A link between Multiuser MMSE and Canonical Correlation Analysis

Mohamed Salah Ibrahim, Paris A. Karakasis, and Nicholas D. Sidiropoulos

***Abstract*—Recent work has shown that repetition coding followed by interleaving induces signal structure that can be exploited to separate multiple co-channel user transmissions, without need for pilots or coordination/synchronization between the users. This is accomplished via a statistical learning technique known as canonical correlation analysis (CCA), which works even when the channels are time-varying. Previous analysis has established that it is possible to identify the user signals up to complex scaling in the noiseless case. This letter goes one important step further to show that CCA in fact yields the linear MMSE estimate of the user signals up to complex scaling, without using any explicit training. Instead, CCA relies only on the repetition and interleaving structure. This is particularly appealing in asynchronous ad-hoc and unlicensed setups, where tight user coordination is not practical.**

***Index Terms*—Multiuser interference, canonical correlation analysis (CCA), repetition coding, minimum mean square error (MMSE) estimation.**

I. Introduction

As modern wireless systems move towards higher spectral efficiencies with more aggressive frequency reuse, mitigating interference becomes an increasingly important challenge and opportunity at the same time. Whereas traditional thinking posits that effective channel estimation for both the desired and the interfering signals is necessary for effective interference mitigation, recent work has shown that directly aiming to extract a signal of interest may be feasible if one exploits the freedom to design the sought transmission line code [1], [2]. Simple repetition coding at the transmitter and a multi-antenna receiver can enable this when coupled with the right receive processing. The latter is rather unconventional: rather than using the repetition structure as a rudimentary error control or spreading leading to very modest gain, the idea is to use the repetition structure to build two matrices that share a common one-dimensional subspace; the one spanned by the signal of interest. This is the key idea and message in [1], which allows recovering the signal of interest even under strong and time-varying analog or digital interference, without any channel estimation or any need for coordination with the interfering co-channel transmitter. The approach even works with time-varying channels and intermittent interference, and it comes with rigorous identifiability guarantees.

When more than one co-channel transmitters wish to communicate using the above protocol, an additional step is needed at each participating transmitter, where after repetition a transmitter-specific pseudo-random interleaver is used to scramble the symbols before transmission. At the receiver, applying the inverse permutation corresponding to a transmitter of interest reinstates the block-repetition structure, but the remaining transmissions are simply double-scrambled. This idea has been exploited in [2] to come up with multiuser uplink decoders and downlink precoders that offer guaranteed signal recovery / isolation respectively.

These claims are not only theoretically sound but also practically verified in a lab setting using software radios [1], [2]. The main theoretical claims of [1], [2] pin down what happens in noiseless scenarios subject to potentially strong interference; they also help explain why the approach is successful in high signal to noise ratio (SNR) scenarios. Experimental results in [1], [2] though indicate impressive performance across a wide range of SNRs. This letter closes this gap in the analysis. It shows that CCA in fact yields the linear MMSE estimate of the user signals up to complex scaling, without using any explicit training. This is particularly appealing in asynchronous ad-hoc and unlicensed setups, where tight user coordination is not practical.

II. System Model and Problem Description

Consider a multiuser communication system comprising K single-antenna transmitters and a receiver equipped with M antennas. Note that a single receiver scenario is adopted for ease of notation and simplicity of exposition; but the proposed method is broadly applicable to cell-free and cellular interference scenarios with potentially strong intra- and inter-cell interference. Let the channel response vector between the k-th user transmitter and the multi-antenna receiver be denoted by $\mathbf{h}_k \in \mathbb{C}^M$ and expressed as

$$\mathbf{h}_k = \sqrt{\alpha_k}\mathbf{z}_k, \tag{1}$$

where $\mathbf{z}_k \in \mathbb{C}^M$ is the vector representing the small scale fading coefficients, while $\alpha_k \in \mathbb{R}_+$ models large scale fading (e.g., path loss). We assume that the receiver has no prior information about the user channels or their statistics.

All users are assumed to be communicating over the same time-frequency resource blocks, potentially in asynchronous and intermittent fashion. Let $\mathbf{x}_k \in \mathbb{C}^N$ represent the k-th user's

Manuscript submitted 4 September 2023; accepted 20 September 2023. This work was supported by NSF ECCS-2118002.

Mohamed Salah Ibrahim was with the Department of Electrical and Computer Engineering, University of Virginia, Charlottesville, VA, 22904 USA (e-mail: salah@virginia.edu); he is now with Interdigital Communications, Conshohocken, PA 19428, USA.
Paris A. Karakasis, and Nicholas D. Sidiropoulos are with the Department of Electrical and Computer Engineering, University of Virginia, Charlottesville, VA, 22904 USA (email: karakasis@virginia.edu, nikos@virginia.edu).

transmitted signal, where $k \in [K] := \{1, \cdots, K\}$ and N is the number of transmitted symbols. Without loss of generality, we assume that $\mathbb{E}[|\mathbf{x}_k(n)|^2] = 1, \forall k \in [K]$ and $n \in [N]$. The user signals can be either analog or digital, and different waveforms can be used across different users. The discrete-time baseband model of the received signal at the receiver, $\mathbf{Y} \in \mathbb{C}^{M \times N}$, can be written as

$$\mathbf{Y} = \sum_{k=1}^{K} \sqrt{p_k} \mathbf{h}_k \mathbf{x}_k^T + \mathbf{W}, \tag{2}$$

where $p_k \in \mathbb{R}_+$ denotes the allocated power of the k-th user, $\forall k \in [K]$. We shall henceforth absorb $\sqrt{p_k}$ in $\mathbf{h}_k$ for simplicity. The matrix $\mathbf{W} \in \mathbb{C}^{M \times N}$ represents the additive white Gaussian noise term with i.i.d. entries with zero mean and variance σ^2 each. Note that the received signals are taken to be synchronous in the above equation, but this is only for simplicity of exposition. The received signals can be asynchronous, and the proposed method will still work, as long as it is possible to synchronize with the signal of the user(s) of interest. This can be accomplished by exploiting the induced signal structure; see [2] for details.

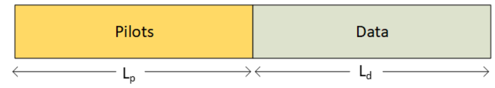
Given the received signal $\mathbf{Y}$, the goal is to decode the K user signals $\{\mathbf{x}_k\}_{k=1}^K$. Traditional methods rely on first estimating the channels associated with the different user transmissions, $\{\mathbf{h}_k\}_{k=1}^K$, followed by designing the equalizer as a function of the estimated channels to recover the signals of interest. The linear minimum mean square error (MMSE, for short) equalizer is the widely adopted equalization technique in actual wireless communication systems, owing to its ability to mitigate both interference and noise impacts, but also its relative simplicity of implementation. After letting $\mathbf{H} := [\mathbf{h}_1, \cdots, \mathbf{h}_K] \in \mathbb{C}^{M \times K}$ represent the channel matrix associated with the K user channels, it can be shown that the MMSE equalizer, given by

$$\mathbf{W}_{\text{MMSE}} = \mathbf{H}^H (\mathbf{H}\mathbf{H}^H + \sigma^2 \mathbf{I})^{-1} \in \mathbb{C}^{K \times M}, \tag{3}$$

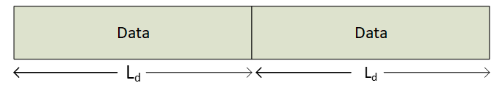
minimizes the mean squared error between the transmitted signal $\mathbf{X} := [\mathbf{x}_1, \ldots, \mathbf{x}_K]^T$ and the equalized signal $\mathbf{W}_{\text{MMSE}}\mathbf{Y}$.

As it can be seen, the performance of the MMSE equalizer depends on the quality of the channel and noise power estimates. Hence, any degradation in the channel estimates along with any under/over-estimation of the noise power may impact the detection performance. In practice, accurate channel estimates can be acquired using orthogonal pilot sequences of length L_p. The pilot sequence length L_p is a key parameter, as large values may hurt the overall spectral efficiency and can even degrade the channel estimates in the case of time-varying channels; short values may also degrade the channel estimates due to insufficient averaging, while also render the so-called pilot contamination [3] problem more severe.

In what follows, we briefly summarize a novel communication paradigm, originally proposed in [2], which relies on repeating the user data followed by interleaving prior to transmission, and applying CCA at the receiver to recover the user signals. In [2], it was shown that the receiver can uniquely unravel the user transmissions under ideal conditions. Herein, we show that the proposed framework can be used to directly find the MMSE equalizer without first performing channel or noise estimation.



(a) Data structure for traditional equalization.



(b) Proposed data structure for CCA equalization.

Fig. 1: (a) Traditional pilot-based TDD data structure. (b) Proposed pilot-free TDD data structure

III. Canonical Correlation Analysis

Canonical correlation analysis (CCA) falls under the umbrella of linear dimensionality reduction techniques and aims at finding linear relationships between two zero-mean[1] random vectors $\mathbf{y}^{(1)} \in \mathbb{C}^{M_1}$ and $\mathbf{y}^{(2)} \in \mathbb{C}^{M_2}$ via their second order statistics (auto-correlation and cross-correlation). Consider the case of two datasets, $\mathbf{Y}^{(1)} := [\mathbf{y}_1^{(1)}, \cdots, \mathbf{y}_N^{(1)}] \in \mathbb{C}^{M_1 \times N}$ and $\mathbf{Y}^{(2)} := [\mathbf{y}_1^{(2)}, \cdots, \mathbf{y}_N^{(2)}] \in \mathbb{C}^{M_2 \times N}$, that emerge after collecting N joint realizations of $\mathbf{y}^{(1)}$ and $\mathbf{y}^{(2)}$. Then, simply put, CCA seeks to find two vectors $\mathbf{q}^{(1)} \in \mathbb{C}^{M_1}$ and $\mathbf{q}^{(2)} \in \mathbb{C}^{M_2}$, known as canonical vectors, that extract two maximally correlated random variables from individual linear combinations of random vectors $\mathbf{y}^{(1)}$ and $\mathbf{y}^{(2)}$. From a mathematical point of view, the CCA problem can be posed as [4], [5]

$$\begin{aligned} \max_{\mathbf{q}^{(1)}, \mathbf{q}^{(2)}} \quad & \operatorname{Re}\left\{ \mathbf{q}^{(1)^H} \mathbf{Y}^{(1)} \mathbf{Y}^{(2)^H} \mathbf{q}^{(2)} \right\} \\ \text{s.t.} \quad & \mathbf{q}^{(i)^H} \mathbf{Y}^{(i)} \mathbf{Y}^{(i)^H} \mathbf{q}^{(i)} = 1, \quad i \in \{1, 2\}, \end{aligned} \tag{4}$$

where the two scaling constraints in (4) are used to make the problem well-defined and avoid any trivial solutions.

One appealing feature of CCA that makes it a favorable tool in practice is that solving problem (4) admits a simple algebraic solution via eigendecomposition [5]. Upon defining $\mathbf{\Sigma}^{(i)} := \frac{1}{N} \mathbf{Y}^{(i)} \mathbf{Y}^{(i)^H}$, as the sample auto-correlation of the random vector $\mathbf{y}^{(i)}$, and $\mathbf{\Sigma}^{(ij)} := \frac{1}{N} \mathbf{Y}^{(i)} \mathbf{Y}^{(j)^H}$, as the sample cross-correlation of the two random vectors $\mathbf{y}^{(i)}$ and $\mathbf{y}^{(j)}$ for $i, j \in \{1, 2\}$ and $i \neq j$, solving (4) is tantamount to solving the generalized eigenvalue problem

$$\mathbf{\Sigma}^{(12)} \mathbf{\Sigma}^{(2)^{-1}} \mathbf{\Sigma}^{(21)} \mathbf{q}^{(1)} = \lambda \mathbf{\Sigma}^{(1)} \mathbf{q}^{(1)}. \tag{5}$$

It can be easily verified that the largest eigenvalue, $\lambda^\star$, represents the square of the correlation coefficient

$$\rho(\mathbf{q}^{(1)^\star}, \mathbf{q}^{(2)^\star}) := \operatorname{Re}\{\mathbf{q}^{(1)^{\star H}} \mathbf{Y}^{(1)} \mathbf{Y}^{(2)^H} \mathbf{q}^{(2)^\star}\} = \sqrt{\lambda^\star}. \tag{6}$$

Once the optimal $\mathbf{q}^{(1)^\star}$ and $\lambda^\star$ are obtained from solving (5), the optimal $\mathbf{q}^{(2)^\star}$ can be obtained via direct substitution using

$$\mathbf{q}^{(2)^\star} = \frac{1}{\sqrt{\lambda^\star}} \mathbf{\Sigma}^{(2)^{-1}} \mathbf{\Sigma}^{(21)} \mathbf{q}^{(1)^\star}. \tag{7}$$

[1]This assumption comes without loss of generality as their means, or sample based estimates of them, can be always subtracted as a pre-processing step.

Following its probabilistic interpretation [5], CCA has demonstrated promising performance in solving various problems in wireless communications and signal processing, including direction-of-arrival estimation [6], equalization [7], spectrum sharing [1], and distributed blind source separation [8]. Taking a step away from statistical and probabilistic viewpoints on CCA, the authors of [9] came up with a novel algebraic interpretation of CCA as a method that can discover a common subspace between two data matrix views under a linear generative model. It was demonstrated that if the two views have a single common component plus linearly independent "private" components for each view, applying CCA to those views will recover the common component up to a global complex scaling ambiguity. In a communications context, the scaling ambiguity can be resolved using one or few pilot symbols. In what follows, we will use this viewpoint to show that CCA in fact yields the MMSE equalizer (up to scaling) without channel or noise power estimation.

IV. Main Result

A. Data Repetition

In this section, we explain how CCA can be exploited in order to solve the problem described in Section II. Recall that traditional equalization techniques require transmitting pilot sequences to first estimate the user channels, as shown in Fig. 1a, and then designing the equalizers as a function of the estimated channel to recover the desired signals. Our proposed CCA-based equalization framework is fundamentally different from conventional equalization techniques, in the sense that it does not require any pilot transmission for channel estimation. The proposed framework consists of two steps: repeating the data at the transmitter, as shown in Fig. 1b, and then leveraging the repetition at the receiver to derive the CCA-based equalizers needed for recovering the desired signals. Following data repetition, we assume that each user permutes its repeated data using a unique (i.e., user-specific) code (e.g., derived from the user or hardware ID). This simple permutation step is needed to ensure a single common component when constructing the two user-specific data views at the receiver. Having a one-dimensional common subspace has two important advantages. It provides more flexibility on the transmitted user waveforms (which could in fact be analog), and also, it reduces the receiver complexity considerably – see [2].

Let $\mathbf{x}_{\text{c}k} \in \mathbb{C}^{L_d}$ denote the common signal associated with the k-th user, where L_d represents the length of the common signal, and "c" refers to common. Thus, the transmitted signal $\mathbf{x}_k$, appearing in (2), can be expressed as

$$\mathbf{x}_k = \mathbf{\Pi}_k [\mathbf{x}_{\text{c}k}^\top \ \mathbf{x}_{\text{c}k}^\top]^\top, \tag{8}$$

where $\mathbf{\Pi}_k$ is the k-th user permutation matrix, and it is assumed to be known at the base station (BS), $\forall k \in [K]$. Notice that the block repetition structure, on the right of (8), can be recovered for a specific user, k, only when the corresponding permutation matrix, $\mathbf{\Pi}_k$, is used by the BS. As for the transmitted signals of all the other users, $\neq k$, applying $\mathbf{\Pi}_k$ will keep them randomly permuted, since the permutation matrices are distinct across the different users. In other words, the matrix $\mathbf{\Pi}_j^\top \mathbf{\Pi}_k$ will be equal to identity only if $j = k$, otherwise, it will be a different permutation matrix for every $j \neq k$. Assuming that the BS applies the permutation matrix associated with the k-th user and upon constructing the k-th user views, the baseband equivalent model of the received signal of the i-th view, can be expressed as

$$\mathbf{Y}_k^{(i)} = \mathbf{h}_k \mathbf{x}_{\text{c}k}^\top + \sum_{j \neq k}^{K} \mathbf{h}_j \mathbf{x}_{jk}^{(i)\top} + \mathbf{W}_k^{(i)}, \tag{9}$$

where $\mathbf{Y}_k^{(i)} \in \mathbb{C}^{M \times L_d}$ is the received signal associated with the i-th view of the k-th user, for $i \in \{1, 2\}$, and is obtained as follows. First, the received signal $\mathbf{Y} \in \mathbb{C}^{M \times 2L_d}$ is multiplied by the k-th user permutation matrix $\mathbf{\Pi}_k$ to obtain $\mathbf{Y}_k := \mathbf{Y} \times \mathbf{\Pi}_k$, and then the first view of $\mathbf{Y}_k$, denoted as $\mathbf{Y}_k^{(i)}$, is constructed as $\mathbf{Y}_k[:, 1 : L_d]$ while the second view of $\mathbf{Y}_k$, denoted as $\mathbf{Y}_k^{(2)}$, is constructed as $\mathbf{Y}_k[:, L_d + 1 : 2L_d]$. The term $\mathbf{W}_k^{(i)} \in \mathbb{C}^{M \times L_d}$ represents the resulting noise term associated with the k-th user signal in its i-th view, while the term $\mathbf{x}_{jk}^{(i)}$ denotes the interleaved signal $\mathbf{x}_j$ of the j-th user in the i-th view of the k-th user, i.e., upon getting further permuted by applying the k-th user's permutation matrix. Recall that the power allocation terms are absorbed in the respective channel vectors.

When the channel coherence time is of order L_d, we may permute each L_d block separately (i.e., use two shorter interleavers). As a result, the model of (9) becomes

$$\mathbf{Y}_k^{(i)} = \mathbf{h}_k^{(i)} \mathbf{x}_{\text{c}k}^\top + \sum_{j \neq k}^{K} \mathbf{h}_j^{(i)} \mathbf{x}_{jk}^{(i)\top} + \mathbf{W}_k^{(i)}, \tag{10}$$

where $\mathbf{h}_u^{(i)} \in \mathbb{C}^M$ now denotes the channel coefficients of the u-th user in the i-th view. Our CCA based framework can deal with this more general model as well.

B. CCA based Equalization

Looking at the two views in (10), it is clear that there is only one shared/common component associated with the k-th user signal, $\mathbf{x}_{ck}$. This renders the algebraic interpretation of CCA applicable to the two constructed views in (10). In other words, it can be shown that by solving the CCA problem in (4) given the two views associated with the k-th user in (10), in the noise-free case and under mild linear independence conditions, the common signal $\mathbf{x}_{\text{c}k}$ can be identified up to a global complex scaling ambiguity that can be easily resolved using one pilot symbol (see Theorem 1 in [9]).

In this work, we aim at going a further step forward to investigate what the CCA canonical vectors or equalizers $\{\mathbf{q}^{(i)}\}_{i=1,2}$ represent. Interestingly, it turns out that the resulting CCA equalizer is the MMSE solution given in (3), without going through channel and noise estimation. In other words, the traditional approach using the data structure in Fig. 1a with training symbols/pilots transmitted first, followed by estimating the different user channels, and then designing and applying the MMSE equalizer on the transmitted data, can be replaced by the proposed data structure in Fig. 1b – repeating and interleaving the user data followed by applying CCA at the receiver. What is more, the CCA approach has distinct advantages, as we will see.

To show this, let us define $\mathbf{X}_{jk}^{(i)} \in \mathbb{C}^{L_d \times K}$ as the matrix holding the common signal $\mathbf{x}_{ck}$ along with the other $K-1$ randomly permuted signals $\{\mathbf{x}_{jk}^{(i)}\}$ in its columns, $\forall k \in [K]$, $j \neq k$, and $i \in \{1,2\}$. Moreover, let $\mathbf{H}^{(i)} \in \mathbb{C}^{M \times K}$ hold the channel vectors $\mathbf{h}_u^{(i)}$ of (10) in its columns. Then, given the two constructed data views in (10), consider the following CCA problem

$$\min_{\{\mathbf{q}_k^{(i)}\}_{i=1,2}} \quad \|\mathbf{Y}_k^{(1)\top}\mathbf{q}_k^{(1)} - \mathbf{Y}_k^{(2)\top}\mathbf{q}_k^{(2)}\|_2^2, \tag{11a}$$

$$\text{s.t.} \quad \mathbf{q}_k^{(i)H}\mathbf{Y}_k^{(i)}\mathbf{Y}_k^{(i)H}\mathbf{q}_k^{(i)} = 1, \quad i \in \{1,2\}. \tag{11b}$$

We have the following result regarding the optimal CCA equalizers $\mathbf{q}_k^{(1)^\star}$ and $\mathbf{q}_k^{(2)^\star}$.

Theorem 1. *In the presence of interference and noise, if the different user transmissions are uncorrelated, the matrices* $\mathbf{X}_{jk}^{(i)} \in \mathbb{C}^{L_d \times K}$ *and* $\mathbf{H}^{(i)} \in \mathbb{C}^{M \times K}$ *are full column rank, and problem* (11) *is solved using the ensemble (instead of the sample) auto- and cross-correlation matrices, then the optimal solutions* $\mathbf{q}_k^{(i)^\star}$ *of problem* (11) *are scaled versions of the corresponding MMSE equalizers, as they are given by*

$$\mathbf{q}_k^{(i)^\star} = \gamma_k(\mathbf{H}^{(i)}\mathbf{H}^{(i)^H} + \sigma^2\mathbf{I})^{-1}\mathbf{h}_k^{(i)}, \tag{12}$$

where $\gamma_k \in \mathbb{C}_{\neq 0}$ *is a nonzero complex scaling factor.*

Proof. The proof is relegated to Appendix A. □

Remark 1. *If sample correlation matrices are used in place of the ensemble ones, the (standard) assumption that the different user transmissions are uncorrelated implies that the sample correlation matrices will converge to the ensemble ones, hence the result will hold asymptotically for large enough* L_d.

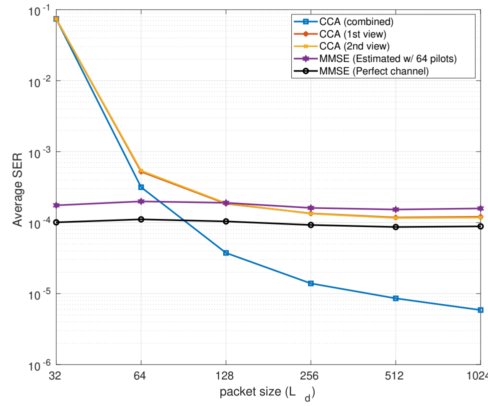
Remark 2. *We can see that in the case where the channel coefficients of all the users are constant during the transmissions of their whole packets, i.e.* $\mathbf{H}^{(1)} = \mathbf{H}^{(2)} = \mathbf{H}$, *also the CCA based pairs of equalizers will be equal for each user and they will match the corresponding MMSE equalizers, namely*

$$\mathbf{q}_k^{(1)^\star} = \mathbf{q}_k^{(2)^\star} = \gamma_k(\mathbf{H}\mathbf{H}^H + \sigma^2\mathbf{I})^{-1}\mathbf{h}_k, \text{ for a } \gamma_k \in \mathbb{C}_{\neq 0}.$$

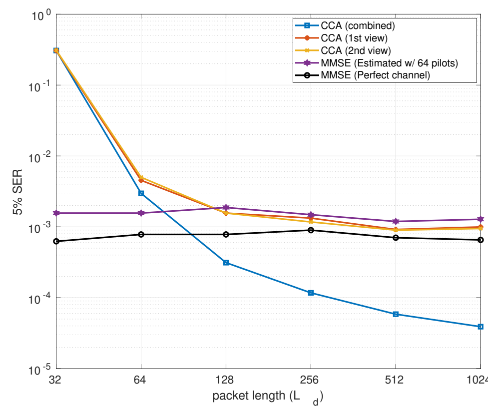
V. Simulation Results

To support the proposed theoretical claim, we simulate a scenario with a single cell of radius 100 meters and a BS equipped with $M = 16$ antennas located at its center. We use 1000 different user drops, each with $K = 10$ users and 100 Monte Carlo (MC) trials for each user, where the UE transmit signals, noise matrix, and small-scale fading coefficients are generated randomly in each MC trial, according to the model described in [2]. The carrier frequency is set to 2.4 GHz, the user transmit power to 23 dBm, and the noise power to -80 dBm. This makes the resulting SNR range approximately between 3 dB and 40 dB. We consider the MMSE equalizers based on perfect channel and noise power information, but also the ones based on estimated channels using orthogonal pilot sequences of length 64 each.

We varied the user packet length L_d from 32 to 1024, with QPSK signal transmissions for all users. Figure 2a and Figure 2b show the mean and cell-edge, i.e., 5%, symbol error rate (SER) performance, respectively. Since CCA uses packet repetition, the desired signal can be obtained either from the first view or the second view or more efficiently by combining/averaging out the outcome from both views. It can be seen that the CCA single-view based estimates approach the MMSE equalizer with perfect channel knowledge (black curve) asymptotically when L_d exceeds 128. Furthermore, CCA with averaging out the output from the two views (blue curve) considerably outperforms the MMSE performance when T_d exceeds 128. This primarily comes from the combining gain obtained from sending the user data twice.



(a) Average SER.



(b) Cell edge (5%) SER.

Fig. 2: Performance comparison between CCA and MMSE.

VI. Discussion and Conclusions

The proposed framework provides a lot of flexibility from a system perspective, making it practically appealing. For instance, traditional training-based estimation of the multiuser MMSE equalizer requires the use of orthogonal pilot sequences. Our proposed framework requires a far less restrictive condition: the user repetition patterns merely need to be distinct, and this can be easily achieved using different user permutation codes. Thus the proposed framework can be more bandwidth-efficient and is immune to so-called pilot contamination. Another key advantage of the proposed framework is that it does not require user coordination (e.g., for training) or synchronization. Users can come in and drop out at will, and their channels can vary on a per-packet basis. In fact, not all users need to employ the proposed repetition and interleaving protocol. We may have some "incumbent" users transmitting (possibly analog) signals, and one or more "secondary" users that employ the proposed protocol. Our claim then applies on the latter users only.

Another key strength of the proposed framework is its robustness to potentially strong intermittent interference from other systems, including from other cells in cellular scenarios. As we have shown, CCA yields the MMSE solution that is naturally accounting for all interference on a per-packet basis. On the other hand, traditional training-based MMSE solutions only learn the channels of users within the cell and average out interference from other systems or cells. They cannot account for potentially harmful intermittent interference, unless they perform training on a per-packet basis, which is hard when packets are short and the number of users is large.

APPENDIX A
PROOF OF THEOREM 1

Let $\mathbf{y}_k^{(i)}$ be a random vector, the realizations of which are described by (10). Moreover, let $\mathbf{x}_k^{(i)}$ be the random vector that consists of random variables $x_{jk}^{(i)}$, for $j \neq k$, expressing the transmitted symbols of the j-th user in the i-th view of the k-th user, and $\mathbf{w}_k^{(i)}$ be the random vector of the additive white Gaussian noise, in the i-th view, consisting of i.i.d. random variables with zero mean and variance σ^2. Finally, let x_{ck} denote a random variable, the realizations of which form the repeated signal associated with the k-th user, $\mathbf{x}_{ck}$.

Then, the sample auto-correlation matrix of $\mathbf{y}_k^{(i)}$ is given by

$$\mathbf{\Sigma}_k^{(i)} = \frac{1}{L_d}\mathbf{Y}_k^{(i)}\mathbf{Y}_k^{(i)^H}. \tag{13}$$

As $L_d \to \infty$, $\mathbf{\Sigma}_k^{(i)}$ converges to the auto-correlation matrix

$$\mathbf{R}_{\mathbf{y}_k^{(i)}} := \mathbb{E}\left[\mathbf{y}_k^{(i)}\mathbf{y}_k^{(i)^H}\right] = \mathbf{H}^{(i)}\mathbf{H}^{(i)^H} + \sigma^2\mathbf{I}, \tag{14}$$

where we have used that

$$\mathbb{E}\left[\begin{bmatrix} x_{ck} \\ \mathbf{x}_k^{(i)} \end{bmatrix}\begin{bmatrix} x_{ck} \\ \mathbf{x}_k^{(i)} \end{bmatrix}^H\right] = \mathbf{I}_K, \quad \mathbb{E}\left[\begin{bmatrix} x_{ck} \\ x_{jk}^{(i)} \end{bmatrix}\mathbf{w}_k^{(i)^H}\right] = \mathbf{0},$$

and $\mathbb{E}[\mathbf{w}_k^{(i)}\mathbf{w}_k^{(i)^H}] = \sigma^2\mathbf{I}_M$. Similarly, the sample cross-correlation matrix $\mathbf{\Sigma}_k^{(ij)}$ can be expressed as

$$\mathbf{\Sigma}_k^{(ij)} = \frac{1}{L_d}\mathbf{Y}_k^{(i)}\mathbf{Y}_k^{(j)^H} \tag{15}$$

for $i, j \in \{1, 2\}$ and $i \neq j$. Again, as $L_d \to \infty$, we have that $\mathbf{\Sigma}_k^{(ij)}$ approaches the cross-correlation matrix

$$\mathbf{R}_{\mathbf{y}_k^{(i)},\mathbf{y}_k^{(j)}} := \mathbb{E}[\mathbf{y}_k^{(i)}\mathbf{y}_k^{(j)^H}] = \mathbf{h}_k^{(i)}\mathbf{h}_k^{(j)^H}, \tag{16}$$

where we have used that

$$\mathbb{E}\left[\begin{bmatrix} x_{ck} \\ \mathbf{x}_k^{(i)} \end{bmatrix}\begin{bmatrix} x_{ck} \\ \mathbf{x}_k^{(j)} \end{bmatrix}^H\right] = \begin{bmatrix} 1 & 0 \\ 0 & \mathbf{0}_{K-1} \end{bmatrix}$$

and the uncorrelatedless of the pairs of random vectors $(\mathbf{w}_k^{(i)}, \mathbf{w}_k^{(j)})$, $(\begin{bmatrix} x_{ck}, & \mathbf{x}_k^{(i)} \end{bmatrix}^T, \mathbf{w}_k^{(j)})$, and $(\begin{bmatrix} x_{ck}, & \mathbf{x}_k^{(j)} \end{bmatrix}^T, \mathbf{w}_k^{(i)})$.

Next, we consider the formulation of CCA in (5), as an eigenvalue problem, but in terms of the ensemble auto-correlation and cross-correlation matrices, i.e.,

$$\mathbf{R}_{\mathbf{y}_k^{(i)}}^{-1}\mathbf{R}_{\mathbf{y}_k^{(i)},\mathbf{y}_k^{(j)}}\mathbf{R}_{\mathbf{y}_k^{(j)}}^{-1}\mathbf{R}_{\mathbf{y}_k^{(j)},\mathbf{y}_k^{(i)}}\mathbf{q}_k^{(i)} = \lambda^*\mathbf{q}_k^{(i)} \tag{17}$$

Let

$$\nu^{(i)} := \mathbf{h}_k^{(i)^H}\left(\mathbf{H}^{(i)}\mathbf{H}^{(i)^H} + \sigma^2\mathbf{I}\right)^{-1}\mathbf{h}_k^{(i)} > 0 \tag{18}$$

and

$$\begin{aligned}\mathbf{B}_k^{(ij)} &:= \mathbf{R}_{\mathbf{y}_k^{(i)}}^{-1}\mathbf{R}_{\mathbf{y}_k^{(i)},\mathbf{y}_k^{(j)}}\mathbf{R}_{\mathbf{y}_k^{(j)}}^{-1}\mathbf{R}_{\mathbf{y}_k^{(j)},\mathbf{y}_k^{(i)}} \\ &= \nu^{(j)}\left(\mathbf{H}^{(i)}\mathbf{H}^{(i)^H} + \sigma^2\mathbf{I}\right)^{-1}\mathbf{h}_k^{(i)}\mathbf{h}_k^{(i)^H},\end{aligned} \tag{19}$$

after substituting the derived expressions for the correlation matrices and pulling the scalar term $\nu^{(j)}$ to the left. Because of (16), matrices $\mathbf{B}_k^{(ij)}$ and $\mathbf{B}_k^{(ji)}$ are rank one, while their unique nonzero eigenvalues are positive, equal, and given by

$$\begin{aligned}\lambda_{\max}\left(\mathbf{B}_k^{(ij)}\right) &= \operatorname{Tr}\left(\mathbf{B}_k^{(ij)}\right) \\ &= \nu^{(i)}\nu^{(j)} \\ &= \operatorname{Tr}\left(\mathbf{B}_k^{(ji)}\right) = \lambda_{\max}\left(\mathbf{B}_k^{(ji)}\right).\end{aligned}$$

As a result, it can be shown that (17), which can be rewritten as

$$\mathbf{B}_k^{(ij)}\mathbf{q}_k^{(i)} = \nu^{(i)}\nu^{(j)}\mathbf{q}_k^{(i)}, \tag{20}$$

is equivalent to $\mathbf{y}_k^{(i)}\mathbf{h}_k^{(i)^H}\mathbf{q}_k^{(i)} = \nu^{(i)}\mathbf{q}_k^{(i)}$, where $\mathbf{y}_k^{(i)} := (\mathbf{H}^{(i)}\mathbf{H}^{(i)^H} + \sigma^2\mathbf{I})^{-1}\mathbf{h}_k^{(i)}$. It can now be easily verified that the solution of (17) is given by

$$\mathbf{q}_k^{(i)^\star} = \gamma_k\mathbf{y}_k^{(i)} = \gamma_k(\mathbf{H}^{(i)}\mathbf{H}^{(i)^H} + \sigma^2\mathbf{I})^{-1}\mathbf{h}_k^{(i)}, \tag{21}$$

where γ_k is a complex scaling factor.

REFERENCES

[1] M. S. Ibrahim, P. A. Karakasis, and N. D. Sidiropoulos, "A simple and practical underlay scheme for short-range secondary communication," *IEEE Transactions on Wireless Communications*, vol. 21, no. 11, pp. 9990 – 10 004, June 2022.

[2] M. S. Ibrahim, A. Hussain, and N. D. Sidiropoulos, "A novel linear precoder design for reliable UL/DL detection in tdd cellular networks," *IEEE Transactions on Communications*, vol. 70, no. 12, pp. 8167–8180, Oct. 2022.

[3] O. Elijah, C. Y. Leow, T. A. Rahman, S. Nunoo, and S. Z. Iliya, "A comprehensive survey of pilot contamination in massive MIMO—5G system," *IEEE Communications Surveys & Tutorials*, vol. 18, no. 2, pp. 905–923, Nov. 2015.

[4] H. Hotelling, "Relations between two sets of variates," *Biometrika*, vol. 28, no. 3/4, pp. 321–377, Dec. 1936.

[5] D. R. Hardoon, S. Szedmak, and J. Shawe-Taylor, "Canonical correlation analysis: An overview with application to learning methods," *Neural Computation*, vol. 16, no. 12, pp. 2639–2664, Dec. 2004.

[6] Q. Wu and K. M. Wong, "Un-music and un-cle: An application of generalized correlation analysis to the estimation of the direction of arrival of signals in unknown correlated noise," *IEEE Transactions on Signal Processing*, vol. 42, no. 9, pp. 2331–2343, Sep. 1994.

[7] A. Dogandzic and A. Nehorai, "Finite-length MIMO equalization using canonical correlation analysis," *IEEE Transactions on Signal Processing*, vol. 50, no. 4, pp. 984–989, Aug. 2002.

[8] A. Bertrand and M. Moonen, "Distributed canonical correlation analysis in wireless sensor networks with application to distributed blind source separation," *IEEE Transactions on Signal Processing*, vol. 63, no. 18, pp. 4800–4813, June 2015.

[9] M. S. Ibrahim and N. D. Sidiropoulos, "Reliable detection of unknown cell-edge users via canonical correlation analysis," *IEEE Transactions on Wireless Communications*, vol. 19, no. 6, pp. 4170–4182, Mar. 2020.