A Massive MIMO Stochastic Geometry Analysis of Various Beamforming Designs with Partial CSIT

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

***Abstract*—We consider coordinated beamforming (BF) for the Multi-Input Single-Output (MISO) Interfering Broadcast Channel (IBC). The beamformers are optimized for the Ergodic Weighted Sum Rate (EWSR) or various approximations and bounds thereof, for the case of Partial Channel State Information at the Transmitters (CSIT). Gaussian (posterior) partial CSIT can optimally combine channel estimate and channel covariance information. With Gaussian partial CSIT, the beamformers only depend on the means (estimates) and (error) covariances of the channels. We extend a recently introduced large system analysis for optimized beamformers with partial CSIT, by a stochastic geometry inspired randomization of the channel covariance eigen spaces, leading to much simpler analytical results, which depend only on some essential channel characteristics. In the Massive MISO (MaMISO) limit, we obtain deterministic approximations of the signal and interference plus noise powers at the receivers for various BFs, which are tight as the number of BS antennas and the total user subspace dimension tend to infinity at fixed ratio. Simulation results exhibit the correctness of the large system results and the performance superiority of optimal BF designs based on both the MaMISO limit of the EWSR and using Linear Minimum Mean Squared Error (LMMSE) channel estimates.**

***Index terms*—** Massive MIMO, stochastic geometry, partial CSIT, ergodic weighted sum rate, optimal beamforming

I. Introduction

In this paper, Tx may denote transmit/transmitter /transmission and Rx may denote receive/receiver/reception. The recent development of Massive MIMO (MaMIMO) [1] opens new possibilities for increased system capacity while at the same time simplifying system design. 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). One of the pioneering work which talks about the effect of imperfect channel knowledge on the capacity is [2].

Indeed, in Massive MISO (MaMISO) systems, the received signal and interference powers converge to their expected value (channel hardening effect) due to the law of large numbers. We refer the readers to a more detailed discussion on the state of the art on large system analysis to [3]–[5] and instead focus on the stochastic geometric aspects in this section.

An introduction to the literature on stochastic geometry can be found in [6]. In stochastic geometry, generally the location of the nodes in the wireless network is modeled as random, following for example a poisson point process. In stochastic geometry based methods [7], [8], the location of the users being random, their geographic distribution then induces a certain probability distribution for the channel attenuations. This leads to results on the coverage probability, the capacity, the outage probability and other fundamental limits in wireless networks. Whereas most stochastic geometry work focuses on the distribution of the attenuations, here we consider an extension to multi-antenna systems. The multipath propagation for the various users leads to randomized angles of arrival at the BS which can be translated into spatial channel response contributions that depend on the antenna array response. In the massive MIMO regime in which the number of BS antennas gets very large, it has been observed and exploited that despite complex multipath propagation, the channel covariance matrix tends to be low rank. Exploiting the randomized nature of the user and scatterer positions and making abstraction of the antenna array response, we proposed to model the user channel subspaces as isotropically randomly oriented. This allows us to assume the eigen vectors of the channel covariance matrix to be Haar distributed, and this identically and independently for all users.

A. Contributions of this paper

- We first consider the various channel estimates such as linear minimum mean square error (LMMSE), least sqaures (LS) and subspace projected. Further we review the optimal BF design for the expected weighted sum rate (EWSR) criterion in the MaMISO limit.
- We evaluate the ergodic sum rate performance for LS, LMMSE and subspace projection channel estimators with optimal EWSR BF. Numerical results suggest that there is substantial gain by exploiting the channel covariance information compared to just using the LS estimates.
- Compared to our previous work [3], we derive simplified sum rate expressions at high SNR for the various BFs (optimal EWSR, naive and EWSMSE) for the various channel estimates, which clearly shows the SNR offset for the sub-optimal BFs compared to the proposed optimal EWSR BF.

Notation: In the following, boldface lower-case and upper-case characters denote vectors and matrices respectively. The operators $\mathrm{E}(\cdot)$, $tr(\cdot)$, $(\cdot)^H$, $(\cdot)^T$ represents expectation, trace , conjugate transpose and transpose respectively. diag$(\cdot)$ represents the diagonal matrix formed by the elements $(\cdot)$. A circularly complex Gaussian random vector with mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Theta}$ is distributed as $\boldsymbol{x} \sim \mathcal{CN}(\boldsymbol{\mu}, \boldsymbol{\Theta})$. $\mathbf{V}_{max}(\mathbf{A}, \mathbf{B})$ or $\mathbf{V}_{max}(\mathbf{A})$ represents (normalized) dominant generalized eigen vector of $\mathbf{A}$ and $\mathbf{B}$ or (normalized) dominant eigen vector of $\mathbf{A}$ respectively and $\lambda_{max}(\mathbf{A})$ being the max eigen value.

II. MISO IBC Signal Model

We consider here an IBC with C cells with a total of K single antenna users. We shall consider a system-wide numbering of the users. User k is served by BS b_k. The received signal at user k in cell b_k is

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

where x_k is the intended (white, unit variance) scalar signal stream, $\mathbf{h}_{k,b_k}$ is the $M_{b_k} \times 1$ channel from BS b_k to user k. The Rx signal (and hence the channel) is assumed to be scaled so that we get for the noise $v_k \sim \mathcal{CN}(0,1)$. BS c serves $K_c = \sum_{i:b_i=c} 1$ users. The $M_{b_k} \times 1$ spatial Tx filter or beamformer (BF) is $\mathbf{g}_k$.

A. Channel and CSIT Model

For simplicity, we omit all the user indices k. Each MISO channel is modeled according to a correlation structure [9] as

$$\mathbf{h} = \mathbf{C}\mathbf{D}^{1/2}\mathbf{c}\,,\ R_{\mathbf{hh}} = \mathbf{C}\mathbf{D}\mathbf{C}^H \quad (2)$$

where $\mathbf{c} \sim \mathcal{CN}(0, \mathbf{I}_L)$ are the Rayleigh fading multipath gains in the eigen domain. Here $\mathbf{C}$ is the $M \times L$ eigenvector matrix of the reduced rank channel covariance $R_{\mathbf{hh}}$ with diagonal eigenvalue matrix $\mathbf{D}$. This reduced rank covariance matrix of user channels typically occurs in realistic MaMISO channels due to the limited angular spread of the multipath components [10]. The rank corresponds to an equivalent number of linearly independent multipath components. The total sum rank across all users $L_t = \sum_{k=1}^{K} L_{k,c}$ is assumed to be less than M_c, where $L_{k,c}$ is the channel rank between user k and BS c.

For estimation, we start from a deterministic Least-Squares (LS) channel estimate

$$\widehat{\mathbf{h}}_{LS} = \mathbf{h} + \widetilde{\mathbf{h}}, \quad (3)$$

where $\mathbf{h}$ is the true MISO channel, and the error is modeled as circularly symmetric white Gaussian noise $\widetilde{\mathbf{h}} \sim \mathcal{CN}(0, \widetilde{\sigma}^2\mathbf{I})$. Now, assuming the channel covariance subspace is known, the LMMSE channel estimate can be obtained as $\widehat{\mathbf{h}} = \mathbf{C}\mathbf{D}\mathbf{C}^H \left(\mathbf{C}\mathbf{D}\mathbf{C}^H + \widetilde{\sigma}^2\mathbf{I}\right)^{-1} \widehat{\mathbf{h}}_{LS}$. Applying the matrix inversion lemma and exploiting $\mathbf{C}^H\mathbf{C} = \mathbf{I}_L$, this simplifies to

$$\widehat{\mathbf{h}} = \mathbf{C}\left(\widetilde{\sigma}^2\mathbf{D}^{-1} + \mathbf{I}\right)^{-1}\mathbf{C}^H\widehat{\mathbf{h}}_{LS} = \mathbf{C}\widehat{\mathbf{D}}^{1/2}\widehat{\mathbf{c}}, \quad (4)$$

where $\widehat{\mathbf{D}} = \left(\widetilde{\sigma}^2\mathbf{D}^{-1} + \mathbf{I}\right)^{-1}\mathbf{D}$ and $\widehat{\mathbf{c}} = \mathbf{D}^{-1/2}(\sigma^2\mathbf{D}^{-1} + \mathbf{I})^{-1/2}\mathbf{C}^H\widehat{\mathbf{h}}_{LS}$ with $R_{\widehat{\mathbf{c}}\widehat{\mathbf{c}}} = \mathbf{I}$.

$$R_{\widetilde{\mathbf{h}}\widetilde{\mathbf{h}}} = \mathbf{C}\widetilde{\mathbf{D}}\mathbf{C}^H = \mathbf{C}\left[\mathbf{D} - \left(\widetilde{\sigma}^2\mathbf{D}^{-1} + \mathbf{I}\right)^{-1}\mathbf{D}\right]\mathbf{C}^H. \quad (5)$$

So we can write for $\mathbf{S} = \mathrm{E}_{\mathbf{h}|\widehat{\mathbf{h}}}\left(\mathbf{h}\mathbf{h}^H\right) = \widehat{\mathbf{h}}\,\widehat{\mathbf{h}}^H + R_{\widetilde{\mathbf{h}}\widetilde{\mathbf{h}}} = \mathbf{C}\mathbf{W}_L\mathbf{C}^H$, where $\mathbf{W}_L = \widehat{\mathbf{D}}^{1/2}\widehat{\mathbf{c}}\widehat{\mathbf{c}}^H\widehat{\mathbf{D}}^{1/2} + \widetilde{\mathbf{D}}$.

III. Partial CSIT BF based on Different Channel Estimates

In the MaMIMO limit, (the EWSR upper bound based) BF design with partial CSIT will depend on the quantities $\mathbf{S} = \mathrm{E}_{\mathbf{h}|\widehat{\mathbf{h}}}\left(\mathbf{h}\mathbf{h}^H\right) = \widehat{\mathbf{h}}\,\widehat{\mathbf{h}}^H + \widetilde{\mathbf{\Theta}}$. We shall consider three possible channel estimates.

(i) *LS Channel Estimate*
We have $\widetilde{\mathbf{\Theta}} = \widetilde{\sigma}^2\mathbf{I}$, $\widehat{\mathbf{h}}_{LS} = \mathbf{h} + \widetilde{\mathbf{h}}$ where $\mathbf{h}$ and $\widetilde{\mathbf{h}}$ are independent.

(ii) *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. $\widetilde{\mathbf{\Theta}} = R_{\widetilde{\mathbf{h}}\widetilde{\mathbf{h}}}$ is the posterior covariance. The resulting $\mathbf{S} = \mathrm{E}_{\mathbf{h}|\widehat{\mathbf{h}}}\mathbf{h}\mathbf{h}^H = \widehat{\mathbf{h}}\widehat{\mathbf{h}}^H + \widetilde{\mathbf{\Theta}}$ is the (nonlinear) MMSE estimate of $\mathbf{h}\mathbf{h}^H$ (nonlinear because quadratic in $\widehat{\mathbf{h}} = \widehat{\mathbf{h}}_{LS}$ plus a constant). It is unbiased : $E_{\widehat{\mathbf{h}}}\mathbf{S} = E_{\widehat{\mathbf{h}}}E_{\mathbf{h}|\widehat{\mathbf{h}}}\mathbf{h}\mathbf{h}^H = \mathrm{E}_{\mathbf{h}}\mathbf{h}\mathbf{h}^H$ and it is MMSE, hence minimum variance since unbiased. In particular, it also minimizes the variance of $|\mathbf{g}^H\mathbf{h}|^2 = \mathbf{g}^H\mathbf{h}\mathbf{h}^H\mathbf{g} = \mathbf{g}^T \otimes \mathbf{g}^H vec(\mathbf{h}\mathbf{h}^H)$ where $vec(\mathbf{h}\mathbf{h}^H) = \mathbf{h}^* \otimes \mathbf{h}$.

(iii) *Subspace Projection based Channel Estimate*
We also investigate the effect of limiting channel estimation error to the covariance subspace (LMMSE without weighting). The subspace channel estimate is given as,

$$\widehat{\mathbf{h}}_S = \boldsymbol{P}_C\widehat{\mathbf{h}}_{LS} = \mathbf{h} + \boldsymbol{P}_C\widetilde{\mathbf{h}}_{LS}\,,\ R_{\widetilde{\mathbf{h}}_S\widetilde{\mathbf{h}}_S} = \widetilde{\sigma}^2\boldsymbol{P}_C, \quad (6)$$

where $\boldsymbol{P}_C = \mathbf{C}\mathbf{C}^H$ represents the projection onto the covariance subspace. Further we can write the estimate for $\mathbf{h}\mathbf{h}^H$, $\mathbf{S} = \widehat{\mathbf{h}}_S\widehat{\mathbf{h}}_S^H + R_{\widetilde{\mathbf{h}}_S\widetilde{\mathbf{h}}_S} = \mathbf{C}\mathbf{W}_S\mathbf{C}^H$ with $\mathbf{W}_S = \widehat{\mathbf{c}}\widehat{\mathbf{c}}^H + \widetilde{\sigma}^2\mathbf{I}$. One remark here is that subspace channel estimator represents a simplification of the LMMSE channel estimator, since it doesn't require the knowledge of the eigen value matrix $\mathbf{D}$ and without negligible performance loss compared to the LMMSE estimator as is validated in our numerical simulations. Another point to be noted is that, combining subspace channel estimator and LMMSE estimator, we can write $\widehat{\mathbf{h}} = \mathbf{C}\boldsymbol{U}\mathbf{C}^H\widehat{\mathbf{h}}_{LS}$, where for LMMSE $\boldsymbol{U}_L = (\mathbf{I} + \widetilde{\sigma}^2\mathbf{D}^{-1})^{-1}$ and for subspace $\boldsymbol{U}_S = \mathbf{I}$. This observation also hints at the possibility of optimizing $\boldsymbol{U}$ (LMMSE is not necessarily the best) to maximize the ergodic capacity, but this is left for future work.

A. Max EWSR BF (ESIP-WSR Upper Bound)

In the following $\mathbf{h}_{k,b_i}, \widehat{\mathbf{h}}_{k,b_i}, \widetilde{\mathbf{h}}_{k,b_i}$ denote the actual channel, channel estimate and estimation error resp. between user k and BS b_i. Once the CSIT is imperfect, various optimization criteria such as outage capacity can be considered. Motivated by the ergodic capacity formulations in [11] for point to point MIMO systems, and in [12] for multi-user MISO systems, the design here is based on expected weighted sum rate (EWSR) (and normally with LMMSE channel estimates). In a first stage, the WSR is averaged over the channels given the channel estimates and covariance information (i.e. the partial CSIT), leading to a cost function that can be optimized by the Tx. The optimized result then needs to be averaged over the channel estimates to obtain the final ergodic WSR. From the law of total expectation

$$EWSR = \mathrm{E}_{\widehat{\mathbf{h}}} \max_{\mathbf{g}} EWSR(\mathbf{g}), \text{where}$$

$$EWSR(\mathbf{g}) = \mathrm{E}_{\mathbf{h}|\widehat{\mathbf{h}}} WSR(g) = \sum_{k=1}^{K} u_k\, \mathrm{E}_{\mathbf{h}|\widehat{\mathbf{h}}} \ln(s_k/s_{\overline{k}})$$

$$= \mathrm{E}_{\mathbf{h}|\widehat{\mathbf{h}}} \sum_{k=1}^{K} u_k \ln(1 + \frac{|\mathbf{h}_{k,b_k}^H\mathbf{g}_k|^2}{s_{\overline{k}}})$$

$$\overset{(a)}{\approx} \mathrm{E}_{\mathbf{h}|\widehat{\mathbf{h}}} \sum_{k=1}^{K} u_k \ln(1+\frac{|\mathbf{h}_{k,b_k}^H \mathbf{g}_k|^2}{\mathrm{E}_{\mathbf{h}}\, s_{\overline{k}}})$$
$$\overset{(b)}{\leq} \sum_{k=1}^{K} u_k \;\ln(1+\frac{\mathrm{E}_{\mathbf{h}|\widehat{\mathbf{h}}}|\mathbf{h}_{k,b_k}^H\, \mathbf{g}_k|^2}{\mathrm{E}_{\mathbf{h}|\widehat{\mathbf{h}}}\, s_{\overline{k}}})$$
$$= \sum_{k=1}^{K} u_k \;\ln(r_{\overline{k}}^{-1} r_k) = ESIP-WSR(\mathbf{g}) \quad (7)$$

where u_k are the rate weights, $\mathbf{g}$ represents the collection of BFs $\mathbf{g}_k$. Transition (a) is due to the MaMISO limit ($K \to \infty$) and (b) is due to the concavity of $\ln(.)$ and Jensen's inequality. This leads to the ESIP-WSR (Expected Signal and Interference Power WSR) upper bound. $s_{\overline{k}}$ is the (channel dependent) interference plus noise power and s_k is the total received power, with conditional expectations $r_{\overline{k}}$, r_k:

$$s_{\overline{k}} = 1 + \sum_{i \neq k} |\mathbf{h}_{k,b_i}^H \mathbf{g}_i|^2, s_k = s_{\overline{k}} + |\mathbf{h}_{k,b_k}^H \mathbf{g}_k|^2$$
$$r_{\overline{k}} = \mathrm{E}_{\mathbf{h}|\widehat{\mathbf{h}}} s_{\overline{k}} = 1 + \sum_{i \neq k} \mathbf{g}_i^H \mathbf{S}_{k,b_i} \mathbf{g}_i,$$
$$r_k = \mathrm{E}_{\mathbf{h}|\widehat{\mathbf{h}}} s_k = r_{\overline{k}} + \mathbf{g}_k^H \mathbf{S}_{k,b_k} \mathbf{g}_k,\;\; \mathbf{S}_{k,b_k} = \mathbf{C}_{k,b_k} \mathbf{W}_{k,b_k} \mathbf{C}_{k,b_k}. \quad (8)$$

The ESIP-WSR upper bound can be somewhat loose (gap is maximal at high SNR, see [13]), because inspite of $\mathbf{h}_k$ being MISO, $\mathbf{g}_k^H \mathbf{h}_k$ is only a simple complex Gaussian scalar. Nevertheless, this gap is upper bounded by the Euler constant $\gamma = 0.58$, regardless of SNR. And for the case of only coCSIT ($\widehat{\mathbf{h}} = 0$), the gap is exactly γ, which means that it has no influence on the optimal $\mathbf{g}_k$.

Now optimizing $ESIP-WSR(\mathbf{g})$ w.r.t. $\mathbf{g}_k$ leads to the following generalized eigenvector.

$$\mathbf{g}_k' = \mathbf{V}_{max}(\widehat{\mathbf{B}}_k, \widehat{\mathbf{A}}_k + \mu_{b_k} \mathbf{I})\;. \quad (9)$$

where $\widehat{\mathbf{A}}_k = \sum_{i=1,\neq k}^{K} \beta_i \mathbf{S}_{i,b_k}, \widehat{\mathbf{B}}_k = r_{\overline{k}}^{-1} \mathbf{S}_{k,b_k}$. We skip here the details of the derivation of the BFs and user powers and refer to our paper [3] due to space limitations. (9) comes from the difference of convex functions programming (DCP) [14] and $\beta_k = u_k(\frac{1}{r_{\overline{k}}} - \frac{1}{r_k})$). Substituting $\mathbf{g}_k = \sqrt{p_k}\, \mathbf{g}_k'$ and optimizing using DCP leads to the following *interference leakage* ($\sigma_k^{(2)}$) *aware water filling*

$$p_k = \left(\frac{u_k}{\sigma_k^{(2)} + \mu_{b_k}} - \frac{1}{\sigma_k^{(1)}}\right)^+, \quad (10)$$

where $(x)^+ = \max\{0, x\}$, $\sigma_k^{(1)} = \mathbf{g}_k'^H \widehat{\mathbf{B}}_k \mathbf{g}_k'$, $\sigma_k^{(2)} = \mathbf{g}_k'^H \widehat{\mathbf{A}}_k \mathbf{g}_k'$, and the Lagrange multipliers μ_c are adjusted (e.g. by bisection) to satisfy the power constraints.

B. Further Considerations on EWSR Bounds

We observed that ESIP-WSR represents an upper bound to the massive MIMO ergodic capacity. Three types of BF design with partial CSIT can actually be analyzed. We get for the Rx signal,

$$y_k = \widehat{\mathbf{h}}_{k,b_k}^H\, \mathbf{g}_k\, x_k + \underbrace{\widetilde{\mathbf{h}}_{k,b_k}^H\, \mathbf{g}_k\, x_k}_{\text{sig. ch. error}}$$
$$+ \sum_{i=1,\neq k}^{K} (\widehat{\mathbf{h}}_{k,b_i}^H\, \mathbf{g}_i\, x_i + \underbrace{\widetilde{\mathbf{h}}_{k,b_i}^H\, \mathbf{g}_i\, x_i}_{\text{interf. ch. error}}) + v_k\,. \quad (11)$$

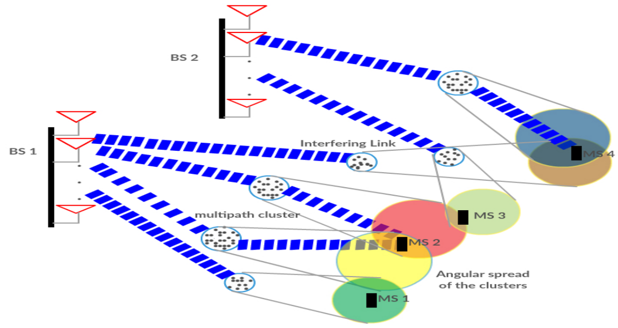

Fig. 1. COST2100 MIMO Channel Model.

1) Naive BF EWSR: just replace $\mathbf{h}$ by $\widehat{\mathbf{h}}$ in a perfect CSIT approach, i.e. ignore $\widetilde{\mathbf{h}}$ everywhere.
2) EWSMSE BF (Expected Weighted Sum MSE) [15]: accounts for covariance CSIT in the interference terms, but also associates the signal $\widetilde{\mathbf{h}}$ term with the interference ! EWSMSE, also called the "use and forget lower bound" in [16], can indeed be shown to be a lower bound for EWSR.
3) EWSR upper bound ESIP-WSR: also accounts for covariance CSIT in the interference term but, unlike EWSMSE, associates the signal $\widetilde{\mathbf{h}}$ term with the signal power.

IV. Stochastic Geometry MaMIMO Regime

The channel model (2) results from multipath propagation and the use of BS side antenna arrays. An example of a geometry based stochastic channel model is provided by the COST2100 channel model [17], and can be depicted as in Figure 1. One particular application of this model for the user covariance matrices is considered in [18]. [18] considers the scenario in which the support of the multipath angle of arrival or departure (AoA/AoD) for any desired user does not overlap with that of the interfering users. The authors show that the multipath components with AoA/AoD outside the angular support of the desired user tend to fall in the null space of its covariance matrix in the large antenna limit, leading to orthogonal subspaces $\mathbf{C}$ in the MaMISO regime. Here we add a stochastic geometry regime, in which the random positions of users and scatterers lead to antenna array responses at random angles. In reality, the antenna array responses will be more complex than the Vandermonde vectors for Uniform Linear Arrays considered in [18] due to mutual antenna coupling and various other effects. As a result of this randomness of angles and antenna array responses, and due to limited angular support, the multipath channels live in subspaces that are of limited dimension and uniformly randomly oriented in array response space. As a result, an appropriate random model for the semi-unitary matrices $\mathbf{C}$ spanning these subspaces is a Haar distribution. We shall consider that as the number of antennas M grows unboundedly, the subspace dimensions L also go to infinity (leading to hardening of the signal power), but slower than M. As a result, for the large system analysis we may equivalently consider the elements of $\mathbf{C}$ as i.i.d. with zero mean and variance $1/M$ so that asymptotically such a $\mathbf{C}$ is still semi-unitary: $\mathbf{C}^H\mathbf{C} \overset{M\to\infty}{\longrightarrow} \mathbf{I}_L$. The subspaces $\mathbf{C}$ of different channels will be considered independent.

In this section, we summarize the large system results from [19] we use and directly write the simplified form of the BF expression of $\mathbf{g}_k$ in (9) and refer to [3] for the derivation. For the large system analysis, we use Theorem 1, Lemma $1, 4, 6$ from [19]. We briefly summarize the Lemma's here. Lemma 4 in Appendix VI of [19] states that $\boldsymbol{x}_N^H \mathbf{A}_N \boldsymbol{x}_N \xrightarrow{N\to\infty} (1/N)tr\mathbf{A}_N$ when the elements of $\boldsymbol{x}_N$ are iid with variance $1/N$ and independent of $\mathbf{A}_N$, and similarly when $\boldsymbol{y}_N$ is independent of $\boldsymbol{x}_N$, that $\boldsymbol{x}_N^H \mathbf{A}_N \boldsymbol{y}_N \xrightarrow{N\to\infty} 0$. Lemma 6 from [19] (rank 1 perturbation lemma) states that $\frac{1}{N}\text{tr}\{\mathbf{A}_N^{-1}\} - \frac{1}{N}\text{tr}\{(\mathbf{A}_N + \boldsymbol{v}\boldsymbol{v}^H)^{-1}\} \xrightarrow{N\to\infty} 0$ to approximate terms of the form $[\sum\limits_{i\neq k} \beta_i \mathbf{C}_{i,c}\mathbf{W}_{i,c}\mathbf{C}_{i,c}^H + \mu_c\,\mathbf{I}]^{-1} = [\sum\limits_{i=1} \beta_i \mathbf{C}_{i,c}\mathbf{W}_{i,c}\mathbf{C}_{i,c}^H + \mu_c\,\mathbf{I}]^{-1}$. Theorem 1 from [19] implies that any term of the form $\frac{1}{N}\text{tr}\{(\mathbf{A}_N - z\mathbf{I}_N)^{-1}\}$, where $\mathbf{A}_N$ is the summation of independent rank one matrices with covariance matrix $\boldsymbol{\Theta}_i$ is equal to the unique positive solution of $e_j = \frac{1}{N}\text{tr}\{(\sum\limits_{i=1}^{K} \frac{\boldsymbol{\Theta}_i}{1+e_i} - z\mathbf{I}_N)^{-1}\}$. Thus $\text{tr}\{[\sum\limits_{i\neq k} \beta_i \mathbf{C}_{i,c}\mathbf{W}_{i,c}\mathbf{C}_{i,c}^H + \mu_c\,\mathbf{I}]^{-1}\}$ can be simplified as e_c, which is defined as the solution to the following fixed point equation,

$$e_c = \left(\frac{1}{M_c}\sum_{i=1}^{K}\sum_{r=1}^{L_{i,c}} \frac{\beta_i \zeta_{i,c}^{(r)}}{1+\beta_i \zeta_{i,c}^{(r)} e_c} + \mu_c\right)^{-1}, \tag{12}$$

where $\zeta_{i,c}^{(r)}$ follows from the eigen decomposition of $\mathbf{W}_{k,c} = \mathbf{V}_{k,c}\mathbf{\Lambda}_{k,c}\mathbf{V}_{k,c}^H$, where $\mathbf{\Lambda}_{k,c} = \text{diag}(\zeta_{k,c}^{(1)}, ..., \zeta_{k,c}^{(L_{k,c})})$. We also remark that $\mathbf{C}_{k,c}\mathbf{V}_{k,c}$ which is the product of two matrices (also remains unitary) has the identity covariance matrix, so theorem 1 is still valid here. From [3], we write the optimized BF w.r.t. partial CSIT, in the stochastic geometry MaMIMO regime as,

$$\mathbf{g}_k' = \frac{\mathbf{g}_k''}{\|\mathbf{g}_k''\|},\ \ \mathbf{g}_k'' = [\sum_{i\neq k} \beta_i \mathbf{S}_{i,b_k} + \mu_{b_k}\,\mathbf{I}]^{-1}\mathbf{C}_{k,b_k}\boldsymbol{v}_{k,b_k}, \\ \text{where,}\ \ \boldsymbol{v}_{k,b_k} = \mathbf{V}_{max}(\mathbf{W}_{k,b_k}). \tag{13}$$

A. *Computation of eigen values of* $\mathbf{W}_{k,b_i}$

For the convenience of analysis we omit the user and BS index here. We represent $\mathbf{W}$ by $\mathbf{W}_L, \mathbf{W}_S$ for LMMSE and subspace channel estimators respectively. From Section III, we define $\widehat{\mathbf{d}} = \mathbf{C}^H\widehat{\mathbf{h}}_{LS}$, $\mathbf{W}_L = (\mathbf{I} + \widetilde{\sigma}^2\mathbf{D}^{-1})^{-1}\widehat{\mathbf{d}}\widehat{\mathbf{d}}^H(\mathbf{I} + \widetilde{\sigma}^2\mathbf{D}^{-1})^{-1} + \mathbf{D} - \mathbf{D}(\widetilde{\sigma}^2\mathbf{I} + \mathbf{D})^{-1}\mathbf{D}$. At both high and low SNR, we can replace the error covariance $\mathbf{D} - \mathbf{D}(\widetilde{\sigma}^2\mathbf{I} + \mathbf{D})^{-1}\mathbf{D}$ by its dominating term $\mathbf{D}_{1,1}\widetilde{\sigma}^2/(\mathbf{D}_{1,1} + \widetilde{\sigma}^2)\boldsymbol{e}_1\boldsymbol{e}_1^H$, assuming $\mathbf{D}_{1,1}$ is the largest diagonal element of $\mathbf{D}$ and $\boldsymbol{e}_1$ is vector of all zeros with the first element 1. Thus $\mathbf{W}_L$ becomes the sum of two rank one matrices and we propose the resulting rank 2 approximation for $\mathbf{W}_L$ for all SNR. It will be all the more precise at intermediate SNR if $\mathbf{D}_{1,1}$ dominates the rest of $\mathbf{D}$. Further we look at the computation of the eigen value matrix $\mathbf{\Lambda}$ of $\mathbf{W}_L$. For the LMMSE,

$$\mathbf{W}_L = \boldsymbol{U}_L(\widehat{\mathbf{d}}\widehat{\mathbf{d}}^H + \widetilde{\sigma}^2\mathbf{I})\boldsymbol{U}_L^H + (\mathbf{I} - \boldsymbol{U}_L)\mathbf{D}(\mathbf{I} - \boldsymbol{U}_L)^H \tag{14}$$

where $\boldsymbol{U}_L = (\mathbf{I} + \widetilde{\sigma}^2\mathbf{D}^{-1})^{-1}$, $\boldsymbol{U}_L\widehat{\mathbf{d}} = \widehat{\mathbf{c}}_L$ for short. At high SNR, we can approximate $(\mathbf{I} + \widetilde{\sigma}^2\mathbf{D}^{-1})^{-1} = \mathbf{I} - \widetilde{\sigma}^2\mathbf{D}^{-1}$. So up to first order in $\widetilde{\sigma}^2$, we obtain $\mathbf{W}_L \approx \widehat{\mathbf{c}}_L\widehat{\mathbf{c}}_L^H + \widetilde{\sigma}^2\mathbf{I}$, where the first term contains first-order terms also. At low SNR, $\boldsymbol{U}_L \approx \widetilde{\sigma}^{-2}\mathbf{D}$ and we can obtain $\mathbf{W}_L \approx \widehat{\mathbf{c}}_L\widehat{\mathbf{c}}_L^H + \mathbf{D}$. For any SNR, these two extremes can be connected by the following approximation:

$$\begin{aligned} \mathbf{\Lambda} &= \widetilde{\sigma}^2\mathbf{D}(\widetilde{\sigma}^2\mathbf{I} + \mathbf{D})^{-1} + ||\widehat{\mathbf{c}}_L||^2\boldsymbol{e}_1\boldsymbol{e}_1^H \\ &= \widetilde{\sigma}^2\mathbf{D}(\widetilde{\sigma}^2\mathbf{I} + \mathbf{D})^{-1} + \text{tr}\{\mathbf{D}(\mathbf{I} + \widetilde{\sigma}^2\mathbf{D}^{-1})^{-1}\}\,\boldsymbol{e}_1\boldsymbol{e}_1^H \end{aligned} \tag{15}$$

where the last equality is due to the law of large numbers. For subspace channel estimator, $\mathbf{W}_S = \widehat{\mathbf{d}}\widehat{\mathbf{d}}^H + \widetilde{\sigma}^2\mathbf{I}$, the eigen value matrix becomes $\mathbf{\Lambda} = \widetilde{\sigma}^2\mathbf{I} + \text{tr}\{\mathbf{D}\}\,\boldsymbol{e}_1\boldsymbol{e}_1^H$.

V. Various BF High SNR Sum Rate Expressions with LMMSE/Subspace Channel Estimator

In the Appendix A, we derive the sum rate expressions ($u_k = 1, \forall k$) applicable at all SNR for the various BFs (optimal EWSR, naive and EWSMSE). In this section, using the results from the Appendix A, we consider the simplified high SNR sum rate expressions for naive, EWSMSE and optimal EWSR BFs for LMMSE/Subspace/LS channel estimators under multi cell (C cells), with identical parameters, $\widetilde{\sigma}_{k,c}^2 = \widetilde{\sigma}^2$, $L_{k,c} = L$, $\mathbf{D}_{k,c} = \frac{\eta_{k,c}}{L_{k,c}}\mathbf{I}$ and $M_c = M, \forall k, c$. Number of users in cell c is denoted as $K_c = K/C, \forall c$. We obtain,

$$\zeta_{k,c}^{(1)} = \frac{\eta_{k,c}^2}{L\widetilde{\sigma}^2 + \eta_{k,c}} + \frac{\widetilde{\sigma}^2\eta_{k,c}}{L\widetilde{\sigma}^2 + \eta_{k,c}}, \tag{16}$$

and rest of the eigen values $\zeta_{k,c}^{(r)} = \frac{\widetilde{\sigma}^2\eta_{k,c}}{L\widetilde{\sigma}^2 + \eta_{k,c}}, \forall r = 2, ..., L$. For the subspace channel estimator, the eigen values are $\zeta_{k,c}^{(1)} = (\eta_{k,c} + L\widetilde{\sigma}^2) + \widetilde{\sigma}^2, \zeta_{k,c}^{(r)} = \widetilde{\sigma}^2, \forall r \neq 1$. For LS only channel estimate, the eigen values are $\zeta_{k,c}^{(1)} = (\eta_{k,c} + L\widetilde{\sigma}^2) + \widetilde{\sigma}^2$ and $\zeta_{k,c}^{(r)} = \widetilde{\sigma}^2$. We define $\rho_{k,c} = \eta_{k,c}p_k$, where $\rho_{k,c}$ is the received SNR at user k from BS c. Substituting these values, we observe that the sum rate expressions at high SNR can be expressed as,

$$\overline{R} = \sum_{k=1}^{K} \ln(1 + \alpha\rho_{k,b_k}), \tag{17}$$

where $\alpha = \frac{1-z}{1+y}$, where z, y varies w.r.t the channel estimator and the type of BF design. The corresponding α for the 9 different combinations of channel estimator and BFs are depicted in the Table I below. If we consider the simplified case of identical channel attenuation for all users in the system, $\eta_{k,c} = \eta, \forall k, c$, for which the sum rate simplifies to $\overline{R} = K\ln(1 + \alpha\rho), \rho = \eta\frac{P}{K_c}$. Note that at high SNR, EWSR BF with subspace channel estimator converges to the performance of the LMMSE estimator, hence we have merged the values of subspace and LMMSE estimator in the tables.

A. *Sum Rate Analysis at Low SNR for LMMSE and Subspace Channel Estimators*

For simplicity of notation, we drop user and BS indices in this section. So, at low SNR, all the interference are negligible and the BF gets simplified as, $\mathbf{g} = \mathbf{V}_{max}(\widehat{\mathbf{h}}\widehat{\mathbf{h}}^H + R_{\widetilde{\mathbf{h}}\widetilde{\mathbf{h}}}) = \mathbf{C}\mathbf{V}_{max}(\mathbf{W})$, where $\mathbf{W} = \boldsymbol{U}(\widehat{\mathbf{d}}\widehat{\mathbf{d}}^H + \widetilde{\sigma}^2\mathbf{I}_L)\boldsymbol{U}^H + (\boldsymbol{U} - \mathbf{I})\mathbf{D}(\boldsymbol{U} - \mathbf{I})^H$ where $\widehat{\mathbf{d}} = \mathbf{C}^H\widehat{\mathbf{h}}_{LS} = \mathbf{d} + \widetilde{\mathbf{d}}$. And at low SNR, the following simplifications can be done for $\boldsymbol{U}$ and $\mathbf{W}$ for LMMSE estimator,

TABLE I
HIGH SNR SUM RATE OFFSET FOR VARIOUS BFS

α	naive	EWSMSE	ESIP-EWSR
LS	$\frac{1-\frac{K-1}{M}}{1+\widetilde{\sigma}^2 P}$	$\frac{1-\frac{K-1}{M}}{1+\widetilde{\sigma}^2 P}$	$\frac{1-\frac{K-1}{M}}{1+\widetilde{\sigma}^2 P}$
LMMSE/Subspace	$\frac{1-\frac{K-1}{M}}{1+\frac{1}{M}\widetilde{\sigma}^2 P}$	$1-\frac{(K-1)L}{M}$	$1-\frac{(K-1)L}{M}$

$$\begin{aligned} \boldsymbol{U}_L &= (\mathbf{I}+\widetilde{\sigma}^2\mathbf{D}^{-1})^{-1} \approx \widetilde{\sigma}^{-2}\mathbf{D}, \\ \text{So, } \mathbf{W}_L &= \widetilde{\sigma}^{-2}\mathbf{D}(\widetilde{\sigma}^{-2}\widetilde{\mathbf{d}}\widetilde{\mathbf{d}}^H+\mathbf{I}_L)\mathbf{D}+\mathbf{D}. \end{aligned} \quad (18)$$

There is no signal concentration along $\mathbf{h}$ or $\mathbf{d}$, $\mathbf{V}_{max}(\mathbf{W}_L)$ remains a random projection in the channel subspace $\mathbf{C}$, if $\mathbf{D}$ is a multiple of identity. If $\mathbf{D}$ is not a multiple of identity, $\mathbf{V}_{max}(\mathbf{W}_L)$ is a function of $\widetilde{\mathbf{d}}$, $\mathbf{D}$ and $\widetilde{\sigma}^2$, independent of $\mathbf{d}$ which appears in $\mathbf{h}$. Further considering the signal part, $\mathrm{E}(|\mathbf{g}^H\mathbf{h}|^2) = \mathrm{tr}\{\mathbf{D}\mathbf{E}\mathbf{V}_{max}(\mathbf{W}_L)\mathbf{V}_{max}(\mathbf{W}_L)^H\}$, for example, in the extreme case, $\mathbf{D} = \mathrm{tr}\{\mathbf{D}\}\boldsymbol{e}_1\boldsymbol{e}_1^H$. Then $\mathbf{W}_L$ is proportional to $\boldsymbol{e}_1\boldsymbol{e}_1^H$, hence $\mathbf{g} = \mathbf{C}\boldsymbol{e}_1$. Then $\mathrm{E}(|\mathbf{g}^H\mathbf{h}|^2) = \mathrm{tr}\{\mathbf{D}\}$ but with $\|\mathbf{g}\| = 1$. For subspace channel estimator, $\mathbf{W}_S = \widetilde{\mathbf{d}}\widetilde{\mathbf{d}}^H + \widetilde{\sigma}^2\mathbf{I}_L$ and $\mathbf{V}_{max}(\mathbf{W}_S) = \widetilde{\mathbf{d}}$. Substituting for $\mathbf{g} = \mathbf{C}\widetilde{\mathbf{d}}$ in the signal part and applying law of large numbers leads to $\mathrm{E}(|\mathbf{g}^H\mathbf{h}|^2) \xrightarrow[a.s]{L,M\to\infty} \widetilde{\mathbf{d}}^H\mathbf{D}\widetilde{\mathbf{d}} \xrightarrow[a.s]{L\to\infty} \widetilde{\sigma}^2\mathrm{tr}\{\mathbf{D}\}$. Similarly computing $\|\mathbf{g}\|^2 = \widetilde{\mathbf{d}}^H\widetilde{\mathbf{d}} \xrightarrow[a.s]{L\to\infty} \mathrm{tr}\{\widetilde{\sigma}^2\mathbf{I}_L\} = \widetilde{\sigma}^2 L$. So, we conclude that the signal power equals $\mathrm{tr}\{\mathbf{D}\}$ for optimal EWSR BF with LMMSE instead of $\mathrm{tr}\{\mathbf{D}\}/L$ for subspace channel estimator. This explains why LMMSE performs better than subspace estimator at low SNR, also illustrated by our simulations.

Further we consider the simplified case of $\mathbf{D}_{k,c} = \frac{\eta}{L}\mathbf{I}, \forall k, c$. Doing a similar analysis as above at low SNR for naive and EWSMSE BFs for the various channel estimates, we observe that the sum rate can be written as follows,

$$\overline{R} = C\ln(1+\gamma\rho) \overset{a}{\approx} C\gamma\rho, \text{ where, } \rho = \eta P, \quad (19)$$

where in (a), we made the approximation $\ln(1+x) \approx x$, when $x \ll 1$ and γ represents the SNR offset for various BFs. In Table II, we show the γ (detailed derivations are omitted) for different BF and channel estimator combination to explain the SNR offset for sub-optimal BFs compared to the ESIP-WSR BF. Note that at low SNR, ILA-WF allocates all the power to the strongest (in terms of channel attenuation) user resulting in the received SNR $\rho = \eta P$ for the corresponding user.

VI. 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_c, number of (single-antenna) users in cell c and $K = \sum_c K_c$. M, number of transmit antennas in each cell. We consider a path-wise or low rank channel model as in section II-A, with L = number of paths = channel covariance rank. d : scale factor in the LS

TABLE II
LOW SNR SUM RATE OFFSET FOR VARIOUS BFS ($\eta_k = \eta$)

γ	naive	EWSMSE	ESIP-EWSR
LS	$\frac{1+\frac{\widetilde{\sigma}^2 M}{\eta}}{1+P\frac{\widetilde{\sigma}^4 M^2}{\eta+\widetilde{\sigma}^2 M}}$	$\frac{1+\frac{\widetilde{\sigma}^2 M}{\eta}}{1+P\frac{\widetilde{\sigma}^4 M^2}{\eta+\widetilde{\sigma}^2 M}}$	$\frac{\eta}{\eta+\widetilde{\sigma}^2 M}$
LMMSE	$\frac{\eta}{L\widetilde{\sigma}^2(1+\rho\frac{\eta}{L\widetilde{\sigma}^2})}$	$\frac{\eta}{L\widetilde{\sigma}^2(1+\rho\frac{\eta}{L\widetilde{\sigma}^2})}$	1
Subspace	$\frac{L\widetilde{\sigma}^2}{\eta(1+\rho\frac{L\widetilde{\sigma}^2}{\eta})}$	$\frac{L\widetilde{\sigma}^2}{\eta(1+\rho\frac{L\widetilde{\sigma}^2}{\eta})}$	$\frac{1}{L}$

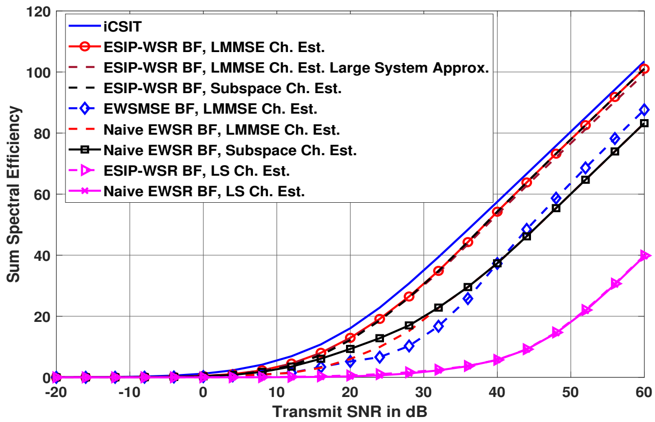

Fig. 2. EWSR for $C = 1$ cell, $K_1 = K = 10$ users, $M = 64$, $L = 2$, $\widetilde{\sigma}^2 \propto 1/SNR$.

channel estimation error variance $\widetilde{\sigma}^2 = d/SNR$. Notations: in the figures, iCSIT refers to the optimal BF design for the instantaneous CSIT case [20]. In Figure 2, we plot the optimal BF performance with LMMSE channel estimator comparing to the optimal BF performance for the case of large system approximation. It is evident that the deterministic approximations are accurate even for finite M, K. It is evident from the figure that exploiting the channel estimation error covariance information has significant performance gain compared to the sub-optimal methods such as EWSMSE and naive BFs. In Figure 3, we plot the EWSMSE beamforming performance also and it is evident from the figure that ESIP-WSR based beamformers (i.e. MaMISO limit based) perform better EWSR approximations than a EWSMSE design. From the numerical simulations in both Figures 2,3, it is quite evident that just using LS channel estimates may lead to substantial EWSR loss. In Massive MIMO, the exploitation of channel subspaces

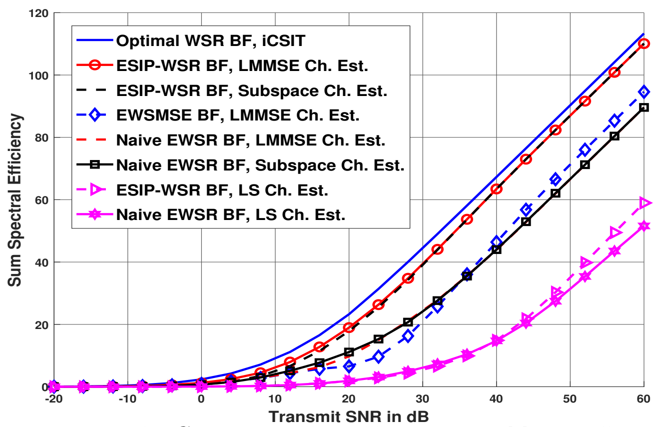

Fig. 3. EWSR for $C = 2$ cells, $K_1 = K_2 = 5$ users, $M = 32$, $L = 2$, $\widetilde{\sigma}^2 \propto 1/SNR$.

(reduced rank covariances) in channel estimates may lead to substantial reductions in SNR loss due to partial CSIT. Moreover, there is significant gain from exploiting (error) channel covariances in addition to (LMMSE) channel estimates and proper handling of channel error covariance in the direct link in the BF design.

VII. Conclusion

This paper investigated the optimal linear precoder based on partial CSIT in the multi-cell MU-MISO downlink. We introduced a stochastic geometry inspired randomization of the channel covariance eigen spaces and analyzed the large system behavior. Moreover, we show the improvement in performance by using an LMMSE channel estimate compared to just having LS estimates, and by furthermore properly exploiting all covariance information. Numerical simulations suggest that the large system approximations are accurate even for finite values of M, K. We provided simple and elegant expressions for the sum rate at high and low SNR, providing useful analytical insights into the SNR offsets between different suboptimal BFs which matches with our simulations.

Appendix A
Sum Rate Evaluation

A. Optimal EWSR BF with LMMSE/Subspace Estimators

In Section V, for simplicity of analysis, we consider only the case of $\mathbf{D}_{k,c} = \frac{\eta_{k,c}}{L_{k,c}}\mathbf{I}$, the random user positions leads to random attenuation factors $\eta_{k,c}$ for each channel. However, we consider that this random attenuation is known at the BS. In this simplified case, we obtain $\mathbf{V}_{max}(\mathbf{W}_{k,c}) \propto \boldsymbol{U}_{k,c}\widehat{\mathbf{d}}_{k,c}$. For the convenience of analysis, we write $\mathbf{g}_k = \boldsymbol{\Gamma}_k^{-1}\mathbf{C}_{k,b_k}\boldsymbol{v}_{k,b_k}$. Here $\boldsymbol{\Gamma}_k = \sum_{i \neq k} \beta_i \mathbf{S}_{i,b_k} + \mu_{b_k}\,\mathbf{I}$. Also, let $\gamma_{k,L}^{(s)}, \gamma_{k,S}^{(s)}$ denotes the SINRs for user k in the case of LMMSE and subspace channel estimators respectively. The superscript s can be 'Opt' or 'N' or 'E' which represents the optimal/naive/EWSMSE BFs respectively. By large system limit, we implies that $L, M, K \rightarrow \infty$. First we compute the deterministic equivalent for the signal power P_{S_k},

$$
\begin{aligned}
&\mathbf{g}'_k = \mathbf{g}''_k / \|\mathbf{g}''_k\|\,, \mathbf{g}''_k = \boldsymbol{\Gamma}_k^{-1}\mathbf{C}_{k,b_k}\boldsymbol{v}_{k,b_k},\\
&P_{S_k} = p_k \mathbf{g}'^{\,H}_k \mathbf{C}_{k,b_k}\mathbf{d}_{k,b_k}\mathbf{d}_{k,b_k}^H \mathbf{C}_{k,b_k}^H \mathbf{g}'_k,\\
&\mathbf{g}''^{\,H}_k \mathbf{C}_{k,b_k}\mathbf{d}_{k,b_k} = \widehat{\mathbf{d}}_{k,b_k}^H \boldsymbol{U}_{k,b_k}\mathbf{C}_{k,b_k}^H \boldsymbol{\Gamma}_k^{-1}\mathbf{C}_{k,b_k}\mathbf{d}_{k,b_k},\\
&= (\mathbf{d}_{k,b_k} + \widetilde{\mathbf{d}}_{k,b_k})^H \boldsymbol{U}_{k,b_k}\mathbf{C}_{k,b_k}\boldsymbol{\Gamma}_k^{-1}\mathbf{C}_{k,b_k}\mathbf{d}_{k,b_k},\\
&\overset{(a)}{=} e_{b_k}\mathrm{E}(\mathrm{tr}\{\boldsymbol{U}_{k,b_k}\mathbf{d}_{k,b_k}\mathbf{d}_{k,b_k}^H\}) = e_{b_k}\mathrm{tr}\{\boldsymbol{U}_{k,b_k}\mathbf{D}_{k,b_k}\},
\end{aligned}
\tag{20}
$$

where (a) follows from the law of large numbers as $L_{k,b_k}, M_{b_k} \rightarrow \infty$ and using the large system analysis simplifications shown in (12), $\mathbf{C}_{k,b_k}^H\boldsymbol{\Gamma}_k^{-1}\mathbf{C}_{k,b_k} = e_{b_k}\mathbf{I}$. Also note that $\mathrm{E}(\mathbf{d}_{k,b_k}\widetilde{\mathbf{d}}_{k,b_k}^H) = 0$ since $\mathbf{d}_{k,b_k}$ and $\widetilde{\mathbf{d}}_{k,b_k}$ are zero mean and independent. Further,

$$
\begin{aligned}
&\mathbf{g}'^{\,H}_k \mathbf{h}_{k,b_k}\mathbf{h}_{k,b_k}^H \mathbf{g}'_k = e_{b_k}^2 \mathrm{tr}\{\boldsymbol{U}_{k,b_k}\mathbf{D}_{k,b_k}\}^2 / \|\mathbf{g}''_k\|^2 = \mathbf{g}_k,\\
&\|\mathbf{g}''_k\|^2 = \widehat{\mathbf{d}}_{k,b_k}^H \boldsymbol{U}_{k,b_k}\mathbf{C}_{k,b_k}^H \boldsymbol{\Gamma}_k^{-2}\mathbf{C}_{k,b_k}\boldsymbol{U}_{k,b_k}\widehat{\mathbf{d}}_{k,b_k}.
\end{aligned}
\tag{21}
$$

Further in the large system limit, $\|\mathbf{g}''_k\|^2$ gets simplified as $\|\mathbf{g}''_k\|^2 = \frac{1}{M_{b_k}}\mathrm{tr}\{\boldsymbol{\Gamma}_k^{-2}\}\mathrm{tr}\{\boldsymbol{U}_{k,b_k}^2 \mathrm{E}(\widehat{\mathbf{d}}_{k,b_k}\widehat{\mathbf{d}}_{k,b_k}^H)\}$. Further this can be simplified as $\|\mathbf{g}''_k\|^2 = e'_{b_k}\mathrm{tr}\{\boldsymbol{U}_{k,b_k}^2(\mathbf{D}_{k,b_k} + \widetilde{\sigma}_k^2\mathbf{I})\}$. From [19], in the large system limit, for $(1/M_{b_k})\mathrm{tr}\{\boldsymbol{\Gamma}_{b_k}^{-2}\}$, we have an almost sure convergence value as e'_{b_k}, where e'_{b_k} is the derivative of e_{b_k} w.r.t. μ_{b_k}, and thus $\|\mathbf{g}''_k\|^2 = e'_{b_k}$,

$$
\begin{aligned}
&e'_c = e_c^2\Big(\frac{1}{M_c}\sum_{i=1}^{K}\sum_{r=1}^{L_{i,c}} \frac{\beta_i^2 \zeta_{i,c}^{(r),\,2} e'_c}{(1+\beta_i\zeta_{i,c}^{(r)}e_c)^2} + 1\Big)\\
&\implies e'_c = \frac{e_c^2}{1 - \frac{e_c^2}{M_c}\sum\limits_{i=1}^{K}\sum\limits_{r=1}^{L_{i,c}}\frac{\beta_i^2\zeta_{i,c}^{(r),\,2}}{(1+\beta_i\zeta_{i,c}^{(r)}e_c)^2}}.
\end{aligned}
\tag{22}
$$

Finally we write the signal power as,

$$
P_{S_k} = p_k \frac{e_{b_k}^2 \mathrm{tr}\{\mathbf{D}_{k,b_k}\boldsymbol{U}_{k,b_k}\}^2}{e'_{b_k}\mathrm{tr}\{\boldsymbol{U}_{k,b_k}^2(\mathbf{D}_{k,b_k}+\widetilde{\sigma}_k^2\mathbf{I})\}},
\tag{23}
$$

$$
\begin{aligned}
&x_c = \frac{e_c^2}{M_c}\sum_{i=1}^{K}\sum_{r=1}^{L_{i,c}}\frac{\beta_i^2\zeta_{i,c}^{(r),\,2}}{(1+\beta_i\zeta_{i,c}^{(r)}e_c)^2},\; e'_c = \frac{e_c^2}{1-x_c},\\
&\text{therefore, } P_{S_k} = p_k(1-x_{b_k})\frac{\mathrm{tr}\{\mathbf{D}_{k,b_k}\boldsymbol{U}_{k,b_k}\}^2}{\mathrm{tr}\{\boldsymbol{U}_{k,b_k}^2(\mathbf{D}_{k,b_k}+\widetilde{\sigma}_k^2\mathbf{I})\}},
\end{aligned}
\tag{24}
$$

Deterministic limit for interference power P_{I_k}: Each term in P_{I_k} is of the form, $\mathbf{g}''^{\,H}_i \mathbf{h}_{k,b_i}\mathbf{h}_{k,b_i}^H\mathbf{g}''_i = \boldsymbol{v}_{i,b_i}^H\mathbf{C}_{i,b_i}^H\boldsymbol{\Gamma}_i^{-1}\mathbf{C}_{k,b_i}\mathbf{d}_{k,b_i}\mathbf{d}_{k,b_i}^H\mathbf{C}_{k,b_i}^H\boldsymbol{\Gamma}_i^{-1}\mathbf{C}_{i,b_i}\boldsymbol{v}_{i,b_i}$. Since $\mathbf{C}_{i,b_i}\boldsymbol{v}_{i,b_i}$ is independent of all other random quantities in this expression, we apply Lemma 4 and then Lemma 6 to get, $\boldsymbol{v}_{i,b_i}^H\mathbf{C}_{i,b_i}^H\boldsymbol{\Gamma}_i^{-1}\mathbf{C}_{k,b_i}\mathbf{d}_{k,b_i}\mathbf{d}_{k,b_i}^H\mathbf{C}_{k,b_i}^H\boldsymbol{\Gamma}_i^{-1}\mathbf{C}_{i,b_i}\boldsymbol{v}_{i,b_i} = \frac{1}{M_{b_i}}\mathrm{tr}\{\boldsymbol{\Gamma}_i^{-1}\mathbf{C}_{k,b_i}\mathrm{E}(\mathbf{d}_{k,b_i}\mathbf{d}_{k,b_i}^H)\mathbf{C}_{k,b_i}^H\boldsymbol{\Gamma}_i^{-1}\}\mathrm{tr}\{\boldsymbol{U}_{i,b_i}^2(\mathbf{D}_{i,b_i} + \widetilde{\sigma}_i^2\mathbf{I})\}$. Applying Lemma 1 to each of the rows of $\mathbf{C}_{k,b_i}^H\boldsymbol{\Gamma}_i^{-1}$, then Lemma 4 and 6, we obtain the following simplified expression,

$$
\begin{aligned}
&p_i \mathbf{g}''^{\,H}_i \mathbf{h}_{k,b_i}\mathbf{h}_{k,b_i}^H\mathbf{g}''_i = p_i\frac{1}{M_{b_i}}\mathrm{tr}\{\mathbf{D}_{k,b_i}\mathbf{B}_{k,b_i}^{-2}\},\\
&\text{where, } \mathbf{B}_{k,b_i} = \mathrm{diag}(1+\beta_k\zeta_{k,b_i}^{(1)}e_{b_i}, ..., 1+\beta_k\zeta_{k,b_i}^{(L_{k,b_i})}e_{b_i}).
\end{aligned}
\tag{25}
$$

Finally, we write the SINR as,

$$
\begin{aligned}
&\gamma_{k,L}^{(Opt)} = \frac{p_k(1-x_{b_k}^{(L)})\mathrm{tr}\{\mathbf{D}_{k,b_k}(\mathbf{I}+\widetilde{\sigma}_k^2\mathbf{D}_{k,b_k}^{-1})^{-1}\}}{\frac{1}{M_{b_i}}\sum\limits_{i\neq k}p_i\mathrm{tr}\{\mathbf{D}_{k,b_i}\mathbf{B}_{k,b_i}^{-2}\}+1},\\
&\gamma_{k,S}^{(Opt)} = \frac{p_k(1-x_{b_k}^{(S)})\frac{\mathrm{tr}\{\mathbf{D}_{k,b_k}\}^2}{\mathrm{tr}\{\mathbf{D}_{k,b_k}+\widetilde{\sigma}_k^2\mathbf{I}\}}}{\frac{1}{M_{b_i}}\sum\limits_{i\neq k}p_i\mathrm{tr}\{\mathbf{D}_{k,b_i}\mathbf{B}_{k,b_i}^{-2}\}+1}.
\end{aligned}
\tag{26}
$$

Note that $x_{b_k}^{(S)} > x_{b_k}^{(L)}$ since the eigen values of $\mathbf{W}_{k,b_k}$ for the subspace estimator is greater than that of the LMMSE channel estimator. This leads to a reduction in signal power for the subspace channel estimator case at low to mid SNR range (while the interference power remains approximately the same for both) which explains the sub-optimal performance of subspace channel estimators. Further considering the simplifications at high SNR, we observe that e_c increases with SNR since μ_c converges to zero with high SNR and $e_c \gg 1$. So $\beta_i \zeta_{i,c}^{(r)} e_c >> 1$, thus we simplify,

$$e_c = (\frac{1}{e_c} \sum_{i=1}^{K} \sum_{r=1}^{L_{i,c}} \frac{1}{M_c} + \mu_c)^{-1} \implies e_c = \frac{1 - \frac{\sum_{i=1}^{K} L_{i,c}}{M_c}}{\mu_c},$$
$$\text{Similarly, } x_{b_k} = \frac{1}{M_{b_k}} \sum_{i=1}^{K} L_{i,b_k} = x_{b_k}^{(S)} = x_{b_k}^{(L)}. \tag{27}$$

Note that in the case when $\frac{K \sum_i L_{i,c}}{M} \to 0$, $e_c \to \mu_c^{-1}$ or $e_c \to SNR^{-1}$. Finally we show that at very high SNR, in the large antenna limit, all the interference powers converge to zero or in other words LMMSE or subspace estimator does ZF at high SNR. We also remark that depending on the large system regime and behaviour of channel estimation error with SNR, ESIP-WSR BF with LMMSE/Subspace channel estimator ZFs to either the interfering covariance subspaces or the actual channel estimates of the interfering users, which is left as a future work.

$$\frac{1}{M_{b_i}} \text{tr}\{\mathbf{D}_{k,b_i} \mathbf{B}_{k,b_i}^{-2}\} = \frac{1}{M_{b_i}} [\sum_{r=1}^{L_{k,b_i}} \frac{\eta_{k,b_i}^{(r)}}{(1+\beta_i \zeta_{k,b_i}^{(r)} e_{b_i})^2}], \tag{28}$$

As $\mu_c \to 0$, it is clear from (28) that, $\frac{1}{M_{b_i}} \text{tr}\{\mathbf{D}_{k,b_i} \mathbf{B}_{k,b_i}^{-2}\} \to 0$, since each of the summation term becomes proportional to $1/SNR$ or $\frac{1}{M_{b_i}} \text{tr}\{\mathbf{D}_{k,b_i} \mathbf{B}_{k,b_i}^{-2}\} \to \frac{K \sum_i L_{k,b_i}}{M} \frac{1}{SNR}$. Finally we can write $r_{\overline{k}}$ in the high SNR regime as below,

$$p_i \mathbf{g}_i^H \mathbf{h}_{k,b_i} \mathbf{h}_{k,b_i}^H \mathbf{g}_i = 0, \ \overline{r}_{\overline{k}} = 1. \tag{29}$$

B. Naive EWSR BF with LMMSE/Subspace Channel Estimators

We denote superscript (N) to denote the SINRs for the case naive BFs (eg. $\gamma_{k,L}^{(N)}$). Also, $x_{b_k}^{(N_L)}, x_{b_K}^{(N_S)}$ represents x_{b_k} for the LMMSE and Subspace channel estimates respectively. We split $\mathbf{h} = \widehat{\mathbf{h}} + \widetilde{\mathbf{h}}$ or $\widehat{\mathbf{d}}_U = \mathbf{d} + \widetilde{\mathbf{d}}_U, \widehat{\mathbf{d}}_U = \boldsymbol{U}\widehat{\mathbf{d}}, \widetilde{\mathbf{d}}_U = (\boldsymbol{U} - \mathbf{I})\mathbf{d} + \boldsymbol{U}\widetilde{\mathbf{d}}$ and refer to the received signal model (11). For naive BF, we redefine $\boldsymbol{\Gamma}_k = \sum_{i \neq k} \beta_i \mathbf{C}_{i,b_k} \boldsymbol{U}_{i,b_k} \widehat{\mathbf{d}}_{i,b_k} \boldsymbol{U}_{i,b_k} \widehat{\mathbf{d}}_{i,b_k}^H \mathbf{C}_{i,b_k}^H + \mu_{b_k} \mathbf{I}$. First we compute the signal power,

$$\begin{aligned} & P_{S_k} = p_k \mathbf{g}_k'^H \mathbf{C}_{k,b_k} \widehat{\mathbf{d}}_{U,k,b_k} \widehat{\mathbf{d}}_{U,k,b_k}^H \mathbf{C}_{k,b_k}^H \mathbf{g}_k', \\ & \mathbf{g}_k''^H \mathbf{C}_{k,b_k} \widehat{\mathbf{d}}_{U,k,b_k} = \widehat{\mathbf{d}}_{k,b_k}^H \boldsymbol{U}_{k,b_k} \mathbf{C}_{k,b_k}^H \boldsymbol{\Gamma}_k^{-1} \mathbf{C}_{k,b_k} \boldsymbol{U}_{k,b_k} \widehat{\mathbf{d}}_{k,b_k}, \\ & = e_{b_k} \widehat{\mathbf{d}}_{k,b_k}^H \boldsymbol{U}_{k,b_k}^2 \widehat{\mathbf{d}}_{k,b_k} = e_{b_k} \text{E}(\text{tr}\{\boldsymbol{U}_{k,b_k}^2 \widehat{\mathbf{d}}_{k,b_k} \widehat{\mathbf{d}}_{k,b_k}^H\}) = \\ & e_{b_k} \text{tr}\{\boldsymbol{U}_{k,b_k}^2 (\mathbf{D}_{k,b_k} + \widetilde{\sigma}_k^2 \mathbf{I})\}, \end{aligned} \tag{30}$$

Finally, we write the signal power as,

$$P_{S_k} = p_k (1 - x_{b_k}^{(N_L)}) \text{tr}\{\boldsymbol{U}_{k,b_k}^2 (\mathbf{D}_{k,b_k} + \widetilde{\sigma}_k^2 \mathbf{I})\}. \tag{31}$$

Now consider the interference power P_{I_k}. First we consider the interference power due to the channel estimation error part of the direct channel for user k.

$$\begin{aligned} & \mathbf{g}_k'^H \mathbf{C}_{k,b_k} \widetilde{\mathbf{d}}_{U,k,b_k} \widetilde{\mathbf{d}}_{U,k,b_k}^H \mathbf{C}_{k,b_k}^H \mathbf{g}_k' \xrightarrow[a.s]{M \to \infty} \\ & \text{E}(\mathbf{g}_k'^H \mathbf{C}_{k,b_k} \widetilde{\mathbf{d}}_{U,k,b_k} \widetilde{\mathbf{d}}_{U,k,b_k}^H \mathbf{C}_{k,b_k}^H \mathbf{g}_k') \xrightarrow[a.s]{M \to \infty} = \\ & e_{b_k}^2 \text{tr}\{\boldsymbol{U}_{k,b_k} (\boldsymbol{U}_{k,b_k} - \mathbf{I}) \mathbf{D} + \boldsymbol{U}_{k,b_k} \boldsymbol{U}_{k,b_k} \widetilde{\sigma}^2\}^2. \end{aligned} \tag{32}$$

Therefore $p_k \mathbf{g}_k'^H \mathbf{C}_{k,b_k} \widetilde{\mathbf{d}}_{U,k,b_k} \widetilde{\mathbf{d}}_{U,k,b_k}^H \mathbf{C}_{k,b_k}^H \mathbf{g}_k' = p_k (1 - x_{b_k}^{(N_L)}) \text{tr}\{\boldsymbol{U}_{k,b_k} (\boldsymbol{U}_{k,b_k} - \mathbf{I}) \mathbf{D} + \boldsymbol{U}_{k,b_k} \boldsymbol{U}_{k,b_k} \widetilde{\sigma}^2\}^2 = E_{I_k}$. Expressions for the interfering powers due to rest of the user's channel remains of the same form as in the previous section for optimal EWSR BFs. Finally, we can write the SINR expressions for the optimal EWSR naive BFs with LMMSE and subspace channel estimator becomes,

$$\gamma_{k,L}^{(N)} = \frac{p_k (1 - x_{b_k}^{(N_L)}) \text{tr}\{\boldsymbol{U}_{k,b_k} (\mathbf{D}_{k,b_k} + \widetilde{\sigma}_k^2 \mathbf{I})\}}{\frac{1}{M_{b_i}} \sum_{i \neq k} p_i \text{tr}\{\mathbf{D}_{k,b_i} \mathbf{B}_{k,b_i}^{-2}\} + \frac{E_{I_k}}{\text{tr}\{\boldsymbol{U}_{k,b_k}^2 (\mathbf{D}_{k,b_k} + \widetilde{\sigma}^2 \mathbf{I})\}} + 1}. \tag{33}$$

In (33), compared to the optimal EWSR estimators, the interference power got increased by the direct channel estimation error being moved to the interference part, which explains the degradation in performance.

C. EWSMSE BF with LMMSE/Subspace Channel Estimators

We denote superscript (E) to denote the SINRs for the case of EWSMSE BFs (eg. $\gamma_{k,L}^{(E)}$). Also, $x_{b_k}^{(E_L)}, x_{b_K}^{(E_S)}$ represents x_{b_k} for the respective channel estimates. For the EWSMSE BFs, since we consider the desired user channel estimation error as part of the interference terms, the signal model remains same as in (11). We need to redefine $\boldsymbol{\Gamma}_k = \sum_{i \neq k} \beta_i \mathbf{C}_{i,b_k} \mathbf{W}_{i,b_k} \mathbf{C}_{i,b_k}^H + \alpha_k \mathbf{C}_{i,b_k} \widetilde{\mathbf{D}}_{i,b_k} \mathbf{C}_{i,b_k}^H + \mu_{b_k} \mathbf{I}$.The large system analysis will have similar expression as for the naive BFs. However, since we take into account the error covariance information also into the BF optimization, $x_{b_k}^{(E_L)}, x_{b_k}^{(E_S)}$ will be different for EWSMSE compared to the naive BF due to the change in the eigen value matrix $\boldsymbol{\Lambda}_{k,b_i}$.

D. Optimal EWSR BF with LS only Channel Estimate

For LS only estimation, optimizing $EWSR(\mathbf{g})$ leads to the following generalized eigen value problem,

$$\begin{aligned} & \nu_k (\sum_{i \neq k} \mathbf{S}_{i,b_k} + \mu_{b_k} \mathbf{I}) \mathbf{g}_k = (\widehat{\mathbf{h}}_{k,b_k,LS} \widehat{\mathbf{h}}_{k,b_k,LS}^H + \widetilde{\sigma}_k^2 \mathbf{I}) \mathbf{g}_k, \\ & \widehat{\mathbf{h}}_{k,b_k,LS} \widehat{\mathbf{h}}_{k,b_k,LS}^H \mathbf{g}_k = \nu_k (\sum_{i \neq k} \mathbf{S}_{i,b_k} + (\mu_{b_k} - \frac{\widetilde{\sigma}_k^2}{\nu_k}) \mathbf{I}) \mathbf{g}_k. \end{aligned} \tag{34}$$

where $\mathbf{S}_{i,b_k} = \widehat{\mathbf{h}}_{i,b_k,LS} \widehat{\mathbf{h}}_{i,b_k,LS}^H + \widetilde{\sigma}_i^2 \mathbf{I}$ here. This leads to $\mathbf{g}_k \propto (\sum_{i \neq k} \mathbf{S}_{i,b_k} + (\mu_{b_k} - \frac{\widetilde{\sigma}_k^2}{\nu_k}) \mathbf{I})^{-1} \widehat{\mathbf{h}}_{k,b_k,LS}$. To find ν_k, we multiply by $\mathbf{g}_k$ on both sides of (34) and obtain ν_k,

$$\begin{aligned} & \mathbf{g}_k^H \widehat{\mathbf{h}}_{k,b_k,LS} \widehat{\mathbf{h}}_{k,b_k,LS}^H \mathbf{g}_k = \nu_k \mathbf{g}_k^H (\sum_{i \neq k} \mathbf{S}_{i,b_k} + (\mu_{b_k} - \frac{\widetilde{\sigma}_k^2}{\nu_k}) \mathbf{I}) \mathbf{g}_k, \\ & \text{So, } \nu_k = \widehat{\mathbf{h}}_{k,b_k,LS}^H (\sum_{i \neq k} \mathbf{S}_{i,b_k} + (\mu_{b_k} - \frac{\widetilde{\sigma}_k^2}{\nu_k}) \mathbf{I})^{-1} \widehat{\mathbf{h}}_{k,b_k,LS} \end{aligned} \tag{35}$$

First we compute the deterministic equivalent for ν_k. We define $\boldsymbol{\Gamma}_k = \sum_{i \neq k} \mathbf{S}_{i,b_k} + (\mu_{b_k} - \frac{\widetilde{\sigma}_k^2}{\nu_k}) \mathbf{I}$. In the large system limit, $\nu_k \xrightarrow[a.s]{M \to \infty} \text{E}(\text{tr}\{\widehat{\mathbf{h}}_{k,b_k,LS}^H \boldsymbol{\Gamma}_k^{-1} \widehat{\mathbf{h}}_{k,b_k,LS}\}) =$

$\frac{1}{M_{b_k}}\text{tr}\{\mathbf{\Gamma}_k^{-1}\}\text{tr}\{\mathbf{D}_{k,b_k}\} + \widetilde{\sigma}_k^2\text{E}(\text{tr}\{\mathbf{\Gamma}_k^{-1}\}) = \text{tr}\{\mathbf{D}_{k,b_k}\}e_{b_k} + \widetilde{\sigma}_k^2 M_{b_k}e_{b_k}$, where e_c is defined as,

$$e_c = (\frac{1}{M_c}\sum_{i=1}^{K}\sum_{r=1}^{L_{i,c}}\frac{\beta_i\zeta_{i,c}^{(r)}}{1+\beta_i\zeta_{i,c}^{(r)}e_c} + \mu_c - \frac{\widetilde{\sigma}_k^2}{\nu_k})^{-1}. \quad (36)$$

Now $\mathbf{g}_k'' = \mathbf{\Gamma}_k^{-1}\widehat{\mathbf{h}}_{k,b_k,LS}$, considering the signal part and substituting for $\widehat{\mathbf{h}}_{k,b_k,LS} = \mathbf{h}_{k,b_k} + \widetilde{\mathbf{h}}_{k,b_k}$ and using the fact that $\mathbf{h}_{k,b_k}$ and $\widetilde{\mathbf{h}}_{k,b_k}$ are independent,

$$\begin{aligned}&\mathbf{g}_k''^H\mathbf{h}_{k,b_k} = \\ &\widehat{\mathbf{h}}_{k,b_k,LS}^H\mathbf{\Gamma}_k^{-1}\mathbf{h}_{k,b_k} \xrightarrow[a.s]{M\to\infty} \text{E}(\text{tr}\{\mathbf{C}_{k,b_k}^H\mathbf{\Gamma}_k^{-1}\mathbf{C}_{k,b_k}\mathbf{d}_{k,b_k}\mathbf{d}_{k,b_k}^H\}) \\ &= e_{b_k}\text{tr}\{\mathbf{D}_{k,b_k}\}, \;\; \mathbf{g}_k''^H\mathbf{h}_{k,b_k}\mathbf{h}_{k,b_k}\mathbf{g}_k'' = e_{b_k}^2\text{tr}\{\mathbf{D}_{k,b_k}\}^2.\end{aligned} \quad (37)$$

Further, $\|\mathbf{g}_k''\|^2 = \widehat{\mathbf{h}}_{k,b_k,LS}^H\mathbf{\Gamma}_k^{-2}\widehat{\mathbf{h}}_{k,b_k,LS} \xrightarrow[a.s]{M\to\infty} \text{tr}\{\mathbf{\Gamma}_k^{-2}(\mathbf{C}_{k,b_k}\mathbf{D}_{k,b_k}\mathbf{C}_{k,b_k}^H + \widetilde{\sigma}_k^2\mathbf{I})\} = e_{b_k}'\text{tr}\{\mathbf{D}_{k,b_k}\} + \widetilde{\sigma}_k^2 M_{b_k}e_{b_k}'$. Finally we obtain the deterministic equivalent of the signal power as,

$$P_{S_k} = (1 - x_{b_k}^{(LS)})\frac{\text{tr}\{\mathbf{D}_{k,b_k}\}^2}{\text{tr}\{\mathbf{D}_{k,b_k}\}+\widetilde{\sigma}_k^2 M_{b_k}}. \quad (38)$$

Note that $x_{b_k}^{(LS)}$ has the same definition as in (24), but with the eigenvalues, $\zeta_{k,b_i}^{(1)} = \text{tr}\{\mathbf{D}_{k,b_i} + \widetilde{\sigma}_k^2\mathbf{I}\} + \widetilde{\sigma}_k^2, \zeta_{k,b_i}^{(r)} = \widetilde{\sigma}_k^2, \forall r = 2, ..., L_{k,b_i}$. Further considering the interfering user channel powers,

$$\begin{aligned}&\mathbf{g}_i''^H\mathbf{C}_{k,b_i}\mathbf{d}_{k,b_i}\mathbf{d}_{k,b_i}^H\mathbf{C}_{k,b_i}^H\mathbf{g}_i'' = \\ &\widehat{\mathbf{h}}_{i,b_i,LS}^H\mathbf{\Gamma}_i^{-1}\mathbf{C}_{k,b_i}\mathbf{d}_{k,b_i}\mathbf{d}_{k,b_i}^H\mathbf{C}_{k,b_i}^H\mathbf{\Gamma}_i^{-1}\widehat{\mathbf{h}}_{i,b_i,LS} = \\ &\text{tr}\{\mathbf{D}_{i,b_i}\}\frac{1}{M_{b_i}}\text{tr}\{\mathbf{\Gamma}_i^{-1}\mathbf{C}_{k,b_i}\mathbf{d}_{k,b_i}\mathbf{d}_{k,b_i}^H\mathbf{C}_{k,b_i}^H\mathbf{\Gamma}_i^{-1}\} + \\ &\widetilde{\sigma}_i^2\frac{L_{i,b_i}}{M_{b_i}}\text{tr}\{\mathbf{\Gamma}_i^{-1}\mathbf{C}_{k,b_i}\mathbf{d}_{k,b_i}\mathbf{d}_{k,b_i}^H\mathbf{C}_{k,b_i}^H\mathbf{\Gamma}_i^{-1}\} = \\ &\text{tr}\{\mathbf{D}_{i,b_i}\}\frac{1}{M_{b_i}}e_{b_i}'\text{tr}\{\mathbf{B}_{k,b_i}^{-2}\mathbf{D}_{k,b_i}\} + e_{b_i}'\widetilde{\sigma}_i^2\frac{L_{i,b_i}}{M_{b_i}}\text{tr}\{\mathbf{B}_{k,b_i}^{-2}\mathbf{D}_{k,b_i}\}\end{aligned} \quad (39)$$

Finally we write the interference power as,

$$P_{I_k} = \frac{1}{M_{b_i}}\sum_{i\neq k}[\text{tr}\{\mathbf{B}_{k,b_i}^{-2}\mathbf{D}_{k,b_i}\} + L_{i,b_i}\widetilde{\sigma}_i^2\text{tr}\{\mathbf{B}_{k,b_i}^{-2}\mathbf{D}_{k,b_i}\}] \quad (40)$$

Finally we write the SINR expression as,

$$\gamma_{k,LS}^{(Opt)} = \frac{(1-x_{b_k}^{(LS)})\frac{\text{tr}\{\mathbf{D}_{k,b_k}\}^2}{\text{tr}\{\mathbf{D}_{k,b_k}\}+\widetilde{\sigma}_k^2 M_{b_k}}}{\frac{1}{M_{b_i}}\sum_{i\neq k}p_i[\text{tr}\{\mathbf{B}_{k,b_i}^{-2}\mathbf{D}_{k,b_i}\}+L_{i,b_i}\widetilde{\sigma}_i^2\text{tr}\{\mathbf{B}_{k,b_i}^{-2}\mathbf{D}_{k,b_i}\}]+1} \quad (41)$$

From (41), it can be observed that with LS only channel estimator, signal power gets decreased by a factor $\text{tr}\{\mathbf{D}_{k,b_k}\} + \widetilde{\sigma}_k^2 M_{b_k}$ and similarly the interference power gets added by $\frac{L_{i,b_i}}{M_{b_i}}\widetilde{\sigma}_i^2\text{tr}\{\mathbf{B}_{k,b_i}^{-2}\mathbf{D}_{k,b_i}\}$. This leads to the sub-optimal performance compared to BFs with LMMSE or Subspace channel estimators. Now we consider the naive BF for the LS only channel estimator. Here we need to consider the interference power due to the desired channel estimation error part,

$$\begin{aligned}&\mathbf{g}_k'^H\widetilde{\mathbf{h}}_{k,b_k} \xrightarrow[a.s]{M\to\infty} \text{E}(\widehat{\mathbf{h}}_{k,b_k,LS}^H\mathbf{\Gamma}_k^{-1}\widetilde{\mathbf{h}}_{k,b_k}) = \\ &\text{E}(\widetilde{\mathbf{h}}_{k,b_k,LS}^H\mathbf{\Gamma}_k^{-1}\widetilde{\mathbf{h}}_{k,b_k}) = e_{b_k}M_{k,b_k}\widetilde{\sigma}^2, \\ &\text{So, } \mathbf{g}_k'^H\widetilde{\mathbf{h}}_{k,b_k}\widetilde{\mathbf{h}}_{k,b_k}^H\mathbf{g}_k'^H = \frac{(1-x_{b_k}^{(LS)})M_{k,b_k}^2\widetilde{\sigma}^4}{\text{tr}\{\mathbf{D}_{k,b_k}+\widetilde{\sigma}^2\mathbf{I}\}}.\end{aligned} \quad (42)$$

Finally we write the SINR expression for the naive BF as,

$$\gamma_{k,LS}^{(N)} = \frac{(1-x_{b_k}^{(NLS)})\frac{\text{tr}\{\mathbf{D}_{k,b_k}\}^2}{\text{tr}\{\mathbf{D}_{k,b_k}\}+\widetilde{\sigma}_k^2 M_{b_k}}}{\frac{1}{M_{b_i}}\sum_{i\neq k}p_i[\text{tr}\{\mathbf{B}_{k,b_i}^{-2}\mathbf{D}_{k,b_i}\}+L_{i,b_i}\widetilde{\sigma}_i^2\text{tr}\{\mathbf{B}_{k,b_i}^{-2}\mathbf{D}_{k,b_i}\}]+\frac{p_k(1-x_{b_k}^{(NLS)})M_{k,b_k}^2\widetilde{\sigma}^4}{\text{tr}\{\mathbf{D}_{k,b_k}+\widetilde{\sigma}^2\mathbf{I}\}}+1} \quad (43)$$

Note that $x_{b_k}^{(NLS)}$ has the same definition as in (24), but with the eigenvalues, $\zeta_{k,b_i}^{(1)} = \text{tr}\{\mathbf{D}_{k,b_i}\}, \zeta_{k,b_i}^{(r)} = 0, \forall r = 2, ..., L_{k,b_i}$.

ACKNOWLEDGEMENTS

EURECOM's research is partially supported by its industrial members: ORANGE, BMW, Symantec, SAP, Monaco Telecom, iABG, and by the projects MASS-START (French FUI) and DUPLEX (French ANR).

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] M. Medard, "The effect upon channel capacity in wireless communications of perfect and imperfect knowledge of the channel," *IEEE Trans. on Info. Theo.*, vol. 46, no. 3, 2000.

[3] C. K. Thomas and Dirk Slock, "Massive MISO IBC Beamforming - a Multi-Antenna Stochastic Geometry Perspective," in *IEEE Globecom Wkshps. on Emerg. Techn. for 5G and Bey. Wire. and Mob. Netw.*, Abu Dhabi, UAE, 2018.

[4] C. K. Thomas and Dirk Slock, "Reduced-order zero-forcing beamforming vs optimal beamforming and dirty Paper coding and massive MIMO Analysis," in *10th IEEE Sens. Arr. and Mul.chnl. Sig. Process. Wkshp.*, Sheffield, UK, 2018.

[5] C. K. Thomas and Dirk Slock, "Massive MISO IBC reduced order zero forcing beamforming - a multi-antenna stochastic geometry perspective," in *Intl. Conf. on Comp., Netwrk. and Commun. (ICNC)*, Honolulu, Hawaii, USA, 2019.

[6] Martin Haenggi, Jeffrey G. Andrews, , François Baccelli, Olivier Dousse, and Massimo Franceschetti, "Stochastic Geometry and Random Graphs for the Analysis and Design of Wireless Networks ," *IEEE Jrnl. on Sel. Areas in Commun.*, Sept. 2009.

[7] G. George, R. K. Mungara, A. Lozano, and M. Haenggi, "Ergodic Spectral Efficiency in MIMO Cellular Networks," *IEEE Trans. on Wireless Communications*, May 2017.

[8] Z. Chen and Emil Bjrnson, "Channel hardening and favorable propagation in cell-free massive MIMO with stochastic geometry," *IEEE Trans. on Comm.*, vol. 66, no. 11, 2018.

[9] A. Adhikary, J. Nam, J. Y. Ahn, and G. Caire, "Joint spatial division and multiplexing: the large-scale array regime," *IEEE Trans. on Info. Theo.*, vol. 59, no. 10, 2013.

[10] R. B. Ertel, P. Cardieri, K. W. Sowerby, T. S. Rappaport, and J. H. Reed, "Overview of spatial channel models for antenna array communication systems ," *IEEE Pers. Commun.*, Feb. 1998.

[11] G. Caire and K. R. Kumar, "Information theoretic foundations of adaptive coded modulation," *Proc. IEEE*, vol. 95, no. 12, Dec. 2007.

[12] H. Joudeh and B. Clerckx, "Sum-rate maximization for linearly precoded downlink multiuser MISO systems with partial CSIT: A rate-splitting approach," *IEEE Trans. on Commun.*, vol. 64, no. 11, November 2016.

[13] Kalyana Gopala and Dirk Slock, "A refined analysis of the gap between expected rate for partial CSIT and the massive MIMO rate limit," in *IEEE Proc. IEEE Int'l Conf. Acoustics Speech and Sig. Proc. (ICASSP)*, Calgary, Canada, 2018.

[14] S. J. Kim and G. B. Giannakis, "Optimal resource allocation for MIMO ad-hoc cognitive radio networks," in *IEEE Trans. on Info. Theory*, May 2011, vol. 57.

[15] F. Negro, I. Ghauri, and Dirk T. M. Slock, "Sum rate maximization in the noisy MIMO interfering broadcast channel with partial CSIT via the expected weighted MSE," in *ISWCS, Paris, France*, 2012.

[16] T.L. Marzetta and E.G. Larsson and H. Yang and H. Quoc Ngo, *Fundamentals of Massive MIMO*, Cambridge U. Press, 2016.

[17] L. Liu, C. Oestges, J. Poutanen, K. Haneda, P. Vainikainen, F. Quitin, F. Tufvesson, and P. De Doncker, "The COST 2100 MIMO channel model ," *IEEE Wire. Commun.*, 2012.

[18] Haifan Yin, David Gesbert, Miltiades Filippou, and Yingzhuang Liu, "A coordinated approach to channel estimation in large-scale multiple-antenna systems ," *IEEE Jrnl. on Sel. Areas in Commun.*, Feb. 2013.

[19] S. Wagner, R. Couillet, M. Debbah, and D. T. M. Slock, "Large system analysis of linear precoding in MISO broadcast channels with limited feedback," *IEEE Trans. Inf. Theory*, vol. 58, no. 7, July 2012.

[20] S. S. Christensen, R. Agarwal, E. de Carvalho, and J. Cioffi, "Weighted sum-rate maximization using weighted MMSE for MIMO-BC beamforming design," *IEEE Trans. on Wireless Commun.*, December 2008.