Subset Regularized Zero-Forcing Precoders for Cell-Free C-RANs

Mohammad M. Mojahedian
Univ. Pompeu Fabra
Barcelona, Spain
mohammadmahdi.mojahedian@upf.edu

Angel Lozano
Univ. Pompeu Fabra
Barcelona, Spain
angel.lozano@upf.edu

Abstract—This paper formulates regularized zero-forcing (RZF) precoders that are both network- and user-centric for the downlink of centralized radio-access networks operating in a cell-free fashion. The transmission to every user involves a distinct subset of access points (APs), and every AP participates in the transmission to a distinct subset of users, hence the moniker *subset precoders*. These subsets, defined on the basis of the large-scale channel gains between users and APs, capture the most relevant signal and interference contributions while disregarding those whose processing is cost-ineffective and whose associated channel estimations would incur unnecessary overheads. With that, subset precoding approaches the performance of network-wide RZF, for the power allocation strategies of interest, while being scalable.

I. INTRODUCTION

The progressive move to ultradense software-defined wireless architectures drives the interest in centralized radio access networks (C-RANs) [1]. Such C-RANs feature access points (APs) that consist solely of antennas and RF stages, and that are fronthauled to an edge datacenter where the baseband processing is hosted. Also, C-RANs are naturally cell-free, meaning that every AP potentially transmits to every user, stretching the principles of cell cooperation [2]. Focusing on the downlink, much of the cell-free literature postulates conjugate-beamforming precoders [3]–[7], yet the potential of C-RANs is fully realized only with more sophisticated precoders that exploit the centralized nature of such networks. In particular, a regularized zero-forcing (RZF) precoder is decidedly superior because, rather than neglected, interference is then a central consideration [4], [8]–[10]. Moreover, while channel pseudo-inversion is not strictly the best possible form of ZF because of the separate power constraints at each AP, it is pleasingly close [11] and regularization can only strengthen its already satisfying behavior.

Altogether, a network-wide RZF precoder based on pseudo-inversion is a powerful solution, but it is not scalable and, in large deployments with thousands of APs and users, it is outright unfeasible. Furthermore, a network-wide precoder is an unnecessary overkill because (*i*) an AP's transmission elicits negligible signal levels at faraway users and it causes negligible interference to the transmissions from faraway APs, and (*ii*) reliable channel estimates for faraway users are so burdensome overhead-wise that it is preferable to forgo those estimations altogether. This suggests a degree of containment in the channel estimation and in the data transmission, with the goal of approaching the network-wide RZF performance in a scalable fashion.

Building on ideas as to how to curtail the APs and users whose signals should be jointly processed in cell-free networks [12]–[15], an approach to define receivers that are both AP- and user-centric was proposed for the C-RAN uplink [16], [17]; since they involve subsets of APs and users, these were termed *subset receivers*.

This paper tackles the formulation of subset RZF precoders for the C-RAN downlink, where additional challenges related to channel estimation and to power allocation compound the difficulty of the uplink problem.

II. NETWORK AND CHANNEL MODELS

A. Large-scale Modeling

We consider networks having N APs and K users, all featuring a single omnidirectional antenna. Distance-dependent pathloss with exponent η, combined with shadowing, gives rise to a large-scale gain $G_{n,k}$ between the kth user and the nth AP. The corresponding large-scale signal-to-noise ratio equals $\mathsf{SNR}_{n,k} = G_{n,k}P/\sigma^2$ with P the per-AP transmit power and σ^2 the noise power. While P/σ^2 is taken as equal for uplink and downlink, asymmetries could be readily incorporated by differentiating the respective SNR variables throughout.

B. Small-scale Modeling

Besides $G_{n,k}$, the channel between the kth user and the nth AP includes a small-scale fading coefficient $h_{n,k} \sim \mathcal{N}_{\mathbb{C}}(0,1)$, independent across APs and users.

For any snapshot of the large-scale parameters, with channel estimation errors and interference treated by the decoder as additional Gaussian noise, user k can attain a spectral efficiency of

$$C_k = \mathbb{E}\big[\log_2(1 + \mathsf{sinr}_k)\big], \tag{1}$$

where sinr_k denotes its signal-to-interference-plus-noise ratio and the expectation is over the distribution induced by the small-scale fading. Adding (1) over all K users, we obtain the sum spectral efficiency. The spectral efficiencies in this paper are *gross*, meaning that uplink and downlink pilot overheads are yet to be subtracted out.

C. Simulation Environment

Our wrapped-around simulation universe contains $N = 200$ APs elevated 2 m above the users to prevent distance singularities. With the premise of AP positions agnostic to the radio propagation, shadow fading renders the network approximately Poisson-like from the perspective of any user [18]. This effect sharpens as the shadowing grows strong, being precise for relevant values thereof [19]. Based on this phenomenon, the AP locations are drawn uniformly at random and, likewise, the user positions are uniformly random.

We set P/σ^2 such that $\mathsf{SNR}_{n,k} = 20$ dB at a distance d, where d would be the inter-AP spacing if the network were a hexagonal grid of the same spatial density. Under reasonable values for P, the bandwidth, and the pathloss intercept, this is compatible with ultradense deployments ($d \approx 10$–20 m).

III. SCALABILITY

Scalability is considered in terms of those aspects that are inherent to cell-free operation, namely (*i*) precoder obtainment cost, (*ii*) precoder application cost, and (*iii*) channel estimation. The encoding and remaining pre-processing tasks are as in a cellular network, one chain per user, inherently scalable.

We measure the computational cost in number of complex multiply-and-accumulate (MA) operations per time-frequency coherence block. For fixed K/N, we want the precoder obtainment and application costs, as well as the number of channel coefficients to estimate, to be $\mathcal{O}(N)$.

IV. PRECODING, CHANNEL ESTIMATION, AND SINRs

Let C be the channel matrix that combines large- and small-scale components, i.e., with (n,k)th entry given by

$$c_{n,k} = \sqrt{G_{n,k}} h_{n,k} \tag{2}$$

such that the column vector of observations at the K users is

$$\boldsymbol{y} = \boldsymbol{C}^* \sqrt{P} \boldsymbol{T} \boldsymbol{s} + \boldsymbol{v}, \tag{3}$$

where $\boldsymbol{s}$ is a vector containing the unit-power symbols intended for the K users, $\boldsymbol{T}$ is the $N \times K$ precoding matrix, and $\boldsymbol{v} \sim \mathcal{N}_{\mathbb{C}}(\boldsymbol{0}, \sigma^2 \boldsymbol{I})$. For the power constraints to be satisfied at the APs, it must hold that

$$\left[\boldsymbol{T}\boldsymbol{T}^*\right]_{n,n} \leq 1 \qquad n = 1, \ldots, N. \tag{4}$$

A. Uplink Channel Estimation

Denote by N_{p} the number of symbols reserved for uplink pilot transmissions on every coherence block, with each user being allocated N_{p}/K of those pilots. Disregarding pilot contamination [5], the MMSE fading estimate $\hat{h}_{n,k}$ gathered by the network upon observation at the nth AP of the pilots emitted by user k satisfies [20]

$$\mathbb{E}\left[|\hat{h}_{n,k}|^2\right] = \frac{\frac{N_{\mathrm{p}}}{K}\mathsf{SNR}_{n,k}}{1 + \frac{N_{\mathrm{p}}}{K}\mathsf{SNR}_{n,k}} \tag{5}$$

while

$$\left(h_{n,k} - \hat{h}_{n,k}\right) \sim \mathcal{N}_{\mathbb{C}}\left(0, \frac{1}{1 + \frac{N_{\mathrm{p}}}{K}\mathsf{SNR}_{n,k}}\right) \tag{6}$$

is uncorrelated error. Based on the uplink channel estimate $\hat{C}$ with (n,k)th entry given by $\hat{c}_{n,k} = \sqrt{G_{n,k}}\hat{h}_{n,k}$, the precoder can be obtained.

B. Downlink Precoded Channel Estimation

Once the precoder is set, downlink pilots enable each user to estimate its own precoded channel, i.e., the effective channel undergone by its data symbols. From (3), the kth user observes

$$y_k = \sqrt{P}\, \boldsymbol{c}_k^* \boldsymbol{T} \boldsymbol{s} + v_k \tag{7}$$

$$= \sqrt{P}\, \boldsymbol{c}_k^* \boldsymbol{t}_k s_k + \sum_{\ell \neq k} \sqrt{P}\, \boldsymbol{c}_k^* \boldsymbol{t}_\ell s_\ell + v_k \tag{8}$$

$$= \sqrt{P}\, a_{k,k} s_k + \sum_{\ell \neq k} \sqrt{P}\, a_{k,\ell} s_\ell + v_k, \tag{9}$$

where $\boldsymbol{c}_k$ and $\boldsymbol{t}_k$ are the kth columns of $\boldsymbol{C}$ and $\boldsymbol{T}$, respectively, while $a_{k,k} = \boldsymbol{c}_k^* \boldsymbol{t}_k$ and $a_{k,\ell} = \boldsymbol{c}_k^* \boldsymbol{t}_\ell$. User k is to estimate $a_{k,k}$.

Let $N_{\mathrm{d}} \geq K$ be the number of downlink pilot symbols and let $\boldsymbol{\phi}_1, \ldots, \boldsymbol{\phi}_K$ be orthonormal pilot sequences of length N_{d}. Upon the transmission of $\boldsymbol{\phi}_1^*, \ldots, \boldsymbol{\phi}_K^*$, the $1 \times N_{\mathrm{d}}$ observation at user k is

$$\boldsymbol{y}_k = \sqrt{N_{\mathrm{d}} P}\, \boldsymbol{c}_k^* \boldsymbol{T} \begin{bmatrix} \boldsymbol{\phi}_1^* \\ \vdots \\ \boldsymbol{\phi}_K^* \end{bmatrix} + \boldsymbol{v}_k. \tag{10}$$

Projecting $\boldsymbol{y}_k$ onto $\boldsymbol{\phi}_k$, user k obtains

$$\boldsymbol{y}_k \boldsymbol{\phi}_k = \sqrt{N_{\mathrm{d}} P}\, \boldsymbol{c}_k^* \boldsymbol{T} \begin{bmatrix} \boldsymbol{\phi}_1^* \\ \vdots \\ \boldsymbol{\phi}_K^* \end{bmatrix} \boldsymbol{\phi}_k + \boldsymbol{v}_k \boldsymbol{\phi}_k \tag{11}$$

$$= \sqrt{N_{\mathrm{d}} P}\, a_{k,k} + \boldsymbol{v}_k \boldsymbol{\phi}_k, \tag{12}$$

and, from that, the least-squares estimate [21]

$$\hat{a}_{k,k} = \frac{1}{\sqrt{N_{\mathrm{d}} P}}\, \boldsymbol{y}_k \boldsymbol{\phi}_k, \tag{13}$$

with an estimation error of variance

$$\mathbb{E}\left[|a_{k,k} - \hat{a}_{k,k}|^2\right] = \mathbb{E}\left[\left|\frac{1}{\sqrt{N_{\mathrm{d}} P}}\, \boldsymbol{v}_k \boldsymbol{\phi}_k\right|^2\right] \tag{14}$$

$$= \frac{\sigma^2}{N_{\mathrm{d}} P}. \tag{15}$$

C. User SINRs

With $\hat{a}_{k,k}$ at hand, the observation at user k upon subsequent downlink data transmission can be written as

$$y_k = \sqrt{P}\, \hat{a}_{k,k} s_k + \sqrt{P}\, (a_{k,k} - \hat{a}_{k,k}) s_k + \sum_{\ell \neq k} \sqrt{P}\, a_{k,\ell} s_\ell + v_k$$

giving at user k, conditioned on the known $\hat{a}_{k,k}$,

$$\mathsf{sinr}_k = \frac{|\hat{a}_{k,k}|^2}{\mathbb{E}\left[|a_{k,k} - \hat{a}_{k,k}|^2\right] + \sum_{\ell \neq k} \mathbb{E}\left[|a_{k,\ell}|^2\right] + \frac{\sigma^2}{P}} \tag{16}$$

$$= \frac{|\hat{a}_{k,k}|^2}{\sum_{\ell \neq k} \mathbb{E}\left[|a_{k,\ell}|^2\right] + \frac{\sigma^2}{P}\frac{N_{\mathrm{d}}+1}{N_{\mathrm{d}}}}. \tag{17}$$

This SINR expression, which holds regardless of the precoder, specializes to RZF when the precoder is set correspondingly.

V. BENCHMARK: NETWORK-WIDE RZF

From $\hat{C}$, the network-wide RZF precoder obtained by regularizing the channel's pseudo-inverse is

$$\boldsymbol{T} = \hat{C}(\hat{C}^*\hat{C} + \varrho\boldsymbol{I})^{-1}\mathrm{diag}(\sqrt{p_1}, \ldots, \sqrt{p_K}) \qquad (18)$$

$$= (\hat{C}\hat{C}^* + \varrho\boldsymbol{I})^{-1}\hat{C}\,\mathrm{diag}(\sqrt{p_1}, \ldots, \sqrt{p_K}) \qquad (19)$$

where $p_1, \ldots, p_K$ determine the power allocation while ϱ is the regularizing term. The optimization of ϱ is generally NP-hard [22], out of the scope of this paper. We let $\varrho = \frac{K}{N}\frac{\sigma^2}{P}$, a value that is known to be robust [23, sec. 9.9] and that is perfectly appropriate in the course of our pursuit of scalability.

From (19), the precoding vector for user k is

$$\boldsymbol{t}_k = \sqrt{p_k}\,(\hat{C}\hat{C}^* + \varrho\boldsymbol{I})^{-1}\hat{c}_k \qquad (20)$$

and, defining $\tilde{c}_k = c_k - \hat{c}_k$, we have that user k observes

$$y_k = \sqrt{P}\,(\hat{c}_k^* + \tilde{c}_k^*)\,\boldsymbol{T}s + v_k \qquad (21)$$

$$= \underbrace{\sqrt{P}\hat{c}_k^*\boldsymbol{t}_k s_k}_{\text{Desired Signal}} + \underbrace{\sqrt{P}\sum_{\ell \neq k}\hat{c}_k^*\boldsymbol{t}_\ell s_\ell + \sqrt{P}\,\tilde{c}_k^*\boldsymbol{T}s}_{\text{Interference}} + v_k \qquad (22)$$

and, from the network's perspective, i.e., conditioned on $\boldsymbol{T}$,

$$\mathsf{sinr}_k^{\text{RZF}} = \frac{\left|\hat{c}_k^*\boldsymbol{t}_k\right|^2}{\sum\limits_{\ell \neq k}\left|\hat{c}_k^*\boldsymbol{t}_\ell\right|^2 + \mathbb{E}\left[\tilde{c}_k^*\boldsymbol{T}\boldsymbol{T}^*\tilde{c}_k \,|\, \boldsymbol{T}\right] + \frac{\sigma^2}{P}} \qquad (23)$$

where

$$\mathbb{E}\left[\tilde{c}_k^*\boldsymbol{T}\boldsymbol{T}^*\tilde{c}_k \,|\, \boldsymbol{T}\right] = \qquad (24)$$
$$\mathrm{Tr}\left(\boldsymbol{T}\boldsymbol{T}^*\,\mathrm{diag}\left(\frac{G_{k,1}}{1 + \frac{N_\mathrm{p}}{K}\mathsf{SNR}_{k,1}}, \ldots, \frac{G_{k,N}}{1 + \frac{N_\mathrm{p}}{K}\mathsf{SNR}_{k,N}}\right)\right).$$

It is important to distinguish between the following.

- sinr_k in (17), the actual SINR experienced by user k and the one that determines its performance. This SINR is affected by uplink and downlink channel estimation errors: the former nudge the precoder away from true RZF, reducing the desired-signal coefficient $|a_{k,k}|$ and increasing the cross-interference coefficients $|a_{k,\ell}|$ for $\ell \neq k$; the latter reduce $|\hat{a}_{k,k}|$ while increasing $|a_{kk} - \hat{a}_{kk}|$.
- $\mathsf{sinr}_k^{\text{RZF}}$ in (23), which is the network's perception of the SINR at user k, based on which the precoder is computed. It is affected only by uplink channel estimation errors.

From (23), given any objective function $f(\cdot)$, the optimum $p_1, \ldots, p_K$ can be obtained as

$$\max_{p_1, \ldots, p_K} f\left(\mathsf{sinr}_1^{\text{RZF}}, \ldots, \mathsf{sinr}_K^{\text{RZF}}\right) \qquad (25)$$

$$\text{s.t. } \left[\hat{C}(\hat{C}^*\hat{C} + \varrho\boldsymbol{I})^{-1}\mathrm{diag}(\{p_k\})(\hat{C}^*\hat{C} + \varrho\boldsymbol{I})^{-1}\hat{C}^*\right]_{n,n} \leq 1.$$

The two extreme objective functions in terms of fairness are:

- Maximum fairness across users, disregarding the aggregate performance. This corresponds to

$$f\left(\mathsf{sinr}_1^{\text{RZF}}, \ldots, \mathsf{sinr}_K^{\text{RZF}}\right) = \min_k \mathsf{sinr}_k^{\text{RZF}}. \qquad (26)$$

- Maximum sum performance, disregarding fairness across users. Recalling (1), this is well represented by the sum spectral efficiency

$$f\left(\mathsf{sinr}_1^{\text{RZF}}, \ldots, \mathsf{sinr}_K^{\text{RZF}}\right) = \sum_{k=1}^{K}\log_2\left(1 + \mathsf{sinr}_k^{\text{RZF}}\right). \qquad (27)$$

As the cost of the matrix inversion in (18) is $\mathcal{O}(K^3)$, the obtention of the network-wide RZF precoder is decidedly not scalable irrespective of the power allocation. Likewise, the cost of applying $\boldsymbol{T}$ to s is KN, also not scalable, and the channel estimates including uplink and downlink are at least $KN + K$, again not scalable.

VI. SUBSET ZF PRECODING

Although, because of pathloss and shadowing, C has most of its mass concentrated on a small share of its entries, a network-wide ZF precoder processes every such entry, hence the unscalability. A workaround could be to zero-out all but the dominant entries of C, obtaining a sparse matrix C whose estimate would then be plugged into the various expressions in lieu of C itself. However, such sparse matrix might be unbalanced, with some APs and users heavily favored by many connections while others are outright disconnected from the network. Thus, the goal is to generate a sparse channel matrix that is balanced across rows and columns.

A. Formulation

Let us restrict to a subset $\mathcal{K}_n$ the users whose channels are estimated by the nth AP. Then, rather than $\hat{C}$, the channel matrix estimated by the network is $\hat{\mathsf{C}}$, with

$$[\hat{\mathsf{C}}]_{n,k} = \begin{cases} \sqrt{G_{n,k}}\,\hat{h}_{n,k} & k \in \mathcal{K}_n \\ 0 & \text{otherwise.} \end{cases} \qquad (28)$$

Next, let us curtail to a subset $\mathcal{N}_k$ the APs involved in transmitting to user k, identifying the submatrix $\hat{\mathsf{C}}_k$ obtained by selecting those rows of $\hat{\mathsf{C}}$ that correspond to APs in $\mathcal{N}_k$. From $\hat{\mathsf{C}}_k$, we compute the $|\mathcal{N}_k| \times 1$ vector $(\hat{\mathsf{C}}_k\hat{\mathsf{C}}_k^* + \varrho\boldsymbol{I})^{-1}\hat{\mathsf{c}}_k$ and, inserting zeros appropriately, we finally obtain the $N \times 1$ vector f_k to apply in lieu of $(\hat{C}\hat{C}^* + \varrho\boldsymbol{I})^{-1}\hat{c}_k$. Precisely,

$$[\mathsf{f}_k]_n = \begin{cases} \left[(\hat{\mathsf{C}}_k\hat{\mathsf{C}}_k^* + \varrho\boldsymbol{I})^{-1}\hat{\mathsf{c}}_k\right]_{\mathbb{I}_k(n)} & n \in \mathcal{N}_k \\ 0 & \text{otherwise} \end{cases} \qquad (29)$$

where $\mathbb{I}_k(n) \in \{1, \ldots, |\mathcal{N}_k|\}$ is the index of n within $\mathcal{N}_k$.

Taking all users into account, $(\hat{C}\hat{C}^* + \varrho\boldsymbol{I})^{-1}\hat{C}$ is replaced by $\mathsf{F} = [\mathsf{f}_1, \ldots, \mathsf{f}_K]$ in the computation of $p_1, \ldots, p_K$ as well as in the precoder itself, which now becomes

$$\mathsf{T} = \mathsf{F}\,\mathrm{diag}(\sqrt{p_1}, \ldots, \sqrt{p_K}). \qquad (30)$$

Applying T to the entire network, the users receive, in place of $\boldsymbol{y}$,

$$\mathsf{y} = \sqrt{P}C^*\mathsf{T}s + v. \qquad (31)$$

The precoder T is sparse as per the subsets $\mathcal{N}_1, \ldots, \mathcal{N}_K$ and $\mathcal{K}_1, \ldots, \mathcal{K}_N$. And, provided these subsets are defined properly,

the performance on the basis of **y** is close to the one made possible by y. Precisely, user k now observes

$$y_k = \sqrt{P}\left(\hat{\boldsymbol{c}}_k^* + \tilde{\boldsymbol{c}}_k^*\right)\mathbf{T}\boldsymbol{s} + v_k \tag{32}$$

$$= \sqrt{P}\,\hat{\boldsymbol{c}}_k^*\mathbf{t}_k s_k + \sqrt{P}\tilde{\boldsymbol{c}}_k^*\mathbf{T}\boldsymbol{s} + \sqrt{P}\sum_{\ell \neq k}\hat{\boldsymbol{c}}_k^*\mathbf{t}_\ell s_\ell + v_k \tag{33}$$

such that (23) now becomes

$$\mathrm{sinr}_k^{\mathrm{ZF}} = \frac{\left|\hat{\boldsymbol{c}}_k^*\mathbf{t}_k\right|^2}{\sum_{\ell \neq k}\left|\hat{\boldsymbol{c}}_k^*\mathbf{t}_\ell\right|^2 + \mathbb{E}\left[\tilde{\boldsymbol{c}}_k^*\mathbf{T}\mathbf{T}^*\tilde{\boldsymbol{c}}_k \mid \mathbf{T}\right] + \frac{\sigma^2}{P}} \tag{34}$$

which accounts for the additional interference from users belonging to other subsets. Likewise, (17), holds simply with **T** in place of $\boldsymbol{T}$ within $a_{k,k}$ and $a_{k,\ell}$.

B. Interpretation

From the sparsified channel matrix $\hat{\mathbf{C}}$, decimated row-wise as per $\mathcal{N}_k$, we obtain the precoding vector for user k. Then:

- If any other user has its channels estimated by all the APs in $\mathcal{N}_k$, transmissions to that user are projected away—in the RZF sense, toned by the regularization—from user k.
- Users whose channels are estimated by some (but not all) the APs in $\mathcal{N}_k$ are partially projected away from user k.
- Users whose channels are not estimated by any AP within $\mathcal{N}_k$ are interfered by transmissions to user k; no precoding effort is made to shield those users.

The interference management with the proposed subset structure exhibits this desirable feature: the stronger an interference situation, the more resources devoted to its avoidance.

This is exemplified in Fig. 1 for a toy network. The transmissions to users 1 and 2 project away from each other fully, project away partially from the transmission to user 3, and are oblivious to the transmission to user 4. The transmission to user 3, in turn, projects partially away from those to users 1 and 2 while ignoring the transmission to user 4.

C. Scalability

Contingent on the subsets $\mathcal{K}_1, \ldots, \mathcal{K}_N$ and $\mathcal{N}_1, \ldots, \mathcal{N}_K$ not growing with K and N, the cost of obtaining **F** is

$$\mathcal{O}\left(\sum_{k=1}^{K}|\mathcal{N}_k|^3\right) = \mathcal{O}(K) = \mathcal{O}(N), \tag{35}$$

whereby the scalability of the precoder obtention boils down to that of the power allocation. If the optimization in (25) is itself scalable, so is the precoder obtention, and vice versa.

As of the application of the precoder to the signals, its cost is $\sum_{n=1}^{N}|\mathcal{K}_n| = \mathcal{O}(N)$. In turn, the number of channel coefficients to estimate, combining uplink and downlink, is $\sum_{n=1}^{N}|\mathcal{K}_n| + K = \mathcal{O}(N)$.

D. Subset Selection Policy

We choose to determine $\mathcal{K}_n$ on the basis of large-scale gains, to avoid having to update the subsets at the small-scale fading rate. Precisely, we have $\mathcal{K}_n$ contain the $|\mathcal{K}_n|$ users with largest $G_{n,k}$ to the nth AP. In duality with that, $\mathcal{N}_k$ contains the $|\mathcal{N}_k|$ APs with the largest $G_{n,k}$ to user k. Moreover, the nth AP always estimates the channel coefficients corresponding to users in $\mathcal{K}_n$, to all of which it has to transmit.
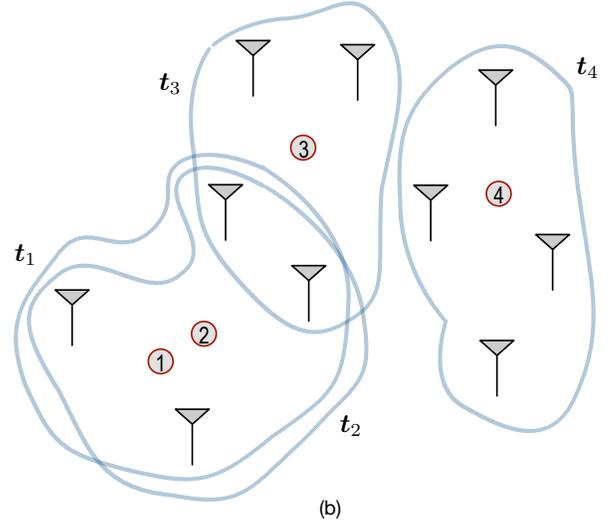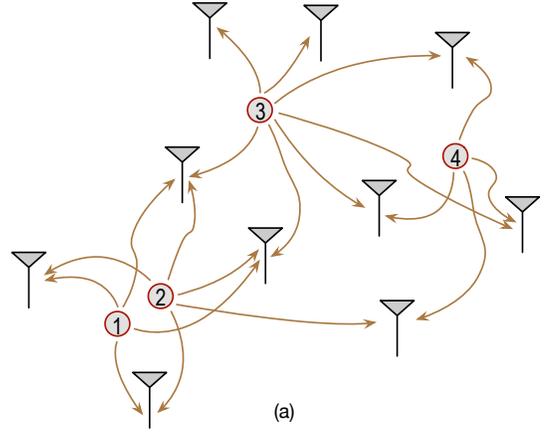

Fig. 1. Toy network having $N = 10$ APs and $K = 4$ users, with $\mathcal{N}_1 = \mathcal{N}_2 = \mathcal{N}_3 = \mathcal{N}_4 = 4$ and with $1 \leq \mathcal{K}_n \leq 3$. Every AP estimates the channels indicated in (a) and accounts for the corresponding users in its precoding. The APs that transmit signals to each of the users are delineated in (b).

E. Evaluation

For the broadest possible scope in terms of power allocation, we consider the two extreme objectives in (26) and (27). The procedures required by these objectives are not themselves scalable, but complete scalability is guaranteed for any power allocation policy between these extremes that is itself scalable.

The subset sizes $|\mathcal{K}|$ and $|\mathcal{N}|$ are made identical for all APs and users, with $|\mathcal{K}|/|\mathcal{N}| = K/N$. Shown in Fig. 2 is the CDF over the locations of APs and users of $\mathbb{E}[\mathrm{sinr}_k]$, with expectation over the small-scale fading, for maximum fairness power allocation; with reasonably small subsets, the performance approaches that of a network-wide RFZ precoder. The performance is even more robust to subsetting when the power allocation maximizes the sum spectral efficiency, as can be appreciated in Fig. 3.

Subset precoding preserves, to a large extent, the RZF's advantage over conjugate beamforming (whose performance is included in Fig. 3). The delivery of a completely scalable
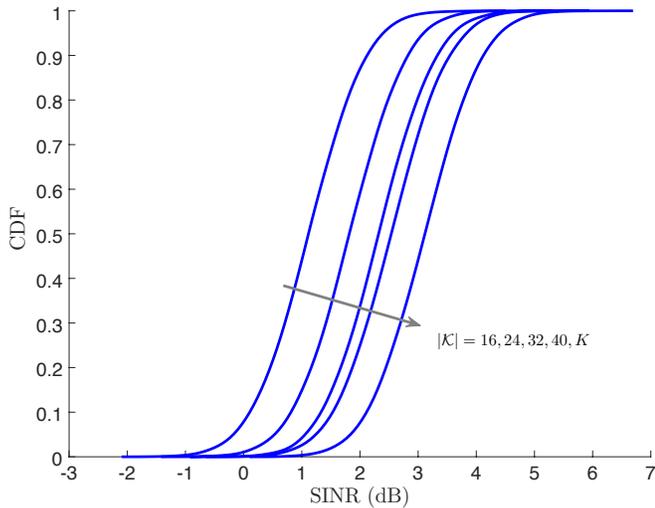
Fig. 2. CDF of $\mathbb{E}[\mathsf{sinr}_k]$ for $\eta = 4$, $K/N = 0.8$, $N_{\mathrm{p}} = N_{\mathrm{d}} = K$, and maximum fairness power allocation: network-wide RZF vs subset RZF with $|\mathcal{K}| = 16$, 24, 32, and 40.

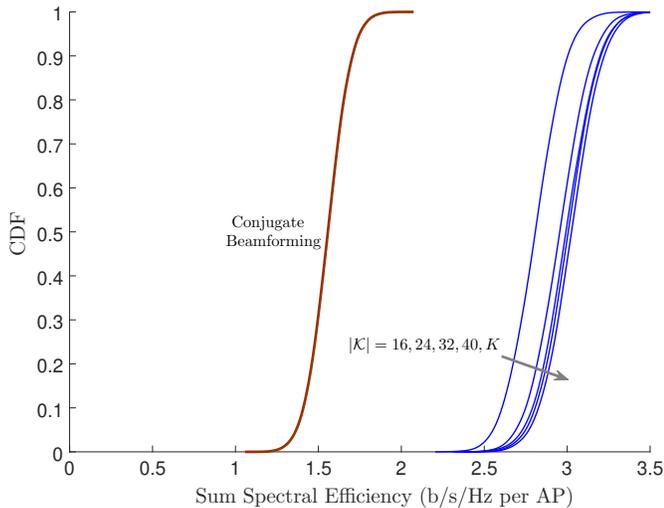Fig. 3. CDF of the sum spectral efficiency for $\eta = 4$, $K/N = 0.8$, $N_{\mathrm{p}} = N_{\mathrm{d}} = K$, and maximum sum spectral efficiency power allocation: network-wide RZF vs subset RZF with $|\mathcal{K}| = 16$, 24, 32, and 40. Also shown is the performance under conjugate beamforming.

precoding solution is then contingent on the power allocation, and identifying suitable scalable such policies is an appealing topic for subsequent research. Precisely, it would be of great interest to formulate RZF counterparts to the fractional power allocation policy derived for conjugate beamforming [24].

Acknowledgment

References

[1] A. Checko, H. L. Christiansen, Y. Yan, L. Scolari, G. Kardaras, M. S. Berger, and L. Dittmann, "Cloud RAN for mobile networks—A tech-nology overview," *IEEE Commun. Surveys & Tutorials*, vol. 17, no. 1, pp. 405–426, 2014.

[2] S. Venkatesan, A. Lozano, and R. Valenzuela, "Network MIMO: Over-coming intercell interference in indoor wireless systems," in *Asilomar Conf. on Signals, Systems and Computers*, 2007, pp. 83–87.

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

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

[5] M. Attarifar, A. Abbasfar, and A. Lozano, "Random vs structured pilot assignment in cell-free massive MIMO wireless networks," in *IEEE Int'l Conf. Commun. Workshops (ICCW'18)*, 2018, pp. 1–6.

[6] M. Attarifar, A. Abbasfar, and A. Lozano, "Modified conjugate beam-forming for cell-free massive MIMO," *IEEE Wireless Commun. Letters*, vol. 8, no. 2, pp. 616–619, 2019.

[7] G. Interdonato, H. Q. Ngo, P. Frenger, and E. G. Larsson, "Downlink training in cell-free massive MIMO: A blessing in disguise," *IEEE Trans. Wireless Commun.*, vol. 18, no. 11, pp. 5153–5169, 2019.

[8] F. Rezaei, C. Tellambura, A. A. Tadaion, and A. R. Heidarpour, "Rate analysis of cell-free massive MIMO-NOMA with three linear precoders," *IEEE Trans. Commun.*, vol. 68, no. 6, pp. 3480–3494, 2020.

[9] G. Interdonato, M. Karlsson, E. Björnson, and E. G. Larsson, "Local partial zero-forcing precoding for cell-free massive MIMO," *IEEE Trans. Wireless Commun.*, 2020.

[10] H. Nguyen, V.-D. Nguyen, O. Dobre, S. Sharma, S. Chatzinotas, B. Ottersten, and O.-S. Shin, "On the spectral and energy efficiencies of full-duplex cell-free massive MIMO," *IEEE J, Sel. Areas Commun.*, vol. 38, no. 8, pp. 1698–1718, 2020.

[11] M. Mojahedian, R. Mosayebi, and A. Lozano, "Pseudo-inverse vs generalized inverse for C-RAN downlink precoding," in *IEEE GLOBE-COM'20*, Dec. 2020.

[12] E. Bjornson, N. Jalden, M. Bengtsson, and B. Ottersten, "Optimality properties, distributed strategies, and measurement-based evaluation of coordinated multicell OFDMA transmission," *IEEE Trans. Signal Process.*, vol. 59, no. 12, pp. 6086–6101, 2011.

[13] J. Wang and L. Dai, "Downlink rate analysis for virtual-cell based large-scale distributed antenna systems," *IEEE Trans. Wireless Commun.*, vol. 15, no. 3, pp. 1998–2011, 2015.

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

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

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

[17] R. Mosayebi, M. Mojahedian, and A. Lozano, "Linear interference cancellation for the cell-free C-RAN uplink," *IEEE Trans. Wireless Commun.*, vol. 20, 2021.

[18] B. Błaszczyszyn, M. K. Karray, and H. P. Keeler, "Wireless networks appear Poissonian due to strong shadowing," *IEEE Trans. on Wireless Communications*, vol. 14, no. 8, pp. 4379–4390, Aug. 2015.

[19] G. George, R. K. Mungara, A. Lozano, and M. Haenggi, "Ergodic spectral efficiency in MIMO cellular networks," *IEEE Trans. on Wireless Communications*, vol. 16, no. 5, pp. 2835–2849, 2017.

[20] A. Lozano, "Interplay of spectral efficiency, power and Doppler spec-trum for reference-signal-assisted wireless communication," *IEEE Trans. on Wireless Communications*, vol. 7, no. 12, pp. 5020–5029, Dec. 2008.

[21] A. Zhou, J. Wu, E. Larsson, and P. Fan, "Max-min optimal beamforming for cell-free massive MIMO," *IEEE Commun. Letters*, vol. 24, no. 10, pp. 2344–2348, 2020.

[22] Y. Liu, Y. Dai, and Z. Luo, "Coordinated beamforming for MISO interference channel: Complexity analysis and efficient algorithms," *IEEE Trans. Signal Proc.*, vol. 59, no. 3, pp. 1142–1157, 2011.

[23] R. W. Heath Jr. and A. Lozano, *Foundations of MIMO communication*. Cambridge University Press, 2019.

[24] R. Nikbakht, R. Mosayebi, and A. Lozano, "Uplink fractional power control and downlink power allocation for cell-free networks," *IEEE Wireless Commun. Letters*, vol. 9, no. 6, pp. 774–777, 2020.