Beamforming Design with Combined Channel Estimate and Covariance CSIT via Random Matrix Theory

Wassim Tabikh[‡¶], Yi Yuan-Wu[¶], Dirk Slock[‡],
[‡]EURECOM, Sophia-Antipolis, France, Email: {slock,wassim.tabikh}@eurecom.fr
[¶]Orange Labs, Issy-les-Moulineaux, France, Email: {yi.yuan,wassim.tabikh}@orange.com

***Abstract*—The Interfering Broadcast Channel (IBC) applies to the downlink of (cellular and/or heterogeneous) 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 transmission. 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 MU systems, is always imperfect in practice. We consider the joint optimal exploitation of mean (channel estimates) and covariance Gaussian partial CSIT. Indeed, in a Massive MIMO (MaMIMO) setting (esp. when combined with mmWave) the channel covariances may exhibit low rank and zero-forcing might be possible by just exploiting the covariance subspaces. But the question is the optimization of beamformers for the expected weighted sum rate (EWSR) at finite SNR. We propose explicit beamforming solutions and indicate that existing large system analysis can be extended to handle optimized beamformers with the more general partial CSIT considered here.**

***Index Terms*—Massive MIMO; multi-user; multi-cell; sum rate; beamforming; partial CSIT; large system analysis**

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 User Equipments (UEs) simultaneously. However, MU systems have precise requirements for CSIT which is more difficult to acquire than CSI at the Rx (CSIR). Hence we focus here on the more challenging downlink (DL) and we talk about the so-called maximizing the weighted sum rate with partial CSIT. Earlier works have attempted optimal partial CSIT designs for multi-user MIMO, e.g. the Expected Weighted Sum MSE (EWSMSE) approach applied in [1] for MIMO IBC BF design. However, the EWSMSE approach is suboptimal and cannot even be used in the zero channel mean case (case of covariance CSIT only). In spite of that, it has been mistakenly considered optimal as recently as in [2]. We treat the Gaussian CSIT case with both mean and covariance information. The goal here is to go beyond the extreme of ZF and to introduce a meaningful BF design at finite SNR and with partial CSIT. A significant push for large system analysis in MaMIMO systems appeared in [3]. It allows to obtain deterministic (instead of fast fading channel dependent) expressions for various scalar quantities, facilitating the analysis of wireless systems. E.g. it may allow to evaluate beamforming performance without computing explicit beamformers. The analysis in [3] allowed e.g. the determination of the optimal regularization factor in R-ZF BF. A little known extension appeared in [4] for optimal beamformers, but only for the perfect CSIT case. Some other extensions appeared recently in [5] or [6] where MISO IBC is considered with perfect CSIT and weighted R-ZF BF, with two optimized weight levels, for intracell or intercell interference.

The contributions of this paper are: a new look at Rx/Tx design with partial CSI, derivation of BF design exploiting both mean and covariance CSIT for the MIMO IBC, which becomes optimal in the MaMIMO limit.

II. The IBC Signal Model

Let us consider an IBC system 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

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

where $\mathbf{x}_k$ are the intended $d_k \times 1$ signals (each white and unit variance), d_k is the number of intended streams, $\mathbf{H}_{k,b_k}$ 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 consider the noise as $\mathbf{v}_k \sim \mathcal{CN}(0, \sigma^2 I_{N_k})$. The $M_{b_k} \times d_k$ spatial Tx filter or beamformer (BF) is $\mathbf{G}_k$.

III. Max WSR with Perfect CSIT

This section stems entirely from [7]. Consider as a starting point for the optimization the weighted sum rate (WSR)

$$WSR = WSR(\mathbf{Q}) = \sum_{k=1}^{K} u_k \ \ln\det(\mathbf{I}_{N_k} + \mathbf{R}_{\overline{k}}^{-1}\mathbf{H}_{k,b_k}\mathbf{Q}_k\mathbf{H}_{k,b_k}^H) \tag{2}$$

where $\mathbf{Q}$ represents the collection of transmit covariance matrices $\mathbf{Q}_k$, the u_k are rate weights

$$\begin{aligned}\mathbf{R}_k &= \mathbf{H}_{k,b_k}\mathbf{Q}_k\,\mathbf{H}^H_{k,b_k} + \mathbf{R}_{\overline{k}}\,,\ \mathbf{Q}_i = \mathbf{G}_i\mathbf{G}_i^H\,,\\ \mathbf{R}_{\overline{k}} &= \sum_{i\neq k}\mathbf{H}_{k,b_i}\mathbf{Q}_i\,\mathbf{H}^H_{k,b_i} + \sigma^2\mathbf{I}_{N_k}\,.\end{aligned} \tag{3}$$

$\mathbf{R}_k$, $\mathbf{R}_{\overline{k}}$ are the total and the interference plus noise Rx covariance matrices resp. The WSR cost function needs to be augmented with the power constraints

$$\sum_{k:b_k=j}\text{tr}\{\mathbf{Q}_k\} \leq P_j^{BS}\,. \tag{4}$$

So our optimization problem can be expressed as the following:

$$\begin{aligned}&\max_{Q} WSR(\mathbf{Q})\\ s.t. &\sum_{k:b_k=j}\text{tr}\{\mathbf{Q}_k\} \leq P_j^{BS}\end{aligned} \tag{5}$$

where $WSR(\mathbf{Q})$ is given in (2). In a classical difference of convex functions (DC programming) approach, Kim and Giannakis [7] 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

$$\begin{aligned}WSR &= u_k\ln\det(\mathbf{R}_{\overline{k}}^{-1}\mathbf{R}_k) + WSR_{\overline{k}}\,,\\ WSR_{\overline{k}} &= \textstyle\sum_{i=1,\neq k}^{K} u_i\ln\det(\mathbf{R}_{\overline{i}}^{-1}\mathbf{R}_i)\end{aligned} \tag{6}$$

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

$$\begin{aligned}&WSR_{\overline{k}}(\mathbf{Q}_k,\widehat{\mathbf{Q}}) \approx WSR_{\overline{k}}(\widehat{\mathbf{Q}}_k,\widehat{\mathbf{Q}}) - \text{tr}\{(\mathbf{Q}_k-\widehat{\mathbf{Q}}_k)\widehat{\mathbf{A}}_k\}\\ &\widehat{\mathbf{A}}_k = -\left.\frac{\partial WSR_{\overline{k}}(\mathbf{Q}_k,\widehat{\mathbf{Q}})}{\partial \mathbf{Q}_k}\right|_{\widehat{\mathbf{Q}}_k,\widehat{\mathbf{Q}}} = \sum_{i\neq k}^{K} u_i\mathbf{H}^H_{i,b_k}(\widehat{\mathbf{R}}_{\overline{i}}^{-1} - \widehat{\mathbf{R}}_i^{-1})\mathbf{H}_{i,b_k}\end{aligned} \tag{7}$$

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

$$\begin{aligned}&WSR(\mathbf{G},\hat{\mathbf{G}},\lambda) = \sum_{j=1}^{C}\lambda_j P_j^{BS} +\\ &\sum_{k=1}^{K} u_k\ln\det(I_{d_k} + \mathbf{G}_k^H\widehat{\mathbf{B}}_k\mathbf{G}_k) - \text{tr}\{\mathbf{G}_k^H(\widehat{\mathbf{A}}_k + \lambda_{b_k}I_{M_{b_k}})\mathbf{G}_k\}\end{aligned} \tag{8}$$

where
$$\widehat{\mathbf{B}}_k = \mathbf{H}^H_{k,b_k}\widehat{\mathbf{R}}_{\overline{k}}^{-1}\mathbf{H}_{k,b_k}\,. \tag{9}$$

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 eigenmatrix condition, thus $\mathbf{G}_k^{'} = eigenmatrix(\widehat{\mathbf{B}}_k, \widehat{\mathbf{A}}_k + \lambda_{b_k}I_{M_{b_k}})$ is the (normalized) generalized eigenmatrix of the two indicated matrices, with eigenvalues $\Sigma_k = eigenvalues(\widehat{\mathbf{B}}_k, \widehat{\mathbf{A}}_k + \lambda_{b_k}I_{M_{b_k}})$. 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 (8) is that it allows straightforward power adaptation: introducing diagonal power matrices $\mathbf{P}_k \geq 0$ and substituting $\mathbf{G}_k = \mathbf{G}_k^{'}\mathbf{P}_k^{\frac{1}{2}}$ in (8) yields

$$\begin{aligned}WSR = \sum_{j}^{C}\lambda_j P_j^{BS} + \sum_{k=1}^{K}\Big[u_k\ln\det\Big(\mathbf{I}_{d_k} + \mathbf{P}_k\Sigma_k^{(1)}\Big)\\ - \text{tr}\{\mathbf{P}_k(\Sigma_k^{(2)} + \lambda_{b_k}I)\}\Big]\end{aligned} \tag{10}$$

which leads to the following interference leakage aware water filling

$$\mathbf{P}_k(l,l) = \left(\frac{1}{\Sigma_k^{(1)}(l,l)}\left(\frac{u_k\Sigma_k^{(1)}(l,l)}{\Sigma_k^{(2)}(l,l)+\lambda_{b_k}} - 1\right)\right)^{+} \tag{11}$$

for all l s.t. $\Sigma_k^{(1)} > 0$ where $z^+ = max(0,z)$ and the Lagrange multipliers λ_{b_k} are adjusted to satisfy the power constraints $\sum_{k:b_k=j}\sum_{l=1}^{d_k}\mathbf{P}_k(l,l) = P_j^{BS}$. 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 $\mathbf{P}_k$ and the λ_l. The advantage of the DC approach is that it works for any number of streams/user d_k, by simply taking more eigenvectors. In other words, we can take the d_k max eigenvectors of the eigenmatrix $\mathbf{G}_k^{'}$. We mean by the max eigenvectors, the eigenvectors corresponding to the highest eigenvalues. The waterfilling then automatically determines (at each iteration) how many streams can be sustained.

IV. Joint Mean and Covariance Gaussian CSIT

In this section we drop the user index k for simplicity. Assume that the channel has a (prior) Gaussian distribution with zero mean and separable correlation model

$$\mathbf{H} = \mathbf{C}_r^{1/2}\,\widetilde{\mathbf{H}}\,\mathbf{C}_t^{1/2} \tag{12}$$

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

$$\begin{aligned}\mathrm{E}\,\mathbf{H}\,\mathbf{H}^H &= \text{tr}\{\mathbf{C}_t\}\,\mathbf{C}_r\\ \mathrm{E}\,\mathbf{H}^H\mathbf{H} &= \text{tr}\{\mathbf{C}_r\}\,\mathbf{C}_t\end{aligned} \tag{13}$$

Now, the Tx dispose of a (deterministic) channel estimate

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

where again the elements of $\widetilde{\mathbf{H}}_d$ are i.i.d. $\sim\mathcal{CN}(0,1)$, and typically $\mathbf{C}_d = \sigma^2_{\widetilde{\mathbf{H}}}I_M$. The combination of the estimate with the prior information leads to the (posterior) LMMSE estimate

$$\begin{aligned}\overline{\mathbf{H}} &= \hat{\mathbf{H}}_d\,(\mathbf{C}_t+\mathbf{C}_d)^{-1}\mathbf{C}_t = \mathbf{H} + \mathbf{C}_r^{1/2}\,\widetilde{\mathbf{H}}_p\,\mathbf{C}_p^{1/2}\\ \mathbf{C}_p &= \mathbf{C}_d\,(\mathbf{C}_t+\mathbf{C}_d)^{-1}\mathbf{C}_t\end{aligned} \tag{15}$$

where $\widetilde{\mathbf{H}}$ and $\widetilde{\mathbf{H}}_p$ are independent. Note that we get for the MMSE estimate of a quadratic quantity of the form

$$\mathrm{E}_{\mathbf{H}|\hat{\mathbf{H}}_d}\mathbf{H}^H\mathbf{H} = \overline{\mathbf{H}}^H\overline{\mathbf{H}} + \text{tr}\{\mathbf{C}_r\}\mathbf{C}_p = \mathbf{S}\,. \tag{16}$$

Let us emphasize that this MMSE estimate implies $S = \arg\min_{\mathbf{T}} \mathrm{E}_{\mathbf{H}|\hat{\mathbf{H}}_d} ||\mathbf{H}^H\mathbf{H} - \mathbf{T}||^2$. It averages out to

$$\mathrm{E}_{\hat{\mathbf{H}}_d}\mathbf{S} = \mathrm{E}_{\mathbf{H},\hat{\mathbf{H}}_d}\mathbf{H}^H\mathbf{H} = \mathrm{E}_{\mathbf{H}}\mathbf{H}^H\mathbf{H} = \mathrm{tr}\{\mathbf{C}_r\}\mathbf{C}_t \ . \tag{17}$$

Hence, if we want the best estimate for $\mathbf{H}^H\mathbf{H}$ (which appears in the signal or interference powers), it is not sufficient to replace $\mathbf{H}$ by $\overline{\mathbf{H}}$ but also the covariance information should be exploited. Note that $\mu = \frac{\mathrm{tr}\{\overline{\mathbf{H}}^H\overline{\mathbf{H}}\}}{\mathrm{tr}\{\mathbf{C}_r\}\mathrm{tr}\{\mathbf{C}_p\}}$ is a form of Ricean factor that represents channel estimation quality.

V. Expected WSR (EWSR)

For the WSR criterion, we have assumed so far that the channel $\mathbf{H}$ is known. Once the CSIT is imperfect, various optimization criteria could be considered, such as outage capacity. Here we shall consider the expected weighted sum rate $\mathrm{E}_{\mathbf{H}} WSR(\mathbf{Q},\mathbf{H}) =$

$$EWSR(\mathbf{Q}) = \mathrm{E}_{\mathbf{H}} \sum_k u_k \ln\det(I_{M_{b_k}} + \mathbf{H}^H_{k,b_k}\mathbf{R}^{-1}_{\overline{k}}\mathbf{H}_{k,b_k}\mathbf{Q}_k) \tag{18}$$

VI. MaMIMO Limit

In the MaMIMO limit where the number of Tx antennas M becomes very large, 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

$$\mathbf{H}\mathbf{Q}\mathbf{H}^H \overset{M\to\infty}{\longrightarrow} \mathrm{E}_{\mathbf{H}}\, \mathbf{H}\mathbf{Q}\mathbf{H}^H = \overline{\mathbf{H}}\mathbf{Q}\overline{\mathbf{H}}^H + \mathrm{tr}\{\mathbf{Q}\mathbf{C}_p\}\,\mathbf{C}_r \, . \tag{19}$$

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

$$\begin{aligned}
&\mathbf{H}_k = [\mathbf{H}_{k,1}\cdots\mathbf{H}_{k,C}] = \overline{\mathbf{H}}_k - \mathbf{C}^{1/2}_{r,k}\widetilde{\mathbf{H}}_k\mathbf{C}^{1/2}_{p,k}\\
&\mathbf{Q} = \begin{bmatrix} \sum\limits_{i:b_i=1}\mathbf{Q}_i & & \\ & \ddots & \\ & & \sum\limits_{i:b_i=C}\mathbf{Q}_i \end{bmatrix} = \sum_{j=1}^{C}\sum_{i:b_i=j}\mathbf{I}_j\mathbf{Q}_i\mathbf{I}_j^H\\
&\mathbf{Q}_{\overline{k}} = \mathbf{Q} - \mathbf{I}_{b_k}\mathbf{Q}_k\mathbf{I}^H_{b_k}
\end{aligned} \tag{20}$$

where $\mathbf{C}_{p,k} = \mathrm{blockdiag}\{\mathbf{C}_{p,k,1},\ldots,\mathbf{C}_{p,k,C}\}$, and $\mathbf{I}_j$ is an all zero block vector except for an identity matrix in block j. Using (19), let us define

$$\begin{aligned}
\breve{\mathbf{R}}_k &= \sigma^2\mathbf{I}_{N_k} + \overline{\mathbf{H}}_k\mathbf{Q}\overline{\mathbf{H}}_k^H + \mathrm{tr}\{\mathbf{Q}\mathbf{C}_{p,k}\}\,\mathbf{C}_{r,k}\\
\breve{\mathbf{R}}_{\overline{k}} &= \sigma^2\mathbf{I}_{N_k} + \overline{\mathbf{H}}_k\mathbf{Q}_{\overline{k}}\overline{\mathbf{H}}_k^H + \mathrm{tr}\{\mathbf{Q}_{\overline{k}}\mathbf{C}_{p,k}\}\,\mathbf{C}_{r,k}
\end{aligned} \tag{21}$$

which represent the total and the interference plus noise Rx covariance matrices in the MaMIMO regime respectively. This leads to

$$WSR = u_k \ln\det(\mathbf{I}_{M_{b_k}} + \mathbf{H}^H_{k,b_k}\breve{\mathbf{R}}^{-1}_{\overline{k}}\mathbf{H}_{k,b_k}\mathbf{Q}_k) + WSR_{\overline{k}} \tag{22}$$

Thus,

$$\begin{aligned}
&WSR = u_k \ln\det(\mathbf{I}_{d_k} + \mathbf{G}_k^H\breve{\mathbf{B}}_k\mathbf{G}_k) + WSR_{\overline{k}} \quad \text{with}\\
&\breve{\mathbf{B}}_k = \mathrm{E}_{\mathbf{H}|\overline{\mathbf{H}}}\mathbf{H}^H_{k,b_k}\breve{\mathbf{R}}^{-1}_{\overline{k}}\mathbf{H}_{k,b_k}\\
&\quad\; = \overline{\mathbf{H}}^H_{k,b_k}\breve{\mathbf{R}}^{-1}_{\overline{k}}\overline{\mathbf{H}}_{k,b_k} + \mathrm{tr}\{\mathbf{C}_{r,k}\breve{\mathbf{R}}^{-1}_{\overline{k}}\}\mathbf{C}_{p,k,b_k}
\end{aligned} \tag{23}$$

which is now similar to the corresponding expression 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 $\breve{\mathbf{R}}^{-1/2}_{\overline{k}}\mathbf{H}_{k,b_k}$. The linearization $\breve{\mathbf{A}}_k$ of $WSR_{\overline{k}}$ w.r.t. $\mathbf{Q}_k$ (7) gives in the MaMIMO regime:

$$\begin{aligned}
\breve{\mathbf{A}}_k = \sum_{i\neq k}^{K} u_i[&\breve{\mathbf{A}}^C_{i,k}(\mathbf{I}_{M_{b_k}} + \mathbf{Q}_k\breve{\mathbf{A}}^C_{i,k})^{-1}\\
&- \breve{\mathbf{A}}^D_{i,k}(\mathbf{I}_{M_{b_k}} + \mathbf{Q}_k\breve{\mathbf{A}}^D_{i,k})^{-1}]
\end{aligned} \tag{24}$$

with

$$\begin{aligned}
\breve{\mathbf{A}}^C_{i,k} &= \overline{\mathbf{H}}^H_{i,b_k}\breve{\mathbf{R}}^{-1}_{\overline{i},\overline{k}}\overline{\mathbf{H}}_{i,b_k} + \mathrm{tr}\{\breve{\mathbf{R}}^{-1}_{\overline{i},\overline{k}}\mathbf{C}_{r,i}\}\mathbf{C}_{p,i,b_k};\\
\breve{\mathbf{A}}^D_{i,k} &= \overline{\mathbf{H}}^H_{i,b_k}\breve{\mathbf{R}}^{-1}_{i,\overline{k}}\overline{\mathbf{H}}_{i,b_k} + \mathrm{tr}\{\breve{\mathbf{R}}^{-1}_{i,\overline{k}}\mathbf{C}_{r,i}\}\mathbf{C}_{p,i,b_k};\\
\breve{\mathbf{R}}_{\overline{i},\overline{k}} &= \sum_{j\neq i,j\neq k}^{K}\overline{\mathbf{H}}_{i,b_j}\mathbf{Q}_j\overline{\mathbf{H}}^H_{i,b_j} + \mathrm{tr}\{\mathbf{Q}_j\mathbf{C}_{p,i,b_j}\}\mathbf{C}_{r,i} + \sigma^2\mathbf{I}_{N_i};\\
\breve{\mathbf{R}}_{i,\overline{k}} &= \sum_{j\neq k}^{K}\overline{\mathbf{H}}_{i,b_j}\mathbf{Q}_j\overline{\mathbf{H}}^H_{i,b_j} + \mathrm{tr}\{\mathbf{Q}_j\mathbf{C}_{p,i,b_j}\}\mathbf{C}_{r,i} + \sigma^2\mathbf{I}_{N_i}.
\end{aligned}$$

The proof is given in the Appendix. As in section III, to get the normalized precoder we use

$$\mathbf{G}^{'}_k = eigenmatrix(\breve{\mathbf{B}}_k, \breve{\mathbf{A}}_k + \lambda_{b_k}I_{M_{b_k}}) \tag{25}$$

with eigenvalues $\Sigma_k = eigenvalues(\breve{\mathbf{B}}_k, \breve{\mathbf{A}}_k + \lambda_{b_k}I_{M_{b_k}})$. Let $\Sigma^{(1)}_k = \mathbf{G}^{'H}_k\breve{\mathbf{B}}_k\mathbf{G}^{'}_k$, $\Sigma^{(2)}_k = \mathbf{G}^{'H}_k\breve{\mathbf{A}}_k\mathbf{G}^{'}_k$. Powers $\mathbf{P}_k \geq 0$ are defined as in (11). λ_{b_k} is determined also as described in section III. We can take also only d_k max eigenvectors as described in section III. The algorithm can be then summarized as in Table I.

VII. Numerical Results

In this section, the performance of our proposed scheme, denoted by "with Combined Channel Estimate and Covariance CSIT", is evaluated through numerical simulations. We compare it to the "EWSMSE" approach in [1] and to the approach given in section III with channel estimate considered as true channel. The latter approach is denoted by "with Channel Estimate Only". For the initialization of our approach we use the transmit covariance matrix output from the approach "with Channel Estimate Only". Figures 1 and 2 show the achievable sum rate versus SNR for a Mutliple-Input Single-Output (MISO) system composed of 2 cells, with 8 single-antenna users in total and 8 antennas per BS , the Tx covariance matrices $\mathbf{C}_{t,i,b_k}\forall i,\forall k$ and $\mathbf{C}_{p,i,b_k}\forall i,\forall k$ are considered as identity matrices in Figure 1 and as low rank matrices (rank = 2) in Figure 2. The low-rank property of the correlation matrices is motivated in the work in [8]. Since we are in a

TABLE I: The Iterative EWSR Algorithm

For k = 1 ... K, initialize Q_k
Repeat until convergence
For j = 1 ... C
Set $\underline{\lambda} = 0$, $\overline{\lambda} = \lambda_{max}$
For k such that $b_k = j$
Compute $\breve{\mathbf{A}}_k$ using (25) and the interference terms using (21)
Next k
Repeat until convergence
$\lambda = \frac{1}{2}(\underline{\lambda} + \overline{\lambda})$
For k such that $b_k = j$
Compute $\breve{\mathbf{B}}_k$ using (23)
Compute the generalized eigenmatrix $\mathbf{G}_k$ of $\breve{\mathbf{B}}_k$ and $\breve{\mathbf{A}}_k + \lambda \mathbf{I}_M$
Normalize the generalized eigenmatrix so as to have $\mathbf{G}_k^{'}$
Compute $\Sigma_k^{(1)} = \mathbf{G}_k^{'H} \breve{\mathbf{B}}_k \mathbf{G}_k^{'}$, $\Sigma_k^{(2)} = \mathbf{G}_k^{'H} \breve{\mathbf{A}}_k \mathbf{G}_k^{'}$
Compute $\mathbf{P}_k$ as in (11)
Next k
Compute $\mathbf{P}$=$\sum_k \mathbf{P}_k$
if $\mathtt{tr}(\mathbf{P}) \geq P_j^{BS}$, set $\underline{\lambda} = \lambda$, otherwise set $\overline{\lambda} = \lambda$
For all k such that $b_k = j$, set $\mathbf{Q}_k = \mathbf{G}_k^{'} \mathbf{P}_k \mathbf{G}_k^{',H}$
Next j

MISO scenario, the Rx covariance scalars as considered as one by default. So for a user k we have the following:

$$\overline{\mathbf{h}}_{k,b_k} = \sqrt{\alpha^2}\left(\frac{M}{tr(\mathbf{C}_{t,k,b_k})}\right)^{\frac{1}{2}} \widetilde{\mathbf{h}}_{k,b_k} \mathbf{C}_{t,k,b_k}^{\frac{1}{2}} + \sqrt{(1-\alpha^2)}\left(\frac{M}{tr(\mathbf{C}_{p,k,b_k})}\right)^{\frac{1}{2}} \widetilde{\mathbf{h}}_{p,k,b_k} \mathbf{C}_{p,k,b_k}^{\frac{1}{2}}; \quad (26)$$

$$\mathbf{h}_{k,b_k} = \widetilde{\mathbf{h}}_{k,b_k} \mathbf{C}_{t,k,b_k}^{\frac{1}{2}}. \quad (27)$$

The model in (26) and (27) is the same as (15) and (12) but conceived in a way to preserve the channel gain. In other words the real channel $\mathbf{h}_{k,b_k}$ in (27) and the channel estimate $\overline{\mathbf{h}}_{k,b_k}$ in (26) have the same gain. Proof:

$$E((\alpha^2)\frac{M^2}{tr(\mathbf{C}_{t,k,b_k})^2} \widetilde{\mathbf{h}}_{k,b_k} \mathbf{C}_{t,k,b_k} \widetilde{\mathbf{h}}_{k,b_k}^H) = (\alpha^2)\frac{M}{tr(\mathbf{C}_{t,k,b_k})} tr(\mathbf{C}_{t,k,b_k}) = (\alpha^2)M; \quad (28)$$

$$E((1-\alpha^2)\frac{M^2}{tr(\mathbf{C}_{p,k,b_k})^2} \widetilde{\mathbf{h}}_{p,k,b_k} \mathbf{C}_{p,k,b_k} \widetilde{\mathbf{h}}_{p,k,b_k}^H) = (1-\alpha^2)\frac{M}{tr(\mathbf{C}_{p,k,b_k})} tr(\mathbf{C}_{p,k,b_k}) = (1-\alpha^2)M; \quad (29)$$

$$(28) + (29) = M = E(\overline{\mathbf{h}}_{k,b_k} \overline{\mathbf{h}}_{k,b_k}^H) \quad (30)$$

which completes the proof. For the simulations of Figures 1 and 2, we have used 100 channel realizations with $\alpha^2 = \frac{3}{4}$, i.e. 100 different realizations of the couple $\widetilde{\mathbf{h}}_{k,b_k}$ and $\widetilde{\mathbf{h}}_{p,k,b_k}$ which both follow the zero-mean unit-variance Gaussian distribution. Thus, in Figure 1 and Figure 2 we have in total $1 - \alpha^2 = \frac{1}{4}$ as estimation error variance. As Figures 1 and 2 suggest, our algorithm surpasses the EWSMSE approach in [1], and we perceive a huge gain for our approach with respect to EWSMSE in the case of low rank correlation matrices as depicted in Figure 2. The reason of the gain can be explained as follows. The Expected Signal-to-Interference plus Noise ratio (ESINR) given in (31)

$$ESINR_k = E[\frac{\mathbf{h}_{k,b_k}^H \mathbf{Q}_k \mathbf{h}_{k,b_k}}{\sigma^2 + \sum_{i \neq k} \mathbf{h}_{i,b_k}^H \mathbf{Q}_i \mathbf{h}_{i,b_k}}] \quad (31)$$

converges to (32) when the number of Tx antennas becomes very large and induces quantities containing the posterior covariance matrix $\mathbf{C}_p$.

$$\begin{aligned} ESINR_k &= \frac{E[\mathbf{h}_{k,b_k}^H \mathbf{Q}_k \mathbf{h}_{k,b_k}]}{\sigma^2 + \sum_{i \neq k} E[\mathbf{h}_{i,b_k}^H \mathbf{Q}_i \mathbf{h}_{i,b_k}]} \\ &= \frac{\overline{\mathbf{h}}_{k,b_k}^H \mathbf{Q}_k \overline{\mathbf{h}}_{k,b_k} + \mathrm{tr}\{\mathbf{Q}_k \mathbf{C}_{p,k,b_k}\}}{\sigma^2 + \sum_{i \neq k} \overline{\mathbf{h}}_{k,b_i}^H \mathbf{Q}_i \overline{\mathbf{h}}_{k,b_i} + \mathrm{tr}\{\mathbf{Q}_i \mathbf{C}_{p,k,b_i}\}} \end{aligned} \quad (32)$$

Furthermore, the EWSMSE approach maximizes the WSR via the minimization of the MSE which in the signal terms contains a term linear in $\mathbf{h}$ as shown in [1, Equation (3)] and its expectation in the same section of [1], which only contains the (posterior) mean of $\mathbf{h}$, hence does not account for the posterior covariance. However, our approach maximizes directly the expression in (32) and hence is better than the EWMSE approach.

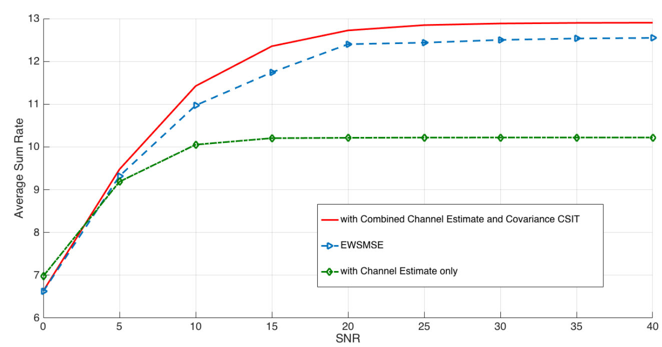

Fig. 1: Sum rate Comparison for $C = 2, K = 8, M_{b_k} = M = 8\, N_k = N = 1 \forall k$ and $C_{t,i,b_k} = C_{p,i,b_k} = I_M\ \forall i, \forall k$

VIII. Conclusion

In this work, we introduced a novel algorithm which maximizes the weighted sum rate in the presence of partial CSIT, but with combined channel estimate and covariance CSIT, using tools from RMT. We compared ours to the existing approach in [1] and demonstrated its superiority for both cases of full and low rank channel correlation matrices.

Acknowledgements

EURECOMs research is partially supported by its industrial members: ORANGE, BMW, SFR, ST Microelectronics, Symantec, SAP, Monaco Telecom, iABG, and by the projects

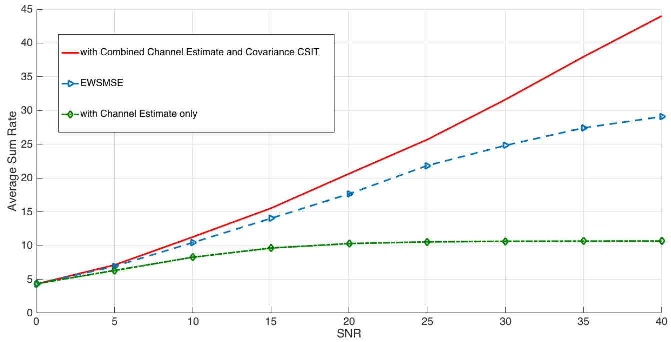

Fig. 2: Sum rate Comparison for $C = 2, K = 8, M_{b_k} = M = 8\, N_k = N = 1 \forall k$ and $rank(\mathbf{C}_{t,i,b_k}) = rank(\mathbf{C}_{p,i,b_k}) = 2\ \forall i, \forall k$

ADEL (EU FP7) and DIONISOS (French ANR). The research of Orange Labs is partially supported by the EU H2020 project ONE5G.

References

[1] 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.

[2] X. Li, E. Björnson, E.G. Larsson, S. Zhou, and J. Wang, "A Multi-cell MMSE Precoder for Massive MIMO Systems and New Large System Analysis," in *IEEE Global Comm's Conf. (Globecom)*, San Diego, CA, USA, 2015.

[3] 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.

[4] S. Wagner and D.T.M. Slock, "Weighted Sum Rate Maximization of Correlated MISO Broadcast Channels under Linear Precoding : a Large System Analysis," in *IEEE Int'l Workshop on Sig. Proc. Advances in Wireless Comm's (SPAWC)*, San Fransisco, CA, USA, 2011.

[5] S. Lakshminarayana, M. Assaad, and M. Debbah, "Coordinated Multicell Beamforming for Massive MIMO: A Random Matrix Approach," *IEEE Trans. Information Theory*, Jun. 2015.

[6] A. Müller, R. Couillet, E. Björnson, S. Wagner, and M. Debbah, "Interference-Aware RZF Precoding for Multi Cell Downlink Systems," *IEEE Trans. Signal Proc.*, Aug. 2015.

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

[8] 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*, October 2014.

Appendix

The proof of (25) is based on lemmas from Random Matrix Theory (RMT) in [3]. The equation (7) can be reformulated as

$$\begin{aligned}
&\mathbf{A}_k = \sum_{i \neq k}^{K} u_i \left[\mathbf{A}^A_{i,k} - \mathbf{A}^B_{i,k}\right];\\
&\mathbf{A}^A_{i,k} = \mathbf{H}^H_{i,b_k}\mathbf{R}^{-1}_{\bar{i}}\mathbf{H}_{i,b_k};\, \mathbf{A}^B_{i,k} = \mathbf{H}^H_{i,b_k}\mathbf{R}^{-1}_{i}\mathbf{H}_{i,b_k}\\
&\mathbf{R}^{-1}_{\bar{i}} = \mathbf{R}^{-1}_{\bar{i},\bar{k}} + (\mathbf{R}^{-1}_{\bar{i}} - \mathbf{R}^{-1}_{\bar{i},\bar{k}}) with\\
&\mathbf{R}_{\bar{i},\bar{k}} = \sum_{j\neq i, j\neq k} \mathbf{H}_{i,b_j}\mathbf{Q}_j\,\mathbf{H}^H_{i,b_j} + \sigma^2 I_{N_i}
\end{aligned} \tag{33}$$

Applying [3, Lemma 2],

$$\begin{aligned}
(\mathbf{R}^{-1}_{\bar{i}} - \mathbf{R}^{-1}_{\bar{i},\bar{k}}) &= -\mathbf{R}^{-1}_{\bar{i}}(\mathbf{R}_{\bar{i}} - \mathbf{R}_{\bar{i},\bar{k}})\mathbf{R}^{-1}_{\bar{i},\bar{k}}\\
&= -\mathbf{R}^{-1}_{\bar{i}}(\mathbf{H}_{i,b_k}\mathbf{Q}_k\mathbf{H}^H_{i,b_k})\mathbf{R}^{-1}_{\bar{i},\bar{k}};
\end{aligned}$$

$$\begin{aligned}
&\mathbf{A}^A_{i,k} = \mathbf{H}^H_{i,b_k}\mathbf{R}_{\bar{i},\bar{k}}\mathbf{H}_{i,b_k}\\
&-\mathbf{H}^H_{i,b_k}\mathbf{R}^{-1}_{\bar{i}}(\mathbf{H}_{i,b_k}\mathbf{Q}_k\mathbf{H}^H_{i,b_k})\mathbf{R}^{-1}_{\bar{i},\bar{k}}\mathbf{H}_{i,b_k};\\
&\mathbf{A}^A_{i,k} = \mathbf{A}^C_{i,k} - \mathbf{A}^A_{i,k}\mathbf{Q}_k\mathbf{A}^C_{i,k} \text{ with } \mathbf{A}^C_{i,k} = \mathbf{H}^H_{i,b_k}\mathbf{R}^{-1}_{\bar{i},\bar{k}}\mathbf{H}_{i,b_k}.
\end{aligned}$$

Similarly,

$$\begin{aligned}
&\mathbf{A}^B_{i,k} = \mathbf{A}^D_{i,k} - \mathbf{A}^B_{i,k}\mathbf{Q}_k\mathbf{A}^D_{i,k}\\
&\text{with } \mathbf{A}^D_{i,k} = \mathbf{H}^H_{i,b_k}\mathbf{R}^{-1}_{i,\bar{k}}\mathbf{H}_{i,b_k}\\
&\text{and } \mathbf{R}_{i,\bar{k}} = \sum_{j\neq k}\mathbf{H}_{i,b_j}\mathbf{Q}_j\,\mathbf{H}^H_{i,b_j}\sigma^2 I_{N_i}.
\end{aligned}$$

Using the channel model in section IV,

$$\begin{aligned}
&\mathbf{A}^C_{i,k} = \overline{\mathbf{H}}^H_{i,b_k}\mathbf{R}^{-1}_{\bar{i},\bar{k}}\overline{\mathbf{H}}_{i,b_k}\\
&+\mathbf{C}^{\frac{1}{2},H}_{p,i,b_k}\widetilde{\mathbf{H}}^H_{p,i,b_k}\mathbf{C}^{\frac{1}{2},H}_{r,i}\mathbf{R}^{-1}_{\bar{i},\bar{k}}\mathbf{C}^{\frac{1}{2}}_{r,i}\widetilde{\mathbf{H}}_{p,i,b_k}\mathbf{C}^{\frac{1}{2}}_{p,i,b_k}\\
&-\mathbf{C}^{\frac{1}{2},H}_{p,i,b_k}\widetilde{\mathbf{H}}^H_{p,i,b_k}\mathbf{C}^{\frac{1}{2},H}_{r,i}\mathbf{R}^{-1}_{\bar{i},\bar{k}}\overline{\mathbf{H}}_{i,b_k}\\
&-\overline{\mathbf{H}}^H_{i,b_k}\mathbf{R}^{-1}_{\bar{i},\bar{k}}\mathbf{C}^{\frac{1}{2}}_{r,i}\widetilde{\mathbf{H}}_{p,i,b_k}\mathbf{C}^{\frac{1}{2}}_{p,i,b_k} \xrightarrow{(a)}\\
&\overline{\mathbf{H}}^H_{i,b_k}\breve{\mathbf{R}}^{-1}_{\bar{i},\bar{k}}\overline{\mathbf{H}}_{i,b_k} + tr\{\breve{\mathbf{R}}^{-1}_{\bar{i},\bar{k}}\mathbf{C}_{r,i}\}\mathbf{C}_{p,i,b_k} = \breve{\mathbf{A}}^C_{i,k};\\
&\mathbf{R}_{\bar{i},\bar{k}} = \sigma^2\mathbf{I}_{N_i} + \sum_{j\neq i, j\neq k}^{K}\overline{\mathbf{H}}_{i,b_j}\mathbf{Q}_j\overline{\mathbf{H}}^H_{i,b_j}\\
&\quad + \sum_{j\neq i, j\neq k}^{K}\mathbf{C}^{\frac{1}{2}}_{r,i}\widetilde{\mathbf{H}}_{p,i,b_j}\mathbf{C}^{\frac{1}{2}}_{p,i,b_j}\mathbf{Q}_j\mathbf{C}^{\frac{1}{2},H}_{p,i,b_j}\widetilde{\mathbf{H}}^H_{p,i,b_j}\mathbf{C}^{\frac{1}{2},H}_{r,i}\\
&\quad - \sum_{j\neq i, j\neq k}^{K}[\mathbf{C}^{\frac{1}{2}}_{r,i}\widetilde{\mathbf{H}}_{p,i,b_j}\mathbf{C}^{\frac{1}{2}}_{p,i,b_j}\mathbf{Q}_j\overline{\mathbf{H}}^H_{i,b_j}\\
&\qquad +\overline{\mathbf{H}}_{i,b_j}\mathbf{Q}_j\mathbf{C}^{\frac{1}{2},H}_{p,i,b_j}\widetilde{\mathbf{H}}^H_{p,i,b_j}\mathbf{C}^{\frac{1}{2},H}_{r,i}]\\
&\xrightarrow{(b)} \sum_{j\neq i, j\neq k}^{K}\overline{\mathbf{H}}_{i,b_j}Q_j\overline{\mathbf{H}}^H_{i,b_j} + tr\{\mathbf{Q}_j\mathbf{C}_{p,i,b_j}\}\mathbf{C}_{r,i}\\
&\quad +\sigma^2\mathbf{I}_{N_i} = \breve{\mathbf{R}}_{\bar{i},\bar{k}};
\end{aligned}$$

Moreover,

$$\begin{aligned}
&\mathbf{A}^D_{i,k} \xrightarrow{(c)} \overline{\mathbf{H}}^H_{i,b_k}\breve{\mathbf{R}}^{-1}_{i,\bar{k}}\overline{\mathbf{H}}_{i,b_k} + tr\{\breve{\mathbf{R}}^{-1}_{i,\bar{k}}\mathbf{C}_{r,i}\}\mathbf{C}_{p,i,b_k} = \breve{\mathbf{A}}^D_{i,k};\\
&\mathbf{R}_{i,\bar{k}} \xrightarrow{(d)} \sum_{j\neq k}^{K}\overline{\mathbf{H}}_{i,b_j}\mathbf{Q}_j\overline{\mathbf{H}}^H_{i,b_j} + tr\{\mathbf{Q}_j\mathbf{C}_{p,i,b_j}\}\mathbf{C}_{r,i}\\
&\qquad +\sigma^2\mathbf{I}_{N_i} = \breve{\mathbf{R}}_{i,\bar{k}}.
\end{aligned} \tag{34}$$

Note that (a) and (c) above correspond to "using the expected value of the matrix and (34)", and "using the expected value of the matrix and (34)" respectively, while (b) and (d) correspond both to "using the expected value of the matrix". Now, the proof is completed.