Variational Bayesian Learning for Channel Estimation and Transceiver Determination

Christo Thomas Kurisummoottil, Dirk Slock,
EURECOM, Sophia-Antipolis, France, Email: {kurisumm,slock}@eurecom.fr

***Abstract*—We consider the design of linear precoders for the MISO Interfering Broadcast Channel (IBC) with partial Channel State Information at the Transmitter (CSIT) in the form of both channel estimates and channel covariance information. Most of the results can also be transposed to the SIMO Interfering Multiple Access Channel with linear receivers. We first point out that in the case of reduced rank covariance matrices, there is significant gain in sum rate by using LMMSE as opposed to Least-Squares (LS) channel estimates. We also analyze various beamforming designs exploiting partial CSIT. Then we go beyond assuming the availability of covariance CSIT and propose variational Bayesian learning (VBL) techniques to acquire it assuming TDD channel reciprocity. In particular a Space Alternating version of Variational Estimation (SAVE) allows a well founded alternative to AMP based techniques while being of similar complexity. The SAVE techniques can also be applied to obtain reduced complexity iterative techniques for determining the transmit/receive signals or beamformers themselves.**

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 recent development of Massive MIMO (MaMIMO) [1] opens new possibilities for increased system capacity while at the same time simplifying system design. We refer to [2] for a further discussion of the state of the art, in which MIMO Interference Alignment (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. [3] proposes optimal beamformers (BFs) in the case of partial CSIT in the massive MIMO limit.

Gaussian (Posterior) partial CSIT can optimally combine channel estimate and channel covariance information. However, the posterior covariance computation requires matrix inversion on the $O(M^3)$ which is computationally cumbersome. The huge number of antennas M at the base station and K, the number of users imposes a very high computational complexity on the massive MIMO systems. Computing the precoding/detection matrices from the estimated channel also require matrix inversions which consists of $O(K^3)$ and $O(MK^2)$ operations. This is the main motivation behind searching for low complexity solutions with close to optimal performance. [4] proposes truncated polynomial expansion (PE) for reducing precoder complexity. [5] introduces a non-parametric algorithm called NOPE that doesn't require any knowledge of the signal and noise powers. The authors also prove that in the large system limit, NOPE achieves the same performance as that of the Linear Minimum Mean Squared Error (LMMSE) equalizer. [6] showed that the design of all variants of linear precoder/combiners for the downlink (DL) and uplink (UL) can be posed as the solution of a set of linear equations. Furthermore, this is solved using Kaczmarz method, which is essentially the Normalized LMS algorithm from adaptive filtering, applied to a randomized selection of the normal equations to be satisfied.

In this paper, we propose Bayesian learning techniques in compressed sensing to tackle the two major MaMIMO issues discussed above, namely computational complexity and (LMMSE) CSIT acquisition. In compressed sensing, the system model is $\boldsymbol{y} = \mathbf{A}\boldsymbol{x} + \boldsymbol{w}$, where $\mathbf{A} \in \mathcal{R}^{N\times M}$ is the measurement matrix, $\boldsymbol{x} \in \mathcal{R}^{M}$ is a possibly sparse vector, $\boldsymbol{w}$ is the noise vector and $\boldsymbol{y}$ the observation or data. In sparse Bayesian learning (SBL), a hierarchical prior distribution structure is assumed, in which each element of $\boldsymbol{x}$ is conditionally Gaussian. The inverse variance α_i of $\boldsymbol{x}_i$ is Gamma distributed which leads to a sparsity inducing distribution for $\boldsymbol{x}$. Despite its superior performance, SBL has high computational complexity since it requires matrix inversions [7]. In [8], the authors propose a Fast Marginalized Maximum Likelihood (FMML) by alternating maximization of the hyperparameters α_i. [9] introduces a Fast version of SBL by alternatingly maximizing the variational posterior lower bound with respect to single (hyper)parameters. They analytically show that the stationary points for the α_i are the same as those of FMML, provide the pruning conditions and thus accelerate the convergence. Both previous approaches allow for a greedy initialization (OMP-like) which improves convergence speed and handles initialization issues. [10] uses AMP to approximate matrix inversions in SBL and uses EM to update the precision parameters α_i. [11] introduces inverse-free SBL via a Taylor series expansion.

A. Contributions of this paper

In this paper:

- We first review optimal BFs for the expected weighted sum rate (EWSR) criterion in the MaMIMO limit.
- We evaluate the ergodic sum rate performance for LS, LMMSE and subspace projection channel estimators. Numerical results suggest that there is substantial gain by exploiting the channel covariance information compared to just using the LS estimates.
- We propose a novel Space Alternating Variational Estimation (SAVE) based SBL technique called SAVE.
- We suggest the use of SAVE to compute the LMMSE Tx/Rx signal in the UL/DL.
- We also extend the SAVE techniques to compute the LMMSE channel estimates, by estimating the parameters for both the instantaneous channels and their covariances.

II. Streamwise IBC Signal Model

We start with 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

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

where x_k is the intended (white, unit variance) scalar signal stream, $\mathbf{H}_{k,b_k}$ is the $N_k \times M_{b_k}$ channel from BS b_k to user k. BS b_k serves $K_{b_k} = \sum_{i:b_i=b_k} 1$ users. We are considering a noise whitened signal representation so that we get for the noise $\boldsymbol{v}_k \sim \mathcal{CN}(0, I_{N_k})$. 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 $\widehat{x}_k = \mathbf{f}_k^H \boldsymbol{y}_k$.

III. Max WSR with Perfect CSIT

We consider the optimization of the beamformers based on the max WSR criterion subject to a per base station power constraint,

$$\begin{aligned} &WSR = WSR(\mathbf{g}) = \textstyle\sum_{k=1}^K u_k\ \ln \frac{1}{e_k} \\ &\textstyle\sum_{k:b_k=j} \text{tr}\{\boldsymbol{Q}_k\} \leq P_j\,, \end{aligned} \tag{2}$$

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:

$$\begin{aligned} &\frac{1}{e_k} = 1 + \mathbf{g}_k^H \mathbf{H}_{k,b_k}^H \boldsymbol{R}_{\overline{k}}^{-1} \mathbf{H}_{k,b_k} \mathbf{g}_k = (1 - \mathbf{g}_k^H \mathbf{H}_{k,b_k}^H \boldsymbol{R}_k^{-1} \mathbf{H}_{k,b_k} \mathbf{g}_k)^{-1} \\ &\boldsymbol{R}_k = \mathbf{H}_{k,b_k} \boldsymbol{Q}_k\, \mathbf{H}_{k,b_k}^H + \boldsymbol{R}_{\overline{k}}\,,\ \boldsymbol{Q}_i = \mathbf{g}_i \mathbf{g}_i^H\,, \\ &\boldsymbol{R}_{\overline{k}} = \sum_{i\neq k} \mathbf{H}_{k,b_i} \boldsymbol{Q}_i\, \mathbf{H}_{k,b_i}^H + I_{N_k}\,. \end{aligned} \tag{3}$$

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

$$\mathbf{f}_k = \boldsymbol{R}_k^{-1} \mathbf{H}_{k,b_k} \mathbf{g}_k = \boldsymbol{R}_k^{-1} \mathbf{h}_{k,k}\,. \tag{4}$$

A. From Max WSR to Min WSMSE

The expression for MSE with a general Rx filter $\mathbf{f}_k$ is,

$$\begin{aligned} &e_k(\mathbf{f}_k, \mathbf{g}) = (1 - \mathbf{f}_k^H \mathbf{H}_{k,b_k} \mathbf{g}_k)(1 - \mathbf{g}_k^H \mathbf{H}_{k,b_k}^H \mathbf{f}_k) \\ &+ \textstyle\sum_{i\neq k} \mathbf{f}_k^H \mathbf{H}_{k,b_i} \mathbf{g}_i \mathbf{g}_i^H \mathbf{H}_{k,b_i}^H \mathbf{f}_k + ||\mathbf{f}_k||^2 = 1 - \mathbf{f}_k^H \mathbf{H}_{k,b_k} \mathbf{g}_k \\ &- \mathbf{g}_k^H \mathbf{H}_{k,b_k}^H \mathbf{f}_k + \sum_i \mathbf{f}_k^H \mathbf{H}_{k,b_i} \mathbf{g}_i \mathbf{g}_i^H \mathbf{H}_{k,b_i}^H \mathbf{f}_k + ||\mathbf{f}_k||^2. \end{aligned} \tag{5}$$

The $WSR(\mathbf{g})$ is a non-convex and complicated function of $\mathbf{g}$. Inspired by [12], an augmented cost function is introduced in [13], [14], the Weighted Sum MSE,

$$\begin{aligned} &WSMSE(\mathbf{g}, \mathbf{f}, w) = \\ &\textstyle\sum_{k=1}^K u_k (w_k\ e_k(\mathbf{f}_k, \mathbf{g}) - \ln w_k) + \sum_{i=1}^C \lambda_i (\sum_{k:b_k=i} ||\mathbf{g}_k||^2 - P_i) \end{aligned} \tag{6}$$

where λ_i are Lagrange multipliers. Optimization over the aggregate auxiliary Rx filters $\mathbf{f}$ and weights w, lead to the original WSR expression:

$$\min_{\mathbf{f}, w} WSMSE(\mathbf{g}, \mathbf{f}, w) = -WSR(\mathbf{g}) + \sum_{k=1}^K u_k \tag{7}$$

From the augumented cost function, alternating optimization leads to solving simple quadratic or convex functions as follows,

$$\begin{aligned} &\min_{w_k} WSMSE\ \Rightarrow\ w_k = 1/e_k \\ &\min_{\mathbf{f}_k} WSMSE \Rightarrow \mathbf{f}_k = (\sum_i \mathbf{H}_{k,b_i} \mathbf{g}_i \mathbf{g}_i^H \mathbf{H}_{k,b_i}^H + I_{N_k})^{-1} \mathbf{H}_{k,b_k} \mathbf{g}_k \\ &\min_{\mathbf{g}_k} WSMSE\ \Rightarrow \\ &\mathbf{g}_k = (\textstyle\sum_i u_i w_i \mathbf{H}_{i,b_k}^H \mathbf{f}_i \mathbf{f}_i^H \mathbf{H}_{i,b_k} + \lambda_{b_k} I_M)^{-1} \mathbf{H}_{k,b_k}^H \mathbf{f}_k u_k w_k \end{aligned} \tag{8}$$

UL/DL duality: the optimal Tx filter g_k can also be interpreted as an 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. Minorization (DC Programming)

In a classical difference of convex functions (DC programming) approach (also called Successive Convex Approximation (SCA)) as in [15], the cost function is written as the sum of a convex and concave terms. The concave signal terms are kept and the convex interference terms are replaced by the linear (and hence concave) tangent approximation. This linearization is with respect to the Tx covariance matrix $\boldsymbol{Q}_k$. However, after substituting $\boldsymbol{Q}_k = \boldsymbol{G}_k \boldsymbol{G}_k^H$ in terms of

BF matrices $\boldsymbol{G}_k$, the concave character is less clear. But in any case, this DC programming/SCA approximation allows to construct a minorizer cost function, and minorization is a well established optimization approach [16].

So, consider the WSR as a function of $\boldsymbol{Q}_k$ alone. Then

$$\begin{aligned} WSR &= u_k \ln\det(\boldsymbol{R}_{\overline{k}}^{-1}\boldsymbol{R}_k) + WSR_{\overline{k}}\,, \\ WSR_{\overline{k}} &= \textstyle\sum_{i=1,\neq k}^{K} u_i \ln\det(\boldsymbol{R}_{\overline{i}}^{-1}\boldsymbol{R}_i) \end{aligned} \tag{9}$$

where $\ln\det(\boldsymbol{R}_{\overline{k}}^{-1}\boldsymbol{R}_k)$ is concave in $\boldsymbol{Q}_k$ and $WSR_{\overline{k}}$ is convex in $\boldsymbol{Q}_k$. A linear function is simultaneously convex and concave. So we consider the first order Taylor series expansion of $WSR_{\overline{k}}$ in $\boldsymbol{Q}_k$ around the current[1] $\boldsymbol{Q}^{'}$ (i.e. all $\boldsymbol{Q}_i^{'}$) with e.g. $\boldsymbol{R}_i = \boldsymbol{R}_i(\boldsymbol{Q}^{'})$, then

$$\begin{aligned} &WSR_{\overline{k}}(\boldsymbol{Q}_k,\boldsymbol{Q}^{'}) \approx WSR_{\overline{k}}(\boldsymbol{Q}_k^{'},\boldsymbol{Q}^{'}) - \mathrm{tr}\{(\boldsymbol{Q}_k-\boldsymbol{Q}_k^{'})\mathbf{A}_k\} \\ &\mathbf{A}_k = - \left.\frac{\partial WSR_{\overline{k}}(\boldsymbol{Q}_k,\boldsymbol{Q}^{'})}{\partial \boldsymbol{Q}_k}\right|_{\boldsymbol{Q}_k^{'},\boldsymbol{Q}^{'}} = \sum_{i\neq k}^{K} u_i \mathbf{H}_{i,b_k}^H(\boldsymbol{R}_{\overline{i}}^{-1} - \boldsymbol{R}_i^{-1})\mathbf{H}_{i,b_k} \end{aligned} \tag{10}$$

Note that the linearized (tangent) expression for $WSR_{\overline{k}}$ constitutes a lower bound for it. Now, dropping constant terms, reparameterizing the $Q_k = \boldsymbol{G}_k\boldsymbol{G}_k^H$ and perform this linearization for all users. We augment the WSR cost function with the Tx power constraints, resulting in the Lagrangian,

$$\begin{aligned} &WSR(\boldsymbol{G},\boldsymbol{G}^{'},\lambda) = \sum_{j=1}^{C}\lambda_j P_j + \\ &\sum_{k=1}^{K} u_k \ln\det(1+\boldsymbol{G}_k^H\mathbf{B}_k\boldsymbol{G}_k) - \boldsymbol{G}_k^H(\mathbf{A}_k+\lambda_{b_k}I)\boldsymbol{G}_k \end{aligned} \tag{11}$$

where

$$\mathbf{B}_k = \mathbf{H}_{k,b_k}^H \boldsymbol{R}_{\overline{k}}^{-1}\mathbf{H}_{k,b_k}\;. \tag{12}$$

The gradient (w.r.t. $\boldsymbol{G}_k$) of this concave WSR lower bound is actually still the same as that of the original WSR criterion! And it can be interpreted as a generalized eigenvector condition,

$$\mathbf{B}_k\,\boldsymbol{G}_k = (\mathbf{A}_k+\lambda_{b_k}I)\boldsymbol{G}_k\,\frac{1}{u_k}(I+\boldsymbol{G}_k^H\mathbf{B}_k\boldsymbol{G}_k) \tag{13}$$

or hence $\overline{\boldsymbol{G}}_k = V_{max}(\mathbf{B}_k,\mathbf{A}_k+\lambda_{b_k}I)$ are the (normalized) "max" generalized eigenvectors of the two indicated matrices, with eigenvalues $\Sigma_k = \Sigma_{max}(\mathbf{B}_k,\mathbf{A}_k+\lambda_{b_k}I)$. Let $\Sigma_k^{(1)} = \overline{\boldsymbol{G}}_k^H\mathbf{B}_k\overline{\boldsymbol{G}}_k$ and $\Sigma_k^{(2)} = \overline{\boldsymbol{G}}_k^H\mathbf{A}_k\boldsymbol{G}_k$. The advantage of formulation (11) is that it allows straightforward power adaptation: introducing stream powers in the diagonal matrices $\boldsymbol{P}_k \geq 0$ and substituting $\boldsymbol{G}_k = \overline{\boldsymbol{G}}_k\,\boldsymbol{P}_k^{\frac{1}{2}}$ in (11) yields

$$\begin{aligned} &WSR(\boldsymbol{P},\lambda) = \textstyle\sum_j^C \lambda_j P_j + \\ &\sum_{k=1}^{K}[u_k\ln\det(I+\boldsymbol{P}_k\Sigma_k^{(1)}) - \mathrm{tr}\{\boldsymbol{P}_k(\Sigma_k^{(2)}+\lambda_{b_k}I)\}] \end{aligned} \tag{14}$$

[1]To keep notation light, we shall not denote $\boldsymbol{R}_i$, $\mathbf{A}_k$ as $\boldsymbol{R}_i^{'}$, $\mathbf{A}_k^{'}$ etc.

optimization of which leads to the following interference leakage aware water filling (WF) (jointly for the P_k and λ_c)

$$\boldsymbol{P}_k = \left(u_k(\Sigma_k^{(2)}+\lambda_{b_k}I)^{-1} - \Sigma_k^{-(1)}\right)^+,\ \sum_{k:b_k=c}\mathrm{tr}\{\boldsymbol{P}_k\} = P_c \tag{15}$$

where the Lagrange multipliers (to satisfy the power constraints) are computed by bisection and gets executed per BS. It is possible 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 $\boldsymbol{P}_k$ and the λ_c. Note that the minorization 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. The WF operation then decide how many streams can actually be sustained.

In contrast, in [15], for given λ, the $\boldsymbol{G}$ get iterated till convergence and the λ are found by duality (line search):

$$\begin{aligned} &\min_{\lambda\geq 0}\max_{\boldsymbol{G}}[\sum_j^C\lambda_j P_j + \sum_k\{u_k\ln\det(\boldsymbol{R}_{\overline{k}}^{-1}\boldsymbol{R}_k) - \lambda_{b_k}\mathrm{tr}\{\boldsymbol{P}_k\}\}] \\ &= \min_{\lambda\geq 0} WSR(\lambda). \end{aligned} \tag{16}$$

This typically leads to higher computational complexity for a given convergence precision.

IV. EWSR BF IN THE MaMIMO LIMIT

In this section, we consider the case when there is only partial channel state information. Here we consider the BF design based on the expected weighted sum rate, $\mathrm{E}_{\mathbf{H}|\widehat{\mathbf{H}}}WSR(\boldsymbol{G},\mathbf{H})$,

$$\mathrm{E}_{\mathbf{H}|\widehat{\mathbf{H}}}WSR(\boldsymbol{G},\mathbf{H}) = \mathrm{E}_{\mathbf{H}|\widehat{\mathbf{H}}}\sum_k u_k\ln(I+\boldsymbol{G}_k^H\mathbf{H}_{k,b_k}^H\boldsymbol{R}_{\overline{k}}^{-1}\mathbf{H}_{k,b_k}\boldsymbol{G}_k). \tag{17}$$

If the number of Tx antennas M becomes very large, we get a convergence for any quadratic term of the form,

$$\mathbf{H}\mathbf{Q}\mathbf{H}^H \xrightarrow{M\rightarrow\infty} \mathrm{E}_{\mathbf{H}|\widehat{\mathbf{H}}}\,\mathbf{H}\mathbf{Q}\mathbf{H}^H = \widehat{\mathbf{H}}Q\widehat{\mathbf{H}}^H + \mathrm{tr}\{\mathbf{Q}\mathbf{C}_p\}\,\mathbf{C}_r\,. \tag{18}$$

and hence we get the following MaMIMO limit matrices,

$$\begin{aligned} \widehat{\boldsymbol{R}}_k &= \mathbf{I}_{N_k} + \sum_{i=1}^{K}\left\{\widehat{\mathbf{H}}_{k,b_i}\mathbf{Q}_i\widehat{\mathbf{H}}_{k,b_i}^H + \mathrm{tr}\{\mathbf{Q}_i\mathbf{C}_{p,k,b_i}\}\,\mathbf{C}_{r,k}\right\} \\ \widehat{\boldsymbol{R}}_{\overline{k}} &= \mathbf{I}_{N_k} + \sum_{i=1,\neq k}^{K}\left\{\widehat{\mathbf{H}}_{k,b_i}\mathbf{Q}_i\widehat{\mathbf{H}}_{k,b_i}^H + \mathrm{tr}\{\mathbf{Q}_i\mathbf{C}_{p,k,b_i}\}\,\mathbf{C}_{r,k}\right\} \\ \widehat{\mathbf{B}}_k &= \widehat{\mathbf{H}}_{k,b_k}^H\bar{\mathbf{R}}_{\overline{k}}^{-1}\widehat{\mathbf{H}}_{k,b_k} + \mathrm{tr}\{\mathbf{C}_{r,k}\bar{\mathbf{R}}_{\overline{k}}^{-1}\}\mathbf{C}_{p,k,b_k} \\ \widehat{\mathbf{A}}_k &= \sum_{i\neq k}^{K} u_i\left[\widehat{\mathbf{H}}_{i,b_k}^H\left(\bar{\mathbf{R}}_{\overline{i}}^{-1} - \bar{\mathbf{R}}_i^{-1}\right)\widehat{\mathbf{H}}_{i,b_k}\right. \\ &\qquad \left. +\mathrm{tr}\{\left(\bar{\mathbf{R}}_{\overline{i}}^{-1} - \bar{\mathbf{R}}_i^{-1}\right)\mathbf{C}_{r,i}\}\mathbf{C}_{p,i,b_k}\right]\,. \end{aligned} \tag{19}$$

Here $\mathbf{C}_{p,k,b_i}$ corresponds to the error covariance matrix as in Section V, for the channel between user k and BS b_i. It suffices now to replace the matrices $\mathbf{A}_k$, $\mathbf{B}_k$ in the DC programming approach of Section III-B by the matrices $\widehat{\mathbf{A}}_k$, $\widehat{\mathbf{B}}_k$ above to get a maximum EWSR design. Thus the digital beamformers could be written as:

$$\overline{\boldsymbol{G}}_k = \mathbf{V}_{max}\left(\widehat{\mathbf{B}}_k, \left(\widehat{\mathbf{A}}_k + \lambda_{b_k}\mathbf{I}_{N_t^{b_k}}\right)\right). \quad (20)$$

V. MISO CASE : CHANNEL ESTIMATES

In this case $\boldsymbol{C}_r = 1$ and we shall denote the matrices $\boldsymbol{R}, \widehat{\mathbf{H}}$ as the scalar r and the vector $\mathbf{h}$. The channel $\mathbf{h}$, its estimate $\widehat{\mathbf{h}}$ and the estimation error $\widetilde{\mathbf{h}}$ have covariance matrices $\boldsymbol{\theta} = \boldsymbol{C}_{\mathbf{hh}}$, $\widehat{\boldsymbol{\theta}} = \boldsymbol{C}_{\widehat{\mathbf{h}}\widehat{\mathbf{h}}}$ and $\widetilde{\boldsymbol{\theta}} = \boldsymbol{C}_{\widetilde{\mathbf{h}}\widetilde{\mathbf{h}}}$, all three of which are arbitrary. Two special cases:
(i) In the case of a deterministic channel estimate (we abbreviate it as LS (Least Squares)), we have $\widehat{\mathbf{h}} = \mathbf{h} + \widetilde{\mathbf{h}}$ where $\mathbf{h}$ and $\widetilde{\mathbf{h}}$ are independent.
(ii) In the case of a LMMSE channel estimate, we have $\mathbf{h} = \widehat{\mathbf{h}} + \widetilde{\mathbf{h}}$ in which $\widehat{\mathbf{h}}$ and $\widetilde{\mathbf{h}}$ are decorrelated and hence independent in the Gaussian case. In the partial CSIT case, the term $\widehat{\mathbf{h}}\widehat{\mathbf{h}}^H$ of the perfect CSIT case gets replaced by its estimate $\mathbf{S} = \mathrm{E}_{\mathbf{h}|\widehat{\mathbf{h}}}\mathbf{h}\mathbf{h}^H = \widehat{\mathbf{h}}\widehat{\mathbf{h}}^H + \widetilde{\boldsymbol{\theta}}$, where
i) In the LS $\widehat{\mathbf{h}}$ case, $\widetilde{\boldsymbol{\theta}} = \boldsymbol{C}_{\widetilde{\mathbf{h}}\widetilde{\mathbf{h}}} = \sigma^2_{\widetilde{\mathbf{h}}}\boldsymbol{I}$. For the unbiased LS, $\widetilde{\boldsymbol{\theta}} = \boldsymbol{C}_{\widetilde{\mathbf{h}}\widetilde{\mathbf{h}}} = -\sigma^2_{\widetilde{\mathbf{h}}}\boldsymbol{I}$.
ii) In the LMMSE case, $\widetilde{\boldsymbol{\theta}} = C_{\widetilde{\mathbf{h}}\widetilde{\mathbf{h}}}$ is the posterior covariance, $\widetilde{\boldsymbol{\theta}} = \boldsymbol{\theta} - \boldsymbol{\theta}(\boldsymbol{\theta} + \sigma^2_{\widetilde{\mathbf{h}}}\boldsymbol{I})^{-1}\boldsymbol{\theta}$. Three types of channel estimates can be analyzed. In the case of partial CSIT we get for the Rx signal,

$$\begin{aligned} y_k = {} & \widehat{\mathbf{h}}_{k,b_k}\,\mathbf{g}_k\,x_k + \underbrace{\widetilde{\mathbf{h}}_{k,b_k}\,\mathbf{g}_k\,x_k}_{\text{sig. ch. error}} \\ & + \sum_{i=1,\neq k}^{K}(\widehat{\mathbf{h}}_{k,b_i}\,\mathbf{g}_i\,x_i + \underbrace{\widetilde{\mathbf{h}}_{k,b_i}\,\mathbf{g}_i\,x_i}_{\text{interf. ch. error}}) + v_k\,. \end{aligned} \quad (21)$$

Naive EWSR : just replace $\mathbf{h}$ by $\widehat{\mathbf{h}}$ in a perfect CSIT approach. Ignore $\widetilde{\mathbf{h}}$ everywhere. EWSMSE: accounts 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, EWSMSE also moves the signal $\widetilde{h}$ term to the interference plus noise!

A. Subspace Projection based Channel Estimator

In this section, we investigate the effect of reducing channel estimation error to the covariance subspace (without the LMMSE weighting, this is a simplification of the LMMSE estimate). Subspace channel estimate is given as,

$$\widehat{\mathbf{h}}_S = \boldsymbol{P}_A\widehat{\mathbf{h}}_{LS} = \mathbf{h} + \boldsymbol{P}_A\widetilde{\mathbf{h}}_{LS}\,,\; C_{\widetilde{\mathbf{h}}_S\widetilde{\mathbf{h}}_S} = \sigma^2_{\widetilde{\mathbf{h}}}\boldsymbol{P}_A, \quad (22)$$

where $\boldsymbol{P}_A = \mathbf{A}(\mathbf{A}^H\mathbf{A})^{-1}\mathbf{A}^H = \boldsymbol{P}_{\boldsymbol{C}_t}$, $\boldsymbol{C}_t = \mathbf{A}\mathbf{A}^H$, where $\mathbf{A}^H = \boldsymbol{D}\mathbf{H}_t^H$ and $\boldsymbol{D}$ and $\mathbf{H}_t^H$ are as defined in Section VIII. So, in WSR expressions $\mathbf{h}\mathbf{h}^H$ gets replaced by,
(i) naive Subspace Channel Estimator, $\mathbf{S} = \widehat{\mathbf{h}}_S\widehat{\mathbf{h}}_S^H$.
(ii) Subspace Channel Estimator in the MaMIMO limit, $\mathbf{S} = \widehat{\mathbf{h}}_S\widehat{\mathbf{h}}_S^H + \boldsymbol{C}_{\widetilde{h}_S\widetilde{h}_S}$.
(iii) unbiased Subspace Channel Estimator, $\mathbf{S} = \widehat{\mathbf{h}}_S\widehat{\mathbf{h}}_S^H - \boldsymbol{C}_{\widetilde{h}_S\widetilde{h}_S}$.

VI. VB-SBL

In this section we assume the system model as defined in section I. In Bayesian compressive sensing, a two-layer hierarchical prior is assumed for the $\boldsymbol{x}$ as in [7]. The parameters of the hierarchical prior are such that it supports the sparsity property of $\boldsymbol{x}$. $\boldsymbol{x}$ is assumed to have a Gaussian distribution parameterized by $\boldsymbol{\alpha} = [\alpha_1\ \alpha_2\ \ldots\ \alpha_M]$, where α_i represents the inverse variance of x_i.

$$p(\boldsymbol{x}/\boldsymbol{\alpha}) = \prod_{i=1}^{M} p(x_i/\alpha_i) = \prod_{i=1}^{M}\mathcal{N}(x_i/0, \alpha_i^{-1}). \quad (23)$$

Further a Gamma prior is considered over $\boldsymbol{\alpha}$,

$$p(\boldsymbol{\alpha}) = \prod_{i=1}^{M} p(\alpha_i/a,b) = \prod_{i=1}^{M}\Gamma^{-1}(a)b^a\alpha_i^{a-1}e^{-b\alpha_i}. \quad (24)$$

The inverse of noise variance γ is also assumed to have a Gamma prior,

$$p(\gamma) = \Gamma^{-1}(c)d^c\alpha_i^{c-1}e^{-d\gamma}. \quad (25)$$

A. Variational bayes

The computation of the posterior distribution of the parameters is usually intractable. In order to address this issue, in variational bayesian framework, the posterior distribution $p(\boldsymbol{x},\boldsymbol{\alpha},\gamma)$ is approximated as the product of the marginals:

$$p(\boldsymbol{x},\boldsymbol{\alpha},\gamma/\boldsymbol{y}) \approx q_\gamma(\gamma)\prod_{i=1}^{M}q_{x_i}(x_i)\prod_{i=1}^{M}q_{\alpha_i}(\alpha_i) = q(\boldsymbol{x},\boldsymbol{\alpha},\gamma/\boldsymbol{y}) \quad (26)$$

Variational bayes compute the factors q by minimizing the Kullback-Leibler distance between posterior distribution $p(\boldsymbol{x},\boldsymbol{\alpha},\gamma/\boldsymbol{y})$ and the $q(\boldsymbol{x},\boldsymbol{\alpha},\gamma/\boldsymbol{y})$. From [17],

$$KLD_{VB} = KL\left(p(\boldsymbol{x},\boldsymbol{\alpha},\gamma/\boldsymbol{y})||q(\boldsymbol{x},\boldsymbol{\alpha},\gamma/\boldsymbol{y})\right) \quad (27)$$

The KL divergence minimization is equivalent to maximizing the evidence lower bound (ELBO) [18],

$$p(\boldsymbol{y},\boldsymbol{\theta}) = p(\boldsymbol{y}/\boldsymbol{x},\boldsymbol{\alpha})p(\boldsymbol{x}/\boldsymbol{\alpha})p(\boldsymbol{\alpha})p(\gamma) \quad (28)$$

$$\ln(q_i(\theta_i)) = <\ln p(\boldsymbol{y},\boldsymbol{\theta})>_{k\neq i} + c_i, \quad (29)$$

where $\boldsymbol{\theta} = \{\boldsymbol{x},\boldsymbol{\alpha},\gamma\}$ and θ_i represents each scalar in $\boldsymbol{\theta}$. Here $<>_{k\neq i}$ represents the expectation operator over the distributions q_k for all $k \neq i$.

VII. SAVE - SPACE ALTERNATING VARIATIONAL ESTIMATION

In this section, we propose a Space Alternating Variational Estimation (SAVE) based alternating optimization between each elements of $\boldsymbol{\theta}$. We assume that the probability distribution of each x_i factorizes independently in q. The joint distribution can be written as,

$$\begin{aligned} & \ln p(\boldsymbol{y},\boldsymbol{\theta}) = \tfrac{N}{2}\ln\gamma - \tfrac{\gamma}{2}\,||\boldsymbol{y} - \mathbf{A}\boldsymbol{x}|\,|^2 + \\ & \sum_{i=1}^{N}\left(\ln\alpha_i - \frac{\boldsymbol{\alpha}_i}{2}\,|\boldsymbol{x}_i|^2\right) + \sum_{i=1}^{N}(\ln b - b\alpha_i) + \ln d - d\gamma \\ & q = \textstyle\prod_{i=1}^{M} q_{\boldsymbol{x}_i}(x_i)\prod_{i=1}^{M} q_{\alpha_i}(\alpha_i)\prod_{i=1}^{M} q_\gamma(\gamma)\,. \end{aligned} \quad (30)$$

Update of $q_{x_i}(x_i)$**:** $\ln q_{x_i}(x_i)$ is quadratic in x_i and thus can be represented as a Gaussian distribution as follows,

$$\begin{aligned}
&\ln q_{x_i}(x_i) = \\
&-\frac{\gamma}{2}\Big\{ <||\boldsymbol{y} - \mathbf{A}_{\bar{i}}\boldsymbol{x}_{\bar{i}}||^2> - (\boldsymbol{y} - \mathbf{A}_{\bar{i}} < x_{\bar{i}} >)^T \mathbf{A}_i x_i - \\
&x_i \mathbf{A}_i^T(\boldsymbol{y} - \mathbf{A}_{\bar{i}} < \boldsymbol{x}_{\bar{i}} >) + ||A_i||^2 x_i^2 \Big\} - \frac{<\alpha_i>}{2} x_i^2 + c_{x_i} \\
&= -\frac{1}{2\sigma_i^2}\left(x_i - \mu_i\right)^2 + c'_{x_i}
\end{aligned} \tag{31}$$

Note that we split $\mathbf{A}\boldsymbol{x}$ as, $\mathbf{A}\boldsymbol{x} = \mathbf{A}_i x_i + \mathbf{A}_{\bar{i}}\boldsymbol{x}_{\bar{i}}$, where $\mathbf{A}_i$ represents the i^{th} column of $\mathbf{A}$, $\mathbf{A}_{\bar{i}}$ represents the matrix with i^{th} column of $\mathbf{A}$ removed, x_i is the i^{th} element of $\boldsymbol{x}$, and $\boldsymbol{x}_{\bar{i}}$ is the vector without x_i. Clearly, the mean of the variance of the resulting Gaussian distribution becomes,

$$\sigma_i^2 = \frac{1}{\gamma||\mathbf{A}_i||^2 + \alpha_i}, \;\; \mu_i = \sigma_i^2 A_i^T\left(\boldsymbol{y} - \mathbf{A}_{\bar{i}} < \boldsymbol{x}_{\bar{i}} >\right) < \gamma > \tag{32}$$

Update of $q_{\alpha_i}(\alpha_i)$**:** The variational approximation leads to the following Gamma distribution for the $q_{\alpha_i}(\alpha_i)$,

$$\begin{aligned}
&\ln q_{\alpha_i}(\alpha_i) = \frac{1}{2}\ln\alpha_i - \alpha_i\left(\frac{<x_i^2>}{2} + b\right) + c_{\alpha_i}, \\
&q_{\alpha_i}(\alpha_i) \propto \alpha_i^{1/2} e^{-\alpha_i\left(\frac{<x_i^2>}{2} + b\right)}.
\end{aligned} \tag{33}$$

The mean of the Gamma distribution is given by,

$$\begin{aligned}
&<\alpha_i> = \frac{3/2}{\left(\frac{<x_i^2>}{2} + b\right)}, \\
&\text{where} \;\; <x_i^2> = \mu_i^2 + \sigma_i^2.
\end{aligned} \tag{34}$$

Update of $q_\gamma(\gamma)$**:** Similarly, the Gamma distribution from the variational bayesian approximation for the $q_\gamma(\gamma)$ can be written as,

$$\begin{aligned}
&\ln q_\gamma(\gamma) = \frac{N}{2}\ln\gamma - \gamma\left(\frac{<||\boldsymbol{y} - \mathbf{A}\boldsymbol{x}||^2>}{2} + d\right) + c_\gamma, \\
&q_\gamma(\gamma) \propto \gamma^{N/2} e^{-\gamma\left(\frac{<||\boldsymbol{y} - A\boldsymbol{x}||^2>}{2} + d\right)}.
\end{aligned} \tag{35}$$

The mean of the gamma distribution for γ is given by,

$$\begin{aligned}
&<\gamma> = \frac{N/2 + 1}{\left(\frac{<||\boldsymbol{y} - \mathbf{A}\boldsymbol{x}||^2>}{2} + d\right)}, \\
&\text{where,} \\
&<||\boldsymbol{y} - \mathbf{A}\boldsymbol{x}||^2> = \\
&||\boldsymbol{y}||^2 - 2\boldsymbol{y}^T\mathbf{A}\boldsymbol{\mu} + \text{tr}\mathbf{A}^T\mathbf{A}\left(\boldsymbol{\mu}\boldsymbol{\mu}^T + \boldsymbol{\Sigma}\right), \\
&\boldsymbol{\Sigma} = diag(\sigma_1^2, \sigma_2^2, ..., \sigma_M^2) \\
&\boldsymbol{\mu} = diag(\mu_1, \mu^2, ..., \mu_M).
\end{aligned} \tag{36}$$

A. Reduced Complexity Linear Tx/Rx Computation

An optimal linear Tx/Rx filter in MU MIMO is of the form

$$\boldsymbol{F} = (\mathbf{A}\,\boldsymbol{D}_1\,\mathbf{A}^H + \lambda\,\boldsymbol{I})^{-1}\mathbf{A}\,\boldsymbol{D}_2. \tag{37}$$

Other sub-optimal beamformers are special cases of this, where, for the R-ZF, $\boldsymbol{D}_1 = \boldsymbol{I}$ and ZF $\lambda \to 0$. LMMSE Tx/Rx can also be found by SAVE. Consider the case of a multi-user UL system, with $\mathbf{A} \in \mathcal{C}^{N \times M}$ SIMO channel, x as the $M \times 1$ transmit signal from all users and $y \in C^{N \times 1}$ is the received signal at the BS. From (32), it can be seen that the estimate of $\boldsymbol{x} = \mu$ converges to the L-MMSE equalizer,

$$\begin{aligned}
&\sigma_v^2\mu_i + \sigma_i^2\mathbf{A}_i^T\mathbf{A}\mu = \sigma_i^2\mathbf{A}_i^T\boldsymbol{y} \;\Rightarrow \\
&\widehat{\boldsymbol{x}} = \mu = (\mathbf{A}^T\mathbf{A} + \sigma_v^2\boldsymbol{\Sigma}^{-1})^{-1}\mathbf{A}^T\boldsymbol{y}.
\end{aligned} \tag{38}$$

SAVE recursions are similar to PE [4]. However, PE only converges in case of sufficient diagonal dominance of $\mathbf{A}^T\mathbf{A}$, whereas SAVE is guaranteed to converge, employing implicitly varying damping factors (the σ_i^2).

VIII. Multipath Channel Model

We get for the matrix impulse response of a time-varying frequency-selective MIMO channel $\mathbf{H}(t,\tau)$ [19],

$$\mathbf{H}(t,\tau) = \sum_{i=1}^{L} A_i(t)\, e^{j2\pi\, f_i\, t}\,\mathbf{h}_r(\phi_i)\,\mathbf{h}_t^T(\psi_i)\,p(\tau - \tau_i) \;\; . \tag{39}$$

with L (specular) pathwise contributions where

- A_i: complex attenuation
- f_i: Doppler shift
- ψ_i: direction of departure (AoD) (azimuth, elevation, polar)
- ϕ_i: direction of arrival (AoA) (azimuth, elevation, polarization)
- τ_i: path delay (ToA)
- $\mathbf{h}_t(.)$, $\mathbf{h}_r(.)$: $M/N \times 1$ Tx/Rx antenna array response
- $p(.)$: pulse shape (Tx filter)

In the case of distributed antenna systems (near field), or very wideband regime, the array responses become a function of the position parameters of the (last) path scatterers. The fast variation of the phase in $e^{j2\pi\, f_i\, t}$ 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 (including the Doppler frequency) vary on a slower time scale and correspond to slow fading.

The channel impulse response $\mathbf{H}$ has per path a rank one contribution in four dimensions (Tx and Rx spatial multi-antenna dimensions, delay spread and Doppler spread). Hence, going to the frequency domain, we get

$$\text{vec}(\mathbf{H}(1:t, f_1:f_2)) = \sum_{i=1}^{L} A_i\,\mathbf{h}_t(\psi_i) \otimes \mathbf{h}_r(\phi_i) \otimes \boldsymbol{v}_f(\tau_i) \otimes \boldsymbol{v}_t(f_i). \tag{40}$$

where $\boldsymbol{v}_f(.)$, $\boldsymbol{v}_t(.)$ are appropriate Vandermonde vectors (possibly subsampled in the case of $\boldsymbol{v}_f(.)$). Hence we get a sum of rank one $4D$ tensors. $\mathbf{h}_r$, $\mathbf{h}_t$ could themselves have a Kronecker structure in the case of polarization or in the case of $2D$ antenna arrays with separable structure [20]. In the model above, each of the four Kronecker factors is assumed to be parametric. For instance, $\mathbf{h}_t(.)$ is also a Vandermonde vector in the case of a basic Uniform Linear Array depending on azimuth only, neglecting antenna coupling. Whereas more generally $\mathbf{h}_t(.)$ may be known or learned at the BS side, it is less reasonable to assume a parametric form for $\mathbf{h}_r$ on the UE side, especially in the case of a hand-held device (orientation, way of holding it).

In OFDM, we can factor the channel response at subcarrier n as

$$\mathbf{H}[n] = \sum_{i=1}^{L} |A_i|\, e^{j\xi_i[n]} \mathbf{h}_r(\phi_i)\mathbf{h}_t^T(\psi_i) = \mathbf{H}_r \, \mathbf{\Xi}[n] \, \boldsymbol{D} \, \mathbf{H}_t^H \,,$$

$$\mathbf{H}_r = [\mathbf{h}_r(\phi_1)\, \mathbf{h}_r(\phi_1) \cdots] \,, \; \mathbf{\Xi}[n] = \begin{bmatrix} e^{j\xi_1[n]} & & \\ & e^{j\xi_2[n]} & \\ & & \ddots \end{bmatrix} ,$$

$$\boldsymbol{D} = \begin{bmatrix} |A_1| & & \\ & |A_2| & \\ & & \ddots \end{bmatrix} , \; \mathbf{H}_t^H = \begin{bmatrix} \mathbf{h}_t^T(\psi_1) \\ \mathbf{h}_t^T(\psi_2) \\ \vdots \end{bmatrix} \tag{41}$$

Tx side covariance matrix $\boldsymbol{C}^t$, which only explores the channel correlations as they can be seen from the BS side,

$$\boldsymbol{C}^t = \mathrm{E}\, \mathbf{H}^H \mathbf{H} = \mathbf{H}_t \, \boldsymbol{D}^2 \, \mathbf{H}_t^H . \tag{42}$$

by averaging of the random path phases ξ_i.

A. Non-Separable (Non-Kronecker) Covariance CSIT

Averaging over the (uniform) path phases ψ_i leads to

$$\boldsymbol{C}_{\mathbf{hh}} = \sum_{i=1}^{L} |A_i|^2 \, \mathbf{h}_i \mathbf{h}_i^H = \sum_{i=1}^{L} A_i^2 \, (\mathbf{h}_r(\phi_i)\mathbf{h}_r^H(\phi_i)) \otimes (\mathbf{h}_t(\psi_i)\mathbf{h}_t^H(\psi_i)). \tag{43}$$

where $\boldsymbol{C}_{\mathbf{hh}} = \mathrm{E}\, \mathbf{h}\mathbf{h}^H$, $\mathbf{h} = \mathrm{vec}(\mathbf{H})$ and $\mathbf{h}_i = \mathbf{h}_t(\psi_i) \otimes \mathbf{h}_r(\phi_i)$. NADA (Narrow AoD Aperture): the rank of $\boldsymbol{C}_{\mathbf{hh}}$ can be substantially less than the number of paths, consider e.g. a cluster of paths with narrow AoD spread, then we have

$$\psi_i = \psi + \Delta\psi_i \tag{44}$$

where ψ is the nominal AoD and $\Delta\psi_i$ is small. Hence a first order taylor series approximation leads to,

$$\mathbf{h}_t(\psi_i) \approx \mathbf{h}_t(\psi) + \Delta\psi_i \, \dot{\mathbf{h}}_t(\psi) \,. \tag{45}$$

Such a cluster of paths only adds a rank 2 contribution to $\boldsymbol{C}_{\mathbf{hh}}$. Covariance of $\mathbf{h} = \mathrm{vec}(\mathbf{H})$ is not of Kronecker form.

B. Channel estimation using VB

In this section, we discuss the VB based technique to approximate the LMMSE channel estimate using the LS channel estimate as the observation. Some of the existing works on multi-path parameter estimation can be found in [19]–[22]. [21] focuses on joint channel estimation and data detection in an OFDM system using SBL. The authors in [22] discuss about tracking channel subspaces (motivated by Joint Spatial Division and Multiplexing). This requires to determine dominant subspace dimension, which may vary with SNR. [20] proposes tensor decomposition algorithms to estimate the channel parameters.

We can model the path amplitudes A_i as gaussian. Our aim is to compute the posteriors for path parameters, allowing to express covariance CSIT uncertainty. Hence can use VB-SBL or SAVE with Gaussian posterior approximation. VB (or any sparse technique) for automatic model order detection. For path clusters (NADA), requires to handle multiple atoms jointly.

IX. Simulation Results

In this section, we present the Ergodic Sum Rate Evaluations for BF design for the various channel estimates. Monte Carlo evaluations of ergodic sum rates are done with the following parameters: C, number of cells. K, number of (single-antenna) users per cell. Nt, number of transmit Antennas. We consider a path-wise channel model as in section VIII, with L number of paths/Channel Rank. c, scale factor for LS channel estimation error variance $\sigma_{\tilde{h}}^2 = c/SNR$. In the figures, iCSIT refers to the BF design for the instantaneous CSIT case.

A. Single Cell Simulations

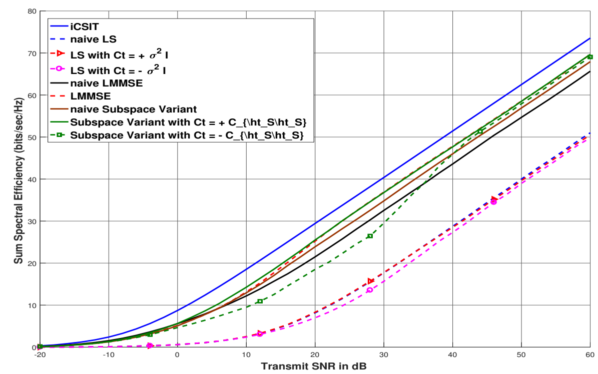

Fig. 1. Expected sum rate for $C = 1$ cell, $K = 5$ users/cell, $Nt = 32$, $L = 2$ paths in each channel, $\sigma_{\tilde{h}}^2 = 1/SNR$.

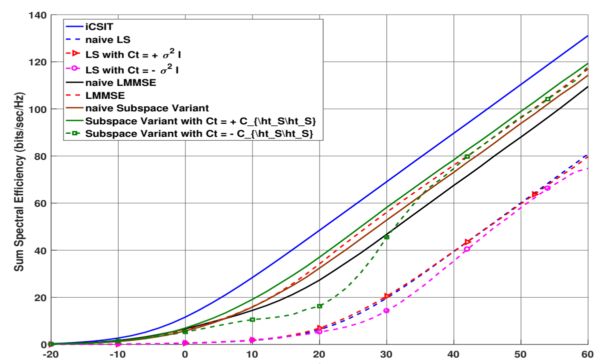

Fig. 2. Expected sum rate for $C = 1$ cell, $K = 10$ users/cell, $Nt = 64$, $L = 2$ paths in each channel, $\sigma_{\tilde{h}}^2 = 1/SNR$.

B. Multi-Cell Simulations

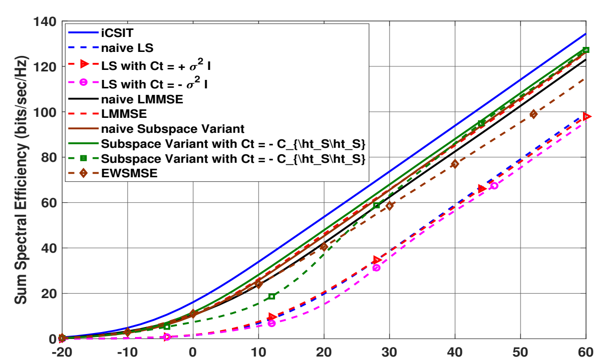

Fig. 3. Expected sum rate for $C = 2$ cells, $K = 5$ users/cell, $Nt = 32$, $L = 2$ paths in each channel, $\sigma_{\tilde{h}}^2 = 1/SNR$.

In Figure 3, we plot the EWSMSE beamforming performance also and it is evident from the figure that EWSR based beamformers such as LMMSE/Subspace projection techniques perform better than EWSMSE design.

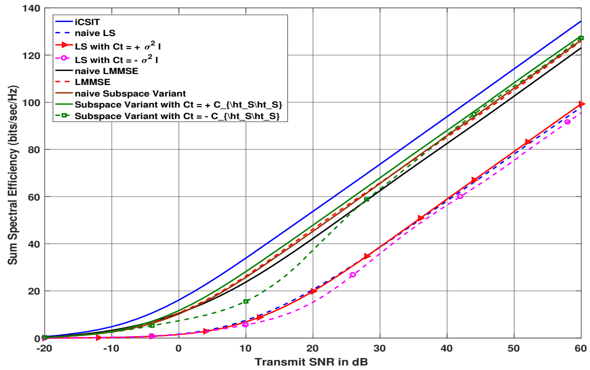

Fig. 4. Expected sum rate for $C = 2$ cells, $K = 5$ users/cell, $Nt = 32$, $L = 2$ paths in each channel, $\sigma^2_{\tilde{h}} = 1/SNR$.

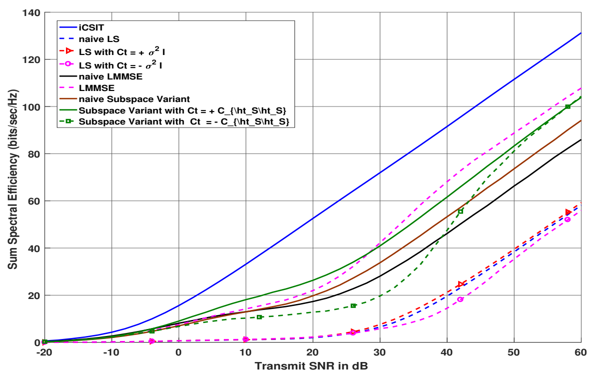

Fig. 5. Expected sum rate for $C = 2$ cells, $K = 5$ users/cell, $Nt = 32$, $L = 2$ paths in each channel, $\sigma^2_{\tilde{h}} = 10/SNR$.

From the numerical simulations, it is quite evident that just using LS channel estimates may lead to substantial EWSR loss. In Massive MIMO, the exploitation of channel subspaces (reduced rank covariances) in channel estimates may lead to substantial reductions in the SNR loss. Moreover, there is significant gain from exploiting (error) channel covariances in addition to channel estimates and proper handling of channel error covariance in the direct link in the BF design.

X. Conclusion

In Massive MIMO, we suggest the application of Compressive Sensing inspired Bayesian learning techniques to reduced complexity LMMSE-style Tx/Rx signal computation. We also apply the SBL technique called SAVE to structured channel estimation, providing both instantaneous and covariance (or even beyond) CSIT. So far implicitly assumed a TDD case with channel estimation on the uplink for Tx computation on the downlink. The Rx computation in the TDD UL can further benefit from joint Rx determination and channel estimation. In FDD, the channel estimation in the UL still allows to use the slow channel parameters for DL covariance CSIT.

Acknowledgments

EURECOM's research is partially supported by its industrial members: ORANGE, BMW, SFR, ST Microelectronics, Symantec, SAP, Monaco Telecom, iABG, and by the projects HIGHTS (EU H2020) and MASS-START (French FUI). The research of Orange Labs is partially supported by the EU H2020 project One5G.

References

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

[2] 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.

[3] Wassim Tabikh, Dirk Slock, and Yi Yuan-Wu, "MIMO IBC beamforming with combined channel estimate and covariance CSIT," in *IEEE International Symposium on Information Theory*, 2017.

[4] A. Kammoun, A. Muller, E. Bjornson, and M. Debbah, "Linear precoding based on polynomial expansion: Large-scale multi-cell MIMO systems," *IEEE Journal of Selected Topics in Signal Processing*, vol. 8, no. 5, pp. 861–875, 2014.

[5] Ramina Ghods, Charles Jeon, Gulnar Mirza, Arian Maleki, and Christoph Studer, "Optimally-Tuned Nonparametric Linear Equalization for Massive MU-MIMO Systems," in *IEEE International Symposium on Information Theory*, 2017.

[6] Mahdi Nouri Boroujerdi, Saeid Haghighatshoar, , and Giuseppe Caire, "Low-Complexity Statistically Robust Precoder/Detector Computation for Massive MIMO Systems," in *arXiv preprint arXiv:1711.11405v2*, December 2017.

[7] Michael E. Tipping, "Sparse bayesian learning and the relevance vector machine," *Journal of Machine Learning Research*, vol. 1, pp. 211–244, June 2001.

[8] Michael E. Tipping and Anita C. Faul, "Fast marginal likelihood maximisation for sparse Bayesian models," in *AISTATS*, January 2003.

[9] Dmitriy Shutin, Thomas Buchgraber, Sanjeev R. Kulkarni, and H. Vincent Poor, "Fast variational sparse bayesian learningwith automatic relevance determination for superimposed signals," *IEEE Transactions on Signal Processing*, vol. 59, no. 12, December 2011.

[10] Maher Al-Shoukairi, Philip Schniter, and Bhaskar D. Rao, "Gamp-based low complexity sparse bayesian learning algorithm," *IEEE Transaction on Signal Processing*, vol. 66, no. 2, January 2018.

[11] Huiping Duan, Linxiao Yang, and Hongbin Li, "Fast inverse-free sparse bayesian learning via relaxed evidence lower bound maximization," *IEEE Signal Processing Letters*, vol. 24, no. 6, June 2017.

[12] 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.

[13] 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.

[14] 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.

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

[16] P. Stoica and Y. Selén, "Cyclic Minimizers, Majorization Techniques, and the Expectation-Maximization Algorithm: A Refresher," *IEEE Sig. Proc. Mag.*, Jan. 2004.

[17] Matthew J. Beal, "Variational algorithms for approximate bayesian inference," in *Thesis, Univeristy of Cambridge, UK*, May 2003.

[18] D. G. Tzikas, A. C. Likas, and N. P. Galatsanos, "The variational approximation for Bayesian inference," *IEEE Signal Process. Mag.*, vol. 29, no. 6, pp. 131–146, November 2008.

[19] Dmitriy Shutin and Bernard H. Fleury, "Sparse Variational Bayesian SAGE Algorithm With Application to the Estimation of Multipath Wireless Channels," *IEEE Transactions on signal processing*, vol. 59, no. 8, pp. 3609–3623, August 2011.

[20] Cheng Qian, Xiao Fu, Nicholas D. Sidiropoulos, and Ye Yang, "Tensor based parameter estimation of double directional massive MIMO channel with dual- polarized antennas," in *Proc. IEEE Int'l Conf. Acoustics Speech and Sig. Proc. (ICASSP)*, 2018.

[21] Ranjitha Prasad, Chandra R. Murthy, and Bhaskar D. Rao, "Joint Channel Estimation and Data Detection in MIMO-OFDM Systems: A Sparse Bayesian Learning Approach," *IEEE Transactions on signal processing*, vol. 63, no. 20, pp. 5369 – 5382, October 2015.

[22] Saeid Haghighatshoar and Giuseppe Caire, "Low-complexity massive MIMO subspace tracking from low-dimensional projections," in *Proc. IEEE Int'l Conf. on Communications (ICC)*, 2017.