Optimal Beamforming with Combined Channel and Path CSIT for Multi-cell Multi-User MIMO

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

***Abstract*—We consider a combined form of partial CSIT (Channel State Information at the Transmitter(s) (Tx)), comprising both channel estimates (mean CSIT) and covariance CSIT. In particular multipath induced structured low rank covariances are considered that arise in Massive MIMO and mmWave settings. For the beamforming optimization, we first revisit Weighted Sum Rate (WSR) maximization with perfect CSIT and introduce yet another equivalent approach: Weighted Sum Unbiased MSE (WSUMSE). We then turn to the partial CSIT case where we consider Expected WSR (EWSR) maximization for which EWSUMSE turns out to be a better approximation compared to the existing EWSMSE approach. These approaches also allow an uplink/downlink duality interpretation for partial CSI.**

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. We refer to [5] for a further discussion of the state of the art, in which MIMO IA requires global MIMO channel CSIT. Recent works focus on intercell exchange of only scalar quantities, at fast fading rate, as also on two-stage approaches in which the intercell interference gets zero-forced (ZF). Also, massive MIMO in most works refers actually to MU MISO.

Whereas the exploitation of covariance CSIT may be beneficial, in a MaMIMO context it may quickly lead to high computational complexity and estimation accuracy issues. Computational complexity may be reduced (and the benefit of covariance CSIT enhanced) in the case of low rank or related covariance structure, but the use and tracking of subspaces may still be cumbersome. In the pathwise approach, these subspaces are very parsimoniously parameterized. In a FDD setting, these parameters may even be estimated from the uplink (UL). In a TDD setting with reciprocity, the channel estimation error may account for time variation also in the UL/DL ping-pong. As opposed to the instantaneous channel CSIT, the path CSIT is not affected by fast fading.

Whereas path CSIT by itself may allow zero forcing (ZF) [6], which is of interest at high SNR, we are particularly concerned here with maximum Weighted Sum Rate (WSR) designs accounting for finite SNR. ZF of all interfering links leads to significant reduction of useful signal strength. Massive MIMO makes the pathwise approach viable: the (cross-link) beamformers (BF) can be updated at a reduced (slow fading) rate, parsimonious channel representation facilitates not only uplink but especially downlink channel estimation, the cross-link BF can be used to significantly improve the downlink direct link channel estimates (in FDD), minimal feedback can be introduced to perform meaningful WSR optimization at a finite SNR (whereas ZF requires much less coordination).

In this paper we first introduce a new approach to maximizing WSR, called minimum Weighted Sum Unbiased MSE (WSUMSE). In the perfect CSIT case, WSUMSE also converges to a local optimum of WSR. We then consider various approaches for maximizing Expected WSR (EWSR) for the case of partial CSIT. The existing EWSMSE approach improves over Naive EWSR (NEWSR) by accounting for covariance CSIT in the interference. This can have significant impact, even on the sumrate prelog (DoF) if the instantaneous channel CSIT quality does not scale with SNR. A further improvement is proposed here in the EWSUMSE approach which represents a better approximation of the EWSR. In a MaMIMO setting, the way mean and covariance CSIT are combined in the EWSMSE or WSUMSE approaches for the interference terms becomes equally optimal as in the EWSR for a large number of users. EWSUMSE represents an improvement over EWSMSE for capturing the signal power (matched filtering and diversity aspects) and only leads to a finite (dB) gain in ESINR, but its remaining approximation error over EWSR may be limited. Strictly speaking, in the large number of users setting, EWSMSE $\leq$ EWSR $\leq$ EWSUMSE. The step from EWSMSE to EWSUMSE also deals with the following question. Covariance CSIT can be used to improve the channel estimate from a basic deterministic estimate to a Bayesian estimate. The question then arises: is that enough? The answer is no and a first take at this issue is proposed here. This paper is a followup on [7] from which we reproduce some sections to ease reading. In [7] we introduced a heuristic to design the Tx separately using path CSIT only. It turns out that this heuristic is recovered by the EWSUMSE approach proposed here, which furthermore provides expressions for a number of auxiliary quantities that are needed and allows the combination of channel estimate and path CSIT.

II. Channel (Information) Models

In this section we drop the user index k for simplicity.

A. Specular Wireless MIMO Channel Model

The MIMO channel transfer matrix at any particular subcarrier of a given OFDM symbol can be written as [8], [9]

$$\mathbf{H} = \sum_{i=1}^{N_p} A_i\, e^{j\psi_i}\, \mathbf{h}_r(\phi_i)\, \mathbf{h}_t^T(\theta_i) = \mathbf{B}\,\mathbf{A}^H \qquad (1)$$

where there are N_p (specular) pathwise contributions with

- $A_i > 0$: path amplitude
- θ_i: direction of departure (AoD)
- ϕ_i: direction of arrival (AoA)
- $\mathbf{h}_t(.)$, $\mathbf{h}_r(.)$: $M/N \times 1$ Tx/Rx antenna array response

with $||\mathbf{h}_t(.)|| = 1$, $||\mathbf{h}_r(.)|| = N$, and

$$\mathbf{B} = [\mathbf{h}_r(\phi_1)\,\mathbf{h}_r(\phi_1)\cdots]\begin{bmatrix} e^{j\psi_1} & & \\ & e^{j\psi_2} & \\ & & \ddots \end{bmatrix}, \mathbf{A}^H = \begin{bmatrix} A_1 & & \\ & A_2 & \\ & & \ddots \end{bmatrix}\begin{bmatrix} \mathbf{h}_t^T(\theta_1) \\ \mathbf{h}_t^T(\theta_2) \\ \vdots \end{bmatrix} \qquad (2)$$

The antenna array responses are just functions of angles AoD, AoA in the case of standard antenna arrays with scatterers in the far field. In the case of distributed antenna systems, the array responses become a function of all position parameters of the path scatterers. The fast variation of the phases ψ_i (due to Doppler) and possibly the variation of the A_i (when the nominal path represents in fact a superposition of paths with similar parameters) correspond to the fast fading. All the other parameters vary on a slower time scale and correspond to slow fading.

B. Dominant Paths Partial CSIT Channel Model

Assuming the Tx disposes of not much more than the information about r dominant path AoDs, we shall consider the following MIMO (Ricean) channel model

$$\mathbf{H} = \mathbf{B}\,\mathbf{A}^H(\theta) + \sqrt{\beta}\widetilde{\mathbf{H}}' \qquad (3)$$

which follows from (1), (2) except restricted to the r strongest paths, with the rest modeled by $\sqrt{\beta}\widetilde{\mathbf{H}}'$ (elements i.i.d. $\sim \mathcal{CN}(0,\beta)$, independent of the ψ_i). Averaging of the path phases ψ_i, we get for the Tx side covariance matrix

$$\mathrm{E}\,\mathbf{H}^H\mathbf{H} = N\,\mathbf{C}_t = N(\mathbf{A}\mathbf{A}^H + \beta\, I_M) \qquad (4)$$

since due to the normalization of the antenna array responses, $\mathrm{E}\,\mathbf{B}^H\mathbf{B} = \mathrm{diag}\{[\mathbf{h}_r(\phi_1)\,\mathbf{h}_r(\phi_2)\cdots]^H\,[\mathbf{h}_r(\phi_1)\,\mathbf{h}_r(\phi_2)\cdots]\} = N\mathbf{I}$. Note that the pathwise channel model, which leads here to a type of Tx covariance CSIT, does not lead to the usual separable covariance case, which is discussed e.g. [5].

C. Combined Channel and Path Partial CSIT

Now, the Tx dispose also of a (det erministic) channel estimate

$$\widehat{\mathbf{H}} = \mathbf{H} + \frac{1}{\sqrt{N}}\widetilde{\mathbf{H}}_d\,\mathbf{C}_d^{1/2} \qquad (5)$$

where 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 channel estimate with the prior information leads to the (posterior) LMMSE estimate. The non-separable prior covariance should in principle be accounted for properly in the LMMSE estimation. However, to simplify beamformer design, we shall force a separable prior model by considering all of $\mathbf{B}$ to be unknown with elements that are i.i.d. $\sim \mathcal{CN}(0,1)$. As a result we can interpret $\mathbf{H}$ to be of the form $\mathbf{H} = \widetilde{\mathbf{H}}\,\mathbf{C}_t^{1/2}$ and we get for the LMMSE estimate

$$\begin{array}{l} \widehat{\widehat{\mathbf{H}}} = \widehat{\mathbf{H}}_d\,(\mathbf{C}_t + \mathbf{C}_d)^{-1}\mathbf{C}_t = \mathbf{H} + \widetilde{\mathbf{H}}_p\,\mathbf{C}_p^{1/2} \\ \mathbf{C}_p = \mathbf{C}_d\,(\mathbf{C}_t + \mathbf{C}_d)^{-1}\mathbf{C}_t \end{array} \qquad (6)$$

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

$$\mathrm{E}_{\mathbf{H}|\widehat{\mathbf{H}}_d}\mathbf{H}^H\mathbf{H} = \widehat{\widehat{\mathbf{H}}}^H\,\widehat{\widehat{\mathbf{H}}} + N\,\mathbf{C}_p = \mathbf{W}\,. \qquad (7)$$

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

$$\mathrm{E}_{\widehat{\mathbf{H}}_d}\mathbf{W} = \mathrm{E}_{\mathbf{H},\widehat{\mathbf{H}}_d}\mathbf{H}^H\mathbf{H} = \mathrm{E}_{\mathbf{H}}\mathbf{H}^H\mathbf{H} = N\,\mathbf{C}_t\,. \qquad (8)$$

III. Streamwise IBC Signal Model

In this paper we shall consider mostly 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

$$\mathbf{y}_k = \underbrace{\mathbf{H}_{k,b_k}\,\mathbf{g}_k\,x_k}_{\text{signal}} + \underbrace{\sum_{\substack{i\neq k \\ b_i=b_k}}\mathbf{H}_{k,b_k}\,\mathbf{g}_i\,x_i}_{\text{intracell interf.}} + \underbrace{\sum_{j\neq b_k}\sum_{i:b_i=j}\mathbf{H}_{k,j}\,\mathbf{g}_i\,x_i}_{\text{intercell interf.}} + \mathbf{v}_k \qquad (9)$$

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 $\widehat{x}_k = \mathbf{f}_k^H \mathbf{y}_k$

$$\begin{aligned}\widehat{x}_k &= \mathbf{f}_k^H \mathbf{H}_{k,b_k}\, \mathbf{g}_k\, x_k + \sum_{i=1,\neq k}^{K} \mathbf{f}_k^H \mathbf{H}_{k,b_i}\, \mathbf{g}_i\, x_i + \mathbf{f}_k^H \mathbf{v}_k \\ &= \mathbf{f}_k^H \mathbf{h}_{k,k}\, x_k + \sum_{i\neq k} \mathbf{f}_k^H \mathbf{h}_{k,i}\, x_i + \mathbf{f}_k^H \mathbf{v}_k\end{aligned} \tag{10}$$

where $\mathbf{h}_{k,i} = \mathbf{H}_{k,b_i}\, \mathbf{g}_i$ is the channel-Tx cascade vector. ZF (IA) feasibility for both the general reduced rank MIMO channels case and the pathwise MIMO case has been discussed in [6], in particular also when only based on Tx side covariance CSIT. Also the role of Rx antennas is highlighted and a comparison with FIR ZF in an asynchronous scenario is presented.

IV. Max WSR with Perfect CSIT

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

$$WSR = WSR(\mathbf{g}) = \sum_{k=1}^{K} u_k \; \ln \frac{1}{e_k} \tag{11}$$

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:

$$\begin{aligned}&\frac{1}{e_k} = 1 + \mathbf{g}_k^H \mathbf{H}_{k,b_k}^H \mathbf{R}_{\overline{k}}^{-1} \mathbf{H}_{k,b_k} \mathbf{g}_k = (1 - \mathbf{g}_k^H \mathbf{H}_{k,b_k}^H \mathbf{R}_k^{-1} \mathbf{H}_{k,b_k} \mathbf{g}_k)^{-1} \\ &\mathbf{R}_k = \mathbf{H}_{k,b_k} \mathbf{Q}_k\, \mathbf{H}_{k,b_k}^H + \mathbf{R}_{\overline{k}}\,,\; \mathbf{Q}_i = \mathbf{g}_i \mathbf{g}_i^H\,, \\ &\mathbf{R}_{\overline{k}} = \sum_{i\neq k} \mathbf{H}_{k,b_i} \mathbf{Q}_i\, \mathbf{H}_{k,b_i}^H + I_{N_k}\,.\end{aligned} \tag{12}$$

$\mathbf{R}_k$, $\mathbf{R}_{\overline{k}}$ 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 $\mathbf{f}_k$,

$$\mathbf{f}_k = \mathbf{R}_k^{-1} \mathbf{H}_{k,b_k} \mathbf{g}_k = \mathbf{R}_k^{-1} \mathbf{h}_{k,k}\,. \tag{13}$$

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

$$\sum_{k:b_k=j} \mathrm{tr}\{\mathbf{Q}_k\} \leq P_j\,. \tag{14}$$

A. From Max WSR to Min WSMSE

For a general Rx filter $\mathbf{f}_k$ we have the MSE

$$\begin{aligned}&e_k(\mathbf{f}_k, \mathbf{g}) = (1 - \mathbf{f}_k^H \mathbf{H}_{k,b_k} \mathbf{g}_k)(1 - \mathbf{g}_k^H \mathbf{H}_{k,b_k}^H \mathbf{f}_k) \\ &+ \textstyle\sum_{i\neq k} \mathbf{f}_k^H \mathbf{H}_{k,b_i} \mathbf{g}_i \mathbf{g}_i^H \mathbf{H}_{k,b_i}^H \mathbf{f}_k + ||\mathbf{f}_k||^2 = 1 - \mathbf{f}_k^H \mathbf{H}_{k,b_k} \mathbf{g}_k \\ &- \mathbf{g}_k^H \mathbf{H}_{k,b_k}^H \mathbf{f}_k + \sum_i \mathbf{f}_k^H \mathbf{H}_{k,b_i} \mathbf{g}_i \mathbf{g}_i^H \mathbf{H}_{k,b_i}^H \mathbf{f}_k + ||\mathbf{f}_k||^2.\end{aligned} \tag{15}$$

The $WSR(\mathbf{g})$ is a non-convex and complicated function of $\mathbf{g}$. Inspired by [10], we introduced [11], [1] an augmented cost function, the Weighted Sum MSE, $WSMSE(\mathbf{g}, \mathbf{f}, w)$

$$= \sum_{k=1}^{K} u_k (w_k\; e_k(\mathbf{f}_k, \mathbf{g}) - \ln w_k) + \sum_{i=1}^{C} \lambda_i (\sum_{k:b_k=i} ||\mathbf{g}_k||^2 - P_i) \tag{16}$$

where λ_i = Lagrange multipliers. After optimizing over the aggregate auxiliary Rx filters $\mathbf{f}$ and weights w, we get the WSR back:

$$\min_{\mathbf{f},w} WSMSE(\mathbf{g}, \mathbf{f}, w) = -WSR(\mathbf{g}) + \sum_{k=1}^{K} u_k \tag{17}$$

The advantage of the augmented cost function: alternating optimization leads to solving simple quadratic or convex functions:

$$\begin{aligned}&\min_{w_k} WSMSE \;\Rightarrow\; w_k = 1/e_k \\ &\min_{\mathbf{f}_k} WSMSE \Rightarrow \mathbf{f}_k = (\sum_i \mathbf{H}_{k,b_i} \mathbf{g}_i \mathbf{g}_i^H \mathbf{H}_{k,b_i}^H + I_{N_k})^{-1} \mathbf{H}_{k,b_k} \mathbf{g}_k \\ &\min_{\mathbf{g}_k} WSMSE \;\Rightarrow \\ &\mathbf{g}_k = (\textstyle\sum_i u_i w_i \mathbf{H}_{i,b_k}^H \mathbf{f}_i \mathbf{f}_i^H \mathbf{H}_{i,b_k} + \lambda_{b_k} I_M)^{-1} \mathbf{H}_{k,b_k}^H \mathbf{f}_k u_k w_k\end{aligned} \tag{18}$$

UL/DL duality: the optimal Tx filter g_k is of the form of a MMSE linear Rx for the dual UL in which λ plays the role of Rx noise variance and $u_k w_k$ plays the role of stream variance.

B. Difference of Convex Functions Programming

In a classical difference of convex functions (DC programming) approach, Kim and Giannakis [12] 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

$$\begin{aligned}&WSR = u_k \ln\det(\mathbf{R}_{\overline{k}}^{-1} \mathbf{R}_k) + WSR_{\overline{k}}\,, \\ &WSR_{\overline{k}} = \textstyle\sum_{i=1,\neq k}^{K} u_i \ln\det(\mathbf{R}_{\overline{i}}^{-1} \mathbf{R}_i)\end{aligned} \tag{19}$$

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 $\mathbf{Q}^{'}$ (i.e. all $\mathbf{Q}_i^{'}$) with e.g. $\mathbf{R}_i^{'} = \mathbf{R}_i(\mathbf{Q}^{'})$, then

$$\begin{aligned}&WSR_{\overline{k}}(\mathbf{Q}_k, \mathbf{Q}^{'}) \approx WSR_{\overline{k}}(\mathbf{Q}_k^{'}, \mathbf{Q}^{'}) - \mathrm{tr}\{(\mathbf{Q}_k - \mathbf{Q}_k^{'}) \mathbf{T}_k^{'}\} \\ &\mathbf{T}_{\overline{k}}^{'} = - \left.\frac{\partial WSR_{\overline{k}}(\mathbf{Q}_k, \mathbf{Q}^{'})}{\partial \mathbf{Q}_k}\right|_{\mathbf{Q}_k^{'}, \mathbf{Q}^{'}} = \sum_{i\neq k}^{K} u_i \mathbf{H}_{i,b_k}^H (\mathbf{R}_{\overline{i}}^{'-1} - \mathbf{R}_i^{'-1}) \mathbf{H}_{i,b_k}\end{aligned} \tag{20}$$

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

$$\begin{aligned}&WSR(\mathbf{g}, \mathbf{g}^{'}, \lambda) = \sum_{j=1}^{C} \lambda_j P_j + \\ &\sum_{k=1}^{K} u_k \ln(1 + \mathbf{g}_k^H \mathbf{S}_k^{'} \mathbf{g}_k) - \mathbf{g}_k^H (\mathbf{T}_{\overline{k}}^{'} + \lambda_{b_k} I) \mathbf{g}_k\end{aligned} \tag{21}$$

where

$$\mathbf{S}_k^{'} = \mathbf{H}_{k,b_k}^H \mathbf{R}_{\overline{k}}^{'-1} \mathbf{H}_{k,b_k}\,. \tag{22}$$

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

$$\mathbf{S}_k^{'}\,\mathbf{g}_k = \frac{1+\mathbf{g}_k^H\mathbf{S}_k^{'}\mathbf{g}_k}{u_k}\,(\mathbf{T}_{\overline{k}}^{'}+\lambda_{b_k}I)\mathbf{g}_k \qquad (23)$$

or hence $\overline{\mathbf{g}}_k = V_{max}(\mathbf{S}_k^{'},\mathbf{T}_{\overline{k}}^{'}+\lambda_{b_k}I)$ is the (normalized) "max" generalized eigenvector of the two indicated matrices, with max eigenvalue $\sigma_k = \sigma_{max}(\mathbf{S}_k^{'},\mathbf{T}_{\overline{k}}^{'}+\lambda_{b_k}I)$. Let $\sigma_k^{(1)} = \overline{\mathbf{g}}_k^H\mathbf{S}_k^{'}\overline{\mathbf{g}}_k$, $\sigma_k^{(2)} = \overline{\mathbf{g}}_k^H\mathbf{T}_{\overline{k}}^{'}\overline{\mathbf{g}}_k$. The advantage of formulation (21) is that it allows straightforward power adaptation: introducing stream powers $p_k \geq 0$ and substituting $\mathbf{g}_k = \sqrt{p_k}\,\overline{\mathbf{g}}_k$ in (21) yields

$$WSR{=}\sum_{j}^{C}\lambda_j P_j + \sum_{k=1}^{K}\{u_k\ln(1+p_k\sigma_k^{(1)}) - p_k(\sigma_k^{(2)}+\lambda_{b_k})\} \qquad (24)$$

which leads to the following interference leakage aware water filling (WF)

$$p_k = \left(\frac{u_k}{\sigma_k^{(2)}+\lambda_{b_k}} - \frac{1}{\sigma_k^{(1)}}\right)^+ \qquad (25)$$

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 $\overline{\mathbf{g}}_k$, the p_k and the λ_l.

Note that the DC programming approach, which avoids introducing Rxs, can at every BF update allow to introduce an arbitrary number of streams per user by determining multiple dominant generalized eigenvectors, and then let the WF operation decide how many streams can actually be sustained.

V. From Max WSR to Min Weighted Sum Unbiased MSE (WSUMSE)

For the Rx output $\widehat{x}_k$ to be an unbiased estimator for the Tx signal x_k, we require

$$\mathrm{E}_{|x_k}\,\widehat{x}_k = x_k \;\Rightarrow\; \mathbf{f}_k^H\mathbf{H}_{k,b_k}\,\mathbf{g}_k = 1\;. \qquad (26)$$

If the Tx/Rx filters satisfy the unbiasedness constraint, then we get the Unbiased MSE (UMSE)

$$e_k^u(\mathbf{f}_k,\mathbf{g}) = \sum_{i\neq k}\mathbf{f}_k^H\mathbf{H}_{k,b_i}\mathbf{g}_i\mathbf{g}_i^H\mathbf{H}_{k,b_i}^H\mathbf{f}_k + ||\mathbf{f}_k||^2\,. \qquad (27)$$

In the complex case, it is more convenient to work with the alternative unbiasedness constraint

$$|\mathbf{f}_k^H\mathbf{H}_{k,b_k}\,\mathbf{g}_k|^2 = 1 \qquad (28)$$

in which the phase uncertainty does not affect SINR or Gaussian capacity. Now consider the following augmented cost function, the Weighted Sum UMSE, $WSUMSE(\mathbf{g},\mathbf{f},w_1,w_2)$

$$\begin{aligned}&{=}\sum_{k=1}^{K}u_k(w_{1,k}e_k^u(\mathbf{f}_k,\mathbf{g}){-}\ln w_{1,k} - w_{2,k}(e_k^u(\mathbf{f}_k,\mathbf{g}){+}1){+}\ln w_{2,k})\\ &+\sum_{k=1}^{K}\mu_k(1-|\mathbf{f}_k^H\mathbf{H}_{k,b_k}\mathbf{g}_k|^2) + \sum_{i=1}^{C}\lambda_i(\sum_{k:b_k=i}||\mathbf{g}_k||^2{-}P_i)\end{aligned} \qquad (29)$$

where the λ_i, μ_k are Lagrange multipliers. After optimizing over the aggregate auxiliary Rx filters $\mathbf{f}$ and weights w_1, w_2, we get the WSR back:

$$\min_{\mathbf{f}}\min_{w_1}\max_{w_2} WSUMSE(\mathbf{g},\mathbf{f},w_1,w_2) = -WSR(\mathbf{g}) \quad (30)$$

where $WSR = \sum_k u_k\ln(1+\frac{1}{e_k^u})$. The augmented cost function can again be conveniently optimized by alternating optimization:

$$\begin{aligned}&\min_{w_{1,k}}\max_{w_{2,k}} WSUMSE \;\Rightarrow\; w_{1,k} = 1/e_k^u, w_{2,k} = 1/(e_k^u+1)\\ &\min_{\mathbf{f}_k} WSUMSE \Rightarrow \mathbf{f}_k{=}\mathbf{R}_{\overline{k}}^{-1}\mathbf{H}_{k,b_k}\mathbf{g}_k\,\frac{\mu_k\mathbf{g}_k^H\mathbf{H}_{k,b_k}^H\mathbf{f}_k}{u_k w_k}\\ &\min_{\mathbf{g}_k} WSUMSE \;\Rightarrow\\ &\mathbf{g}_k{=}(\mathbf{T}_{\overline{k}}{+}\lambda_{b_k}I_M)^{-1}\mathbf{H}_{k,b_k}^H\mathbf{f}_k\,\mu_k\mathbf{f}_k^H\mathbf{H}_{k,b_k}\mathbf{g}_k\end{aligned} \qquad (31)$$

where $w_k = w_{1,k} - w_{2,k} > 0$. Note that $|\mathbf{f}_k^H\mathbf{H}_{k,b_k}\mathbf{g}_k| = 1$. We can choose the phase such that $\mathbf{f}_k^H\mathbf{H}_{k,b_k}\mathbf{g}_k = 1$. If we also reparameterize $\mu_k = u_k w_k\mu_k^{'}$, where the $\mu_k^{'}$ are chosen to satisfy $\mathbf{f}_k^H\mathbf{H}_{k,b_k}\mathbf{g}_k = 1$, then we can rewrite
$\mathbf{f}_k = \mathbf{R}_{\overline{k}}^{-1}\mathbf{H}_{k,b_k}\mathbf{g}_k\,\mu_k^{'}$, $\mathbf{g}_k = (\mathbf{T}_{\overline{k}}{+}\lambda_{b_k}I_M)^{-1}\mathbf{H}_{k,b_k}^H\mathbf{f}_k u_k w_k\,\mu_k^{'}$ which are proportional to the Tx/Rx found by the WSMSE approach.

A word about convergence. Note that given the $\mathbf{f}$ and $\mathbf{g}$ (or in other words the e_k^u), the WSUMSE decomposes into the w_1 and w_2 dependencies, which can be optimized in parallel. The optimization over w_1, w_2 jointly decreases the cost function since $\ln(\frac{e_k^u}{1+e_k^u})$ is an increasing function of e_k^u. Now, as long as the optimization over the w_1, w_2 is done jointly, every alternating optimization in (31) decreases the WSUMSE cost function.

VI. 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. Once the CSIT is imperfect, various optimization criteria could be considered, such as outage capacity. Here we shall consider the expected weighted sum rate $\mathrm{E}_{\mathbf{H}|\widehat{\widehat{\mathbf{H}}}}WSR(\mathbf{g},\mathbf{H}) =$

$$EWSR(\mathbf{g}) = \mathrm{E}_{\mathbf{H}|\widehat{\widehat{\mathbf{H}}}}\sum_k u_k\ln(1+\mathbf{g}_k^H\mathbf{H}_{k,b_k}^H\mathbf{R}_{\overline{k}}^{-1}\mathbf{H}_{k,b_k}\mathbf{g}_k) \qquad (32)$$

where we now underlign the dependence of various quantities on $\mathbf{H}$. The EWSR in (32) 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. Now we consider various deterministic approximations for the EWSR.

VII. EWSR Lower Bound: EWSMSE

The criterion $EWSR(\mathbf{g})$ is difficult to compute and to maximize directly. It is much more attractive to consider $\mathrm{E}_{\mathbf{H}|\hat{\hat{\mathbf{H}}}} e_k(\mathbf{f}_k, \mathbf{g}, \mathbf{H})$ as in [13] since $e_k(\mathbf{f}_k, \mathbf{g}, \mathbf{H})$ is quadratic in $\mathbf{H}$. Hence consider optimizing the expected weighted sum MSE $\mathrm{E}_{\mathbf{H}|\hat{\hat{\mathbf{H}}}} WSMSE(\mathbf{g}, \mathbf{f}, w, \mathbf{H})$.

$$\begin{aligned}&\min_{\mathbf{f},w} \mathrm{E}_{\mathbf{H}|\hat{\hat{\mathbf{H}}}} WSMSE(\mathbf{g}, \mathbf{f}, w, \mathbf{H}) \\ &\geq \mathrm{E}_{\mathbf{H}|\hat{\hat{\mathbf{H}}}} \min_{\mathbf{f},w} WSMSE(\mathbf{g}, \mathbf{f}, w, \mathbf{H}) = -EWSR(\mathbf{g})\end{aligned} \tag{33}$$

or hence

$$EWSR(\mathbf{g}) \geq -\min_{\mathbf{f},w} \mathrm{E}_{\mathbf{H}|\hat{\hat{\mathbf{H}}}} WSMSE(\mathbf{g}, \mathbf{f}, w, \mathbf{H})\ . \tag{34}$$

So now only a *lower bound* to the EWSR gets maximized, which corresponds in fact to the *CSIR being equally partial as the CSIT* (whereas the EWSR criterion corresponds to partial CSIT but perfect CSIR). Now, since

$$\begin{aligned}\mathrm{E}_{\mathbf{H}|\hat{\hat{\mathbf{H}}}} \mathbf{H}\mathbf{Q}\mathbf{H}^H &= \hat{\hat{\mathbf{H}}}\mathbf{Q}\hat{\hat{\mathbf{H}}}^H + \mathrm{tr}\{\mathbf{Q}\mathbf{C}_p\}\, I_N \\ \mathrm{E}_{\mathbf{H}|\hat{\hat{\mathbf{H}}}} \mathbf{H}^H\mathbf{Q}\mathbf{H} &= \hat{\hat{\mathbf{H}}}^H\mathbf{Q}\hat{\hat{\mathbf{H}}} + \mathrm{tr}\{\mathbf{Q}\}\, \mathbf{C}_p\end{aligned} \tag{35}$$

we get $\mathrm{E}_{\mathbf{H}|\hat{\hat{\mathbf{H}}}} e_k =$

$$\begin{aligned}&\hat{\hat{e}}_k = 1 - 2\Re\{\mathbf{f}_k^H \hat{\hat{\mathbf{H}}}_{k,b_k}\mathbf{g}_k\} + \textstyle\sum_{i=1}^K \mathbf{f}_k^H \hat{\hat{\mathbf{H}}}_{k,b_i}\mathbf{g}_i\mathbf{g}_i^H \hat{\hat{\mathbf{H}}}_{k,b_i}^H \mathbf{f}_k \\ &+ ||\mathbf{f}_k||^2 \textstyle\sum_{i=1}^K \mathbf{g}_i^H \mathbf{C}_{p,k,b_i}\mathbf{g}_i + ||\mathbf{f}_k||^2.\end{aligned} \tag{36}$$

where $\mathbf{C}_{t,k,b_i}$ are Tx side (LMMSE error) covariance matrices of $\mathbf{H}_{k,b_i}$. Note that the signal term disappears if $\hat{\hat{\mathbf{H}}}_k = 0$! Hence the EWSMSE lower bound is (very) loose unless the Rice factor is high, and is useless in the absence of channel estimates. Alternating optimization as before leads to

$$\begin{aligned}&\min_{w_k} EWSMSE \ \Rightarrow\ w_k = 1/\hat{\hat{e}}_k \\ &\min_{\mathbf{f}_k} EWSMSE \Rightarrow \mathbf{f}_k = \hat{\hat{\mathbf{R}}}_k^{-1} \hat{\hat{\mathbf{H}}}_{k,b_k}\mathbf{g}_k \\ &\min_{\mathbf{g}_k} EWSMSE \ \Rightarrow \mathbf{g}_k = (\hat{\hat{\mathbf{T}}}_k + \lambda_{b_k} I_M)^{-1} \hat{\hat{\mathbf{H}}}_{k,b_k}^H \mathbf{f}_k u_k w_k\end{aligned} \tag{37}$$

where

$$\begin{aligned}\hat{\hat{\mathbf{R}}}_k &= \textstyle\sum_i \hat{\hat{\mathbf{H}}}_{k,b_i}\mathbf{g}_i\mathbf{g}_i^H \hat{\hat{\mathbf{H}}}_{k,b_i}^H + (1 + \sum_i \mathbf{g}_i^H \mathbf{C}_{p,k,b_i}\mathbf{g}_i)\, I_{N_k} \\ \hat{\hat{\mathbf{T}}}_k &= \textstyle\sum_{i=1}^K u_i w_i (\hat{\hat{\mathbf{H}}}_{i,b_k}^H \mathbf{f}_i\mathbf{f}_i^H \hat{\hat{\mathbf{H}}}_{i,b_k} + ||\mathbf{f}_i||^2 \mathbf{C}_{p,k,b_i})\ .\end{aligned} \tag{38}$$

VIII. Expected Weighted Sum Unbiased MSE (EWSUMSE)

Consider now the Expected WSUMSE (EWSUMSE) criterion obtained by averaging (29) over $\mathrm{E}_{\mathbf{H}|\hat{\hat{\mathbf{H}}}}$. Alternating optimization leads to

$$\begin{aligned}w_k &= 1/(\hat{\hat{e}}_k^u(\hat{\hat{e}}_k^u + 1)) \\ \mathbf{f}_k &= V_{max}(\hat{\hat{\mathbf{H}}}_{k,b_k}\mathbf{g}_k\mathbf{g}_k^H \hat{\hat{\mathbf{H}}}_{k,b_k}^H + \mathbf{g}_k^H \mathbf{C}_{p,k,b_k}\mathbf{g}_k\, I_{N_k}, \hat{\hat{\mathbf{R}}}_{\overline{k}}) \\ \mathbf{g}_k &= V_{max}(\hat{\hat{\mathbf{H}}}_{k,b_k}^H \mathbf{f}_k\mathbf{f}_k^H \hat{\hat{\mathbf{H}}}_{k,b_k} + ||\mathbf{f}_k||^2 \mathbf{C}_{p,k,b_k}, \hat{\hat{\mathbf{T}}}_{\overline{k}} + \lambda_{b_k} I_M)\end{aligned} \tag{39}$$

where $\hat{\hat{\mathbf{R}}}_{\overline{k}}$ and $\hat{\hat{\mathbf{T}}}_{\overline{k}}$ are like $\hat{\hat{\mathbf{R}}}_k$ and $\hat{\hat{\mathbf{T}}}_k$ in (38) but without the term for user k. Note that in case of no channel estimate, $\hat{\hat{\mathbf{H}}}_{k,b_k} = 0$, then

$$\begin{aligned}\mathbf{f}_k &= V_{min}(\hat{\hat{\mathbf{R}}}_{\overline{k}}) \\ \mathbf{g}_k &= V_{max}(\mathbf{C}_{p,k,b_k}, \hat{\hat{\mathbf{T}}}_{\overline{k}} + \lambda_{b_k} I_M)\ .\end{aligned} \tag{40}$$

The Rx and Tx filters optimizing the EWSUMSE criterion also optimize the ESEINR (Expected Signal to Expected Interference plus Noise Ratio) and (optimally weighted) ESELNR (Expected Signal to Expected Leakage plus Noise Ratio):

$$\begin{aligned}\mathrm{ESEINR}_k &= \frac{\mathbf{f}_k^H(\hat{\hat{\mathbf{H}}}_{k,b_k}\mathbf{g}_k\mathbf{g}_k^H \hat{\hat{\mathbf{H}}}_{k,b_k}^H + \mathbf{g}_k^H \mathbf{C}_{p,k,b_k}\mathbf{g}_k\, I_{N_k})\mathbf{f}_k}{\mathbf{f}_k^H \hat{\hat{\mathbf{R}}}_{\overline{k}}\mathbf{f}_k} \\ \mathrm{ESELNR}_k &= \frac{\mathbf{g}_k^H(\hat{\hat{\mathbf{H}}}_{k,b_k}^H \mathbf{f}_k\mathbf{f}_k^H \hat{\hat{\mathbf{H}}}_{k,b_k} + ||\mathbf{f}_k||^2 \mathbf{C}_{p,k,b_k})\mathbf{g}_k}{\mathbf{g}_k^H(\hat{\hat{\mathbf{T}}}_{\overline{k}} + \lambda_{b_k} I_M)\mathbf{g}_k}\ .\end{aligned} \tag{41}$$

Of course, the scale factors of the Tx and Rx filters in (39) need to be adjusted, which could be done as in [10] or (with the Lagrange multipliers also) using an appropriate adaptation of the DC programming waterfilling. In that case, $\mathbf{f}_k$ is left to be unit norm as in (39) whereas $\mathbf{g}_k$ from (39) is interpreted as $\overline{\mathbf{g}}_k$. The stream powers and Lagrange multipliers get adapted from (24),(25) with $\hat{\hat{\mathbf{T}}}_{\overline{k}}$ replacing $\mathbf{T}_{\overline{k}}$ and $\mathbf{S}_k$ being replaced by

$$\hat{\hat{\mathbf{S}}}_k = \hat{\hat{\mathbf{H}}}_{k,b_k}^H \hat{\hat{\mathbf{R}}}_{\overline{k}}^{-1} \hat{\hat{\mathbf{H}}}_{k,b_k} + \mathrm{tr}\{\hat{\hat{\mathbf{R}}}_{\overline{k}}^{-1}\}\mathbf{C}_{p,k,b_k}\ . \tag{42}$$

IX. Discussion EWSR Approximations

In the case of partial CSIT we get for the symbol estimate

$$\begin{aligned}\hat{x}_k &= \mathbf{f}_k^H \hat{\hat{\mathbf{H}}}_{k,b_k}\mathbf{g}_k\, x_k + \underbrace{\mathbf{f}_k^H \widetilde{\mathbf{H}}_{k,b_k}\mathbf{g}_k\, x_k}_{\text{sig. ch. error}} \\ &+ \sum_{i=1,\neq k}^K (\mathbf{f}_k^H \hat{\hat{\mathbf{H}}}_{k,b_i}\mathbf{g}_i\, x_i + \underbrace{\mathbf{f}_k^H \widetilde{\mathbf{H}}_{k,b_i}\mathbf{g}_i\, x_i}_{\text{interf. ch. error}}) + \mathbf{f}_k^H \mathbf{v}_k\ .\end{aligned} \tag{43}$$

A first approach would be Naive EWSR (NEWSR) which would just replace $\mathbf{H}$ by $\hat{\hat{\mathbf{H}}}$ in a perfect CSIT approach. It ignores the channel estimation error in both signal and interference terms. The EWSMSE approach improves over NEWSR by accounting for covariance CSIT in the interference. This can have significant impact, even on the DoF if the instantaneous channel CSIT quality does not scale with SNR. However, note from (38) that EWSMSE also moves the channel estimation error in the signal term to the interference plus noise. A further improvement is proposed here in the

EWSUMSE approach which represents a better approximation of the EWSR. In EWSUMSE, the channel estimation error in the signal term is accounted for in the signal power. In a MaMIMO setting, the way mean and covariance CSIT are combined in the WSUMSE approach for the interference terms becomes equally optimal as in the EWSR for a large number of users. Indeed, in the MaMIMO setting the interference becomes Gaussian due to the Central Limit Theorem and hence is characterized by its covariance. EWSUMSE furthermore represents an improvement over EWSMSE for capturing the signal power (matched filtering and diversity aspects) but only leads to a finite (dB) gain in ESEINR, though its remaining approximation error over EWSR may be limited. In the MaMIMO setting, EWSUMSE represents a EWSR upper bound due to the concavity of $\ln(.)$.

Both EWSMSE and EWSUMSE lead to UL/DL duality for partial CSIT, but only EWSUMSE treats signal and interference terms similarly.

Acknowledgments

EURECOM's research is partially supported by its industrial members: ORANGE, BMW, SFR, ST Microelectronics, Symantec, SAP, Monaco Telecom, iABG, by the projects ADEL (EU FP7) and DIONISOS (French ANR). The research of Orange Labs is partially supported by the EU H2020 project Fantastic5G.

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. Ghauri, "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] Y. Lejosne, M. Bashar, D. Slock, and Y. Yuan-Wu, "From MU Massive MISO to Pathwise MU Massive MIMO," in *Proc. IEEE Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*, Toronto, Canada, June 2014.

[6] Y. Lejosne, M. Bashar, D. Slock, and Y. Yuan-Wu, "Decoupled, Rank Reduced, Massive and Frequency-Selective Aspects in MIMO Interfering Broadcast Channels," in *Proc. IEEE Int'l Symp. on Communications, Control and Sig. Proc. (ISCCSP)*, Athens, Greece, May 2014.

[7] W. Tabikh, D. Slock, and Y. Yuan-Wu, "The Pathwise MIMO Interfering Broadcast Channel," in *IEEE 16th Workshop Signal Proc. Advances in Wireless Comm's (SPAWC)*, 2015.

[8] A. Medles and D. Slock, "Matched Filter Bounds without Channel Knowledge at the Receiver," in *Proc. IEEE Asilomar Conf. Signals, Systems and Computers*, Pacific Grove, CA, USA, Nov. 2003.

[9] M. Bashar, Y. Lejosne, D. Slock, and Y. Yuan-Wu, "MIMO Broadcast Channels with Gaussian CSIT and Application to Location based CSIT," in *Proc. IEEE Workshop on Information Theory and Applications (ITA)*, San Diego, CA, USA, Feb. 2014.

[10] S.S. Christensen, R. Agarwal, E. Carvalho, and J. Cioffi, "Weighted sum-rate maximization using weighted MMSE for MIMO-BC beamforming design," *IEEE Trans. on Wireless Communications*, vol. 7, no. 12, pp. 4792–4799, December 2008.

[11] F. Negro, S. Prasad Shenoy, I. Ghauri, and D.T.M. Slock, "On the MIMO Interference Channel," in *Proc. IEEE Information Theory and Applications workshop (ITA)*, San Diego, CA, USA, Feb. 2010.

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

[13] 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.