proposed as a promising alternative for the current 5G cellular networks. In CFmMIMO systems a large number of APs are connected to a network center (NC) and cooperate via a fronthaul network to serve a large number of UEs in the network simultaneously as shown in Fig. 1. The joint operation of the APs then allows for implementing interference-rejecting combining and precoding algorithms effective for all UEs in the network. CFmMIMO thus effectively eliminates cell borders while still reusing the network layout as rolled out previously, which offers the potential to spectacularly increase performance as compared to current cellular mMIMO systems.

CFmMIMO has its roots in works on distributed MIMO [7, 8] and coordinated multipoint [9, 10], but with the distinction that a UE can be served by several APs, instead of only by the one AP serving the cell in which the UE is located and that the transmission is coordinated by a NC. To allow for a scalable and low-complexity implementation, CFmMIMO has to operate under the following constraints [6]: (a) the time division duplex (TDD) protocol is used, exploiting channel reciprocity between UL and DL; (b) UL channel estimates are computed locally at each AP and will be used only locally, which means that they are not directly communicated to the NC over the fronthaul links; (c) combining/precoding vectors are computed locally at

the APs and not at the NC; (d) the fronthaul network is only used to send data symbols and the necessary channel statistics to perform centralized data decisions.

### 1.1 Related works

The authors in [4, 11] proposed to use maximum ratio combining and conjugate beamforming for UL and DL data transmission in CFmMIMO systems, presented closed-form expressions for the spectral efficiency in UL and DL for each UE and formulated max–min power control algorithms, which were shown to outperform a cellular setup. However, using more complicated combining and precoding algorithms combined with heuristic power allocation methods [5, 6, 10, 12, 13], are shown to lead to higher throughputs. Optimal UL power allocation and receive combining was investigated in [14, 15] using minimum-mean-squared error (MMSE) processing by formulating the original max–min signal-to-interference-plus-noise ratio problem for the optimization of combining vectors at a central point (like the NC). Unfortunately centralized computations increase the fronthaul communication and are not complying with the scalable characteristics of CFmMIMO systems as defined before. Optimal DL power allocation was considered in [5] combined with zero-forcing transmit precoding: although the gains were significant, the zero-forcing transmit precoding again requires a fully centralized computation with significant fronthaul communications. In [16], optimal power allocation algorithms for local partial zero-forcing were proposed as a scalable counterpart and showed to achieve a performance close to the performance of zero-forcing transmit precoding and the centralized regularized zero-forcing transmit precoding, which is a benchmark in DL transmit precoding and can be related to MMSE combining in the UL via UL/DL duality [17, 18]. In [19], a max sum SE problem in an uplink CFmMIMO system was also investigated. Artificial neural networks were used, with the UE positions as input and the power control policy as the output. An UL CFmMIMO system with limited fronthaul was taken into consideration in [20], where zero-forcing was employed locally to heuristically estimate the max sum SE power control scheme. For single antenna APs, a deep neural network-based power allocation method was proposed in [21].

### 1.2 Contributions

The missing point in the research field is the fact that (optimal) MMSE-based combining/precoding in its original form can only be performed at the NC if all antenna signals are transmitted to it, which implies a large fronthaul communication overhead. Otherwise, heuristic combining/precoding schemes can be used at the APs, reducing the fronthaul communication overhead strongly, but optimal performance can no longer be guaranteed. The same can be stated for the UL and DL power allocation. Therefore, in this paper an MMSE-based distributed UL receive combing and power allocation algorithm (D-UL-RCPA) and DL transmit precoding and power allocation algorithm (D-DL-TPPA) are presented that are scalable and yet attain this optimal performance. The combining/precoding vectors are computed only at the APs, such that the computation are distributed over the network and are not only performed by the NC. The necessary fronthaul communications to estimate the combining/precoding vectors and the necessary large-scale channel statistics are reduced to a minimum and rely on in-network

summation that can be accomplished whenever the APs can be arranged into a tree-topology. The reduction in fronthaul communications is especially large when the total number of antennas in the network is larger than the total number of UEs.

The algorithms are derived based on general lower bounds of the system achievable rate, taking into account the effect of channel estimation errors. Non-weighted max-sum and max–min are used as utilities for the power allocation, but the algorithms are not limited to these cases. The former power allocation strategy maximizes the proposed lower bound for the total system sum rate, while the latter, targeting fairness across users, maximizes the minimum spectral efficiency lower bound across all the UEs. Simulations show that the proposed algorithms outperform heuristic power allocation methods, both in UL and DL.

### 1.3 Paper outline and notation

The paper is organized as follows. The signal model is presented in Sect. 2. In Sect. 3 an optimal UL receive combining and power allocation strategy is developed and an efficient distributed algorithm for a CFmMIMO system is presented, A similar algorithm is developed in Sect. 4 for DL transmit precoding and power allocation, using UL/DL duality. Some further considerations about the power allocation strategies and network topologies are presented in Sect. 5. The performance of the proposed algorithms is numerically evaluated and compared in Sect. 6. Conclusions and further research directions are provided in Sect. 7.

Superscripts $.^{\mathrm{T}}$ and $.^{H}$ are used to denote the transpose and conjugate transpose operation. Bold lower case letters $\mathbf{b} = [b_1 \ \ldots \ b_K]^{\mathrm{T}}$ are used to denote complex vectors and bold upper case letters are used to represent complex matrices $\mathbf{A} = [\mathbf{a}_1 \ \ldots \ \mathbf{a}_K]$. $\mathbf{I}_N$ is the $N \times N$ identity matrix and $\mathbf{0}$ denotes an all zeros vectors whose dimensions are clear from the context. The multi-variate circularly symmetric complex Gaussian distribution with correlation matrix $\mathbf{R}$ is denoted by $\mathcal{NC}(0, \mathbf{R})$, $\mathrm{diag}\{a_1, \ldots, a_K\}$ is used to denote a diagonal matrix with $a_1, \ldots, a_K$ on its diagonal and $\mathrm{Blkdiag}\{\mathbf{A}_1, \ldots, \mathbf{A}_K\}$ is used for a block-diagonal matrix with the square matrices $\mathbf{A}_1, \ldots, \mathbf{A}_K$ on the diagonal. Finally $||.||$ and $E\{.\}$ denote the Euclidean norm and expected value operator respectively.

## 2 Signal model

Consider a CFmMIMO system consisting of $K$ single-antenna UEs and $L$ APs, each having $N$ antennas and local processing capabilities, that are randomly deployed over the coverage area. The APs are connected to a NC via a fronthaul network. This setup allows for coherent transmission and reception of data to and from the UEs. In the CFmMIMO literature [5, 22] it is often assumed that the number of antennas $NL$ is much larger than the number of UEs $K$ and that both $L$ and $K$ are large.

The time-frequency resources are divided into coherence blocks of $\tau_c$ samples during which the channels are assumed to remain constant [18]. APs and UEs operate using a TDD protocol and a channel coherence interval is divided into three phases: UL channel estimation, UL data transmission and DL data transmission. These phases consist of $\tau_p$, $\tau_u$ and $\tau_d$ samples respectively, such that $\tau_c = \tau_p + \tau_u + \tau_d$. Since the TDD protocol is used, the UL channel and the DL channel are assumed to be each other's conjugate due to channel reciprocity, so that indeed only UL channel estimation has to be performed.

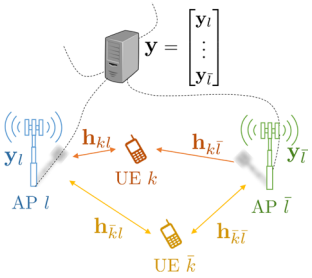



**Fig. 2** Schematic overview of the signal model with the used symbols

The channel from UE $k$ to AP $l$ is given by $\mathbf{h}_{kl} \in \mathbb{C}^N$ and the channel from UE $k$ to all the APs is denoted by $\mathbf{h}_k = [\mathbf{h}_{k1}^{\mathrm{T}} \ \ldots \ \mathbf{h}_{kL}^{\mathrm{T}}]^{\mathrm{T}} \in \mathbb{C}^{NL}$. The channels $\mathbf{h}_{kl}$ are assumed to remain constant during a coherence block and drawn from an independent correlated Rayleigh fading realization $\mathcal{NC}(\mathbf{0}, \mathbf{R}_{kl})$ in each coherence block.[1] $\mathbf{R}_{kl} \in \mathbb{C}^{N \times N}$ is the positive semi-definite spatial correlation matrix describing the large-scale fading, including geometric pathloss, shadowing, antenna gains, and spatial channel correlation [24]. The large-scale fading is assumed to remain constant over different coherence blocks. The Gaussian distribution models the small-scale fading. Due to the spatial distribution of the APs in the network, the channel vectors of different APs are independently distributed, i.e. $E\{\mathbf{h}_{kl}\mathbf{h}_{kn}^H\} = \mathbf{0}_{N \times N}$ for $l \neq n$, such that channel estimation can be performed independently at each AP. A schematic overview of the signal model is provided in Fig. 2.

### 2.1 UL channel estimation

In the UL channel estimation phase, the UEs send UL pilots to allow for simultaneous channel estimation at the APs. There are $\tau_{\mathrm{p}}$ mutually orthogonal pilot signals $\{\phi_1, \ldots, \phi_{\tau_{\mathrm{p}}}\} \in \mathbb{C}^{\tau_{\mathrm{p}}}$ of $\tau_{\mathrm{p}}$ samples, normalized to have unit power. This implies that $\phi_i^H \phi_j$ is equal to $\tau_{\mathrm{p}}$ if $i = j$ and 0 otherwise. Each UE is assigned to a pilot signal when it gains access to the network (see [4, 17] for pilot assignment protocols). As $K$ is often much larger than $\tau_{\mathrm{p}}$, the same pilot signal will be assigned to multiple UEs, leading to pilot contamination and the associated negative effects on the channel estimation [3, 25, 26]. The received signal $\mathbf{y}_l^{\mathrm{pilot}} \in \mathbb{C}^{N \times \tau_{\mathrm{p}}}$ at AP $l$ is given as:

$$\mathbf{y}_l^{\mathrm{pilot}} = \sum_{k=1}^{K} \sqrt{p_k^{\mathrm{pilot}}} \mathbf{h}_{kl} \phi_k^H + \mathbf{n}_l^{\mathrm{pilot}} \tag{1}$$

where $p_k^{\mathrm{pilot}}$ is the pilot transmit power of UE $k$ and $\mathbf{n}_l^{\mathrm{pilot}} \in \mathbb{C}^{N \times \tau_{\mathrm{p}}}$ is thermal noise. Let $\mathcal{S}_t \subset \{1, \ldots, K\}$ denote the subset of UEs assigned to the same pilot signal $\phi_t$ for $t \in \{1, \ldots, \tau_{\mathrm{p}}\}$. After despreading, i.e. multiplying the signal coherently with $\phi_t / ||\phi_t||$, the received signal $\mathbf{y}_{tl}^{\mathrm{pilot}} \in \mathbb{C}^N$ at AP $l$ is given as

$$\mathbf{y}_{tl}^{\mathrm{pilot}} = \sum_{i \in \mathcal{S}_t} \sqrt{\tau_{\mathrm{p}} p_i^{\mathrm{pilot}}} \mathbf{h}_{il} + \mathbf{n}_{tl}^{\mathrm{pilot}} \tag{2}$$

[1] One can consider the more general Rician fading channel [23], but this is not considered here for simplicity.

where the thermal noise $\mathbf{n}_{tl}^{\text{pilot}}$ for each pilot signal $t$ is distributed as $\mathcal{NC}(\mathbf{0}, \mathbf{R}_{\mathbf{n}_l\mathbf{n}_l}^{\text{pilot}})$. Assuming that AP $l$ is aware of the local large-scale fading correlation matrix $\mathbf{R}_{kl}$ (see e.g., [27–30] for correlation matrix estimation methods) of all the UEs, the MMSE estimate of $\mathbf{h}_{kl}$ for $k \in \mathcal{S}_t$ is

$$\hat{\mathbf{h}}_{kl} = \sqrt{\tau_{\text{p}} p_k^{\text{pilot}}} \mathbf{R}_{kl} \boldsymbol{\Psi}_{tl}^{-1} \mathbf{y}_{tl}^{\text{pilot}} \tag{3}$$

where

$$\boldsymbol{\Psi}_{tl} = E\left\{ \mathbf{y}_{tl}^{\text{pilot}} \mathbf{y}_{tl}^{\text{pilot},H} \right\} = \sum_{i \in \mathcal{S}_t} \tau_{\text{p}} p_i^{\text{pilot}} \mathbf{R}_{il} + \mathbf{R}_{\mathbf{n}_l\mathbf{n}_l}^{\text{pilot}}. \tag{4}$$

The MMSE estimate $\hat{\mathbf{h}}_{kl}$ of $\mathbf{h}_{kl}$ is an unbiased estimate with estimation error $\tilde{\mathbf{h}}_{kl} = \mathbf{h}_{kl} - \hat{\mathbf{h}}_{kl}$. The distribution of both random variables is given by

$$\hat{\mathbf{h}}_{kl} \sim \mathcal{NC}\left(\mathbf{0}, \tau_{\text{p}} p_k^{\text{pilot}} \mathbf{R}_{kl} \boldsymbol{\Psi}_{tl}^{-1} \mathbf{R}_{kl}\right), \tag{5}$$

$$\tilde{\mathbf{h}}_{kl} \sim \mathcal{NC}\left(\mathbf{0}, \mathbf{R}_{kl} - \tau_{\text{p}} p_k^{\text{pilot}} \mathbf{R}_{kl} \boldsymbol{\Psi}_{tl}^{-1} \mathbf{R}_{kl}\right) \triangleq \mathcal{NC}(\mathbf{0}, \mathbf{C}_{kl}). \tag{6}$$

It is noted that the $N \times N$ matrices $\sqrt{\tau_{\text{p}} p_k^{\text{pilot}}} \mathbf{R}_{kl} \boldsymbol{\Psi}_{tl}^{-1}$ and $\mathbf{C}_{kl}$ can be precomputed for complexity reduction at AP $k$, since they only depend on the channel statistics, which are changing slowly throughout the communication.[2]

### 2.2 UL data transmission

In the UL data transmission phase, the received signal $\mathbf{y}_l^{\text{UL}} \in \mathbb{C}^N$ at AP $l$ is given by

$$\mathbf{y}_l^{\text{UL}} = \sum_{k=1}^{K} \mathbf{h}_{kl} s_k + \mathbf{n}_l^{\text{UL}} = \mathbf{H}_l \mathbf{s} + \mathbf{n}_l^{\text{UL}} \tag{7}$$

where $s_k \in \mathbb{C}$ is the signal transmitted by UE $k$ with UL transmit power $p_k = E\{s_k s_k^H\}$ and $\mathbf{n}_l^{\text{UL}} \in \mathbb{C}^N \sim \mathcal{NC}(\mathbf{0}, \mathbf{R}_{\mathbf{n}_l\mathbf{n}_l}^{\text{UL}})$ is an additive Gaussian noise component, including thermal antenna noise and quantization noise during UL transmission. Furthermore, $\mathbf{H}_l = [\mathbf{h}_{1\,l} \ \ldots \ \mathbf{h}_{Kl}] \in \mathbb{C}^{N \times K}$, $\mathbf{s} = [s_1 \ \ldots \ s_K]^{\text{T}} \in \mathbb{R}^K$ and $\mathbf{p} = [p_1 \ \ldots \ p_K]^{\text{T}} \in \mathbb{C}^K$ are the concatenation of the channels from all the UEs to AP $l$, the signals transmitted by all the UEs and the UL transmit powers respectively. Stacking the received signals of all APs in $\mathbf{y}^{\text{UL}} = [\mathbf{y}_1^{\text{UL,T}} \ \ldots \ \mathbf{y}_L^{\text{UL,T}}]^{\text{T}} \in \mathbb{C}^{NL}$ as well as the noise components in $\mathbf{n}^{\text{UL}} = [\mathbf{n}_1^{\text{UL,T}} \ \ldots \ \mathbf{n}_L^{\text{UL,T}}]^{\text{T}} \in \mathbb{C}^{NL} \sim \mathcal{NC}(0, \mathbf{R}_{\mathbf{nn}}^{\text{UL}})$ where $\mathbf{R}_{\mathbf{nn}}^{\text{UL}} = \text{Blkdiag}\{\mathbf{R}_{\mathbf{n}_1\mathbf{n}_1}^{\text{UL}}, \ldots, \mathbf{R}_{\mathbf{n}_L\mathbf{n}_L}^{\text{UL}}\}$, results in the network-wide signal model:

$$\mathbf{y}^{\text{UL}} = \mathbf{H}\mathbf{s} + \mathbf{n}^{\text{UL}} \tag{8}$$

with $\mathbf{H} = [\mathbf{h}_1 \ \ldots \ \mathbf{h}_K] = [\mathbf{H}_1^{\text{T}} \ \ldots \ \mathbf{H}_L^{\text{T}}]^{\text{T}} \in \mathbb{C}^{NL \times K}$.

[2] Changes are due to UE mobility or new scheduling decisions, but these channel statistics changes are typically 50 times slower than the small scale fading statistics changes, so are assumed negligible for several coherence blocks.

In network-wide UL receive combining the signal $s_k$ is estimated by linearly combining the received signal $\mathbf{y}^{\mathrm{UL}}$ with a combining vector $\mathbf{v}_k \in \mathbb{C}^{NL}$, i.e. $\hat{s}_k = \mathbf{v}_k^H \mathbf{y}^{\mathrm{UL}}$. Note that this linear combining can be performed in the network if AP $l$ selects the local combining vector $\mathbf{v}_{kl} \in \mathbb{C}^N$ in $\mathbf{v}_k = [\mathbf{v}_{k1}^{\mathrm{T}} \ \ldots \ \mathbf{v}_{kL}^{\mathrm{T}}]^{\mathrm{T}}$ and computes the local estimate $\mathbf{v}_{kl}^H \mathbf{y}_l$. The NC then estimates $s_k$ by combining the local estimates as:

$$\hat{s}_k = \sum_{l=1}^{L} \mathbf{v}_{kl}^H \mathbf{y}_l^{\mathrm{UL}} = \mathbf{v}_k^H \mathbf{h}_k s_k + \sum_{i \neq k} \mathbf{v}_k^H \mathbf{h}_i s_i + \mathbf{v}_k^H \mathbf{n}^{\mathrm{UL}}. \tag{9}$$

The goal is then to choose the combining vector $\mathbf{v}_k$ that provides a good estimate $\hat{s}_k$, but where the APs use only local CSI. A popular choice in CFmMIMO literature is maximum-ratio (MR) combing with $\mathbf{v}_{kl} = \hat{\mathbf{h}}_{kl}$ [4, 31, 32]. An alternative heuristic scheme, which performs generally better but requires more processing power of the AP is local MMSE combining [6]. Network-wide UL power allocation algorithms for MR combining can be found in [14, 15, 22]. In this paper, network-wide MMSE combining [17] and associated power allocation algorithms will be considered, requiring typically network-wide CSI. However this paper shows that if a limited number of parameters can be exchanged between the NC and the APs, this network-wide MMSE combining and associated power allocation can still be obtained locally at each AP.

### 2.3 DL data transmission

In the DL data transmission phase, the received signal $y_k^{\mathrm{DL}}$ at UE $k$ is given by

$$\begin{aligned} y_k^{\mathrm{DL}} &= \sum_{l=1}^{L} \mathbf{h}_{kl}^H \sum_{i=1}^{K} \mathbf{w}_{il} \sqrt{\rho_i} \zeta_i + n_k^{\mathrm{DL}} \\ &= \mathbf{h}_k^H \mathbf{w}_k \sqrt{\rho_k} \zeta_k + \sum_{i \neq k} \mathbf{h}_k^H \mathbf{w}_i \sqrt{\rho_i} \zeta_i + n_k^{\mathrm{DL}} \end{aligned} \tag{10}$$

where $\zeta_k \in \mathbb{C}$ is the signal transmitted to UE $k$ with unit DL transmit power, $\rho_k$ is the network-wide DL transmit power allocated to UE $k$, $n_k^{\mathrm{DL}} \sim \mathcal{NC}\left(0, \sigma_k^{2,\mathrm{DL}}\right)$ is an additive Gaussian noise component, $\mathbf{w}_k = [\mathbf{w}_{k1}^{\mathrm{T}} \ldots \mathbf{w}_{kL}^{\mathrm{T}}]^{\mathrm{T}} \in \mathbb{C}^{NL}$ is the concatenation of the precoding vectors used for UE $k$. Also denote the DL transmit powers as $\boldsymbol{\rho} = [\rho_1 \ \ldots \ \rho_K]^{\mathrm{T}} \in \mathbb{R}^K$ and the predocing matrix as $\mathbf{W} = [\mathbf{w}_1 \ \ldots \ \mathbf{w}_K] \in \mathbb{C}^{NL \times K}$. Again the selection of the precoding vector should only depend on local CSI for CFmMIMO. The most popular choice is MR precoding with

$$\mathbf{w}_{kl} = \frac{\hat{\mathbf{h}}_{kl}}{\sqrt{E\{||\hat{\mathbf{h}}_{kl}||^2\}}}. \tag{11}$$

DL power allocation algorithms for different local precoding schemes can be found in [4, 5, 33]. In this paper, network-wide MMSE precoding [34] and associated power allocation algorithms will be considered, requiring typically network-wide CSI. However this paper shows that using a similar approach as for UL data transmission, this

network-wide MMSE precoding and associated power allocation can still be obtained locally at each AP.

## 3 UL receive combing and power allocation

### 3.1 UL receive combining

A standard lower bound for the ergodic capacity of the UL data transmission for UE $k$ can be obtained by rewriting (9) using the estimated channels as

$$\hat{s}_k = \mathbf{v}_k^H \hat{\mathbf{h}}_k s_k + \mathbf{v}_k^H \tilde{\mathbf{h}}_k s_k + \sum_{i \neq k} \mathbf{v}_k^H (\hat{\mathbf{h}}_i + \tilde{\mathbf{h}}_i) s_i + \mathbf{v}_k^H \mathbf{n}^{\mathrm{UL}} \tag{12}$$

and is given by the achievable UL spectral efficiency (SE) [24, 35]

$$\mathrm{SE}_k^{\mathrm{UL},1} = \frac{\tau_{\mathrm{u}}}{\tau_{\mathrm{c}}} E\left\{ \log_2 \left( 1 + \mathrm{SINR}_k^{\mathrm{UL,inst}} \right) \right\} \text{ [bits/s/Hz]} \tag{13}$$

with the instantaneous effective signal-to-interference-and-noise ratio $\mathrm{SINR}_k^{\mathrm{UL,inst}}$ given as

$$\frac{p_k |\mathbf{v}_k^H \hat{\mathbf{h}}_k|^2}{\sum_{i \neq k} p_i |\mathbf{v}_k^H \hat{\mathbf{h}}_i|^2 + \sum_{i=1}^{K} p_i \mathbf{v}_k^H \mathbf{C}_i \mathbf{v}_k + \mathbf{v}_k^H \mathbf{R}_{\mathbf{nn}}^{\mathrm{UL}} \mathbf{v}_k}. \tag{14}$$

where $\mathbf{C}_i = \mathrm{Blkdiag}\{\mathbf{C}_{i,1}, \ldots, \mathbf{C}_{iL}\}$. The expectation is with respect to the different channel realizations.

The UL SE is valid for any combining vector $\mathbf{v}_k$. In this paper optimal receive combining is considered and it can be seen from expression (13) that the optimal instantaneous combining vector can be obtained by maximizing the generalized Rayleigh quotient in the instantaneous SINR in (14). The optimal combining vector is unique up to a scalar multiplication, and so to make the solution unique, the solution that minimizes the mean-squared error $E\{|s_k - \hat{s}_k|^2 | \hat{\mathbf{H}}\}$ is chosen. The MMSE combining vector for UE $k$ is then given by

$$\mathbf{v}_k^{\mathrm{MMSE}} = p_k \left( \hat{\mathbf{H}} \mathbf{P} \hat{\mathbf{H}}^H + \sum_{i=1}^{K} p_i \mathbf{C}_i + \mathbf{R}_{\mathbf{nn}}^{\mathrm{UL}} \right)^{-1} \hat{\mathbf{h}}_k \tag{15}$$

with $\mathbf{P} = \mathrm{diag}\{p_1, \ldots, p_K\}$ as shown in [27].

The following theorem shows that the MMSE combining vectors for UEs $k = 1 \ldots K$, i.e., the combining matrix $\mathbf{V}^{\mathrm{MMSE}} = [\mathbf{v}_1^{\mathrm{MMSE}}, \ldots, \mathbf{v}_K^{\mathrm{MMSE}}]$, can be computed in a way that allows for an efficient distributed implementation.

**Theorem 3.1** *The MMSE combining matrix* $\mathbf{V}^{\mathrm{MMSE}}$ *can be written as*

$$\mathbf{V}^{\mathrm{MMSE}} = \begin{bmatrix} \mathbf{X}_1 \\ \vdots \\ \mathbf{X}_L \end{bmatrix} \mathbf{U} = \mathbf{X}\mathbf{U} \tag{16}$$

*with*

$$\mathbf{U} = \left(\mathbf{P}^{-1} + \mathbf{X}^H \hat{\mathbf{H}}\right)^{-1} = \left(\mathbf{P}^{-1} + \sum_{l=1}^{L} \mathbf{X}_l^H \hat{\mathbf{H}}_l\right)^{-1}, \tag{17}$$

$$\mathbf{X}_l = \left(\sum_{k=1}^{K} p_k \mathbf{C}_{kl} + \mathbf{R}^{\mathrm{UL}}_{\mathbf{n}_l \mathbf{n}_l}\right)^{-1} \hat{\mathbf{H}}_l. \tag{18}$$

***Proof*** See "Appendix 1". □

In [27], the network-wide distributed MMSE receive combining (N-DRC) algorithm is proposed based on the computation of the MMSE combining vectors in Theorem 3.1. The algorithm can be summarized in the following steps for each coherence block:

1 Each AP $l$ estimates $\hat{\mathbf{H}}_l$ and computes $\mathbf{X}_l$ as in (18) and $\mathbf{X}_l^H \hat{\mathbf{H}}_l$. It transmits the combined signals $\mathbf{x}_l = \mathbf{X}_l^H \mathbf{y}_l^{\mathrm{UL}}$ and the parameters $\mathbf{X}_l^H \hat{\mathbf{H}}_l$ to the NC.
2 The available links in the network are used efficiently such that the NC obtains the in-network sums $\mathbf{x} = \sum_{l=1}^{L} \mathbf{x}_l$ and $\mathbf{X}^H \hat{\mathbf{H}} = \sum_{l=1}^{L} \mathbf{X}_l^H \hat{\mathbf{H}}_l$.
3 The NC uses $\mathbf{X}^H \hat{\mathbf{H}}$ to compute $\mathbf{U} = [\mathbf{u}_1 \ \ldots \ \mathbf{u}_K]$ as in (17) and computes the estimates $\hat{s}_k$ as $\mathbf{u}_k^H \mathbf{x}$ for all UEs.

Some important advantages of this distributed algorithm compared to a centralized algorithm, i.e. when each AP serves as a receiver that directly transmits its received signals to the NC, are repeated here:

- The NC only has to invert a $K \times K$ matrix in (17) instead of the $LN \times LN$ matrix in (15). However, this requires that each AP inverts an $N \times N$ matrix to compute $\mathbf{X}_l$, where the necessary local matrices can be precomputed and kept fixed as long as the transmit powers $\mathbf{p}$ remain unchanged.
- If the in-network sums in step 2 are computed efficiently exploiting the available network topology, this will reduce the network signaling strongly compared to the transmission of all $NL$ signals.
- The method is also robust against link failures: if the data ($\mathbf{x}_l$ and $\mathbf{X}_l^H \hat{\mathbf{H}}_l$) from a certain AP $l$ is not received as a term in the in-network sums, the obtained estimate will still be optimal for a network setup with AP $l$ removed.

It is important to note that the MMSE combining vectors depend on the UL transmit powers, the noise statistics and the channel estimation $\hat{\mathbf{H}}$, which in turn depends on the pilot transmit powers, the pilot assignment strategy and the large-scale statistics. The UL SE in (13) is as a consequence also influenced by these parameters. In the

following subsection a way to improve this achievable bound is provided, based on power allocation.

### 3.2 UL power allocation preliminaries

With UL power allocation[3] the UL transmit powers $\mathbf{p}$ are set to maximize some utility function for some UL receive combining strategy. The problem can be formulated mathematically as:

$$\begin{aligned} \max_{\mathbf{p}\in\mathcal{P}} \quad & U(\mathrm{SE}_1,\ldots,\mathrm{SE}_K) \\ \text{s.t.} \quad & \mathrm{SE}_k = \mathrm{SE}(\mathbf{p},\mathbf{v}_k(\mathbf{p})) \quad \forall k \end{aligned} \tag{19}$$

where, for instance, (13) is used for $\mathrm{SE}_k$ and $\mathcal{P}$ is a convex region of allowed UL transmit powers $\mathbf{p}$ for the UEs, e.g. corresponding to a total power constraint or a per UE power constraint. The utility function is assumed to be (quasi-)concave and nondecreasing. In this paper, two specific utility functions are considered, but the proposed algorithm is not limited to these cases. The channel hardening makes power allocation different in CFmMIMO systems compared to single-antenna systems. While the combining vectors $\mathbf{v}_k$ will be adopted depending on small-scale variations of the channels, $\mathbf{p}$ will only be adopted to large-scale variations to effectively compute the expectations in e.g. (13). Since the expected value operation in the expression for the SE in (13) contains the logarithm, finding the optimal power allocation is difficult due to the non-trivial relation between $\mathbf{p}$ and the $\mathrm{SE}_k^{\mathrm{UL},1}(\mathbf{p},\mathbf{v}_k)$. Therefore a different SE formulation will be considered for the UL power allocation, which is obtained under the assumption of channel hardening[4]. For this, only the part of the desired signal received over the average channel $E\{\mathbf{v}_k^H\mathbf{h}_k\}$ is treated as the true desired signal, i.e.

$$\begin{aligned} \hat{s}_k = {} & E\{\mathbf{v}_k^H\mathbf{h}_k\}s_k + \left(\mathbf{v}_k^H\mathbf{h}_k - E\{\mathbf{v}_k^H\mathbf{h}_k\}\right)s_k \\ & + \sum_{i\neq k}\mathbf{v}_k^H\mathbf{h}_i s_i + \mathbf{v}_k^H\mathbf{n}. \end{aligned} \tag{20}$$

This results in a deterministic channel under uncorrelated interference and noise for which the following SE (also referred to as the use-and-then-forget bound [24]) can be derived:

$$\mathrm{SE}_k^{\mathrm{UL},2} = \frac{\tau_\mathrm{u}}{\tau_\mathrm{c}}\log_2\left(1+\mathrm{SINR}_k^{\mathrm{UL},2}\right) \text{ [bits/s/Hz]} \tag{21}$$

where

$$\frac{p_k\left|E\{\mathbf{v}_k^H\mathbf{h}_k\}\right|^2}{\sum_{i=1}^K p_i E\left\{\left|\mathbf{v}_k^H\mathbf{h}_i\right|^2\right\} - p_k\left|E\{\mathbf{v}_k^H\mathbf{h}_k\}\right|^2 + E\left\{\left|\mathbf{v}_k^H\mathbf{n}\right|^2\right\}} \tag{22}$$

[3] The pilot transmit powers are assumed to remain fixed. The estimation of the quantities $\mathbf{C}_k$ and $\hat{\mathbf{h}}_k$ is thus not depending on the assigned UL transmit powers. Joint pilot and UL transmit power allocation methods can be found in [36, 37], but are not considered here.

[4] When the number of antennas per AP is large, the effective channels to the desired UEs are almost deterministic after combining/precoding, although the channel responses are random. This phenomenon is called channel hardening and can be expressed as $\frac{\mathbf{v}_k^H\mathbf{h}_k}{NL} \to \frac{E\{\mathbf{v}_k^H\mathbf{h}_k\}}{NL}$ if $NL \to \infty$.

**Table 1** Overview of used symbols for power allocation

| Symbol | Meaning | Place of computation |
|---|---|---|
| $\gamma_k$ | $E\{\mathbf{v}_k^H \hat{\mathbf{h}}_k\}$ | NC |
| $\upsilon_{ki}$ | $E\{\lvert\mathbf{v}_k^H \hat{\mathbf{h}}_i\rvert^2\}$ | NC |
| $\xi_{ki}$ | $E\{\mathbf{v}_k^H \mathbf{C}_i \mathbf{v}_k\}$ | NC |
| $\nu_k$ | $E\{\mathbf{v}_k^H \mathbf{R}_{\mathbf{nn}}^{\mathrm{UL}} \mathbf{v}_k\}$ | NC |
| $\mu_k$ | $E\{\mathbf{v}_k^H \mathbf{v}_k\}$ | NC |
| $\xi_{ki}^l$ | $E\{\mathbf{v}_{kl}^H \mathbf{C}_{jl} \mathbf{v}_{kl}\}$ | AP |
| $\nu_k^l$ | $E\{\mathbf{v}_{kl}^H \mathbf{R}_{\mathbf{n}_j\mathbf{n}_j}^{\mathrm{UL}} \mathbf{v}_{kl}\}$ | AP |
| $\mu_k^l$ | $E\{\mathbf{v}_{kl}^H \mathbf{v}_{kl}\}$ | AP |

defines the signal-to-interference-and-noise ratio $\mathrm{SINR}_k^{\mathrm{UL},2}$. The expectations are with respect to the different channel realizations. The expression is in theory only depending on the large-scale fading.

It is intuitively clear that the $\mathrm{SE}_k^{\mathrm{UL},2}$ is a less tight lower bound for the ergodic capacity than the achievable lower bound given by $\mathrm{SE}_k^{\mathrm{UL},1}$, since the channel estimates are not utilized in the signal estimation, i.e.

$$\mathrm{SE}_k^{\mathrm{UL},1} \geq \mathrm{SE}_k^{\mathrm{UL},2}. \tag{23}$$

However, simulations in [24, Sect. 4.2] show that the bound is still tight for $LN \gg 1$, i.e. when channel hardening has its effect. The bound is also tight for a deterministic channel, i.e. when only 1 channel realization is used to compute the expectations over the channel realizations. The simulations also show that the bound is tight for any combining scheme, but that using the MMSE combining vector leads to a tighter bound than any other multiple of this MMSE combining vector, e.g. when a constraint $\mathbf{v}_k^H \hat{\mathbf{h}}_k = 1$ is used in each coherence block.

Since closed form expressions for the expectations in (22) are difficult to derive for the optimal MMSE combining vector (15),[5] the next theorem provides a way to estimate the expectations without having access to the true channels $\mathbf{H}$.

**Theorem 3.2** *If the combining vector* $\mathbf{v}_k$ *is independent of the instantaneous channel estimation error* $\{\tilde{\mathbf{h}}_k\}_{\forall k}$ *and also independent of the instantaneous UL noise* $\mathbf{n}^{\mathrm{UL}}$*, then the expectations can be computed without knowledge of the true channel* $\mathbf{H}$:

$$E\left\{\mathbf{v}_k^H \mathbf{h}_k\right\} = \gamma_k, \tag{24}$$

$$E\left\{\left|\mathbf{v}_k^H \mathbf{h}_i\right|^2\right\} = \upsilon_{ki} + \xi_{ki} = \upsilon_{ki} + \sum_{l=1}^{L} \xi_{ki}^l, \tag{25}$$

[5] There are closed form expressions available in [15, 38] when the number of antennas in the network grows to infinity, but these do not allow for an efficient distributed computation.

$$E\left\{\left|\mathbf{v}_k^H \mathbf{n}^{\mathrm{UL}}\right|^2\right\} = \nu_k = \sum_{l=1}^{L} \nu_k^l. \tag{26}$$

*The right-hand side symbols are defined in Table 1 and these only depend on the channel estimates and the large-scale statistics* $\{\mathbf{C}_k\}_{\forall k}$ *and* $\mathbf{R}_{\mathbf{nn}}^{\mathrm{UL}}$.

***Proof*** See "Appendix 2" □

The requirements for Theorem 3.2 are valid for the MMSE combining vector (15), as $\mathbf{v}_k^{\mathrm{MMSE}}$ only depends on the instantaneous channel estimates $\hat{\mathbf{H}}$ and the fixed matrices $\{\mathbf{C}\}_{\forall k}$ and $\mathbf{R}_{\mathbf{nn}}^{\mathrm{UL}}$.

The (ergodic) statistics in Table 1 can be estimated using (recursive) time-averaging at the NC or in a distributed way using in-network sums for $\xi_{ki}$ and $\nu_k$, if each AP $l$ has access to the transformation matrix $\mathbf{U}$, since $\mathbf{v}_{kl} = \mathbf{X}_l \mathbf{u}_k$.

### 3.3 UL receive combining and power allocation

To be able to perform the power allocation problem (19) with $\mathrm{SE}(\mathbf{p}, \mathbf{v}_k(\mathbf{p}))$ defined as $\mathrm{SE}_k^{\mathrm{UL},2}$ in (21), the following alternating optimization method is proposed.

1 Set $i \leftarrow 0$ and initialize $\mathbf{p}^0$ randomly such that $\mathbf{p}^0 \in \mathcal{P}$.

2 Determine the combining matrix $\mathbf{V}^{i+1}$ for each coherence block that maximizes:

$$\begin{aligned} \max_{\mathbf{V}} \quad & U(\mathrm{SE}_1, \ldots, \mathrm{SE}_K) \\ \text{s.t.} \quad & \mathrm{SE}_k = \mathrm{SE}_k^{\mathrm{UL},2}(\mathbf{p}^i, \mathbf{v}_k) \quad \forall k \end{aligned} \tag{27}$$

for fixed transmit powers $\mathbf{p}^i$. Since the relation between $\mathrm{SE}_k^{\mathrm{UL},2}(\mathbf{p}^i, \mathbf{v}_k)$ and $\mathbf{v}_k$ is non-trivial, (23) is used to obtain a tractable optimization problem:

$$\begin{aligned} \max_{\mathbf{V}} \quad & U(\mathrm{SE}_1, \ldots, \mathrm{SE}_K) \\ \text{s.t.} \quad & \mathrm{SE}_k = \mathrm{SE}_k^{\mathrm{UL},1}(\mathbf{p}^i, \mathbf{v}_k) \left( \geq \mathrm{SE}_k^{\mathrm{UL},2}(\mathbf{p}^i, \mathbf{v}_k) \right) \forall k. \end{aligned} \tag{28}$$

Due to the nondecreasing function $U(\mathrm{SE}_1, \ldots, \mathrm{SE}_K)$, this is equivalent to maximizing $\mathrm{SINR}_k^{\mathrm{UL,inst}}$ in (14) for each $k$ in each coherence block. The solution is thus given by (15) and (16).

3 Determine the optimal transmit powers $\mathbf{p}^{i+1}$ that maximize:

$$\begin{aligned} \max_{\mathbf{p} \in \mathcal{P}} \quad & U(\mathrm{SE}_1, \ldots, \mathrm{SE}_K) \\ \text{s.t.} \quad & \mathrm{SE}_k = \mathrm{SE}_k^{\mathrm{UL},2}\left(\mathbf{p}, \mathbf{v}_k^{i+1}\right) \quad \forall k. \end{aligned} \tag{29}$$

4 Set $i \leftarrow i+1$ and return to step 2.

To following theorem is relevant to determine the convergence of the proposed alternating optimization method.

**Theorem 3.3** *Suppose that $f(\mathbf{x}_1, \mathbf{x}_2)$ is continuously differentiable over a closed convex set $\mathcal{X} = \mathcal{X}_1 \times \mathcal{X}_2$. Furthermore, suppose that for each $\mathbf{x}^i \in \mathcal{X}$, the maximizations*

$$\mathbf{x}_1^{i+1} = \arg\max_{\mathbf{z}_1 \in \mathcal{X}_1} f\left(\mathbf{z}_1, \mathbf{x}_2^i\right) \tag{30}$$

*and*

$$\mathbf{x}_2^{i+1} = \arg\max_{\mathbf{z}_2 \in \mathcal{X}_2} f\left(\mathbf{x}_1^{i+1}, \mathbf{z}_2\right) \tag{31}$$

*are uniquely attained. Let $\{x^i\}$ be the sequence generated from a random initial point $\mathbf{x}^0 \in \mathcal{X}$ using the above equations. Then every limit point of $\{x^i\}$ is a stationary point of $\max_{\mathbf{x} \in \mathcal{X}} f(\mathbf{x})$.*

***Proof*** The proof is a simplification of the proof in [39, Sect. 2.7] for block coordinate descent methods. Only two blocks are considered here and the maximizations in the theorem need to be replaced with a minimization of $-f(\mathbf{x})$ to obtain the same result. □

Although not fully applicable to the proposed alternating optimization method, this theorem can still be used to show that the alternating optimization method converges to a stationary point of (19) when the number of antennas $NL$ goes to infinity. In that case, the bound (23) is tight, i.e. $\mathrm{SE}_k^{\mathrm{UL},1} \approx \mathrm{SE}_k^{\mathrm{UL},2}$ and this implies that step 3 of the alternating optimization method is optimizing (19) for a fixed $\mathbf{p}^i$. The maximum in optimization problems (27) is uniquely attained whenever $U$ is nondecreasing, so it still needs to be checked if a unique solution to (29) can be found for the given utility function.

Based on this, a distributed UL receive combining and power allocation algorithm (D-UL-RCPA) is proposed in Algorithm 1, where the (16)–(18) is used in each coherence block and the statistics to determine $\mathrm{SE}_k^{\mathrm{UL},2}(\mathbf{p}, \mathbf{v}_k^{i+1})$ are gathered over $B$ coherence blocks before a better power allocation is computed. If the channel statistics $\{\mathbf{R}_{kl}\}_{\forall k,l}$ remain constant and the necessary large-scale statistics are estimated without estimation error, i.e. $B \to \infty$, one can see that the proposed algorithm is basically performing the alternating optimization method with (27)–(28). It exhibits the same convergence behavior in that case. For a finite window $B$ the large-scale statistics for the power allocation step are only obtained approximately but the algorithm will still provide a good performance compared to other methods as will be shown in the simulations sections. Simulations show also that $\mathbf{p}$ is not changing much after 3 iterations if the large-scale statistics remain constant, demonstrating the good tracking performance of the proposed algorithm.

Algorithm 1 also has the advantages listed in Sect. 3.1: The NC only has to deal with $K \times K$ matrices, the communications over the fronthaul network can be done efficiently using in-network sums and the algorithm is robust against link failures. The extra communication of $\mathbf{U}$ and the parameters $\{\xi_{ki}, \nu_k\}_{\forall k,i}$ are only necessary for the power allocation step, so whenever the performance of the estimation task at hand is sufficient, these steps can be removed to save network resources. Furthermore, it can be seen as an adaptive algorithm, since is adapts to slow changes in the channel statistics over time.

**Algorithm 1: Distributed UL MMSE-based receive combining and power allocation (D-UL-RCPA)**

1 Initialize $\mathbf{p} \in \mathcal{P}$.

2 **for** *each coherence block in* $\{1, ..., B\}$ **do**

1 - Each AP $l$ estimates $\hat{\mathbf{H}}_l$ locally and computes $\mathbf{X}_l$

$$\mathbf{X}_l = \left( \sum_{k=1}^{K} p_k \mathbf{C}_{kl} + \mathbf{R}^{\mathsf{UL}}_{\mathbf{n}_l \mathbf{n}_l} \right)^{-1} \hat{\mathbf{H}}_l. \tag{32}$$

- Each AP $l$ transmits the parameters $\mathbf{X}_l^H \hat{\mathbf{H}}_l \in \mathbb{C}^{K \times K}$ and the transformed signals $\mathbf{x}_l = \mathbf{X}_l^H \mathbf{y}_l^{\mathsf{UL}} \in \mathbb{C}^K$ for all received signals in the coherence block to the NC.

2 The network is used to perform in-network sums to obtain

$$\mathbf{x} = \sum_{l=1}^{L} \mathbf{x}_l, \quad \mathbf{X}^H \hat{\mathbf{H}} = \sum_{l=1}^{L} \mathbf{X}_l^H \hat{\mathbf{H}}_l. \tag{33}$$

3 - The NC computes $\mathbf{U}$ using (17) and broadcasts $\mathbf{U}$ to all APs.
- The NC estimates the UL signals as

$$\hat{s}_k = \mathbf{u}_k^H \mathbf{x} \quad \forall k. \tag{34}$$

- The NC updates the large-scale statistics as

$$\gamma_k \leftarrow \gamma_k + \frac{1}{B} \mathbf{u}_k^H \mathbf{X}^H \hat{\mathbf{h}}_k \qquad \forall k, \tag{35}$$

$$\upsilon_{ki} \leftarrow \upsilon_{ki} + \frac{1}{B} |\mathbf{u}_k^H \mathbf{X}^H \hat{\mathbf{h}}_i|^2 \qquad \forall k, i. \tag{36}$$

4 Each AP $l$ updates the local large-scale statistics as

$$\xi_{ki}^l \leftarrow \xi_{ki}^l + \frac{1}{B} \mathbf{u}_k^H \mathbf{X}_l^H \mathbf{C}_{il} \mathbf{X}_l \mathbf{u}_k \qquad \forall k, i, \tag{37}$$

$$\nu_k^l \leftarrow \nu_k^l + \frac{1}{B} \mathbf{u}_k^H \mathbf{X}_l^H \mathbf{R}^{\mathsf{UL}}_{\mathbf{n}_l \mathbf{n}_l} \mathbf{X}_l \mathbf{u}_k \qquad \forall k. \tag{38}$$

**end**

3 The NC gathers the large-scale statistics $\xi_{ki}, \nu_k \ \forall k, i$ (using in-network sums).

4 The NC solves power allocation problem (28) to obtain new values for $\mathbf{p}$. Examples op optimization algorithms will be provided in Subsection 5.1.

5 The NC transmits the new values for $\mathbf{p}$ to all APs and $p_k$ to UE $k \ \forall k$.

6 Reset $\gamma_k, \upsilon_{ki}, \xi_{ki}^l, \nu_k^l \ \forall k, i, l$ and return to step 2.

## 4 DL transmit precoding and power allocation

### 4.1 DL power allocation preliminaries

When a precoding vector $\mathbf{w}_k$ is used to encode data for UE $k$, the desired signal is received at UE $k$ propagates with a gain of $\mathbf{h}_k^H \mathbf{w}_k$. The UE $k$ does not know this gain a priori, but it can estimate its mean value $E\{\mathbf{h}_k^H \mathbf{w}_k\}$, which is a common approach in massive MIMO systems. The received DL signal can then be expressed as

$$y_k^{\mathrm{DL}} = E\{\mathbf{h}_k^H \mathbf{w}_k\}\sqrt{\rho_k}\zeta_k + \left(\mathbf{h}_k^H \mathbf{w}_k - E\{\mathbf{h}_k^H \mathbf{w}_k\}\right)\sqrt{\rho_k}\zeta_k + \sum_{i \neq k} \mathbf{h}_k^H \mathbf{w}_i \sqrt{\rho_i}\zeta_i + n_k^{\mathrm{DL}}. \tag{39}$$

Here $E\{\mathbf{h}_k^H \mathbf{w}_k\}$ represents a deterministic channel and the other terms represent uncorrelated interference and noise. The following achievable SE can then be obtained using the hardening bound [4, 5, 24]:

$$\mathrm{SE}_k^{\mathrm{DL},2} = \frac{\tau_{\mathrm{d}}}{\tau_{\mathrm{c}}} \log_2\left(1 + \mathrm{SINR}_k^{\mathrm{DL}}\right) \text{ [bits/s/Hz]} \tag{40}$$

with $\mathrm{SINR}_k^{\mathrm{DL}}$ given as

$$\frac{\rho_k |E\{\mathbf{w}_k^H \mathbf{h}_k\}|^2}{\sum_{i=1}^K \rho_i E\{|\mathbf{w}_i^H \mathbf{h}_k|^2\} - \rho_k |E\{\mathbf{w}_k^H \mathbf{h}_k\}|^2 + \sigma_k^{2,\mathrm{DL}}}. \tag{41}$$

Here the assumption $E\{||\mathbf{w}_k||^2\} = 1$ is not implied, while this is the case in [4, 5, 24]. The total transmit power allocated to UE $k$ is still given by $\rho_k E\{||\mathbf{w}_k||^2\}$, which is no longer only determined by $\rho_k$, but is also influenced by the statistics of $E\{||\mathbf{w}_k||^2\}$.

The DL power allocation problem with a per AP total power constraint, is finally formulated as

$$\begin{aligned} \max_{\boldsymbol{\rho} \geq \mathbf{0}} \quad & U(\overline{\mathrm{SE}}_1, \ldots, \overline{\mathrm{SE}}_K) \\ \text{s.t.} \quad & \overline{\mathrm{SE}}_k = \mathrm{SE}_k^{\mathrm{DL},2}(\boldsymbol{\rho}, \mathbf{W}(\boldsymbol{\rho})) \quad \forall k, \\ & \sum_{k=1}^K \rho_k \mu_k^l \leq P_t^l \quad \forall l \end{aligned} \tag{42}$$

with $P_t^l$ the maximum transmit power of AP $l$ and

$$\mu_k^l = E\{\mathbf{w}_{kl}^H \mathbf{w}_{kl}\}. \tag{43}$$

### 4.2 UL–DL duality

Instead of solving the primal problem (42) directly, it is proposed to solve the dual problem using a dual subgradient descent procedure. The dual problem to (42) can be found by introducing $L$ dual variables $\boldsymbol{\sigma} = [\sigma_1 \ \ldots \ \sigma_L]^{\mathrm{T}}$ for the $L$ power constraints and an extra common dual variable $\lambda$ resulting in

$$\min_{\boldsymbol{\sigma}^\nu \geq \mathbf{0}, \lambda \geq 0} \max_{\boldsymbol{\rho} \geq \mathbf{0}, \mathbf{W}} \mathcal{L}^{\mathrm{DL}}(\boldsymbol{\rho}, \mathbf{W}, \boldsymbol{\sigma}^\nu, \lambda) \tag{44}$$

where the Lagrangian $\mathcal{L}^{\mathrm{DL}}(\boldsymbol{\rho}, \mathbf{W}, \boldsymbol{\sigma}^{\nu}, \lambda)$ is given as

$$U\left(\mathrm{SE}_1^{\mathrm{DL},2}, \ldots, \mathrm{SE}_K^{\mathrm{DL},2}\right) - \lambda \sum_{l=1}^{L} \sigma_l^{\nu} \left( \sum_{k=1}^{K} \rho_k \mu_k^l - P_t^l \right). \tag{45}$$

For some specific cases with deterministic channels (e.g. [40, 41]), it is possible to show that the duality gap between the primal and dual problem is zero, so that solving the dual problem will result in a primal optimal solution. Unfortunately this is not the case for the general CFmMIMO DL power allocation problem in (42), so convergence to a global optimum can not be guaranteed. However, the proposed algorithm will provide an efficient way of obtaining a set of locally optimal DL power allocation and precoding vectors.

To be able to find the optimal precoding strategy for $\mathbf{W}$, a change of variables is introduced. $\mathrm{SE}_k^{\mathrm{DL},2}$ in (42) is replaced with the expression for the SE of a virtual UL network $\mathrm{SE}_k^{\mathrm{UL},\nu} = \frac{\tau_{\mathrm{d}}}{\tau_{\mathrm{c}}} \log_2(1 + \mathrm{SINR}_k^{\mathrm{UL},\nu})$, depending on the newly introduced variables $\mathbf{p}^{\nu}$, which will replace the DL transmit powers $\boldsymbol{\rho}$. The virtual UL network $\mathrm{SINR}_k^{\mathrm{UL},\nu}$ depends on $\mathbf{p}^{\nu}$, $\mathbf{W}$ and $\boldsymbol{\sigma}$ and is defined as:

$$\frac{p_k^{\nu} \left| E\left\{ \mathbf{w}_k^H \mathbf{h}_k \right\} \right|^2}{\sum_{i=1}^{K} p_i^{\nu} E\left\{ \left| \mathbf{w}_k^H \mathbf{h}_i \right|^2 \right\} - p_k^{\nu} \left| E\left\{ \mathbf{w}_k^H \mathbf{h}_k \right\} \right|^2 + \underbrace{\sum_{l=1}^{L} \sigma_l^{\nu} \mu_k^l}_{\nu_k}}. \tag{46}$$

In order to show that this results in the same allowed region for $\overline{\mathrm{SE}}_k$, it can be shown that any set of DL SEs $\{\mathrm{SE}_k^{\mathrm{DL},2}\}_{\forall k}$ obtained with specific DL transmit powers $\boldsymbol{\rho} \geq \mathbf{0}$ can also be obtained as UL SEs $\{\mathrm{SE}_k^{\mathrm{UL},\nu}\}_{\forall k}$ with specific virtual transmit powers $\mathbf{p}^{\nu} \geq \mathbf{0}$ and visa versa. Equating the virtual UL SE expression $\mathrm{SE}_k^{\mathrm{UL},\nu}$ with the DL SE expression $\mathrm{SE}_k^{\mathrm{DL},2}$ for all UEs, as done in expression (47),

$$\begin{aligned} &\cancel{\rho_k \left| E\{\mathbf{w}_k^H \mathbf{h}_k\} \right|^2} \left( \sum_{i \neq k} p_i^{v} E\{|\mathbf{w}_k^H \mathbf{h}_i|^2\} + \cancel{p_k^{v} (E\{|\mathbf{w}_k^H \mathbf{h}_k|^2\} - |E\{\mathbf{w}_k^H \mathbf{h}_k\}|^2)} + \sum_{l=1}^{L} \sigma_l^{v} \mu_k^l \right) = \\ &\cancel{p_k^{v} \left| E\{\mathbf{w}_k^H \mathbf{h}_k\} \right|^2} \left( \sum_{i \neq k} \rho_i E\{|\mathbf{w}_i^H \mathbf{h}_k|^2\} + \cancel{\rho_k (E\{|\mathbf{w}_k^H \mathbf{h}_k|^2\} - |E\{\mathbf{w}_k^H \mathbf{h}_k\}|^2)} + \sigma_k^{2,\mathrm{DL}} \right) \quad \forall k, \end{aligned} \tag{47}$$

result in the following relation between the UL and DL transmit powers:

$$\boldsymbol{\rho} = \left( \mathbf{M}(\mathbf{p}^{\nu}, \mathbf{W}, \boldsymbol{\sigma}^{\nu}) \right)^{-1} \mathrm{diag}\left\{ \sigma_1^{2,\mathrm{DL}}, \ldots, \sigma_K^{2,\mathrm{DL}} \right\} \mathbf{p}^{\nu} \tag{48}$$

with

$$[\mathbf{M}]_{ki} = \begin{cases} \sum_{i \neq k} p_i^{\nu} E\left\{ \left| \mathbf{w}_k^H \mathbf{h}_i \right|^2 \right\} + \nu_k & \text{if } k = i, \\ -p_k^{\nu} E\left\{ \left| \mathbf{w}_i^H \mathbf{h}_k \right|^2 \right\} & \text{if } k \neq i. \end{cases} \tag{49}$$

The matrix $\mathbf{M}$ is real-valued with positive diagonal elements and non-positive off-diagonal elements, and is column diagonally dominant. Therefore, $\mathbf{M}$ is an $M$-matrix [42]. As

M-matrices are inverse positive, it is clear that for each $\mathbf{p}^\nu \geq \mathbf{0}$, solving (48) will yield DL transmit powers $\boldsymbol{\rho} \geq \mathbf{0}$. Similar results can be shown for the inverse relation, where $\mathbf{p}^\nu$ is computed from $\boldsymbol{\rho}$.

Summing all left- and right-hand sides of (47) for all UEs, results in the following relation between the virtual UL transmit powers and DL transmit powers:

$$\sum_{k=1}^{K} \rho_k \sum_{l=1}^{L} \sigma_l^\nu \mu_k^l = \sum_{k=1}^{K} p_k^\nu \sigma_k^{2,\mathrm{DL}}. \tag{50}$$

As such the dual problem (44) can be written as

$$\min_{\boldsymbol{\sigma}^\nu \geq \mathbf{0}} \min_{\lambda \geq 0} \max_{\mathbf{p}^\nu \geq \mathbf{0}, \mathbf{W}} \mathcal{L}^{\mathrm{UL},\nu}(\mathbf{p}^\nu, \mathbf{W}, \boldsymbol{\sigma}^\nu, \lambda) \tag{51}$$

where the Lagrangian $\mathcal{L}^{\mathrm{UL},\nu}(\mathbf{p}^\nu, \mathbf{W}, \boldsymbol{\sigma}^\nu, \lambda)$ is given as

$$U(\mathrm{SE}_1^{\mathrm{UL},\nu}, \ldots, \mathrm{SE}_K^{\mathrm{UL},\nu}) - \lambda \left( \sum_{k=1}^{K} p_k^\nu \sigma_k^{2,\mathrm{DL}} - \sum_{l=1}^{L} \sigma_l^\nu P_t^l \right). \tag{52}$$

If the expressions for estimating the statistics in Theorem 3.2 are used in (46) and the expectations are replaced by immediate channel estimations, an expression for the virtual instantaneous $\mathrm{SINR}_k^{\mathrm{UL},\nu,\mathrm{inst}}$ is obtained as

$$\frac{p_k^\nu |\mathbf{w}_k^H \hat{\mathbf{h}}_k|^2}{\sum_{i \neq k} p_i^\nu |\mathbf{w}_k^H \hat{\mathbf{h}}_i|^2 + \sum_{i=1}^{K} p_i^\nu \mathbf{w}_k^H \mathbf{C}_i \mathbf{w}_k + \mathbf{w}_k^H \mathbf{R}_{\mathbf{nn}}^{\mathrm{UL},\nu} \mathbf{w}_k} \tag{53}$$

with $\mathbf{R}_{\mathbf{nn}}^{\mathrm{UL},\nu} = \mathrm{Blkdiag}\{\sigma_1^\nu \mathbf{I}_N, \ldots, \sigma_L^\nu \mathbf{I}_N\}$. Again the optimal value for $\mathbf{w}_k$ that maximizes this Rayleigh quotient is given by a closed form formula similar to (15), but with $\mathbf{p}$ and $\mathbf{R}_{\mathbf{nn}}^{\mathrm{UL}}$ replaced by $\mathbf{p}^\nu$ and $\mathbf{R}_{\mathbf{nn}}^{\mathrm{UL},\nu}$.

This instantaneous expression for $\mathrm{SINR}_k^{\mathrm{UL},\nu,\mathrm{inst}}$ can again be used to define a $\mathrm{SE}_k^{\mathrm{UL},\nu,1}$ as

$$\mathrm{SE}_k^{\mathrm{UL},\nu,1} = \frac{\tau_\mathrm{d}}{\tau_\mathrm{c}} E\left\{ \log_2 \left( 1 + \mathrm{SINR}_k^{\mathrm{UL},\nu,\mathrm{inst}} \right) \right\} \text{ [bits/s/Hz]} \tag{54}$$

which provides as before a tight upper bound for the virtual $\mathrm{SE}_k^{\mathrm{UL},\nu}$, i.e.:

$$\mathrm{SE}_k^{\mathrm{UL},\nu} \leq \mathrm{SE}_k^{\mathrm{UL},\nu,1} \tag{55}$$

and this bound is tight for $LN \gg 1$.

Taking the equivalence defined above into account, the combined DL transmit precoding and power allocation problem can be formulated as

$$\begin{aligned} \min_{\substack{\boldsymbol{\sigma}^\nu \geq \mathbf{0} \\ \lambda \geq 0}} \max_{\substack{\mathbf{p}^\nu \geq \mathbf{0} \\ \mathbf{W}}} \; & U(\mathrm{SE}_1, \ldots, \mathrm{SE}_K) - \lambda \left( \sum_{k=1}^{K} p_k^\nu \sigma_k^{2,DL} - \sum_{l=1}^{L} \sigma_l^\nu P_t^l \right) \\ \text{s.t. } & \mathrm{SE}_k = \mathrm{SE}_k^{\mathrm{UL},\nu}(\mathbf{p}^\nu, \mathbf{w}_k, \boldsymbol{\sigma}^\nu) \leq \mathrm{SE}_k^{\mathrm{UL},\nu,1}(\mathbf{p}^\nu, \mathbf{w}_k, \boldsymbol{\sigma}^\nu) \; \forall k \end{aligned} \tag{56}$$

with the optimal value for $\boldsymbol{\rho}$ defined using (48).

### 4.3 DL transmit precoding and power allocation

Since the dual problems (56) is convex in $\boldsymbol{\sigma}^{\nu}$, it can be solved via a subgradient projection algorithm, where the subgradient for $\boldsymbol{\sigma}^{\nu}$ can be found using the proposed UL-DL duality. A subgradient can be derived from (45) as

$$\lambda^{\star}(\boldsymbol{\sigma}^{\nu})\left(P_t^l - \sum_{k=1}^{K} \rho_k^{\star}(\boldsymbol{\sigma}^{\nu}) E\left\{\mathbf{w}_{kl}^{\star,H}(\boldsymbol{\sigma}^{\nu})\mathbf{w}_{kl}^{\star}(\boldsymbol{\sigma}^{\nu})\right\}\right) \tag{57}$$

where $\boldsymbol{\rho}^{\star}(\boldsymbol{\sigma}^{\nu}), \lambda^{\star}(\boldsymbol{\sigma}^{\nu})$ and $\mathbf{W}^{\star}(\boldsymbol{\sigma}^{\nu})$ are the optimal solutions of $\min_{\lambda\geq 0} \max_{\boldsymbol{\rho}\geq\mathbf{0},\mathbf{W}} \mathcal{L}^{\mathrm{DL}}(\boldsymbol{\rho}, \mathbf{W}, \boldsymbol{\sigma}^{\nu})$ and can also be obtained using the optimal solutions of $\min_{\lambda\geq 0} \max_{\boldsymbol{\rho}\geq\mathbf{0},\mathbf{W}} \mathcal{L}^{\mathrm{UL},\nu}(\mathbf{p}^{\nu}, \mathbf{W}, \boldsymbol{\sigma}^{\nu})$ and using (48) to obtain $\boldsymbol{\rho}^{\star}(\boldsymbol{\sigma}^{\nu})$. By inspecting $\min_{\lambda\geq 0} \max_{\mathbf{p}^{\nu}\geq\mathbf{0},\mathbf{W}} \mathcal{L}^{\mathrm{UL},\nu}(\mathbf{p}^{\nu}, \mathbf{W}, \boldsymbol{\sigma}^{\nu})$, it can be shown that this is the dual problem of the following virtual UL problem:

$$\begin{aligned} \max_{\mathbf{p}^{\nu}\geq\mathbf{0},\mathbf{W}} \quad & U\left(\mathrm{SE}_1^{\mathrm{UL},\nu}, \ldots, \mathrm{SE}_K^{\mathrm{UL},\nu}\right) \\ \text{s.t.} \quad & \sum_{k=1}^{K} p_k^{\nu}\sigma^{2,\mathrm{DL}} \leq \sum_{l=1}^{L} \sigma_l^{\nu} P_t^l. \end{aligned} \tag{58}$$

The optimal solutions for $\boldsymbol{\rho}^{\star}(\boldsymbol{\sigma}^{\nu})$ and $\mathbf{W}^{\star}(\boldsymbol{\sigma}^{\nu})$ can alternatively be found from this virtual UL problem, which is a particular instance of problem (19) and can thus be solved using Algorithm 1.

However, as computing the required maximization for (57) in the CFmMIMO setup will require multiple iterations of Algorithm 1 over different coherence blocks, this will result in a slow tracking speed. Therefore, it is proposed to use the solutions after one iteration of the alternating optimization method to update the dual variables. Also all the subgradients (57) are scaled with a common factor $\lambda^{\star}(\boldsymbol{\sigma}^{\nu})$, which can be compensated by the stepsize, so its computation can be avoided. The simulations in Sect. 6 will show that the convergence behavior is not affected.

The proposed algorithm to solve the DL transmit precoding and power allocation problem can thus be summarized as follows:

1. Set $i \rightarrow 0$ and initialize $\mathbf{p}^{\nu,0} \geq \mathbf{0}, \boldsymbol{\sigma}^{\nu,0} \geq \mathbf{0}$ randomly
2. Determine the precoding matrix $\mathbf{W}^{i+1}$ for each coherence block that maximizes:

$$\begin{aligned} \max_{\mathbf{W}} \quad & U(\mathrm{SE}_1, \ldots, \mathrm{SE}_K) \\ \text{s.t.} \quad & \mathrm{SE}_k = \mathrm{SE}_k^{\mathrm{UL},\nu,1}\left(\mathbf{p}^{\nu,i}, \mathbf{w}_k, \boldsymbol{\sigma}^{\nu,i}\right) \quad \forall k \end{aligned} \tag{59}$$

   for a fixed transmit power $\mathbf{p}^{\nu,i}$ and $\boldsymbol{\sigma}^{\nu,i}$ by using the virtual quantities in the MMSE-expression (15) or (16).
3. Determine the optimal transmit powers $\mathbf{p}^{\nu,i+1}$ that maximizes:

$$\begin{aligned} \max_{\mathbf{p}^\nu \geq \mathbf{0}} \quad & U(\mathrm{SE}_1, \ldots, \mathrm{SE}_K) \\ \text{s.t.} \quad & \mathrm{SE}_k = \mathrm{SE}_k^{\mathrm{UL},\nu}\left(\mathbf{p}^\nu, \mathbf{w}_k^{i+1}, \boldsymbol{\sigma}^{\nu,i}\right) \quad \forall k \\ & \sum_{k=1}^{K} p_k^\nu \sigma_k^{2,DL} \leq \sum_{l=1}^{L} \sigma_l P_t^l \end{aligned} \tag{60}$$

and compute $\boldsymbol{\rho}^{i+1}$ using (48).

4. Update the dual variables with stepsize $t_i$ using the current estimate for the subgradient:

$$\sigma_l^{i+1} = \max\left(\sigma_l^i - t_i\left(P_t^l - \sum_{k=1}^{K} \rho_k^{i+1}\mu_k^{l,i+1}\right), 0\right) \forall l. \tag{61}$$

5. Set $i \leftarrow i+1$ and return to step 2.

Algorithm 2 presents the corresponding distributed DL tranmsit beamforming and power allocation algorithm (DL-D-DTTP). Algorithm 2 also has the advantages of Algorithm 1, since the communications over the fronthaul network and the local computations are similar. The introduction of the dual variables $\boldsymbol{\sigma}$ and the virtual uplink network allows to make abstraction of the DL transmit power constraints in (44) and transforms them into a virtual noise source using $\mathbf{R}_{\mathbf{nn}}^{\mathrm{UL},\nu}$. The virtual uplink problem has only one constraint related to the virtual uplink powers and can then be efficiently solved using Algorithm 1. The dual subgradient method is used to update the dual variables, which can be done efficiently and locally at each AP, without requiring the transmission of the transmit precoder $\mathbf{W}_l = \mathbf{X}_l\mathbf{U}$ to the other APs or NC.

It is possible for an AP to include additional constraints on its antenna powers. Examples are per-antenna power constraints or per-cluster constraints when an AP is a concatenation of different antenna arrays with local power amplifiers. The proposed algorithm can be extended by introducing an additional dual variable $\sigma_{l,m}^\nu$ for each additional constraint locally at AP $l$ and redefining $\nu_k^l$ as $\sum_m \sigma_{l,m}^\nu \nu_k^{l,m}$ where the sum is over all the local constraints and $\nu_k^{l,m}$ is the measured power relevant for the $m$-th constraint. The dual variable update can be extended in a similar way. The virtual noise correlation matrix $\mathbf{R}_{\mathbf{n}_l\mathbf{n}_l}^{\mathrm{UL},\nu}$ in (62) should then include the extra dual variables on the corresponding positions. The introduction of the additional constraints at an AP increases the computational load of an AP, but not of the NC.[6]

[6] Note that the per-AP power constraint will only hold upon convergence of the subgradient algorithm. Therefore, it may be necessary to rescale the DL transmit powers locally at each AP $l$ such that $\sum_{k=1}^{K} \rho_{kl}\mu_k^l \leq P_t^l$ before transmitting the DL signal. This will not influence the convergence of the algorithm.

**Algorithm 2: Distributed DL MMSE-based transmit precoding and power allocation (D-DL-DTPA)**

Initialize $\mathbf{p}^v, \boldsymbol{\sigma}^v$ and $\boldsymbol{\rho}$ randomly[6].

**for** *each coherence block in* $\{1, ..., B\}$ **do**

1 - Each AP $l$ obtains a local estimate of $\hat{\mathbf{H}}_l$ and computes $\mathbf{X}_l$ as

$$\mathbf{X}_l = \left( \sum_{k=1}^{K} p_k^v \mathbf{C}_{kl} + \sigma_l^v \mathbf{I}_N \right)^{-1} \hat{\mathbf{H}}_l. \tag{62}$$

- Each AP $l$ transmits the parameters $\mathbf{X}_l^H \hat{\mathbf{H}}_l \in \mathbb{C}^{K \times K}$ to the NC.

2 The network is used to perform in-network sums to obtain

$$\mathbf{X}^H \hat{\mathbf{H}} = \sum_{l=1}^{L} \mathbf{X}_l^H \hat{\mathbf{H}}_l. \tag{63}$$

3 - The NC computes $\mathbf{U}$ using (17) and broadcasts $\mathbf{U}$ to all APs.
- The NC updates the large-scale statistics as

$$\gamma_k \leftarrow \gamma_k + \frac{1}{B} \mathbf{u}_k^H \mathbf{X}^H \hat{\mathbf{h}}_k \qquad \forall k, \tag{64}$$

$$\upsilon_{ki} \leftarrow \upsilon_{ki} + \frac{1}{B} |\mathbf{u}_k^H \mathbf{X}^H \hat{\mathbf{h}}_i|^2 \qquad \forall k, i. \tag{65}$$

4 - Each AP $l$ transmits the DL signal

$$\sum_{k=1}^{K} \mathbf{X}_l \mathbf{u}_k \sqrt{\rho_k} \zeta_k. \tag{66}$$

- Each AP $l$ updates the large-scale statistics

$$\xi_{ki}^l \leftarrow \xi_{ki}^l + \frac{1}{B} \mathbf{u}_k^H \mathbf{X}_l^H \mathbf{C}_{il} \mathbf{X}_l \mathbf{u}_k \qquad \forall k, i, \tag{67}$$

$$\mu_k^l \leftarrow \mu_k^l + \frac{1}{B} \mathbf{u}_k^H \mathbf{X}_l^H \mathbf{X}_l \mathbf{u}_k \qquad \forall k, \tag{68}$$

$$\nu_k^l \leftarrow \sigma_l^v \mu_k^l \qquad \forall k. \tag{69}$$

**end**

3 The NC gathers the large-scale statistics $\xi_{ki}, \nu_k, \sum_{l=1}^{L} \sigma_l P_t^l$ for all UEs $k, i$ (using in-network sums).

4 The NC solves power allocation problem (60) for fixed precoding vectors to obtain new values for $\mathbf{p}^v$ and find the new value for $\boldsymbol{\rho}$ using (48).

5 The NC transmits the new values for $\mathbf{p}^v, \boldsymbol{\rho}$ to all APs.

6 Each AP $l$ performs a dual update as in (61)

7 Reset $\gamma_k, \upsilon_{ki}^l, \xi_{ki}, \nu_k^l \ \forall k, i, l$ and return to step 2.

## 5 Further considerations

### 5.1 Examples of power allocation strategies

In this subsection, two different utility functions are introduced, which will be used for the simulations in Sect. 6. It is repeated that the requirements are that the utility function is monotonically nondecreasing in each argument $\mathrm{SE}_k$ and that the power allocation problem in (29) and (60) have a unique maximum that can be obtained. The details of

the algorithms used to obtain the optimal solution for (29) and (60) can be found in the provide references.

The first utility function is the *non-weighted max-sum SE*, which represents the total number of bits per second that are transmitted without considering how these bits are assigned between the UEs. The max-sum SE problem can be expressed as

$$U(\mathrm{SE}_1,\ldots,\mathrm{SE}_K) = \sum_{k=1}^{K} \mathrm{SE}_k. \tag{70}$$

Note that the above problem is usually not convex in $\mathbf{p}$ or $\boldsymbol{\rho}$ for given combining/precoding matrices $\mathbf{V}^{i+1}, \mathbf{W}^{i+1}$ and, hence, it is hard to obtain the optimal solution [43]. There exist global optimization methods [44, 45], but their computational complexity is unsuitable for real-time applications. A pragmatic solution is to instead settle for a local optimum. Common approaches for finding a local optimum to the max-sum SE problem are the weighted MMSE method [46] or successive convex optimization [47], which result in iterative algorithms. Schemes using machine learning can potentially find better solutions [20, 21], but require some off-line training and will not provide guaranteed convergence to the optimal solution. Therefore, successive convex optimization [47] will be used for the simulations in Sect. 6.

The max-sum SE intrinsically provides a higher SE for users with good channel conditions, while leaving users with worse channel conditions at a disadvantage. Therefore the goal of the second utility function is to provide uniform fairness for all users. The aim of the *non-weighted max–min SE fairness* is to maximize the minimum SE among all the UEs in the network:

$$U(\mathrm{SE}_1,\ldots,\mathrm{SE}_K) = \min_k \mathrm{SE}_k. \tag{71}$$

Since the SE of UE $k$ is an increasing function of the effective $\mathrm{SINR}^{\mathrm{UL},2}$ or $\mathrm{SINR}^{\mathrm{UL},\nu}$ respectively, maximizing the minimum SE is the same as maximizing the minimum effective SINR among all the UEs. This quasi-concave problem can be solved using the epigraph trick [48] combined with a bisection method [4, 5, 47, 49] or using fixed-point algorithms [50]. Since all these algorithms converge to a global maximum, the methods in [5] is chosen for the simulations in Sect. 6.

Other utility functions that result in convex optimization problems for fixed combining/precoding matrices $\mathbf{V}^{i+1}, \mathbf{W}^{i+1}$ [44], are for example the weighted geometric mean $\prod_k \mathrm{SE}_k^{w_k}$ and the weighted harmonic mean $(\sum_k w_k/\mathrm{SE}_k)^{-1}$ for some positive weights $\{w_k\}_{\forall k}$.

### 5.2 Network topologies

The proposed algorithms have been described as if the network has a tree topology, where all the APs are connected and the NC is the root node. However, they are applicable to any network topology since selecting a tree topology is always possible whenever all the APs are connected [51]. The tree topology allows for an efficient way of computing the required in-network sums in D-UL-RCPA and D-DL-TPPA by starting from the leaf nodes and summing up all the required data when moving towards the root node. Therefore, many current technologies like radio-stripes [52], very large

**Table 2** Number of complex scalars communicated over each fronthaul link (either from AP to NC or from NC to AP) during $B$ coherence blocks

| | Channel estimation | UL signal estimation | UL power allocation | DL data transmission | DL power allocation |
|---|---|---|---|---|---|
| Network-wide SEPA and DTPA | $B\tau_p NL$ | $B\tau_u NL$ | – | $B(\tau_d K + NL)$ | – |
| D-UL-SEPA and D-DL-DTPA | – | $B(\tau_u K + K^2)$ | $K^2 + K$ | $B\tau_d K + K^2$ | $K^2 + K$ |

**Table 3** Key parameters of considered simulation setup, based on [17]

| Parameter | Value |
|---|---|
| Network area | 1 km × 1 km |
| Number of APs | $L = 50$ |
| Number of antennas per AP | $N = 2, 4, 8, 16$ |
| Frounthaul communication gain $\mathcal{O}(\frac{NL}{K})$ | $\mathcal{O}(2), \mathcal{O}(4), \mathcal{O}(8), \mathcal{O}(16)$ |
| Bandwidth | 20 MHz |
| Noise model | $\mathbf{R}_{\mathbf{nn}}^{\mathrm{UL}} = \sigma^2 \mathbf{I}_M, \sigma_k^{2,\mathrm{DL}} = \sigma^2 \; \forall k$ |
| Noise power | $\sigma^2 = -94$ dBm |
| Maximum UL transmit power | 100 mW |
| Maximum DL transmit power | $P_t = 1000$ mW |
| Samples per coherence block | $\tau_c = 200$ |
| Channel gain at 1 km | − 140.6 dB |
| Pathloss exponent | 3.67 |
| Height difference between AP and UE | 10 m |
| Standard deviation of shadow fading | 4 |
| Number of coherence block | $B = 100$ |
| Number of Monte Carlo simulations | 50 |

aperture mMIMO [53], distributed mMIMO [54], etc. can be used to realize the proposed algorithms.

Table 2 provides an overview of the required communications in one iteration $i$ of the proposed algorithms compared to a network-wide setup, where each AP merely transmits its received signals to the NC. Note that quantities like $\mathbf{X}_l^H \hat{\mathbf{H}}_l$ are Hermitian symmetric, so only transmitting the upper blockdiagonal part suffices. The big advantage of the proposed algorithms is that the training signals do not need to be transmitted to the NC and that only $K$ signals are broadcast over the different links instead of $NL$. Therefore the gain in fronthaul communication can be expressed as $\mathcal{O}(\frac{NL}{K})$. Another advantage of the proposed algorithms is that the required large-scale statistics $\mathbf{R}_{kl}$ have to be available only at AP $l$ and that the processing requirement of the NC is relaxed, i.e. it only has to invert $K \times K$ matrices instead of $NL \times NL$ matrices.

## 6 Numerical simulations

Numerical simulations are used in this section to demonstrate the performance and convergence properties of the proposed algorithms. The parameters used for the setup are presented in Table 3. The propagation model in [24, Sec. 4.1.3] with spatially correlated fading is used. $\tau_{\mathrm{u}} = 190$ and $\tau_{\mathrm{d}} = 190$ are used respectively when evaluating

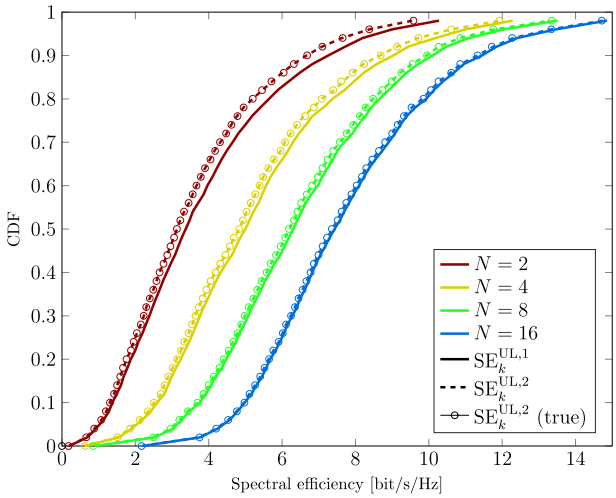



**Fig. 3** Different UL SE for MMSE combining with fixed UE transmit power

the SE expressions for UL and DL. A per UE transmit power constraint is used for UL power allocation and a per AP transmit power constraint for DL power allocation. The presented figures and tables are the results averaged over 50 Monte Carlo simulations. The APs and UEs are distributed uniformly at random over the area with a wrap-around technique to determine the shortest distance between a UE and the APs in each Monte Carlo simulation. $B = 100$ channel realizations are used to estimate the expectations and to estimate the large-scale statistic of Table 1.

## 6.1 Numerical simulations UL

### *6.1.1 Estimation method*

The first UL simulations are conducted to investigate the validity of Theorem 3.2 and the tightness of (23). This is done for a scenario where all UEs are transmitting with maximum power $p_k = 100$ mW$\forall k$ and MMSE-based combining is used. The results for different numbers of antennas per AP are shown in Fig. 3.

$\mathrm{SE}_k^{\mathrm{UL},2}$ and the $\mathrm{SE}_k^{\mathrm{UL},2}$ (true), which is the result obtained when the true channel is used in the expectation, are almost perfectly overlapping for all the considered setups. This shows that when a sufficiently large number of samples are used to estimate the quantities of Theorem 3.2, the required assumptions hold for the MMSE-based combining vector. Here $B = 100$ samples are used for the expectations, so this provides already a good estimate of the average channel statistics.

Regarding the tightness of the bound in (23) which can be observed by comparing $\mathrm{SE}_k^{\mathrm{UL},1}$ and the $\mathrm{SE}_k^{\mathrm{UL},2}$, it can be observed that $\mathrm{SE}_k^{\mathrm{UL},2}$ is always a lower bound for $\mathrm{SE}_k^{\mathrm{UL},1}$, but this bound becomes tighter as the number of antennas per AP increases. For the setup with $N = 16$, there is almost no difference between the different SEs, which validates the assumption of equality in the proposed algorithms.

### *6.1.2 Convergence behavior*

To investigate the convergence behavior of D-UL-RCPA, the utility function over different iterations is plotted in Fig. 4, both for max-sum and max–min power allocation. The

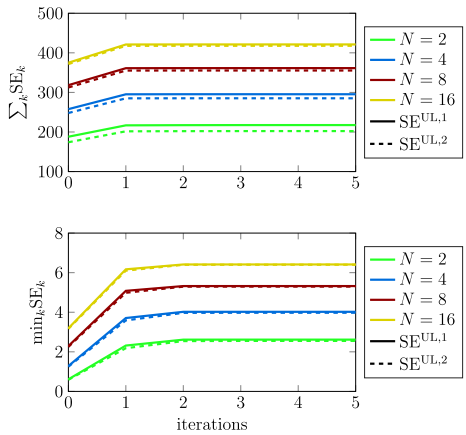



**Fig. 4** Convergence behavior of $\sum_{k=1}^{K} \mathrm{SE}_k$ and $\min_k \mathrm{SE}_k$ for D-UL-RCPA over different iterations with max-sum (upper figure) and max–min (lower figure) power allocation and a per UE transmit power constraint

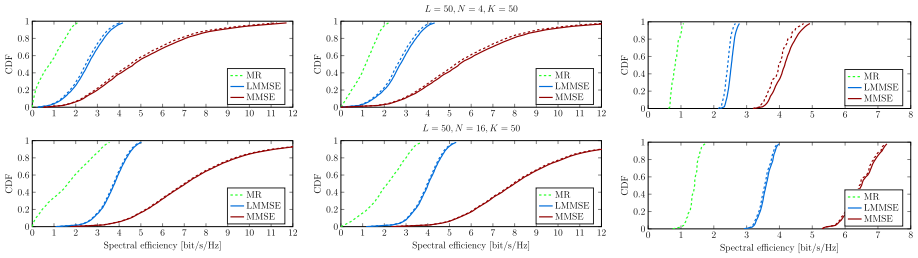
iterations are started from a random feasible power allocation $\mathbf{p}^0 \geq \mathbf{0}$. The results are plotted for both $\mathrm{SE}_k^{\mathrm{UL},1}$ and $\mathrm{SE}_k^{\mathrm{UL},2}$ and for different numbers of antennas.

As expected, both $\sum_{k=1}^{K} \mathrm{SE}_k^{\mathrm{UL},1}$ and $\sum_{k=1}^{K} \mathrm{SE}_k^{\mathrm{UL},2}$ for max-sum power allocation and $\min_k \mathrm{SE}_k^{\mathrm{UL},1}$ and $\min_k \mathrm{SE}_k^{\mathrm{UL},2}$ for max–min power allocation are monotonically increasing and reach an equilibrium point after 3 iterations for each setup. This is even the case for the max-sum power allocation, although the power allocation problem is here only solved to local optimality. The fast convergence and significant performance improvement are desirable properties of D-UL-RCPA.

#### *6.1.3 Performance*

The performance of the combining vectors and power allocation obtained after convergence of D-UL-RCPA is compared with heuristic distributed combining methods, namely MR combining and Local MMSE (LMMSE) combining. In MR combining, the local combining vector is chosen as $\mathbf{v}_{kl} = \hat{\mathbf{h}}_{kl}$ for UE $k$ at AP $l$. In LMMSE, the combining vector is chosen to minimize the local MMSE criterion $E\{||s_k - \mathbf{v}_{kl}^H \mathbf{y}_l||^2 | \hat{\mathbf{H}}_l\}$ and is given as $\mathbf{v}_{kl} = p_k \left( \hat{\mathbf{H}}_l \mathbf{P} \hat{\mathbf{H}}_l^H + \sum_{i=1}^{K} p_i \mathbf{C}_{il} + \mathbf{R}_{\mathbf{n}_l \mathbf{n}_l}^{\mathrm{UL}} \right)^{-1} \hat{\mathbf{h}}_{kl}$. The optimal power allocation for both combining schemes is chosen similar as in D-UL-RCPA, but with $\mathbf{U}$ fixed to an identity matrix $\mathbf{I}_K$. The optimal power allocation for MR combining is then found in one iteration, since $\mathbf{v}_{kl}$ is independent of $\mathbf{p}$. For LMMSE combining this will result in an iterative algorithm, but convergence is here also observed after 3 iterations. The cumulative distribution function (CDF) of the UL SE per UE is provided in Fig. 5 for 3 kinds of power allocation: uniform power allocation (where each UE transmits with maximal power), max-sum and max–min power allocations and for setups with either $N = 4$ antennas per AP or $N = 16$ antennas per AP. The values of the utility function compared to uniform power allocation are given in Table 4 for max-sum and Table 5 for max–min power allocation.

It is clear that the proposed power allocation algorithm improves the utility function for all the presented receive combining methods. MR combining requires no inversion, but its performance is always inferior to LMMSE combining. Even though LMMSE




(a) Uniform power allocation.

(b) Max-sum power allocation.

(c) Max-min power allocation.

**Fig. 5** UL SE of $\mathrm{SE}^{\mathrm{UL},1}$ (–) and $\mathrm{SE}^{\mathrm{UL},2}$ (- -) for different combining schemes with uniform, max-sum and max–min power allocation

**Table 4** $\sum_{k=1}^{K}\mathrm{SE}_k$ after convergence of D-UL-RCPA (uniform power allocation | max-sum power allocation)

| | $L=50, N=8, K=50$ | | $L=50, N=16, K=50$ | | $L=100, N=8, K=50$ | |
|---|---|---|---|---|---|---|
| | $\mathrm{SE}^{\mathrm{UL},2}$ | $\mathrm{SE}^{\mathrm{UL},1}$ | $\mathrm{SE}^{\mathrm{UL},2}$ | $\mathrm{SE}^{\mathrm{UL},1}$ | $\mathrm{SE}^{\mathrm{UL},2}$ | $\mathrm{SE}^{\mathrm{UL},1}$ |
| MR | 42.5\|55.1 | – | 82.0\|104.1 | – | 89.2\|121.3 | – |
| LMMSE | 122.3\|129.3 | 127.7\|135.0 | 184.7\|197.4 | 187.0\|199.7 | 195.5\|206.6 | 198.3\|209.7 |
| MMSE | 248.0\|285.2 | 257.7\|295.4 | 371.6\|417.8 | 375.0\|421.2 | 389.0\|441.3 | 393.4\|447.0 |

**Table 5** $\min_k \mathrm{SE}_k$ after convergence of D-UL-RCPA (uniform power allocation | max–min power allocation)

| | $L=50, N=8, K=50$ | | $L=50, N=16, K=50$ | | $L=100, N=8, K=50$ | |
|---|---|---|---|---|---|---|
| | $\mathrm{SE}^{\mathrm{UL},2}$ | $\mathrm{SE}^{\mathrm{UL},1}$ | $\mathrm{SE}^{\mathrm{UL},2}$ | $\mathrm{SE}^{\mathrm{UL},1}$ | $\mathrm{SE}^{\mathrm{UL},2}$ | $\mathrm{SE}^{\mathrm{UL},1}$ |
| MR | 0.0\|0.8 | – | 0.0\|1.4 | – | 0.0\|1.6 | – |
| LMMSE | 0.7\|2.3 | 0.7\|2.4 | 2.0\|3.4 | 2.1\|3.4 | 2.2\|3.7 | 2.3\|3.7 |
| MMSE | 1.2\|4.0 | 1.3\|4.0 | 3.2\|6.4 | 3.2\|6.4 | 3.5\|7.0 | 3.5\|7.1 |

combining requires an inversion of an $N \times N$ matrix in each coherence block at each AP, while for MMSE-based combining this inversion only has to be performed after $B$ coherence blocks at an AP $l$, its performance is always inferior to MMSE-based combining.

## 6.2 Numerical simulations DL

### *6.2.1 Convergence behavior*

D-DL-TPPA is run for 40 iterations using a stepsize of $t_i = \min(1, 10/i)$ and initialized with virtual UL powers $p_k^{\mathrm{v},0} = 100$ mW $\forall k$ and a random feasible DL power allocation $\boldsymbol{\rho}^0 \geq \mathbf{0}$. The utility function over different iterations is plotted in Fig. 6, both for max-sum and max–min power allocation. To verify if the constraints in all the APs are satisfied, the maximal constraint variation $\max_l \sum_{k=1}^{K} \rho_k E\{\mathbf{w}_{kl}^H \mathbf{w}_{kl}^H\}$ is shown in Fig. 7 over the different iterations. The per AP transmit power constraints are satisfied when this is below 1000 mW.

While difficult the observe in Fig. 6, the utility function is no longer monotonically increasing since the proposed method is based on a dual subgradient projection algorithm. However, the objective remains approximately constant after 4 iterations.

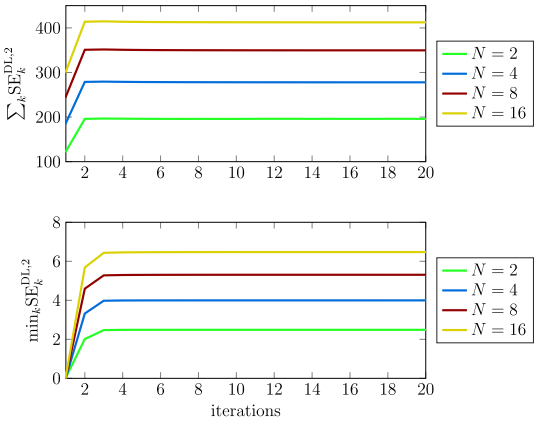



**Fig. 6** Convergence behavior of $\sum_{k=1}^{K} \mathrm{SE}_k^{\mathrm{DL},2}$ and $\min_k \mathrm{SE}_k^{\mathrm{DL},2}$ for D-DL-TPPA over different iterations with max-sum (upper figure) and max–min (lower figure) power allocation and a per AP transmit power constraint

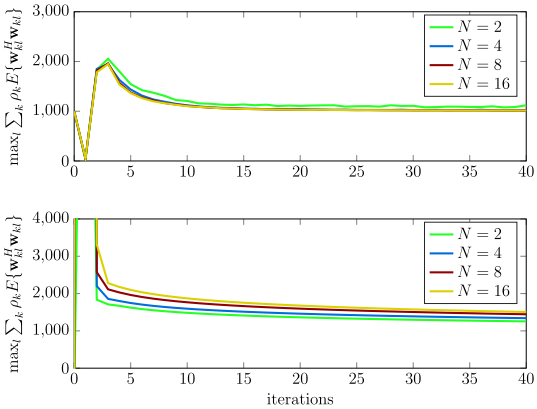
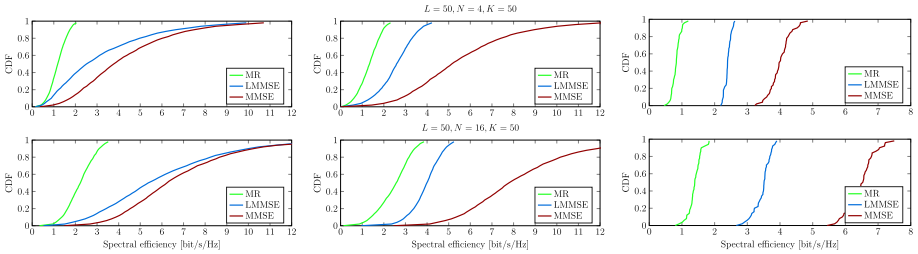



**Fig. 7** Convergence behavior of maximal constraint violation $\max_l \sum_{k=1}^{K} \rho_k E\{\mathbf{w}_{kl}^H \mathbf{w}_{kl}^H\}$ of D-DL-TPPA over different iterations with max-sum and max–min power allocation and a per AP transmit power constraint

Looking at the constraint violation in Fig. 7, it can be observed that only for $N = 4, 8, 16$ and max-sum power allocation convergence is reached after 40 iterations. For $N = 2$ is fluctuating around 1100 mW, probably because the APs have a lower degree of freedom (fewer antennas) so need more iterations with smaller stepsizes to converge. The constraint violations for max–min power allocation are converging very slowly, which is typical for dual subgradient methods. However, as will be shown in the next subsection, the performance after 40 iterations is still superior even when the DL transmit powers are normalized to satisfy the constraints.

#### *6.2.2 Performance*

Finally, the performance of the precoding vectors and power allocation obtained after 40 iterations of D-DL-TPPA is again compared to MR precoding and LMMSE precoding. To make sure that the per AP transmit power constraint is not violated during the evaluation, the DL transmit powers are normalized at the APs where this constraint is violated. The heuristic power allocation method proposed in [17] is used to determine the DL transmit powers for full power transmission. The cumulative distribution function




(a) Full power transmission. (b) Max-sum power allocation. (c) Max-min power allocation.

**Fig. 8** **a** Full power transmission. **b** Max-sum power allocation. **c** Max–min power allocation

**Table 6** $\sum_{k=1}^{K} \mathrm{SE}^{\mathrm{DL},2}$ after convergence of D-DL-TPPA (full power transmission | max-sum power allocation)

| | $L = 50, N = 8, K = 50$ | $L = 50, N = 16, K = 50$ | $L = 100, N = 8, K = 50$ |
|---|---|---|---|
| MR | 59.8\|65.7 | 111.3\|124.2 | 124.4\|145.0 |
| LMMSE | 162.0\|128.0 | 297.2\|197.4 | 325.9\|216.3 |
| MMSE | 214.3\|278.0 | 336.0\|412.5 | 369.7.0\|453.2 |

**Table 7** $\min_k \mathrm{SE}^{\mathrm{DL},2}$ after convergence of D-DL-TPPA (full power transmission | max–min power allocation)

| | $L = 50, N = 8, K = 50$ | $L = 50, N = 16, K = 50$ | $L = 50, N = 16, K = 50$ |
|---|---|---|---|
| MR | 0.4\|0.8 | 0.9\|1.4 | 1.0\|1.5 |
| LMMSE | 0.4\|2.4 | 1.4\|3.5 | 1.6\|3.9 |
| MMSE | 0.9\|4.0 | 2.6\|6.5 | 2.8\|7.2 |

(CDF) of the DL SE per UE are provided in Fig. 8 for 3 different power allocations: full power transmission, max-sum and max–minpower allocation, and for setups with either $N = 4$ antennas per AP or $N = 16$ antennas per AP. The values of the utility function compared to full power transmission are given in Table 6 for max-sum and Table 7 for max–min power allocation.

Similar observations as for the UL simulations can be made concerning the performance of D-DL-TPPA. The proposed power allocation algorithm improves the utility function for all the presented precoding methods, MR combining requires less computational effort, but its performance is inferior and MMSE-based combining outperforms LMMSE combining.

## 7 Conclusion

This paper has presented distributed MMSE-based receive combining, transmit beamforming and power allocation strategies for both UL as well as DL data transmission in CFmMIMO systems. The necessary fronthaul communications to estimate the combining/precoding vectors and the necessary large-scale channel statistics are reduced to a minimum and rely on in-network summation that can be accomplished whenever the

APs can be arranged into a tree-topology. The computations are also distributed over the different APs and the NC. It is shown that the MMSE-based algorithms reach optimal performance.

In this paper it is assumed that all APs serve all the UEs in the network and the data of all the UEs is transmitted to one central NC. This is not scalable when the network size and the number of UEs grow larger. Therefore in future work, the current algorithms will be used as benchmark algorithms when designing future scalable algorithms in which not all APs serve all users and certain APs take some of the functiona;ity of the NC. To not revert to cellular systems, a user-centric approach [17, 32] will be followed, in which each UE can determine its serving APs.

## Appendix 1: Proof of Theorem 3.1

***Proof*** The combining matrix $\mathbf{V}^{\text{MMSE}} = [\mathbf{v}_1^{\text{MMSE}} ... \mathbf{v}_K^{\text{MMSE}}]$ can be written as

$$\mathbf{V}^{\text{MMSE}} = \left( \hat{\mathbf{H}}\mathbf{P}\hat{\mathbf{H}}^H + \underbrace{\sum_{i=1}^{K} p_i \mathbf{C}_i + \mathbf{R}_{\mathbf{nn}}^{\text{UL}}}_{\mathbf{T}} \right)^{-1} \hat{\mathbf{H}}\mathbf{P}. \tag{72}$$

The matrix $\mathbf{V}^{\text{MMSE}}$ can further be written as

$$\begin{aligned}
&= \left(\hat{\mathbf{H}}\mathbf{P}\hat{\mathbf{H}}^H + \mathbf{T}\right)^{-1} \hat{\mathbf{H}}\mathbf{P} \\
&= \left[\mathbf{T}^{-1} - \mathbf{T}^{-1}\hat{\mathbf{H}}\left(\mathbf{P}^{-1} + \hat{\mathbf{H}}^H\mathbf{T}^{-1}\hat{\mathbf{H}}\right)^{-1}\hat{\mathbf{H}}^H\mathbf{T}^{-1}\right]\hat{\mathbf{H}}\mathbf{P} \\
&= \mathbf{T}^{-1}\hat{\mathbf{H}}\left(\mathbf{P}^{-1} + \hat{\mathbf{H}}^H\mathbf{T}^{-1}\hat{\mathbf{H}}\right)^{-1} \\
&\quad \underbrace{\left(\mathbf{P}\left(\mathbf{P}^{-1} + \hat{\mathbf{H}}^H\mathbf{T}^{-1}\hat{\mathbf{H}}\right)^{-1} - \hat{\mathbf{H}}^H\mathbf{T}^{-1}\hat{\mathbf{H}}\mathbf{P}\right)}_{\mathbf{I}} \\
&= \begin{bmatrix} \mathbf{X}_1 \\ \vdots \\ \mathbf{X}_L \end{bmatrix} \left(\mathbf{P}^{-1} + \mathbf{X}^H\hat{\mathbf{H}}\right)^{-1}
\end{aligned} \tag{73}$$

with $\mathbf{X}_k$ and $\mathbf{X}^H\hat{\mathbf{H}}$ defined in (18) and (17) respectively. In (73) the Sherman-Morrison-Woodbury formula is used together with the fact that $\mathbf{T}$ is a block-diagonal matrix. □

## Appendix 2: Proof of Theorem 3.2

***Proof*** By using $\mathbf{h}_k = \hat{\mathbf{h}}_k + \tilde{\mathbf{h}}_k$ with $\hat{\mathbf{h}}_k$ and $\tilde{\mathbf{h}}_k$ independent random variables, the required expectations for Theorem 3.2 can be written as:

$$\begin{aligned} E\{\mathbf{v}_k^H \mathbf{h}_k\} =& E\left\{\mathbf{v}_k^H \hat{\mathbf{h}}_k\right\} + E\left\{\mathbf{v}_k^H \tilde{\mathbf{h}}_k\right\} \\ =& E\left\{\mathbf{v}_k^H \hat{\mathbf{h}}_k\right\}, \end{aligned} \tag{74}$$

$$\begin{aligned} E\left\{\left|\mathbf{v}_k^H \mathbf{h}_i\right|^2\right\} =& E\left\{\mathbf{v}_k^H \left(\hat{\mathbf{h}}_i + \tilde{\mathbf{h}}_i\right)\left(\hat{\mathbf{h}}_i + \tilde{\mathbf{h}}_i\right)^H \mathbf{v}_k\right\} \\ =& E\left\{\left|\mathbf{v}_k^H \hat{\mathbf{h}}_i\right|^2\right\} + E\left\{\mathbf{v}_k^H \tilde{\mathbf{h}}_i \tilde{\mathbf{h}}_i^H \mathbf{v}_k\right\} \\ =& E\left\{\left|\mathbf{v}_k^H \hat{\mathbf{h}}_i\right|^2\right\} + E\left\{\mathbf{v}_k^H E\left\{\tilde{\mathbf{h}}_i \tilde{\mathbf{h}}_i^H\right\} \mathbf{v}_k\right\} \\ =& E\left\{\left|\mathbf{v}_k^H \hat{\mathbf{h}}_i\right|^2\right\} + E\{\mathbf{v}_k^H \mathbf{C}_i \mathbf{v}_k\} \\ =& E\left\{\left|\mathbf{v}_k^H \hat{\mathbf{h}}_i\right|^2\right\} + E\left\{\sum_{l=1}^{L} \mathbf{v}_{kl}^H \mathbf{C}_{il} \mathbf{v}_{kl}\right\} \\ =& E\left\{\left|\mathbf{v}_k^H \hat{\mathbf{h}}_i\right|^2\right\} + \sum_{l=1}^{L} E\{\mathbf{v}_{kl}^H \mathbf{C}_{il} \mathbf{v}_{kl}\} \end{aligned} \tag{75}$$

where in (74) and (75) independence between $\mathbf{v}_k$ and the zero-mean channel estimate $\tilde{\mathbf{h}}_k$ is used. Similarly, it is possible to rewrite the last expectation as

$$\begin{aligned} E\left\{\left|\mathbf{v}_k^H \mathbf{n}^{\mathrm{UL}}\right|^2\right\} &= E\left\{\left|\mathbf{v}_k^H \mathbf{n}^{\mathrm{UL}} \mathbf{n}^{\mathrm{UL},H} \mathbf{v}_k\right|^2\right\} \\ &= E\left\{\mathbf{v}_k^H E\left\{\mathbf{n}^{\mathrm{UL}} \mathbf{n}^{\mathrm{UL},H}\right\} \mathbf{v}_k\right\} \\ &= E\left\{\mathbf{v}_k^H \mathbf{R}_{\mathbf{nn}}^{\mathrm{UL}} \mathbf{v}_k\right\} \\ &= E\left\{\sum_{l=1}^{L} \mathbf{v}_{kl}^H \mathbf{R}_{\mathbf{n}_l \mathbf{n}_l}^{\mathrm{UL}} \mathbf{v}_{kl}\right\} \\ &= \sum_{l=1}^{L} E\left\{\mathbf{v}_{kl}^H \mathbf{R}_{\mathbf{n}_l \mathbf{n}_l}^{\mathrm{UL}} \mathbf{v}_{kl}\right\} \end{aligned} \tag{76}$$

where this time independence between $\mathbf{v}_k$ and the additive uplink noise vector $\mathbf{n}^{\mathrm{UL}}$ is used. □

**Acknowledgements**
Not applicable.

**Author contributions**
RVR and MM conceived of the presented idea. RVR developed the theoretical formalism, performed the analytic calculations and performed the numerical simulations. All authors discussed the results and contributed to the final manuscript.

**Funding**
The work of R. Van Rompaey was supported by a PhD Fellowship of the Research Foundation Flanders (FWO-Vlaanderen No. 40062). This research work was carried out at the ESAT Laboratory of KU Leuven, in the frame of Research Council KU Leuven Project C3-19-00221 "Cooperative Signal Processing Solutions for IoT-based Multi-User Speech Communication Systems", Research Council KU Leuven Project C24/16/019 "Distributed Digital Signal Processing for Ad-hoc Wireless Local Area Audio Networking" and Fonds de la Recherche Scientifique - FNRS and the Fonds Wetenschappelijk Onderzoek - Vlaanderen under EOS Project no 30452698 '(MUSE-WINET) MUlti-SErvice WIreless NETwork'. The scientific responsibility is assumed by its authors.

**Availability of data and materials**
Not applicable.

## Declarations

**Competing interests**
The authors declare that they have no competing interests.

**Consent for publication**
Text for this section.

Received: 24 February 2022 Accepted: 17 May 2023
Published online: 11 June 2023

**References**

1. E. Björnson, L. Sanguinetti, H. Wymeersch, J. Hoydis, T.L. Marzetta, Massive MIMO is a reality-What is next?: Five promising research directions for antenna arrays. Digit. Signal Process. **94**, 3–20 (2019)
2. D. Gesbert, S. Hanly, H. Huang, S. Shamai Shitz, O. Simeone, W. Yu, Multi-cell MIMO cooperative networks: a new look at interference. IEEE J. Sel. Areas Commun. **28**(9), 1380–1408 (2010)
3. T.L. Marzetta, Noncooperative cellular wireless with unlimited numbers of base station antennas. IEEE Trans. Wirel. Commun. **9**(11), 3590–3600 (2010)
4. H.Q. Ngo, A. Ashikhmin, H. Yang, E.G. Larsson, T.L. Marzetta, Cell-free massive MIMO versus small cells. IEEE Trans. Wirel. Commun. **16**(3), 1834–1850 (2017)
5. E. Nayebi, A. Ashikhmin, T.L. Marzetta, H. Yang, B.D. Rao, Precoding and power optimization in cell-free massive MIMO systems. IEEE Trans. Wirel. Commun. **16**(7), 4445–4459 (2017)
6. E. Björnson, L. Sanguinetti, Making cell-free massive MIMO competitive with MMSE processing and centralized implementation. IEEE Trans. Wirel. Commun. **19**(1), 77–90 (2020)
7. S. Shamai, B.M. Zaidel, Enhancing the cellular downlink capacity via co-processing at the transmitting end. IEEE Veh. Technol. Conf. **3**(53ND), 1745–1749 (2001)
8. S. Zhou, M. Zhao, X. Xu, J. Wang, Y. Yao, Distributed wireless communication system: a new architecture for future public wireless access. IEEE Commun. Mag. **41**(3), 108–113 (2003)
9. M. Boldi, A. Tölli, M. Olsson, E. Hardouin, T. Svensson, F. Boccardi, L. Thiele, V. Jungnickel, Coordinated multipoint (comp) systems, in *Mobile and Wireless Communications for IMT-Advanced and Beyond* (Wiley, 2011), pp. 121–155
10. E. Björnson, R. Zakhour, D. Gesbert, B. Ottersten, Cooperative multicell precoding: rate region characterization and distributed strategies with instantaneous and statistical CSI. IEEE Trans. Signal Process. **58**(8), 4298–4310 (2010)
11. S. Chen, J. Zhang, E. Bjornson, J. Zhang, B. Ai, Structured massive access for scalable cell-free massive MIMO systems. IEEE J. Sel. Areas Commun. **39**(4), 1086–1100 (2021)
12. G. Interdonato, P. Frenger, E.G. Larsson, Scalability aspects of cell-free massive MIMO, in *IEEE International Conference on Communications* (2019), pp. 1–5
13. S. Chen, J. Zhang, J. Zhang, E. Björnson, B. Ai, A survey on user-centric cell-free massive MIMO systems. Digit. Commun. Netw. **8**, 695–719 (2021)
14. M. Bashar, K. Cumanan, A.G. Burr, M. Debbah, H.Q. Ngo, On the uplink max–min SINR of cell-free massive MIMO systems. IEEE Trans. Wirel. Commun. **18**(4), 2021–2036 (2019)
15. E. Nayebi, A. Ashikhmin, T.L. Marzetta, B.D. Rao, Performance of cell-free massive MIMO systems with MMSE and LSFD receivers, in *2016 59th Asilomar Conference on Signals, Systems and Computers* (IEEE, 2017), pp. 203–207
16. G. Interdonato, M. Karlsson, E. Björnson, E.G. Larsson, Local partial zero-forcing precoding for cell-free massive MIMO. IEEE Trans. Wirel. Commun. **19**(7), 4758–4774 (2020)
17. E. Björnson, L. Sanguinetti, scalable cell-free massive MIMO systems. IEEE Trans. Commun. **68**(7), 4247–4261 (2020). http://arxiv.org/abs/1908.03119
18. D. Tse, P. Viswanath, *Fundamentals of Wireless Communication* (Cambridge University Press, New York, 2005)
19. C. D'Andrea, A. Zappone, S. Buzzi, M. Debbah, Uplink power control in cell-free massive MIMO via deep learning, in *2019 IEEE 8th International Workshop on Computational Advances in Multi-sensor Adaptive Processing (CAMSAP)* (2019), pp. 554–558
20. M. Bashar, A. Akbari, K. Cumanan, H.Q. Ngo, A.G. Burr, P. Xiao, M. Debbah, J. Kittler, Exploiting deep learning in limited-fronthaul cell-free massive MIMO uplink. IEEE J. Sel. Areas Commun. **38**(8), 1678–1697 (2020)
21. F. Liang, C. Shen, W. Yu, F. Wu, Towards optimal power control via ensembling deep neural networks. IEEE Trans. Commun. **68**(3), 1760–1776 (2020)
22. H.Q. Ngo, A. Ashikhmin, H. Yang, E.G. Larsson, T.L. Marzetta, Cell-free massive MIMO: uniformly great service for everyone, in *IEEE Workshop on Signal Processing Advances in Wireless Communications, SPAWC* (2015), pp. 201–205
23. Ö. Özdogan, E. Björnson, E.G. Larsson, Massive MIMO with spatially correlated Rician fading channels. IEEE Trans. Commun. **67**(5), 3234–3250 (2019)
24. E. Björnson, J. Hoydis, L. Sanguinetti, Massive MIMO networks: spectral, energy, and hardware efficiency. Found. Trends® Signal Process. **11**(3–4), 154–655 (2017)
25. H. Yin, L. Cottatellucci, D. Gesbert, R.R. Müller, G. He, Robust pilot decontamination based on joint angle and power domain discrimination. IEEE Trans. Signal Process. **64**(11), 2990–3003 (2016)
26. E. Björnson, J. Hoydis, L. Sanguinetti, Massive MIMO has unlimited capacity. IEEE Trans. Wirel. Commun. **17**(1), 574–590 (2018)
27. R. Van Rompaey, M. Moonen, Scalable and distributed MMSE algorithms for uplink receive combining in cell-free massive MIMO systems, in *ICASSP 2021—2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (2021), pp. 4445–4449
28. D. Neumann, M. Joham, W. Utschick, Covariance matrix estimation in massive MIMO. IEEE Signal Process. Lett. **25**(6), 863–867 (2018)

29. H. Xie, F. Gao, S. Jin, An overview of low-rank channel estimation for massive MIMO systems. IEEE Access **4**, 7313–7321 (2016)
30. E. Björnson, L. Sanguinetti, M. Debbah, Massive MIMO with imperfect channel covariance information, in *2016 50th Asilomar Conference on Signals, Systems and Computers* (2016), pp. 974–978
31. J. Zhang, S. Chen, Y. Lin, J. Zheng, B. Ai, L. Hanzo, Cell-free massive MIMO: a new next-generation paradigm. IEEE Access **7**, 99 878-99 888 (2019)
32. S. Buzzi, C. D'Andrea, Cell-free massive MIMO: user-centric approach. IEEE Wirel. Commun. Lett. **6**(6), 706–709 (2017)
33. J. Zhang, Y. Wei, E. Björnson, S. Member, Performance analysis and power control of cell-free massive MIMO systems with hardware impairments. IEEE Access **6**, 55302–55314 (2018)
34. C.B. Peel, B.M. Hochwald, A.L. Swindlehurst, A vector-perturbation technique for near-capacity multiantenna multi-user communication—Part I : channel inversion and regularization. IEEE Trans. Commun. **53**(1), 195–202 (2005)
35. E. Biglieri, J. Proakis, S. Shlomo, Fading channels : information-theoretic and communications aspects. IEEE Trans. Inf. Theory **44**(6), 2619–2692 (1998)
36. T.V. Chien, E. Björnson, E.G. Larsson, Joint pilot design and uplink power allocation in multi-cell massive MIMO systems. IEEE Trans. Wirel. Commun. **17**(3), 2000–2015 (2018)
37. T.C. Mai, H.Q. Ngo, M. Egan, T.Q. Duong, S. Member, Pilot power control for cell-free massive MIMO. IEEE Trans. Veh. Technol. **67**(11), 11 264-11 268 (2018)
38. J. Hoydis, S. Ten Brink, M. Debbah, Massive MIMO in the UL/DL of cellular networks: How many antennas do we need? IEEE J. Sel. Areas Commun. **31**(2), 160–171 (2013)
39. D.P. Bertsekas, W.W. Hager, O.L. Mangasarian, *Nonlinear Programming* (Athena Scientific, Belmont, 1998)
40. W. Yu, T. Lan, Transmitter optimization for the multi-antenna downlink with per-antenna power constraints. IEEE Trans. Signal Process. **55**(6 I), 2646–2660 (2007)
41. V. Le Nir, M. Moonen, M. Guenach, J. Verlinden, Broadcast channel optimal spectrum balancing (BC-OSB) with per-modem total power constraints for downstream DSL, in *European Signal Processing Conference* (2007), pp. 2189–2193
42. R.J. Plemmons, M-matrix characterizations. I-nonsingular M-matrices. Linear Algebra Appl. **18**(2), 175–188 (1977)
43. Z.Q. Luo, S. Zhang, Dynamic spectrum management: complexity and duality. IEEE J. Sel. Topics Signal Process. **2**(1), 57–73 (2008)
44. E. Björnson, E. Jorswieck, Optimal resource allocation in coordinated multi-cell systems. Found. Trends Inf. Theory **9**(2–3), 113–381 (2012)
45. H. Tuy, Monotonic optimization: problems and solution approaches. SIAM J. Optim. **11**(2), 464–494 (2000)
46. Q. Shi, M. Razaviyayn, Z.Q. Luo, C. He, An iteratively weighted MMSE approach to distributed sum-utility maximization for a MIMO interfering broadcast channel. IEEE Trans. Signal Process. **59**(9), 4331–4340 (2011)
47. S. Buzzi, C. D'Andrea, A. Zappone, User-centric 5G cellular networks: resource allocation and comparison with the cell-free massive MIMO approach. IEEE Trans. Wirel. Commun. **19**(2), 1250–1264 (2020)
48. S. Boyd, L. Vandenberghe, *Convex Optimization* (Cambridge University Press, Cambridge, 2004)
49. M. Bashar, K. Cumanan, A.G. Burr, M. Debbah, H.Q. Ngo, Enhanced max–min SINR for uplink cell-free massive MIMO systems, in *IEEE International Conference on Communications* (2018)
50. Y.W. Hong, C.W. Tan, L. Zheng, C.L. Hsieh, C.H. Lee, A unified framework for wireless max–min utility optimization with general monotonic constraints, in *IEEE INFOCOM 2014—IEEE Conference on Computer Communications* (IEEE, 2014), pp. 2076–2084
51. H. Chen, A. Campbell, B. Thomas, A. Tamir, Minimax flow tree problems. Networks **54**, 117–129 (2009)
52. G. Interdonato, E. Björnson, H.Q. Ngo, P. Frenger, E.G. Larsson, Ubiquitous cell-free massive MIMO communications. EURASIP J. Wirel. Commun. Netw. **197**, 2019 (2019)
53. O. Mart, E.D. Carvalho, J.Ø. Nielsen, *Towards Very Large Aperture Massive MIMO: A Measurement Based Study* (2014), 281–286
54. U. Madhow, D.R. Brown, S. Dasgupta, R. Mudumbai, Distributed massive MIMO: algorithms, architectures and concept systems, in *2014 Information Theory and Applications Workshop, ITA 2014—Conference Proceedings* (2014)


## Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Submit your manuscript to a SpringerOpen journal and benefit from:**

- ► Convenient online submission
- ► Rigorous peer review
- ► Open access: articles freely available online
- ► High visibility within the field
- ► Retaining the copyright to your article

**Submit your next manuscript at ► springeropen.com**