Decentralized User Scheduling and Beamforming in Multi-cell MIMO Networks

Zehua Li, Tyler Gamvrelis, Hussein A. Ammar and Raviraj Adve
Dept. of Electrical and Computer Engineering
University of Toronto
Toronto, Canada
{samzehuali.li, ty.gamvrelis}@mail.utoronto.ca, {ammarhus, rsadve}@ece.utoronto.ca

ICC 2022 - IEEE International Conference on Communications | 978-1-5386-8347-7/22/$31.00 ©2022 IEEE | DOI: 10.1109/ICC45855.2022.9839273

***Abstract*—We study the problem of distributed user scheduling and beamforming in multi-user, multi-cell, multiple-input multiple-output (MIMO) networks to maximize the weighted sum-rate. While previous work has focused on optimizing the signal-to-leakage-plus noise ratio (SLNR) or the signal-to-interference-plus-noise ratio (SINR), we propose a new signal-to-leakage-plus-interference-plus-noise ratio (SLINR) metric which hybridizes the SINR and SLNR by incorporating the intra-cell interference and inter-cell leakage. Using fractional programming and the Hungarian algorithm, we construct an iterative resource allocator that performs user scheduling and beamforming while accounting for channel estimation errors. Furthermore, we show different approaches for calculating the leakage which vary in terms of practicality and scalability. These approaches decrease the complexity compared to the standard method of leakage calculation, while providing comparable performance. Our results show that resource allocation based on the SLINR metric is a promising solution for decentralized implementation.**

I. Introduction

With the tremendous growth of the wireless connectivity market, service providers are finding ways to deliver higher data rates to denser populations. One of the main factors that limits data rates is interference from neighbouring transmitters. To reduce this undesired effect, a key technique in multiple-input multiple-output (MIMO) systems involves jointly optimizing the scheduling of users and the beams used to serve them [1]–[4]. For example, the authors in [4] propose a signal-to-interference-plus-noise ratio (SINR)-based resource allocation algorithm which uses fractional programming and the Hungarian algorithm to maximize the weighted sum-rate. Nevertheless, SINR-based optimization problems require base stations (BSs) to have global knowledge of the channel state information (CSI) over the entire network, and necessitates centralized optimization since the interference term is naturally coupled to the decisions made across all BSs.

To avoid the dependence on CSI, researchers have proposed alternative approaches based on different metrics, such as the signal-to-leakage-plus-noise ratio (SLNR) [5]–[10]. In these approaches, the power leaked to other users and cells is treated as a proxy for interference. The key benefit of this approach is that it allows for distributed implementation since the SLNR only depends on the local decisions of each BS.

The authors would like to acknowledge the support of Telus Canada and the Natural Sciences and Engineering Research Council of Canada.

However, optimizing SLNR has important drawbacks. For example, since the power of the beamformers scales equally in both the numerator and denominator of the SLNR, it is not effective at guiding power allocation. Furthermore, even local scheduling decisions under SLNR are difficult, since the decision of whether to schedule a user affects the leakage of all beams. In this paper, we develop a resource allocation approach based on the so-called signal-to-leakage-plus-interference-plus-noise ratio (SLINR) metric, which comprises *intra-cell* interference, *inter-cell* leakage and noise. The contributions of this paper are:

- We formulate a weighted *pseudo-rate* maximization problem based on the SLINR metric, which retains the benefits of the SLNR (via inter-cell leakage) and SINR (via intra-cell interference) metrics.
- We develop an iterative and decentralized user scheduling and beamforming approach using fractional programming and the Hungarian algorithm. Furthermore, this approach accounts for channel estimation errors.
- We propose and test three approaches to calculate the leakage, which vary in terms of computational complexity and demand for real-time information. All approaches produce comparable performance.

It is worth noting that the work in [11] used our proposed approach in the context of distributed networks.

This paper is organized as follows. Section II presents the system model, while Section III formulates the optimization problem. Section IV presents our proposed resource allocation algorithm and Section V proposes three approaches for leakage computation. Finally, Section VI reports our simulation results, and we conclude in Section VII.

II. System Model

A. Network Model

We consider the downlink of a time-division duplex (TDD) system comprising several BSs, represented by the set $\mathcal{B}$. Each BS b is equipped with M antennas and serves K_b single-antenna users with its cell. All BSs operate on the same frequency band. We define the channel between the b^{th} BS and the k^{th} user of the $(b')^{\text{th}}$ BS as $\mathbf{h}_{kb',b} \triangleq \mathbf{g}_{kb',b}\sqrt{\beta_{kb',b}} \in \mathbb{C}^{M\times 1}$. The term $\mathbf{g}_{kb',b} \sim \mathcal{CN}(\mathbf{0}, \mathbf{I}_M)$ denotes the Rayleigh fading channel component and $\beta_{kb',b} = (1 + d_{kb',b}/d_0)^{-\alpha}$ the large-scale path loss, where $d_{kb',b}$ is the distance between BS b and user k of BS b', d_0 is the path loss reference distance and α is the path loss exponent.

We define u_{kb} as the binary scheduling variable that represents whether user k of BS b is scheduled ($u_{kb} = 1$) or not ($u_{kb} = 0$). Furthermore, we define $\mathbf{v}_{kb} \in \mathbb{C}^{M\times 1}$ as the user's beamforming vector. The SINR and achievable rate are therefore given by, respectively,

$$\gamma_{kb} = \frac{u_{kb}|\mathbf{h}_{kb,b}^H \mathbf{v}_{kb}|^2}{\sigma_Z^2 + \sum_{b'=1}^{|\mathcal{B}|}\sum_{k'=1,(b',k')\neq(b,k)}^{K_{b'}} u_{k'b'}|\mathbf{h}_{kb,b'}^H \mathbf{v}_{k'b'}|^2}, \tag{1a}$$

$$R_{kb} = W\log(1+\gamma_{kb}). \tag{1b}$$

where σ_Z^2 is the power of the additive white Gaussian noise (AWGN), and W is the bandwidth.

Using (1b), the network-wide weighted sum-rate achieved can be written as

$$\sum_{b=1}^{|\mathcal{B}|}\sum_{k=1}^{K_b} w_{kb}R_{kb}, \tag{2}$$

where w_{kb} denotes the rate weighting factor, interpreted as the relative priority of each user within the network.

B. Channel Estimation

The first step in a communication cycle is estimating the channels using a pilot-based scheme, which requires assigning the users predefined pilot sequences. Since we are studying a cellular network, it is convenient to cluster the users (for pilot assignment) based on which BS they are served by. This is applicable when the length of the pilot sequence is more than the number of users in each cell.

We perform channel estimation through an uplink pilot training phase of length τ_p. During this phase of length τ_p, the signal received at the b^{th} BS can be represented as the matrix $\mathbf{Y}_b \in \mathbb{C}^{M\times\tau_p}$ and written as

$$\mathbf{Y}_b = \sum_{b'=1}^{|\mathcal{B}|}\sum_{k=1}^{K_{b'}} \sqrt{p}\mathbf{h}_{kb',b}\mathbf{\Phi}_{kb'} + \mathbf{Z}_b, \tag{3}$$

where BSs are receiving both intra-cell and inter-cell pilot signals from all users in the network. The term $\mathbf{\Phi}_{kb} \in \mathbb{C}^{1\times\tau_p}$ is the pilot sequence used by the k^{th} user of the b^{th} BS, p is the transmit power of the user, and $\mathbf{Z}_b$ is the AWGN with entries distributed according to $\mathcal{CN}\left(0,\sigma_Z^2\right)$. The power budget p of the user is defined per time instant, so the norm of the pilot sequence would be equal to the length of the training phase ($\mathbf{\Phi}_{kb}\mathbf{\Phi}_{kb}^H = \tau_p$).

Similar to [12], we assume each BS has knowledge of each user's transmit power and large-scale path loss. The b^{th} BS can extract the channel of the k^{th} user of the $(b')^{\text{th}}$ BS by first projecting the received signal (3) on the pilot sequence used by user k as $\mathbf{y}_{kb',b} = \frac{1}{\tau_p\sqrt{p}}\mathbf{Y}_b\mathbf{\Phi}_{kb'}^H$. This eliminates all the contributions of the users using pilot sequences orthogonal to pilot sequence $\mathbf{\Phi}_{kb'}$. Then, the BS uses the linear minimum mean square error (MMSE) estimator to calculate the estimate $\hat{\mathbf{h}}_{kb',b}$ of the channel $\mathbf{h}_{kb',b}$ as

$$\hat{\mathbf{h}}_{kb',b} = \mathbf{R}_{\mathbf{y}_{kb',b}\mathbf{h}_{kb',b}}\mathbf{R}_{\mathbf{y}_{kb',b}\mathbf{y}_{kb',b}}^{-1}\mathbf{y}_{kb',b}, \tag{4}$$

where $\mathbf{R}_{\mathbf{y}_{kb',b}\mathbf{y}_{kb',b}} \in \mathbb{C}^{M\times M}$ is the covariance matrix of $\mathbf{y}_{kb',b}$, and $\mathbf{R}_{\mathbf{y}_{kb',b}\mathbf{h}_{kb',b}} \in \mathbb{C}^{M\times M}$ is the cross-covariance matrix between $\mathbf{y}_{kb',b}$ and the parameter to be estimated, i.e., the uplink channel $\mathbf{h}_{kb',b}$.

Assuming spatially uncorrelated channels, the cross-covariance matrix $\mathbf{R}_{\mathbf{y}_{kb',b}\mathbf{h}_{kb',b}}$ can be written as a diagonal matrix $\mathbf{D}_{kb',b}$, whose non-zero entries are the large-scale path loss $[\mathbf{D}_{kb',b}]_{mm} \triangleq \beta_{kb',b}$. Similarly, the covariance matrix can be obtained as

$$\mathbf{R}_{\mathbf{y}_{kb',b}\mathbf{y}_{kb',b}} = \mathbf{C}_{kb',b} = \sum_{(k',b')\in\mathcal{U}_k}\mathbf{D}_{k'b',b} + \frac{\sigma_z^2}{\tau_p p}\mathbf{I}_M, \tag{5}$$

where the set $\mathcal{U}_k$ represents all users that are using the same pilot sequence as user k of BS b' (including this user). Hence,

$$\hat{\mathbf{h}}_{kb',b} = \mathbf{D}_{kb',b}{\mathbf{C}_{kb',b}}^{-1}\mathbf{y}_{kb',b}. \tag{6}$$

From the properties of MMSE estimation, the estimated channel $\hat{\mathbf{h}}_{kb',b}$ is distributed according to $\mathcal{CN}(\mathbf{0},\mathbf{A}_{kb',b})$, where the covariance matrix $\mathbf{A}_{kb',b}$ is given by [13]

$$\mathbf{A}_{kb',b} = \mathbf{D}_{kb',b}{\mathbf{C}_{kb',b}}^{-1}\mathbf{D}_{kb',b}. \tag{7}$$

Since channels are estimated using linear MMSE, the channel estimation error $\mathbf{e}_{kb',b} = \mathbf{h}_{kb',b} - \hat{\mathbf{h}}_{kb',b}$ is uncorrelated with $\hat{\mathbf{h}}_{kb',b}$, and can be modelled as $\mathbf{e}_{kb',b} \sim \mathcal{CN}(\mathbf{0},\mathbf{\Theta}_{kb',b})$, where $\mathbf{\Theta}_{kb',b} \triangleq \mathbf{D}_{kb',b} - \mathbf{A}_{kb',b}$.

III. Problem Formulation

A. Problem Definition

As discussed in Section I, SLNR-based optimization algorithms face issues such as ineffective power allocation and difficulty with scheduling. Below, we develop a weighted sum *pseudo*-rate optimization problem with a metric that incorporates intra-cell interference and inter-cell leakage. We introduce the hybrid leakage- and interference-based metric as the SLINR, given by

$$\zeta_{kb} = \frac{u_{kb}|\mathbf{h}_{kb,b}^H\mathbf{v}_{kb}|^2}{\sigma_Z^2 + L_{kb}^{\text{inter}}(\mathbf{v}_{kb}) + \sum_{k'=1\,(k'\neq k)}^{K_b} u_{k'b}|\mathbf{h}_{kb,b}^H\mathbf{v}_{k'b}|^2}. \tag{8}$$

The summation term in the denominator is the intra-cell interference term. If we have perfect knowledge of scheduling variables between different BSs through information exchange, then the leakage term in (8) can be written as

$$L_{kb}^{\text{inter}}(\mathbf{v}_{kb}) = \sum_{\substack{b'=1\\(b'\neq b)}}^{|\mathcal{B}|}\sum_{k'=1}^{K_{b'}} u_{k'b'}|\mathbf{h}_{k'b',b}^H\mathbf{v}_{kb}|^2. \tag{9}$$

The SLINR defined in (8) has a strong intuition behind it. For example, if BS b minimizes the leakage (9) while serving user k, it helps enhance the performance of users in the network who experience interference due to this leakage. A similar argument applies for the intra-cell interference term in (8). Hence, generally speaking, maximizing (8) can act as a proxy for maximizing the SINR. However, unlike the SINR, the formula in (8) decouples the beamforming decisions across BSs, which helps in implementing a decentralized resource allocation scheme.

Using (8), we construct a weighted sum pseudo-rate objective function and formulate the following resource allocation problem

$$\max_{\mathbf{U},\mathbf{V}} \quad \sum_{b=1}^{|\mathcal{B}|}\sum_{k=1}^{K_b} w_{kb}\log\left(1+\zeta_{kb}\right) \tag{10a}$$

$$\text{s.t.} \quad u_{kb}\in\{0,1\} \tag{10b}$$

$$\sum_{k=1}^{K_b} u_{kb}\le M, \qquad b=1,...|\mathcal{B}| \tag{10c}$$

$$\sum_{k=1}^{K_b} ||\mathbf{v}_{kb}||^2\le P_T, \qquad b=1,...|\mathcal{B}| \tag{10d}$$

where $b=1,...,|\mathcal{B}|,\ k=1,...,K_b$. The first set of constraints in (10b) enforces binary scheduling decisions. The second set of constraints in (10c) ensures that the number of users scheduled at a BS does not exceed the number of antennas M. Constraints (10d) impose a limit of P_T on the transmit power at each BS. The optimization is over the scheduling variables and beamformers, which are collected into matrices $\mathbf{U}$ and $\mathbf{V}$, respectively.

Finally, it is true that problem (10) is used as a proxy to enhance performance; ultimately, the system performance must be evaluated through real SINR-based rates using (2).

B. Problem Analysis

From (9), the only coupling between the BSs is through the scheduling variable $u_{k'b'}$. Instead of setting these variables to 1 or acquiring perfect knowledge of them through information exchange, we estimate the inter-cell leakage for beam $\mathbf{v}_{kb}$ as

$$L_{kb}^{\text{inter}}(\mathbf{v}_{kb})=\sum_{\substack{b'=1\\(b'\neq b)}}^{|\mathcal{B}|}\sum_{k'=1}^{K_{b'}}\Pr\left(u_{k'b'}=1\right)|\mathbf{h}_{k'b',b}^H\mathbf{v}_{kb}|^2, \tag{11}$$

where we adopt a flexible approach based on our knowledge of $\Pr\left(u_{k'b'}=1\right)$. Note that this probabilistic knowledge could vary with respect to time; however, we assume that it changes on a slow time scale, so that the impact of real-time knowledge on the inter-cell leakage calculation is small.

Under the condition of imperfect CSI and using the definition $\mathbf{h}_{kb',b}=\hat{\mathbf{h}}_{kb',b}+\mathbf{e}_{kb',b}$, we use the estimated channels instead of the actual ones (which cannot be known in practice) and we treat the estimation errors as extra noise. Therefore, the SLINR given by (8) becomes

$$\tilde{\zeta}_{kb}=\frac{u_{kb}|\hat{\mathbf{h}}_{kb,b}^H\mathbf{v}_{kb}|^2}{\hat{\text{den}}\{\zeta_{kb}\}+\tilde{L}_{kb}^{\text{inter}}(\mathbf{v}_{kb})+\tilde{Y}_{kb}^{\text{intra}}(\mathbf{\Theta}_{kb,b})}, \tag{12}$$

where $\hat{\text{den}}\{\zeta_{kb}\}$ is almost equivalent to the denominator of ζ_{kb}, with the exception that all perfect channels $\mathbf{h}_{kj,i}$ are replaced with estimated channels $\hat{\mathbf{h}}_{kj,i}$. If we adopt the leakage term in (11) and consider imperfect CSI in a way similar to [14, eq. (6)], the second term in the denominator of (12) is the estimation error for the inter-cell channels and it takes the form of

$$\tilde{L}_{kb}^{\text{inter}}(\mathbf{v}_{kb})=\sum_{\substack{b'=1\\(b'\neq b)}}^{|\mathcal{B}|}\sum_{k'=1}^{K_{b'}}\Pr\left(u_{k'b'}=1\right)\mathbf{v}_{kb}^H\mathbf{\Theta}_{k'b',b}\mathbf{v}_{kb}, \tag{13}$$

where $\mathbf{\Theta}_{kb',b}=\mathbb{E}\{\mathbf{e}_{kb',b}\mathbf{e}_{kb',b}^H\}$ is the covariance of the estimation error for the channel between the b^{th} BS and the k^{th} user of the $(b')^{\text{th}}$ BS. The third term in the denominator of (12) is

$$\tilde{Y}_{kb}^{\text{intra}}(\mathbf{\Theta}_{kb,b})=\sum_{k'=1}^{K_b}u_{k'b}\mathbf{v}_{k'b}^H\mathbf{\Theta}_{kb,b}\mathbf{v}_{k'b}, \tag{14}$$

which uses the estimation error of the intra-cell channels.

IV. Decentralized Resource Allocation

A. Solution Approach

We employ an iterative optimization approach based on fractional programming [15] (FP) to find a fixed point for the weighted sum pseudo-rate maximization problem (10). We will provide a high-level overview of the solution method, then list the resulting equations without presenting full derivations.

First, we introduce an auxiliary variable ϕ_{kb} to replace the fractional SLINR term in the logarithm (10a), and introduce a Lagrange multiplier for its equality constraint. Differentiating with respect to ϕ_{kb}, we obtain a revised objective function with the numerators of the fractional terms being non-negative and the denominators being strictly positive. Therefore, we can use the quadratic transform [15] to decouple the numerator and denominator of each fraction, resulting in an additional auxiliary variable y_{kb} being introduced [16]. The new optimization problem's objective function will have four variables to be optimized, denoted by u_{kb}, $\mathbf{v}_{kb}$, ϕ_{kb}, and y_{kb} (introduced in the FP step).

This problem can now be solved via cyclic maximization with closed-form updates. First, by keeping u_{kb}, $\mathbf{v}_{kb}$, and y_{kb} fixed and setting the derivative of the objective function with respect to ϕ_{kb} to zero, we obtain the optimal value of ϕ_{kb} as

$$\phi_{kb}=\tilde{\zeta}_{kb}. \tag{15}$$

Similarly, by keeping u_{kb}, $\mathbf{v}_{kb}$, and ϕ_{kb} fixed and setting the derivative of the objective function with respect to y_{kb} equal to zero, we obtain the optimal value of y_{kb} as

$$y_{kb}=\frac{\sqrt{w_{kb}u_{kb}(1+\phi_{kb})}\mathbf{v}_{kb}^H\hat{\mathbf{h}}_{kb,b}}{\text{den}\{\phi_{kb}\}+u_{kb}|\hat{\mathbf{h}}_{kb,b}^H\mathbf{v}_{kb}|^2}, \tag{16}$$

where $\text{den}\{\phi_{kb}\}$ represents the denominator of (12).

To find the optimal $\mathbf{v}_{kb}$, we first introduce a dual variable $\mu_{kb}\ge 0$ to capture the per-BS power constraint (10d) and form its Lagrangian formulation. Differentiating the Lagrangian with respect to $\mathbf{v}_{kb}^*$ and setting it to zero, we find

$$\mathbf{v}_{kb}^o=y_{kb}^*\sqrt{w_{kb}u_{kb}\left(1+\phi_{kb}\right)}\mathbf{X}_b^{-1}(\mu_b)\hat{\mathbf{h}}_{kb,b}, \tag{17}$$

where the intermediate matrix $\mathbf{X}_b^{-1}$ depends on μ_b as

$$\mathbf{X}_b(\mu_b)=\mu_b\mathbf{I}_{M_t}+\sum_{k'=1}^{K_b}u_{kb}|y_{k'b}|^2\left(\hat{\mathbf{h}}_{k'b,b}\hat{\mathbf{h}}_{k'b,b}^H+\mathbf{\Theta}_{k'b,b}\right)$$
$$+\sum_{\substack{b'=1\\(b'\neq b)}}^{|\mathcal{B}|}\sum_{k'=1}^{K_{b'}}|y_{kb}|^2\Pr\left(u_{k'b'}=1\right)\left(\hat{\mathbf{h}}_{k'b',b}\hat{\mathbf{h}}_{k'b',b}^H+\mathbf{\Theta}_{k'b',b}\right). \tag{18}$$

The dual variable μ_b can be chosen using a bisection search in order to satisfy the beamformer power constraint (10d).

The final step in the cyclic maximization is to update u_{kb}. Note that once the set of beams is fixed, the intra-cell interference pattern and inter-cell leakage are independent of local scheduling decisions. Thus, we can represent the SLINR of the k^{th} user of the b^{th} BS served on the n^{th} non-zero beam by substituting $\mathbf{v}_{kb}$ with $\boldsymbol{\nu}_{nb}$ where $n = 1, ..., N_b$ and $N_b \leq M$. Here, $\boldsymbol{\nu}_{nb}$ denotes the n^{th} non-zero beam at BS b. Furthermore, we can obtain the pseudo-rate as $R_{kb,n} = W \log(1 + \tilde{\zeta}_{kb})$.

Updating the local scheduling variables is then a matter of solving a linear sum assignment problem:

$$\max_{\chi} \quad \sum_{k=1}^{K_b} \sum_{n=1}^{N_b} R_{kb,n} \chi_{kb,n} \tag{19a}$$

$$\text{s.t.} \quad \chi_{kb,n} \in \{0, 1\} \tag{19b}$$

$$\sum_{k=1}^{K_b} \chi_{kb,n} = 1 \tag{19c}$$

$$\sum_{n=1}^{N_b} \chi_{kb,n} \leq 1, \tag{19d}$$

where $b = 1, ..., |\mathcal{B}|,\ k = 1, ..., K_b,\ n = 1, ..., N_b$ and the binary variables $\chi_{kb,n}$ indicate whether the k^{th} user of the b^{th} BS is served on the n^{th} beam, which determines u_{kb}. Using the Hungarian algorithm [17], we can solve this linear sum assignment problem with $O(N_b^2 K_b)$ time complexity.

Note that by setting $\hat{\mathbf{h}}_{kj,i} = \mathbf{h}_{kj,i}$ and $\boldsymbol{\Theta}_{kj,i} = \mathbf{0}_M$, where $\mathbf{0}_M$ is the $M \times M$ zero matrix, the formulas for the cyclic maximization updates reduce to the case of perfect CSI.

B. Optimization Algorithm

Algorithm 1: Resource Allocation Algorithm.

1: Initialize $u_{kb}, \mathbf{v}_{kb}$ for all users
2: Initialize $N_{\text{iterations}}$ and set $i = 1$
3: **repeat** for all users
4: Update ϕ_{kb} using (15)
5: Update y_{kb} using (16)
6: Update $\mathbf{v}_{kb}$ using (17)
7: Update u_{kb} and $\mathbf{v}_{kb}$ jointly by solving (19)
8: Increment i
9: **until** convergence or $i = N_{\text{iterations}}$

The downlink decentralized resource allocation algorithm is presented in Algorithm 1. This algorithm executes independently at each BS b and it optimizes the SLINR-based weighted sum-rate. The algorithm initializes scheduling decisions and beamformers (Step 1) and then for all users, $k = 1, ..., K_b$ of BS b, it updates the variables ϕ_{kb}, y_{kb}, $\mathbf{v}_{kb}$, and u_{kb} iteratively until convergence or the maximum number of iterations is reached.

V. Scalability of Leakage Computation

A. Standard CSI Estimation

The main issue in calculating the leakage expression in (9) is that each BS does not know the scheduling decisions of the other BSs, i.e., $u_{kb'}$ is not known by the b^{th} BS. Recall that this is why we adopted user-based discrete probabilistic knowledge in (11), to replace the knowledge of scheduling decisions in neighbouring cells. However, this method still requires BSs to estimate the channels of all users in the network. This is likely impractical and unscalable. One main performance limitation results from training sequences being reused across cells: from each BS's perspective, the training signals from users in neighbouring cells end up being contaminated by the training signals from the users it is serving. In an ideal setting, the BS would only estimate the channels to τ_p users closest to it, where τ_p is the number of orthogonal pilot signals.

B. Statistical CSI

Similar to discussions made in [11], this method of calculating the leakage is similar to the standard CSI Estimation method discussed in Section V-A, except that we only use the large-scale path loss, thus eliminating the need to estimate the leakage channels. Note that for calculating intra-cell interference, we would still use channel estimation.

Each BS uses the large-scale path loss statistics to calculate the leakage to all users other than the ones it serves, which corresponds to the inter-cell leakage term in the SLINR metric. Hence, the inter-cell leakage term found in (8) can be alternatively defined as

$$L_{kb}^{\text{inter}}(\mathbf{v}_{kb}) = \sum_{\substack{b'=1 \\ (b' \neq b)}}^{|\mathcal{B}|} \sum_{k'=1}^{K_{b'}} \mathbf{v}_{kb}^H \mathbf{D}_{k'b',b} \mathbf{v}_{kb}, \tag{20}$$

where $\mathbf{D}_{kb',b}$ is as defined in Section II-B.

When updating the variables described in Section IV-A, we would replace the inter-cell leakage terms with (20) and ignore terms related to leakage channel estimation error. In (18), the matrix formed with leakage channels $\hat{\mathbf{h}}_{k'b',b}\hat{\mathbf{h}}_{k'b',b}^H$ is replaced with $\mathbf{D}_{k'b',b}$.

C. Traffic Distribution

Similar to discussions made in [11], this method is based on the statistical distribution of traffic and completely eliminates the needs for real-time leakage channel information.

Instead of calculating the leakage to all neighbouring users that may or may not be scheduled, each BS can use a spatial traffic distribution for the cells around it. A traffic probability density function (PDF), $\Upsilon_b(x, y)$, can be constructed from a traffic survey around the b^{th} BS. Here, we model the traffic distribution as a mix of a uniform distribution with a probability P_h and a number of N_h hotspots modeled as bivariate normal distributions. Hence, we define the PDF of the traffic used to calculate the inter-cell leakage of the b^{th} BS as

$$\boldsymbol{\Gamma}_b(x, y) = f_h P_h \left(\frac{1}{a_b}\right) + (1 - P_h) \frac{f_h}{N_h 2\pi\sigma_h^2} \sum_{i=1}^{N_b} \exp\left(-\frac{(x - x_{h_i})^2 + (y - y_{h_i})^2}{2\sigma_h^2}\right). \tag{21}$$

The value of P_h can be varied to control the density of the hotpots centred at (x_{h_i}, y_{h_i}) with equal variance of σ_h^2. The term a_b is the sum of areas of adjacent cells around b^{th} BS and f_h is the normalizing factor.

Similar to the Statistical CSI method described in Section V-B, we can rewrite the inter-cell leakage terms by using the PDF of the region around the BS and ignoring the channel estimation error terms. Hence, the new inter-cell leakage term can be written as

$$L_{kb}^{\text{inter}}(\mathbf{v}_{kb}) = N_b \mathbf{v}_{kb}^H \mathbf{\Gamma}_b \mathbf{v}_{kb}, \tag{22}$$

where N_b is the total number of users in the cells that are next to the b^{th} BS and $\mathbf{\Gamma}_b$ is an auxiliary diagonal matrix with the diagonal entries defined as

$$[\mathbf{\Gamma}_b]_{mm} = \iint_{(x,y)\in\iota_b} \beta\left(\sqrt{(x-x_b)^2+(y-y_b)^2}\right)\Upsilon_b(x,y)\,dx\,dy, \tag{23}$$

where ι_b is the boundary of region considered for the inter-cell leakage of the b^{th} BS, (x_b, y_b) is the Cartesian coordinate for the b^{th} BS, and $\beta(d) = (1 + d/d_0)^{-\alpha}$ is the large-scale path loss. For (18), the matrix formed with leakage channels $\hat{\mathbf{h}}_{k'b',b}\hat{\mathbf{h}}_{k'b',b}^H$ is replaced with $\mathbf{\Gamma}_b$.

VI. Numerical Results and Analysis

We simulate a network with 7 hexagonal cells, with BSs located at the cell center. To eliminate network border effects, we use a wrap-around structure. Each BS is 1000 m apart, equipped with $M = 8$ antennas, and given a maximum transmit power of $P_T = 43$ dBm. Each user's channel is a resource block ($W = 180$ kHz) within a broadband transmission of 20 MHz, the noise power spectral density is -174 dBm/Hz, and the noise figure is 9 dBm. We use a path loss exponent of $\alpha = 3.76$ and a reference distance of $d_o = 0.3920$ m. For the channel estimation phase, the user transmit power is set to $p = 17$ dBm and we use the discrete Fourier Transform matrix to generate the set of orthogonal pilot sequences.

In an ideal setting, the number of orthogonal pilot signals τ_p equals K_b, i.e., the number of users within a cell. We set $\tau_d = 200$ as the length of downlink transmission phase [18]. The achievable rate, taking into account the pilot training overhead, is the product of the rate (1) and the fraction of the time that is used to transmit the data, i.e., $(\tau_d - \tau_p)/\tau_d$. Finally, we use a unity weighting scheme, i.e., $w_{kb} = 1$ for all k and b in (10a).

For Fig. 1, we first run Algorithm 1 using the Standard CSI Estimation scheme for both the perfect and imperfect CSI case. After obtaining the beamformers and scheduling decisions, we plot the sum-rate, using (1), as a function of the number of users in each cell, K_b. We average our results over 30 network topologies and 400 channel realizations. For the case of perfect CSI, we assume all BSs have knowledge of all users' channel without training, so we do not consider any overhead. For the imperfect CSI case, without considering the overhead, we observe a decrease in network sum data rate ranging from 12.5% to 14.6% and the performance worsens as the number of users per cell increases. Due to the greater impact of channel estimation errors, the weaker users suffer

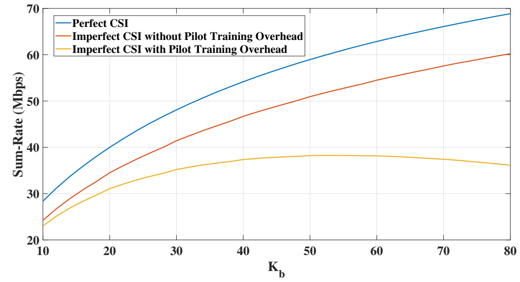

Fig. 1: Average sum-rate achieved as a function of the number of users per cell.

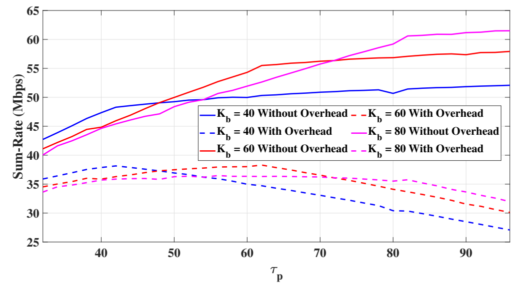

Fig. 2: Average sum-rate achieved as a function of the number of orthogonal pilot signals with $K_b = 40$, 60, and 80.

a larger percentage decrease in rate. This is expected since more estimated channels are used by the resource allocation algorithm, causing it to converge to a less desirable point.

If we employ the ideal setting for pilot signal lengths, the data rate is multiplied by $(\tau_d - K_b)/\tau_d$ to compensate for the pilot training overhead. Starting from $K_b = 50$, the data rate achieved decreases as the number of users per cell increases, due to more resources being used during the training phase. The performance reduces by 1.9 times as K_b approaches 80.

For Fig. 2, we no longer consider the ideal setting for channel estimation. Instead, we fix the number of users within one cell and vary the number of orthogonal pilot signals τ_p. Here, we use a simple pilot assignment policy that arbitrarily assigns pilots to users such that the number of repetitive pilots within one cell as well as the number of inter-cell repetitive pilots are minimized. We average our results over 60 network topologies and 400 channel realizations.

We simulate three cases: $K_b = 40$, 60, and 80. Since the resulting plots follow a similar pattern, we only discuss the case where $K_b = 60$. Our observations show that there are three different cases that result in different pilot assignments. For the case where $\tau_p < K_b$, there are not enough pilot sequences to assign to the K_b users, so the same pilots are used by several users in the same cell. This assignment increases pilot contamination, thus resulting in inaccurate estimated channels. This in turn decreases the data rate obtained by the resource allocation algorithm. As observed in Fig. 2, when $K_b = 60$ and $\tau_p = 32$, the data rate suffers a 1.32-fold decrease compared to $\tau_p = 60$, while it improves as τ_p approaches K_b. However, if we consider the training overhead, $\tau_p = 32$ only suffers a 1.10-fold decrease in data rate due to the fact that it uses less time for training.

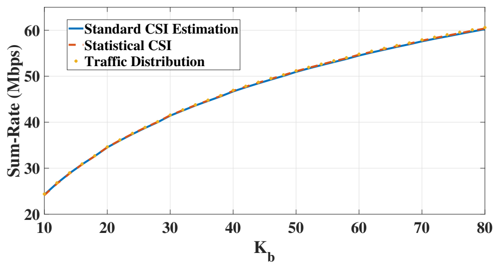

Fig. 3: Average sum-rate achieved as a function of the number of users per cell with different leakage computation methods.

For the case where $\tau_p > K_b$, the users in each cell will not use up all the orthogonal pilots. As a result, some users in another cell will use these pilots. This will decrease interference, resulting in better estimated channels but at a cost of increased training overhead. We observe that at $\tau_p = 96$ there is a 1.07-fold increase in data rates compared to $\tau_p = 60$ without considering overhead. However, the data rate suffers a 1.26-fold decrease, making it worse than the case of $\tau_p < K_b$. This implies that it is not worthwhile to overuse resources in the training phase, as the improved channel estimation accuracy does not compensate for the additional training overhead.

In Fig. 3, we simulate the different leakage computation schemes described in Section V. For the Standard CSI estimation, the setup used is as described for Fig. 1. Since its procedure involves estimating all the channels and using them for computing inter-cell leakage, it is used as a reference to evaluate the performance of the other two methods. For the traffic distribution method, we consider a uniform distribution for the users and ignore the hotspots by setting $P_h = 1$ in (21). The details of selecting hotspot locations and parameters (from a spatial traffic survey) will be analyzed in future work.

Both the Statistical CSI and the Traffic Distribution methods in Fig. 3 show matching data rates relative to the Standard CSI estimation method. Interestingly, with these more practical and scalable methods that do not estimate the inter-cell channels, we can still achieve almost the same data rate as the standard CSI method, which seems promising for future studies.

VII. Conclusions

This paper studied user scheduling and resource allocation on the downlink in a multi-cell MIMO network using a decentralized weighted sum pseudo-rate problem formulation. We used fractional programming to provide closed-form expressions for the optimized variables, then used the Hungarian algorithm to solve an assignment problem to determine user scheduling variables. This allowed us to construct an effective iterative resource allocation algorithm based on the SLINR metric, which takes both leakage and interference into consideration. Additionally, we proposed three approaches to calculate the leakage, with each requiring a different level of CSI estimation.

Our numerical results show that resource allocation based on SLINR does not suffer significant performance losses if we limit the number of resources used in the uplink training phase. Furthermore, the proposed methods to calculate the leakage are scalable and result in competitive data rates relative to the standard method.

References

[1] Q. Shi, M. Razaviyayn, Z.-Q. Luo, and C. He, "An iteratively weighted mmse approach to distributed sum-utility maximization for a MIMO interfering broadcast channel," *IEEE Transactions on Signal Processing*, vol. 59, no. 9, pp. 4331–4340, 2011.

[2] K. Hosseini, C. Zhu, A. A. Khan, R. S. Adve, and W. Yu, "Optimizing the MIMO cellular downlink: Multiplexing, diversity, or interference nulling?," *IEEE Trans. on Comm.*, vol. 66, no. 12, pp. 6068–6080, 2018.

[3] A. A. Khan, R. Adve, and W. Yu, "Optimizing multicell scheduling and beamforming via fractional programming and hungarian algorithm," in *2018 IEEE Globecom Workshops (GC Wkshps)*, pp. 1–6, 2018.

[4] A. Khan, R. Adve, and W. Yu, "Optimizing downlink resource allocation in multiuser MIMO networks via fractional programming and the hungarian algorithm," *IEEE Trans. on Comm.*, vol. 19, no. 8, pp. 5162–5175, 2020.

[5] H. Shen, W. Xu, A. Lee Swindlehurst, and C. Zhao, "Transmitter optimization for per-antenna power constrained multi-antenna downlinks: An slnr maximization methodology," *IEEE Transactions on Signal Processing*, vol. 64, no. 10, pp. 2712–2725, 2016.

[6] B. Boiadjieva and A. Klein, "Distributed two-stage beamforming with power allocation in multicell massive MIMO," in *ICC 2019 - 2019 IEEE International Conference on Communications (ICC)*, pp. 1–7, 2019.

[7] J. Zhao, H. Dai, X. Wu, X. Xu, and W.-P. Zhu, "Joint beamforming design for energy efficient wireless communications in heterogeneous intelligent connected vehicles networks," *IEEE Access*, vol. 7, pp. 170134–170143, 2019.

[8] P. Patcharamaneepakorn, S. Armour, and A. Doufexi, "On the equivalence between slnr and mmse precoding schemes with single-antenna receivers," *IEEE Comm. Letters*, vol. 16, no. 7, pp. 1034–1037, 2012.

[9] X. Xia, G. Wu, J. Liu, and S. Li, "Leakage-based user scheduling in MU-MIMO broadcast channel," *Science in China Series F: Information Sciences*, vol. 52, pp. 2259–2268, 12 2009.

[10] Y. Liu, J. X. Wang, P. Wang, and R. K. Yu, "Joint scheduling and precoding based on slnr criteria in MU-MIMO system," in *Mechanical Components and Control Engineering III*, vol. 668 of *Applied Mechanics and Materials*, pp. 1273–1277, Trans Tech Publications Ltd, 11 2014.

[11] H. A. Ammar, R. Adve, S. Shahbazpanahi, G. Boudreau, and K. Srinivas, "Distributed resource allocation optimization for user-centric cell-free MIMO networks," *IEEE Trans. on Wireless Comm.*, 2021.

[12] H. Q. Ngo, A. E. Ashikhmin, H. Yang, E. G. Larsson, and T. L. Marzetta, "Cell-free massive MIMO versus small cells," *CoRR*, vol. abs/1602.08232, 2016.

[13] S. M. Kay, *Fundamentals of Statistical Signal Processing, vol. 1: Estimation Theory*. Prentice Hall, 1997.

[14] H. A. Ammar, R. Adve, S. Shahbazpanahi, G. Boudreau, and K. V. Srinivas, "Downlink resource allocation in multiuser cell-free MIMO networks with user-centric clustering," *IEEE Trans. on Wireless Comm.*, pp. 1–1, 2021.

[15] K. Shen and W. Yu, "Fractional programming for communication systems—part II: Uplink scheduling via matching," *IEEE Transactions on Signal Processing*, vol. 66, no. 10, pp. 2631–2644, 2018.

[16] K. Shen and W. Yu, "Fractional programming for communication systems—part I: Power control and beamforming," *IEEE Transactions on Signal Processing*, vol. 66, no. 10, pp. 2616–2630, 2018.

[17] H. W. Kuhn and B. Yaw, "The hungarian method for the assignment problem," *Naval Res. Logist. Quart*, pp. 83–97, 1955.

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