Rate-balancing in Massive MIMO using Statistical Precoding

David Neumann, Andreas Gründinger, Michael Joham, and Wolfgang Utschick
Associate Institute for Signal Processing, Technische Universität München, 80290 Munich, Germany
{d.neumann,gründinger,joham,utschick}@tum.de

***Abstract*—In a typical massive MIMO system, the limited coherence time of the wireless channels leads to interference during the uplink training phase, since pilot sequences have to be reused between different users. This interference ultimately limits the achievable rate when basic linear beamforming is used. We show how to push this limit by exploiting the statistical properties of the channel. Specifically, we use a lower bound on the achievable rate to formulate a rate balancing problem which is independent of the instantaneous channel state information but only depends on the second order information. The rate balancing problem is then solved by methods known from classical MIMO systems leading to beamforming vectors and power allocations which significantly outperform the standard matched filter or zero-forcing approaches.**

I. Introduction

Massive MIMO is a promising technology for the next generation of cellular wireless networks [1]–[3]. The idea is basically to deploy such a large number of antennas at each base station that the number of antennas is at least an order of magnitude larger than the number of simultaneously served users. The large number of antennas has several advantages. The array gain leads to an increased energy efficiency. Even more importantly, for a typical wireless channel, the large number of antennas leads to approximately orthogonal channel vectors for two different users due to the law of large numbers (e.g. [4]). These properties enable robust spatial multiplexing with simple signal processing methods [1].

In other words, to fully exploit the potentials of a massive MIMO system, we want to serve serve multiple users simultaneously, relying on simple spatial multiplexing. As a result, a significant number of interfering data streams is transmitted simultaneously throughout the massive MIMO network. For perfect channel state information (CSI), the interference is negligible due to the afore mentioned asymptotic orthogonality of different channel vectors [4]. However, for a block fading channel model with limited coherence time and frequency, the limited number of available channel accesses per channel realization results in a dimensionality bottleneck. In fact, for a simple channel model where all channel coefficients are i.i.d. complex Gaussian, it can be shown that for sufficiently large numbers of users and antennas, the coherence time poses an upper limit on the achievable degrees of freedom [5]. In this case, the optimal amount of channel accesses used for training is half of the coherence block [5].

This dimensionality bottleneck gives rise to the pilot contamination effect which is observed in the classical massive MIMO setup [4]. Pilot contamination describes the interference in the channel estimates obtained in the uplink that is caused by the fact that the number of channel accesses in one coherence block is too small to give every user an orthogonal training sequence. The interference in the channel estimates in turn leads to interference during data transmission, which ultimately limits the performance of a massive MIMO system even for an unlimited number of antennas at each base station [4].

This is not the final word on massive MIMO, however, since the upper bound for the degrees of freedom in [5] does not take the channel structure into account. In [6], a precoding technique is introduced that is capable of completely suppressing the interference during data transmission for an unlimited number of base station antennas. The precoding is based on a linear transformation of the contaminated channel estimates, where the transformation only depends on the channel covariance matrices of the different users.

In a practical setting, with a large but not infinite number of antennas, this statistical zero-forcing might lead to suboptimal results, since inter-user interference that is not related to pilot-contamination cannot be neglected in this case. Thus, we propose a precoder design which is also based on a linear transformation of the contaminated channel estimates, but where the transformations are optimized for a finite number of antennas with a rate balancing approach. The optimization is performed with respect to a lower bound on the achievable rate based on the bound introduced by Medard [7], which only depends on the covariance matrices of the channels. In principle, the proposed algorithm is a generalization of the pilot-contamination precoding method introduced in [8] which was extended to SINR balancing in [9]. Notably, our method is also applicable in an uncoordinated setting, whereas pilot-contamination precoding is applicable in a network MIMO setting only.

A. Notation

Throughout this paper, $(\cdot)^{\mathrm{H}}$ denotes the conjugate transpose of a matrix and $\mathrm{E}[\cdot]$ denotes the expectation. The covariance matrix of a vector $\boldsymbol{x}$ is denoted as

$$\boldsymbol{C}_{\boldsymbol{x}} = \mathrm{E}[\boldsymbol{x}^{\mathrm{H}}\boldsymbol{x}]. \tag{1}$$

II. System Model

For the ease of exposition and notation, we only consider a single-cell scenario, but the extension to the multi-cell case is straightforward (see [6]). We exploit the reciprocity of the wireless system in time-division-duplex (TDD) operation by using the channel estimate obtained during the uplink training phase to design the downlink precoder. Let $\boldsymbol{h}_k \in \mathbb{C}^M$ denote the channel vector from user k to the base station and $\boldsymbol{\psi}_k^{\mathrm{H}} \in \mathbb{C}^{1\times T_{\mathrm{tr}}}$ the unit-norm pilot sequence assigned to user k. The number of channel accesses used for training T_{tr} is assumed to be smaller than the number of simultaneously served users K.

The least squares estimate of $\boldsymbol{h}_k$ is then given by

$$\boldsymbol{y}_k = \boldsymbol{h}_k + \sum_{n\neq k} \boldsymbol{h}_n \boldsymbol{\psi}_n^{\mathrm{H}} \boldsymbol{\psi}_k + \frac{1}{\sqrt{\rho_{\mathrm{tr}}}} \boldsymbol{n}_k \tag{2}$$

$$= \boldsymbol{h}_k + \sum_{n\neq k} \boldsymbol{h}_n c_{nk} + \frac{1}{\sqrt{\rho_{\mathrm{tr}}}} \boldsymbol{n}_k \tag{3}$$

where the coefficients $c_{nk} = \boldsymbol{\psi}_n^{\mathrm{H}} \boldsymbol{\psi}_k$ capture the correlations between the pilot sequences, $\boldsymbol{n}_k \sim \mathcal{N}_{\mathbb{C}}(\boldsymbol{0}, \mathbf{I})$ is additive white Gaussian noise and ρ_{tr} denotes the effective training SNR.

Given the decomposition

$$\boldsymbol{C}_{\boldsymbol{y}_k} = \mathrm{E}[\boldsymbol{y}_k \boldsymbol{y}_k^{\mathrm{H}}] = \boldsymbol{L}_k \boldsymbol{L}_k^{\mathrm{H}} \tag{4}$$

the beamforming vector for user k is of the form

$$\boldsymbol{w}_k = \boldsymbol{A}_k \boldsymbol{L}_k^{-1} \boldsymbol{y}_k \tag{5}$$

for some linear transformation $\boldsymbol{A}_k$ that only depends on the statistics of the channel. The whitening $\boldsymbol{L}_k^{-1}$ is introduced to make the derivations below less cumbersome. Note that the precoder structure in (5) is a generalization of several previously introduced approaches, such as minimum mean squared error (MMSE) channel estimation [10], [11], pilot-contamination precoding [8] or projection based approaches (e.g. [12]). In contrast to these typically heuristic approaches, we will derive an algorithm that leads to optimal transformations $\boldsymbol{A}_k$ with respect to a rate balancing problem.

With the independent transmit symbols $s_k \sim \mathcal{N}_{\mathbb{C}}(0,1)$ for the users, the transmit signal for the downlink is given by

$$\boldsymbol{x} = \sqrt{\rho_{\mathrm{dl}}} \sum_k \boldsymbol{w}_k s_k. \tag{6}$$

where ρ_{dl} denotes the maximum average transmit power. Consequently, the beamforming vectors are constrained by

$$\frac{\mathrm{E}[\boldsymbol{x}^{\mathrm{H}}\boldsymbol{x}]}{\rho_{\mathrm{dl}}} = \sum_k \mathrm{E}[\boldsymbol{w}_k^{\mathrm{H}} \boldsymbol{w}_k] = \sum_k \mathrm{tr}[\boldsymbol{A}_k \boldsymbol{A}_k^{\mathrm{H}}] \leq 1. \tag{7}$$

To design the transformation matrices, we use the lower bound on the achievable rate introduced in [7], which leads to the achievable rate

$$r_k = \log_2(1+\gamma_k) \tag{8}$$

for user k, with the equivalent SINR

$$\gamma_k = \frac{|\mathrm{E}[\boldsymbol{h}_k^{\mathrm{H}} \boldsymbol{w}_k]|^2}{\frac{1}{\rho_{\mathrm{dl}}} + \mathrm{var}[\boldsymbol{h}_k^{\mathrm{H}} \boldsymbol{w}_k] + \sum_{n\neq k} \mathrm{E}[|\boldsymbol{h}_k^{\mathrm{H}} \boldsymbol{w}_n|^2]}. \tag{9}$$

Due to the independence of the channels belonging to different users, when using the beamforming vector from (5), and with

$$\boldsymbol{T}_{kn} = \boldsymbol{L}_n^{-1} \boldsymbol{C}_{\boldsymbol{h}_k} c_{nk} \tag{10}$$

the equivalent SINRs can be written as

$$\gamma_k = \frac{\left|\mathrm{tr}[\boldsymbol{T}_{kk}\boldsymbol{A}_k]\right|^2}{\frac{1}{\rho_{\mathrm{dl}}} + \sum_n \mathrm{tr}\left[\boldsymbol{A}_n^{\mathrm{H}} \boldsymbol{C}_{\boldsymbol{h}_k} \boldsymbol{A}_n\right] + \sum_{n\neq k} \left|\mathrm{tr}[\boldsymbol{T}_{kn}\boldsymbol{A}_n]\right|^2}. \tag{11}$$

The derivation for the expectations in the denominator uses higher order moments of the complex Gaussian distribution and can be found in [6].

III. Problem Formulation

The rate balancing problem with a simple sum power constraint is given by

$$\max_{\boldsymbol{A}_1,\ldots,\boldsymbol{A}_K} \beta \ \text{ s.t. } \ \log_2(1+\gamma_k) \geq \beta\tau_k \ \ \forall k, \tag{12}$$

$$\sum_k \mathrm{tr}[\boldsymbol{A}_k \boldsymbol{A}_k^{\mathrm{H}}] \leq 1 \tag{13}$$

for some given rate targets τ_k. For identical rate targets $\tau_k = \tau \ \forall i$, the rate balancing problem degrades to the simpler quasi-convex SINR balancing problem. Either problem can be solved with a bisection on β and the solution of a convex problem for each candidate β within the bisection. This approach is prohibitive in our case due to the high dimensionality of the problem (the $\boldsymbol{A}_k$ are $M \times M$). Another approach is the transformation of the given problem into the dual uplink formulation. To this end, we first write the SINR expression in (11) in vector form. Let $\boldsymbol{a}_k$ denote the vector containing the stacked columns of $\boldsymbol{A}_k$ and let $\boldsymbol{t}_{kn}$ denote the vector containing the stacked columns of $\boldsymbol{T}_{kn}^{\mathrm{H}}$. With the Kronecker product $\boldsymbol{B}_k = \mathbf{I} \otimes \boldsymbol{C}_{\boldsymbol{h}_k}$ the vectorized reformulation of the SINR reads as

$$\gamma_k = \frac{|\boldsymbol{t}_{kk}^{\mathrm{H}} \boldsymbol{a}_k|^2}{\frac{1}{\rho_{\mathrm{dl}}} + \sum_n \boldsymbol{a}_n^{\mathrm{H}} \boldsymbol{B}_k \boldsymbol{a}_n + \sum_{n\neq k} |\boldsymbol{t}_{kn}^{\mathrm{H}} \boldsymbol{a}_n|^2}. \tag{14}$$

The SINR in the dual uplink for user k with filter vector $\boldsymbol{g}_k$ can be identified to be (see e.g., [13], [14])

$$\gamma_k^{\mathrm{UL}} = \frac{q_k |\boldsymbol{t}_{kk}^{\mathrm{H}} \boldsymbol{g}_k|^2}{\boldsymbol{g}_k^{\mathrm{H}} \left(\frac{1}{\rho_{\mathrm{dl}}} \mathbf{I} + \sum_n q_n \boldsymbol{B}_n + \sum_{n\neq k} \boldsymbol{t}_{nk} \boldsymbol{t}_{nk}^{\mathrm{H}} q_n\right) \boldsymbol{g}_k} \tag{15}$$

where the q_n are the uplink transmit powers. Since the uplink performance is independent of the norm of the filters, we define the uplink filters to have unit norm, i.e., $\boldsymbol{g}_k^{\mathrm{H}} \boldsymbol{g}_k = 1$. The rate balancing problem for the uplink is thus given by

$$\max_{\boldsymbol{q},\boldsymbol{g}_1,\ldots,\boldsymbol{g}_K} \beta \ \text{ s.t. } \ \log_2(1+\gamma_k^{\mathrm{UL}}) \geq \beta\tau_k \ \ \forall k, \tag{16}$$

$$\boldsymbol{g}_k^{\mathrm{H}} \boldsymbol{g}_k = 1 \ \ \forall k$$

$$\boldsymbol{q}^{\mathrm{T}} \boldsymbol{q} = 1$$

$$\boldsymbol{q} \geq \boldsymbol{0}$$

where $\boldsymbol{q} = [q_1, \ldots, q_K]^{\mathrm{T}}$. For fixed uplink transmit powers $\boldsymbol{q}$, the optimal filter is given by

$$\boldsymbol{g}_k = \frac{\tilde{\boldsymbol{g}}_k}{\|\tilde{\boldsymbol{g}}_k\|} \tag{17}$$

with

$$\tilde{\boldsymbol{g}}_k = \Big(\frac{1}{\rho_{\mathrm{dl}}}\mathbf{I} + \sum_n q_n \boldsymbol{B}_n + \sum_{n\neq k} \boldsymbol{t}_{nk}\boldsymbol{t}_{nk}^{\mathrm{H}} q_n\Big)^{-1} \boldsymbol{t}_{kk}. \tag{18}$$

Incorporating the optimal filters into the uplink SINR yields

$$\bar{\gamma}_k^{\mathrm{UL}}(\boldsymbol{q}) = q_k \boldsymbol{t}_{kk}^{\mathrm{H}} \left(\frac{1}{\rho_{\mathrm{dl}}}\mathbf{I} + \sum_n q_n \boldsymbol{B}_n + \sum_{n\neq k} \boldsymbol{t}_{nk}\boldsymbol{t}_{nk}^{\mathrm{H}} q_n\right)^{-1} \boldsymbol{t}_{kk} \tag{19}$$

Given the optimal filters, the optimal uplink power allocation $\boldsymbol{q}$ can be calculated using a fixed-point iteration [15]. To this end, we decompose the uplink SINR with optimal choice for the filters in (19) into

$$\bar{\gamma}_k^{\mathrm{UL}}(\boldsymbol{q}) = \frac{q_k}{\mathcal{I}_k(\boldsymbol{q})} \tag{20}$$

where the effective interference

$$\mathcal{I}_k(\boldsymbol{q}) = \frac{1}{\boldsymbol{t}_{kk}^{\mathrm{H}}\big(\frac{1}{\rho_{\mathrm{dl}}}\mathbf{I} + \sum_n q_n \boldsymbol{B}_n + \sum_{n\neq k} \boldsymbol{t}_{nk}\boldsymbol{t}_{nk}^{\mathrm{H}} q_n\big)^{-1} \boldsymbol{t}_{kk}} \tag{21}$$

captures the impact of the interference and noise on the SINR at user k. The interference in our case has an equivalent structure to the one in [14], thus

$$\mathcal{I}(\boldsymbol{q}) = [\mathcal{I}_1(\boldsymbol{q}), \ldots, \mathcal{I}_1(\boldsymbol{q})]^{\mathrm{T}} \tag{22}$$

satisfies the properties of a standard interference function [14], [16]. Consequently, for given feasible SINR targets

$$\boldsymbol{\sigma}(\beta) = [2^{\beta\tau_k} - 1, \ldots, 2^{\beta\tau_K} - 1]^{\mathrm{T}} \tag{23}$$

the fixed point iteration

$$\boldsymbol{q} \leftarrow \mathrm{diag}(\boldsymbol{\sigma}(\beta))\mathcal{I}(\boldsymbol{q}) \tag{24}$$

converges globally to the unique fixed point [16]. In [17], an algorithm for the sum power constrained problem is introduced. In this method, the rate factor β is adapted in every step to ensure convergence to the unique optimal solution for both β and $\boldsymbol{q}$ that fulfills the transmit power constraint $\boldsymbol{q}^{\mathrm{H}}\boldsymbol{q} = 1$. In our case, the iteration of the following two steps leads to the optimal solution from an arbitrary initial $\boldsymbol{q}$:

$$\beta \leftarrow \beta' : \boldsymbol{\sigma}(\beta')^{\mathrm{T}}\mathcal{I}(\boldsymbol{q}) - 1 = 0 \tag{25}$$

$$\boldsymbol{q} \leftarrow \mathrm{diag}(\boldsymbol{\sigma}(\beta))\mathcal{I}(\boldsymbol{q}) \tag{26}$$

Basically, we adapt β such that the result of the fixed point iteration in (24) always fulfills the transmit power constraint. Note that the root in (25) is unique since the components $\sigma_k(\beta)$ of $\boldsymbol{\sigma}(\beta)$ are strictly increasing in β. Additionally, the derivatives $\sigma_k'(\beta)$ are strictly positive and thus the Newton-Raphson method is globally convergent in this case.

Once we found the optimal power allocation $\boldsymbol{q}^\star$ and uplink filters $\boldsymbol{g}_k^\star$ leading to the optimal uplink SINRs $\gamma_k^{\mathrm{UL}\star}$, we find the optimal power allocation for the downlink by setting

$$\gamma_k = \gamma_k^{\mathrm{UL}\star}\ \forall k \tag{27}$$

and substituting $\boldsymbol{a}_k = \sqrt{p_k}\boldsymbol{g}_k$ in the downlink SINRs. As a result, we obtain K equations

$$\frac{1}{\rho_{\mathrm{dl}}} + \sum_n p_n \boldsymbol{g}_n^{\mathrm{H}}\boldsymbol{B}_k\boldsymbol{g}_n + \sum_{n\neq k} p_n|\boldsymbol{t}_{kn}^{\mathrm{H}}\boldsymbol{g}_n|^2 = p_k\frac{|\boldsymbol{t}_{kk}^{\mathrm{H}}\boldsymbol{g}_k|^2}{\gamma_k \mathrm{UL}^\star} \tag{28}$$

which are linear in the downlink power allocations p_n. That is, we need to solve the linear systems of equations

$$\boldsymbol{\Phi}\boldsymbol{p} = \frac{1}{\rho_{\mathrm{dl}}}\mathbf{1} \tag{29}$$

where $\boldsymbol{p} = [p_1, \ldots, p_K]^{\mathrm{T}}$ and the entries of $\boldsymbol{\Phi}$ can be identified from the equations in (28). Note that due to the structure of $\boldsymbol{\Phi}$ we have $\boldsymbol{p} \geq \mathbf{0}$ [14]. It can further be verified that $\mathbf{1}^{\mathrm{T}}\boldsymbol{\Phi}^{-1}\mathbf{1}\frac{1}{\rho_{\mathrm{dl}}} = 1$ [14]. All steps of the rate balancing algorithm are summarized in Algorithm 1.

Algorithm 1 Statistical Rate-balancing for Massive MIMO

$\boldsymbol{T}_{kn} \leftarrow \boldsymbol{L}_n^{-1}\boldsymbol{C}_{\boldsymbol{h}_k} c_{nk}\ \ \forall k, n$
$\boldsymbol{t}_{kn} \leftarrow \mathrm{vec}(\boldsymbol{T}_{kn})$
$\boldsymbol{B}_k \leftarrow \mathbf{I} \otimes \boldsymbol{C}_{\boldsymbol{h}_k}$
$\boldsymbol{q} \leftarrow \frac{1}{K}\mathbf{1}$
repeat
  $\tilde{\boldsymbol{g}}_k \leftarrow \Big(\frac{1}{\rho_{\mathrm{dl}}}\mathbf{I} + \sum_n q_n \boldsymbol{B}_n + \sum_{n\neq k} \boldsymbol{t}_{nk}\boldsymbol{t}_{nk}^{\mathrm{H}} q_n\Big)^{-1} \boldsymbol{t}_{kk}\ \forall k$
  $\mathcal{I}_k \leftarrow 1/(\boldsymbol{t}_{kk}^{\mathrm{H}}\tilde{\boldsymbol{g}}_k)$
  $\beta \leftarrow \beta : \boldsymbol{\sigma}(\beta)^{\mathrm{T}}\mathcal{I} - 1 = 0$
  $\boldsymbol{q} \leftarrow \mathrm{diag}(\boldsymbol{\sigma}(\beta))\mathcal{I}(\boldsymbol{q})$
until convergence
$\boldsymbol{g}_k \leftarrow \tilde{\boldsymbol{g}}_k / \|\tilde{\boldsymbol{g}}_k\|\ \ \forall k$
$\boldsymbol{p} \leftarrow \boldsymbol{\Phi}^{-1}\mathbf{1}\frac{1}{\rho_{\mathrm{dl}}}$ ▷ Entries of $\boldsymbol{\Phi}$ from (28)
$\boldsymbol{a}_k \leftarrow p_k\boldsymbol{g}_k\ \ \forall k$

IV. Complexity Reduction

The design of the beamformers can also be done with respect to a different basis $\boldsymbol{Q} = [\boldsymbol{q}_1, \ldots, \boldsymbol{q}_M]$. To this end, we transform all inputs into the desired basis

$$\boldsymbol{y}' \leftarrow \boldsymbol{Q}^{-1}\boldsymbol{y}. \tag{30}$$

After the beamformer design, the transmit signal is transformed back into the original basis

$$\boldsymbol{x} \leftarrow \boldsymbol{Q}\boldsymbol{x}'. \tag{31}$$

For an orthonormal basis ($\boldsymbol{Q}^{\mathrm{H}}\boldsymbol{Q} = \mathbf{I}$), the transformation has no effect on the transmit power. Thus, the beamformer design is equivalent to the one without transformation

The idea is to use the transformation to get a more convenient structure of the covariance matrices. For example, for a uniform linear array at the base station, the transformation with the unitary DFT matrix $\boldsymbol{F}$ leads to approximately diagonal covariance matrices. If most of the structural information in

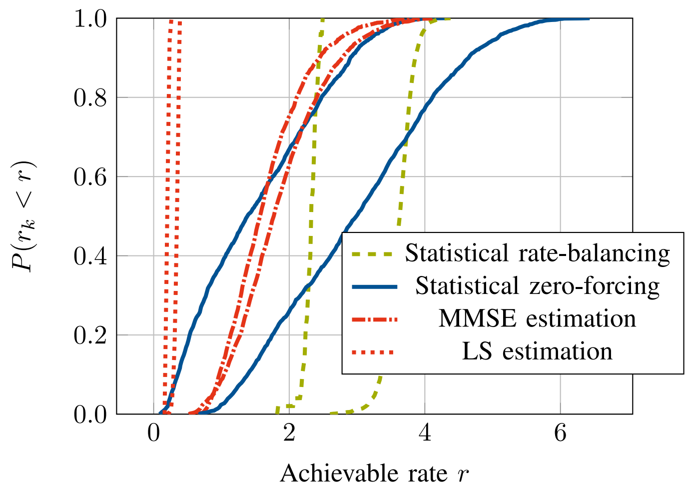

Fig. 1. Experimental cumulative distribution functions of the achievable rates for different precoders in a single cell with $M = 400$ antennas, $K = 40$ users, and $T_{\mathrm{tr}} = 10$ orthonormal pilot sequences. Both, the upper and lower bounds on the achievable rates, are depicted.

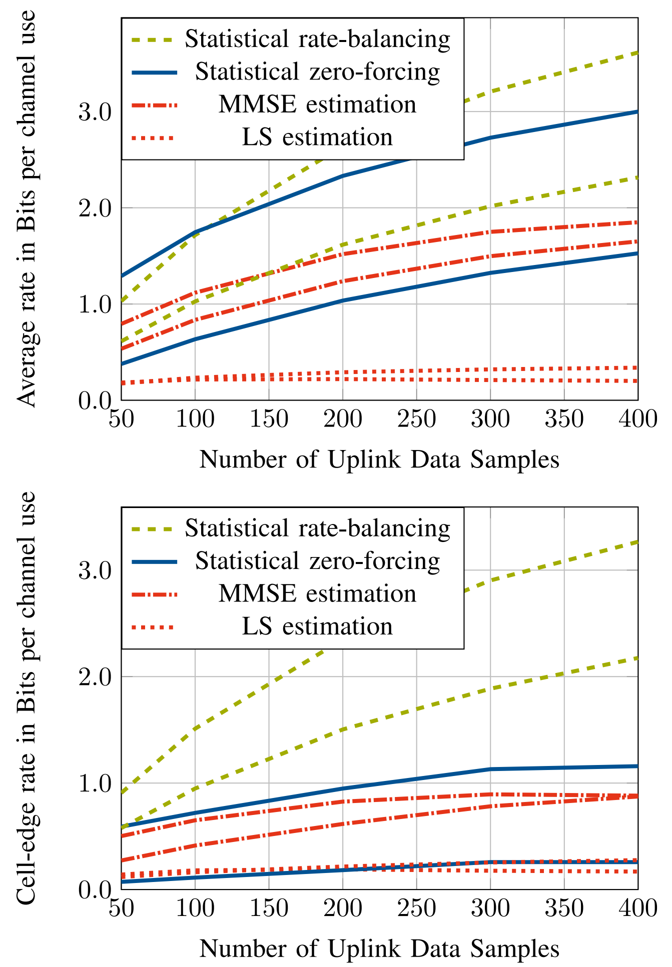

Fig. 2. Achievable rates for different precoders in a single cell with $K = 40$ users and $T_{\mathrm{tr}} = 10$ orthonormal pilot sequences. The upper graph depicts the average rates and the lower one the rate at the 5th percentile. Both, upper and lower bounds, are shown.

the covariance matrix is captured in the diagonal entries, we can also restrict the linear transformations $\boldsymbol{A}_i$ to be diagonal without sacrificing performance.

For diagonal $\boldsymbol{A}_i = \operatorname{diag}(\boldsymbol{a}_i)$, we get expressions that have an equivalent structure as in the case of full matrices, but with significantly reduced computational complexity. Replacing $\boldsymbol{A}_i$ with $\operatorname{diag}(\boldsymbol{a}_i)$ in the SINR expression in (11), we note that only the diagonals $\boldsymbol{c}_{\boldsymbol{h}_i}$ of $\boldsymbol{C}_{\boldsymbol{h}_i}$ and $\boldsymbol{t}_{ij}$ of $\boldsymbol{T}_{ij}$ are relevant for the computation. With $\boldsymbol{B}_i = \operatorname{diag}(\boldsymbol{c}_{\boldsymbol{h}_i})$, we obtain a formulation identically to the one in (14). The dimensionality is reduced form M^2 to M and the matrices $\boldsymbol{B}_i$ are diagonal.

It is clear that the main complexity of Algorithm 1 is the solution of the (now $M \times M$) system of equations to obtain

$$\tilde{\boldsymbol{g}}_k = \Big(\frac{1}{\rho_{\mathrm{dl}}}\mathbf{I} + \sum_n q_n \boldsymbol{B}_n + \sum_{n\neq k} \boldsymbol{t}_{nk}\boldsymbol{t}_{nk}^{\mathrm{H}} q_n\Big)^{-1}\boldsymbol{t}_{kk}. \tag{32}$$

Assume that all pilot sequences are drawn from a set of orthonormal vectors, i.e., the correlations c_{nk} are either 0 or 1. Thus, only those vectors $\boldsymbol{t}_{nk}$ are non-zero for which users k and n use the same pilot sequence. Let $\boldsymbol{\varGamma}_k$ denote the matrix which contains the non-zero $\boldsymbol{t}_{nk}$ of the users n interfering with user k, excluding $\boldsymbol{t}_{kk}$. The vector $\tilde{\boldsymbol{q}}_k$ contains the corresponding uplink powers. With

$$\bar{\boldsymbol{B}} = \frac{1}{\rho_{\mathrm{dl}}}\mathbf{I} + \sum_k q_k \boldsymbol{B}_k \tag{33}$$

we can simplify (32) to

$$\tilde{\boldsymbol{g}}_i = \big(\bar{\boldsymbol{B}} + \boldsymbol{\varGamma}_i \operatorname{diag}(\tilde{\boldsymbol{q}}_k)\boldsymbol{\varGamma}_i^{\mathrm{H}}\big)^{-1}\boldsymbol{t}_{ii}. \tag{34}$$

Due to the diagonal structure of $\bar{\boldsymbol{B}}$ we can reduce the complexity by applying the matrix inversion lemma to get

$$\tilde{\boldsymbol{g}}_k = \bar{\boldsymbol{B}}^{-1}\boldsymbol{t}_{kk} + \bar{\boldsymbol{B}}^{-1}\boldsymbol{\varGamma}_k(\operatorname{diag}(\tilde{\boldsymbol{q}}_k)^{-1} + \boldsymbol{\varGamma}_k^{\mathrm{H}}\bar{\boldsymbol{B}}^{-1}\boldsymbol{\varGamma}_k)^{-1}\boldsymbol{\varGamma}_k^{\mathrm{H}}\bar{\boldsymbol{B}}^{-1}\boldsymbol{t}_{kk}. \tag{35}$$

Consequently, the order of complexity of calculating the optimal filters for all users reduces from $O(KM^3)$ to $O(K(K/T_{\mathrm{tr}})^3) + O(KM(K/T_{\mathrm{tr}}))$ for the inverses and the matrix vector multiplications respectively.

V. Results

We compare the performance of the statistical rate-balancing with other statistical precoding methods and classical precoding. To this end, we present simulations for a single-cell scenario where the number of simultaneously served users is larger than the number of available orthonormal pilot sequences. We compare the proposed rate balancing approach with the statistical zero-forcing approach from [6], simple matched filter precoding based on MMSE estimates, and matched filter precoding based on least squares estimation. To analyze the performance, we use the lower bound on the achievable rate in (8) and the case of perfect CSI at the receivers as an upper bound. The latter is evaluated by Monte-Carlo simulations where channel realizations are generated based on the given covariance matrices. For both bounds,

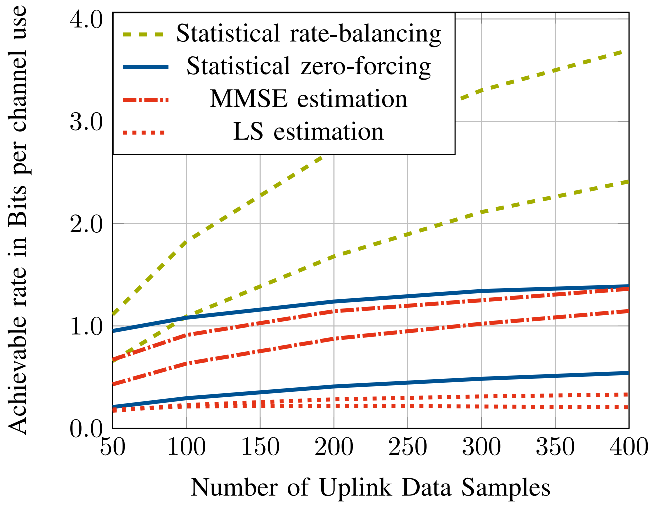

Fig. 3. Achievable rate for different precoders in a single cell with $K = 40$ users and $T_{\text{tr}} = 10$ orthonormal pilot sequences averaged over several simulations. Here the power allocation is optimized towards a balancing solution for all methods. Both, upper and lower bounds, are shown.

several iterations are performed with different uniform user placements in the cell. The covariance matrices for each placement of users are generated according to the urban macro model in [18].

Empirical cumulative distribution functions for both the upper and lower bounds are shown in Fig. 1. For one placement of users, by design, all users have the same lower bound in the rate-balancing case. Since each simulation includes several user placements there is a small variation in the achievable rates, but we observe that a similar rate is achievable for all placements. Also the upper bound ends up to be almost balanced for the proposed rate-balancing approach. We note that the statistical rate-balancing not only leads to a huge gain for the cell-edge users but also to a significant improvement of the average rate. This is further illustrated in Fig. 2, where we present the average achievable rate and the rate of the 5th percentile with respect to the number of transmit antennas. Note that for the approaches that are optimized with respect to the lower bound, namely the statistical rate-balancing and zero-forcing, there is a significant gap between upper and lower bound, i.e., the lower bound might be too pessimistic for these approaches.

In the simulations above, we use a uniform power allocation for all methods except the proposed rate-balancing approach. In Fig. 3, we additionally illustrate the performance when the downlink power allocation is optimized for all methods to achieve rate balanced solutions. As we can see, the proposed rate-balancing significantly outperforms the other approaches due to the fact that not only the power allocation, but also the spatial direction of the statistical precoding is optimized.

VI. Conclusion

We showed how the idea of statistical precoding that was introduced in [6] can be extended to a rate-balancing formulation. To this end, an algorithm for optimal downlink beamforming for the rate-balancing problem was derived by exploiting the uplink-downlink duality. Simulation results demonstrate that the proposed approach significantly outperforms previous methods both in terms of fairness and average spectral efficiency.

References

[1] F. Rusek, D. Persson, B. K. Lau, E. Larsson, T. Marzetta, O. Edfors, and F. Tufvesson, "Scaling Up MIMO: Opportunities and Challenges with Very Large Arrays," *IEEE Signal Processing Magazine*, vol. 30, no. 1, pp. 40 –60, Jan. 2013.

[2] L. Lu, G. Li, A. Swindlehurst, A. Ashikhmin, and R. Zhang, "An Overview of Massive MIMO: Benefits and Challenges," *IEEE Journal of Selected Topics in Signal Processing*, vol. 8, no. 5, pp. 742–758, Oct. 2014.

[3] E. Larsson, O. Edfors, F. Tufvesson, and T. Marzetta, "Massive MIMO for next generation wireless systems," *IEEE Communications Magazine*, vol. 52, no. 2, pp. 186–195, Feb. 2014.

[4] T. Marzetta, "Noncooperative Cellular Wireless with Unlimited Numbers of Base Station Antennas," *IEEE Transactions on Wireless Communications*, vol. 9, no. 11, pp. 3590 –3600, Nov. 2010.

[5] L. Zheng and D. Tse, "Communication on the Grassmann manifold: a geometric approach to the noncoherent multiple-antenna channel," *IEEE Transactions on Information Theory*, vol. 48, no. 2, pp. 359–383, Feb. 2002.

[6] D. Neumann, M. Joham, and W. Utschick, "CDI Precoding for Massive MIMO," in *Conference on Systems Communications and Coding (SCC)*, Feb. 2015.

[7] M. Medard, "The effect upon channel capacity in wireless communications of perfect and imperfect knowledge of the channel," *IEEE Transactions on Information Theory*, vol. 46, no. 3, pp. 933–946, May 2000.

[8] A. Ashikhmin and T. Marzetta, "Pilot contamination precoding in multi-cell large scale antenna systems," in *International Symposium on Information Theory Proceedings (ISIT)*, Jul. 2012, pp. 1137 –1141.

[9] L. Li, A. Ashikhmin, and T. Marzetta, "Pilot contamination precoding for interference reduction in large scale antenna systems," in *2013 51st Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, Oct. 2013, pp. 226–232.

[10] D. Neumann, M. Joham, L. Weiland, and W. Utschick, "Low-Complexity Computation of LMMSE Channel Estimates in Massive MIMO," in *WSA 2015; 19th International ITG Workshop on Smart Antennas; Proceedings of*, Mar. 2015, pp. 1–6.

[11] H. Yin, D. Gesbert, M. Filippou, and Y. Liu, "A Coordinated Approach to Channel Estimation in Large-Scale Multiple-Antenna Systems," *IEEE Journal on Selected Areas in Communications*, vol. 31, no. 2, pp. 264 –273, Feb. 2013.

[12] H. Yin, D. Gesbert, and L. Cottatellucci, "Dealing With Interference in Distributed Large-Scale MIMO Systems: A Statistical Approach," *IEEE Journal of Selected Topics in Signal Processing*, vol. 8, no. 5, pp. 942–953, Oct. 2014.

[13] M. Schubert and H. Boche, "Solution of the multiuser downlink beamforming problem with individual SINR constraints," *IEEE Transactions on Vehicular Technology*, vol. 53, no. 1, pp. 18–28, Jan. 2004.

[14] M. Joham, A. Grundinger, A. Pastore, J. Fonollosa, and W. Utschick, "Rate balancing in the vector BC with erroneous CSI at the receivers," in *2013 47th Annual Conference on Information Sciences and Systems (CISS)*, Mar. 2013, pp. 1–6.

[15] M. Schubert and H. Boche, "A Generic Approach to QoS-Based Transceiver Optimization," *IEEE Transactions on Communications*, vol. 55, no. 8, pp. 1557–1566, Aug. 2007.

[16] R. Yates, "A framework for uplink power control in cellular radio systems," *IEEE Journal on Selected Areas in Communications*, vol. 13, no. 7, pp. 1341–1347, Sep. 1995.

[17] C. Hellings, M. Joham, and W. Utschick, "Gradient-based rate balancing for MIMO broadcast channels with linear precoding," in *2011 International ITG Workshop on Smart Antennas (WSA)*, Feb. 2011, pp. 1–8.

[18] ITU-R, "Guidelines for evaluation of radio interface technologies for IMT-Advanced," International Telecommunication Union (ITU), Tech. Rep. Report ITU-R M.2135-1, 2009.