Multi-Cell Multi-User MIMO Downlink with Partial CSIT and Decentralized Design

Yohan Lejosne[‡¶], Atef Ben Nasser[¶], Dirk Slock[‡], Yi Yuan-Wu[¶]
[‡]EURECOM, Sophia-Antipolis, France, Email: {lejosne,slock}@eurecom.fr
[¶]Orange Labs, Issy-les-Moulineaux, France, Email: {yohan.lejosne,atef.bennasser,yi.yuan}@orange.com

***Abstract*—The Interfering Broadcast Channel (IBC) applies to the downlink of multi-cell networks, which are limited by multi-user (MU) interference. The interference alignment (IA) concept has shown that interference does not need to be inevitable. In particular spatial IA in the MIMO IBC allows for low latency. However, IA requires perfect and typically global Channel State Information at the Transmitter(s) (CSIT), whose acquisition does not scale well with network size. Also, the design of transmitters (Txs) and receivers (Rxs) is coupled and hence needs to be centralized (cloud) or duplicated (distributed approach). CSIT, which is crucial in multi-user systems, is always imperfect in practice, especially for the intercell links. We consider mean and covariance Gaussian partial CSIT, and the special case of a (possibly location based) MIMO Ricean channel model. In this paper we focus on the optimization of beamformers for the expected weighted sum rate (EWSR) under per BS power constraints. We apply a perfect CSI technique, based on a difference of convex functions approach, to a number of deterministic approximations of the EWSR, involving the Massive MIMO (MaMIMO) limit (large number of transmit antennas), and the large MIMO limit (both large transmit and receive antenna numbers). We then focus on distributed techniques that exploit local CSIT, feedback of a limited number of scalars, and only one or few iterations.**

I. Introduction

In this paper, Tx may denote transmit/transmitter/transmission and Rx may denote receive/receiver/reception. Interference is the main limiting factor in wireless transmission. Base stations (BSs) disposing of multiple antennas are able to serve multiple Mobile Terminals (MTs) simultaneously, which is called Spatial Division Multiple Access (SDMA) or Multi-User (MU) MIMO. However, MU systems have precise requirements for Channel State Information at the Tx (CSIT) which is more difficult to acquire than CSI at the Rx (CSIR). Hence we focus here on the more challenging downlink (DL).

The main difficulty in realizing linear IA for MIMO I(B)C is that the design of any BS Tx filter depends on all Rx filters whereas in turn each Rx filter depends on all Tx filters [1]. As a result, all Tx/Rx filters are globally coupled and their design requires global CSIT. To carry out this Tx/Rx design in a distributed fashion, global CSIT is required at all BS [2]. The overhead required for this global distributed CSIT is substantial, even if done optimally, leading to substantially reduced Net Degrees of Freedom (DoF) [3].

The recent development of Massive MIMO (MaMIMO) [4] opens new possibilities for increased system capacity while at the same time simplifying system design. From a DoF point of view it may seem like a suboptimal use of antennas. However, as shown in [5], section V, Fig. 6, the (MaMIMO asymptotics based analytical expression for the) optimal number of users decreases below the DoF as the SNR decreases. Furthermore, Net DoF considerations and CSI acquisition make the optimal number of users decrease further. In [6], [7], MISO was considered in a single cell. Statistical CSIT between user groups was considered and instantaneous CSIT within user groups. The hypothesis is that some users overlap strongly in terms of covariance subspaces but not in terms of instantaneous CSIT. In [8] a hierarchical approach is considered.MISO is considered and (high SNR based) user selection also. Intercell zero-forcing (ZF) beamforming (BF) is considered based on statistical CSIT, treating interfering links in a binary fashion (either ZF or ignore). Intracell BF is based on instantaneous CSIT and performs Regularized-ZF, which is claimed to be asymptotically optimal (which is only true for uniform user power profile). In [9], following up on work in [10], beamspace processing is proposed, which is the basic form of hierarchical BF. As argued in [7] also, mmWave communications, which we target here also, facilitate MaMIMO, and lead to a limited number of dominant paths as they approach optics.

What is known as MaMIMO is more appropriately called MU Massive MISO whereas here we consider actual MU MC MaMIMO. In this paper the objective is to find the set of beamforming (BF) vectors that maximize the Weighted Sum Rate (WSR) of the IBC network. In [11] the alternative problem formulation of SINR balancing is considered.

Partial CSIT formulations can typically be categorized as either bounded error / worst case (relevant for quantization error in digital feedback) or Gaussian error (relevant for analog feedback, prediction error, second-order statistics information etc.). The Gaussian CSIT formulation with mean and covariance information was first introduced for SDMA (a Direction of Arrival (DoA) based historical precedent of MU MIMO), in which the channel outer product was typically replaced by the transmit side channel correlation matrix, and worked out in more detail for single user (SU) MIMO, e.g. [12]. The use of covariance CSIT has recently reappeared in the context of MaMIMO [13], where a not so rich propagation environment leads to subspaces (slow CSIT) for the channel vectors so that the fast CSIT can be reduced to the smaller dimension of the subspace. Such CSIT (feedback) reduction is especially crucial for MaMIMO.

The contributions here are significantly better partial CSIT approaches compared to the Expected Weighted Sum MSE (EWSMSE) approach in [14] (which cannot even be used in the zero channel mean case), and to present deterministic alternatives to the stochastic approximation solution of [15]. We first treat the general Gaussian CSIT case. We also focus on a location aided CSIT case with zero mean and identity plus rank one Tx side covariance matrix and no Rx side correlations. The goal here is to go beyond the extreme of zero-forcing (ZF) and to introduce a meaningful beamforming design at finite SNR and with partial CSIT, for e.g. a finite Ricean factor when not much more than the (location based) LoS information of the intercell links is available at the BS. The other goal is to arrive at distributed approaches in which global CSIT (channels from all BS) gets replaced by local CSIT (channels from own BS only) plus feedback of a limited number of scalars, as in [16], [17]. However, the design of MIMO systems, as opposed to MISO systems, complicates distributed designs if one wants to keep the feedback low. Another issue is to keep the number of iterations low, for performance and feedback considerations.

II. Streamwise IBC Signal Model

In the rest of this paper we shall consider a per stream approach (which in the perfect CSI case would be equivalent to per user). In an IBC formulation, one stream per user can be expected to be the usual scenario. In the development below, in the case of more than one stream per user, treat each stream as an individual user. So, consider again an IBC with C cells with a total of K users. We shall consider a system-wide numbering of the users. User k is served by BS b_k. The $N_k \times 1$ received signal at user k in cell b_k is

$$\mathbf{y}_k = \underbrace{\mathbf{H}_{k,b_k}\,\mathbf{g}_k\,x_k}_{\text{signal}} + \underbrace{\sum_{\substack{i\neq k\\ b_i=b_k}} \mathbf{H}_{k,b_k}\,\mathbf{g}_i\,x_i}_{\text{intracell interf.}} + \underbrace{\sum_{j\neq b_k}\sum_{i:b_i=j} \mathbf{H}_{k,j}\,\mathbf{g}_i\,x_i}_{\text{intercell interf.}} + \mathbf{v}_k \tag{1}$$

where x_k is the intended (white, unit variance) scalar signal stream, $\mathbf{H}_{k,b_k}$ is the $N_k \times M_{b_k}$ channel from BS b_k to user k. BS b_k serves $K_{b_k} = \sum_{i:b_i=b_k} 1$ users. We considering a noise whitened signal representation so that we get for the noise $\mathbf{v}_k \sim \mathcal{CN}(0, I_{N_k})$. The $M_{b_k} \times 1$ spatial Tx filter or beamformer (BF) is $\mathbf{g}_k$. Treating interference as noise, user k will apply a linear Rx filter $\mathbf{f}_k$ to maximize the signal power (diversity) while reducing any residual interference that would not have been (sufficiently) suppressed by the BS Tx. The Rx filter output is $\widehat{x}_k = \mathbf{f}_k^H \mathbf{y}_k$

$$\begin{aligned} \widehat{x}_k &= \mathbf{f}_k^H \mathbf{H}_{k,b_k}\,\mathbf{g}_k\,x_k + \sum_{i=1,\neq k}^{K} \mathbf{f}_k^H \mathbf{H}_{k,b_i}\,\mathbf{g}_i\,x_i + \mathbf{f}_k^H \mathbf{v}_k \\ &= \mathbf{f}_k^H \mathbf{h}_{k,k}\,x_k + \sum_{i\neq k} \mathbf{f}_k^H \mathbf{h}_{k,i}\,x_i + \mathbf{f}_k^H \mathbf{v}_k \end{aligned} \tag{2}$$

where $\mathbf{h}_{k,i} = \mathbf{H}_{k,b_i}\,\mathbf{g}_i$ is the channel-Tx cascade vector.

III. Max WSR with Perfect CSIT

Consider as a starting point for the optimization the weighted sum rate (WSR)

$$WSR = WSR(\mathbf{g}) = \sum_{k=1}^{K} u_k\ \ln\frac{1}{e_k} \tag{3}$$

where $\mathbf{g}$ represents the collection of BFs $\mathbf{g}_k$, the u_k are rate weights, the $e_k = e_k(\mathbf{g})$ are the Minimum Mean Squared Errors (MMSEs) for estimating the x_k:

$$\begin{aligned} &\frac{1}{e_k} = 1 + \mathbf{g}_k^H \mathbf{H}_{k,b_k}^H \mathbf{R}_{\overline{k}}^{-1} \mathbf{H}_{k,b_k} \mathbf{g}_k = (1 - \mathbf{g}_k^H \mathbf{H}_{k,b_k}^H \mathbf{R}_k^{-1} \mathbf{H}_{k,b_k} \mathbf{g}_k)^{-1} \\ &\mathbf{R}_k = \mathbf{H}_{k,b_k} \mathbf{Q}_k\, \mathbf{H}_{k,b_k}^H + \mathbf{R}_{\overline{k}}\,,\ \mathbf{Q}_i = \mathbf{g}_i \mathbf{g}_i^H\,, \\ &\mathbf{R}_{\overline{k}} = \sum_{i\neq k} \mathbf{H}_{k,b_i} \mathbf{Q}_i\, \mathbf{H}_{k,b_i}^H + I_{N_k}\,. \end{aligned} \tag{4}$$

$\mathbf{R}_k$, $\mathbf{R}_{\overline{k}}$ are the total and interference plus noise Rx covariance matrices resp. and e_k is the MMSE obtained at the output $\widehat{x}_k = \mathbf{f}_k^H \mathbf{y}_k$ of the optimal (MMSE) linear Rx $\mathbf{f}_k$,

$$\mathbf{f}_k = \mathbf{R}_k^{-1} \mathbf{H}_{k,b_k} \mathbf{g}_k = \mathbf{R}_k^{-1} \mathbf{h}_{k,k}\,. \tag{5}$$

The WSR cost function needs to be augmented with the power constraints

$$\sum_{k:b_k=j} \text{tr}\{\mathbf{Q}_k\} \leq P_j\,. \tag{6}$$

In a classical difference of convex functions (DC programming) approach, Kim and Giannakis [18] propose to keep the concave signal terms and to replace the convex interference terms by the linear (and hence concave) tangent approximation. More specifically, consider the dependence of WSR on $\mathbf{Q}_k$ alone. Then

$$\begin{aligned} WSR &= u_k \ln\det(\mathbf{R}_{\overline{k}}^{-1} \mathbf{R}_k) + WSR_{\overline{k}}\,, \\ WSR_{\overline{k}} &= \textstyle\sum_{i=1,\neq k}^{K} u_i \ln\det(\mathbf{R}_{\overline{i}}^{-1} \mathbf{R}_i) \end{aligned} \tag{7}$$

where $\ln\det(\mathbf{R}_{\overline{k}}^{-1} \mathbf{R}_k)$ is concave in $\mathbf{Q}_k$ and $WSR_{\overline{k}}$ is convex in $\mathbf{Q}_k$. Since a linear function is simultaneously convex and concave, consider the first order Taylor series expansion in $\mathbf{Q}_k$ around $\widehat{\mathbf{Q}}$ (i.e. all $\widehat{\mathbf{Q}}_i$) with e.g. $\widehat{\mathbf{R}}_i = \mathbf{R}_i(\widehat{\mathbf{Q}})$, then

$$\begin{aligned} &WSR_{\overline{k}}(\mathbf{Q}_k, \widehat{\mathbf{Q}}) \approx WSR_{\overline{k}}(\widehat{\mathbf{Q}}_k, \widehat{\mathbf{Q}}) - \text{tr}\{(\mathbf{Q}_k - \widehat{\mathbf{Q}}_k) \widehat{\mathbf{A}}_k\} \\ &\widehat{\mathbf{A}}_k = -\left.\frac{\partial WSR_{\overline{k}}(\mathbf{Q}_k, \widehat{\mathbf{Q}})}{\partial \mathbf{Q}_k}\right|_{\widehat{\mathbf{Q}}_k, \widehat{\mathbf{Q}}} = \sum_{i\neq k}^{K} u_i \mathbf{H}_{i,b_k}^H (\widehat{\mathbf{R}}_{\overline{i}}^{-1} - \widehat{\mathbf{R}}_i^{-1}) \mathbf{H}_{i,b_k} \end{aligned} \tag{8}$$

Note that the linearized (tangent) expression for $WSR_{\overline{k}}$ constitutes a lower bound for it. Now, dropping constant terms, reparameterizing the $Q_k = \mathbf{g}_k \mathbf{g}_k^H$, performing this linearization for all users, and augmenting the WSR cost function with the constraints, we get the Lagrangian

$$\begin{aligned} &WSR(\mathbf{g}, \widehat{\mathbf{g}}, \lambda) = \sum_{j=1}^{C} \lambda_j P_j + \\ &\sum_{k=1}^{K} u_k \ln(1 + \mathbf{g}_k^H \widehat{\mathbf{B}}_k \mathbf{g}_k) - \mathbf{g}_k^H (\widehat{\mathbf{A}}_k + \lambda_{b_k} I) \mathbf{g}_k \end{aligned} \tag{9}$$

where

$$\widehat{\mathbf{B}}_k = \mathbf{H}_{k,b_k}^H \widehat{\mathbf{R}}_{\overline{k}}^{-1} \mathbf{H}_{k,b_k} \ . \tag{10}$$

The gradient (w.r.t. $\mathbf{g}_k$) of this concave WSR lower bound is actually still the same as that of the original WSR criterion! And it allows an interpretation as a generalized eigenvector condition

$$\widehat{\mathbf{B}}_k \, \mathbf{g}_k = \frac{1 + \mathbf{g}_k^H \widehat{\mathbf{B}}_k \mathbf{g}_k}{u_k} \, (\widehat{\mathbf{A}}_k + \lambda_{b_k} I) \mathbf{g}_k \tag{11}$$

or hence $\mathbf{g}_k^{'} = V_{max}(\widehat{\mathbf{B}}_k, \widehat{\mathbf{A}}_k + \lambda_{b_k} I)$ is the (normalized) "max" generalized eigenvector of the two indicated matrices, with max eigenvalue $\sigma_k = \sigma_{max}(\widehat{\mathbf{B}}_k, \widehat{\mathbf{A}}_k + \lambda_{b_k} I)$. Let $\sigma_k^{(1)} = \mathbf{g}_k^{'H} \widehat{\mathbf{B}}_k \mathbf{g}_k^{'}$, $\sigma_k^{(2)} = \mathbf{g}_k^{'H} \widehat{\mathbf{A}}_k \mathbf{g}_k^{'}$. The advantage of formulation (9) is that it allows straightforward power adaptation: introducing stream powers $p_k \geq 0$ and substituting $\mathbf{g}_k = \sqrt{p_k}\, \mathbf{g}_k^{'}$ in (9) yields

$$WSR = \sum_{j}^{C} \lambda_j P_j + \sum_{k=1}^{K} \{ u_k \ln(1 + p_k \sigma_k^{(1)}) - p_k (\sigma_k^{(2)} + \lambda_{b_k}) \} \tag{12}$$

which leads to the following interference leakage aware water filling

$$p_k = \left(\frac{u_k}{\sigma_k^{(2)} + \lambda_{b_k}} - \frac{1}{\sigma_k^{(1)}} \right)^{+} \tag{13}$$

where the Lagrange multipliers are adjusted to satisfy the power constraints $\sum_{k:b_k=j} p_k = P_j$. This can be done by bisection and gets executed per BS. Note that some Lagrange multipliers could be zero. Note also that as with any alternating optimization procedure, there are many updating schedules possible, with different impact on convergence speed. The quantities to be updated are the $\mathbf{g}_k^{'}$, the p_k and the λ_l.

IV. Mean and Covariance Gaussian CSIT

In this section we drop the user index k for simplicity. The separable correlation model is

$$\mathbf{H} = \overline{\mathbf{H}} + \mathbf{C}_r^{1/2} \, \widetilde{\mathbf{H}} \, \mathbf{C}_t^{1/2} \tag{14}$$

where $\overline{\mathbf{H}} = \mathrm{E}\,\mathbf{H}$, and $\mathbf{C}_r^{1/2}$, $\mathbf{C}_t^{1/2}$ are Hermitian square-roots of the Rx and Tx side covariance matrices

$$\begin{aligned} \mathrm{E}(\mathbf{H} - \overline{\mathbf{H}})(\mathbf{H} - \overline{\mathbf{H}})^H &= \mathrm{tr}\{\mathbf{C}_t\} \ \mathbf{C}_r \\ \mathrm{E}(\mathbf{H} - \overline{\mathbf{H}})^H(\mathbf{H} - \overline{\mathbf{H}}) &= \mathrm{tr}\{\mathbf{C}_r\} \ \mathbf{C}_t \end{aligned} \tag{15}$$

and the elements of $\widetilde{\mathbf{H}}$ are i.i.d. $\sim \mathcal{CN}(0,1)$. It is also of interest to consider the total Tx side correlation matrix

$$\mathbf{R}_t = \mathrm{E}\,\mathbf{H}^H \mathbf{H} = \overline{\mathbf{H}}^H \overline{\mathbf{H}} + \mathrm{tr}\{\mathbf{C}_r\} \, \mathbf{C}_t \ . \tag{16}$$

A. Location Aided Partial CSIT LoS Channel Model

Assuming that for certain (e.g. intercell) links a BS disposes of not much more than the LoS component information, consider the following MIMO channel model

$$\mathbf{H} = \mathbf{h}_r \, \mathbf{h}_t^H(\theta) + \widetilde{\mathbf{H}}^{'} \tag{17}$$

where θ is the LoS Angle of Departure (AoD) and the BS side array response is normalized: $||\mathbf{h}_t(\theta)||^2 = 1$. We shall model the unknown Rx side LoS array response $\mathbf{h}_r$ as a vector of i.i.d. complex Gaussian variables

$$\begin{aligned} &\mathbf{h}_r \ \text{ i.i.d. } \sim \mathcal{CN}(0, \tfrac{\mu}{\mu+1}) \ \text{ and} \\ &\widetilde{\mathbf{H}}^{'} \ \text{ i.i.d. } \sim \mathcal{CN}(0, \tfrac{1}{\mu+1} \tfrac{1}{M}) \text{ , independent of } \mathbf{h}_r, \end{aligned} \tag{18}$$

where the matrix $\widetilde{\mathbf{H}}$ represents the aggregate NLoS components. Note that $(\mathrm{E}||\mathbf{h}_r \, \mathbf{h}_t^T(\theta)||_F^2)/(\mathrm{E}||\widetilde{\mathbf{H}}^{'}||_F^2) = \mu$ can be considered as a Rice factor. In fact the only parameter additional to the LoS AoD θ assumed in (17) is μ. So, this is a case of zero mean CSIT and Tx side covariance CSIT

$$\mathbf{C}_t = \mathrm{E}\,\mathbf{H}^H \mathbf{H} = \frac{\mu N}{\mu + 1} \mathbf{h}_t(\theta) \mathbf{h}_t^H(\theta) + \frac{N}{\mu + 1} \frac{1}{M} I_M \ . \tag{19}$$

V. Expected WSR (EWSR)

For the WSR criterion, we have assumed so far that the channel $\mathbf{H}$ is known. The scenario of interest however is that of partial CSIT, e.g. perfect or good partial intracell CSIT but very partial (zero mean, e.g. LoS) CSIT of the intercell links. Once the CSIT is imperfect, various optimization criteria could be considered, such as outage capacity. Here we shall consider the expected weighted sum rate $\mathrm{E}_{\mathbf{H}} WSR(\mathbf{g}, \mathbf{H}) =$

$$EWSR(\mathbf{g}) = \mathrm{E}_{\mathbf{H}} \sum_k u_k \ln(1 + \mathbf{g}_k^H \mathbf{H}_{k,b_k}^H \mathbf{R}_{\overline{k}}^{-1} \mathbf{H}_{k,b_k} \mathbf{g}_k) \tag{20}$$

where we now underlign the dependence of various quantities on $\mathbf{H}$. The EWSR in (20) corresponds to perfect CSIR since the optimal Rx filters $\mathbf{f}_k$ as a function of the aggregate $\mathbf{H}$ have been substituted, namely $WSR(\mathbf{g}, \mathbf{H}) = \max_{\mathbf{f}} \sum_k u_k(-\ln(e_k(\mathbf{f}_k, \mathbf{g})))$. At high SNR, max EWSR attempts ZF.

In [15] a **stochastic approximation** approach for maximizing the EWSR was introduced. In this approach the statistical average gets replaced by a sample average (samples of $\mathbf{H}$ get generated according to its Gaussian CSIT distribution in a Monte Carlo fashion), and one iteration of the min WSMSE (Weighted Sum MSE) approach gets executed per term added in the sample average.

Some issues with this approach are that in this case the number of iterations may get dictated by a sufficient size for the sample average rather than by a convergence requirement for the iterative approach. Another issue is that this approach converges to a local maximum of the EWSR. It is not immediately clear how to combine this stochastic approximation approach with deterministic annealing. Deterministic annealing can be used as in [1] for a deterministic algorithm as in Section III to track the global optimum from SNR ≈ 0 (where the solution is clear analytically) to the desired SNR. This is essentially a homotopy method in which the problem gets resolved for an SNR that increases in small steps. At each higher SNR, the global optimum will be in the region of attraction of the global optimum at the lower SNR.

In the rest of this paper we discuss **various deterministic approximations for the EWSR**, which can then be optimized as in the full CSI case.

VI. MaMIMO Limit

If the number of Tx antennas M becomes very large, then quantities of the form $\mathbf{H}\mathbf{R}^{-1}\mathbf{H}^H$ converge to their mean (by the LLN). Hence in the MaMIMO limit, the WSR converges to a deterministic limit that depends on the distribution of the channels. The actual statistical distribution of the channel is one thing. The CSIT distribution as in Section IV is another. The Txs have no choice but to design their BFs according to their partial CSIT. Then to get the actual resulting WSR, the BFs designed with the partial CSIT need to be evaluated with the actual channel distribution.

Now, for the design with partial CSIT, the WSR will also converge to a deterministic limit in the MaMIMO regime. We get a convergence for any term of the form

$$\mathbf{H}\mathbf{Q}\mathbf{H}^H \overset{M\to\infty}{\longrightarrow} \mathrm{E}_{\mathbf{H}}\,\mathbf{H}\mathbf{Q}\mathbf{H}^H = \overline{\mathbf{H}}\mathbf{Q}\overline{\mathbf{H}}^H + \mathrm{tr}\{\mathbf{Q}\mathbf{C}_t\}\,\mathbf{C}_r\,. \quad (21)$$

In what follows we shall go one step further in the separable channel correlation model and assume $\mathbf{C}_{r,k,b_i} = \mathbf{C}_{r,k}$, $\forall b_i$. This leads us to introduce

$$\begin{aligned}
&\mathbf{H}_k = [\mathbf{H}_{k,1}\cdots\mathbf{H}_{k,C}] = \overline{\mathbf{H}}_k + \mathbf{C}_{r,k}^{1/2}\widetilde{\mathbf{H}}_k\mathbf{C}_{t,k}^{1/2}\\
&\mathbf{Q} = \begin{bmatrix} \sum_{i:b_i=1}\mathbf{Q}_i & & \\ & \ddots & \\ & & \sum_{i:b_i=C}\mathbf{Q}_i \end{bmatrix} = \sum_{j=1}^{C}\sum_{i:b_i=j}\mathbf{I}_j\mathbf{Q}_i\mathbf{I}_j^H \\
&\mathbf{Q}_{\overline{k}} = \mathbf{Q} - \mathbf{I}_{b_i}\mathbf{Q}_i\mathbf{I}_{b_i}^H
\end{aligned} \quad (22)$$

where $\mathbf{C}_{t,k} = \mathrm{blockdiag}\{\mathbf{C}_{t,k,1},\ldots,\mathbf{C}_{t,k,C}\}$, and $\mathbf{I}_j$ is an all zero block vector except for an identity matrix in block j. Then we get for the WSR ($= EWSR$), using (21),

$$WSR = \sum_{k=1}^{K} u_k \ln\det(\breve{\mathbf{R}}_{\overline{k}}^{-1}\breve{\mathbf{R}}_k) \quad (23)$$

where

$$\begin{aligned}
\breve{\mathbf{R}}_k &= I_{N_k} + \overline{\mathbf{H}}_k\mathbf{Q}\overline{\mathbf{H}}_k^H + \mathrm{tr}\{\mathbf{Q}\mathbf{C}_{t,k}\}\,\mathbf{C}_{r,k}\\
\breve{\mathbf{R}}_{\overline{k}} &= I_{N_k} + \overline{\mathbf{H}}_k\mathbf{Q}_{\overline{k}}\overline{\mathbf{H}}_k^H + \mathrm{tr}\{\mathbf{Q}_{\overline{k}}\mathbf{C}_{t,k}\}\,\mathbf{C}_{r,k}
\end{aligned} \quad (24)$$

This leads to

$$\begin{aligned}
WSR = u_k \ln\det(I + \breve{\mathbf{R}}_{\overline{k}}^{-1}\Big(&\overline{\mathbf{H}}_{k,b_k}\mathbf{g}_k\mathbf{g}_k^H\overline{\mathbf{H}}_{k,b_k}^H \\
&+\mathrm{tr}\{\mathbf{g}_k\mathbf{g}_k^H\mathbf{C}_{t,k,b_k}\}\mathbf{C}_{r,k})\big) + WSR_{\overline{k}}\,.
\end{aligned} \quad (25)$$

Whereas the generalized eigenvector condition for the BF $\mathbf{g}_k$ can be worked out from here, the adaptation of the stream power p_k gets a bit complex for this case of general Gaussian CSIT. So, for what follows, consider the simplified case of a "Ricean factor" μ in the Gaussian partial CSIT model that behaves proportionally to the SNR, for the direct links $\mathbf{H}_{k,b_k}$ (only). This corresponds to the case of $\overline{\mathbf{H}}_{k,b_k}$ representing a channel estimate, with properly organized (intracell) channel estimation and feedback. In that case the dependence of the user k rate on its BF in (25) can be approximated by

$$\begin{aligned}
&WSR = u_k \ln\det(I + \mathbf{g}_k^H\breve{\mathbf{B}}_k\mathbf{g}_k) + WSR_{\overline{k}} \quad \text{with}\\
&\breve{\mathbf{B}}_k = \overline{\mathbf{H}}_{k,b_k}^H\breve{\mathbf{R}}_{\overline{k}}^{-1}\overline{\mathbf{H}}_{k,b_k} + \mathrm{tr}\{\mathbf{C}_{r,k}\breve{\mathbf{R}}_{\overline{k}}^{-1}\}\mathbf{C}_{t,k,b_k}
\end{aligned} \quad (26)$$

which is now similar to the corresponding interpretation in (9). Note that $\breve{\mathbf{B}}_k$ corresponds to a total Tx side correlation matrix as in (16), but now for an interference plus noise whitened channel $\breve{\mathbf{R}}_{\overline{k}}^{-1/2}\mathbf{H}_{k,b_k}$. The linearization of $WSR_{\overline{k}}$ w.r.t. $\mathbf{Q}_k$ now involves

$$\begin{aligned}
\breve{\mathbf{A}}_k = \sum_{i\neq k}^{K} u_i \Big[&\overline{\mathbf{H}}_{i,b_k}^H\left(\breve{\mathbf{R}}_{\overline{i}}^{-1} - \breve{\mathbf{R}}_i^{-1}\right)\overline{\mathbf{H}}_{i,b_k}\\
&+\mathrm{tr}\{\left(\breve{\mathbf{R}}_{\overline{i}}^{-1} - \breve{\mathbf{R}}_i^{-1}\right)\mathbf{C}_{r,i}\}\mathbf{C}_{t,i,b_k}\Big]\,.
\end{aligned} \quad (27)$$

The rest of the development is now completely analogous to the case of perfect CSIT. Note that for general Gaussian partial CSIT we get e.g.

$$\begin{aligned}
\breve{\mathbf{R}}_{\overline{i}}^{-1} - \breve{\mathbf{R}}_i^{-1} &= \breve{\mathbf{R}}_i^{-1}(\breve{\mathbf{R}}_i - \breve{\mathbf{R}}_{\overline{i}})\breve{\mathbf{R}}_{\overline{i}}^{-1}\\
&= \breve{\mathbf{R}}_i^{-1}(\mathbf{g}_i^H\mathbf{C}_{t,i,b_i}\mathbf{g}_i\,I + \overline{\mathbf{H}}_{i,b_i}\mathbf{g}_i\mathbf{g}_i^H\overline{\mathbf{H}}_{i,b_i}^H)\breve{\mathbf{R}}_{\overline{i}}^{-1}\,.
\end{aligned} \quad (28)$$

VII. Large MIMO Asymptotics

The large MIMO asymptotics from [19], [20], in which both $M, N \to \infty$ at constant ratio, tend to give more precise approximations when M is not so large. For the general case of Gaussian CSIT with separable (Kronecker) covariance structure, [19], [20] lead to asymptotic expressions of the form

$$\begin{aligned}
&\mathrm{E}_{\mathbf{H}}\ln\det(I + \mathbf{H}\mathbf{Q}\mathbf{H}^H)\\
&= \max_{z\geq 0,\, w\geq 0}\left\{\ln\det\begin{bmatrix} I + w\mathbf{C}_r & \overline{\mathbf{H}} \\ -\mathbf{Q}\overline{\mathbf{H}}^H & I + z\mathbf{Q}\mathbf{C}_t\end{bmatrix} - zw\right\}.
\end{aligned} \quad (29)$$

where the maximization over z and w should be carried out alternatingly (and not jointly: the joint optimization may correspond to a global maximum or a saddle point; the cost function is concave however in z or w separately). We shall assume the same fully separable correlation Gaussian channel model as in (22). The EWSR with large MIMO asymptotics now becomes

$$\begin{aligned}
EWSR = \sum_{k=1}^{K} u_k \Big(&\max_{z_k,w_k}\{\ln\det\mathbf{S}_k(\mathbf{Q}, z_k, w_k) - z_k w_k\}\\
&- \max_{z_{\overline{k}},w_{\overline{k}}}\{\ln\det\mathbf{S}_k(\mathbf{Q}_{\overline{k}}, z_{\overline{k}}, w_{\overline{k}}) - z_{\overline{k}}w_{\overline{k}}\}\Big)
\end{aligned} \quad (30)$$

where

$$\mathbf{S}_k(\mathbf{Q}, z, w) = \begin{bmatrix} I + w\mathbf{C}_{r,k} & \overline{\mathbf{H}}_k \\ -\mathbf{Q}\overline{\mathbf{H}}_k^H & I + z\mathbf{Q}\mathbf{C}_{t,k}\end{bmatrix}. \quad (31)$$

Note that

$$\begin{aligned}
&\ln\det\mathbf{S}_k(\mathbf{Q}, z, w) = \ln\det(I + w\mathbf{C}_{r,k}) + \ln\det(I + \mathbf{Q}\mathbf{T}_k(z, w))\\
&\quad\text{with } \mathbf{T}_k(z, w) = z\mathbf{C}_{t,k} + \overline{\mathbf{H}}_k^H(I + w\mathbf{C}_{r,k})^{-1}\overline{\mathbf{H}}_k
\end{aligned} \quad (32)$$

where $\mathbf{T}_k$ plays the role of some kind of total Tx side channel correlation matrix. Note that the weighting coefficients z, w depend on the BFs also though. The EWSR expression in (30) can be maximized alternatingly over the $\{\mathbf{g}_k\}$, the $\{z_k, w_k\}$ and the $\{z_{\overline{k}}, w_{\overline{k}}\}$. For the optimization of the BFs $\mathbf{g}_k$, for given

z, w, introduce

$$\begin{aligned}\widehat{\widehat{\mathbf{R}}}_{k,\overline{k}} &= I + \widehat{\mathbf{Q}}_{\overline{k}}\, \mathbf{T}_k(z_k, w_k)\\ \widehat{\widehat{\mathbf{R}}}_{\overline{k}} &= I + \widehat{\mathbf{Q}}_{\overline{k}}\, \mathbf{T}_k(z_{\overline{k}}, w_{\overline{k}})\\ \widehat{\widehat{\mathbf{R}}}_{k} &= I + \widehat{\mathbf{Q}}_{k}\, \mathbf{T}_k(z_k, w_k)\ .\end{aligned} \tag{33}$$

Then the generzalized eigenvector approach and the interference aware WF of the perfect CSI case can be applied with

$$\begin{aligned}\widehat{\mathbf{B}}_k &= \mathbf{T}_k(z_k, w_k)\, \widehat{\widehat{\mathbf{R}}}_{k,\overline{k}}^{-1}\\ \widehat{\mathbf{A}}_k &= \sum_{i\neq k} u_i \left[\mathbf{T}_i(z_i, w_i)\, \widehat{\widehat{\mathbf{R}}}_{i}^{-1} - \mathbf{T}_i(z_{\overline{i}}, w_{\overline{i}})\, \widehat{\widehat{\mathbf{R}}}_{\overline{i}}^{-1}\right] .\end{aligned} \tag{34}$$

Note that in spite of their appearance, matrices of the form $\mathbf{T}\mathbf{R}^{-1}$ are symmetric. Indeed, if e.g. $\mathbf{T}$ is invertible then $\mathbf{T}(I + \mathbf{Q}\mathbf{T})^{-1} = (\mathbf{T}^{-1} + \mathbf{Q})^{-1}$. For the optimization $\max\limits_{z\geq 0, w\geq 0} \{\ln\det \mathbf{S}(\mathbf{Q}, z, w) - zw\}$, we get from the extremum conditions

$$\begin{aligned}w &= f(\mathbf{Q}, z, w) = \mathrm{tr}\{\mathbf{Q}\mathbf{C}_t[I + \mathbf{Q}\mathbf{T}(z, w)]^{-1}\}\\ z &= g(\mathbf{Q}, z, w) = \mathrm{tr}\{\mathbf{C}_r[I + w\mathbf{C}_r\overline{\mathbf{H}}(I + z\mathbf{Q}\mathbf{C}_t)^{-1}\mathbf{Q}\overline{\mathbf{H}}^H]^{-1}\}\end{aligned} \tag{35}$$

which can be iterated until a fixed point.

For the simpler case of zero channel means $\overline{\mathbf{H}}_k = 0$ and no Rx side correlations $\mathbf{C}_r = I$, and with per user Tx side correlations $\mathbf{C}_k$, the EWSR can be rewritten with large MIMO asymptotics as

$$\begin{aligned}&EWSR =\\ &\sum_{k=1}^{K}\Big\{u_k \max_{z_k, w_k} [\ln\det(I + z_k\mathbf{Q}\mathbf{C}_k) + N\ln(1 + w_k) - z_k w_k]\\ &-u_k \max_{z_{\overline{k}}, w_{\overline{k}}} \left[\ln\det(I + z_{\overline{k}}\mathbf{Q}_{\overline{k}}\mathbf{C}_k) + N\ln(1 + w_{\overline{k}}) - z_{\overline{k}} w_{\overline{k}}\right]\Big\} .\end{aligned} \tag{36}$$

This criterion can be used to evaluate the EWSR for given $\mathbf{Q}$. It can also be used to optimize $\mathbf{Q}$, in which case we can apply again the algorithm of Section III, with the following conventions:

$$\begin{aligned}&\widehat{\widehat{\mathbf{R}}}_k(z) = I + z\, \widehat{\mathbf{Q}}\, \mathbf{C}_k\ ,\ \widehat{\widehat{\mathbf{R}}}_{\overline{k}}(z) = I + z\, \widehat{\mathbf{Q}}_{\overline{k}}\, \mathbf{C}_k\\ &\widehat{\mathbf{B}}_k = z_k \mathbf{C}_k \widehat{\widehat{\mathbf{R}}}_{\overline{k}}^{-1}(z_k)\ ,\ \widehat{\mathbf{A}}_k = \sum_{i\neq k} u_i \mathbf{C}_i (z_{\overline{i}} \widehat{\widehat{\mathbf{R}}}_{\overline{i}}^{-1}(z_{\overline{i}}) - z_i \widehat{\widehat{\mathbf{R}}}_{i}^{-1}(z_i))\end{aligned} \tag{37}$$

where z_k, $z_{\overline{k}}$ are obtained from

$$\begin{aligned}&\max_{z_k, w_k} g(z_k, w_k, \mathbf{Q}, \mathbf{C}_k)\ ,\ \max_{z_{\overline{k}}, w_{\overline{k}}} g(z_{\overline{k}}, w_{\overline{k}}, \mathbf{Q}_{\overline{k}}, \mathbf{C}) \quad \text{where}\\ &g(z, w, \mathbf{Q}, \mathbf{C}) = \ln\det(I + z\mathbf{Q}\mathbf{C}) + N\ln(1 + w) - z\, w\ .\end{aligned} \tag{38}$$

For these optimizations, we get from $\partial g/\partial w = 0$ that $z = N/(1 + w)$. From this and $\partial g/\partial z = 0$ we get

$$w = f(w) = \frac{1 + w}{N}\mathrm{tr}\{(\frac{1 + w}{N} I_M + \mathbf{Q}\mathbf{C})^{-1}\mathbf{Q}\mathbf{C}\} \tag{39}$$

The curves $y = w$ and $y = f(w)$ have a unique intersection in the first quadrant, with $y = f(w)$ lying initially above $y = w$. Hence the optimal w can be found by iterating $w^{(i)} = f(w^{(i-1)})$. The first time, one can initialize with $w^{(0)} = 0$. In the iterative algorithm from Section III, w can be initialized with the value obtained in the previous iteration for $\mathbf{g}$. The corresponding optimal z is then $z = N/(1 + w)$.

VIII. Centralized IBC Design

We shall first consider centralized optimization approaches (as in e.g. cloud RAN) in which CSIT information and BF computation gets centralized. In this case we consider the BSs in one cluster, and the users they serve, and one cluster forms an IBC. In a centralized IBC design, the CSIT training and feedback (for the FDD case) leads to a CSIT acquisition overhead that remains linear in C, the number of BS [2]. Also in the TDD case, though in this case only uplink training is required. The proper partial CSIT model instance in this case will have a strong non-zero mean $\overline{\mathbf{H}}$ which represents the channel estimate, and the weaker covariance part represents estimation (and feedback) noise. In practice however, the cluster will not be isolated, and the intercluster interference can be handled similarly to intercell interference. However, the intercluster CSIT may be more degraded, and is possibly only composed of Tx side covariance information. Note that in the MaMIMO regime, strong direct link CSIT with zero-mean cross link CSIT induces quite some structure e.g. in the computation of the matrices $\breve{\mathbf{A}}_k$ (see (28)) and leads to even more simplification in the Interference Channel (not IBC) case. The MaMIMO IBC with zero mean and $\mathbf{C}_r = I$ intercell CSIT scenario leads to the same rate expressions as if the intercell interference is modeled as Gaussian while preserving the covariance matrix:

$$\begin{aligned}&\mathbf{y}_k = \mathbf{H}_{k,b_k}\, \mathbf{g}_k\, x_k + \textstyle\sum_{i:b_i=b_k} \mathbf{H}_{k,b_k}\, \mathbf{g}_i\, x_i + \widetilde{\mathbf{v}}_k\\ &\mathbf{C}_{\widetilde{\mathbf{v}}_k\widetilde{\mathbf{v}}_k} = (1 + \textstyle\sum_{i:b_i\neq b_k} \mathbf{g}_i^H \mathbf{C}_{t,k,b_i} \mathbf{g}_i)\, I\ .\end{aligned} \tag{40}$$

In general, modeling the intercell interference plus noise as a Gaussian noise $\widetilde{\mathbf{v}}_k$ would lead to a mutual information lower bound. Also the large MIMO formulation of Section VII can be used.

IX. Distributed IBC Design

In this case, global intracluster CSIT can also be gathered but it takes an overhead that evolves with C^2 [2]. Hence, high quality (high Ricean factor) intracell CSIT and Tx covariance only intercell CSIT may be a more appropriate setting. For what follows we shall assume the LoS Tx intercell CSIT. We shall focus on a MaMIMO setting.

A. Initialization Phase

The approach considered here is non-iterative, or could be taken as initialization for further iterations.

1) Iteration 0: To properly gauge the intercell interference caused by the other BS, we shall start with a per cell design. In the case of multiple Tx and Rx antennas, different WSR local optima correspond at high SNR to different distributions of the zero forcing (ZF) roles between the various Txs and Rxs. To simplify design, we shall assume here that Rx antennas are used to handle intracell interference. Hence all intercell interference needs to be handled by Tx (BS) antennas. In that case, the crosslinks (cascades of channel and Rx) can be considered as independent from the intracell channels. In a MaMIMO setting, the ZF by BS j towards $K-K_j$ crosslink channels (or LoS components in fact) will tend to have a deterministic effect of reducing the effective number of Tx antennas by this amount and hence of reducing the Tx power by a factor $\frac{M_j}{M_j-(K-K_j)}$. Hence a per BS design can be carried out with (partial) intracell CSIT, with BS Tx power P_j replaced by $\frac{M_j}{M_j-(K-K_j)}P_j$, and with all intercell links $\mathbf{H}_{k,b_i}=0$, $b_i \neq b_k$. The power reduction factor considered corresponds to ZF or hence a high SNR assumption. At a finite SNR, the user k SINR reduction will be less because the optimal BF would do some regularized ZF of intercell interference caused, but the SINR would on the other hand be smaller because the intercell interference is not ZF'd. We shall assume that these two opposite effects roughly compensate each other. This first step (which is itself an iterative design for the scenario considered with reduced Tx power and no intercell links) leads to BFs $\mathbf{g}^{(0)}$ which, from (24), lead to

$$\breve{\mathbf{R}}_k = (1+\text{tr}\{\mathbf{Q}_{b_k}^{(0)}\mathbf{C}_{t,k,b_k}\})I_{N_k} + \overline{\mathbf{H}}_{k,b_k}\mathbf{Q}_{b_k}^{(0)}\overline{\mathbf{H}}_{k,b_k}^H\,, \\ \text{where} \quad \mathbf{Q}_{b_k}^{(0)} = \sum_{i:b_i=b_k} \mathbf{g}_i^{(0)}\mathbf{g}_i^{(0)H} \tag{41}$$

and similarly for $\breve{\mathbf{R}}_{\overline{k}}$.

2) Iteration 1*:* The idea is now to do one iteration in order to adjust the Tx filters for the intercell interference. So, with the initial BFs $\mathbf{g}^{(0)}$, the local intercell CSIT $\mathbf{C}_{t,i,b_k}$ also, the correct power constraints, and $\breve{\mathbf{R}}_k$, $\breve{\mathbf{R}}_{\overline{k}}$ as in (41), we get $\breve{\mathbf{B}}_k$ as in (26) and $\breve{\mathbf{A}}_k$ from (27) becomes

$$\begin{aligned}\breve{\mathbf{A}}_k = &\sum_{i\neq k: b_i=b_k} u_i \Big[\overline{\mathbf{H}}_{i,b_k}^H\left(\breve{\mathbf{R}}_{\overline{i}}^{-1}-\breve{\mathbf{R}}_i^{-1}\right)\overline{\mathbf{H}}_{i,b_k} \\ &\qquad +\text{tr}\{\left(\breve{\mathbf{R}}_{\overline{i}}^{-1}-\breve{\mathbf{R}}_i^{-1}\right)\mathbf{C}_{r,i}\}\mathbf{C}_{t,i,b_k}\Big] \\ &+\sum_{i:b_i\neq b_k}\underbrace{u_i\text{tr}\{\breve{\mathbf{R}}_{\overline{i}}^{-1}-\breve{\mathbf{R}}_i^{-1}\}}_{=\mu_i}\mathbf{C}_{t,i,b_k}\,.\end{aligned} \tag{42}$$

Hence the only information that needs to be fed back from user i in another cell is the positive scalar μ_i. This is related to the interference pricing in game theory [17]. The normalized BFs are then computed as $\mathbf{g}_k^{'}=V_{max}(\breve{\mathbf{B}}_k,\breve{\mathbf{A}}_k+\lambda_{b_k}I)$ where the λ_{b_k} are taken from the previous iteration. The stream powers are obtained from (13).

B. Continuous Adaptation

The solution discussed above should be adequate in many scenarios. Further refinements can be considered. For instance, staying with covariance intercell CSIT but going beyond LoS, the subspaces of the $\mathbf{C}_t$ to be accounted for in the initialization stage would be more than one dimensional, leading to a further power reduction in the initial per cell design. Further iterations beyond the first iteration discussed above require for each iteration the feedback of the scalars μ_i characterizing the intercell links, and within each cell, the feedback of the interference plus noise whitened channels $\breve{\mathbf{R}}_{\overline{k}}^{-1/2}\overline{\mathbf{H}}_{k,b_k}$ and the scalars $\text{tr}\{\mathbf{C}_{r,k}\breve{\mathbf{R}}_{\overline{k}}^{-1}\}$.

Acknowledgments

EURECOM's research is partially supported by its industrial members: ORANGE, BMW Group, SFR, ST Microelectronics, Symantec, SAP, Monaco Telecom, iABG, and also by the EU FP7 projects ADEL and NEWCOM#.

References

[1] F. Negro, I. Ghauri, and D.T.M. Slock, "Deterministic Annealing Design and Analysis of the Noisy MIMO Interference Channel," in *Proc. IEEE Information Theory and Applications workshop (ITA)*, San Diego, CA, USA, Feb. 2011.

[2] F. Negro, D.T.M. Slock, and I. Ghari, "On the Noisy MIMO Interference Channel with CSI through Analog Feedback," in *Int'l Symp. Comm's, Control and Sig. Proc. (ISCCSP)*, Rome, Italy, May 2012.

[3] Y. Lejosne, D. Slock, and Y. Yuan-Wu, "Foresighted Delayed CSIT Feedback for Finite Rate of Innovation Channel Models and Attainable NetDoFs of the MIMO Interference Channel," in *Proc. Wireless Days*, Valencia, Spain, Nov. 2013.

[4] E.G. Larsson, O. Edfors, F. Tufvesson, and T.L. Marzetta, "Massive MIMO for Next Generation Wireless Systems," *IEEE Comm's Mag.*, Feb. 2014.

[5] S. Wagner, R. Couillet, M. Debbah, and D. Slock, "Large System Analysis of Linear Precoding in Correlated MISO Broadcast Channels Under Limited Feedback," *IEEE Trans. Info. Theory*, Jul. 2012.

[6] A. Adhikary, J. Nam, J.-Y. Ahn, and G. Caire, "Joint Spatial Division and Multiplexing: the Large-Scale Array Regime," *IEEE Trans. Info. Theory*, Oct. 2013.

[7] A. Adhikary, E. El Sadafi, M. Samimi, R. Wang, G. Caire, T. Rappaport, and A. Molisch, "Joint Spatial Division and Multiplexing for mm-Wave Channels," [Online]. Available: http://arxiv.org/abs/1312.2045.

[8] A. Liu and V.K.N. Lau, "Hierarchical Interference Mitigation for Large MIMO Cellular Networks," *IEEE Trans. Signal Processing*, Apr. 2013.

[9] J. Brady and A. Sayeed, "Beamspace MU-MIMO for High-Density Gigabit Smallm Cell Access at Millimeter-Wave Frequencies," in *Proc. IEEE Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*, Toronto, Canada, Jun. 2014.

[10] A. Sayeed, "Deconstructing Multiantenna Fading Channels," *IEEE Trans. Sig. Proc.*, Oct. 2002.

[11] F. Negro, M. Cardone, I. Ghauri, and D.T.M. Slock, "SINR Balancing and Beamforming for the MISO Interference Channel," in *Proc. IEEE Int'l Symp. Personal Indoor and Mobile Radio Communications (PIMRC)*, Sept. 2011.

[12] R. de Francisco and D.T.M. Slock, "Spatial Transmit Prefiltering for Frequency-Flat MIMO Transmission with Mean and Covariance Information," in *IEEE Asilomar Conf, on Signals, Systems, and Computers*, Pacific Grove, US, Oct. 2005.

[13] H. Yin, D. Gesbert, M. Filippou, and Y. Liu, "A coordinated approach to channel estimation in large-scale multiple-antenna systems," *IEEE J. on Selected Areas in Comm's (JSAC), Special Issue on: Large Scale Antenna Systems*, Jan. 2013.

[14] F. Negro, I. Ghauri, and D.T.M. Slock, "Sum Rate Maximization in the Noisy MIMO Interfering Broadcast Channel with Partial CSIT via the Expected Weighted MSE," in *IEEE Int'l Symp. on Wireless Commun. Systems (ISWCS)*, Paris, France, Aug. 2012.

[15] M. Razaviyayn, M.S. Boroujeni, and Z.-Q. Luo, "A Stochastic Weighted MMSE Approach to Sum Rate Maximization for a MIMO Interference Channel," in *IEEE workshop on Signal Processing Advances in Wireless Communications (SPAWC)*, Darmstadt, Germany, June 2013.

[16] Y. Huang, G. Zheng, M. Bengtsson, K.K. Wong, L. Yang, B. Ottersten, "Distributed Multicell Beamforming Design Approaching Pareto Boundary with Max-Min Fairness," *IEEE T. Wireless Comm.*, Aug. 2012.

[17] W. Xu and X. Wang, "Pricing-Based Distributed Downlink Beamforming in Multi-Cell OFDMA Networks," *IEEE J. Sel. Areas Comm.*, Oct. 2012.

[18] S.-J. Kim and G.B. Giannakis, "Optimal Resource Allocation for MIMO Ad Hoc Cognitive Radio Networks," *IEEE T. Info. Theory*, May 2011.

[19] J. Dumont, W. Hachem, S. Lasaulce, P. Loubaton, and J. Najim, "On the Capacity Achieving Covariance Matrix for Rician MIMO Channels: An Asymptotic Approach," *IEEE Trans. Info. Theory*, Mar. 2010.

[20] G. Taricco, "Asymptotic mutual information statistics of separately correlated Rician fading MIMO channels," *IEEE Trans. Info. Theory*, Aug. 2008.