Novel Combinatorial Results on Downlink MU-MIMO Scheduling with applications

Yasaman Ghasempour* Narayan Prasad† Mohammad Khojastepour† Sampath Rangarajan†
* Rice University {Ghasempour}@rice.edu; † NEC Labs, {prasad,amir,sampath}@nec-labs.com.

***Abstract*—In this paper we consider the classical problem of downlink (DL) multiuser (MU) multi-input-multi-output (MIMO) scheduling with linear transmit precoding. This problem was formulated over a decade ago and has been deeply studied since then. Moreover, MU-MIMO with linear transmit precoding is being increasingly pursued as a key technology by the industry, with a strong emphasis on efficient scheduling algorithms. However, the intractable combinatorial nature of the problem has mostly restricted algorithm design to the realm of simple greedy heuristics. Such algorithms do not exploit any underlying structure in the problem. Recently, it has been formally shown that in general this problem is even hard to approximate. Our significant contribution in this work is to consider the practically most important choices of linear precoding and power allocation, and show that the resulting problems can be cast as ones where a difference of two submodular set functions has to be maximized. This opens up a new framework for MU-MIMO scheduler design. We use this framework to design an algorithm and demonstrate that significant gains can be achieved at a reasonable complexity. Our framework can also incorporate analog receive beamforming which is deemed to be essential in mmWave MIMO systems.**

I. Introduction

The huge promise of MU-MIMO was revealed in the seminal work of [1] which derived the theoretical limits for a broadcast channel under ideal conditions. This led to investigations that have continued for the past decade. Notable works in these investigations have considered more practical yet asymptotically optimal linear transmit precoding [2], [3] and have also considered the impact of imperfect (quantized) channel state information (CSI) [4]. In industry, the effort to standardize MU-MIMO in 3GPP LTE was initiated right at the onset of Rel. 8 and has continued since then. It has so far resulted in the adoption of precoded pilots, which obviate the need for restricting transmit precoders to a finite codebook, thereby enabling the full optimization of linear transmit precoding schemes. However, the performance results of MU-MIMO in FDD systems equipped with a modest number of transmit antennas (typically 2 or 4) have so far been underwhelming. This is because the available coarse CSI and the small number of cross-polarized transmit antennas are not conducive to creating beams that enable good separation of different users in the signal space. The advent of massive MIMO has once again energized MU-MIMO. Indeed, simultaneous transmission to several different users is the main benefit promised by massive MIMO which is a key 5G technology [5]. The emphasis now is on realizing low complexity scheduling algorithms.

In this paper, we consider the classical DL MU-MIMO scheduling problem, which must be solved by the base station in each cell of an LTE-A network, every sub-frame. We focus on the narrowband (frequency non-selective) model and note that our obtained results can be directly applied to a wideband model by using the approach in [6]. The problem at hand involves selecting users on each narrowband resource block (RB), where a selected user can be assigned more than one stream (a.k.a. a transmit rank greater than 1). Note that allowing multiple streams per user is important especially in practical massive MIMO scenarios where the number of active users is not large compared to the number of transmit antennas at the base station (BS). The narrowband user selection problem has been studied mainly for the case when each user has one antenna and hence can be assigned one stream [2], [3]. Due to the intractability of this problem, several heuristics have been proposed (cf. [7]). A combinatorial optimization view has been adopted in [8], [9]. More recently in an important work [10] it was shown that this problem in general is even hard to approximate. To the best of our knowledge no prior work has discovered any underlying exploitable structure in the problem without making further simplifying assumptions. Our contributions in this paper are as follows:

- We propose a new framework for designing MU-MIMO scheduling algorithms. In particular, we show that the scheduling problem can be cast as one where a difference of submodular (DS) set functions has to be maximized. This framework, a.k.a. DS framework, was proposed to solve machine learning problems in [11] and can be viewed as the discrete analogue of maximizing the difference of concave functions, where we note that the latter continuous optimization technique has been successfully used to optimize transmit precoder matrices [12].
- We establish that our framework can incorporate the practically most important choices of linear transmit precoding methods as well as power allocations. We then use our framework to design an algorithm and demonstrate that significant gains can be achieved at a reasonable complexity.
- Our framework can also incorporate analog receive beamforming which is deemed to be essential in mmWave MIMO systems, where we note that mmWave MIMO systems are another key 5G technology [13].

In the following sections we present our main theoretical results with full proofs and a few representative simulation results. Due to space constraints some auxiliary proofs and additional description and simulation results are deferred to a longer version [14].

Notation: we will use boldface uppercase (lowercase) alphabets to denote matrices (vectors). Further, $|.|$ is used to denote the determinant of its matrix argument as well as cardinality of its input set. $(.)^\dagger$ is used to denote the conjugate transpose of its matrix argument while $\|.\|$ denotes the Frobenius (ℓ_2) norm of its matrix (vector) argument.

II. System Model

We consider the classical DL MU-MIMO system with M_t transmit antennas at the base station (BS) and M_r receive antennas at each user. We assume K active users in the cell of interest and focus on data transmission on a resource block in each scheduling interval. Without loss of generality, in the following analysis we assume each resource block to be of unit size, on which each user sees a frequency non-selective channel. Then, the signal received by user k is modeled as

$$\mathbf{y}_k = \breve{\mathbf{H}}_k \mathbf{x} + \boldsymbol{\eta}_k, \quad k = 1, \cdots, K, \tag{1}$$

where $\breve{\mathbf{H}}_k \in \mathbb{C}^{M_r \times M_t}$ is the channel matrix and $\boldsymbol{\eta}_k \sim \mathcal{CN}(\mathbf{0}, \mathbf{I})$ is the additive noise. The signal vector $\mathbf{x}$ transmitted by the BS can be expanded as $\mathbf{x} = \mathbf{T}\sum_{k\in\mathcal{U}} \mathbf{V}_k \mathbf{s}_k$, where $\mathcal{U}$ is the set of users that are co-scheduled (or grouped) together. $\mathbf{T}$ is the $M_t \times M$ analog transmit precoding matrix such that $\mathbf{T}^\dagger\mathbf{T} = \mathbf{I}$. Here we adopt the framework from [15] where the cell specific precoding matrix $\mathbf{T}$ is optimized at a coarse time scale and can thus be regarded as a constant matrix in the fine time-scale problem of interest to us. $\mathbf{V}_k,\ k \in \mathcal{U}$ is the $M \times r_k$ digital transmit precoding matrix used to transmit to the kth user and has r_k columns. $\mathbf{s}_k$ is the $r_k \times 1$ symbol vector intended for the kth user. Furthermore, let $S = \sum_{k\in\mathcal{U}} r_k$ be the total number of co-scheduled streams or total rank and let the total power for all streams be ρ. We consider the practically most important power allocation, which is to equally split the available power among all transmitted streams. In this case we have $E[\mathbf{s}\mathbf{s}^\dagger] = (\rho/S)\mathbf{I}$ with each column of $\mathbf{V}_k,\ \forall\, k \in \mathcal{U}$ (and hence of $\mathbf{T}\mathbf{V}_k,\ \forall\ k \in \mathcal{U}$) normalized to unit norm. We note that this power allocation is also referred to as column-norm based normalization [16] and is mandated for MU-MIMO performance evaluations in 3GPP LTE. We also consider another practical power allocation referred to as the matrix norm based normalization [16] in which the sum of all squared column norms (across all $\mathbf{V}_k,\ \forall\, k \in \mathcal{U}$) is normalized to unity, together with $E[\mathbf{s}\mathbf{s}^\dagger] = \rho\mathbf{I}$. We define $\mathbf{H}_k$ to be the estimate of $\breve{\mathbf{H}}_k\mathbf{T},\ \forall\, k$ available at the BS and adopt the conventional approach where the BS assumes the estimates at hand to be sufficiently accurate.

Define $\mathbf{A} = [\mathbf{A}_k]_{k\in\mathcal{U}}$, where $\mathbf{A}_k = \sqrt{\rho'}\mathbf{V}_k,\ \forall\ k \in \mathcal{U}$, as the scaled and concatenated digital transmit precoding matrix of size $M \times S$ for MU-MIMO transmission, where the scaling factor $\rho' = \rho/S$ for column-norm and $\rho' = \rho$ for matrix-norm based normalizations, respectively. Each user in order to receive its data, employs an RF analog receive beamforming front-end followed by baseband linear detection. Such an architecture is significantly preferred in mmWave systems [13]. In this work, we incorporate the practically meaningful scenario in which each user uses a codebook $\mathcal{W}$ for analog receive beamforming. To describe the data reception we focus on any user k. To receive data sent on each one of its r_k streams that user k employs r_k unit-norm beamforming vectors from $\mathcal{W}$. Let $\mathbf{G}_k$ denote the $M_r \times r_k$ matrix whose columns are these beamforming vectors. The received signal post receive beamforming is down-converted and detected at baseband. We consider two types of detection methods at baseband. The first one is a simple linear detector at the baseband, in which no further mitigation is carried out to suppress inter-stream residual interference. We refer to this method as the matched filter (MF) baseband detector. The resulting signal-to-interference plus noise ratio (SINR) for the ith stream (or layer) of the kth user is given by

$$\gamma_{i,k} = \frac{|[\mathbf{G}_k^\dagger \mathbf{H}_k \mathbf{A}_k]_{i,i}|^2}{1 + \sum_{j=1}^{S} |[\mathbf{G}_k^\dagger \mathbf{H}_k \mathbf{A}]_{i,j}|^2 - |[\mathbf{G}_k^\dagger \mathbf{H}_k \mathbf{A}_k]_{i,i}|^2}, \tag{2}$$

where $i = 1, \cdots, r_k$ and $[.]_{i,j}$ is the $(i,j)^{th}$ element of the matrix argument. The corresponding information rate is given by

$$\eta_{i,k} = \log(1 + \gamma_{i,k}), \tag{3}$$

so that the information rate over all the streams of user k can be written as $R_k = \sum_{i=1}^{r_k} \eta_{i,k}$. The second detection method that we consider is the non-linear optimal method of detection at the baseband, for which the corresponding information rate over all streams is given by

$$R_k = \log|\mathbf{I} + \mathbf{Q}_k^{-1}\mathbf{G}_k^\dagger \mathbf{H}_k \mathbf{A}_k (\mathbf{G}_k^\dagger \mathbf{H}_k \mathbf{A}_k)^\dagger| \tag{4}$$

where $\mathbf{Q}_k = \mathbf{G}_k^\dagger \mathbf{G}_k + \sum_{\ell\in\mathcal{U}\setminus k} \mathbf{G}_k^\dagger \mathbf{H}_k \mathbf{A}_\ell (\mathbf{G}_k^\dagger \mathbf{H}_k \mathbf{A}_\ell)^\dagger$ represents the covariance matrix of additive noise and interference from streams intended for other users. Note that the additive noise is colored by the analog receive beamforming operation.

We outline the three (digital) linear transmit precoding methods considered in this paper which cover all the main practical ones. Consider any given user set $\mathcal{U}$ along with a rank vector $\mathbf{r}$. In all these methods it is assumed for precoder construction that each user $k \in \mathcal{U}$ that is assigned rank r_k will receive data only in the span of its chosen r_k receive beamforming vectors in $\mathbf{G}_k$. Consequently, we define the $r_k \times M$ matrix $\tilde{\mathbf{H}}_k = \mathbf{G}_k^\dagger \mathbf{H}_k$.

The construction of the transmit precoder matrices then proceeds by using the matrices $\{\tilde{\mathbf{H}}_k\}_{k\in\mathcal{U}}$.

- *Zero Forcing (ZF):* Let $\tilde{\mathbf{H}} = \left(\left[\tilde{\mathbf{H}}_k^\dagger\right]_{k\in\mathcal{U}}\right)^\dagger$ denote the $S \times M$ composite matrix, where we recall $S = \sum_{k\in\mathcal{U}} r_k$. We obtain the matrix $\mathbf{V} = \tilde{\mathbf{H}}^\dagger(\tilde{\mathbf{H}}\tilde{\mathbf{H}}^\dagger)^{-1}\mathbf{D}$ where $\mathbf{D}$ is a diagonal matrix which normalizes all columns of $\tilde{\mathbf{H}}^\dagger(\tilde{\mathbf{H}}\tilde{\mathbf{H}}^\dagger)^{-1}$ to have unit norm. Then $\mathbf{V}_k$ is obtained as the sub-matrix of $\mathbf{V}$ formed by the r_k columns corresponding to user k.
- *Block Diagonalization (BD):* Let $\tilde{\mathbf{H}}_{\bar{k}} = ([\tilde{\mathbf{H}}_j^\dagger]_{j\in\mathcal{U}:j\neq k})^\dagger$ denote the $(\sum_{j\in\mathcal{U},j\neq k} r_j) \times M$ composite matrix that excludes user k. We then obtain the matrix $(\mathbf{I} -$

$\tilde{\mathbf{H}}_{\bar{k}}^{\dagger}(\tilde{\mathbf{H}}_{\bar{k}}\tilde{\mathbf{H}}_{\bar{k}}^{\dagger})^{-1}\tilde{\mathbf{H}}_{\bar{k}})\tilde{\mathbf{H}}_{k}^{\dagger}(\mathbf{G}_{k}^{\dagger}\mathbf{G}_{k})^{-1/2}$ and then form the $M \times r_k$ matrix $\mathbf{V}_k$ by choosing its first r_k dominant left singular vectors corresponding to its first r_k dominant singular values.

- *Maximal Ratio Transmission (MRT):* Let $\tilde{\mathbf{H}} = \left(\left[\tilde{\mathbf{H}}_k^{\dagger}\right]_{k\in\mathcal{U}}\right)^{\dagger}$ denote the $S \times M$ composite matrix as before. Here we consider the matrix norm based normalization. The transmit precoder $\mathbf{V}_k$ used for any user $k \in \mathcal{U}$ with rank r_k, is simply the matrix $\tilde{\mathbf{H}}_k^{\dagger}/\|\tilde{\mathbf{H}}\|$. Notice that the choice of the precoder depends on the co-scheduled users via the normalization factor.

We note that column-norm based normalization can outperform matrix norm based normalization for ZF transmit precoding, whereas the converse holds true for MRT precoding [16]. This observation has motivated the above power allocation choices. Also note that for BD precoding the optimal detector can be implemented as a linear detector, whereas for ZF precoding MF detector is optimal whenever $\mathbf{G}_k^{\dagger}\mathbf{G}_k = \mathbf{I}$, $\forall\, k \in \mathcal{U}$ [14].

III. Problem formulation

Our objective in the subsequents sections is to design efficient algorithms to optimize $\sum_{k\in\mathcal{U}} w_k R_k$, where w_k is the weight or priority assigned to user k, under certain practical constraints. Due to space constraints, we consider only the most natural pairings of precoder construction and receiver detection, which are to use either MRT or ZF transmit precoding with the MF basedband detection. On the other hand, we suppose that BD precoding is used in conjunction with optimal baseband detection. Indeed, these pairings were assumed (and subsequently widely adopted) in the notable papers that proposed the three transmit precoding methods [2], [4]. Note that for each such combination of the aforementioned transmit precoder construction and receiver detection methods, the resulting weighted sum rate depends on the choice of user set $\mathcal{U}$ as well as the choice of their assigned transmit ranks and the receive beamforming vectors. Moreover, there can be a non-linear dependence (or coupling) between the choice of receive beamforming vectors and the transmit precoder construction. As a result, the optimization problem at hand appears to be intractable at the first glance. Indeed, [10] has considered another variation of this problem (with ZF precoding and full power optimization) and shown that the objective function therein is not submodular (cf. definitions in the appendix). We first verified that the objective functions in our problems are also not submodular and then proceeded to unearth and exploit the structure hidden in them.

IV. Structure in the Rate Expression

We begin by considering both MRT and ZF transmit precoders with matched filter baseband detection. Our first observation then is that we can regard each user and receive beamfomer combination as a *virtual user*. In particular, consider any stream of any user k that is received along any beamformer $\mathbf{w} \in \mathcal{W}$, and define ψ as the corresponding virtual user with its channel being the $1 \times M$ vector, $\mathbf{z}_{\psi}^{\dagger} = \mathbf{w}^{\dagger}\mathbf{H}_k$. The received statistic for this virtual user can be written as

$$y_{\psi} = \mathbf{z}_{\psi}^{\dagger}\mathbf{x} + \eta_{\psi} \tag{5}$$

where $\eta_{\psi} \sim \mathcal{CN}(0,1)$. Define a ground set Ψ of all virtual users ψ such that $\|\mathbf{z}_{\psi}\|^2 \geq \epsilon > 0$ so that the size of Ψ is at-most $K|\mathcal{W}|$. Note here that $\epsilon > 0$ is any arbitrarily chosen threshold. [1] Consider any choice of co-scheduled virtual users $\mathcal{A} \subseteq \Psi$. Suppose MRT precoding at the BS. Consider any choice of co-scheduled virtual users $\mathcal{A} \subseteq \Psi$ and define the matrix $\mathbf{Z}_{\mathcal{A}} = [\mathbf{z}_{\psi}]_{\psi\in\mathcal{A}}$ of size $M \times |\mathcal{A}|$. For this choice using (5) and (3) the rate for virtual user $\psi \in \mathcal{A}$ is given by

$$R_{\psi}(\mathcal{A}) = \log\left(1 + \frac{\rho\|\mathbf{z}_{\psi}\|^4/\|\mathbf{Z}_{\mathcal{A}}\|^2}{1 + \sum_{\psi'\in\mathcal{A}\setminus\psi}\rho\mathbf{z}_{\psi}^{\dagger}\mathbf{z}_{\psi'}\mathbf{z}_{\psi'}^{\dagger}\mathbf{z}_{\psi}/\|\mathbf{Z}_{\mathcal{A}}\|^2}\right) \tag{6}$$

On the other hand, for any $\psi \in \Psi \setminus \mathcal{A}$, we set $R_{\psi}(\mathcal{A}) = 0$. We offer the following result that is proved in the appendix and reveals the structure in the rate expression.

Proposition 1. *The rate achieved by any virtual user $\psi \in \Psi$ under MRT precoding and set $\mathcal{A} \subseteq \Psi : \mathcal{A} \neq \phi$, where ϕ denotes the empty set, can be expressed as*

$$R_{\psi}(\mathcal{A}) = \underbrace{\log\left(\|\mathbf{Z}_{\mathcal{A}}\|^2 + \sum_{\psi'\in\mathcal{A}}\rho\mathbf{z}_{\psi}^{\dagger}\mathbf{z}_{\psi'}\mathbf{z}_{\psi'}^{\dagger}\mathbf{z}_{\psi}\right)}_{\triangleq f_{\psi}^{\mathrm{MRT}}(\mathcal{A})} - \underbrace{\log\left(\|\mathbf{Z}_{\mathcal{A}}\|^2 + \sum_{\psi'\in\mathcal{A}\setminus\psi}\rho\mathbf{z}_{\psi}^{\dagger}\mathbf{z}_{\psi'}\mathbf{z}_{\psi'}^{\dagger}\mathbf{z}_{\psi}\right)}_{\triangleq g_{\psi}^{\mathrm{MRT}}(\mathcal{A})} \tag{7}$$

Further, for $\mathcal{A} = \phi$ we define $R_{\psi}(\phi) = 0$ with $f_{\psi}^{\mathrm{MRT}}(\phi) = g_{\psi}^{\mathrm{MRT}}(\phi) = -\log(2/\epsilon)$. Then, the set functions $f_{\psi}^{\mathrm{MRT}}(.), g_{\psi}^{\mathrm{MRT}}(.)$ are both submodular set functions over the set Ψ.

We now consider the more complicated case of ZF precoding. The key complication here that we need to overcome is that the transmit precoder for each user depends not only on its channel matrix and choice of receive beamformers, but also on those of other co-scheduled users. Moreover, the latter dependence is non-linear and not just via a normalization factor. We again use the virtual user concept and recall the model in (5) for some virtual user $\psi \in \Psi$. Consider any choice of co-scheduled virtual users $\mathcal{A} \subseteq \Psi$ and define the matrix $\mathbf{Z}_{\mathcal{A}} = [\mathbf{z}_{\psi}]_{\psi\in\mathcal{A}}$ as before, along with $\mathbf{Z}_{\mathcal{A}\setminus\psi} = [\mathbf{z}_{\psi'}]_{\psi'\in\mathcal{A}\setminus\psi}$. Further, suppose that the matrix $\mathbf{Z}_{\mathcal{A}}^{\dagger}\mathbf{Z}_{\mathcal{A}}$ is invertible which is required for the zero forcing operation to be defined. The ZF matrix is given by $\mathbf{Z}_{\mathcal{A}}(\mathbf{Z}_{\mathcal{A}}^{\dagger}\mathbf{Z}_{\mathcal{A}})^{-1}\mathbf{D}$, where $\mathbf{D}$ is the diagonal matrix normalizing the columns of $\mathbf{Z}_{\mathcal{A}}(\mathbf{Z}_{\mathcal{A}}^{\dagger}\mathbf{Z}_{\mathcal{A}})^{-1}$. We can

[1] Such a threshold is anyway necessary in practice since a virtual user whose channel vector norm is poorer than a minimum threshold will never be selected as it cannot support the smallest available MCS.

then express the rate for virtual user $\psi \in \mathcal{A}$ as

$$R_\psi(\mathcal{A}) = \log\big(1 + \rho\|\mathbf{z}_\psi\|^2/|\mathcal{A}| - \rho\mathbf{z}_\psi^\dagger \mathbf{Z}_{\mathcal{A}\setminus\psi}(\mathbf{Z}_{\mathcal{A}\setminus\psi}^\dagger \mathbf{Z}_{\mathcal{A}\setminus\psi})^{-1}\mathbf{Z}_{\mathcal{A}\setminus\psi}^\dagger \mathbf{z}_\psi/|\mathcal{A}|\big) \quad (8)$$

On the other hand, for any $\psi \in \Psi \setminus \mathcal{A}$, we set $R_\psi(\mathcal{A}) = 0$. Note here that the *residual*

$$\mathrm{Res}(\psi, \mathcal{A}\setminus\psi) \triangleq \|\mathbf{z}_\psi\|^2 - \mathbf{z}_\psi^\dagger \mathbf{Z}_{\mathcal{A}\setminus\psi}(\mathbf{Z}_{\mathcal{A}\setminus\psi}^\dagger \mathbf{Z}_{\mathcal{A}\setminus\psi})^{-1}\mathbf{Z}_{\mathcal{A}\setminus\psi}^\dagger \mathbf{z}_\psi$$

is the squared norm of the component of $\mathbf{z}_\psi$ in the orthogonal complement of $\mathbf{Z}_{\mathcal{A}\setminus\psi}$ and hence is always non-negative.

We now proceed to unearth the structure in this rate expression. Towards this end, let us first define the matrix

$$\mathbf{B} = \rho\mathbf{Z}_\Psi^\dagger \mathbf{Z}_\Psi \quad (9)$$

with the understanding that $\mathbf{B}_\mathcal{A},\ \forall\ \mathcal{A} \subseteq \Psi$ is the principal submatrix of $\mathbf{B}$ with row and column indices drawn from $\mathcal{A}$. Note that $\mathbf{B}_\mathcal{A} = \rho\mathbf{Z}_\mathcal{A}^\dagger \mathbf{Z}_\mathcal{A}\ \forall\ \mathcal{A} \subseteq \Psi$. Along similar lines, for each virtual user $\psi \in \Psi$ and any scalar $a \geq 0$, let us define the matrix,

$$\mathbf{C}(a,\psi) = a\mathbf{e}_\psi\mathbf{e}_\psi^\dagger + \rho\mathbf{Z}_\Psi^\dagger \mathbf{Z}_\Psi \quad (10)$$

where $\mathbf{e}_\psi$ is a $|\Psi| \times 1$ vector that has a one as its ψ^{th} element and zeros everywhere else. As before, let $\mathbf{C}_\mathcal{A}(a,\psi),\ \forall\ \mathcal{A} \subseteq \Psi$ be the principal submatrix of $\mathbf{C}(a,\psi)$ with row and column indices drawn from $\mathcal{A}$. Let us next define a family of subsets, $\underline{\mathcal{I}}$, of Ψ such that $\phi \in \underline{\mathcal{I}}$ and all subsets $\mathcal{A}$ of Ψ for which $\mathbf{B}_\mathcal{A}$ is invertible are members of $\underline{\mathcal{I}}$ and conversely for any non-empty member $\mathcal{A} \in \underline{\mathcal{I}}$, $\mathbf{B}_\mathcal{A}$ is invertible. It is readily seen that this family is *downward closed* and that all singleton sets $\{\psi\}_{\psi\in\Psi}$ are members of $\underline{\mathcal{I}}$.

Our next result proved in the appendix, reveals that it is possible to write (8) in a more amenable form. We adopt the convention that $0\log(0) = 0$ and that $\log|.|$ returns zero whenever the input matrix is empty or null matrix.

Proposition 2. *The rate achieved by any virtual user under ZF precoding can be expressed as*

$$R_\psi(\mathcal{A}) = \underbrace{\big(\log|\mathbf{C}_\mathcal{A}(|\mathcal{A}|,\psi)| - |\mathcal{A}|\log|\mathcal{A}|\big)}_{f_\psi^{\mathrm{ZF}}(\mathcal{A})} - \underbrace{\big(\log\left|\mathbf{B}_{\mathcal{A}\setminus\psi}\right| - |\mathcal{A}\setminus\psi|\log|\mathcal{A}|\big)}_{g_\psi^{\mathrm{ZF}}(\mathcal{A})} \quad (11)$$

The functions $f_\psi^{\mathrm{ZF}}(.), g_\psi^{\mathrm{ZF}}(.)$ are both submodular over the family $\underline{\mathcal{I}}$.

We now proceed to analyze the case where the BS employs BD transmit precoding and each user employs the optimal baseband detection. In this case we need to jointly consider the rate across all virtual users that correspond to the same (real) user. Furthermore, we need to account for the coloring of the noise due to receive beamforming. To make the problem tractable, we follow an approach where we first assume that the power per stream (virtual user) is given and does not vary with the number of selected virtual users. This assumption results in no loss of optimality if we also consider all possible total number of streams that can be scheduled, and solve the problem at hand for each such total number. In particular, for each value, S, of the total number of streams, we fix the power per stream to be $\hat{\rho} = \rho/S$ and solve the weighted sum rate maximization under the constraint that no more than S streams can be scheduled. Then, suppose that we are any given a value for the power per stream, $\hat{\rho}$. Let $u : \Psi \to \{1,\cdots,K\}$ denote a scalar valued function which returns the actual user corresponding to any virtual user in Ψ. Similarly, let $\mathbf{w} : \Psi \to \mathcal{W}$ denote a vector valued function which returns the receive beamforming vector corresponding to any virtual user in Ψ. We will use the index $k \in \{1,\cdots,K\}$ to denote an actual user. For each user $k \in \{1,\cdots,K\}$, define the matrix

$$\mathbf{F}^{(k)} = \hat{\rho}\mathbf{Z}_\Psi^\dagger \mathbf{Z}_\Psi + \mathbf{L}^{(k)}, \quad (12)$$

where $\mathbf{L}^{(k)} = [\ell_{\psi,\psi'}^{(k)}]_{\psi,\psi'\in\Psi}$ is a $|\Psi| \times |\Psi|$ matrix whose $(\psi,\psi')^{th}$ entry is given by

$$\ell_{\psi,\psi'}^{(k)} = \begin{cases} \mathbf{w}(\psi)^\dagger\mathbf{w}(\psi'), & u(\psi) = u(\psi') = k \\ 0, & \text{else} \end{cases} \quad (13)$$

As done previously, we let $\mathbf{F}_\mathcal{A}^{(k)}$ ($\mathbf{L}_\mathcal{A}^{(k)}$), $\forall\ \mathcal{A} \subseteq \Psi$ denote the principal submatrix of $\mathbf{F}^{(k)}$ ($\mathbf{L}^{(k)}$) with row and column indices drawn from $\mathcal{A}$. We offer the following result whose proof is deferred to [14].

Proposition 3. *The rate achieved by any user under BD precoding can be expressed as*

$$R_k(\mathcal{A}) = \underbrace{\log\left|\mathbf{F}_\mathcal{A}^{(k)}\right|}_{f_k^{\mathrm{BD}}(\mathcal{A})} - \underbrace{\left(\log|\mathbf{F}_{\mathcal{A}\setminus\{\psi:u(\psi)=k\}}^{(k)}| + \log|\mathbf{L}_{\mathcal{A}\cap\{\psi:u(\psi)=k\}}^{(k)}|\right)}_{g_k^{\mathrm{BD}}(\mathcal{A})} \quad (14)$$

The functions $f_k^{\mathrm{BD}}(.), g_k^{\mathrm{BD}}(.)$ are both submodular over the family $\underline{\mathcal{I}}$.

V. Algorithm Design Framework

We will illustrate the design frame work that is based on optimizing the *difference of submodular (DS) set functions* [11]. We proceed to explain the DS framework for ZF precoding, while noting that other precoding methods can be handled similarly. Then, the optimization problem at hand can be posed as

$$\max_{\mathcal{A}\in\underline{\mathcal{I}}\ \&\ \mathcal{A}\in\underline{\mathcal{J}}} \left\{\sum_{\psi\in\Psi} R_\psi(\mathcal{A})\right\}, \quad (15)$$

where we use the family of sets $\underline{\mathcal{J}}$ to impose further constraints. We consider two key practical constraints:

- The total number of selected virtual virtual users should not exceed a bound, i.e., a cardinality constraint $|\mathcal{A}| \leq S_t$ is imposed, where S_t is the number of transmit RF chains.
- The total number of selected virtual virtual users that correspond to the same real user k should not exceed a

bound, i.e., a cardinality constraint $|\{\psi \in \mathcal{A} : u(\psi) = k\}| \leq S_{r,k}, \forall k$ is imposed, where $S_{r,k}$ is the number of receive RF chains at user k.

Let $\mathcal{J}$ be the collection of all subsets of Ψ that meet the aforementioned two constraints. Then, we have the following observation that follows upon verifying the properties stated in the appendix.

Proposition 4. *The family $\mathcal{J}$ defines a matroid over Ψ.*

Using (11) we can re-state (15) as

$$\max_{\mathcal{A} \in \underline{\mathcal{I}} \ \& \ \mathcal{A} \in \underline{\mathcal{J}}} \left\{ \sum_{\psi \in \Psi} (f_\psi^{\rm ZF}(\mathcal{A}) - g_\psi^{\rm ZF}(\mathcal{A})) \right\}. \quad (16)$$

The DS framework entails an iterative approach in which each iteration seeks to improve the current best solution at hand by solving a simpler maximization problem. Suppose at any iteration, the current best solution is given by $\hat{\mathcal{A}}$. Then, let $g(\mathcal{A}/\mathcal{B}) \triangleq g(\mathcal{A} \cup \mathcal{B}) - g(\mathcal{B})$ define the marginal gain obtained upon adding set $\mathcal{A}$ to set $\mathcal{B}$ for any set function $g(.)$, for any subsets $\mathcal{A}, \mathcal{B}$ of a ground set such that $g(\mathcal{B}), g(\mathcal{A} \cup \mathcal{B})$ are both defined. Next, define a modular upper bound as follows

$$\begin{aligned} g_{\hat{\mathcal{A}},\psi}^{\rm ZF,UB}(\mathcal{A}) \triangleq g_\psi^{\rm ZF}(\hat{\mathcal{A}}) - \sum_{\psi' \in \hat{\mathcal{A}}} g_\psi^{\rm ZF}(\psi'/\hat{\mathcal{A}} \setminus \psi') + \\ \sum_{\psi' \in \mathcal{A}} a_{\hat{\mathcal{A}},\psi}(\psi'), \ \forall \, \mathcal{A} \in \underline{\mathcal{I}}, \quad (17) \end{aligned}$$

where

$$a_{\hat{\mathcal{A}},\psi}(\psi') = \begin{cases} g_\psi^{\rm ZF}(\psi'), & \psi' \notin \hat{\mathcal{A}} \\ g_\psi^{\rm ZF}(\psi'/\hat{\mathcal{A}} \setminus \psi') & \text{else} \end{cases} \quad (18)$$

It can be shown that

$$g_{\hat{\mathcal{A}},\psi}^{\rm ZF,UB}(\mathcal{A}) \geq g_\psi^{\rm ZF}(\mathcal{A}), \ \forall \, \mathcal{A} \in \underline{\mathcal{I}}, \quad (19)$$

with equality in (19) at $\hat{\mathcal{A}} = \mathcal{A}$. Thus, $R_{\hat{\mathcal{A}},\psi}(\mathcal{A}) = f_\psi^{\rm ZF}(\mathcal{A}) - g_{\hat{\mathcal{A}},\psi}^{\rm ZF,UB}(\mathcal{A}), \ \forall \, \mathcal{A} \in \underline{\mathcal{I}}$, satisfies $R_{\hat{\mathcal{A}},\psi}(\mathcal{A}) \leq R_\psi(\mathcal{A}), \ \forall \, \mathcal{A} \in \underline{\mathcal{I}}$ with equality at $\hat{\mathcal{A}} = \mathcal{A}$. With this bound in hand, we proceed to solve the following problem

$$\max_{\mathcal{A} \in \underline{\mathcal{I}} \ \& \ \mathcal{A} \in \underline{\mathcal{J}}} \{R_{\hat{\mathcal{A}},\psi}(\mathcal{A})\}, \quad (20)$$

Let $\breve{\mathcal{A}}$ be an obtained optimized solution. Then, if $R_{\hat{\mathcal{A}},\psi}(\breve{\mathcal{A}}) > R_{\hat{\mathcal{A}},\psi}(\hat{\mathcal{A}})$ we can be sure that the current best solution at hand has been improved, i.e., $R_\psi(\breve{\mathcal{A}}) > R_\psi(\hat{\mathcal{A}})$. The key property of (20) is that since the objective is now a *submodular set function* and the constraint is a matroid, (20) can be relatively well optimized via simple methods such as the classical greedy method [17]. An important by-product of the submodularity of the objective is that we can use the *Lazy Greedy* implementation to significantly lower the complexity of the greedy method [18]. The DS procedure terminates if there is no improvement in the current best solution at hand. Otherwise, we proceed to the next iteration using $\hat{\mathcal{A}} \to \breve{\mathcal{A}}$ as the current best solution.

VI. Simulation Results

In this section we present our simulation results to demonstrate the utility of the DS framework. We first consider a simple setup comprising of a single cell with narrowband i.i.d. Rayleigh fast fading and randomly dropped single receive antenna users. The BS is equipped with multiple ($S_t \geq 1$) RF chains and uses a fixed set of columns from the DFT matrix as its analog precoder. In Figs. 1 and 2 we plot the cell average sum rate versus the transmit power and the number of transmit antennas, respectively. In these figures we have compared our proposed algorithm (using the DS framework and lazy greedy efficient implementation) with: the SU-MIMO scheme where only one user is scheduled in each slot, random selection (where a fixed number S_t of users are randomly chosen in each slot), the conventional greedy [7] and the optimal exhaustive search based one, respectively. As seen from the figures our scheme performs quite close to the optimal brute-force one.

We next consider an mmWave network employing OFDMA for DL access. This network comprises of 6 access points or base stations and 57 users, that were dropped following the 3GPP HetNet distribution. Each BS is equipped with $M_t = 16$ transmit antennas, $S_t = 4$ RF chains and uses an analog precoder whose columns are selected from a transmit beamforming codebook. The transmit beamforming codebook comprises of 16 equal norm constant-magnitude vectors that are also mutually orthogonal. The analog precoder matrix employed by each BS remains fixed for the frame duration, where each frame comprises for several subframes. The mmWave channel was emulated by building upon the freely available NYU simulator based on [19] and PHY parameters such as subframe duration, number of OFDM symbols in a subframe etc. were chosen as per the Verizon 5G standard. On the other hand, for convenience we set each user to be equipped with one receive antenna so that the receive beamforming codebook becomes degenerate, i.e., $\mathcal{W} = \{1\}$ and each user can be assigned only one stream (thus one virtual user per actual user). We considered the following two ways of choosing the analog precoder for each BS and associating users with it.

- The baseline approach is to associate each user to the BS from which it sees the strongest average receive power under the omni pattern. Then each BS is assigned all 16 columns from the transmit precoding codebook.
- In the second approach user association and analog transmit precoder optimization is done in an alternating manner. The details of this method are beyond the scope of this paper. Each BS is assigned at-least 4 columns from the transmit precoding codebook, so that $M \geq S_t = 4$ is ensured for each BS.

Once the user association and the analog precoder choice for each BS is obtained, the fine (per sub-frame) scheduling (which is of primary interest in this paper) is done. The fine scheduling seeks to optimize a weighted sum rate in each subframe, where the weights are updated to optimize the PF utility over the frame duration. We considered the following options for the sub-frame scheduling:

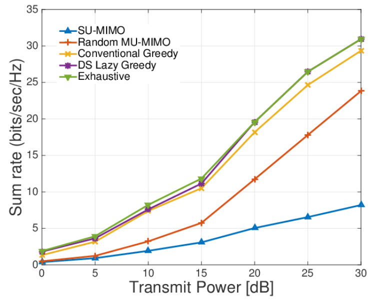

Figure 1: Sum rate vs Transmit power for fixed number of users, RF chains and antennas

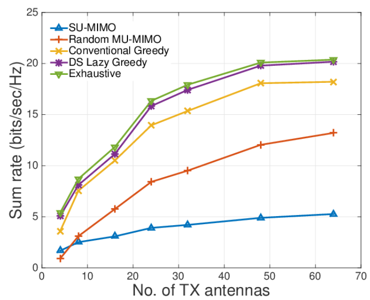

Figure 2: Sum rate vs Transmit antennas for fixed number of users, RF chains and transmit power

- Single-user scheduling: Only implemented for the baseline approach for analog transmit precoder and user association choice. Here at most one user is scheduled on each RB.
- Random user selection: Only implemented for the baseline approach for analog transmit precoder and user association choice. Here 4 users were randomly chosen on each RB.
- Proposed algorithm using the DS framework (with the lazy greedy efficient implementation): Implemented for the baseline approach as well as optimized analog transmit precoder and user association.
- Exhaustive search: Only implemented for the optimized analog transmit precoder and user association.

Our results are shown in Fig. 3. Comparing the first three schemes with baseline approach for analog transmit precoder and user association choice, we note that a significant gain is obtained by our proposed algorithm (which is referred to as Lazy Greedy for this baseline approach) over the single-user and random scheduling. Moving on to the optimized analog transmit precoder and user association, we see that our proposed scheme achieve about 90% of the exhaustive search based one. Here we note that the subframe duration is very small (0.2ms) which makes the exhaustive search based subframe scheduling impractical even with a few users.

VII. Conclusions

In this paper we proposed a new framework for designing MU-MIMO scheduling algorithms. We established that our framework can incorporate practically important choices of linear transmit precoding as well as power allocation. In addition, it can incorporate analog receive beamforming as well. We used this framework to design an algorithm and demonstrated that significant gains can be achieved at a reasonable complexity.

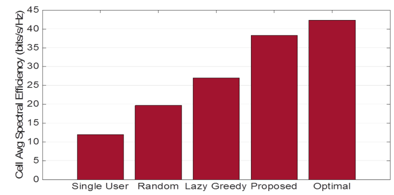

Figure 3: Comparison of Scheduling Algorithms

Appendix

Definition 1. *Let Ω be a ground set and $h : 2^{\Omega} \to \mathbb{R}$ be a real-valued set function defined on the subsets of Ω. The set function $h(.)$ is a* submodular *set function over Ω if it satisfies,*

$$h(\mathcal{B} \cup a) - h(\mathcal{B}) \le h(\mathcal{A} \cup a) - h(\mathcal{A}), \quad \forall \mathcal{A} \subseteq \mathcal{B} \subseteq \Omega \;\&\; a \in \Omega \setminus \mathcal{B}. \tag{21}$$

Definition 2. *$(\Omega, \underline{I})$, where $\underline{I}$ is collection of some subsets of Ω, is said to be a* matroid *if*

- *$\underline{I}$ is downward closed, i.e., $\mathcal{A} \in \underline{I} \;\&\; \mathcal{B} \subseteq \mathcal{A} \Rightarrow \mathcal{B} \in \underline{I}$*
- *For any two members $\mathcal{F}_1 \in \underline{I}$ and $\mathcal{F}_2 \in \underline{I}$ such that $|\mathcal{F}_1| < |\mathcal{F}_2|$, there exists $e \in \mathcal{F}_2 \setminus \mathcal{F}_1$ such that $\mathcal{F}_1 \cup \{e\} \in \underline{I}$. This property is referred to as the exchange property.*

Definition 3. *Let $\underline{\mathcal{I}}$ be any family of subsets of Ω that is downward closed. We say that a real-valued set function $h : \underline{\mathcal{I}} \to \mathbb{R}$ is submodular over $\underline{\mathcal{I}}$, if it satisfies (21) for each choice of $\mathcal{A}, \mathcal{B}, a \in \underline{\mathcal{I}} : \mathcal{A} \subseteq \mathcal{B} \;\&\; a \notin \mathcal{B}$ with $\mathcal{B} \cup a \in \underline{\mathcal{I}}$.*

Lemma 1. *Consider any $N \times N$ positive definite matrix $\mathbf{M}$ and let $\mathbf{M}_{\mathcal{S}}, \; \forall \; \mathcal{S} \subseteq \Omega = \{1, \cdots, N\}$, denote the principal submatrix of $\mathbf{M}$ with row and column indices drawn from $\mathcal{S}$. Then, the set function defined as $h(\mathcal{S}) = \log |\mathbf{M}_{\mathcal{S}}|, \; \forall \; \mathcal{S} \subseteq \Omega$ is a submodular set function over Ω. Thus, for any $j \in \Omega$, the set function defined as $h_j(\mathcal{S}) = \log |\mathbf{M}_{\mathcal{S}\setminus j}|, \; \forall \; \mathcal{S} \subseteq \Omega$ is also a submodular set function over Ω.*

Lemma 2. *Consider any choice of co-scheduled virtual users $\mathcal{A} \subseteq \Psi$ and any virtual user $\psi \in \mathcal{A}$. Define the matrix $\mathbf{Z}_{\mathcal{A}} = [\mathbf{z}_{\psi}]_{\psi \in \mathcal{A}}$ along with $\mathbf{Z}_{\mathcal{A}\setminus\psi} = [\mathbf{z}_{\psi'}]_{\psi' \in \mathcal{A}\setminus\psi}$. Further, define diagonal matrices $\mathbf{E}_{\mathcal{A}} = \mathrm{diag}\{e_{\psi'}\}_{\psi' \in \mathcal{A}}$ and $\mathbf{E}_{\mathcal{A}\setminus\psi} = \mathrm{diag}\{e_{\psi'}\}_{\psi' \in \mathcal{A}\setminus\psi}$. Then, we have,*

$$|\mathbf{E}_{\mathcal{A}} + \mathbf{Z}_{\mathcal{A}}^{\dagger}\mathbf{Z}_{\mathcal{A}}| = |\mathbf{E}_{\mathcal{A}\setminus\psi} + \mathbf{Z}_{\mathcal{A}\setminus\psi}^{\dagger}\mathbf{Z}_{\mathcal{A}\setminus\psi}| \times (e_{\psi} + \|\mathbf{z}_{\psi}\|^2 - \mathbf{z}_{\psi}^{\dagger}\mathbf{Z}_{\mathcal{A}\setminus\psi}(\mathbf{E}_{\mathcal{A}\setminus\psi} + \mathbf{Z}_{\mathcal{A}\setminus\psi}^{\dagger}\mathbf{Z}_{\mathcal{A}\setminus\psi})^{-1}\mathbf{Z}_{\mathcal{A}\setminus\psi}^{\dagger}\mathbf{z}_{\psi}) \tag{22}$$

Note that when $\mathbf{E}_{\mathcal{A}\setminus\psi} = \mathbf{0}$ then,

$$|\mathbf{E}_{\mathcal{A}} + \mathbf{Z}_{\mathcal{A}}^{\dagger}\mathbf{Z}_{\mathcal{A}}| = |\mathbf{Z}_{\mathcal{A}\setminus\psi}^{\dagger}\mathbf{Z}_{\mathcal{A}\setminus\psi}|(e_{\psi} + \mathrm{Res}(\psi, \mathcal{A} \setminus \psi))$$

where $\mathrm{Res}(\psi, \mathcal{A} \setminus \psi) = \|\mathbf{z}_{\psi}\|^2 - \mathbf{z}_{\psi}^{\dagger}\mathbf{Z}_{\mathcal{A}\setminus\psi}(\mathbf{Z}_{\mathcal{A}\setminus\psi}^{\dagger}\mathbf{Z}_{\mathcal{A}\setminus\psi})^{-1}\mathbf{Z}_{\mathcal{A}\setminus\psi}^{\dagger}\mathbf{z}_{\psi}$

Lemma 3. *We collect a few facts that follow after some algebra.*

- *The real-valued functions* $-x\log(x), \forall\ x \geq 0$ *and* $-x\log(x+1), \forall\ x \geq 0$ *are both concave in* x *for all* $x \geq 0$.
- *For any fixed* $a \geq 0$*, the real-valued function* $-(a+1)\log(a+x+1) + a\log(a+x), \forall\ x \geq 0$ *is decreasing in* x *for all* $x \geq 0$.
- *The real-valued function* $-x\log(x+1)+x\log(x), \forall\ x \geq 0$ *is decreasing in* x *for all* $x \geq 1$.

Proof of Proposition 1

Note first that the rate expression in (7) satisfies $R_\psi(\mathcal{A}) = 0,\ \forall\ \psi \notin \mathcal{A}$. Further, for each $\psi \in \mathcal{A}$ it can be readily verified that (7) follows upon expressing the RHS of (6) in a different form. We will prove that the first term $f_\psi^{\mathrm{MRT}} : 2^\Psi \to I\!R$ in the RHS of (7) is a submodular set function over Ψ, for each $\psi \in \Psi$. The second term $g_\psi^{\mathrm{MRT}}(.)$ can be shown to be submodular in an analogous manner [14]. We invoke the following property of the logarithm function,

$$\log(c+e) - \log(c) \leq \log(d+f) - \log(d), \qquad (23)$$
$$\forall\ 0 < d \leq c\ \&\ f \geq e \geq 0.$$

The above property follows from the monotonicity and concavity of the logarithm function. Considering any $\mathcal{E} \subseteq \mathcal{F} \subseteq \Psi : \mathcal{E} \neq \phi$ and any $\psi'' \in \Psi \setminus \mathcal{F}$, we define

$$\begin{aligned} e = f &= \|\mathbf{z}_{\psi''}\|^2 + \rho\mathbf{z}_\psi^\dagger \mathbf{z}_{\psi''}\mathbf{z}_{\psi''}^\dagger \mathbf{z}_\psi, \\ d &= \|\mathbf{Z}_\mathcal{E}\|^2 + \sum_{\psi' \in \mathcal{E}} \rho\mathbf{z}_\psi^\dagger \mathbf{z}_{\psi'}\mathbf{z}_{\psi'}^\dagger \mathbf{z}_\psi, \\ c &= \|\mathbf{Z}_\mathcal{F}\|^2 + \sum_{\psi' \in \mathcal{F}} \rho\mathbf{z}_\psi^\dagger \mathbf{z}_{\psi'}\mathbf{z}_{\psi'}^\dagger \mathbf{z}_\psi \end{aligned} \qquad (24)$$

Note that the scalars so defined satisfy $d \leq c$ and $f \geq e$ so that we can invoke (23) with this choice to verify that the required condition in (21) is satisfied. Now consider the case $\mathcal{E} = \phi$. Clearly, when $\mathcal{F} = \phi$ the required condition is trivially satisfied. Hence, suppose that $\mathcal{F} \neq \phi$ and define the scalars $c, e\ \&\ f$ as in (24). To prove that (21) indeed holds we need to show that

$$\log(c+e) - \log(c) \leq \log(e) - f_\psi^{\mathrm{MRT}}(\phi) \qquad (25)$$

Note that since $c \geq \epsilon$ the LHS in (25) is clearly no greater than $\log(1 + e/\epsilon)$. Therefore, (21) holds if we can show that $-f_\psi^{\mathrm{MRT}}(\phi) \geq \log(1/e + 1/\epsilon)$. Since $e \geq \epsilon$, the latter inequality is true for our choice $f_\psi^{\mathrm{MRT}}(\phi) = -\log(2/\epsilon)$.

Proof of Proposition 2

We first consider the case $\mathcal{A} \in \underline{\mathcal{I}}$ with $\psi \in \mathcal{A}$. Here, we can write (8) as

$$\begin{aligned} R_\psi(\mathcal{A}) = \log\big(&|\mathcal{A}| + \rho\|\mathbf{z}_\psi\|^2 - \\ &\rho\mathbf{z}_\psi^\dagger \mathbf{Z}_{\mathcal{A}\setminus\psi}(\mathbf{Z}_{\mathcal{A}\setminus\psi}^\dagger \mathbf{Z}_{\mathcal{A}\setminus\psi})^{-1}\mathbf{Z}_{\mathcal{A}\setminus\psi}^\dagger \mathbf{z}_\psi\big) - \log|\mathcal{A}| \end{aligned} \qquad (26)$$

Invoking Lemma 2 we can re-write the RHS of (26) to obtain

$$\begin{aligned} R_\psi(\mathcal{A}) = \log|\mathbf{C}_\mathcal{A}(|\mathcal{A}|, \psi)| - \log|\mathcal{A}| - \\ \log\left|\mathbf{C}_{\mathcal{A}\setminus\psi}(|\mathcal{A}|, \psi)\right| \end{aligned} \qquad (27)$$

Then, since $\mathbf{B}_{\mathcal{A}\setminus\psi} = \mathbf{C}_{\mathcal{A}\setminus\psi}(|\mathcal{A}|, \psi)$ and $\log|\mathcal{A}| = |\mathcal{A}|\log|\mathcal{A}| - (|\mathcal{A}|-1)\log|\mathcal{A}|$, we can deduce that (11) holds. On the other hand whenever $\psi \notin \mathcal{A}$, we can verify that (11) yields $R_\psi(\mathcal{A}) = 0$ which is consistent.

We proceed to prove the submodularity of $g_\psi^{\mathrm{ZF}}(.)$ for each $\psi \in \Psi$ over $\underline{\mathcal{I}}$ first. Towards this end we arbitrarily pick any $\psi \in \Psi$ and consider each one of the two terms whose sum gives $g_\psi^{\mathrm{ZF}}(.)$. Considering the first term, if we define $h(\mathcal{A}) = \log\left|\mathbf{B}_{\mathcal{A}\setminus\psi}\right|,\ \forall \mathcal{A} \subseteq \Psi$, then this set function can be verified to be submodular over $\underline{\mathcal{I}}$ upon invoking Lemma 1. Now for the second term we define $h(\mathcal{A}) = -|\mathcal{A}\setminus\psi|\log|\mathcal{A}|,\ \forall \mathcal{A} \subseteq \Psi$. We will show that this set function can be verified to be submodular over Ψ (and hence over $\underline{\mathcal{I}}$). Consider any $\mathcal{E} \subseteq \mathcal{F} \in \Psi$ with any $\psi'' \in \Psi \setminus \mathcal{F}$. To establish submodularity when $\psi \notin \mathcal{F}$ (so that $\psi \notin \mathcal{E}$) and $\psi'' \neq \psi$, we need to show that

$$\begin{aligned} &-(|\mathcal{E}|+1)\log(|\mathcal{E}|+1) + |\mathcal{E}|\log(|\mathcal{E}|) \\ &\quad \geq -(|\mathcal{F}|+1)\log(|\mathcal{F}|+1) + |\mathcal{F}|\log(|\mathcal{F}|) \end{aligned} \qquad (28)$$

(28) holds due to the concavity of $-x\log(x)$ for all $x \geq 0$ stated as the first fact in Lemma 3. Further, when $\psi \notin \mathcal{F}$ but $\psi'' = \psi$, we need to show that

$$\begin{aligned} &-(|\mathcal{E}|)\log(|\mathcal{E}|+1) + |\mathcal{E}|\log(|\mathcal{E}|) \\ &\quad \geq -(|\mathcal{F}|)\log(|\mathcal{F}|+1) + |\mathcal{F}|\log(|\mathcal{F}|) \end{aligned} \qquad (29)$$

(29) follows from the third fact stated in Lemma 3. Next, when $\psi \in \mathcal{E}$ (so that $\psi \in \mathcal{F}$) and $\psi'' \neq \psi$, we need to show that

$$\begin{aligned} &-(|\mathcal{E}|)\log(|\mathcal{E}|+1) + (|\mathcal{E}|-1)\log(|\mathcal{E}|) \\ &\quad \geq -(|\mathcal{F}|)\log(|\mathcal{F}|+1) + (|\mathcal{F}|-1)\log(|\mathcal{F}|) \end{aligned} \qquad (30)$$

(30) holds due to the concavity of $-x\log(x+1)$ for all $x \geq 0$ stated as the first fact in Lemma 3. Finally, when $\psi \notin \mathcal{E}$ but $\psi \in \mathcal{F}$ and $\psi'' \neq \psi$, we need to show that

$$\begin{aligned} &-(|\mathcal{E}|+1)\log(|\mathcal{E}|+1) + (|\mathcal{E}|)\log(|\mathcal{E}|) \\ &\quad \geq -(|\mathcal{F}|)\log(|\mathcal{F}|+1) + (|\mathcal{F}|-1)\log(|\mathcal{F}|) \end{aligned} \qquad (31)$$

(29) follows by first using the concavity of $-x\log(x+1)$ for all $x \geq 0$ to deduce

$$\begin{aligned} &-(|\mathcal{F}|)\log(|\mathcal{F}|+1) + (|\mathcal{F}|-1)\log(|\mathcal{F}|) \\ &\quad \leq -(|\mathcal{E}|+1)\log(|\mathcal{E}|+2) + (|\mathcal{E}|)\log(|\mathcal{E}|+1) \end{aligned}$$

and then using the second fact stated in Lemma 3 to confirm that

$$\begin{aligned} &-(|\mathcal{E}|+1)\log(|\mathcal{E}|+2) + (|\mathcal{E}|)\log(|\mathcal{E}|+1) \\ &\quad \leq -(|\mathcal{E}|+1)\log(|\mathcal{E}|+1) + (|\mathcal{E}|)\log(|\mathcal{E}|). \end{aligned}$$

In summary since $g_\psi^{\mathrm{ZF}}(.)$ is the sum of two terms that are each submodular over $\underline{\mathcal{I}}$, we can confirm that $g_\psi^{\mathrm{ZF}}(.)$ is submodular over $\underline{\mathcal{I}}$.

Now we embark upon the more involved part of proving the submodularity of $f_\psi^{\mathrm{ZF}}(.)$ over $\underline{\mathcal{I}}$, for any given ψ. For convenience in notation and without loss of generality, we set $\rho = 1$. Here although as before $f_\psi^{\mathrm{ZF}}(.)$ is the sum of two terms, we have to consider both the terms in $f_\psi^{\mathrm{ZF}}(.)$ together. This is because the first term in $f_{\psi}^{\mathrm{ZF}}(.)$ need not be submodular. However, as we show below, the second term in $f_{\psi}^{\mathrm{ZF}}(.)$ adequately compensates and makes the sum submodular. Let us define a set function $g(\mathcal{A}) = -(|\mathcal{A}|+1)\log(|\mathcal{A}|+1) + |\mathcal{A}|\log|\mathcal{A}|, \ \forall\ \mathcal{A} \subseteq \Psi$. We consider any $\mathcal{E} \subseteq \mathcal{F} \in \underline{\mathcal{I}}$ with any $\psi'' \in \Psi \setminus \mathcal{F} : \mathcal{F} \cup \psi'' \in \underline{\mathcal{I}}$. Further, it suffices to consider $\mathcal{F} : |\mathcal{F}| = |\mathcal{E}|+1$. Then, we proceed to systematically analyze the first (out of the four possible) case which is the hardest and whose proof captures all the key techniques needed. The proofs of the other three cases are given in [14].

Case I: $\psi \in \mathcal{E}$: Here, we must have $\psi \in \mathcal{F}$ and $\psi'' \neq \psi$. Then, we can expand $\Delta_{\mathcal{F},\psi''} \triangleq f_{\psi}^{\mathrm{ZF}}(\mathcal{F} \cup \psi'') - f_{\psi}^{\mathrm{ZF}}(\mathcal{F})$ using Lemma 2 as

$$\begin{aligned}\Delta_{\mathcal{F},\psi''} = \log(|\mathcal{F}| + 1 + \mathrm{Res}(\psi, \mathcal{F} \setminus \psi \cup \psi'')) - \\ \log(|\mathcal{F}| + \mathrm{Res}(\psi, \mathcal{F} \setminus \psi)) \\ + \log(\mathrm{Res}(\psi'', \mathcal{F} \setminus \psi)) + g(|\mathcal{F}|)\end{aligned} \tag{32}$$

We will add and subtract a term and write $\Delta_{\mathcal{F},\psi''}$ as

$$\begin{aligned}\Delta_{\mathcal{F},\psi''} = \log(|\mathcal{F}| + 1 + \mathrm{Res}(\psi, \mathcal{F} \setminus \psi \cup \psi'')) - \\ \log(|\mathcal{F}| + \mathrm{Res}(\psi, \mathcal{F} \setminus \psi)) \\ + \log(\mathrm{Res}(\psi'', \mathcal{F} \setminus \psi)) + g(|\mathcal{F}|) \\ + \log(|\mathcal{F}| + \mathrm{Res}(\psi, \mathcal{F} \setminus \psi \cup \psi'')) - \\ \log(|\mathcal{F}| + \mathrm{Res}(\psi, \mathcal{F} \setminus \psi \cup \psi''))\end{aligned} \tag{33}$$

Similarly we express $\Delta_{\mathcal{E},\psi''} \triangleq f_{\psi}^{\mathrm{ZF}}(\mathcal{E} \cup \psi'') - f_{\psi}^{\mathrm{ZF}}(\mathcal{E})$ as

$$\begin{aligned}\Delta_{\mathcal{E},\psi''} = \log(|\mathcal{E}| + 1 + \mathrm{Res}(\psi, \mathcal{E} \setminus \psi \cup \psi'')) - \\ \log(|\mathcal{E}| + \mathrm{Res}(\psi, \mathcal{E} \setminus \psi)) + \log(\mathrm{Res}(\psi'', \mathcal{E} \setminus \psi)) + g(|\mathcal{E}|) \\ + \log(|\mathcal{F}| + \mathrm{Res}(\psi, \mathcal{E} \setminus \psi)) - \log(|\mathcal{F}| + \mathrm{Res}(\psi, \mathcal{E} \setminus \psi))\end{aligned} \tag{34}$$

Now, a key observation using Lemma 1, Lemma 2and the fact that $|\mathcal{F}| = |\mathcal{E}| + 1$ is that

$$\begin{aligned}\log(|\mathcal{E}| + 1 + \mathrm{Res}(\psi, \mathcal{E} \setminus \psi \cup \psi'')) + \\ \log(\mathrm{Res}(\psi'', \mathcal{E} \setminus \psi)) - \log(|\mathcal{F}| + \mathrm{Res}(\psi, \mathcal{E} \setminus \psi)) \geq \\ \log(|\mathcal{F}| + \mathrm{Res}(\psi, \mathcal{F} \setminus \psi \cup \psi'')) + \log(\mathrm{Res}(\psi'', \mathcal{F} \setminus \psi)) \\ - \log(|\mathcal{F}| + \mathrm{Res}(\psi, \mathcal{F} \setminus \psi))\end{aligned} \tag{35}$$

In particular, to derive (35) we invoke Lemma 1 on the matrix $\mathbf{M} = \mathbf{Z}_{\mathcal{F}\cup\psi''}^{\dagger}\mathbf{Z}_{\mathcal{F}\cup\psi''} + \mathbf{E}_{\mathcal{F}\cup\psi''}$, where $\mathbf{E}_{\mathcal{F}\cup\psi''}$ is a diagonal matrix with $|\mathcal{E}|+1$ on the diagonal position corresponding to ψ and zeros elsewhere. Then, using (35) with (34) and (33), to prove submodularity, i.e., $\Delta_{\mathcal{F},\psi''} \leq \Delta_{\mathcal{E},\psi''}$, it suffices to show that

$$\begin{aligned}\log(|\mathcal{F}| + 1 + \mathrm{Res}(\psi, \mathcal{F} \setminus \psi \cup \psi'')) - \\ \log(|\mathcal{F}| + \mathrm{Res}(\psi, \mathcal{F} \setminus \psi \cup \psi'')) + g(|\mathcal{F}|) \leq \\ \log(|\mathcal{E}| + 1 + \mathrm{Res}(\psi, \mathcal{E} \setminus \psi)) \\ - \log(|\mathcal{E}| + \mathrm{Res}(\psi, \mathcal{E} \setminus \psi)) + g(|\mathcal{E}|)\end{aligned} \tag{36}$$

Then, since $\log(|\mathcal{E}|+1+\mathrm{Res}(\psi,\mathcal{E}\setminus\psi)) - \log(|\mathcal{E}|+\mathrm{Res}(\psi,\mathcal{E}\setminus\psi)) \geq 0$, it suffices to show that

$$\begin{aligned}\log(|\mathcal{F}| + 1 + \mathrm{Res}(\psi, \mathcal{F} \setminus \psi \cup \psi'')) - \\ \log(|\mathcal{F}| + \mathrm{Res}(\psi, \mathcal{F} \setminus \psi \cup \psi'')) + g(|\mathcal{F}|) \leq g(|\mathcal{E}|)\end{aligned} \tag{37}$$

To show (37) we exploit the concavity of the logarithm function and the non-negativity of the residual to deduce the fact that

$$\begin{aligned}\log(|\mathcal{F}| + 1 + \mathrm{Res}(\psi, \mathcal{F} \setminus \psi \cup \psi'')) - \\ \log(|\mathcal{F}| + \mathrm{Res}(\psi, \mathcal{F} \setminus \psi \cup \psi'')) \leq \\ \log(|\mathcal{F}| + 1) - \log(|\mathcal{F}|)\end{aligned} \tag{38}$$

Using (38) in (37) and recalling that $|\mathcal{F}| = |\mathcal{E}| + 1$, we can see that to establish submodularity in this case, it is enough to show that

$$\begin{aligned}-(|\mathcal{E}|+1)|\log(|\mathcal{E}|+2) + |\mathcal{E}|\log(|\mathcal{E}|+1) \leq \\ -(|\mathcal{E}|+1)|\log(|\mathcal{E}|+1) + |\mathcal{E}|\log(|\mathcal{E}|)\end{aligned} \tag{39}$$

Finally, (39) holds true from the second fact stated in Lemma 3.

References

[1] G. Caire and S. Shamai, "On the achievable throughput of a multiantenna gaussian broadcast channel," *IEEE Transactions on Information Theory*, 2003.

[2] T. Yoo and A. Goldsmith, "On the optimality of multiantenna broadcast scheduling using zero-forcing beamforming," *IEEE JSAC*, 2006.

[3] J. Wang, D. J. Love, and M. D. Zoltowski, "User selection with zeroforcing beamforming achieves the asymptotically optimal sum rate," *IEEE Transactions on Signal Processing*, 2008.

[4] N. Ravindran and N. Jindal, "Limited feedback-based block diagonalization for the mimo broadcast channel," *IEEE JSAC*, 2008.

[5] K. Hosseini, W. Yu, and R. S. Adve, "Optimizing large-scale mimo cellular downlink: Multiplexing, diversity, or interference nulling?" in *Allerton*, 2015.

[6] R. Arnott, K. Oketani, N. Prasad, S. Rangarajan, and P. Wells, "Analysis and evaluation of a practical downlink multiuser mimo scheduler over lte advanced massive mimo systems," in *IEEE Asilomar*, 2016.

[7] G. Dimic and N. Sidiropoulos, "On downlink beamforming with greedy user selection: performance analysis and a simple new algorithm," *IEEE Transactions on Signal Processing*, Oct. 2005.

[8] G. Yue, N. Prasad, and S. Rangarajan, "Downlink multiuser mimo scheduling in LTE," *IEEE ICC*, 2012.

[9] W. Shen, K. C. Lin, M. Chen, and K. Tan, "Sieve: Scalable user grouping for large mu-mimo systems," in *IEEE Infocom*, 2015.

[10] G. Zois, A. Michaloliakos, K. Psounis, V. Vassalos, and I. Mourtos, "Non-asymptotic performance bounds for downlink MU-MIMO scheduling," in *13th Annual Conference on Wireless On-demand Network Systems and Services WONS*, 2016.

[11] R. Iyer and J. Bilmes, "Algorithms for approximate minimization of the difference between submodular functions, with applications," in *Proc. Uncertainty in Artificial Intelligence (UAI)*, 2012.

[12] S. You, L. Chen, and Y. Liu, "Convex-concave procedure for weighted sum-rate maximization in a mimo interference network," in *Proc. IEEE Globecom*, 2014.

[13] A. Alkhateeb, G. Leus, and R. Heath, "Limited feedback hybrid precoding for multi-user millimeter wave systems," *IEEE Trans. Wireless Commun. (revised)*, Mar. 2015.

[14] Y. Ghasempour, N. Prasad, M. Khojastepour, and S. Rangarajan, "Novel combinatorial results on downlink mu-mimo scheduling with applications," *Extended Version, Url: http://tinyurl.com/joe79wz.*

[15] N. Prasad, M. Arslan, and S. Rangarajan, "A two time scale approach for coordinated multi-point transmission and reception over practical backhaul," in *IEEE Comsnets (invited)*, 2014.

[16] Y. Lim, C. Chae, and G. Caire, "Performance analysis of massive mimo for cell-boundary users," *IEEE Trans. Wireless. Comm.*, Dec. 2015.

[17] G. L. Nemhauser and L. A. Wolsey, "Best algorithms for approximating the maximum of a submodular set function," *Math. Operations Research*, 1978.

[18] M. Minoux, "Accelerated greedy algorithms for maximizing submodular set functions," *Optimization Techniques, LNCS*, 1978.

[19] M. Samimi and T. Rappaport, "3-d millimeter-wave statistical channel model for 5g wireless system design," *IEEE Trans. on Microwave Theory and Techniques*, Jul. 2016.