The Pathwise MIMO Interfering Broadcast Channel

Wassim Tabikh[‡¶], Dirk Slock[‡], Yi Yuan-Wu[¶]

[‡]EURECOM, Sophia-Antipolis, France, Email: wtabikh@gmail.com, slock@eurecom.fr

[¶]Orange Labs, Issy-les-Moulineaux, France, Email: yi.yuan@orange.com

***Abstract*—Interference alignment (IA) through beamforming in MIMO Interfering Broadcast Channels (IBC) allows to handle multi-cell interference with low latency. However, with multiple antennas on both ends, the MIMO setting requires global Channel State Information at the Transmitter (CSIT) (i.e. CSIT from the other transmitters (Tx) also). Though global CSIT can be organized within a cluster, it leads to significant fast CSIT acquisition overhead. In this paper we focus on the (dominant) multipath components in the MIMO propagation channels with only the slow fading components known to the Tx, corresponding to a structured form of covariance CSIT. The pathwise approach allows for a decomposition of the alignment tasks between Tx and receivers (Rx), leading to the sufficiency of local pathwise CSIT plus limited coordination overhead. To optimize Ergodic Weighted Sum Rate at finite SNR, we exploit the uplink/downlink duality to design the Tx beamformers as MMSE filters, in which averaging over complex path amplitudes leads to pathwise filters. We furthermore explore a relation between the difference of convex (DC) functions programming and the Weighted Sum MSE (WSMSE) approaches, indicating significant convergence speed potential for the former, and allowing a fixing of the latter for the case of partial CSIT.**

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 multi-user MISO.

Whereas path CSIT by itself may allow zero forcing (ZF), 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, minimal feedback can be introduced to perform meaningful WSR optimization at a finite SNR (whereas ZF requires much less coordination).

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 [6], [7]

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

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 $\mathbf{h}_t(.)$, $\mathbf{h}_r(.)$ of unit norm, and

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

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

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

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

$$\mathbf{C}_t = \mathbf{A}\mathbf{A}^H + N\beta\, I_M \tag{4}$$

since due to the normalization of the antenna array responses, $\mathrm{E}\,\mathbf{B}^H\mathbf{B} = \mathrm{diag}\{[\mathbf{h}_r(\phi_1)\,\mathbf{h}_r(\phi_2)\cdots]^H\,[\mathbf{h}_r(\phi_1)\,\mathbf{h}_r(\phi_2)\cdots]\} = \mathbf{I}$. 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].

III. 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

$$\mathbf{y}_k = \underbrace{\mathbf{H}_{k,b_k}\,\mathbf{g}_k\,x_k}_{\text{signal}} + \underbrace{\sum_{\substack{i\neq k\\ b_i=b_k}} \mathbf{H}_{k,b_k}\,\mathbf{g}_i\,x_i}_{\text{intracell interf.}} + \underbrace{\sum_{j\neq b_k}\sum_{i:b_i=j} \mathbf{H}_{k,j}\,\mathbf{g}_i\,x_i}_{\text{intercell interf.}} + \mathbf{v}_k \tag{5}$$

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

$$\begin{aligned}\widehat{x}_k &= \mathbf{f}_k^H\mathbf{H}_{k,b_k}\,\mathbf{g}_k\,x_k + \sum_{i=1,\neq k}^{K}\mathbf{f}_k^H\mathbf{H}_{k,b_i}\,\mathbf{g}_i\,x_i + \mathbf{f}_k^H\mathbf{v}_k \\ &= \mathbf{f}_k^H\mathbf{h}_{k,k}\,x_k + \sum_{i\neq k}\mathbf{f}_k^H\mathbf{h}_{k,i}\,x_i + \mathbf{f}_k^H\mathbf{v}_k\end{aligned} \tag{6}$$

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 [8], 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)

$$WSR = WSR(\mathbf{g}) = \sum_{k=1}^{K} u_k\ \ln\frac{1}{e_k} \tag{7}$$

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:

$$\begin{aligned}&\frac{1}{e_k} = 1 + \mathbf{g}_k^H\mathbf{H}_{k,b_k}^H\mathbf{R}_{\overline{k}}^{-1}\mathbf{H}_{k,b_k}\mathbf{g}_k = (1 - \mathbf{g}_k^H\mathbf{H}_{k,b_k}^H\mathbf{R}_k^{-1}\mathbf{H}_{k,b_k}\mathbf{g}_k)^{-1}\\ &\mathbf{R}_k = \mathbf{H}_{k,b_k}\mathbf{Q}_k\,\mathbf{H}_{k,b_k}^H + \mathbf{R}_{\overline{k}}\,,\ \mathbf{Q}_i = \mathbf{g}_i\mathbf{g}_i^H\,,\\ &\mathbf{R}_{\overline{k}} = \sum_{i\neq k}\mathbf{H}_{k,b_i}\mathbf{Q}_i\,\mathbf{H}_{k,b_i}^H + I_{N_k}\,.\end{aligned} \tag{8}$$

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

$$\mathbf{f}_k = \mathbf{R}_k^{-1}\mathbf{H}_{k,b_k}\mathbf{g}_k = \mathbf{R}_k^{-1}\mathbf{h}_{k,k}\,. \tag{9}$$

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

$$\sum_{k:b_k=j}\mathrm{tr}\{\mathbf{Q}_k\} \leq P_j\,. \tag{10}$$

A. From Max WSR to Min WSMSE

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

$$\begin{aligned}&e_k(\mathbf{f}_k,\mathbf{g}) = (1 - \mathbf{f}_k^H\mathbf{H}_{k,b_k}\mathbf{g}_k)(1 - \mathbf{g}_k^H\mathbf{H}_{k,b_k}^H\mathbf{f}_k)\\ &+ \textstyle\sum_{i\neq k}\mathbf{f}_k^H\mathbf{H}_{k,b_i}\mathbf{g}_i\mathbf{g}_i^H\mathbf{H}_{k,b_i}^H\mathbf{f}_k + ||\mathbf{f}_k||^2 = 1 - \mathbf{f}_k^H\mathbf{H}_{k,b_k}\mathbf{g}_k\\ &-\mathbf{g}_k^H\mathbf{H}_{k,b_k}^H\mathbf{f}_k + \sum_i\mathbf{f}_k^H\mathbf{H}_{k,b_i}\mathbf{g}_i\mathbf{g}_i^H\mathbf{H}_{k,b_i}^H\mathbf{f}_k + ||\mathbf{f}_k||^2.\end{aligned} \tag{11}$$

The $WSR(\mathbf{g})$ is a non-convex and complicated function of $\mathbf{g}$. Inspired by [9], we introduced [10], [1] an augmented cost function, the Weighted Sum MSE, $WSMSE(\mathbf{g},\mathbf{f},w)$

$$= \sum_{k=1}^{K}u_k(w_k\,e_k(\mathbf{f}_k,\mathbf{g}) - \ln w_k) + \sum_{i=1}^{C}\lambda_i(\sum_{k:b_k=i}||\mathbf{g}_k||^2 - P_i) \tag{12}$$

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

$$\min_{\mathbf{f},w} WSMSE(\mathbf{g},\mathbf{f},w) = -WSR(\mathbf{g}) + \overbrace{\sum_{k=1}^{K}u_k}^{\text{constant}} \tag{13}$$

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

$$\begin{aligned}&\min_{w_k} WSMSE\ \Rightarrow\ w_k = 1/e_k\\ &\min_{\mathbf{f}_k} WSMSE \Rightarrow \mathbf{f}_k = (\sum_i \mathbf{H}_{k,b_i}\mathbf{g}_i\mathbf{g}_i^H\mathbf{H}_{k.b_i}^H + I_{N_k})^{-1}\mathbf{H}_{k,b_k}\mathbf{g}_k\\ &\min_{\mathbf{g}_k} WSMSE\ \Rightarrow\\ &\mathbf{g}_k = (\textstyle\sum_i u_i w_i \mathbf{H}_{i,b_k}^H\mathbf{f}_i\mathbf{f}_i^H\mathbf{H}_{i,b_k} + \lambda_{b_k}I_M)^{-1}\mathbf{H}_{k,b_k}^H\mathbf{f}_k u_k w_k\end{aligned} \tag{14}$$

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 [11] 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

$$\begin{aligned} WSR &= u_k \ln\det(\mathbf{R}_{\overline{k}}^{-1}\mathbf{R}_k) + WSR_{\overline{k}}, \\ WSR_{\overline{k}} &= \textstyle\sum_{i=1,\neq k}^{K} u_i \ln\det(\mathbf{R}_{\overline{i}}^{-1}\mathbf{R}_i) \end{aligned} \quad (15)$$

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

$$\begin{aligned} &WSR_{\overline{k}}(\mathbf{Q}_k, \widehat{\mathbf{Q}}) \approx WSR_{\overline{k}}(\widehat{\mathbf{Q}}_k, \widehat{\mathbf{Q}}) - \mathrm{tr}\{(\mathbf{Q}_k - \widehat{\mathbf{Q}}_k)\widehat{\mathbf{T}}_k\} \\ &\widehat{\mathbf{T}}_k = -\left.\frac{\partial WSR_{\overline{k}}(\mathbf{Q}_k, \widehat{\mathbf{Q}})}{\partial \mathbf{Q}_k}\right|_{\widehat{\mathbf{Q}}_k, \widehat{\mathbf{Q}}} = \sum_{i\neq k}^{K} u_i \mathbf{H}_{i,b_k}^H (\widehat{\mathbf{R}}_{\overline{i}}^{-1} - \widehat{\mathbf{R}}_i^{-1})\mathbf{H}_{i,b_k} \end{aligned} \quad (16)$$

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

$$\begin{aligned} &WSR(\mathbf{g}, \widehat{\mathbf{g}}, \lambda) = \sum_{j=1}^{C} \lambda_j P_j + \\ &\sum_{k=1}^{K} u_k \ln(1 + \mathbf{g}_k^H \widehat{\mathbf{S}}_k \mathbf{g}_k) - \mathbf{g}_k^H (\widehat{\mathbf{T}}_k + \lambda_{b_k} I)\mathbf{g}_k \end{aligned} \quad (17)$$

where

$$\widehat{\mathbf{S}}_k = \mathbf{H}_{k,b_k}^H \widehat{\mathbf{R}}_{\overline{k}}^{-1} \mathbf{H}_{k,b_k} \;. \quad (18)$$

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

$$\widehat{\mathbf{S}}_k \, \mathbf{g}_k = \frac{1 + \mathbf{g}_k^H \widehat{\mathbf{S}}_k \mathbf{g}_k}{u_k} \, (\widehat{\mathbf{T}}_k + \lambda_{b_k} I)\mathbf{g}_k \quad (19)$$

or hence $\mathbf{g}_k^{'} = V_{max}(\widehat{\mathbf{S}}_k, \widehat{\mathbf{T}}_k + \lambda_{b_k} I)$ is the (normalized) "max" generalized eigenvector of the two indicated matrices, with max eigenvalue $\sigma_k = \sigma_{max}(\widehat{\mathbf{S}}_k, \widehat{\mathbf{T}}_k + \lambda_{b_k} I)$. Let $\sigma_k^{(1)} = \mathbf{g}_k^{'H} \widehat{\mathbf{S}}_k \mathbf{g}_k^{'}$, $\sigma_k^{(2)} = \mathbf{g}_k^{'H} \widehat{\mathbf{T}}_k \mathbf{g}_k^{'}$. The advantage of formulation (17) 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 (17) yields

$$WSR = \sum_{j}^{C} \lambda_j P_j + \sum_{k=1}^{K} \{u_k \ln(1 + p_k \sigma_k^{(1)}) - p_k(\sigma_k^{(2)} + \lambda_{b_k})\} \quad (20)$$

which leads to the following interference leakage aware water filling

$$p_k = \left(\frac{u_k}{\sigma_k^{(2)} + \lambda_{b_k}} - \frac{1}{\sigma_k^{(1)}} \right)^{+} \quad (21)$$

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.

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 $\mathrm{E}_{\mathbf{H}} WSR(\mathbf{g}, \mathbf{H}) =$

$$EWSR(\mathbf{g}) = \mathrm{E}_{\mathbf{H}} \sum_k u_k \ln(1 + \mathbf{g}_k^H \mathbf{H}_{k,b_k}^H \mathbf{R}_{\overline{k}}^{-1} \mathbf{H}_{k,b_k} \mathbf{g}_k) \quad (22)$$

where we now underlign the dependence of various quantities on $\mathbf{H}$. The EWSR in (22) 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 [12] we propose **various deterministic approximations for the EWSR**, which can then be optimized as in the full CSI case. In the rest of this paper we propose an application of these techniques to a pathwise design.

VI. Min WSMSE - DC Programming Relation and EWSR

In [7, Section VI 1)], we explain the substantial suboptimality associated with the Expected WSMSE (EWSMSE) approach for which we propose a possible correction here. Consider the min WSMSE at iteration $(i+1)$

$$\begin{aligned} \mathbf{T}_k^{(i)} &= \textstyle\sum_j u_j w_j^{(i)} \mathbf{H}_{j,b_k}^H \mathbf{f}_j^{(i)} \mathbf{f}_j^{(i)H} \mathbf{H}_{j,b_k} + \lambda_{b_k}^{(i)} I_M \\ \mathbf{g}_k^{(i+1)} &= (\mathbf{T}_k^{(i)})^{-1} \mathbf{H}_{k,b_k}^H \mathbf{f}_k^{(i)} u_k w_k^{(i)} \\ &= (\mathbf{T}_k^{(i)})^{-1} \mathbf{S}_k^{(i)} \, \mathbf{g}_k^{(i)} \, u_k w_k^{(i)} \\ \mathbf{S}_k^{(i)} &= \mathbf{H}_{k,b_k}^H \mathbf{R}_k^{-(i)} \mathbf{H}_{k,b_k} \end{aligned} \quad (23)$$

where with some abuse of notation compared to (16) we now integrated the Lagrange multiplier in $\mathbf{T}_k$. In other words, one WSMSE iteration does just one power iteration of the DC programming approach (!) which computes the max eigenvector explicitely:

$$\mathbf{g}_k^{(i+1)} = V_{max}\{(\mathbf{T}_k^{(i)})^{-1} \mathbf{S}_k^{(i)}\} = V_{max}\{\mathbf{S}_k^{(i)}, \mathbf{T}_k^{(i)}\} \quad (24)$$

in which we relate classical and generalized eigenvectors.

For the case of partial CSIT (or MaMIMO) case: we can consider replacing the EWSMSE operation

$$\mathbf{g}_k^{(i+1)} = (\,\mathrm{E}_H \mathbf{T}_k^{(i)})^{-1} \overline{\mathbf{H}}_{k,b_k}^H \mathbf{f}_k^{(i)} u_k w_k^{(i)} \tag{25}$$

by the following modified WSMSE:

$$\mathbf{g}_k^{(i+1)} = (\,\mathrm{E}_H \mathbf{T}_k^{(i)})^{-1} (\,\mathrm{E}_H \mathbf{S}_k^{(i)})\, \mathbf{g}_k^{(i)}\, u_k w_k^{(i)} \tag{26}$$

which acknowledges the quadratic appearance of $\mathbf{H}_{k,b_k}$ in the cascade of $\mathbf{H}_{k,b_k}^H \mathbf{f}_k$ and hence in $\mathbf{S}_k$ and accounts for both channel mean and covariance in the expectation operation. This by itself constitues a significant improvement over the EWSMSE approach for the case of partial CSIT, but the partial CSIT variants of the DC programming approach proposed in [12] are still more powerful.

VII. Pathwise Intercell BF Design

The main issue we want to resolve here is to propose a pathwise design that is optimized at finite SNR. At high SNR, the interference along all intercell paths needs to be forced to zero, but at finite SNR, an optimal weighting and regularization needs to be performed. Note that the paths arriving at a UE from different BS in the multi-cell scenario are different. Hence one cannot simply formulate a SINR at every received path. We shall design the DL BF as an uplink (UL) LMMSE Rx in which we receive different degrees of interference from the different paths.

A. DL BF as Dual UL LMMSE Rx

We recall the DL Rx signal at user k in cell b_k :

$$\mathbf{y}_k = \underbrace{\mathbf{H}_{k,b_k}\, \mathbf{g}_k\, x_k}_{\text{signal}} + \underbrace{\sum_{\substack{i\neq k\\ b_i=b_k}} \mathbf{H}_{k,b_k}\, \mathbf{g}_i\, x_i}_{\text{intracell interf.}} + \underbrace{\sum_{j\neq b_k} \sum_{i:b_i=j} \mathbf{H}_{k,j}\, \mathbf{g}_i\, x_i}_{\text{intercell interf.}} + \mathbf{v}_k \tag{27}$$

or at the output of the Rx we get:

$$\widehat{x}_k = \mathbf{f}_k^H \mathbf{H}_{k,b_k}\, \mathbf{g}_k\, x_k + \sum_{i=1,\neq k}^{K} \mathbf{f}_k^H \mathbf{H}_{k,b_i}\, \mathbf{g}_i\, x_i + \mathbf{f}_k^H \mathbf{v}_k\ . \tag{28}$$

In the dual UL we can consider the Rx signal at BS k

$$\widetilde{\mathbf{y}}_k = \underbrace{\sum_{i:b_i=k} \mathbf{H}_{i,k}^H\, \mathbf{f}_i\, \widetilde{x}_i}_{\text{intracell users}} + \underbrace{\sum_{i:b_i\neq k} \mathbf{H}_{i,k}^H\, \mathbf{f}_i\, \widetilde{x}_i}_{\text{intercell users}} + \widetilde{\mathbf{v}}_k \tag{29}$$

where the ficitious dual UL Tx signals $\widetilde{x}_i$ are uncorrelated zero mean with variance $\sigma^2_{\widetilde{x}_i} = u_i\, w_i$ and the fictitious dual UL Rx noise has covariance matrix $R_{\widetilde{\mathbf{v}}_k \widetilde{\mathbf{v}}_k} = \lambda_k\, I_{M_k}$. The dual UL Rx signal at BS k (29) can be rewritten as

$$\begin{aligned}
\widetilde{\mathbf{y}}_k &= \underbrace{\mathbf{H}_k^H\, \mathbf{F}_k\, \widetilde{\mathbf{x}}_k}_{\text{intracell users}} + \underbrace{\mathbf{H}_{\overline{k}}^H\, \mathbf{F}_{\overline{k}}\, \widetilde{\mathbf{x}}_{\overline{k}}}_{\text{intercell users}} + \widetilde{\mathbf{v}}_k \\
\mathbf{H}_k^H &= [\mathbf{H}_{k,m_k+1}^H \cdots \mathbf{H}_{k,m_k+K_k}^H], \\
\mathbf{H}_{\overline{k}}^H &= [\mathbf{H}_1^H \cdots \mathbf{H}_{k-1}^H\, \mathbf{H}_{k+1}^H \cdots \mathbf{H}_C^H], \\
\mathbf{F}_k &= \text{blockdiag}\,\{\mathbf{f}_{m_k+1}, \ldots, \mathbf{f}_{m_k+K_k}\}, \\
\mathbf{F}_{\overline{k}} &= \text{blockdiag}\,\{\mathbf{F}_1, \ldots, \mathbf{F}_{k-1}\, \mathbf{F}_{k+1}, \ldots, \mathbf{F}_C\}
\end{aligned} \tag{30}$$

$m_k = \sum_{i=1}^{k-1} K_i$ and corresponding block structure for the super vectors $\widetilde{\mathbf{x}}_k$, $\widetilde{\mathbf{x}}_{\overline{k}}$. This leads to the DL BF as an UL LMMSE Rx: (for all intracell users jointly)

$$\mathbf{G}_k^H = R_{\widetilde{\mathbf{x}}_k \widetilde{\mathbf{y}}_k}\, R^{-1}_{\widetilde{\mathbf{y}}_k \widetilde{\mathbf{y}}_k} = (\,\mathrm{E}_{\widetilde{\mathbf{x}},\widetilde{\mathbf{v}}} \widetilde{\mathbf{x}}_k \widetilde{\mathbf{y}}_k)\,(\,\mathrm{E}_{\widetilde{\mathbf{x}},\widetilde{\mathbf{v}}} \widetilde{\mathbf{y}}_k \widetilde{\mathbf{y}}_k)^{-1} \tag{31}$$

which can be seen to correspond to the expressions for the $\mathbf{g}_k$ in (14).

B. Pathwise Dual UL

Substituting the channel response matrices in terms of their pathwise factored form in (29), we get for the pathwise dual UL at BS k

$$\begin{aligned}
\widetilde{\mathbf{y}}_k \;=\;& \sum\nolimits_{i:b_i=k} \mathbf{A}_{i,k}\, \underbrace{\mathbf{B}_{i,k}^H \mathbf{f}_i\, \widetilde{x}_i}_{\widetilde{\mathbf{s}}_{i,k}\ \text{intracell paths}} \\
+\;& \sum\nolimits_{i:b_i\neq k} \mathbf{A}_{i,k}\, \underbrace{\mathbf{B}_{i,k}^H \mathbf{f}_i\, \widetilde{x}_i}_{\widetilde{\mathbf{s}}_{i,k}\ \text{intercell paths}} + \widetilde{\mathbf{v}}_k
\end{aligned} \tag{32}$$

where the $\widetilde{\mathbf{s}}_{i,k}$ are the (vectors of) fictious pathwise UL Tx signals from user i to BS k. The factors $\mathbf{B}_{i,k}$ (see (2)) are now treated as unknown, and are modeled as independent with zero mean i.i.d. elements of variance $\frac{1}{N_i}$. As a result we get for the correlation matrices $R_{\widetilde{\mathbf{s}}_{i,k} \widetilde{\mathbf{s}}_{i,k}} = \frac{||\mathbf{f}_i||^2}{N_i} \sigma^2_{\widetilde{x}_i} I$.

Similarly to (30), the pathwise dual UL Rx signal at BS k (32) can be rewritten as

$$\widetilde{\mathbf{y}}_k = \mathbf{A}_k \underbrace{\widetilde{\mathbf{s}}_k}_{\text{intracell paths}} + \mathbf{A}_{\overline{k}} \underbrace{\widetilde{\mathbf{s}}_{\overline{k}}}_{\text{intercell paths}} + \widetilde{\mathbf{v}}_k \tag{33}$$

where $\widetilde{\mathbf{s}}_k = \mathbf{B}_k^H \mathbf{F}_k \widetilde{\mathbf{x}}_k$, $\widetilde{\mathbf{s}}_{\overline{k}} = \mathbf{B}_{\overline{k}}^H \mathbf{F}_{\overline{k}} \widetilde{\mathbf{x}}_{\overline{k}}$ and $\mathbf{A}_k$,$\mathbf{B}_k$ and $\mathbf{A}_{\overline{k}}$,$\mathbf{B}_{\overline{k}}$ have similar block structure as $\mathbf{H}_k$ and $\mathbf{H}_{\overline{k}}$ resp. except for different block sizes. Hence we get the pathwise DL BF as an UL LMMSE Rx (for all intracell paths jointly)

$$\begin{aligned}
\widetilde{\mathbf{G}}_k^H &= R_{\widetilde{\mathbf{s}}_k \widetilde{\mathbf{y}}_k}\, R^{-1}_{\widetilde{\mathbf{y}}_k \widetilde{\mathbf{y}}_k} = (\,\mathrm{E}_{\widetilde{\mathbf{x}},\widetilde{\mathbf{v}},\mathbf{B}}\, \widetilde{\mathbf{s}}_k \widetilde{\mathbf{y}}_k)\,(\,\mathrm{E}_{\widetilde{\mathbf{x}},\widetilde{\mathbf{v}},\mathbf{B}}\, \widetilde{\mathbf{y}}_k \widetilde{\mathbf{y}}_k)^{-1} \\
&= R_{\widetilde{\mathbf{s}}_k \widetilde{\mathbf{s}}_k} \mathbf{A}_k^H (\mathbf{A}_k R_{\widetilde{\mathbf{s}}_k \widetilde{\mathbf{s}}_k} \mathbf{A}_k^H + \mathbf{A}_{\overline{k}} R_{\widetilde{\mathbf{s}}_{\overline{k}} \widetilde{\mathbf{s}}_{\overline{k}}} \mathbf{A}_{\overline{k}}^H + \lambda_k I)^{-1}
\end{aligned} \tag{34}$$

where $R_{\widetilde{\mathbf{s}}_k \widetilde{\mathbf{s}}_k}$, $R_{\widetilde{\mathbf{s}}_{\overline{k}} \widetilde{\mathbf{s}}_{\overline{k}}}$ are diagonal.

The proposed pathwise approach is similar in spirit to chip equalization. The chip equalizer is a CDMA downlink linear MMSE receiver in which expectations are not only w.r.t. transmitted symbols and noise, but also w.r.t. the scrambling sequence.

C. 2-stage BF Design: Pathwise Intercell + Userwise Intracell

In this subsection we propose two-stage designs in which the pathwise approach is applied for the intercell interference but the intracell interference is handled in a more classical way, based on user channel estimates. In a first stage, the pathwise intercell interference cancellation allows improved intracell channel estimation. Consider the pathwise BF from (34)

$$\begin{aligned}
\widetilde{\mathbf{G}}_k^H &= R_{\widetilde{\mathbf{s}}_k \widetilde{\mathbf{s}}_k} \mathbf{A}_k^H (\mathbf{A}_k R_{\widetilde{\mathbf{s}}_k \widetilde{\mathbf{s}}_k} \mathbf{A}_k^H + \mathbf{A}_{\overline{k}} R_{\widetilde{\mathbf{s}}_{\overline{k}} \widetilde{\mathbf{s}}_{\overline{k}}} \mathbf{A}_{\overline{k}}^H + \lambda_k I)^{-1} \\
&= \underbrace{(\mathbf{A}_k^H R^{-1} \mathbf{A}_k + I)^{-1}}_{\text{stage 2}} \underbrace{\mathbf{A}_k^H R^{-1}}_{\text{stage 1}}
\end{aligned} \tag{35}$$

where $R = \mathbf{A}_{\overline{k}} R_{\widetilde{\mathbf{s}}_{\overline{k}}\widetilde{\mathbf{s}}_{\overline{k}}} \mathbf{A}_{\overline{k}}^H + \lambda_k I$.

1) Stage 1: intercell path suppression: In a straightforward variant,

$$\mathbf{A}_k^H(\mathbf{A}_{\overline{k}} R_{\widetilde{\mathbf{s}}_{\overline{k}}\widetilde{\mathbf{s}}_{\overline{k}}} \mathbf{A}_{\overline{k}}^H + \lambda_k I)^{-1} \tag{36}$$

which allows pilot transmission without intercell path interference, but with intracell interference.

2) Stage 1': intracell and intercell path suppression: In a second variant

$$\mathbf{A}_{i,k}^H(\mathbf{A}_{\bar{i},k} R_{\widetilde{\mathbf{s}}_{\bar{i},k}\widetilde{\mathbf{s}}_{\bar{i},k}} \mathbf{A}_{\bar{i},k}^H + \mathbf{A}_{\overline{k}} R_{\widetilde{\mathbf{s}}_{\overline{k}}\widetilde{\mathbf{s}}_{\overline{k}}} \mathbf{A}_{\overline{k}}^H + \lambda_k I)^{-1} \tag{37}$$

which allows pilot transmission on one user's paths without any interference from paths of any other user. Hence this allows the very short training lenth of just the maximum of the number of paths of a user. Note that the BFs in stage 1' mimimize the dual weighted sum MSE at the path outputs

$$\widetilde{u}_i\widetilde{w}_i\widetilde{f}_i\widetilde{f}_i^H = (R_{\widetilde{\mathbf{s}}\widetilde{\mathbf{s}}})_{i,i}, \widetilde{\mathbf{g}}_i \Rightarrow \widetilde{f}_i, \widetilde{w}_i = 1/\widetilde{e}_i = 1/(1-\widetilde{f}_i\mathbf{A}_i\widetilde{\mathbf{g}}_i)\,. \tag{38}$$

It is not clear if this weighting is optimal for channel estimation also, but intuitively, it goes in the right direction (considering very strong or very weak paths).

3) Stage 2: from paths to user signals (intracell): LMMSE extraction of user signals from Rx signal can be written as a cascade of LMMSE extraction of user signals from pathwise signal estimates and LMMSE extraction of pathwise signals from Rx signal:

$$R_{\widetilde{\mathbf{x}}_k\widetilde{\mathbf{y}}_k} = \underbrace{R_{\widetilde{\mathbf{x}}_k\widetilde{\mathbf{s}}_k} R^{-1}_{\widetilde{\mathbf{s}}_k\widetilde{\mathbf{s}}_k}}_{\text{LMMSE: paths} \to \text{users}} R_{\widetilde{\mathbf{s}}_k\widetilde{\mathbf{y}}_k} \tag{39}$$

where the crosscorrelation matrix of user and path signals is

$$R_{\widetilde{\mathbf{x}}_k\widetilde{\mathbf{s}}_k} = R_{\widetilde{\mathbf{x}}_k\widetilde{\mathbf{x}}_k}\mathbf{F}_k^H\mathbf{B}_k\,. \tag{40}$$

D. Relation to Cognitive Radio Design

The pathwise intercell design can be interpreted as a cellwise intracell design with intercell interference constraints of the form (for cell i)

$$\sum_{k:b_k=i} u_k \,\ln\frac{1}{e_k} + \sum_{k:b_k=i}\sum_{n:b_n\neq i}\mu_{k,n}(|\mathbf{g}_k^H\mathbf{H}_{k,b_n}^H\mathbf{f}_n|^2 - Q_{k,n}) \tag{41}$$

where the $\mu_{k,n} = \sigma^2_{\widetilde{x}_n}$ are Lagrange multipliers and the $Q_{k,n}$ are linkwise interference power constraints. The pathwise approach is obtained by replacing the second term by its expected value w.r.t. the $\mathbf{B}$ factors, leading for the quadratic terms to

$$\mathrm{E}_{\mathbf{B}}\sum_{k:b_k=i}\sum_{n:b_n\neq i}\mu_{k,n}|\mathbf{g}_k^H\mathbf{H}_{k,b_n}^H\mathbf{f}_n|^2 = \sum_{k:b_k=i}\mathbf{g}_k^H\,\mathbf{A}_{\overline{k}} R_{\widetilde{\mathbf{s}}_{\overline{k}}\widetilde{\mathbf{s}}_{\overline{k}}} \mathbf{A}_{\overline{k}}^H\,\mathbf{g}_k\,. \tag{42}$$

The pathwise philosophy corresponds to no intercell exchange of fast fading information. Hence the intercell exchange involves long-term averages for $\sigma^2_{\widetilde{x}_n} = u_n/e_n$ and for the noise variance (which includes residual intercell interference).

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] 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] 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.

[7] 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.

[8] 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.

[9] 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.

[10] 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.

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

[12] Y. Lejosne, A. Ben Nasser, D. Slock, and Y. Yuan-Wu, "Multi-Cell Multi-User MIMO Downlink with Partial CSIT and Decentralized Design," in *Proc. IEEE Broadband Wireless Access (BWA) Workshop at Globecom*, Austin, TX, USA, Dec. 2014.