IEEE TRANSACTIONS ON
WIRELESS
COMMUNICATIONS

FEC Code Anchored Robust Design of Massive MIMO Receivers

Kun Wang, *Student Member, IEEE,* and Zhi Ding, *Fellow, IEEE*

***Abstract*—Massive multiple-input-multiple-output (MIMO) systems have been proposed to support high rate multiple access. As channel estimation in massive MIMO suffers from the well-known impairment of pilot contamination, we propose a novel approach to multi-user detection by exploiting forward error correction (FEC) code diversity. Unlike traditional approaches solely based on worst-case or probabilistic channel estimation errors, we develop a joint quadratic-programming (QP) receiver anchored with a set of FEC code constraints. Exploiting the user signatures presented by FEC channel codes of distinct permutations, our receiver can effectively recover signals from pilot-interfering users. The code-anchored robust design (CARD) method can also be applied to a chance-constrained receiver, which shows further performance gain compared with the direct integration of FEC code constraints in joint QP receiver. The effectiveness of CARD receivers is demonstrated by numerical results that establish substantial performance gain of the proposed receivers over existing robust designs. In addition, we present a distributed multi-cell processing scheme for enhanced performance via alternating direction method of multipliers (ADMM).**

***Index Terms*—Code-anchored robust design (CARD), quadratic programming, second-order cone programming, multi-user detection, massive MIMO, pilot contamination, FEC code signature, multi-cell processing, ADMM.**

I. Introduction

Leveraging the extensive success of multiple-input multiple-output (MIMO) technologies over the past decade in WiFi 802.11 and 3GPP LTE, massive MIMO has generated substantial interest and excitement in wireless communications. In massive MIMO, large number of antennas at base station (BS) can potentially accommodate many more mobile users than ever before. Asymptotic analysis establishes that very large BS antenna array may potentially offer orders of magnitude improvement in spectral efficiency and capacity [1]. For large transmit antenna array, random matrix theory demonstrates advantages against small-scale fading and thermal noise effects [2]. Overall, massive MIMO promises reduced latency, simplified MAC layer, robustness against jamming, and inexpensive low-power components [3].

The many advantages promised by massive MIMO may be limited by key physical factors. Chief among them is the problem *pilot contamination* due to non-orthogonal pilot signals used for channel estimation across different cells. More specifically, obtaining channel state information (CSI) requires sufficient number of training symbols which depends on the number of users served. Applying orthogonal pilots for a large number of users would compromise throughput efficiency. Hence, training sequences may overlap in time-frequency resource blocks and are non-orthogonal. Channel estimates can now be degraded due to training signals from neighboring cells; this is known as pilot contamination [4].

Various research works have been proposed to mitigate pilot contamination. One protocol-based method in [5] proposes to time-shift uplink and downlink transmissions. In [6], the authors carefully designed an amalgam for downlink training and "scheduled" uplink training. Another solution by [7] uses subspace-based techniques for blind pilot decontamination. Nonetheless, these aforementioned approaches cannot completely mitigate pilot contamination in practice, since they either require rather stringent conditions, such as the number of antennas versus mobile users, or may be too costly. As a result, practical massive MIMO systems have to still work under CSI errors. In this work, our goal is to develop a robust receiver in a multi-user massive MIMO system against the CSI mismatch caused by pilot contamination.

For multi-user massive MIMO systems, much insight can be acquired from the studies of code division multiple access (CDMA) because of shared similarities. Analogous to the spreading sequences used in CDMA systems, Rayleigh-fading channel vectors with i.i.d. entries in massive MIMO systems are also user-specific signatures (asymptotically orthogonal) [2], [8]. In this work, although we do not make use of the orthogonality property, we re-apply a well-known receiver design principle that also originates from CDMA for massive MIMO, namely, the *minimum output energy* (MOE) criterion [9] in order to develop a robust receiver that can reject multiple access interferences. It is widely acknowledged that MOE criterion is effective in interference suppression [10].

The MOE principle was first proposed in [9] as a blind approach to signal recovery and later used in CDMA [10] for co-channel interference rejection even if the spreading codes of interfering users are unknown. Additional code information of interfering users can help enhance performance [11]. To combat CDMA code signature mismatch, the authors of [12] devised a robust blind multi-user detector based on second-order cone programming (SOCP). In recent years, MOE has revived in multi-access MIMO systems. Adopting MOE under the assumption of ideal CSI, a linear minimum variance (MV) receiver for space-time coded signals was presented in [13]. Given imperfect CSI, a plurality of robust receivers

This work was presented in part at the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP), Shanghai, China, March 2016.

K. Wang and Z. Ding are with the Department of Electrical and Computer Engineering, University of California, Davis, CA 95616, USA. E-mail: {kunwang, zding}@ucdavis.edu.

This material is based on works supported by the National Science Foundation Grants CNS-1443870, ECCS-1307820, and CIF-1321143.

Digital Object Identifier: 10.1109/TWC.2016.2613516

were proposed [14], including diagonally loaded minimum variance (DLMV) receiver [13], worst-case robust receiver [15] and chance-constrained robust receiver [16]. Generally, these MOE-based receivers can handle reasonably moderate channel estimate errors, their performance tend to be poor under severely impaired CSI caused by pilot contamination. To more effectively combat pilot contamination, additional user signal characteristics may and should be exploited. One of the most striking characteristics left under-utilized is the forward error correction (FEC) codes of mobile transmitters.

FEC codes are effective and nearly universal in modern wireless systems against demodulation errors. Among the most popular FEC codes are Reed Solomon, turbo, convolutional, and low density parity check (LDPC) codes. In particular, LDPC codes proposed in [17] have recently regained widespread popularity since the works by [18] and have demonstrated near capacity performance achieved by soft-decoding algorithms of moderate complexity. The low density of their parity check matrices also delivers another benefit, i.e., the possibility of transforming their code constraints that are linear only in Galois field (GF) to a set of relaxed linear constraints in real field [19]. This transformation of LDPC code constraints makes them amenable to an integrated receiver algorithm that can achieve good performance. Our previous works have demonstrated the powerful joint receivers integrated with code information [20]–[23]. In this work, we propose a new idea by exploiting FEC codes and distinct permutations as unique user signatures in massive MIMO uplink receiver to combat the pilot contamination effect for improved signal recovery [24].

Our novel contribution consists of the reformulation of the MV receiver [13] into an unconstrained quadratic program, while relaxing a strict response-preserving constraint. We then incorporate a set of LDPC equality code constraints [19] to robustify the quadratic programming (QP) receiver. Recognizing that multiple interfering users in practice may be given the same LDPC code for error protection, we shall assign a unique codeword permutation pattern despite a common LDPC code shared by pilot contaminating users. Our code-anchored robust design (CARD) receiver is subsequently able to utilize the user code permutation as a unique signature, thereby achieving much better robustness against channel estimation error over existing receivers. Moreover, when modeling CSI error probabilistically, we apply the CARD concept to the SOCP receiver of [16], providing additional performance enhancement. For distributed BS, we further make use of ADMM algorithm [25] to present a cooperative reception scheme within a cluster of receiving base stations. This collaborative receiver algorithm can distribute major computations to each participating BS with limited information exchange among them. ADMM provides greater flexibility, scalability, and reliability to receiver designs.

We organize the manuscript as follows. First, Section II describes the uplink transmissions in multi-user systems and mathematically models the impact of pilot contamination on channel estimates. We next develop our CARD receivers in Section III by firstly reformulating MOE criterion before integrating channel code as user signature. We further enhance the CARD receiver performance through probabilistic modeling in Section IV. In Section V, we elaborate on the ADMM procedures for single-BS reception followed by distributed cooperative reception in a cluster of BS's. Performance and complexity analyses of the proposed CARD receivers are provided in Section VI. Section VII numerically presents the performance comparisons of the proposed joint QP and joint SOCP receivers against the DLMV and chance-constrained receivers. We also show the design trade-off for clustered reception, before concluding with Section VIII.

II. Pilot-contaminated Massive MIMO System

A. Notations

Throughout this manuscript, $\{\cdot\}^T$ denotes matrix transpose and $\{\cdot\}^H$ denotes conjugate transpose. $\mathrm{tr}\{\cdot\}$ stands for matrix trace. $\mathrm{Re}\{\cdot\}$ and $\mathrm{Im}\{\cdot\}$ denote real and imaginary part of a complex number, respectively. $\|\cdot\|$ denotes the ℓ_2-norm of a vector whereas $\mathbb{E}\{\cdot\}$ means expectation over random variables. Following conventions, we use uppercase and lowercase boldface letters for matrices and vectors, respectively. $\mathbf{A} \otimes \mathbf{B}$ denotes the Kronecker product of matrices $\mathbf{A}$ and $\mathbf{B}$. Notations $\mathbb{R}^{M\times N}$ and $\mathbb{C}^{M\times N}$ represent the set of $M \times N$ matrices over real field and complex field, respectively. $\mathcal{CN}(\mu, \sigma^2)$ denotes circularly symmetric complex Gaussian distribution with mean μ and variance σ^2.

B. Uplink Transmission

In this work, we consider uplink transmissions in a multi-cell multi-user MIMO (MC-MU-MIMO) system. As shown in Fig. 1, the MC-MU-MIMO network consists of L cells, in which the ℓ-th cell BS serves K_ℓ single-antenna mobile users. For simplicity, we suppose that every BS is equipped with N_r antennas. Moreover, uplink channels from mobile users to their BS's are characterized by flat fading. Thus, the channel matrix from all K_ℓ users in the ℓ-th cell to the i-th BS can be represented as

$$\tilde{\mathbf{H}}_{i,\ell} = \begin{bmatrix} \tilde{h}_{1,i,1,\ell} & \dots & \tilde{h}_{1,i,K_\ell,\ell} \\ \vdots & \ddots & \vdots \\ \tilde{h}_{N_r,i,1,\ell} & \dots & \tilde{h}_{N_r,i,K_\ell,\ell} \end{bmatrix}, \tag{1}$$

where $\tilde{h}_{n,i,k,\ell}$ denotes the gain coefficient from the k-th user in the ℓ-th cell to the n-th antenna of the i-th BS. To be specific, $\tilde{h}_{n,i,k,\ell} = \tilde{g}_{n,i,k,\ell}\sqrt{\tilde{d}_{i,k,\ell}}$ consists of small-scale fading coefficient $\tilde{g}_{n,i,k,\ell} \sim \mathcal{CN}(0,1)$ and large-scale attenuation factor $\tilde{d}_{i,k,\ell} \leq 1$. Note that the large-scale fading factor $\tilde{d}_{i,k,\ell}$'s are the same across all antennas at a BS, but are user-dependent [1].

Without loss of generality, we focus on signal reception at one specific BS, e.g., the i-th BS. Thus, we can drop the BS index i in Eq. (1). To further simplify notations, we ignore the cell index ℓ in data transmission equation. We simply let $N_u \triangleq \sum_{\ell=1}^{L} K_\ell$ be the aggregated number of users whose signals can reach this particular BS. Therefore, the received signal vector $\tilde{\mathbf{y}} = [\tilde{y}_1 \dots \tilde{y}_{N_r}]^T \in \mathbb{C}^{N_r\times 1}$ at the BS-of-interest can be expressed as

$$\tilde{\mathbf{y}} = \left[\tilde{\mathbf{H}}_1 \dots \tilde{\mathbf{H}}_L\right]\tilde{\mathbf{x}} + \tilde{\mathbf{n}} = \left[\tilde{\mathbf{h}}_1 \dots \tilde{\mathbf{h}}_{N_u}\right]\tilde{\mathbf{x}} + \tilde{\mathbf{n}}, \tag{2}$$

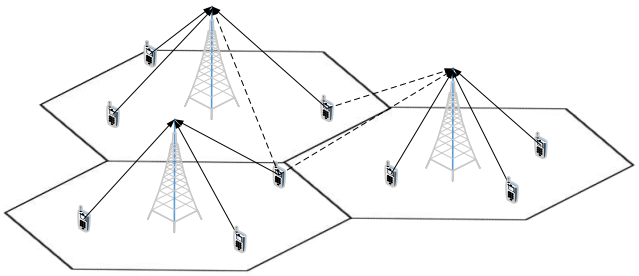


Fig. 1: Uplink transmissions in an MC-MU-MIMO system: solid lines denote intended signals and dashed lines represent interference signals.

where $\tilde{\mathbf{x}} = [\tilde{x}_1 \dots \tilde{x}_{N_u}]^T \in \mathbb{C}^{N_u \times 1}$ is the vector of signals transmitted from all users of interest, $\tilde{\mathbf{n}} = [\tilde{n}_1 \dots \tilde{n}_{N_r}]^T \in \mathbb{C}^{N_r \times 1}$ is a complex Gaussian noise vector $\sim \mathcal{CN}(\mathbf{0}, \sigma_n^2 \mathbf{I})$, and $\tilde{\mathbf{h}}_k$ is the channel vector associated with the k-th user, $1 \leq k \leq N_u$. Note that the data symbols in vector $\tilde{\mathbf{x}}$ are modulated from (possibly permuted) bit streams coded by FEC codes, such as the capacity-approaching LDPC codes.

For notational convenience, we split the real and the imaginary parts of complex symbols as follows $\mathbf{y} = [\text{Re}\{\tilde{\mathbf{y}}\}^T \ \text{Im}\{\tilde{\mathbf{y}}\}^T]^T \in \mathbb{R}^{2N_r \times 1}$, $\mathbf{x} = [\mathbf{x}_1^T \dots \mathbf{x}_{N_u}^T]^T \in \mathbb{R}^{2N_u \times 1}$ where the k-th user's transmitted signal $\mathbf{x}_k = [\text{Re}\{\tilde{x}_k\} \ \text{Im}\{\tilde{x}_k\}]^T \in \mathbb{R}^{2 \times 1}$, $\mathbf{n} = [\text{Re}\{\tilde{\mathbf{n}}\}^T \ \text{Im}\{\tilde{\mathbf{n}}\}^T]^T \in \mathbb{R}^{2N_r \times 1}$, and $\mathbf{H} = [\mathbf{H}_1 \dots \mathbf{H}_{N_u}] \in \mathbb{R}^{2N_r \times 2N_u}$, in which the k-th user's channel matrix is

$$\mathbf{H}_k = \begin{bmatrix} \text{Re}\{\tilde{\mathbf{h}}_k\} & -\text{Im}\{\tilde{\mathbf{h}}_k\} \\ \text{Im}\{\tilde{\mathbf{h}}_k\} & \text{Re}\{\tilde{\mathbf{h}}_k\} \end{bmatrix}. \tag{3}$$

As a result, we reduce Eq. (2) to a real-valued signal transmit-receive equation

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{n}. \tag{4}$$

C. Pilot-contaminated Channel Estimate

Most communication systems nowadays estimate their channel information based on training pilots in each coherence time duration in order for accurate signal detection and decoding. There exist a plurality of schemes for assigning limited training sequences of good properties (such as the Zadoff-Chu sequence). One simple strategy is to reuse the same set of orthogonal pilot sequences among neighboring cells [1], [8]. If those cells that utilize identical pilot sequence are separated far away from each other or the cells are assigned orthogonal frequency bands, then there shall be little or no interference among cells during training periods.

However, practical cellular systems are implementing more and more aggressive frequency reuse and denser cell deployment. One consequence is that the target user's channel estimate based on pilot training is perturbed by channel coefficients of other interfering users. This phenomenon is called *pilot contamination* [4]. Especially in the massive MIMO systems that target to serve many more mobile users, the effect of pilot contamination can exacerbate [2], [3]. Mathematically, when users in nearby cells use the same set of training sequences synchronously, the resulting estimated channel matrix of user k becomes

$$\hat{\mathbf{H}}_k = \mathbf{H}_k + \sum_{j \in \mathcal{I}} \mathbf{H}_j + \hat{\mathbf{N}}, \tag{5}$$

where $\mathcal{I}$ is the set of interfering users during training phase and $\hat{\mathbf{N}}$ is the error term because of noise associated with channel estimates [1], [8]. We call the users in set $\mathcal{I}$ the *pilot-interfering users*. In the following section, given this impaired channel estimate, we will begin to develop receivers that are robust to such CSI uncertainty.

III. FEC Code-anchored Quadratic Programming Receiver

Without loss of generality, let us designate the first user as the target user within the cell. Our goal is to design a linear receiver $\mathbf{W} = [\mathbf{w}_R \ \mathbf{w}_I]$ that can retain target user's signal while effectively suppressing multi-access interference despite pilot contamination.

A. MOE Design Principle

One of the well-established criteria for interference suppression in CDMA is the *minimum output energy* (MOE) criterion, which minimizes the total output energy while preserving a desired response for the target user [9], [10]. This principle has been shown effective in multi-access interference mitigation because any reduction in output energy is obtained by suppressing interference once the desired response is preserved.

Based on the MOE criterion, a minimum variance (MV) receiver in the context of multi-access MIMO systems was proposed in [13]. Specifically, the MV detector is the solution to the following optimization problem

$$\begin{aligned} \min_{\mathbf{W}} \quad & \text{tr}\{\mathbf{W}^T \mathbf{R} \mathbf{W}\} \\ \text{s.t.} \quad & \mathbf{W}^T \hat{\mathbf{H}}_1 = \mathbf{I}_2, \end{aligned} \tag{6}$$

where $\mathbf{R} = \mathbb{E}\{\mathbf{y}\mathbf{y}^T\} \in \mathbb{R}^{2N_r \times 2N_r}$ is the covariance matrix. In practice, the unknown covariance matrix $\mathbf{R}$ is replaced by its estimate from T_R snapshots of received signals $\hat{\mathbf{R}} = \frac{1}{T_R} \sum_{t=1}^{T_R} \mathbf{y}_t \mathbf{y}_t^T$. Furthermore, to provide additional robustness against finite samples and to strengthen the condition number of $\hat{\mathbf{R}}$, diagonally loaded covariance matrix $\check{\mathbf{R}} = \hat{\mathbf{R}} + \gamma \mathbf{I}$ is used to replace $\hat{\mathbf{R}}$, where γ is the diagonal loading (DL) factor. The closed-form receiver for diagonally loaded MV (DLMV) receiver using estimated channel $\hat{\mathbf{H}}_1$ is given in [13] as

$$\mathbf{W}_{\text{DLMV}} = \check{\mathbf{R}}^{-1} \hat{\mathbf{H}}_1 (\hat{\mathbf{H}}_1^T \check{\mathbf{R}}^{-1} \hat{\mathbf{H}}_1)^{-1}. \tag{7}$$

B. QP Reformulation of MV Receiver

Though DLMV receiver is capable of suppressing multi-access interference under mild pilot interference, the formulation in Eq. (6) is not easily amendable for further enhancement even if additional signal or user constraints are available. Moreover, the "exact response-preserving constraint" $\mathbf{W}^T \hat{\mathbf{H}}_1 = \mathbf{I}_2$ in Eq. (6) assumes that the channel estimate $\hat{\mathbf{H}}_1$ is sufficiently accurate. When pilot contamination is substantial, such constraint may degrade the receiver performance

due to imperfect CSI $\hat{\mathbf{H}}_1$. In our work, we reformulate the MV receiver through a quadratic programming approach to remedy the noted shortcomings.

Let $\mathbf{w} \triangleq [\mathbf{w}_R^T \ \mathbf{w}_I^T]^T$ be the vectorized receiver parameter matrix $\mathbf{W}$. We then rewrite the MOE cost function in quadratic form

$$\mathrm{tr}\{\mathbf{W}^T\check{\mathbf{R}}\mathbf{W}\} = \begin{bmatrix}\mathbf{w}_R^T & \mathbf{w}_I^T\end{bmatrix}\begin{bmatrix}\check{\mathbf{R}} & \\ & \check{\mathbf{R}}\end{bmatrix}\begin{bmatrix}\mathbf{w}_R \\ \mathbf{w}_I\end{bmatrix}. \tag{8}$$

For the constraint, we follow the same vectorization strategy

$$\mathbf{W}^T\hat{\mathbf{H}}_1 = \mathbf{I}_2 \iff \begin{bmatrix}\hat{\mathbf{H}}_1^T & \\ & \hat{\mathbf{H}}_1^T\end{bmatrix}\begin{bmatrix}\mathbf{w}_R \\ \mathbf{w}_I\end{bmatrix} = \begin{bmatrix}\mathbf{e}_1 \\ \mathbf{e}_2\end{bmatrix}, \tag{9}$$

where the unit vectors $\mathbf{e}_1 = [1 \ \ 0]^T$ and $\mathbf{e}_2 = [0 \ \ 1]^T$ constitute the identity matrix $\mathbf{I}_2$.

Since CSI mismatch exists, we would like to enforce less stringent constraint on response preservation. Particularly, we lift the interference residual $\|(\mathbf{I}_2 \otimes \hat{\mathbf{H}}_1^T)\mathbf{w} - \mathbf{e}\|^2$, where $\mathbf{e} \triangleq [\mathbf{e}_1^T \ \ \mathbf{e}_2^T]^T$, into the cost function to arrive at an unconstrained QP formulation

$$\min_{\mathbf{w}} \ \mathbf{w}^T(\mathbf{I}_2 \otimes \check{\mathbf{R}})\mathbf{w} + \alpha\|(\mathbf{I}_2 \otimes \hat{\mathbf{H}}_1^T)\mathbf{w} - \mathbf{e}\|^2, \tag{10}$$

where α is a regularization parameter. Choice of α depends on the severity of interference and noise, and it is selected by trial and error in practice.

Next, we shall discuss the incorporation of additional user signal constraints to strengthen the performance of the MOE-based receiver optimization. To differentiate target user from pilot-interfering users, we shall take advantage of the inherent FEC code information. In particular, we propose to integrate a set of linear equality constraints that are generated LDPC parity checks [19] in a unified receiver optimization process.

C. FEC Code Anchoring

In data communication systems, FEC codes can be treated as unique user signatures if user code configurations show distinct features. Such configurations can take the form of unique FEC codes or unique codeword permutations of the same FEC code. We call scheme with such configurations the *FEC code diversity* scheme.

Here we focus on the practical and popular LDPC codes. Consider an (N_c, K_c) LDPC code with parity check matrix $\mathbf{P}$. Let $\mathcal{M}$ and $\mathcal{N}$ be the index set of check nodes and variable nodes, respectively, i.e., $\mathcal{M} = \{1, \ldots, N_c - K_c\}$ and $\mathcal{N} = \{1, \ldots, N_c\}$. Denote the neighbor set of the m-th check node $\mathcal{N}_m = \{n \,|\, \mathbf{P}_{m,n} = 1\}$. Of interest are subsets $\mathcal{S} \subseteq \mathcal{N}_m$ that contain an even number of variable nodes; each such subset corresponds to a local codeword [26]. Let $\mathcal{E}_m \triangleq \{\mathcal{S} \,|\, \mathcal{S} \subseteq \mathcal{N}_m \text{ with } |\mathcal{S}| \text{ even}\}$, and introduce auxiliary variable $v_{m,\mathcal{S}} \in \{0, 1\}$ to indicate the local codeword associated with $\mathcal{S}$. Since each parity check node can only be satisfied with one particular even-sized subset $\mathcal{S}$, we have

$$\sum_{\mathcal{S} \in \mathcal{E}_m} v_{m,\mathcal{S}} = 1, \ \forall m \in \mathcal{M}. \tag{11}$$

Moreover, use $f_n \in \{0, 1\}$ to represent variable node n, indicating a bit value of 0 or 1. The bit variables f_n's must be consistent with each local codeword. Thus,

$$\sum_{\mathcal{S} \in \mathcal{E}_m : n \in \mathcal{S}} v_{m,\mathcal{S}} = f_n, \ \forall n \in \mathcal{N}_m, m \in \mathcal{M}. \tag{12}$$

To see how these code constraints characterize a valid codeword at the m-th parity check, note that, according to constraint (11) and the fact that $v_{m,\mathcal{S}}$ takes integer values, we have $v_{m,\mathcal{S}'} = 1$ for some $\mathcal{S}'$ and $v_{m,\mathcal{S}''} = 0$ for all other $\mathcal{S}'' \neq \mathcal{S}'$, where $\mathcal{S}', \mathcal{S}'' \in \mathcal{E}_m$. Furthermore, from constraint (12), we have $f_n = 1$ for all $n \in \mathcal{S}'$ and $f_n = 0$ for all $n \in \mathcal{N}_m \backslash \mathcal{S}'$. Since $|\mathcal{S}'|$ is even-sized, the m-th parity check is satisfied. Constraints (11) and (12) are enforced for every parity check. Together they define a valid codeword [19]. Notice that the constraint $v_{m,\mathcal{S}} \in \{0, 1\}$ would lead to integer programming, which is computationally expensive. Therefore, we relax it to $0 \leq v_{m,\mathcal{S}} \leq 1$ henceforth. Meanwhile, constraint (12) guarantees that $0 \leq f_n \leq 1$.

Finally, to accomplish the integration of code constraints (11) and (12) into the QP receiver in Eq. (10), we will employ additional (linear) constraints that connect the recovered symbols $\hat{\mathbf{x}}_1 = \mathbf{W}^T\mathbf{y}$ to the bit variables $\{f_n\}_{n \in \mathcal{N}}$. In the following, we outline the linear constraints for widely used square QAM with Gray mapping.

D. QAM Mapping Constraints

The 4-QAM with Gray mapping admits affine relationship between bits and symbols. Without loss of generality, for each bit string, we let the left-most bit denote the most-significant bit and the right-most bit denote the least-significant bit. Thus, the 4-QAM modulation mapping between bit and symbol follows

$$\mathbf{w}_R^T\mathbf{y}_t = (2f_{2t-1} - 1)/\sqrt{2}, \tag{13a}$$

$$\mathbf{w}_I^T\mathbf{y}_t = (1 - 2f_{2t})/\sqrt{2}. \tag{13b}$$

General high order QAM with Gray mapping, however, is not amenable to such nice affine relationships between bits and symbols as shown in 4-QAM. Still, variables for bits and symbols in generic QAM can be linked by introducing indicator variables $q_{t,d}$, where $t \in \mathcal{T} = \{1, \ldots, T\}$ and $d \in \mathcal{Q} = \{1, \ldots, Q\}$ with $\mathcal{T}$ and $\mathcal{Q}$ being the index sets of time instants and signal constellation, respectively. Further, we introduce $\{s_d\}_{d \in \mathcal{Q}}$ as the signal constellation points in the Q-ary QAM. Let $\mathcal{X}_{t,\ell}^1$ denote the index set of constellation symbols whose ℓ-th bit is 1, i.e., $\mathcal{X}_{t,\ell}^1 = \{d \in \mathcal{Q} \,|\, (\ell\text{-th bit of } s_d) = 1\}$. In addition, $\mathcal{B} = \{1, \ldots, \log_2 Q\}$ denotes bit indices in a QAM symbol. Then, the following linear constraints for high-order QAM were given in [27]

$$\mathbf{w}_R^T\mathbf{y}_t = \sum_{d \in \mathcal{Q}} q_{t,d}\mathrm{Re}\{s_d\}, \quad \forall t \in \mathcal{T}, \tag{14a}$$

$$\mathbf{w}_I^T\mathbf{y}_t = \sum_{d \in \mathcal{Q}} q_{t,d}\mathrm{Im}\{s_d\}, \quad \forall t \in \mathcal{T}, \tag{14b}$$

$$\sum_{d \in \mathcal{X}_{t,\ell}^1} q_{t,d} = f_{\log_2 Q \cdot (t-1) + \ell}, \quad \forall \ell \in \mathcal{B}, \ \forall t \in \mathcal{T}, \tag{14c}$$

$$\sum_{d \in \mathcal{Q}} q_{t,d} = 1, \quad \forall t \in \mathcal{T}, \tag{14d}$$

$$0 \leq q_{t,d} \leq 1, \quad \forall t \in \mathcal{T}, d \in \mathcal{Q}, \tag{14e}$$

E. Joint QP Receiver

In order to use the code information for the purpose of distinguishing target user from pilot-interfering users, we can allocate different codes to different users as unique signatures [24]. However, in practice, the number of mobile users may far exceed the channel codes specified in the system standard. Since interleaver and deinterleaver are commonly utilized for time-diversity, we propose to apply different codeword permutations to different users as unique signatures. Specifically, for the 1st user, codeword $\mathbf{f} = [f_1, \ldots, f_{N_c}]$ is permuted by a permutation matrix $\mathbf{\Pi}_1$ before modulation. Therefore, the permuted codeword $\mathbf{f}^p = \mathbf{\Pi}_1\mathbf{f}$. Before finalizing the joint receiver, we point out that the bit variables f_n's in bit-to-symbol mapping constraints [Eqs. (13) and (14)] should be replaced by the permuted variables f_n^p's while those in code constraints [Eqs. (11) and (12)] remain unchanged. Finally, the joint QP receiver anchored with FEC code is summarized below with 4-QAM as a typical case

$$\begin{aligned}
\min_{\mathbf{w},\mathbf{f},\mathbf{v}} \quad & \mathbf{w}^T(\mathbf{I}_2 \otimes \check{\mathbf{R}})\mathbf{w} + \alpha\|(\mathbf{I}_2 \otimes \hat{\mathbf{H}}_1^T)\mathbf{w} - \mathbf{e}\|^2 \\
\text{s.t.} \quad & \mathbf{w}_R^T\mathbf{y}_t = (2f_{2t-1}^p - 1)/\sqrt{2}, \\
& \mathbf{w}_I^T\mathbf{y}_t = (1 - 2f_{2t}^p)/\sqrt{2}, \\
& \mathbf{f}^p = \mathbf{\Pi}_1\mathbf{f}, \\
& \sum_{\mathcal{S}\in\mathcal{E}_m} v_{m,\mathcal{S}} = 1, \ \forall m \in \mathcal{M}, \\
& \sum_{\mathcal{S}\in\mathcal{E}_m : n\in\mathcal{S}} v_{m,\mathcal{S}} = f_n, \ \forall n \in \mathcal{N}_m, m \in \mathcal{M}, \\
& 0 \le v_{m,\mathcal{S}} \le 1, \forall m \in \mathcal{M}, \mathcal{S} \in \mathcal{E}_m.
\end{aligned} \tag{15}$$

We note that the permutation constraint $\mathbf{f}^p = \mathbf{\Pi}_1\mathbf{f}$ merely reorders vector $\mathbf{f}$ to $\mathbf{f}^p$. For practical implementation, we do not wish to incorporate $\mathbf{f}^p = \mathbf{\Pi}_1\mathbf{f}$ by introducing many more constraints as well as more variables (the vector $\mathbf{f}^p$), leading to higher computational complexity. Instead, we just need to ensure the coefficients of the variables are in their correct storage units. Hence, the *de-facto* optimization variables are $\mathbf{w}$, $\mathbf{f}$ and $\mathbf{v} \triangleq \{v_{m,\mathcal{S}} \,|\, \mathcal{S} \in \mathcal{E}_m, m \in \mathcal{M}\}$ for 4-QAM with additional $\{q_{t,d} \,|\, t \in \mathcal{T}, d \in \mathcal{Q}\}$ for 16-QAM. As a final note, the detected symbols $\hat{\mathbf{x}}_1 = \mathbf{W}^T\mathbf{y}$ are not quantized since $0 \le f_n \le 1$, such that they can be further processed by soft decoding for better performance.

IV. Performance Enhancement Through Joint SOCP Receiver

From the perspective of classical robust receiver design, worst-case and/or probabilistic approaches are good at handling CSI uncertainties. Hence, they may be preferred over the aforementioned regularization of the error term. Among these two approaches, worst-case approach is too pessimistic in certain cases, whereas probabilistic approach can balance outage and performance well. Therefore, in this section, we anchor FEC code into the chance-constrained receiver in [16] which finally leads to a (deterministic) joint SOCP receiver of improved robustness.

Let $\tilde{\boldsymbol{\delta}}_1 \triangleq \tilde{\mathbf{h}}_1 - \hat{\mathbf{h}}_1$ denote the channel estimation error $\sim \mathcal{CN}(0, \sigma_\delta^2\mathbf{I}_{N_r})$, where $\hat{\mathbf{h}}_1$ is the estimate of the true complex-valued channel $\tilde{\mathbf{h}}_1$ defined in Eq. (2). Further, we can split and stack the real and imaginary parts into $\boldsymbol{\delta}_1 = [\mathrm{Re}\{\tilde{\boldsymbol{\delta}}_1\}^T \, \mathrm{Im}\{\tilde{\boldsymbol{\delta}}_1\}^T]^T$. Correspondingly, the real-valued error matrix is $\mathbf{\Delta}_1 = \mathbf{H}_1 - \hat{\mathbf{H}}_1$ with $\mathbf{\Delta}_1(:,1) = \mathbf{\Psi}_R\boldsymbol{\delta}_1$ and $\mathbf{\Delta}_1(:,2) = \mathbf{\Psi}_I\boldsymbol{\delta}_1$, where the MATLAB colon notation is used for matrix slicing and the matrices

$$\mathbf{\Psi}_R = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \otimes \mathbf{I}_{N_r}, \quad \mathbf{\Psi}_I = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} \otimes \mathbf{I}_{N_r}.$$

Because of symmetry, we can consider either the real or the imaginary part. Hence in the following formulation, we take the real part formulation as an example without loss of generality. Rather than enforce the straightforward response-preservation and interference-cancelation constraints as in Eq. (9), those constraints are now handled probabilistically. Specifically, the response-preserving constraint takes the form $\Pr\{\mathbf{w}_R^T(\hat{\mathbf{H}}_1(:,1) + \mathbf{\Delta}_1(:,1)) \ge 1\} \ge p$ and the self-interference suppression constraint becomes $\Pr\{|\mathbf{w}_R^T(\hat{\mathbf{H}}_1(:,2) + \mathbf{\Delta}_1(:,2))| \le \epsilon\} \ge p$, where $p \in [0,1]$ is the associated probability and ϵ is a constant quantifying the interference residual. Recalling the MOE cost function, we have the following formulation

$$\begin{aligned}
\min_{\mathbf{w}_R} \quad & \mathbf{w}_R^T\check{\mathbf{R}}\mathbf{w}_R \\
\text{s.t.} \quad & \Pr\{\mathbf{w}_R^T(\hat{\mathbf{H}}_1(:,1) + \mathbf{\Delta_1}(:,1)) \ge 1\} \ge p. \\
& \Pr\{|\mathbf{w}_R^T(\hat{\mathbf{H}}_1(:,2) + \mathbf{\Delta_1}(:,2))| \le \epsilon\} \ge p.
\end{aligned} \tag{16}$$

Through a series of transformations, simplifications and approximations, the above chance-constrained problem can be manipulated into a deterministic SOCP for $p \in (0.5, 1)$ [16]. Moreover, both real and imaginary parts are taken into account with the auxiliary variables τ_R and τ_I. We now have

$$\begin{aligned}
\min_{\mathbf{w}_R,\mathbf{w}_I} \quad & \tau_R + \tau_I \\
\text{s.t.} \quad & ||\mathbf{Z}_R\mathbf{w}_R||_2 \le \tau_R, \ ||\mathbf{Z}_I\mathbf{w}_I||_2 \le \tau_I. \\
& ||\mathbf{\Psi}_R^T\mathbf{w}_R||_2 \le \frac{\mathbf{w}_R^T\hat{\mathbf{H}}_1(:,1) - 1}{\sigma_\delta \, \mathrm{erf}^{-1}(2p-1)}, \\
& ||\mathbf{\Psi}_I^T\mathbf{w}_I||_2 \le \frac{\mathbf{w}_I^T\hat{\mathbf{H}}_1(:,2) - 1}{\sigma_\delta \, \mathrm{erf}^{-1}(2p-1)}.
\end{aligned} \tag{17}$$

Interested readers are referred to [16] for more technical details. Note that, however, we modify the space-time model in [16] to the spatial multiplexing model in our work, and the matrices' definitions are changed correspondingly. For clarity and completeness, we list the matrices used in Eq. (17) as follows

$$\mathbf{Q}_R = \frac{1}{1-p}\left[\hat{\mathbf{H}}_1(:,2)\hat{\mathbf{H}}_1(:,2)^T + \frac{1}{2}\mathbf{\Psi}_I\mathbf{\Psi}_I^T\right],$$
$$\mathbf{Q}_I = \frac{1}{1-p}\left[\hat{\mathbf{H}}_1(:,1)\hat{\mathbf{H}}_1(:,1)^T + \frac{1}{2}\mathbf{\Psi}_R\mathbf{\Psi}_R^T\right],$$

where $\check{\mathbf{R}} + \mathbf{Q}_R = \mathbf{Z}_R^T\mathbf{Z}_R$ and $\check{\mathbf{R}} + \mathbf{Q}_I = \mathbf{Z}_I^T\mathbf{Z}_I$ are the Cholesky decomposition. Furthermore, the error function $\mathrm{erf}(x) = \frac{2}{\sqrt{\pi}}\int_0^x e^{-t^2}dt$ and $\mathrm{erf}^{-1}(\cdot)$ is its inverse.

Finally, by integrating symbol-to-bit mapping and LDPC code constraints in a similar way as the joint QP receiver, we

have the following joint SOCP formulation

$$\begin{aligned}
\min_{\mathbf{w},\mathbf{f},\mathbf{v}} \quad & \tau_R + \tau_I \\
\text{s.t.} \quad & ||\mathbf{Z}_R\mathbf{w}_R||_2 \le \tau_R,\ ||\mathbf{Z}_I\mathbf{w}_I||_2 \le \tau_I. \\
& ||\mathbf{\Psi}_R^T\mathbf{w}_R||_2 \le \frac{\mathbf{w}_R^T\hat{\mathbf{H}}_1(:,1) - 1}{\sigma_\delta\,\text{erf}^{-1}(2p-1)}, \\
& ||\mathbf{\Psi}_I^T\mathbf{w}_I||_2 \le \frac{\mathbf{w}_I^T\hat{\mathbf{H}}_1(:,2) - 1}{\sigma_\delta\,\text{erf}^{-1}(2p-1)}, \\
& \mathbf{w}_R^T\mathbf{y}_t = (2f_{2t-1}^p - 1)/\sqrt{2}, \\
& \mathbf{w}_I^T\mathbf{y}_t = (1 - 2f_{2t}^p)/\sqrt{2}, \\
& \mathbf{f}^p = \mathbf{\Pi}_1\mathbf{f}, \\
& \sum_{\mathcal{S}\in\mathcal{E}_m} v_{m,\mathcal{S}} = 1,\ \forall m \in \mathcal{M}, \\
& \sum_{\mathcal{S}\in\mathcal{E}_m : n\in\mathcal{S}} v_{m,\mathcal{S}} = f_n,\ \forall n \in \mathcal{N}_m, m \in \mathcal{M}, \\
& 0 \le v_{m,\mathcal{S}} \le 1, \forall m \in \mathcal{M}, \mathcal{S} \in \mathcal{E}_m.
\end{aligned} \tag{18}$$

V. Cooperative Reception via Distributed ADMM

In the last two sections, we have introduced CARD receivers at a single BS that attempt to suppress both intra-cell interference and inter-cell interference (ICI). We note, however, given broader deployment of heterogeneous cell, ICI becomes relatively stronger and is thus difficult to fully mitigate. The problem of ICI is compounded in pilot-contaminated scenarios. To cope with ICI, multi-cell cooperative processing (MCP) has been proposed to harness, rather than mitigate, the interference in between small cells [28].

Though MCP is promising in tackling ICI, a straightforward centralized cooperation requires data exchange in large volume, which in turn places heavy burdens on radio backhaul. In consideration of the heavy backhaul traffic in such centralized scheme, many researchers devised their algorithms in a distributed fashion [29], [30]. Among various choices of algorithms, alternating direction method of multipliers (ADMM) is a simple but powerful algorithm that is well suited for distributed convex optimizations [25]. In this section, we apply ADMM to our proposed joint QP receiver for enabling collaborations between BS's while maintaining a low traffic overhead. For joint SOCP receiver, however, it is not as easy to deploy distributed implementations; cf. Section V-C for further discussions.

A. Reception at Single BS

First, we rewrite the joint quadratic program [Eq. (15)] in an ADMM form for single-BS reception. For clarity, concatenate the optimization variables $\mathbf{s} \triangleq [\mathbf{w}^T\ \mathbf{f}^T\ \mathbf{v}^T]^T$. We proceed to fill in the coefficients and appropriately pad zeros in matrices used in both cost function and constraints. In particular, we use matrix $\mathbf{Q}$ for MOE cost function, $\mathbf{A}$ and $\mathbf{a}$ for response-preserving constraint, $\mathbf{C}$ and $\mathbf{c}$ for code constraint, as well as $\mathbf{M}$ and $\mathbf{m}$ for mapping constraint:

$$\begin{aligned}
& \mathbf{w}^T(\mathbf{I}_2 \otimes \check{\mathbf{R}})\mathbf{w} \Rightarrow \mathbf{s}^T\mathbf{Q}\mathbf{s}, \\
& (\mathbf{I}_2 \otimes \hat{\mathbf{H}}_1^T)\mathbf{w} - \mathbf{e} \Rightarrow \mathbf{A}\mathbf{s} - \mathbf{a}, \\
& \text{Eqs. (11) and (12)} \Rightarrow \mathbf{C}\mathbf{s} = \mathbf{c}, \\
& \text{Eqs. (13) or (14)} \Rightarrow \mathbf{M}\mathbf{s} = \mathbf{m}.
\end{aligned} \tag{19}$$

Note that all constraints are equalities except the box constraints for variables in lower and upper bounds. Denote the feasible set of box constraints by $\mathcal{C} \triangleq \{\mathbf{s} \,|\, \mathbf{b}_l \le \mathbf{s} \le \mathbf{b}_u\}$. For an easy fit into the ADMM paradigm, we lift the equality constraints into the objective function as penalty terms. Thus, an ADMM objective function in standard quadratic form is

$$\begin{aligned}
f(\mathbf{s}) &= \mathbf{s}^T\mathbf{Q}\mathbf{s} + \alpha\|\mathbf{A}\mathbf{s} - \mathbf{a}\|^2 + \beta\|\mathbf{M}\mathbf{s} - \mathbf{m}\|^2 + \lambda\|\mathbf{C}\mathbf{s} - \mathbf{c}\|^2 \\
&= \frac{1}{2}\mathbf{s}^T\mathbf{P}\mathbf{s} + \mathbf{q}^T\mathbf{s} + r,
\end{aligned} \tag{20}$$

where α, β and λ are weighting coefficients to control the tightness of each penalty term, and the second equality follows by collecting the corresponding terms. That is, $\mathbf{P} = 2\mathbf{Q} + 2\alpha\mathbf{A}^T\mathbf{A} + 2\beta\mathbf{M}^T\mathbf{M} + 2\lambda\mathbf{C}^T\mathbf{C}$, $\mathbf{q} = -2\alpha\mathbf{A}^T\mathbf{a} - 2\beta\mathbf{M}^T\mathbf{m} - 2\lambda\mathbf{C}^T\mathbf{c}$ and $r = \alpha\|\mathbf{a}\|^2 + \beta\|\mathbf{m}\|^2 + \lambda\|\mathbf{c}\|^2$.

With the box-constrained QP in hand, we form the augmented Lagrangian using parameter ρ and scaled dual variable $\mathbf{u}$, as well as the indicator function $g(\mathbf{z})$ defined on $\mathcal{C}$ [25]

$$L_\rho(\mathbf{s}, \mathbf{z}, \mathbf{u}) = f(\mathbf{s}) + g(\mathbf{z}) + (\rho/2)\|\mathbf{s} - \mathbf{z} + \mathbf{u}\|^2. \tag{21}$$

The iterative updating steps immediately follow as

$$\text{s-step:}\quad \mathbf{s}^{k+1} = (\mathbf{P} + \rho\mathbf{I})^{-1}(\rho(\mathbf{z}^k - \mathbf{u}^k) - \mathbf{q}) \tag{22a}$$

$$\text{z-step:}\quad \mathbf{z}^{k+1} = \Pi_{\mathcal{C}}(\mathbf{s}^{k+1} + \mathbf{u}^k) \tag{22b}$$

$$\text{u-step:}\quad \mathbf{u}^{k+1} = \mathbf{u}^k + \mathbf{s}^{k+1} - \mathbf{z}^{k+1}. \tag{22c}$$

where $\Pi(\cdot)_{\mathcal{C}}$ denotes projection onto $\mathcal{C}$. See (Chapters 3-5, [25]) for details on variable initialization, parameter selection and termination criterion.

B. Cooperative Reception

Consider N_b base stations in cooperative uplink processing. Each BS sees a different uplink channel and receives different signals. Hence, each BS has its own objective function $f_i(\mathbf{s}_i)$ with local variable $\mathbf{s}_i = [\mathbf{w}_i^T\ \mathbf{f}_i^T\ \mathbf{v}_i^T]^T \in \mathbb{R}^{n_i\times 1}$ where $f_i(\mathbf{s}_i) = \frac{1}{2}\mathbf{s}_i^T\mathbf{P}_i\mathbf{s}_i + \mathbf{q}_i^T\mathbf{s}_i + r_i$ is specified in Eq. (20). Note that each BS keeps its $\mathbf{w}_i$ to counter the channel effect. However, the set of cooperating BS's need to reach consensus on unknown variables $\mathbf{f}_i$ and $\mathbf{v}_i$ which are common properties associated with a target user. To establish collaborations, the global objective is set to minimize $\sum_{i=1}^N f_i(\mathbf{s}_i)$, and each local variable $\mathbf{s}_i$ consists of partial global variable $\mathbf{z}$. In another word, each component of a local variable corresponds to some global variable component $\mathbf{z}_g$. The mapping from local variable indices to global variable index can be written as $g = \mathcal{G}(i,j)$, which means that local variable component $(\mathbf{s}_i)_j$ corresponds to global variable component $\mathbf{z}_g$. To achieve consensus among local variables and global variables means that $(\mathbf{s}_i)_j = \mathbf{z}_{\mathcal{G}(i,j)}, i = 1,\ldots,N_b, j = 1,\ldots,n_i$ [25].

In summary, we have the following general-form consensus problem

$$\begin{aligned}
\min.\quad & \sum_{i=1}^{N_b} f_i(\mathbf{s}_i) \\
\text{s.t.}\quad & \mathbf{s}_i = \tilde{\mathbf{z}}_i,\ i = 1,\ldots,N_b,
\end{aligned} \tag{23}$$

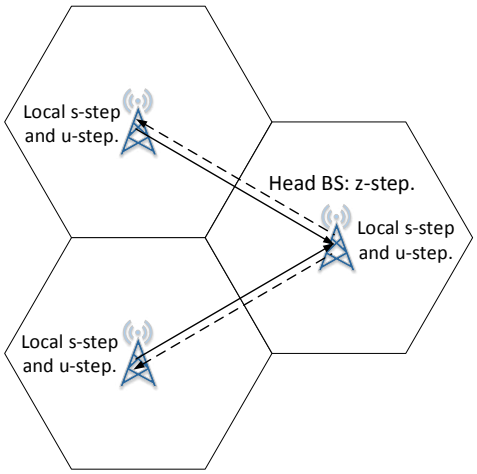


Fig. 2: ADMM process in an MC-MU-MIMO system: solid lines represent the flow direction of s and u; dashed lines denote the flow direction of z.

where $\tilde{\mathbf{z}}_i$ is part of $\mathbf{z}_i$ that relates to $\mathbf{s}_i$. After forming the augmented Lagrangian as in Eq. (21), we have the iterative updating steps shown as below

$$\text{s-step:} \quad \mathbf{s}_i^{k+1} = (\mathbf{P}_i + \rho\mathbf{I})^{-1}(\rho(\tilde{\mathbf{z}}_i^k - \mathbf{u}_i^k) - \mathbf{q}_i),\ \forall i \quad (24a)$$

$$\text{z-step:} \quad \mathbf{z}_g^{k+1} = \Pi_{\mathcal{C}}\left[\frac{\sum_{G(i,j)=g}((\mathbf{s}_i^{k+1})_j + (\mathbf{u}_i^k)_j)}{\sum_{G(i,j)=g} 1}\right] \quad (24b)$$

$$\text{u-step:} \quad \mathbf{u}_i^{k+1} = \mathbf{u}_i^k + \mathbf{s}_i^{k+1} - \tilde{\mathbf{z}}_i^{k+1},\ \forall i = 1, \ldots, N_b. \quad (24c)$$

We note that s-step and u-step are performed locally by each BS, whereas the z-step requires information from each participating BS and should be done at a center or a head BS in the cooperating BS cluster, from which the consensus global information $\mathbf{z}$ is disseminated to every local BS. Fig. 2 illustrates the gathering and disseminating process of ADMM. Also notice that the amount of data exchange in distributed ADMM is significantly less than that needed for gathering all received signals at a central office. This is because the number of served mobile users is much smaller than the number of antennas in a massive MIMO system.

C. Discussion on distributed SOCP

We point out that it is not as easy to deploy distributed implementations for SOCP as for QP. Let $\mathbf{s}$ denote the concatenated variable vector for the joint SOCP (similar to the $\mathbf{s}$ in Section V-A) and $\mathcal{K}$ denote the cone. Then for $\mathbf{s} \in \mathcal{K}$, the z-update step in Eq. (24) involves projection onto $\mathcal{K}$ rather than onto the simple box region $\mathcal{C}$. Notice that the projection onto cone is not a simple computation; see Chapter 5.2.1 in [25]. Instead, we considered a viable alternative approach, which may serve as a "virtually" distributed implementation of SOCP. Our idea is to use joint QP and joint SOCP in conjunction. Basically, a few iterations of cooperative ADMM for joint QP would generate decent intermediate result as a warm-start point for SOCP. Such "distributed warmup" is expected to improve convergence performance and speed of joint SOCP at a single BS. Given the paper length constraint, we leave detailed discussion and refinement of this idea as a possible study direction of our future works. Our emphasis is that the joint QP and joint SOCP receivers formulated in this work can be flexibly combined.

VI. Performance Evaluation and Complexity Analysis

A. Performance Upper Bound

In this section, we evaluate the performance upper bound of the proposed joint receivers. Of particular interest is the performance in terms of coded bit error probability (BEP). To derive the analytical coded BEP, we focus on the complex-valued system model. Recall that the complex-valued received signals $\tilde{\mathbf{y}} = \sum_{i=1}^{N_u} \tilde{\mathbf{h}}_i \tilde{x}_i + \tilde{\mathbf{n}}$ with normalized transmitting power $\mathbb{E}[|\tilde{x}_i|^2] = 1$. Given the receiver vector $\mathbf{c} \in \mathbb{C}^{N_r \times 1}$, the signal-to-interference-and-noise ratio (SINR) after detection is derived in [8]

$$\text{SINR}(\mathbf{c}) = \frac{P_{\text{signal}}(\mathbf{c})}{P_{\text{contam}}(\mathbf{c}) + P_{\text{interf}}(\mathbf{c}) + P_{\text{noise}}(\mathbf{c})}, \quad (25)$$

where the signal power from the user-of-interest is $P_{\text{signal}}(\mathbf{c}) = \mathbf{c}^H \tilde{\mathbf{h}}_1 \tilde{\mathbf{h}}_1^H \mathbf{c}$, the pilot interference power equals $P_{\text{contam}}(\mathbf{c}) = \mathbf{c}^H \left(\sum_{i \in \mathcal{I}} \tilde{\mathbf{h}}_i \tilde{\mathbf{h}}_i^H\right) \mathbf{c}$, other interference power is $P_{\text{interf}}(\mathbf{c}) = \mathbf{c}^H \left(\sum_{i \notin \mathcal{I} \cup \{1\}} \tilde{\mathbf{h}}_i \tilde{\mathbf{h}}_i^H\right) \mathbf{c}$ and the noise power equals $P_{\text{noise}}(\mathbf{c}) = \sigma_n^2 \mathbf{c}^H \mathbf{c}$.

Notice that our proposed receivers can enhance the MV receivers by incorporating the FEC code constraints. Hence, the performance of the integrated joint receivers is at least as good as the basic MV receivers. To circumvent the difficulty in obtaining a closed-form solution of the joint receiver, we view the basic MV receiver performance as an upper bound. Denote the covariance matrix $\tilde{\mathbf{R}} \triangleq \mathbb{E}[\tilde{\mathbf{y}}\tilde{\mathbf{y}}^H]$, and recall that the pilot-contaminated channel estimate $\hat{\mathbf{h}}_1 = \tilde{\mathbf{h}}_1 + \sum_{i \in \mathcal{I}} \tilde{\mathbf{h}}_i$, where the noise effect is neglected. The complex-valued MV receiver is given by $\mathbf{c}_{\text{MV}} = (\hat{\mathbf{h}}_1^H \tilde{\mathbf{R}}^{-1} \hat{\mathbf{h}}_1)^{-1} \tilde{\mathbf{R}}^{-1} \hat{\mathbf{h}}_1$, while the minimum-mean-squared-error (MMSE) receiver $\mathbf{c}_{\text{MMSE}} = \tilde{\mathbf{R}}^{-1} \hat{\mathbf{h}}_1$. We observe that MV receiver is a scaled MMSE receiver [31], and thus $\text{SINR}(\mathbf{c}_{\text{MV}}) \equiv \text{SINR}(\mathbf{c}_{\text{MMSE}})$. The asymptotic (deterministic) characteristics of $\text{SINR}(\mathbf{c}_{\text{MMSE}})$ are analyzed in [8]. Due to space limit, we refer interested readers to [8] for great details. For brevity, let $\gamma_s \triangleq \text{SINR}(\mathbf{c}_{\text{MV}})$. In this analysis, we consider square M-QAM modulations. Its symbol error probability is given in [32]

$$\begin{aligned} p_s = & 4\left(1 - \frac{1}{\sqrt{M}}\right) Q\left(\sqrt{\frac{3\gamma_s}{M-1}}\right) \\ & \times \left(1 - \left(1 - \frac{1}{\sqrt{M}}\right) Q\left(\sqrt{\frac{3\gamma_s}{M-1}}\right)\right), \end{aligned} \quad (26)$$

where Q-function $Q(x) = \frac{1}{\sqrt{2\pi}} \int_x^\infty e^{-t^2/2} dt$. For Gray-mapped QAM, adjacent constellation symbols only differ by one bit. When error occurs, it is very likely that the wrongly detected symbol is the neighbor to the correct one. Such events dominate the BEP. Therefore, one out of the $\log_2 M$ bits will be erroneous. That is to say, the uncoded BEP can be approximated as $p_b^u \approx p_s / \log_2 M$.

Because we use bit-interleaved coded modulation (BICM) in which every bit is treated equally, we can equivalently view the communication chain from modulator to demodulator as a binary symmetric channel (BSC) with crossover probability p_b^u. Assume that an all-zero codeword is transmitted, but a codeword of weight i is decoded. This pairwise error probability (PEP) between two codewords with Hamming distance i is denoted as $P_2(i)$. From [32], it is known that this PEP is upper bounded as $P_2(i) \leq \Delta^i$, where Δ is the Bhatacharyya parameter. For BSC(p_b^u), $\Delta = \sqrt{4p_b^u(1-p_b^u)}$. Furthermore, for a code ensemble, let B_{ij} be the number of codewords with weight i that corresponds to information sequence of weight j. Then, for an (N_c, K_c) code, the expected number of erroneous information bits $\bar{b} \leq \sum_{j=0}^{K_c} j \sum_{i=0}^{N_c} B_{ij} P_2(i)$. Therefore, the coded BEP p_b^c is upper bounded as follows

$$p_b^c = \frac{\bar{b}}{K_c} \leq \frac{1}{K_c} \sum_{j=0}^{K_c} j \sum_{i=0}^{N_c} B_{ij} \Delta^i. \tag{27}$$

As a final note, we currently use the receiver vector $\mathbf{c}_{\text{MV}}$ in this upper bound for the proposed CARD receivers. If the solutions to those integrated formulations can be approximated in some closed forms, the derived bound will become tighter. We leave this as a future work.

B. QP and SOCP Complexity Comparison

In general, both QP and SOCP have time complexity $\mathcal{O}(N^3)$ via interior-point method where N is the length of variable vector [33]. However, SOCP solver imposes a restriction that no variable can be a member of more than one cone [34]. Therefore, we have to introduce more auxiliary variables and more equality constraints to handle the shared variables in the above second-order cones. Given an N_r-antenna BS, the receiver vector $\mathbf{w} \in \mathbb{R}^{4N_r \times 1}$. For an (N_c, K_c) LDPC code, the bit variable vector $\mathbf{f} \in \mathbb{R}^{N_c \times 1}$. If regular LDPC code with row weight w_r is assumed, then length of the set indicator vector $\mathbf{v}$ is $(N_c - K_c) \times 2^{w_r - 1}$. Thus, the final number of variables in joint SOCP is $12N_r + 4 + N_c + (N_c - K_c) \times 2^{w_r - 1}$, whereas joint QP has $4N_r + N_c + (N_c - K_c) \times 2^{w_r - 1}$ variables in total. The extra $8N_r + 4$ variables in joint SOCP (compared to joint QP) would incur substantially higher complexity if the MIMO system really scales up in size, for example, to hundreds of antennas. Moreover, the joint SOCP receiver requires to obtain $\mathbf{Z}_R$ and $\mathbf{Z}_I$ by Cholesky decomposition, which is of complexity $\mathcal{O}(N_r^3)$ in general. On the other hand, as we will show shortly in Section VII-C, joint SOCP receiver achieves superior performance with respect to joint QP receiver under the same settings. This represents a complexity-performance trade-off.

C. Distributed ADMM Complexity Analysis

Recall the optimization variable vector $\mathbf{s} = [\mathbf{w}^T \ \mathbf{f}^T \ \mathbf{v}^T]^T \in \mathbb{R}^{N \times 1}$ in the single-BS ADMM, where $N = 4N_r + N_c + (N_c - K_c) \times 2^{w_r - 1}$. Loosely speaking, the complexity of cooperative scheme with N_b BS's is N_b-fold that of single-BS ADMM. Here, for simplicity and clarity, we concentrate our attention on the complexity of ADMM scheme at one BS. As noted in Eq. (22), each step has complexity $\mathcal{O}(N)$ except for matrix inverse in the s-step, which is the most costly. To solve the large-scale inverse, either direct inverse or indirect iterative method can be applied. For the positive definite matrix $(\mathbf{P} + \rho\mathbf{I})$, the direct method using Cholesky factorization has a one-time cost of $\mathcal{O}(N^3)$ and the subsequent linear system solution through forward-and-backward substitution requires $\mathcal{O}(N^2)$ in each ADMM iteration. Alternatively, iterative method such as conjugate gradient (CG) is suitable for large-scale problem. As analyzed in [35], standard CG has time complexity $\mathcal{O}(N\sqrt{\kappa})$, where κ is the condition number of $(\mathbf{P} + \rho\mathbf{I})$. Hence, if we set maximum iterations M, the overall complexity is $\mathcal{O}(N^3)$ for ADMM with direct method (assuming $M \ll N$) and $\mathcal{O}(MN\sqrt{\kappa})$ for ADMM with CG. In practice, CG with pre-conditioner and warm start can be even faster [36].

VII. Simulation Results

This section presents numerical results of the proposed CARD receivers. We shall consider the DLMV receiver in Eq. (7), the joint QP receiver in Eq. (15), the SOCP receiver in Eq. (17) as well as the joint SOCP receiver in Eq. (18). To make things clear, DLMV and SOCP receivers are the two receivers without code constraints, while joint QP and joint SOCP receivers are integrated with FEC code constraints. We also show the simulation results of distributed ADMM for solving the joint QP with different number of BS's. However, we do not compare numerical results with theoretical upper bound (cf. Section VI-A) at this moment, mainly because the current upper bound can be rather loose due to the use of plain MV receiver. Future works may evaluate the theoretical bound when a better closed-form approximation to CARD receivers solution is found.

Throughout the simulation section, quasi-static Rayleigh fading channels are assumed. As noted in [13], block length T_R determines the accuracy of the sample covariance matrix $\check{\mathbf{R}}$ in MOE cost function, and thus has a certain impact on the receiver performance. In our simulations, all available signal snapshots in a coherence interval are utilized for estimating $\check{\mathbf{R}}$. Recall that the first user is designated as the target user. All the other users (including pilot-interfering users) interfere with the target user during their data transmissions, whereas only pilot-interfering users cause interference to the user-of-interest during training phase. Such pilot contamination causes channel estimation errors. For simplicity, we attribute all pilot interference to the second user (who is the pilot-interfering user), such that the target user's channel estimate $\hat{\mathbf{H}}_1 = \mathbf{H}_1 + \mathbf{H}_2 + \hat{\mathbf{N}}$. Moreover, we assume the pilot-interfering user's transmission power remains unchanged for training and data symbols.

To solve the QPs and SOCPs at a single BS, off-the-shelf MOSEK solver [34] is used, while for the performance comparison of distributed implementations, ADMM is used for both non-cooperative reception at a single BS and cooperative reception at multiple BS's. In particular, we will compare the receivers in terms of coded bit error rate (BER). Standard log-domain sum-product algorithm (SPA) is used for LDPC decoding, where the maximum iterations of SPA is set to 100.

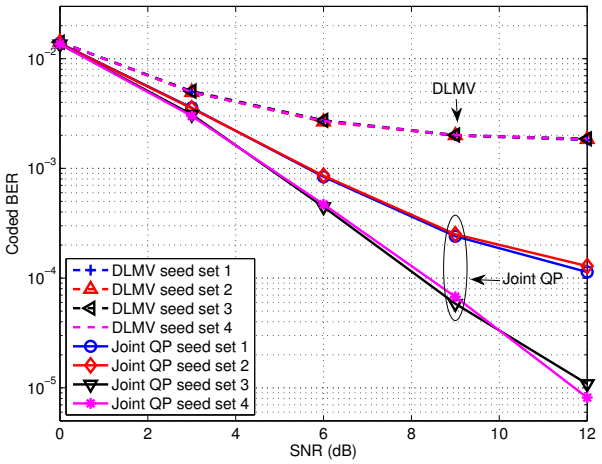


Fig. 3: Demonstration of code diversity effect: DLMV receiver versus joint QP receiver. 4-QAM. $N_r = 32$. LDPC code (256,192). DL factor $\gamma = 200\sigma_n^2$ and regularization parameter $\alpha = 75$. Three users with channel power gains = (1, 0.3, 0.7). Seed set 1 = (568,568,568), set 2 = (568,568,625), set 3 = (568,193,568) and set 4 = (568,193,625).

A. Advantage of FEC Code Anchoring

First, we demonstrate the advantage given by FEC code diversity through code anchoring. Consider 3 interfering uplink users in the network, and their channel power gains to the BS are 1, 0.3 and 0.7, respectively. The BS is equipped with 32 antennas. In this test, we employ interior-point method in MOSEK to solve the joint QP formulation in Eq. (15). The DL factor $\gamma = 200\sigma_n^2$ and regularization parameter $\alpha = 75$. To clearly illustrate the effect of code anchoring, we compare different combinations of code permutations by showing the coded BER versus signal-to-noise ratio (SNR). The permutations are controlled by seeds, which are listed in the figure caption.

Fig. 3 illustrates the achieved BER of the DLMV receiver and our joint QP receiver. By comparing the coded BER of joint QP receiver with that from DLMV receiver under the same code and code permutation conditions, QP receiver consistently provides lower BER. Since DLMV receiver does not exploit FEC code information, the BER results in Fig. 3 of DLMV receiver (in dashed lines) remains the same regardless of FEC codeword permutations.

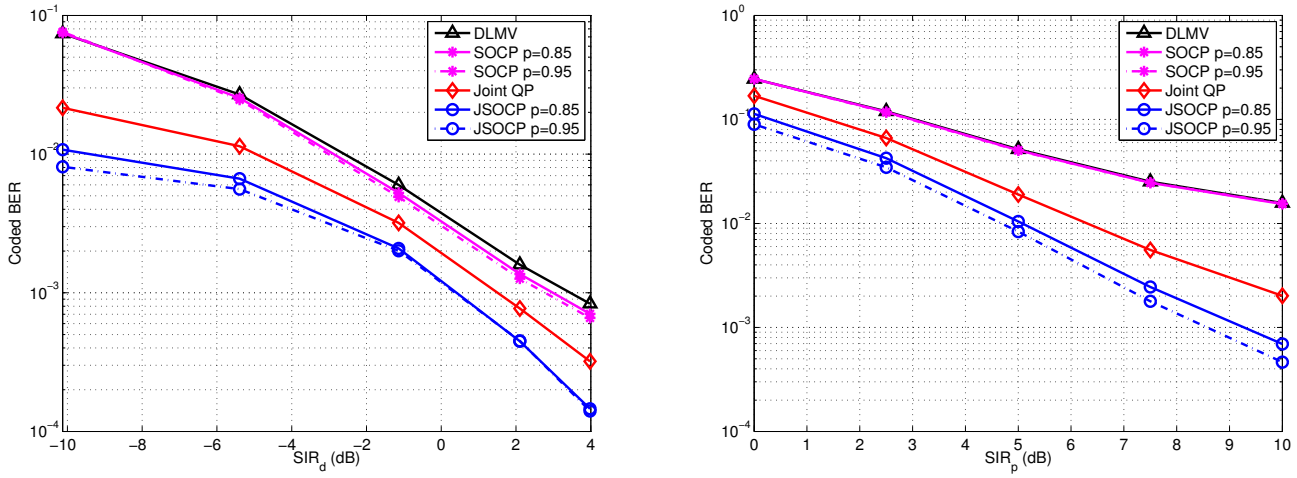
Recall that pilot contamination exists between the first and the second user. Hence, FEC code diversity between these two users is important to our code-anchored QP receiver. Under permutation set 3 and set 4, the target user and pilot-interfering user possess unique FEC permutations as signatures. By exploiting the FEC code diversity through code anchoring, the BER of QP receiver under permutation set 3 and set 4 in Fig. 3 is reduced by as much as 3 orders of magnitude under pilot contamination when compared with DLMV. The similar BER results achieved under these two permutation sets also show that even when user 3 uses the same FEC code and permutation, the BER results remain unchanged because of the lack of pilot contamination from user 3.

On the other hand, when the target user and pilot-interfering user share the same code permutation (in sets 1 and 2), the coded BER of joint QP is severely degraded, regardless of the third user's code permutation. Nevertheless, exploiting FEC code information by the QP receiver can still improve the receiver BER even when the users in pilot contamination are assigned the same FEC code and permutation scheme. One interesting observation from Fig. 3 is that error floors of joint QP receiver are far below those of DLMV receiver, which demonstrates the benefit of applying FEC code constraints. In a word, it is beneficial for us to assign different code permutations among target user and pilot-interfering users to take advantage of code diversity gain in addition to coding gain.

B. Impacts of Payload and Pilot Interference

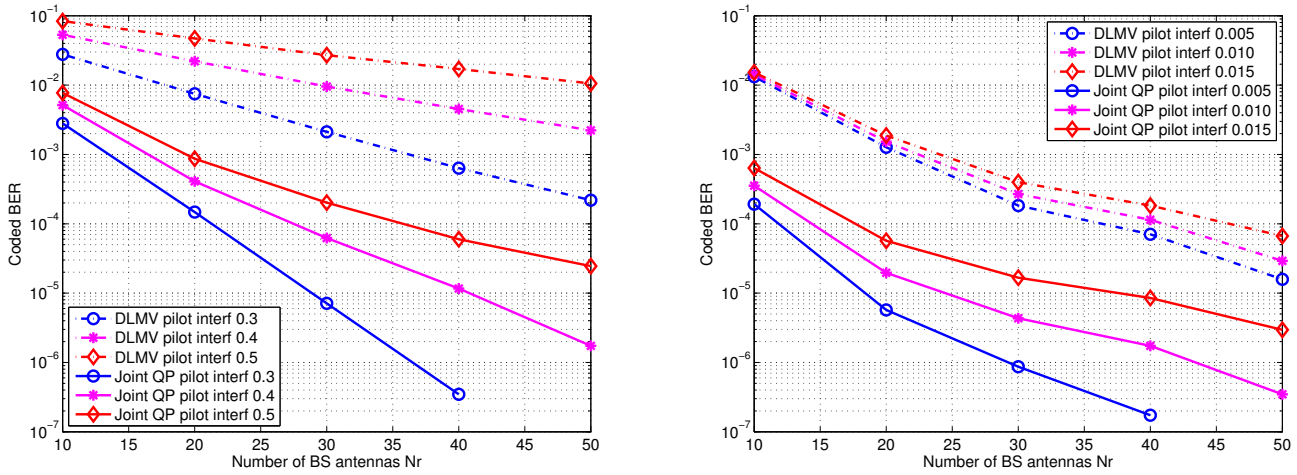
In this section, we design experiments to show the impacts of interference from data payload and pilot sequence. 10 users are considered. The transmit signals of all users are modulated using 4-QAM with unit power. The channel power gains to BS are set to 1 and I_p for the target user (the 1st user) and pilot-interfering user (the 2nd user), respectively. Therefore, the target user's received signal power is $S = 1$. The other interfering users, whose total number is N_d, are assumed to have equal channel gain I_d. To clearly demonstrate the interference effect, we fix the noise level by setting SNR = 5dB. In our simulation tests, we vary the interference powers I_p and I_d. More specifically, we define the signal-to-interference ratio (SIR) of data payload to be $\text{SIR}_d \triangleq 10\log_{10}\frac{S}{N_d \cdot I_d + I_p}$ and the SIR of pilot sequence to be $\text{SIR}_p \triangleq 10\log_{10}\frac{S}{I_p}$. The definitions follow from our settings that the 2nd user is the pilot-interfering user which interferes with the target user during both pilot and data payload transmissions. On the other hand, the remaining N_d users only cause interference during data payload transmissions. Thus, we note that both SIR_d and SIR_p change as I_p varies, whereas I_d only affects SIR_d.

Fig. 4 (a) shows the coded BER versus SIR_d where I_p is fixed at 0.3 but I_d varies. As I_d decreases, performance improves for all the considered receivers. In panel (b) of Fig. 4, all BER curves drop drastically when I_p decreases. In addition, the results illustrate the SOCP receivers' performance for $p = 0.85$ and $p = 0.95$, out of which the larger p-value 0.95 leads to slightly improved performance. We caution that too tight outage constraints (i.e., too large p value) may result in performance degradation. Moreover, it is worth pointing out that, even though the total interference power remains unchanged, BER performance improves for the target user when there are fewer interfering users. This is because the beamformer $\mathbf{w}$ can well "adapt" itself to cancel out the interference channels when there are only a few interfering users. To summarize, DLMV receiver and (disjoint) SOCP receiver are clearly interference-limited. In contrast, the proposed receivers can conquer the limit imposed by interference to a great extent. Thus, the integrated receivers can achieve far lower error rates than the classical ones in all test cases.



(a) (b)

Fig. 4: BER versus data payload and pilot interference. 4-QAM. $N_r = 32$. SNR = 5dB. LDPC code (256,192). DL factor $\gamma = 200\sigma_n^2$ and regularization $\alpha = 75$. (a) Fix $I_p = 0.3$ and vary I_d. (b) Fix $I_d = 0.7$ and vary I_p.



(a) (b)

Fig. 5: BER versus number of BS antennas N_r: DLMV receiver and joint QP receiver. LDPC code (256,192). Three users' seeds = (568,193,625). (a) 4-QAM. SNR = 10 dB. Channel power gains = (1, I_p, 0.8) with pilot interference $I_p = 0.3, 0.4$ and 0.5. (b) 16-QAM. SNR = 15 dB. Channel power gains = (1, I_p, 0.6) with pilot interference $I_p = 0.005, 0.010$ and 0.015.

C. Impact of Number of BS Antennas

As claimed in [2], a clear effect of scaling up the spatial dimensionality is that thermal noise can be averaged out because the coherent signals are received at the antenna array whereas the noises at different antenna elements are independent. In this section, we show the effect of dimensions in large-scale MIMO system by increasing antenna array size at given SIR and SNR. In panels (a) and (b) of Fig. 5, the coded BER results (at different pilot interference levels) are plotted against the number of receive antennas at BS for 4-QAM and 16-QAM, respectively. As N_r increases, the coded BERs decrease without a clear sign of error floor, even though the decreasing slopes are quite different for different array size because of their difference in diversity order. The observation re-affirms the theoretical principle of massive MIMO that improved performance in terms of data rate or link reliability can be achieved with large antenna array. Lastly, same as previous results, joint QP receiver outperforms DLMV receiver with substantial gains.

D. BER Comparisons with Different Code Rates

In general, low-rate codes can provide stronger protections against bit errors. In this test, we compare two codes: a (256,192) code with rate-3/4 and a (256,128) code with rate-1/2. The remaining simulation parameters are listed in the figure caption. Numerical results in Fig. 6 illustrate the performance of joint QP and joint SOCP receivers, where coded BER is shown versus SIR_p. It is clearly seen that the receivers with rate-1/2 code have an advantage of approximately 2–2.5dB over those of rate-3/4 code at BER 10^{-5}. However, the net uplink throughput may not improve as much due to the fact that more redundant bits are used by FEC codes of lower

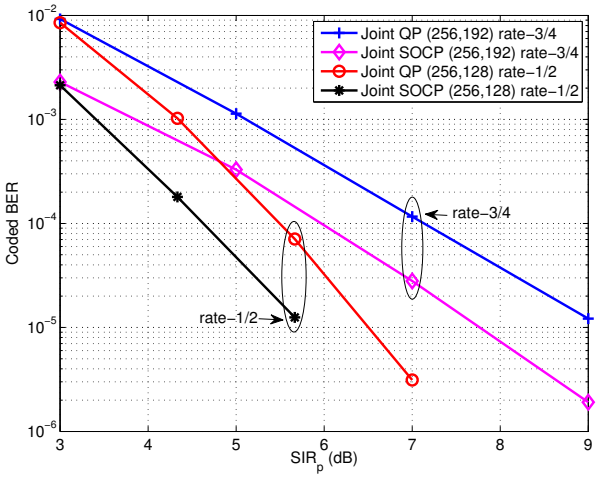


Fig. 6: BER comparisons with different code rates: (256,192) rate-3/4 versus (256,128) rate-1/2. 4-QAM. $N_r = 32$. Three users' seeds = (568,193,625). DL factor $\gamma = 200\sigma_n^2$. regularization parameter $\alpha = 75$. Chance constraints $p = 0.95$. Channel power gains = (1, I_p, 0.7). $\text{SIR}_p \triangleq 10\log_{10}(S/I_p)$.

rates Moreover, we would like to comment that, in addition to lowering code rates, longer codes may also improve BER performance, but at the price of higher complexity, especially for the proposed integrated receivers (cf. the analysis in Section VI-B). One future direction is to explore ways for complexity reduction with little performance loss.

E. Collaborative ADMM Reception at Multiple BS's

In this section, we begin to show the performance advantage of collaborative reception within a cluster of BS's versus non-cooperative reception at a single BS. In particular, we consider a total of 3 BS's covering 5 active mobile users. As before, the 2nd user interferes with the target user during training period. In such a network, each BS can operate on its own, or they can form 3 pairs (each pair consists of 2 BS's), or all of them work together forming a big cluster. We use the ADMM steps in Eq. (22) for single-BS reception, and Eq. (24) for cooperative reception. Specifically, CG is used for the s-step update based on the complexity analysis in Section VI-C. Here we precondition CG with incomplete Cholesky factors and use the solution from last iteration as the initial point for warm start in current iteration. The coded BER results are shown in the first 3 sub-figures (a), (b) and (c) of Fig. 7 for BS 1, 2 and 3, respectively. We use figure legends to depict BS clusterings. As expected, when number of cooperative BS's increases, BER performance improves.

However, the performance gains come at the price of higher complexity. Fig. 7 (d) shows the average runtime of each reception scheme (on machines with identical configurations). Clearly, non-cooperative reception consumes much less time than those collaborative schemes. We point out that the parameter ρ in Eq. (21) has a big impact on convergence speed, where it controls the magnitude of primal residual. The rule-of-thumb is that a larger ρ leads to faster convergence but poorer performance. The trade-off between complexity and performance can also be adjusted by varying the maximum iterations allowed for the iterative updates. After preliminary tests, we assign $\rho = 1, 2, 3$ and maximum iterations $= 300, 250, 200$ for single-BS, two-BS and three-BS reception, respectively. As shown in Fig. 7 (d), a two-BS scheme may even be slower than the three-BS scheme at low SNRs due to the selected ADMM parameters and channel conditions in the simulation setup.

VIII. Conclusion

This work proposes novel receivers based on FEC code-anchored robust design (CARD) for signal recovery in massive MIMO wireless communication systems. To combat the well-known pilot contamination effects, our contribution is the proposal for utilizing FEC codes and codeword permutations as unique user signatures to improve receiver robustness against CSI estimate errors. Unlike existing formulations that only focus on CSI uncertainty, our CARD receivers integrate LDPC code constraints that substantially improve receiver reliability. We also develop a probabilistic approach anchored with code diversity to achieve further performance gain. The proposed receive beamformers are able to extract and recover signals from the user-of-interest with greatly improved robustness. Given channel reciprocity in time-division duplex mode, the receive beamformers can be directly used for transmit beamforming in the downlink direction. For even better design, the detected symbols can serve as pseudo-pilots for more accurate channel information that the downlink beamformer is based on. In such a case, lower-rate FEC codes lead to superior pilot decontamination, which can ameliorate downlink throughput. Faced with strong inter-cell interference, we further present a cooperative reception scheme that works in a distributed fashion enabled by ADMM. Future works may consider quantifying FEC code diversity and further simplify computation complexity.

Acknowledgment

The authors thank the anonymous reviewers for suggestions and comments that substantially improve the manuscript's quality.

References

[1] L. Lu, G. Li, A. Swindlehurst, A. Ashikhmin, and R. Zhang, "An overview of massive MIMO: Benefits and challenges," *IEEE J. Sel. Areas Commun.*, vol. 8, no. 5, pp. 742–758, 2014.

[2] F. Rusek, D. Persson, B. K. Lau, E. G. Larsson, T. L. Marzetta, O. Edfors, and F. Tufvesson, "Scaling up MIMO: Opportunities and challenges with very large arrays," *IEEE Signal Process. Mag.*, vol. 30, no. 1, pp. 40–60, 2013.

[3] E. Larsson, O. Edfors, F. Tufvesson, and T. Marzetta, "Massive MIMO for next generation wireless systems," *IEEE Commun. Mag.*, vol. 52, no. 2, pp. 186–195, 2014.

[4] T. L. Marzetta, "Noncooperative cellular wireless with unlimited numbers of base station antennas," *IEEE Trans. Wireless Commun.*, vol. 9, no. 11, pp. 3590–3600, 2010.

[5] F. Fernandes, A. Ashikhmin, and T. L. Marzetta, "Inter-cell interference in noncooperative tdd large scale antenna systems," *IEEE J. Sel. Areas Commun.*, vol. 31, no. 2, pp. 192–201, 2013.

[6] J. Zhang, B. Zhang, S. Chen, X. Mu, M. El-Hajjar, and L. Hanzo, "Pilot contamination elimination for large-scale multiple-antenna aided OFDM systems," *IEEE J. Sel. Topics Signal Process.*, vol. 8, no. 5, pp. 759–772, 2014.

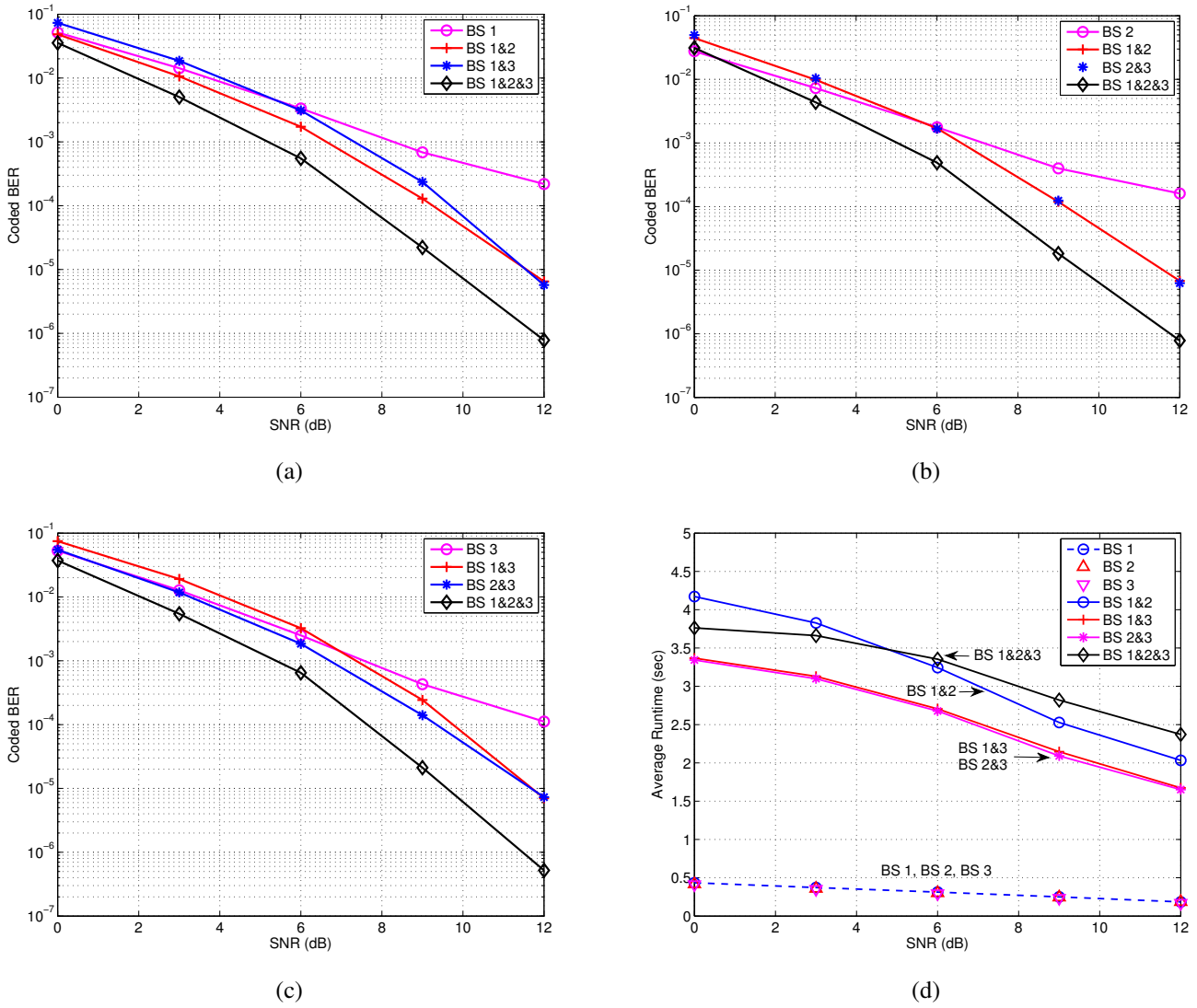


Fig. 7: ADMM implementations of joint QP: single-BS ADMM (BS 1, BS 2, BS 3), two-BS ADMM (BS 1&2, BS 1&3, BS 2&3) and three-BS ADMM (BS 1&2&3). 4-QAM. $N_r = 24$. LDPC code (256,192). $\gamma = 150\sigma_n^2$, $\alpha = 50$, $\beta = 60$ and $\lambda = 70$. Five users' permutation seeds = (987,235,762,398,506). Five users' channel gains to BS 1, 2 and 3 are (0.8, 0.2, 0.9, 0.1, 0.5), (1, 0.25, 0.1, 0.9, 0.3) and (0.7, 0.15, 0.25, 0.1, 0.8), respectively. (a) Coded BER at BS 1. (b) Coded BER at BS 2. (c) Coded BER at BS 3. (d) Average runtime comparisons.

[7] R. R. Muller, L. Cottatellucci, and M. Vehkapera, "Blind pilot decontamination," *IEEE J. Sel. Topics Signal Process.*, vol. 8, no. 5, pp. 773–786, 2014.

[8] N. Krishnan, R. Yates, and N. B. Mandayam, "Uplink linear receivers for multi-cell multiuser MIMO with pilot contamination: Large system analysis," *IEEE Trans. Wireless Commun.*, vol. 13, no. 8, pp. 4360–4373, 2014.

[9] S. Vembu, S. Verdu, R. A. Kennedy, and W. Sethares, "Convex cost functions in blind equalization," *IEEE Trans. Signal Process.*, vol. 42, no. 8, pp. 1952–1960, 1994.

[10] M. Honig, U. Madhow, and S. Verdu, "Blind adaptive multiuser detection," *IEEE Trans. Inform. Theory*, vol. 41, no. 4, pp. 944–960, 1995.

[11] J. B. Schodorf and D. B. Williams, "A constrained optimization approach to multiuser detection," *IEEE Trans. Signal Process.*, vol. 45, no. 1, pp. 258–262, 1997.

[12] S. Cui, M. Kisialiou, Z.-Q. Luo, and Z. Ding, "Robust blind multiuser detection against signature waveform mismatch based on second-order cone programming," *IEEE Trans. Wireless Commun.*, vol. 4, no. 4, pp. 1285–1291, 2005.

[13] S. Shahbazpanahi, M. Beheshti, A. B. Gershman, M. Gharavi-Alkhansari, and K. M. Wong, "Minimum variance linear receives for multiaccess MIMO wireless systems with space-time block coding," *IEEE Trans. Signal Process.*, vol. 52, no. 12, pp. 3306–3313, 2004.

[14] A. B. Gershman, N. D. Sidiropoulos, S. Shahbazpanahi, M. Bengtsson, and B. Ottersten, "Convex optimization-based beamforming: From receive to transmit and network designs," *IEEE Signal Process. Mag.*, vol. 27, no. 3, pp. 62–75, 2010.

[15] Y. Rong, S. Shahbazpanahi, and A. B. Gershman, "Robust linear receivers for space-time block coded multiaccess MIMO systems with imperfect channel state information," *IEEE Trans. Signal Process.*, vol. 53, no. 8, pp. 3081–3090, 2005.

[16] Y. Rong, S. A. Vorobyov, and A. B. Gershman, "Robust linear receivers for multiaccess spacetime block-coded MIMO systems: A probabilistically constrained approach," *IEEE J. Sel. Areas Commun.*, vol. 24, no. 8, pp. 1560–1570, 2006.

[17] R. G. Gallager, "Low-density parity-check codes," *IRE Trans. Inf. Theory*, vol. 8, no. 1, pp. 21–28, Jan. 1962.

[18] D. J. MacKay, "Good error-correcting codes based on very sparse matrices," *IEEE Trans. Inf. Theory*, vol. 45, no. 2, pp. 399–431, Mar. 1999.

[19] J. Feldman, M. J. Wainwright, and D. R. Karger, "Using linear programming to decode binary linear codes," *IEEE Trans. Inf. Theory*, vol. 51, no. 3, pp. 954–972, 2005.

[20] K. Wang, W. Wu, and Z. Ding, "Joint detection and decoding of LDPC coded distributed space-time signaling in wireless relay networks via linear programming," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP), Florence, Italy*, 2014, pp. 1925–1929.

[21] K. Wang, H. Shen, W. Wu, and Z. Ding, "Joint detection and decoding in LDPC-based space-time coded MIMO-OFDM systems via linear programming," *IEEE Trans. Signal Process.*, vol. 63, no. 13, pp. 3411–3424, Jul. 2015.

[22] K. Wang, W. Wu, and Z. Ding, "Diversity combining in wireless relay networks with partial channel state information," in *IEEE Intl. Conf. on Acoust., Speech and Signal Process. (ICASSP), South Brisbane, Queensland*, 2015, pp. 3138–3142.

[23] K. Wang and Z. Ding, "Diversity integration in hybrid-ARQ with Chase combining under partial CSI," *IEEE Trans. Commun.*, DOI: 10.1109/TCOMM.2016.2553664. April 2016.

[24] ——, "Robust receiver design based on FEC code diversity in pilot-contaminated multi-user massive MIMO systems," in *IEEE Intl. Conf. on Acoust., Speech and Signal Process. (ICASSP), Shanghai, China*, 2016.

[25] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Foundations and Trends in Machine Learning*, vol. 3, no. 1, pp. 1–122, 2011.

[26] K. Yang, J. Feldman, and X. Wang, "Nonlinear programming approaches to decoding low-density parity-check codes," *IEEE J. Sel. Areas Commun.*, vol. 24, no. 8, pp. 1603–1613, 2006.

[27] M. F. Flanagan, "A unified framework for linear-programming based communication receivers," *IEEE Trans. Commun.*, vol. 59, no. 12, pp. 3375–3387, Dec. 2011.

[28] D. Gesbert, S. Hanly, H. Huang, S. S. Shitz, O. Simeone, and W. Yu, "Multi-cell MIMO cooperative networks: A new look at interference," *IEEE J. Sel. Areas Commun.*, vol. 28, no. 9, pp. 1380–1408, 2010.

[29] C. Shen, T.-H. Chang, K.-Y. Wang, Z. Qiu, and C.-Y. Chi, "Distributed robust multicell coordinated beamforming with imperfect CSI: an ADMM approach," *IEEE Trans. Signal Process.*, vol. 60, no. 6, pp. 2988–3003, 2012.

[30] S.-J. Kim, S. Jain, and G. B. Giannakis, "Backhaul-constrained multi-cell cooperation using compressive sensing and spectral clustering," in *2012 IEEE 13th Int. Workshop on Signal Process. Advances in Wireless Commun. (SPAWC)*. IEEE, 2012, pp. 65–69.

[31] K. Zarifi and A. B. Gershman, "Asymptotic performance analysis of blind minimum output energy receivers for large DS-CDMA systems," *IEEE Trans. Signal Process.*, vol. 56, no. 2, pp. 650–663, 2008.

[32] J. G. Proakis and M. Salehi, *Digital Communications*. McGraw-Hill Companies, Inc., 2008.

[33] M. S. Lobo, L. Vandenberghe, S. Boyd, and H. Lebret, "Applications of second-order cone programming," *Linear algebra and its applications*, vol. 284, no. 1, pp. 193–228, 1998.

[34] The MOSEK optimization toolbox manual. Version 7.0. [Online]. Available: http://www.mosek.com.

[35] J. R. Shewchuk, "An introduction to the conjugate gradient method without the agonizing pain," 1994.

[36] J. Nocedal and S. Wright, *Numerical optimization*. Springer Science & Business Media, 2006.