Canonical Correlation Analysis for Next-Generation Cellular and Underlay Communication

---

A

Dissertation

Presented to

the faculty of the School of Engineering and Applied Science

University of Virginia

---

in partial fulfillment

of the requirements for the degree

Doctor of Philosophy

by

Mohamed Salaheldeen Youssef Rezk Ibrahim

August 2021

APPROVAL SHEET

This

Dissertation

is submitted in partial fulfillment of the requirements
for the degree of

Doctor of Philosophy

Author: Mohamed Salaheldeen Youssef Rezk

This Dissertation has been read and approved by the examing committee:

Advisor: Nicholas D. Sidiropoulos

Advisor:

Committee Member: Stephen G. Wilson

Committee Member: Cong Shen

Committee Member: Farzad F. Hassanzadeh

Committee Member: Anil Vullikanti

Committee Member:

Committee Member:

Accepted for the School of Engineering and Applied Science:

Jennifer L. West, School of Engineering and Applied Science

August 2021

Acknowledgements

First and foremost, I would like to express my deepest gratitude to my advisor, Professor Nikos Sidiropoulos, for his continuous help, endless support and valuable guidance throughout my PhD journey. His immense knowledge, ingenuity, and dedication to research has served as an inspiration to me. I would like to thank him for the time he spent on developing my writing and presentation skills, and making me not only a qualified researcher but also a better person. Without his advice, guidance, and insightful suggestions, this dissertation would have never been completed.

I am thankful to Prof. Stephen Wilson, Prof. Cong Shen, Prof. Farzad Hassanzadeh and Prof. Anil Vullikanti for serving on my thesis committee and providing valuable feedback on my thesis. I would like to extend my appreciation to the efforts made by the following Professors during my first year at the University of Minnesota: Georgios Giannakis, Youssef Saad and Mingyi Hong; their graduate level courses have helped me building a solid background in essential areas, and played important roles in my research afterwards.

During my PhD, I have had great moments with my current and former labmates: Kejun Huang, Xiao Fu, Aritra Konar, Charilaos Kanatsoulis, Bo Yang, Cheng Qian, Faisal Almutairi, Nikos Kargas, Mikael Sorensen, Magda Amiridi and Paris Karakasis. I would like to thank all of these people for the insightful discussions that greatly benefited this thesis. Special thanks goes to my friend Mahmoud Elnaggar for all the fun times we had together in Virginia. Finally, I would like to express my sincerest thanks to my friend Ahmed Samir, for all the good times we had together during my masters and PhD. I am thankful to have him as a good friend and collaborator.

Last but not least, I would like to give a special thanks to my parents Amany Hafez and Salah Youssef, whose unconditional love and unwavering support provides me with the strength to continue on my journey in life. I hope to make them proud by becoming a doctor now, and I will keep on doing the hard work in the future. I am also grateful to my sister, my grandmother, my uncle, and my mother-in-law for their continuous support and help. I would like to extend my deepest gratitude to my life partner and beloved wife, Raghda, who has been extremely

supportive of me throughout this entire process and has made countless sacrifices to help me get to this point. Life is wonderful with you.

Dedication

To my children, Lareen and Zain, may Allah bless them.

Abstract

Recent years have witnessed a tremendous amount of research in exploiting machine learning tools for handling a wide variety of problems in wireless communications. While data-driven approaches, notably deep neural networks and deep reinforcement learning, have arguably gained center-stage prominence owing to their empirical success in numerous applications when a lot of training data is available, there are in fact several problems that can markedly benefit from classical machine learning tools and latent factor analysis techniques. This dissertation studies canonical correlation analysis (CCA) and its multi-view generalization (GCCA) in the context of modern wireless communications.

One of the main contributions of this dissertation is that it provides a new and broadly useful algebraic interpretation of (G)CCA as a method that can identify a common subspace between two or more matrices, even if the uncommon components are dominant. Beyond identifiability, it develops two performance analyses which show that the common subspace can be accurately estimated via (G)CCA even in the non-ideal case where there is background noise and strong interference of the individual components in the other matrix view(s). These theoretical findings are leveraged to solve the challenging problem of reliably detecting cell-edge (weak) users in cellular wireless systems. It is shown that cell-edge user signals can be reliably decoded via (G)CCA, at very low signal to noise plus interference ratio (SINR), without knowing their channels. The proposed (G)CCA approach, can tolerate strong interference, achieves superior detection performance compared to the state-of-the-art approaches, and is computationally tractable for practical implementation.

The second part of the dissertation introduces a novel framework that enables efficient spectrum utilization by allowing coexistence between two independently operated co-channel networks. Existing methods require some level of primary-secondary coordination, cross-channel state estimation and tracking, or activity detection – which seriously complicate their practical use. This dissertation develops a simple and practical spectrum underlay solution which enables reliable secondary communication in the presence of the primary network, without primary-secondary coordination or channel state information, under potentially strong and time-varying interference from the primary system. It is shown that the proposed approach enjoys theoretical performance guarantees which are corroborated through laboratory experimentation using software defined radios. The proposed approach works with digital or analog modulation, it is computationally cheap, and, as a side-benefit, it provides means for accurate synchronization of the secondary user even at very low SINR.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1 Background and motivation

Mobile wireless networks are the data highways that are now connecting everything, from people, sensors, cloud centers, to smart vehicles and homes. The exponential growth of the number of wireless-connected devices will result in more complicated networks, and will require the wireless networks to transfer much greater amount of data, at much higher speeds, which is in fact a major challenge for the current mobile systems. Furthermore, such a continuous increase of Internet-connected devices in factories, healthcare, homes, and businesses, that is expected to reach to 200 billion by 2021 [1], will render the available spectral resources being overloaded and significant interference issues will arise as a consequence. These considerations urge the need for cutting-edge technologies and efficient approaches that can satisfy new software and hardware constraints, scale with the size of networks, and more importantly, be capable of providing highly reliable and low latency communication across all connected devices, to facilitate the operation of the next-generation wireless systems.

Within the last few years, machine learning has become prominent and rapidly growing field in the broad areas of wireless communications and signal processing. Machine learning tools have shown a lot of success in tackling a wide variety of challenging problems in wireless communications [2, 3], including resource allocation [4], antenna selection [5, 6], channel estimation [7], to name a few. Data-driven approaches, notably deep neural networks and (deep) reinforcement learning, have gained central-stage prominence due to their empirical success in several applications. Though that, there are in fact a lot of problems that can markedly benefit from classical learning tools and latent factor analysis.

Unlike data-driven approaches which usually lack theoretical understanding, factor analysis techniques as principal component analysis (PCA) [8], independent component analysis (ICA) [9], coupled matrix factorization (CMF) [10], non-negative matrix factorization (NMF) [11] and canonical correlation analysis (CCA) [12], not only have the potential to efficiently handle several challenging problems in various fields, but a lot of theoretical aspects can be provided as well, thereby yielding better understanding compared to black-box approaches. While single-view analysis techniques, like PCA, ICA and NMF use different constraints to extract strong components from the given data matrix, multi-view analysis tools such as CMF and CCA, seek to jointly analyze different views of the data. The main difference between CMF and CCA lies in the optimization criterion: whereas CMF uses a data fitting (usually: least squares) criterion, CCA is based on a "differential" criterion that forces it to zoom in only on what is common between the different views. If one of the views includes a very strong component that is absent from the other view(s), a least squares CMF formulation can still be obliged to represent that component. CCA, on the other hand, owing to its use of a differential (balancing) criterion, can ignore principal components no matter how strong they are, as long as they are not common.

Although CCA has been employed in several areas in signal processing, machine learning and wireless communications, theoretical aspects of CCA was in fact very limited prior to our work, and perhaps surprisingly, none of the aforementioned works has proven identifiability of the common (shared) subspace between different data views. In general, identifiability is very important as it provides sufficient conditions under which the recovery of the common signal subspace via CCA is guaranteed under ideal (noiseless) conditions.

The work in this dissertation has filled that gap in the literature through presenting notable theortical contributions of broader interest. Furthermore, this dissertation provides insightful performance analyses, that yield the conditions required for reliably estimating the common subspace in the presence of noise, and provides the reasons why CCA can even work under low signal-to-noise ratio (SNR) and strong interference form the per-view individual components. Besides our theortical findings, this dissertation will leverage our theoretical findings to showcase the power of CCA in solving challenging problems in 5G and beyond. More importantly, laboratory experimentation using real radios will demonstrate that our theory works in practice.

1.2 Thesis Outline and Contributions

Chapter 2 provides a comprehensive introduction to GCCA through covering different (G)CCA formulations together with their computational complexities. Further, we present the complex domain representation of (G)CCA formulations that will be used in some of the subsequent chapters. Our new algebraic interpretation will also be introduced.

Chapter 3 provides a link between CCA and cell-edge user detection in a cellular network with two cooperating base stations. We show that cell-edge user signals can be reliably recovered using CCA, at low SNR, without knowing their channels, and under strong inter- and intra-cell interference. Further, this chapter presents two theortical contributions of broader interest. First, we present a proof that shows the conditions under which CCA can identify a common subspace under a noise-free linear generative model. Second, we develop a performance analysis which shows that CCA works in practical situations where noise and leakage of strong individual components are present. The results in this chapter have been published in [13, 14].

Chapter 4 extends our study of the cell-edge problem to the more general setup that includes large number of base stations. In such a case, generalized canonical correlation analysis (GCCA) is invoked to handle the problem as opposed to CCA in the two cell setup. We address several interesting questions as how many BSs should cooperate to yield the best detection performance of cell-edge users? Does adding more BSs and allowing them to cooperate really help? Further, we provide an insightful performance analysis which shows that GCCA can reliably estimate the common subspace in presence of noise. Also, we propose a GCCA strategy that can be used to differentiate cell center users from cell edge users. The material of this chapter can be found in [15].

Chapter 5 presents a novel framework for efficient spectrum reuse. We propose a practical underlay scheme that permits reliable secondary communication in the presence of strong and time-varying (e.g., intermittent) interference from the primary system. The proposed method is practically appealing as it can efficiently operate without knowing any channel and without any primary-secondary coordination. From the theortical point of view, we show that reliable and computationally efficient recovery of the secondary signal is possible via CCA. From the practical perspective, we demonstrate through laboratory experimentation using a software radio testbed that, for a secondary user with only two receive antennas, reliable detection of the secondary signal is possible for signal to interference plus noise ratio (SINR) in the range of -20 to -40 dB. Further, the approach works with unknown time-varying channels, digital or analog modulation, and, as a side-benefit, it provides means for accurate synchronization of the secondary user even at very low SINR. Part of the content of this chapter can be found in [16].

Finally, Chapter 6 provides a summary of contributions together with future and ongoing research directions.

1.3 Notational Conventions

In this dissertation, we use the following notations. Upper and lower case bold letters are used to denote matrices and column vectors, respectively. For any general matrix $\mathbf{N}$, we use $\mathbf{N}^T$,

$\mathbf{N}^H$, $\mathbf{N}^{-1}$, $\mathbf{N}^\dagger$ and $\mathrm{Tr}(\mathbf{N})$ to denote the transpose, the conjugate-transpose, the inverse (when it exists), the pseudo-inverse, and the trace of $\mathbf{N}$, respectively. $\mathbf{N}(:, m)$ denotes the m-th column of $\mathbf{N}$ (MATLAB notation). Furthermore, $\mathbb{R}\mathrm{e}\{\mathbf{N}\}$ and $\mathbb{I}\mathrm{m}\{\mathbf{N}\}$ extract the real part and the imaginary part of $\mathbf{N}$, respectively. Scalars are represented in the normal face, while calligraphic letters are used to denote sets. $\|.\|_2$ and $\|.\|_F$ denote the ℓ_2-norm and the Frobenius norm, respectively. Finally, $\mathbf{I}_N$ and $\mathbf{0}_{N\times M}$ denote the $N \times N$ identity matrix and the $N \times M$ zero matrix, respectively. $\mathrm{Diag}(\mathbf{x})$ returns a diagonal matrix with the element of $\mathbf{x}$ on its diagonal.

Chapter 2

Canonical Correlation Analysis

2.1 Overview of CCA

Canonical Correlation Analysis (CCA) is a widely-used statistical learning tool that aims at finding directions in the data with maximal cross correlation. Consider, for instance, T samples of the pair $(\mathbf{y}_1, \mathbf{y}_2)$, where $\mathbf{y}_1 \in \mathbb{R}^{M_1}$ and $\mathbf{y}_2 \in \mathbb{R}^{M_2}$ are two "views" of the same entity. For example, $\mathbf{y}_1$ could contain a set of economic indicators, while $\mathbf{y}_2$ could contain crime, corruption, or social welfare data corresponding to the same country or municipality, and we have data for T countries or municipalities. Or, $\mathbf{y}_1$ could be the electroencephalogram (EEG) of a person and $\mathbf{y}_2$ could be the voxels of a functional magnetic resonance (fMRI) scan; or $\mathbf{y}_1$ could be a person's consumer record, while $\mathbf{y}_2$ could reflect his/her social network connections, and we have data for T people. We are interested in discovering what is common between these two views of the same set of entities. Is there a particular 'latent' factor that affects both the economy and crime, for example? Towards this end, we would like to derive 'meta-variables', one from each view, which are strongly correlated with each other. How can we do this?

Let $\mathbf{y}_1[t]$ and $\mathbf{y}_2[t]$ denote the t-th observation of $\mathbf{y}_1$ and $\mathbf{y}_2$, respectively, corresponding to the t-th entity, for $t \in [T] := \{1, \cdots, T\}$. Assume that both $\mathbf{y}_1$ and $\mathbf{y}_2$ are zero-mean, otherwise the sample mean can be subtracted as a pre-processing step. In its simplest form, CCA seeks to find a pair of linear combinations $\mathbf{x}_1 = \mathbf{y}_1^T \mathbf{q}_1$ and $\mathbf{x}_2 = \mathbf{y}_2^T \mathbf{q}_2$ of the elements of the random vectors $\mathbf{y}_1$ and $\mathbf{y}_2$ such that the two derived random variables $\mathbf{x}_1$ and $\mathbf{x}_2$ are maximally correlated – ideally, perfectly correlated. Mathematically, CCA seeks to find two vectors $\mathbf{q}_1 \in \mathbb{R}^{M_1}$ and $\mathbf{q}_2 \in \mathbb{R}^{M_2}$ such that the correlation coefficient between $\mathbf{Y}_1^T \mathbf{q}_1$ and $\mathbf{Y}_2^T \mathbf{q}_2$ is maximized, where $\mathbf{Y}_\ell := [\mathbf{y}_\ell[1], \cdots, \mathbf{y}_\ell[T]] \in \mathbb{R}^{M_\ell \times T}$ and $\ell \in \{1, 2\}$. In an optimization framework, this can be

expressed as

$$\max_{\mathbf{q}_1,\mathbf{q}_2} \frac{\mathbf{q}_1^T\mathbf{Y}_1\mathbf{Y}_2^T\mathbf{q}_2}{\sqrt{\mathbf{q}_1^T\mathbf{Y}_1\mathbf{Y}_1^T\mathbf{q}_1}\sqrt{\mathbf{q}_2^T\mathbf{Y}_2\mathbf{Y}_2^T\mathbf{q}_2}} \tag{2.1}$$

Let $\mathbf{R}_{\mathbf{y}_\ell\mathbf{y}_\ell} := \frac{1}{T}\mathbf{Y}_\ell\mathbf{Y}_\ell^T$ and $\mathbf{R}_{\mathbf{y}_1\mathbf{y}_2} := \frac{1}{T}\mathbf{Y}_1\mathbf{Y}_2^T$ denote the sample auto-correlation matrix of $\mathbf{y}_\ell$ and the sample cross-correlation matrix of $\mathbf{y}_1$ and $\mathbf{y}_2$, respectively. Then, (2.1) can be equivalently written as

$$\max_{\mathbf{q}_1,\mathbf{q}_2} \mathbf{q}_1^T\mathbf{R}_{\mathbf{y}_1\mathbf{y}_2}\mathbf{q}_2 \tag{2.2a}$$

$$\text{s.t.} \quad \mathbf{q}_\ell^T\mathbf{R}_{\mathbf{y}_\ell\mathbf{y}_\ell}\mathbf{q}_\ell = 1,\ \ell = 1,2 \tag{2.2b}$$

where the constraints in (2.2b) arise from the fact that the objective of (2.1) is not affected by re-scaling $\mathbf{q}_1$ and/or $\mathbf{q}_2$. Using the Lagrange duality theorem, a solution of (2.2) can be provided in closed-form. The Lagrangian of (2.2) is

$$\mathcal{L}(\mathbf{q}_1,\mathbf{q}_2,\lambda_1,\lambda_2) = \mathbf{q}_1^T\mathbf{R}_{\mathbf{y}_1\mathbf{y}_2}\mathbf{q}_2 - \sum_{\ell=1}^{2}\frac{\lambda_\ell}{2}(\mathbf{q}_\ell^T\mathbf{R}_{\mathbf{y}_\ell\mathbf{y}_\ell}\mathbf{q}_\ell - 1) \tag{2.3}$$

By taking the derivatives with respect to $\mathbf{q}_1$ and $\mathbf{q}_2$, we obtain

$$\frac{\partial\mathcal{L}}{\partial\mathbf{q}_1} = \mathbf{R}_{\mathbf{y}_1\mathbf{y}_2}\mathbf{q}_2 - \lambda_1\mathbf{R}_{\mathbf{y}_1\mathbf{y}_1}\mathbf{q}_1 = 0 \tag{2.4}$$

$$\frac{\partial\mathcal{L}}{\partial\mathbf{q}_2} = \mathbf{R}_{\mathbf{y}_2\mathbf{y}_1}\mathbf{q}_1 - \lambda_2\mathbf{R}_{\mathbf{y}_2\mathbf{y}_2}\mathbf{q}_2 = 0 \tag{2.5}$$

By left multiplying (2.4) and (2.5) with $\mathbf{q}_1^T$ and $\mathbf{q}_2^T$, respectively, we have

$$\mathbf{q}_1^T\mathbf{R}_{\mathbf{y}_1\mathbf{y}_2}\mathbf{q}_2 = \lambda_1\mathbf{q}_1^T\mathbf{R}_{\mathbf{y}_1\mathbf{y}_1}\mathbf{q}_1 \tag{2.6}$$

$$\mathbf{q}_2^T\mathbf{R}_{\mathbf{y}_2\mathbf{y}_1}\mathbf{q}_1 = \lambda_2\mathbf{q}_2^T\mathbf{R}_{\mathbf{y}_2\mathbf{y}_2}\mathbf{q}_2 \tag{2.7}$$

which together with the constraints in (2.2b) imply that $\lambda_1 = \lambda_2 = \lambda$. By assuming that the matrix $\mathbf{R}_{\mathbf{y}_2\mathbf{y}_2}$ is invertible, the optimal solution, $\mathbf{q}_2^*$, of (2.5) is given by

$$\mathbf{q}_2^* = \frac{1}{\lambda}\mathbf{R}_{\mathbf{y}_2\mathbf{y}_2}^{-1}\mathbf{R}_{\mathbf{y}_2\mathbf{y}_1}\mathbf{q}_1^* \tag{2.8}$$

Then by substituting in (2.4), the optimal solution, $\mathbf{q}_1^*$, can be obtained by solving the following generalized eigenvalue problem

$$\mathbf{R}_{\mathbf{y}_1\mathbf{y}_2}\mathbf{R}_{\mathbf{y}_2\mathbf{y}_2}^{-1}\mathbf{R}_{\mathbf{y}_2\mathbf{y}_1}\mathbf{q}_1 = \lambda^2\mathbf{R}_{\mathbf{y}_1\mathbf{y}_1}\mathbf{q}_1 \tag{2.9}$$

It can be easily seen from (2.4) that the maximum eigenvalue λ^* of (2.9) is nothing but the square of the correlation coefficient, ρ_1, associated with the canonical pair $(\mathbf{q}_1^*, \mathbf{q}_2^*)$.

Considering the generalization to $N \leq \min(M_1, M_2)$ canonical pairs, $\{(\mathbf{q}_1[n], \mathbf{q}_2[n])\}_{n=1}^N$. After identifying $\mathbf{q}_1^*[1] = \mathbf{q}_1^*$ and $\mathbf{q}_2^*[1] = \mathbf{q}_2^*$, we can iteratively solve the following problem

$$\max_{\mathbf{q}_1[n], \mathbf{q}_2[n]} \quad \mathbf{q}_1^T[n]\mathbf{R}_{\mathbf{y}_1\mathbf{y}_2}\mathbf{q}_2[n] \tag{2.10a}$$

$$\text{s.t.} \quad \mathbf{q}_\ell^T[n]\mathbf{R}_{\mathbf{y}_\ell\mathbf{y}_\ell}\mathbf{q}_\ell[n] = 1, \ \ell = 1, 2 \tag{2.10b}$$

$$\mathbf{q}_\ell^T[n]\mathbf{R}_{\mathbf{y}_\ell\mathbf{y}_\ell}\mathbf{q}_\ell[j] = 0, \ j = 1, \cdots, n-1 \tag{2.10c}$$

for $n = \{2, \cdots, N\}$. Instead of solving N sub-problems of type (2.10), we can instead solve one joint problem. Let us stack the vectors $\{\mathbf{q}_\ell[n]\}_{n=1}^N$ in the matrix $\mathbf{Q}_\ell \in \mathbb{R}^{M_\ell \times N}$, for $\ell \in \{1, 2\}$, and rewrite (2.10) in the following compact form

$$\max_{\mathbf{Q}_1, \mathbf{Q}_2} \text{Tr}(\mathbf{Q}_1^T\mathbf{R}_{\mathbf{y}_1\mathbf{y}_2}\mathbf{Q}_2) \tag{2.11a}$$

$$\text{s.t.} \quad \mathbf{Q}_\ell^T\mathbf{R}_{\mathbf{y}_\ell\mathbf{y}_\ell}\mathbf{Q}_\ell = \boldsymbol{I}, \ \ell = 1, 2 \tag{2.11b}$$

which yields simultaneously multiple canonical pairs. Following the same procedures for solving (2.2), it can be shown that the optimal solution $\mathbf{Q}_1^*$ should satisfy the following generalized eigenvalue equation

$$\mathbf{R}_{\mathbf{y}_1\mathbf{y}_2}\mathbf{R}_{\mathbf{y}_2\mathbf{y}_2}^{-1}\mathbf{R}_{\mathbf{y}_2\mathbf{y}_1}\mathbf{Q}_1 = \mathbf{R}_{\mathbf{y}_1\mathbf{y}_1}\mathbf{Q}_1\Lambda^2 \tag{2.12}$$

where $\Lambda = \text{Diag}([\rho_1, \cdots, \rho_N])$ with ρ_n be the n-th correlation coefficient associated with the ℓ-th canonical pair, for $n \in [N] := \{1, \cdots, N\}$. Note that the optimal solution $\mathbf{Q}_2^*$ can be directly obtained from (2.8) after solving (2.12).

The two-view CCA problem in (2.11) can be equivalently formulated as a distance minimization between the low dimensional representations $\mathbf{Y}_1^T\mathbf{Q}_1$ and $\mathbf{Y}_2^T\mathbf{Q}_2$ [17,18], where the distance is measured by the Frobenius norm, i.e.,

$$\min_{\mathbf{Q}_1, \mathbf{Q}_2} \ \|\mathbf{Y}_1^T\mathbf{Q}_1 - \mathbf{Y}_2^T\mathbf{Q}_2\|_F^2 \tag{2.13a}$$

$$\text{s.t.} \quad \mathbf{Q}_\ell^T\mathbf{Y}_\ell\mathbf{Y}_\ell^T\mathbf{Q}_\ell = \boldsymbol{I}, \quad \ell = 1, 2 \tag{2.13b}$$

Note that by expanding the objective in (2.13), the equivalence between (2.12) and (2.13) can be readily verified. Notice that the scaling constraints serve to exclude the trivial (and meaningless) all-zero solution. Problem (2.13) is also known as the sum-of-correlations (SUMCOR) formulation of CCA as it measures the sum of pairwise correlations, hence the name.

An alternative formulation of (2.13) is to search for an orthogonal representation $\mathbf{G} \in \mathbb{R}^{T \times N}$ that is maximally correlated after the linear projections of $\mathbf{Y}_1$ and $\mathbf{Y}_2$ on $\mathbf{Q}_1$ and $\mathbf{Q}_2$, respectively. This leads to the so-called maximum-variance (MAXVAR) CCA formulation, which is

given by

$$\min_{\{\mathbf{Q}_\ell\}_{\ell=1}^L, \mathbf{G}} \sum_{\ell=1}^{2} \|\mathbf{Y}_\ell^T \mathbf{Q}_\ell - \mathbf{G}\|_F^2 \tag{2.14a}$$

$$\text{s.t.} \quad \mathbf{G}^T\mathbf{G} = \mathbf{I}. \tag{2.14b}$$

Both MAXVAR and SUMCOR formulations in (2.14) and (2.13), respectively, are equivalent in the sense that both problems yield the same optimal solutions $\mathbf{Q}_1^\star$ and $\mathbf{Q}_2^\star$ [18], for the two views case considered above.

2.2 Generalized Canonical Correlation Analysis (GCCA)

Both SUMCOR and MAXVAR formulations can be naturally generalized to the case where there are more than two random vectors (and accordingly, more than two data views) involved. This case is known as Generalized CCA (GCCA) or multiview CCA, where there are more than two data views $\{\mathbf{Y}_\ell\}_{\ell=1}^L$ and $\ell = [L] := \{1, \cdots, L\}$.

To generalize the SUMCOR formulation (2.13) to accommodate the case of multiple views ($L > 2$), it is natural to adopt a pair-wise matching criterion [17]. That is,

$$\min_{\{\mathbf{Q}_\ell\}_{\ell=1}^L} \sum_{\ell=1}^{L-1} \sum_{\ell'>\ell}^{L} \|\mathbf{Y}_\ell^T \mathbf{Q}_\ell - \mathbf{Y}_{\ell'}^T \mathbf{Q}_{\ell'}\|_F^2 \tag{2.15a}$$

$$\text{s.t.} \quad \mathbf{Q}_\ell^T \mathbf{Y}_\ell \mathbf{Y}_\ell^T \mathbf{Q}_\ell = \mathbf{I}, \ \forall\ (\ell, \ell') \in [L]. \tag{2.15b}$$

On the other hand, the general form of MAXVAR, for $L > 2$, can be written as

$$\min_{\{\mathbf{Q}_\ell\}_{\ell=1}^L, \mathbf{G}} \sum_{\ell=1}^{L} \|\mathbf{Y}_\ell^T \mathbf{Q}_\ell - \mathbf{G}\|_F^2 \tag{2.16a}$$

$$\text{s.t.} \quad \mathbf{G}^T\mathbf{G} = \mathbf{I}. \tag{2.16b}$$

Although both SUMCOR and MAXVAR formulations aim at finding highly-correlated reduced-dimension views, and their optimal solutions can be shown to coincide in the special case of $L = 2$ (CCA) as mentioned earlier, they are generally not equivalent for the multiview case ($L > 2$). While problem (2.16) introduces an additional TN variables compared to SUMCOR, it replaces the multiple constraints in (2.15) with a single orthonormality constraint on the matrix $\mathbf{G}$. It was shown that SUMCOR in (2.15) is NP–hard (in the worst case) [19, 20]. Hence, several efficient and scalable algorithms have been developed to tackle the SUMCOR GCCA to obtain high-quality approximate solutions [19–24] – also incorporating various constraints.

The MAXVAR GCCA formulation, however, admits a simple algebraic solution via eigenvalue decomposition. To see this, one can fix $\mathbf{G}$ and solve (2.16) with respect to $\mathbf{Q}_\ell$'s. Then,

upon assuming that the matrices $\{\mathbf{Y}_\ell\}_{\ell=1}^L$ are full row rank, it follows that $\mathbf{Q}_\ell^* = (\mathbf{Y}_\ell \mathbf{Y}_\ell^T)^{-1}\mathbf{Y}_\ell \mathbf{G}$. Substituting back $\mathbf{Q}_\ell^*$ in (2.16) and expanding the cost function in (2.16), one can recast (2.16) as

$$\max_{\mathbf{G}\in\mathbb{R}^{N\times K_c}} \ \mathrm{Tr}(\mathbf{G}^T\mathbf{A}\mathbf{G}) \tag{2.17a}$$

$$\text{s.t.} \quad \mathbf{G}^T\mathbf{G} = \mathbf{I}. \tag{2.17b}$$

where $\mathbf{A}$ is defined as $\mathbf{A} := \sum_{\ell=1}^L \mathbf{Y}_\ell^T(\mathbf{Y}_\ell\mathbf{Y}_\ell^T)^{-1}\mathbf{Y}_\ell$. It can be easily seen that (2.17) is nothing but an eigenvalue problem with the optimal solution $\mathbf{G}^\star$ being the first N principal eigenvectors of the matrix $\mathbf{A}$ [25].

In this dissertation, we will focus more on the MAXVAR formulation to show the potential of (G)CCA in handling several problems in wireless communications. In fact, we adopted the MAXVAR formulation over the SUMCOR one as (i) it can be optimally solved using eigendecomposition whereas SUMCOR is an NP-Hard problem when the number of views exceeds two, (ii) is computationally cheaper compared to SUMCOR, and hence, it is practically preferable, and (iii) does not suffer from initialization and tuning parameters issues as SUMCOR does. Therefore, we deem that considering MAXVAR will eventually lead to an overall much simpler approach that can be used to tackle various problems.

2.3 Complex Domain Representation of (G)CCA

In Sections 2.1 and 2.2, we presented all the CCA and GCCA related formulations in the real domain for simplicity, and also because the vast majority of (G)CCA literature only deals with the real case. In wireless communications, however, we usually deal with complex signals mainly because the communication channels are naturally belong to the complex domain. In such a case, the CCA/GCCA formulations have to be slightly modified to accommodate the complex representation. For instance, given the data views $\{\mathbf{Y}_\ell \in \mathbb{C}^{M_\ell\times T}\}_{\ell=1}^L$, the SUMCOR formulation can be expressed in the complex domain as [24],

$$\min_{\{\mathbf{Q}_\ell\}_{\ell=1}^L} \ \sum_{\ell=1}^{L-1}\sum_{\ell'>\ell}^{L} \|\mathbf{Y}_\ell^H\mathbf{Q}_\ell - \mathbf{Y}_{\ell'}^H\mathbf{Q}_{\ell'}\|_F^2 \tag{2.18a}$$

$$\text{s.t.} \quad \mathbf{Q}_\ell^H\mathbf{Y}_\ell\mathbf{Y}_\ell^H\mathbf{Q}_\ell = \mathbf{I}, \ \forall\ (\ell,\ell') \in [L]. \tag{2.18b}$$

while the MAXVAR formulation can be written as

$$\min_{\{\mathbf{Q}_\ell\}_{\ell=1}^L,\mathbf{G}} \ \sum_{\ell=1}^{L} \|\mathbf{Y}_\ell^H\mathbf{Q}_\ell - \mathbf{G}\|_F^2 \tag{2.19a}$$

$$\text{s.t.} \quad \mathbf{G}^H\mathbf{G} = \mathbf{I}. \tag{2.19b}$$

where $\mathbf{Q}_\ell \in \mathbb{C}^{M_\ell \times N}$ and $\mathbf{G} \in \mathbb{C}^{T \times N}$, $\forall \ell \in [L]$. For $L > 2$, the optimal solution $\mathbf{G}^*$ are the first N principal components of the matrix $\mathbf{A} := \sum_{\ell=1}^{L} \mathbf{Y}_\ell^H (\mathbf{Y}_\ell \mathbf{Y}_\ell^H)^{-1} \mathbf{Y}_\ell$, and $\mathbf{Q}_\ell^* = (\mathbf{Y}_\ell \mathbf{Y}_\ell^H)^{-1} \mathbf{Y}_\ell \mathbf{G}^*$. Further, we define the n-th correlation coefficient between views j and ℓ as [24],

$$\rho_{\ell j}^{(n)} = \mathbb{R}\mathrm{e}\{\mathbf{Q}_\ell^H(:,n)\mathbf{Y}_\ell \mathbf{Y}_j^H \mathbf{Q}_j(:,n)\} \tag{2.20}$$

$\forall \ell, j \in [L]$ and $j > \ell$ and $n \in [N]$.

In the subsequent chapters, we will use both the complex and real domain representations of CCA and GCCA.

2.4 New Algebraic Interpretation

With a history of more than three decades, CCA has been employed in several signal processing, communications, and machine learning applications, including array processing [26], multiple-input multiple-output (MIMO) equalization [27, 28], direction-of-arrival (DoA) estimation [29], radar anti-jamming [30, 31] and blind source separation [32–35], biomedical signal processing [36, 37], cell-edge detection [13–15], signal detection in multipath networks [38], underlay communication [16], and multi-view learning [22–24, 39], to name a few applications.

Most of the aforementioned CCA-related works relied on probabilistic interpretations of CCA [40,41]. Recently, we came up with a new and broadly useful generative model and algebraic interpretation of CCA as a method that can identify a common (shared) subspace between two matrices [13, 14]. In particular, we proved that under general (and purely deterministic) linear generative model, CCA can identify the common subspace between two matrices. Even if the common components are much weaker than the individual components associated with each data view (e.g., 40 dB below), our analysis shows that under certain mild condition, the subspace containing the common signals can be identified. In other words, we have recently shown that what CCA does is it computes the intersection $\mathrm{Range}(\mathbf{Y}_1)$ and $\mathrm{Range}(\mathbf{Y}_2)$ [13, 14]. More generally, GCCA computes $\bigcap_{i=1}^{I} \mathrm{Range}(\mathbf{Y}_i)$ [24]. If the ranges have a nontrivial intersection (the common signal of interest), then that can be perfectly recovered, even if the individual components of each matrix are overwhelmingly stronger.

In the next chapters, we will build upon our new algebraic interpretation through establishing identifiability analysis of the common subspace, followed by showing how valuable our theortical findings are, through showcasing the potential of (G)CCA in solving practical engineering problems in next wireless generation.

Chapter 3

Cell-edge Detection via Canonical Correlation Analysis

3.1 Motivation

At the dawn of 5G, providing reliable high-speed service to users on the edge between cells remains a challenge that has persisted through several generations of cellular wireless systems. In 4G and legacy systems, the problem is usually tackled using aggressive power control, multiuser detection, and dynamic base station (BS) assignment / handoff [42, 43]. Multiuser detection (MUD) is computationally complex (optimal MUD is NP-hard) [44,45], requires accurate channel estimates for all users, and while it can tolerate power imbalance, practically tractable multiuser detection does not work well in near-far scenarios, especially when the channels for the far users are not accurately known.

The so-called sphere decoder (SD – a branch-and-bound type implementation of the maximum likelihood detector) features significantly lower complexity than naive implementations at moderately high signal to noise ratios (SNRs), albeit worst-case and average complexities remain exponential [46, 47]. Semidefinite relaxation (SDR) is a polynomial-time alternative to SD, in the low to moderate SNR regime where it yields better error rates and lower complexity than SD [48, 49]. The complexity of SDR remains high for practical implementation [50].

Minimum mean square error (MMSE) [51], and the zero-forcing (ZF – also known as the decorrelating) detector are low-complexity linear detectors, whose performance remains far from optimal in general. ZF and MMSE detectors can be further improved by successively canceling the strong user signals once they are decoded – a technique referred to as successive interference cancellation (SIC), decision feedback (DF) [52,53], or 'turbo' (iterative) interference cancellation

[54].

Although all of the aforementioned detectors have been proven successful in many applications, their detection performance is contingent on the availability of accurate channel estimates. In wireless cellular systems, accurate channel estimates may be acquired for cell-center (strong) users, however, cell-edge (weak) user signals are received at low SNR due to the inverse power law relationship between received signal power and distance. This and the intra- and inter-cell interference (particularly prominent for the cell-edge users) together induce high uncertainty in the cell-edge user channel estimates, degrading their detection performance and even leading to connection drops [55, 56]. Furthermore, the frequent hand-offs of such users further complicate their situation [57]. While power control [58] and scheduling algorithms [59, 60] serve as two possible candidates that can considerably enhance cell-edge user detection performance, this comes at the expense of significantly reducing the rates of cell-center users. These are the ones with the best channels, so throttling their rate has a serious impact on the overall sum rate of the system.

This begs the question whether it is possible to reliably detect cell-edge user signals without knowing their channels or sacrificing cell-center user rates?

We will show that with a suitable base station 'interferometry' strategy inspired from machine learning, together with a well-known algebraic signal processing tool, the cell-edge user signals can be reliably decoded under mild conditions, even at low SNR and when buried under heavy intra-cell and inter-cell interference. Exploiting the fact that cell-edge user signals are *weak but common* to both base stations, while users close to a base station are unique to that base station, reliable detection is enabled by Canonical Correlation Analysis (CCA) [12, 18].

Our approach is very different from multi-user detection using base station cooperation [61], as it capitalizes on CCA. As noted earlier in Chapter 2, CCA has been employed in several signal processing, communications, and machine learning applications , including array processing [26], multiple-input multiple-output (MIMO) equalization [27,28], direction-of-arrival (DoA) estimation [29], radar anti-jamming [30] and blind source separation [32–35], and multi-view learning [39], to name a few applications; but not anywhere close to our present context.

We propose a two-stage learning based approach that leverages base station cooperation to reliably detect cell-edge user signals without knowing their channels. The idea relies on connecting canonical correlation analysis with cell-edge user detection. In the first stage, CCA is invoked to find the common subspace of two space-time matrices, containing the baseband-equivalent signals received at two base stations. A basis for this common subspace is a mixture of the cell-edge user signals. In the second stage, this mixture is unraveled in an unsupervised fashion, using a classical algebraic technique from array signal processing, namely (R)ACMA [62]. (R)ACMA exploits constant modulus structure in the transmitted cell-edge signals, owing to

digital binary/M-ary phase shift keying (BPSK or MPSK) modulation, to recover the individual cell-edge signals. Judicious experiments demonstrate that the proposed approach works remarkably well without any power control under realistic levels of intra-cell and inter-cell interference (following the urban macro scenario from the 3GPP 38.901 standard), delivering order of magnitude error rate improvements compared to 'oracle' multiuser detection that assumes perfect knowledge of the cell-center user channels. Furthermore, the proposed approach does not even require that the signals from the different base stations are synchronized – the right synchronization can be automatically determined as well.

Beyond these compelling contributions to the particular application in cellular communications considered herein, this chapter presents two notable theoretical contributions of broader interest. First, we prove that CCA identifies the common subspace between two matrices, under a rather general (and purely deterministic) linear generative model. Second, we include a performance analysis which shows that CCA works even in the non-ideal case where there is background noise and 'leakage' of the individual components to the other matrix view – e.g., the case where there is thermal noise and realistic adjacent-cell interference from non-cell-edge users that cannot be neglected, in the context of our application herein.

The overall complexity of the proposed method depends on the cost incurred in solving CCA and RACMA. Fortunately, both admit relatively simple algebraic solution via eigenvalue decomposition [18, 62]. This renders the overall approach computationally efficient even when the base station is equipped with a large number of antennas and is serving a large number of users.

The rest of this chapter is organized as follows. Section 3.2 describes the system model and gives a brief review on cell-edge user detection. The proposed detector is presented in Section 3.3, while Section 3.4 explains how our detector can be used to resolve symbol synchronization between the two base stations. Simulation results are provided in Section 3.5, and conclusions are drawn in Section 3.6. Long proofs and derivations are relegated to the Appendix.

3.2 Problem Statement

3.2.1 System Model

Consider a multi-cell multi-user MIMO system comprising two hexagonal cells with a single base station (BS) located at the center of each cell, as shown in Figure 4.1. The ℓ-th BS is equipped with M_ℓ antennas and serves K_ℓ single-antenna users, for $\ell \in \{1, 2\}$. Let $K_e = K_{e_1} + K_{e_2}$ denote the total number of cell-edge users located around the common edge between the two cells, where $K_{e_\ell} < K_\ell$ represents the number of cell-edge users served by the ℓ-th BS. Let $\mathbf{h}_{\ell kj}$ model path-loss and small scale fading between the k-th user in the j-th cell and the ℓ-th BS,

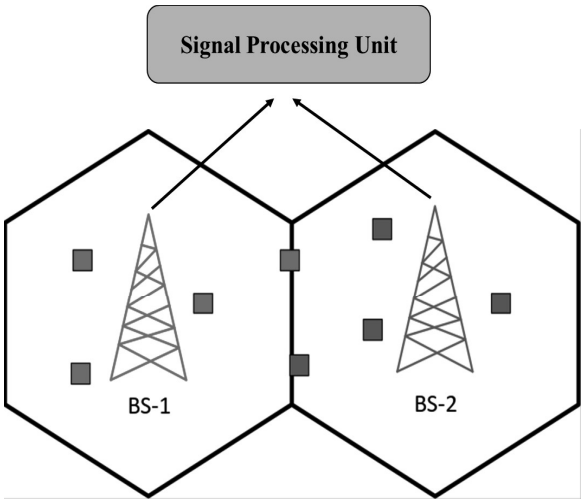


Figure 3.1: System model of two base stations setup serving cell-edge and cell-center users.

given by

$$\mathbf{h}_{\ell kj} = \sqrt{\alpha_{\ell kj}}\mathbf{z}_{\ell kj} \tag{3.1}$$

where $\mathbf{z}_{\ell kj} \in \mathbb{C}^{M_\ell \times 1}$ represents the small scale fading between user k in cell j and BS ℓ, while $\alpha_{\ell kj} \in \mathbb{R}$ models the large scale fading that accounts for the path loss between BS ℓ and user k in cell j. Throughout this work, it is assumed that the uplink channel vectors $\mathbf{h}_{\ell kj}$ for the cell-edge users are not *a priori* known at any BS.

3.2.2 Uplink Transmission

Consider uplink transmission from the users to the BSs where each user aims at transmitting its data to its serving BS. We assume that all users access the same channel without any (sub-) channel allocation or coordination mechanism, thereby creating intra- and inter-cell interference. Define $\mathbf{s}_{kj} \in \mathbb{R}^{T\times 1}$ to be the vector containing symbols transmitted by the k-th user in cell j, where each entry of $\mathbf{s}_{kj}$ belongs to the finite alphabet $\Omega = \{\pm 1\}$ (our approach works for general PSK and other alphabets, with some variations in the second stage). The received signal, $\mathbf{Y}_\ell \in \mathbb{C}^{M_\ell \times T}$, at the ℓ-th BS can be expressed as

$$\mathbf{Y}_\ell = \sum_{j=1}^{2}\sum_{k=1}^{K_j} \sqrt{\beta}_{kj}\mathbf{h}_{\ell kj}\mathbf{s}_{kj}^T + \mathbf{W}_\ell \tag{3.2}$$

where $\mathbf{h}_{\ell kj} \in \mathbb{C}^{M_\ell \times 1}$ is the uplink channel response vector defined in (3.1), $\mathbf{W}_\ell \in \mathbb{C}^{M_\ell \times T}$ contains independent identically distributed (i.i.d.) complex Gaussian entries of zero mean and variance σ^2, and β_{kj} represents the transmit power of the k-th user in the j-th cell.

In this work, we assume that each BS forwards its received signal to a central signal processing unit (CSPU). Although BSs cooperation has been considered before for the sake of mitigating inter-cell interference [63], cooperation here is assumed for a very different purpose. That is, we leverage the joint processing of the BSs signals at the CSPU to provide reliable detection of cell-edge user signals at low SNR, without knowledge of their channels. Furthermore, in contrast to prior cooperation strategies that assume perfect synchronization of the received signals from different BSs [61, 64], this work deals with BS asynchrony as well, rendering the approach more practical. Specifically, it will be shown in Section 3.4 how the proposed method can detect the cell-edge user signals even if there exists a time delay between $\mathbf{Y}_1$ and $\mathbf{Y}_2$.

3.2.3 Cell-edge Challenges

Let us denote the cell-edge user transmitted signals by $\mathbf{S}_c \in \mathbb{R}^{T\times K_e}$ (where the subscript c stands for *common*), and those of the cell-center served by the ℓ-th BS as $\mathbf{S}_{p_\ell} \in \mathbb{R}^{T\times(K_\ell - K_{e_\ell})}$ (where the subscript p stands for *private*). Furthermore, let $\widetilde{\mathbf{W}}_\ell$ represent the noise at the ℓ-th BS plus the inter-cell interference caused by the cell-center users in cell j, where $j \neq \ell$. Therefore, (4.2) can be expressed as follows

$$\mathbf{Y}_\ell = \mathbf{H}_{\ell p_\ell}\mathbf{S}_{p_\ell}^T + \mathbf{H}_{\ell c}\mathbf{S}_c^T + \widetilde{\mathbf{W}}_\ell \tag{3.3}$$

where the matrices $\mathbf{H}_{\ell c} \in \mathbb{C}^{M_\ell \times K_e}$ and $\mathbf{H}_{\ell p_\ell} \in \mathbb{C}^{M_\ell \times (K_\ell - K_{e_\ell})}$ hold on their columns all the channel vectors from cell-edge users to the ℓ-th BS, and the channel vectors from cell-center users to their serving BS, respectively. Moreover, absorb the transmitted signal power, β_{kj}, of the k-th user in the j-th cell in its respective channel vectors, $\forall\ k, j$.

In general, to guarantee reliable detection performance for each cell-edge user, its serving BS requires accurate knowledge about its channel state information (CSI) [65–67]. However, due to the fact that cell-edge user signals are often received intermittently at very low signal to interference plus noise ratio (SINR) and SNR, their channel estimates are inaccurate [55, 56].

One possible approach to detect cell-edge user signals is to apply zero-forcing successive interference cancellation (ZF-SIC) [52], which is based on successively removing the cell-center (strong) user signals once they are detected using ZF. Applying SIC after ZF improves the detection performance of the cell edge user signals as it (ideally) cancels the strong interference that stems from the transmissions of cell-center users, i.e., intra-cell interference. However, cell-center user detection is imperfect, which can lead to error propagation, and in-cell SIC does not address the inter-cell interference, which is particularly prominent for the cell-edge users. In the absence of power control [58] and/or scheduling [59] , cell-edge user detection performance is severely affected by the intra-cell interference from cell-center users. In what follows, we present a novel *blind* detector that can reliably decode cell-edge user signals at low received SNR and without knowing their channels.

3.3 Cell-Edge User Detection via CCA

In this section, it is assumed that the base stations signals are perfectly synchronized at the CSPU. We will explain how to deal with asynchrony later. The goal of the proposed detector is to decode cell-edge user signals $\mathbf{S}_c$ from the received signals $\mathbf{Y}_1$ and $\mathbf{Y}_2$. As a pre-processing step, the signals are transformed to the real domain by forming the matrix $\overline{\mathbf{Y}}_\ell := [\mathbf{Y}_\ell^{(r)}; \mathbf{Y}_\ell^{(i)}] \in \mathbb{R}^{2M_\ell \times T}$, where $\mathbf{Y}_\ell^{(r)} = \mathbb{R}\mathrm{e}\{\mathbf{Y}_\ell\}$ and $\mathbf{Y}_\ell^{(i)} = \mathbb{I}\mathrm{m}\{\mathbf{Y}_\ell\}$ represent the real and imaginary components of the ℓ-th BS signal. Similarly, denote by $\overline{\mathbf{A}}_{\ell p_\ell} := [\mathbf{H}_{\ell p_\ell}^{(r)}; \mathbf{H}_{\ell p_\ell}^{(i)}] \in \mathbb{R}^{2M_\ell \times (K_\ell - K_{e_\ell})}$, $\overline{\mathbf{A}}_{\ell c} := [\mathbf{H}_{\ell c}^{(r)}; \mathbf{H}_{\ell c}^{(i)}] \in \mathbb{R}^{2M_\ell \times K_e}$ and $\overline{\mathbf{W}}_\ell = [\widetilde{\mathbf{W}}_\ell^{(r)}; \widetilde{\mathbf{W}}_\ell^{(i)}] \in \mathbb{R}^{2M_\ell \times T}$. Therefore, (3.3) can be equivalently expressed as

$$\overline{\mathbf{Y}}_\ell = \overline{\mathbf{A}}_{\ell p_\ell} \mathbf{S}_{p_\ell}^T + \overline{\mathbf{A}}_{\ell c} \mathbf{S}_c^T + \overline{\mathbf{W}}_\ell. \tag{3.4}$$

Remark 1. *Due to the broadcast nature of the wireless medium, each BS may (over)hear the transmitted signals of all users. However, due to the inverse relationship between power and distance, the received signal power of the cell-center users associated with the ℓ-th BS is high at the ℓ-th BS (the serving one) and low at the j-th BS (the non-serving one). This power imbalance renders the received SNR of these users to be high at the serving BS and very low at the non-serving one, and hence, one can think of these users as being "private" to their serving BS, as their received signals are around the noise floor at the non-serving BS. On the other hand, cell-edge users are approximately half-way between two different BSs, and so they are received at commensurate power at both BSs. In this sense, cell-edge users are "common" to both BSs. In what follows, we will show theoretically and experimentally that our proposed CCA-based approach can reliably recover these common cell-edge user signals under realistic conditions.*

In what follows, the two-view MAXVAR CCA formulation in (2.14) is exploited to estimate the subspace containing the cell-edge user signals. For the sake of brevity, we refer to this subspace as the *common subspace*. Define the two matrices $\overline{\mathbf{Q}}_1 \in \mathbb{R}^{2M_1 \times N}$ and $\overline{\mathbf{Q}}_2 \in \mathbb{R}^{2M_2 \times N}$, where the n-th column of $\overline{\mathbf{Q}}_\ell$ represents the n-th canonical component of view $\overline{\mathbf{Y}}_\ell$, for $n \in [N] := \{1, \cdots, N\}$. The number of components (pairs) extracted, (N), depends on the minimum value of the correlation coefficient that needs to be considered. Towards this end, we feed the two matrices $\overline{\mathbf{Y}}_1$ and $\overline{\mathbf{Y}}_2$ to the MAXVAR formulation as follows,

$$\min_{\overline{\mathbf{Q}}_1, \overline{\mathbf{Q}}_2, \mathbf{G}} \sum_{\ell=1}^{2} \|\overline{\mathbf{Y}}_\ell^T \overline{\mathbf{Q}}_\ell - \mathbf{G}\|_F^2 \tag{3.5a}$$

$$\text{s.t.} \qquad \mathbf{G}^T \mathbf{G} = \boldsymbol{I} \tag{3.5b}$$

Assume that we are interested in the first K_e canonical components of the matrices $\overline{\mathbf{Q}}_1$ and $\overline{\mathbf{Q}}_2$, i.e., $N = K_e$. We have the following result.

Theorem 1. *In the noiseless case, if matrix* $\mathbf{B} := [\mathbf{S}_c, \mathbf{S}_{p_1}, \mathbf{S}_{p_2}] \in \mathbb{R}^{T\times(K_1+K_2)}$ *is full column rank, and* $\overline{\mathbf{A}}_\ell = [\overline{\mathbf{A}}_{\ell c}, \overline{\mathbf{A}}_{\ell p_\ell}] \in \mathbb{R}^{2M_\ell\times(K_e+K_\ell-K_{e_\ell})}$ *is full column rank for* $\ell \in \{1,2\}$*, then the optimal solution* $\mathbf{G}^\star$ *of problem* (3.5) *is given by* $\mathbf{G}^\star = \mathbf{S}_c\mathbf{P}$*, where* $\mathbf{P}$ *is a* $K_e \times K_e$ *non-singular matrix.*

Remark 2. *The full column rank condition on* $\mathbf{B}$ *requires* T *greater than or equal to* (K_1+K_2)*, and the transmitted sequences from the different users to be linearly independent. For finite-alphabet signals, this occurs with very high probability for modest* T*, since the different user transmissions are independent. The more restrictive condition is full column rank of* $\overline{\mathbf{A}}_\ell$*, which relates the number of base station antennas and signals impinging on each base station. We thus need two times the number of antennas in each base station to be greater than or equal to the number of users assigned to that base station, plus any cell-edge users assigned to the other base station. Other than this dimensionality constraint though, if the channel vectors are drawn from a jointly continuous distribution, the latter condition will be satisfied with probability one.*

Proof. First, let us start with the single cell-edge user case, i.e., $K_e = 1$ and each of $\mathbf{S}_c, \mathbf{G}$ and $\mathbf{Q}_\ell$ is a vector. In such setting (3.5) reduces to the following

$$\min_{\overline{\mathbf{q}}_1,\overline{\mathbf{q}}_2,\mathbf{g}} \sum_{\ell=1}^{2} \|\overline{\mathbf{Y}}_\ell^T\overline{\mathbf{q}}_\ell - \mathbf{g}\|_2^2 \tag{3.6a}$$

$$\text{s.t.} \quad \|\mathbf{g}\|_2^2 = 1 \tag{3.6b}$$

To solve the above problem, we need to find $(\overline{\mathbf{q}}_1^*, \overline{\mathbf{q}}_2^*, \mathbf{g}^*)$ that can together attain a zero-cost. In other words, we need the following two conditions to be satisfied simultaneously

$$\overline{\mathbf{Y}}_1^T\overline{\mathbf{q}}_1 = \mathbf{g} \tag{3.7a}$$

$$\overline{\mathbf{Y}}_2^T\overline{\mathbf{q}}_2 = \mathbf{g} \tag{3.7b}$$

Without loss of generality, we can let $\overline{\mathbf{q}}_\ell = \overline{\mathbf{A}}_\ell(\overline{\mathbf{A}}_\ell^T\overline{\mathbf{A}}_\ell)^{-1}\mathbf{u}_\ell$, where $\mathbf{u}_\ell$ is any vector in $\mathbb{R}^{K_e+K_\ell-K_{e_\ell}}$. The reason is that we can always decompose $\overline{\mathbf{q}}_\ell$ into a component in the subspace spanned by $\overline{\mathbf{A}}_\ell$ and one orthogonal to it. The latter is annihilated anyway after multiplication with $\overline{\mathbf{A}}_\ell^T$. Substituting in (3.7a) and (3.7b) and taking their difference, we obtain

$$\mathbf{B}\mathbf{u} = \mathbf{0}, \tag{3.8}$$

where $\mathbf{B} = [\mathbf{s}_c, \mathbf{S}_{p_1}, \mathbf{S}_{p_2}] \in \mathbb{R}^{T\times(K_1+K_2)}$ and $\mathbf{u} = [\mathbf{u}_1(1) - \mathbf{u}_2(1), \mathbf{u}_1(2:\text{end}), -\mathbf{u}_2(2:\text{end})]^T \in \mathbb{R}^{(K_1+K_2)}$, where $\mathbf{u}_1(2:\text{end})$ is the vector containing all except the first element of $\mathbf{u}$. It can be easily seen that if $\mathbf{B}$ is full column rank, then $\mathbf{u} = \mathbf{0}_{(K_1+K_2)\times 1}$ is the only possible solution of (3.8). This means that $\mathbf{u}_1 = \mathbf{u}_2 = c\mathbf{e}_1$, where c is any constant and $\mathbf{e}_1$ is the first column of

the identity matrix. Consequently, from (3.7), $\mathbf{g}^\star = \alpha \mathbf{s}_c / \|\mathbf{s}_c\|_2$, with $\alpha = \pm 1$, will be the only possible solution for problem (3.6).

The generalization to $K_e > 1$ now follows naturally. Letting $\overline{\mathbf{Q}}_\ell = \overline{\mathbf{A}}_\ell (\overline{\mathbf{A}}_\ell^T \overline{\mathbf{A}}_\ell)^{-1} \mathbf{U}_\ell$, and defining

$$\mathbf{U} := \begin{bmatrix} & \mathbf{U}_1(1:K_e,:) - \mathbf{U}_2(1:K_e,:) \\ & \mathbf{U}_1(K_e+1:\text{end},:) \\ - & \mathbf{U}_2(K_e+1:\text{end},:) \end{bmatrix} \in \mathbb{R}^{(K_1+K_2)\times K_e},$$

where $\mathbf{U}_1(1:K_e,:)$ means rows 1 to K_e and all columns of $\mathbf{U}_1$, we obtain

$$\mathbf{B}\mathbf{U} = \mathbf{0}, \tag{3.9}$$

and when $\mathbf{B}$ is full column rank the solution is unique: $\mathbf{U} = \mathbf{0}$, and therefore $\mathbf{U}_1(1:K_e,:) = \mathbf{U}_2(1:K_e,:) =: \mathbf{P}$, $\mathbf{U}_1(K_e+1:\text{end},:) = \mathbf{0}$, $\mathbf{U}_2(K_e+1:\text{end},:) = \mathbf{0}$, and therefore $\mathbf{G}^\star = \mathbf{S}_c\mathbf{P}$, where $\mathbf{P}$ is $K_e \times K_e$ non-singular such that the orthonormality constraint (3.5b) is satisfied. Note that if the signals themselves are (approximately) orthogonal, then $\mathbf{P}$ will be orthogonal as well, which helps with the next (RACMA) stage. □

Remark 3. *Theorem 1 provides results for an idealized scenario, where at each BS we ignore other-cell signals coming from users that are not close to the given BS's cell-edge. This is a reasonable approximation that we use to prove that the cell-edge signals can be recovered even when buried under very strong cell-center signals. In other words, Theorem 1 says that if we have two multi-antenna signal "views" that include very strong but private components (in our context, the received signals of each group of cell-center users at their serving BS, respectively) and weak but common components (in our context, the received signals of the cell-edge users between the two BSs), then CCA will exactly recover the subspace of the common components irrespective of their relatively low power. We will later present an elegant analysis which shows that what matters is the power (im)balance: signals received at roughly the same power at the two BSs are "common" and recovered via CCA, and signals received at high SNR at one BS and low SNR at the other BS are "private", and cannot be recovered by CCA.*

The next step is to extract the cell-edge user sequences $\mathbf{S}_c$ from $\mathbf{G}^\star = \mathbf{S}_c\mathbf{P}$. This problem can be viewed as a bi-linear factorization of the matrix $\mathbf{G}^\star$ to its factors $\mathbf{P}$ and $\mathbf{S}_c$ under the constraint that the entries of $\mathbf{S}_c$ belong to the finite alphabet $\Omega = \pm 1$. This can be mathematically posed as an optimization problem as follows

$$\min_{\overline{\mathbf{S}}_c, \overline{\mathbf{P}}} \ \|\mathbf{G}^\star - \overline{\mathbf{S}}_c \overline{\mathbf{P}}\|_F^2 \tag{3.10a}$$

$$\text{s.t.} \quad \overline{\mathbf{S}}_c(i,j) \in \Omega \tag{3.10b}$$

In [62], van der Veen proposed an algebraic algorithm called Real Analytical Constant Modulus Algorithm (RACMA) for this problem. RACMA does not claim to optimally solve (3.10), which is NP-hard even if $\mathbf{P}$ is known. Instead, RACMA assumes that noise is small, and reduces (4.10) to a generalized eigenvalue problem. The solution is subject to sign and user permutation ambiguity. This means that the original $\mathbf{S}_c$ can be identified up to permutations and column-wise (user) scaling by ± 1. From the practical point of view, each user has its unique identification sequence, so once the users signals are received correctly each BS can identify each user signal (and sign) via correlation with the identification sequence.

The following Algorithm describes the two-step procedure for cell-edge users detection via CCA followed by RACMA.

Algorithm 1 CCA for Cell-Edge User Detection

Input: $\overline{\mathbf{Y}}_1, \overline{\mathbf{Y}}_2$

1. Solve problem (3.5) for $\overline{\mathbf{Q}}_\ell$ as explained in Section 2.1
2. Compute $\mathbf{G}_\ell = \overline{\mathbf{Y}}_\ell^T \overline{\mathbf{Q}}_\ell \in \mathbb{R}^{T \times K_e},\ \forall \ell = 1, 2$
3. Construct $\mathbf{G} = [\mathbf{G}_1; \mathbf{G}_2] \in \mathbb{R}^{2T \times K_e}$ and pass it to RACMA
4. Compute the BER of cell-edge users by comparing the output of RACMA with the original cell-edge user transmitted sequences

---

Notice that the second step in Algorithm 1 stems out from the fact that the zero-cost solution of problem (3.5) is not guaranteed in the noisy case, and therefore, $\overline{\mathbf{Y}}_1^T \overline{\mathbf{Q}}_1$ is not equal to $\overline{\mathbf{Y}}_2^T \overline{\mathbf{Q}}_2$ in general. Then, it turns out that feeding RACMA with both $\overline{\mathbf{Y}}_1^T \overline{\mathbf{Q}}_1$ and $\overline{\mathbf{Y}}_2^T \overline{\mathbf{Q}}_2$ simultaneously results in much better BER as we will see in Section 3.5.

The overall complexity of the proposed method comes from solving problems (3.5) and (3.10). Fortunately, similar to (3.5), (3.10) also admits simple algebraic solution via eigenvalue decomposition [62]. This means that our end-to-end method requires solving two eigenvalue problems, i.e., the overall complexity is of $O(M^3)$, with $M = \max\{M_1, M_2\}$.

It is important to emphasize that, in the noisy case and under inter-cell interference (i.e., users close to base station B can be overheard at base station A), it turns out that our method can still identify the common subspace, even at low SNR values. In order to show this, we follow a very different path from that is described in the proof of Theorem 1.

First, let us define two channel matrices $\mathbf{H}_1 := [\mathbf{H}_{1c}, \mathbf{H}_{1p_1}, \mathbf{H}_{1p_2}] \in \mathbb{C}^{M_1 \times (K_1 + K_2)}$ and $\mathbf{H}_2 := [\mathbf{H}_{2c}, \mathbf{H}_{2p_1}, \mathbf{H}_{2p_2}] \in \mathbb{C}^{M_2 \times (K_1 + K_2)}$, where $\mathbf{H}_\ell$ holds in its columns all the channel vectors from all users to the ℓ-th BS, for $\ell = 1, 2$. Note that one can factor $\mathbf{H}_\ell = \mathbf{Z}_\ell \mathbf{P}_\ell^{1/2}$, where the

columns of $\mathbf{Z}_\ell$ are the channel vectors representing small scale fading between the ℓ-th BS and all users. Accordingly, the diagonal matrix $\mathbf{P}_\ell$ incorporates the transmitted power and the path-loss between each user and the ℓ-th BS. We have the following result.

Proposition 1. *In the noisy and inter-cell interference case, if* $\frac{1}{T}\mathbf{B}^T\mathbf{B} \approx \mathbf{I}$ *and* $\mathbf{Z}_\ell^H\mathbf{Z}_\ell \approx \mathbf{I}$*, then under certain conditions on the relative SNRs of cell-center and cell-edge users (see the Appendix), the optimal solution* $\mathbf{Q}_\ell^\star$ *of problem* (3.5) *is given by* $\mathbf{Q}_\ell^\star = \mathbf{Z}_\ell\mathbf{V}\mathbf{M}_\ell$*, where* $\mathbf{V}$ *contains the first* K_e *columns of an* $(K_1+K_2)\times(K_1+K_2)$ *identity matrix, and* $\mathbf{M}_\ell$ *is a* $K_e \times K_e$ *non-singular matrix.*

Proof. The proof is relegated to Appendix A.1. □

The approximate semi-$\perp$ constraint on the matrix $\mathbf{B}$ posits that the transmitted sequences of different users are approximately orthogonal. For binary signals, this occurs with high probability for large enough T, since the user transmissions are independent. The approximate orthonormality constraint on $\mathbf{Z}_\ell$ requires the number of base station antennas to be greater than the total number of users assigned to both base stations, and is satisfied if, for example, the entries of $\mathbf{Z}_\ell$ are drawn from an i.i.d. zero-mean complex Gaussian distribution with variance $1/M_\ell$ (which is often assumed in the case of rich scattering).

Remark 4. *Recall that, in Theorem 1 we considered the MAX-VAR formulation in* (3.5)*, and showed that under certain conditions the optimal solution* $\mathbf{G}^*$ *is the subspace containing the cell-edge user signals. In Proposition 1, however, instead of directly estimating the common subspace, we will consider the generalized eigenvalue problem in* (2.12) *to solve for* $\mathbf{Q}_\ell^*$*, for* $\ell = 1, 2$*. Then, we will show in Appendix A.1 that upon applying the resulting* $\mathbf{Q}_\ell^*$ *to the corresponding received signals at BS* ℓ*, we obtain the common subspace corrupted by reduced noise (see* (A.17) *in Appendix A.1). It is worth mentioning that when the noise is absent,* (A.17) *boils down to the result we have in the Theorem 1.*

3.4 Synchronization

In Section 3.3, we proposed a learning-based approach that can identify cell-edge user signals. However, this was under the assumption that the received signals from both BSs are perfectly synchronized at the CSPU. One natural question that can be posed is what if there exists a time delay τ_d between $\overline{\mathbf{Y}}_1$ and $\overline{\mathbf{Y}}_2$ at the CSPU. It turns out that our proposed method not only can recover cell-edge user signals in the synchronized case, it can even detect the time difference, τ_d, between the two signals, re-synchronize the signals and then decode them as explained in Section 3.3.

Assume that the CSPU has received two long sequences $\widetilde{\mathbf{Y}}_1 \in \mathbb{R}^{2M_1 \times \widetilde{T}}$ and $\widetilde{\mathbf{Y}}_2 \in \mathbb{R}^{2M_2 \times \widetilde{T}}$, where $\widetilde{T} > T$, and that the sequence length T is known or has been estimated [68] at the CSPU. The goal is to find the correct delay between the signals $\widetilde{\mathbf{Y}}_1$ and $\widetilde{\mathbf{Y}}_2$ so that we can extract the desired signal $\overline{\mathbf{Y}}_\ell$ from $\widetilde{\mathbf{Y}}_\ell$, and then apply Algorithm 1 to identify cell-edge user signals. Exploiting the fact that communication signals are uncorrelated in time, and thus two copies of the same signal shifted by even one symbol are already uncorrelated, common user signals cannot be extracted via CCA if the $\widetilde{T}$ symbols are misaligned. The correlation coefficient, ρ_n, associated with each pair of canonical directions of $\overline{\mathbf{Q}}_1$ and $\overline{\mathbf{Q}}_2$ will not be at its maximum in this case, $\forall n \in \{1, \cdots, K_e\}$. Based on this key observation, we develop a CCA based algorithm that can re-synchronize and then recover cell-edge user signals.

Define $\widetilde{\mathbf{Y}}_1(\tau_1) := \widetilde{\mathbf{Y}}_1(:, \tau_1 : T + \tau_1 - 1)$ and $\widetilde{\mathbf{Y}}_2(\tau_2) := \widetilde{\mathbf{Y}}_2(:, \tau_2 : T + \tau_2 - 1)$. Furthermore, let us define a search window of size $[w_L, w_R]$ symbols. Upon setting $\tau_2 = 1$, the CSPU solves problem (3.5) using the signals $\widetilde{\mathbf{Y}}_1(\tau_1)$ and $\widetilde{\mathbf{Y}}_2(\tau_2)$ to obtain $\overline{\mathbf{q}}_1^* := \mathbf{Q}_1^*(:, 1)$ and $\overline{\mathbf{q}}_2^* := \overline{\mathbf{Q}}_2^*(:, 1)$. Then, the CSPU computes and stores the corresponding correlation coefficient ρ_1 between $\widetilde{\mathbf{Y}}_1^T(\tau_1)\overline{\mathbf{q}}_1^*$ and $\widetilde{\mathbf{Y}}_2^T(\tau_2)\overline{\mathbf{q}}_2^*$. If $\tau_2 \leq w_s$, increment τ_2 and repeat, where $w_s := w_R - w_L + 1$ is the window size. Finally, pick the value τ_2^* that gives the highest ρ_1. This procedure is summarized in Algorithm 2.

Remark 5. *Note that as the locations of the T symbols are not generally known, the value of τ_1 is chosen such that $\widetilde{\mathbf{Y}}_1(\tau_1)$ includes a sufficient part of $\overline{\mathbf{Y}}_1$. This is guaranteed with a very high probability as long as $w_s << T$. One possible choice is to set $\tau_1 = \widetilde{T}/2 - T/2$ so that one can assure the existence of enough samples from all users in $\widetilde{\mathbf{Y}}_1(\tau_1)$.*

Algorithm 2 CCA SYNC

Input: $\widetilde{\mathbf{Y}}_1 \in \mathbb{R}^{M_1 \times \widetilde{T}}$, $\widetilde{\mathbf{Y}}_2 \in \mathbb{R}^{M_2 \times \widetilde{T}}$

Initialization: $\tau_1 = \widetilde{T}/2 - T/2, \tau_2 := 1$

while $\tau_2 \leq w_s$ **do**

  Compute ρ_1 after solving (3.5) using $\widetilde{\mathbf{Y}}_1(\tau_1)$ and $\widetilde{\mathbf{Y}}_2(\tau_2)$

  Store (τ_2, ρ_1) in a stack

  Set $\tau_2 := \tau_2 + 1$

end

Selection: pick the $\tau_2^* := \tau_2$ corresponding to the highest ρ_1.

While Algorithm 2 returns the correct shift, $\tau_d = \tau_2^* - \tau_1$, between the two sequences, the common part in both $\widetilde{\mathbf{Y}}(\tau_1)$ and $\widetilde{\mathbf{Y}}(\tau_2^*)$ is not necessarily of length T since $\widetilde{\mathbf{Y}}(\tau_1)$ may not be equal to $\overline{\mathbf{Y}}_1$. However, from the practical point of view, each user has its own identification sequence as a preamble, so once we know the correct relative delay, we can run algorithm 1 on

$\widetilde{\mathbf{Y}}(\tau_1)$ and $\widetilde{\mathbf{Y}}(\tau_2^*)$, and then simply find the sample index at which $\overline{\mathbf{Y}}_1$ starts via correlation with the identification sequence of any of the cell-edge users. It is worth mentioning that the computational complexity of Algorithm 2 is that of solving for the principal component (canonical pair) of (3.5) a number of times (equal to the search window). The first canonical pair can be cheaply computed via a power iteration.

3.5 Experimental Results

To evaluate the performance of our proposed method, we consider a scenario with two hexagonal cells; each with radius $R = 500$ meters. Cell edge users are dropped randomly around the common edge between the two cells, i.e., the locations of cell-edge users were chosen randomly between $0.95R$ and $1.05R$. On the other hand, cell-center users are randomly dropped within distance zR from their serving BS, and we vary the value of z to see the effect of inter-cell interference on the proposed method. The transmitted power β_{kj} is set to 25 dBm, $\forall k, j$, i.e., power control is not employed. Furthermore, the transmitted sequence length T is fixed to 800. Additive white Gaussian noise is assumed with variance σ^2 so that the SNR is P_e/σ^2, where P_e is the average received power of cell-edge users. This enables us to see what values of SNR should cell-edge users have to achieve a specific BER. Furthermore, all results are averaged over 1000 channel realizations assuming different user locations in each realization. The uplink channel response vectors $\{\mathbf{h}_{\ell kj}^H\}$ are modeled as

$$\mathbf{h}_{\ell kj}^H = \sqrt{\frac{1}{M_\ell}} \sum_{n=1}^{L} \sqrt{\alpha_{\ell kj}^{(n)}} \mathbf{a}_r(\theta^{(n)})^H \tag{3.11}$$

where L is the number of paths between the ℓ-th BS and the k-th user in cell j, $\forall\{\ell, j\} \in \{1, 2\}$ and $k \in \{1, \cdots, K_\ell\}$. We use the path-loss model of the urban macro (UMa) scenario from the 3GPP 38.901 standard to compute the complex path gain $\alpha_{\ell kj}^{(n)}$, $\forall n, \ell, j, k$. Cell-center users were allowed to have a line of sight (LOS) path according to the LOS probability in the 3GPP 38.901 standard, however, all cell-edge users were non-LOS. The term $\mathbf{a}_r(.)$ is the array response vector at the BS, and $\theta^{(n)} \sim \mathcal{U}[-\pi, \pi]$ denotes the azimuth angle of arrival of the n-th path. Assuming the BS is equipped with a uniform linear array, then

$$\mathbf{a}_r(\theta) = [1, \exp^{ikd\cos(\theta)}, \cdots, \exp^{ikd(M-1)\cos(\theta)}] \tag{3.12}$$

where $k = 2\pi/\lambda$, λ is the carrier wavelength and $d = \lambda/2$ is the spacing between antenna elements.

In order to benchmark the performance of our proposed method, we adopted three baselines. First, we implemented zero-forcing successive interference cancellation (ZF-SIC) where

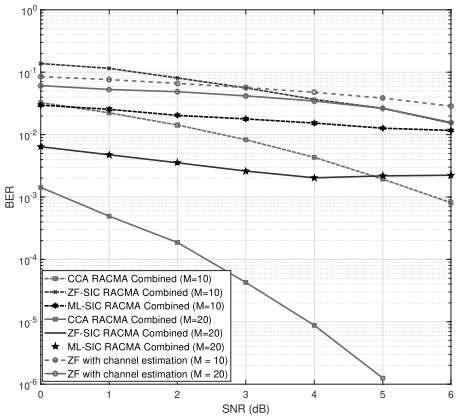


Figure 3.2: BER vs. SNR of cell-edge users, $M_1 = M_2 = 10$, $K_1 = K_2 = 8$ and $K_e = 2$, distance of cell-center users < 0.3R.

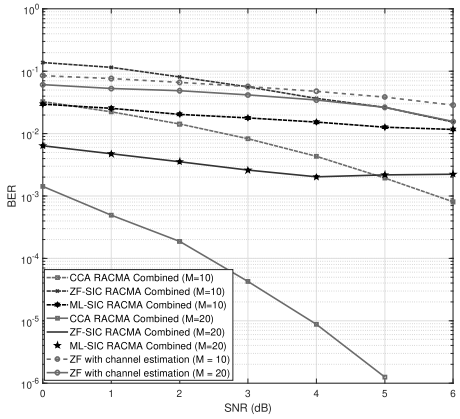


Figure 3.3: BER vs. SNR of cell-edge users, $M_1 = M_2 = M$, $K_1 = K_2 = 8$ and $K_e = 2$, distance of cell-center users < 0.3R.

the channels of the cell-center users were assumed to be perfectly known at their serving BSs. Specifically, each BS decodes its cell-center users signals using ZF, encodes them again and then subtracts them from its received signal. Afterwards, the residual signal from each BS will be passed to RACMA [62] in order to identify cell-edge user signals. Finally, the bit error rate (BER) of the cell-edge users is computed at both BSs and the best of the two BERs is reported. Furthermore, in order to guarantee fairness, since we have assumed joint processing of the BSs received signals, both residual signals from both BSs are further sent simultaneously to RACMA and the resulting BER (from RACMA with "double measurements") is also reported. Second, we estimated the channels of cell-center users and cell-edge users via transmitting orthogonal pilot sequences of length 300 each, then we used a ZF detector at each BS and reported the best of the two BERs computed at the two BSs. Third, we implemented maximum likelihood successive interference cancellation (ML-SIC) to decode and subtract cell-center users signals assuming perfect knowledge of their channels at their serving BS. However, since in the worst-case the ML detector requires enumeration over all possible sequences of cell-center users, we only used this baseline when the number of cell-center users is small. The CCA approach (first stage) was implemented in MATLAB, while the MATLAB codes written by A.-J. van der Veen [62] were utilized for the RACMA (second stage) implementation.

In a preliminary experiment, we consider a scenario with $K_1 = K_2 = 8$, $M_1 = M_2 = 10$,

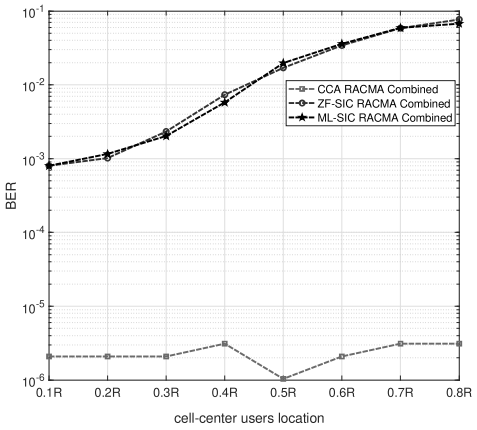


Figure 3.4: BER vs. distance of cell-center users from their serving BS, with $M_1 = M_2 = 20$, $K_1 = K_2 = 8$ and $K_e = 2$, SNR = 5dB.

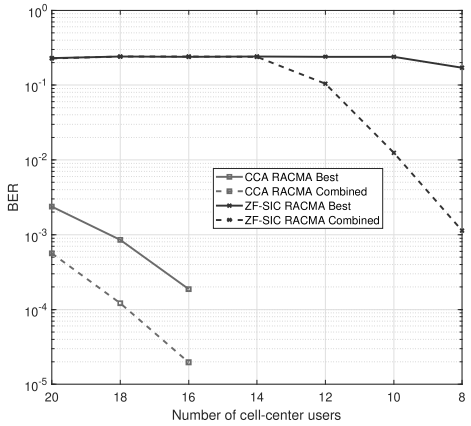


Figure 3.5: BER vs. number of cell-center users at each BS, $M_1 = M_2 = 30$, SNR = 4dB and $K_e = 2$, distance of cell-center users < 0.6R.

$K_e = 2$ and cell-center users are dropped randomly up to distance zR, with $z = 0.3$. The numerical results for BER versus SNR of the cell-edge users is shown in Figure 3.2. It is obvious that our method significantly outperforms ZF-SIC and ML-SIC which assume perfect CSI of the cell-center users, whereas our method does not. For instance, more than one order of magnitude improvement using our CCA-RACMA method is observed at SNR= 6dB. Furthermore, the bad performance of ZF with channel estimation reflects how the inaccurate channel estimates of cell-edge users severely degrade their detection performance.

In order to see the effect of increasing the number of antennas on the performance of the proposed method, we considered the same setting of the previous experiment, however, we increased the number of antennas at each base station to 20, i.e., $M_1 = M_2 = 20$. Figure 3.3 shows that doubling the number of antennas at each base station improves the BER of cell-edge users obtained by all methods. However, a significant improvement gap in the BER obtained by our "blind" method is observed compared to that of ZF-SIC and ML-SIC. For instance, while ZF-SIC achieves an order of magnitude reduction in BER with $M = 20$, CCA-RACMA attains more than three orders of magnitudes improvement in BER at SNR = 5dB. Furthermore, Figure 3.3 shows that our approach does yield measurable BER when the SNR of cell-edge users exceeds 5dB. The reason is that CCA can aggressively suppress the inter-cell interference when the number of antennas exceeds the total number of users, as explained in Appendix A.1.

To test the effect of inter-cell interference, we vary the locations of cell-center users in their cell from $0.1R$ to $0.8R$, and for each setting we measure the BER attained by all methods at SNR = 5dB. Figure 3.4 demonstrates that the proposed CCA-RACMA approach still exhibits a favorable performance relative to that of ZF-SIC and ML-SIC. In particular, two orders of magnitude increase in the BER attained by ZF-SIC and ML-SIC is observed when the cell-center users are spread up to $0.8R$ compared to $0.1R$ from their serving BS, however, a very slight degradation in the performance of CCA-RACMA is observed, even for high spreads. Notice that, while the two baselines assume perfect knowledge of the cell-center user channels, this assumption becomes less realistic when "cell-center" users are in fact fully scattered throughout the cell. This therefore give a big advantage to the baselines over our method; notwithstanding, our method still works the best, even in this case.

We now consider another experiment with $M_1 = M_2 = 30$, $K_e = 2$ and SNR = 5dB. Assuming fixed user positions, we vary the number of cell-center users in each cell from 20 to 8, and for each given number of cell-center users we compute the BER of cell-edge users. In this experiment, all cell-center users are randomly dropped up to distance $0.5R$ from their serving BS. In Figure 3.5, we observe that our proposed blind method can attain BER that is below the detectable threshold for this simulation when the number of cell-center users per cell is less than 16 while the ZF-SIC detector is severely affected by the cancellation errors from cell-center users. This shows how the proposed approach can handle dense scenarios, and hence, it is expected to work well in the case of multiple BSs (more than two).

We next consider $M_1 = M_2 = M = 25$, $K_1 = K_2 = 15$, $Ke = 3$ and cell-center users are randomly located at distance less than $0.8R$ from their serving BS. As shown in Figure 3.6, jointly injecting more users (cell-center and cell-edge) and allowing them to be more spread, yields a noticeable degradation in the BER of cell-edge users achieved by all methods. This makes sense because, for ZF-SIC, there exists a higher chance that the detection performance of some cell-center users will be affected by the interference of cell-edge users resulting in cancellation errors from SIC. On the other hand, our method also exhibits some degradation in the performance because adding more users creates more intercell interference that can contaminate the common subspace estimated by CCA. However, our approach can still achieve much better performance to that obtained by ZF-SIC with perfect cell-center CSI. For example, our method still has more than an order of magnitude lower BER at different SNR values.

We further simulate the previous scenario with double the number of antennas at each base station. As Figure 3.7 depicts, doubling the number of antennas at each base station yields an order of magnitude improvement in the BER of our method, while only slightly improving the BER of ZF-SIC.

Finally, to show how CCA can still detect cell-edge user signals even when the received

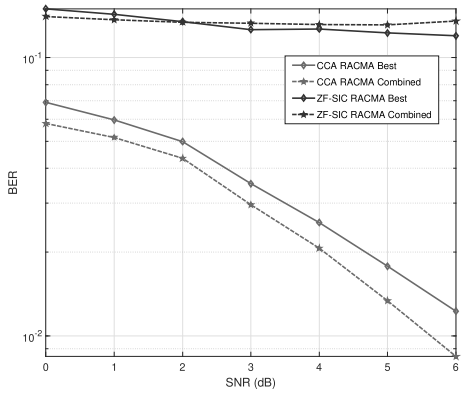


Figure 3.6: BER vs. SNR, with $M_1 = M_2 = 25$, $K_1 = K_2 = 15$ and $K_e = 3$, distance of cell-center users < 0.8R.

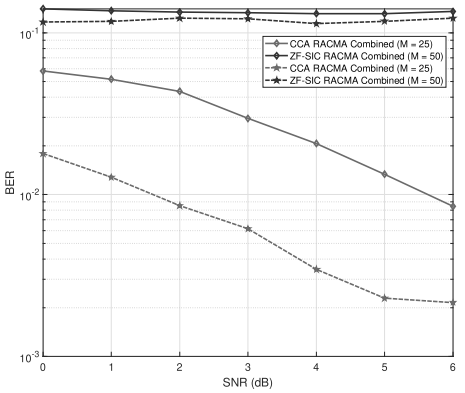


Figure 3.7: BER vs. SNR, with $M_1 = M_2 = M$, $K_1 = K_2 = 15$ and $K_e = 3$, distance of cell-center users < 0.8R.

signals at the two BSs are not perfectly synchronized, we consider a scenario with $K_1 = K_2 = 8$, $M_1 = M_2 = 10$, $K_e = 2$, SNR = 3dB, $T = 800$ and cell-center users are dropped randomly up to distance $0.5R$, and we assume that the received signal at the ℓ-th BS is $\widetilde{\mathbf{Y}}_\ell \in \mathbb{R}^{M_\ell \times \widetilde{T}}$, where $\widetilde{T}$ was set to 830. Then, we applied Algorithm 2 on $\widetilde{\mathbf{Y}}_1$ and $\widetilde{\mathbf{Y}}_2$, and observed the correlation coefficient of the first pair-wise canonical components as a function of the relative shift. Figure 3.8 shows how CCA can clearly identify the correct delay, and hence, detect cell-edge user signals as explained before. Clearly when the BS signals are not synchronized, there is no meaningful common subspace – even the first pair of canonical components exhibits low correlation. When we hit the correct delay, on the other hand, there are common components and the correlation coefficient ρ_1 is very high, as shown in Figure 3.8.

3.6 Summary

This chapter has considered cell-edge user signal detection in the uplink of a multi-cell multi-user MIMO system. The goal is to design a detector that can reliably demodulate cell-edge user signals in the presence of strong intra-cell interference from users close to the base station, without resorting to power control or scheduling algorithms that throttle the cell-center user rates. We have proposed a two-stage approach that leverages base stations cooperation to reliably identify cell-edge user signals at low SNR, without even knowing their channels. First, two-view CCA was brought in to estimate the subspace containing the cell-edge user signals

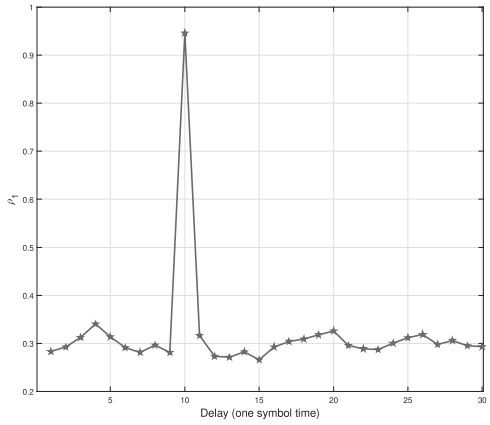


Figure 3.8: Correlation coefficient of the first pair-wise canonical component ρ_1 vs. delay.

shared by both base stations under the assumption that BS signals are synchronized. Then, an efficient analytical method called RACMA that guarantees the identifiability of binary signals from well-conditioned mixtures was utilized to extract the cell-edge user signals from the resulted mixture. We presented theoretical analysis of common subspace identifiability, in both ideal and realistic scenarios that include noise and inter-cell interference.

Furthermore, we developed an algorithm that can identify cell-edge user signals in the case when BS signals are not fully synchronized at the CSPU. Extensive simulations using a realistic path-loss model were carried out to show the superiority of the proposed learning-based method. It was shown that our blind CCA method achieves more than an order of magnitude improvement in the cell-edge user BER compared to the 'oracle' zero forcing and maximum likelihood cell-center multiuser detection followed by interference cancellation of the cell-center users before detecting the cell-edge users.

Chapter 4

Selective Cooperation for Cell-edge Detection: Does Adding Base Stations help?

4.1 Multi-cell Detection

In Chapter 3, we have showed the potential of CCA in handling the cell-edge problem in a setup with two BSs. Considering the general setting case with several BSs (more than two) indeed complicates the cell-edge problem, and makes cell-edge detection much more challenging due to the deleterious impact of inter- and intra-cell interference. Even with advanced technologies such as multiple-input-multiple-output (MIMO) and orthogonal frequency division multiplexing (OFDM) [69, 70] in place, nomadic users who are close to the cell edge are still prone to suffer from significant performance degradation especially in dense networks with large number of BSs and users [71, 72].

In this chapter, we will present an unsupervised learning-based method that leverages *selective* base station cooperation to recover cell-edge users signals at low SNR subject to strong inter- and intra-cell interference. Relying on fact that cell-edge users are located at approximately equal distances from different base stations, and hence their received signals are weak but *common* (meaning: they are received at low but roughly equal power at different base stations), it shows that reliable detection is possible via (generalized) canonical correlation analysis (G)CCA [73] under mild conditions.

While base station cooperation [74] has been considered before for several tasks such as

coordinated power control [75], coordinated scheduling [76], and inter-cell interference mitigation [77], cooperation here is utilized for a completely different purpose: as a means for cell-edge user detection, via GCCA.

Our contributions are as follows:

- We extend our previous work [14], which proposed using classical two-view CCA to detect cell-edge user signals in a cellular network with two cells, to the more general setting. That is, we consider a scenario with L cells / views ($L > 2$) which involves GCCA [17,78] as opposed to CCA for $L = 2$. We propose a two-stage approach that uses cooperation to jointly detect cell-edge users signals without prior knowledge of their channel state information. In particular, we first consider the so-called MAXVAR formulation of GCCA [17], and show that it yields the range space of the cell-edge user signals. We present identifiability conditions under which the common subspace can be recovered. While identifiability conditions for the common subspace of two views have been obtained in [14], the conditions we provide here for the general case are more relaxed. Upon identifying the subspace comprising the cell-edge users signals via GCCA, we utilize the (R)ACMA [62] algorithm, which exploits the finite alphabet constraint of the user transmitted signals to retrieve the original cell-edge user signals from the resulting mixture. Fortunately, both MAXVAR GCCA and RACMA admit relatively simple algebraic solution via eigenvalue decomposition. This renders our approach computationally favorable in practice, because the proposed method for solving the cell-edge problem is tantamount to solving two eigenvalue decomposition problems.

- We present an insightful theoretical analysis which shows that GCCA can reliably estimate the common subspace in the presence of thermal noise and cross interference from users in adjacent cells, under realistic assumptions on the SNR of the different users.

- We provide a GGCA strategy that can be used to classify users as cell-edge or cell-center, thereby determining the correct dimension of the common subspace.

- To showcase the effectiveness of our proposed method for cell-edge user detection, we provide a comprehensive suite of simulations that employs a realistic path-loss model from the 3GPP 38.901 standard. Experiments reveal that our approach attains a considerable improvement in the BER at low SNR under realistic levels of inter-cell interference and dense scenarios with a large number of cell-center users. We compare our proposed method with our previous CCA-based one and different multi-user detection techniques including ZF-SIC and MMSE-SIC which assume perfect knowledge of the cell-center user channels. We show that the proposed GCCA method achieves significant reduction in the BER compared to all baselines. Moreover, our simulations show that using GCCA with the

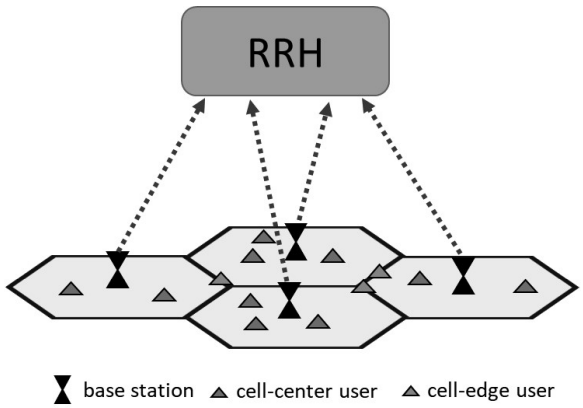


Figure 4.1: System model.

three closest BSs always yields the best detection performance for the cell-edge users. That is, not only does using the three closest BSs always improves the results of using the two closest BSs; but also that using more than the three closest BSs never helps – neither of which was obvious *a priori*.

The outline of this chapter is as follows. Section 4.2 defines the problem statement and highlights the major limitations of the prior cell-edge user detection methods. The proposed detector and the main results are presented in Section 4.3. Then, numerical simulations are provided in Section 4.4. Conclusions are drawn in Section 4.5.

4.2 System Model

Consider an uplink transmission scenario in a cellular network with L regular hexagonal cells – each cell has a base station (BS) located at its center, as shown in Fig. 4.1. The ℓ-th BS is equipped with M_ℓ antennas, and serves K_ℓ single-antenna users. Let K_{e_ℓ} denote the number of cell-edge users served by the ℓ-th BS, with $K_{e_\ell} < K_\ell, \forall\, \ell \in \mathcal{L} := \{1, \cdots, L\}$. The uplink channel vector representing the small-scale fading between the k-th user located in the j-th cell and the ℓ-th BS is given by $\mathbf{z}_{\ell kj} \in \mathbb{C}^{M_\ell}$. The entries of $\mathbf{z}_{\ell kj}$ are independent identically distributed (i.i.d.) complex Gaussian random variables with zero mean and variance $1/M_\ell$. This corresponds to a favorable propagation medium with rich scattering. The coefficient that accounts for the signal attenuation due to distance (path-loss) between the ℓ-th BS and the k-th user in the j-th cell is given by $\alpha_{\ell kj} \in \mathbb{R}$. Accordingly, the overall uplink channel vector is given by

$$\mathbf{h}_{\ell kj} = \sqrt{\alpha_{\ell kj}}\mathbf{z}_{\ell kj}, \tag{4.1}$$

Throughout this chapter, we assume that the channel vectors are *not* known a priori at the BSs.

4.2.1 Uplink Transmission

The considered users in the system are assumed to be allocated the same time-frequency resource. Also, assume that all user transmissions are heard at all BSs, thereby introducing both intra- and inter-cell interference at each BS. Also, all user transmissions are assumed to be synchronized at the BSs (this assumption can be lifted; see below and in [14]). Define $\mathbf{x}_{kj} \in \mathbb{C}^N$ as the vector transmitted by the k-th user in the j-th cell. The received signal, $\mathbf{Y}_\ell \in \mathbb{C}^{M_\ell \times N}$, at the ℓ-th BS is given by

$$\mathbf{Y}_\ell = \sum_{j=1}^{L} \sum_{k=1}^{K_j} \sqrt{p_{kj}} \mathbf{h}_{\ell kj} \mathbf{x}_{kj}^T + \mathbf{N}_\ell, \tag{4.2}$$

where $\mathbf{h}_{\ell kj}$ is the uplink channel response vector as defined in (4.1), p_{kj} is the transmitted signal power of the k-th user in the j-th cell. The term $\mathbf{N}_\ell \in \mathbb{C}^{M_\ell \times N}$ contains i.i.d. entries with zero mean and variance σ^2/N, i.e., $\mathbb{E}[\mathbf{N}_\ell \mathbf{N}_\ell^H] = \sigma^2 \mathbf{I}$. Throughout this work, we assume that neither scheduling algorithms nor power control is employed. In other words, all users are always active, and all users are assigned the same transmission power, i.e., $p_{kj} = p, \forall\ k, j$. Scheduling and power control algorithms can still be employed on top of the proposed framework for additional traffic shaping and other system considerations.

We assume that all BSs are connected to a remote radio head (RRH) via backhaul links which can be either microwave links or high speed optical fiber cables [74]. Each BS forwards its received signal to the RRH. Although base station cooperation has been considered in several earlier papers [75–77], it is adopted here for an entirely different purpose. That is, we exploit the joint processing of the BS signals at the RRH to provide reliable detection of cell-edge users whose signals are received at low SNR without knowledge of any of the user channels. One key challenge for all the BS cooperation techniques in the literature is time synchronization [74] of the received signal at the RRH. Even though all prior works assumed perfect synchronization at the RRH [72], we developed a low-complexity CCA-based algorithm that can handle (lack of) synchronization for two BSs [14], and can be easily modified to deal with the multi-cell case.

4.2.2 Prior Art: Limitations

We now provide a brief discussion of the limitations of the prior art used to detect cell-edge user signals. To this end, it is convenient to write (4.2) as

$$\mathbf{Y}_\ell = \mathbf{H}_{\ell p_\ell} \mathbf{X}_{p_\ell}^T + \mathbf{H}_{\ell e_\ell} \mathbf{X}_{e_\ell}^T + \sum_{j \neq \ell}^{L} \mathbf{H}_{\ell j} \mathbf{X}_j^T + \mathbf{N}_\ell, \tag{4.3}$$

where we collect the transmitted signals of the cell-center users and the cell-edge users served by the ℓ-th BS in the matrices $\mathbf{X}_{p_\ell} \in \mathbb{R}^{N \times (K_\ell - K_{e_\ell})}$ and $\mathbf{X}_{e_\ell} \in \mathbb{R}^{N \times K_{e_\ell}}$, respectively, and the

transmitted signals by the users served by the j-th BS in $\mathbf{X}_j \in \mathbb{R}^{N \times K_j}$. Further, the matrices $\mathbf{H}_{\ell p_\ell} \in \mathbb{C}^{M_\ell \times (K_\ell - K_{e_\ell})}$, $\mathbf{H}_{\ell e_\ell} \in \mathbb{C}^{M_\ell \times K_{e_\ell}}$ and $\mathbf{H}_{\ell j} \in \mathbb{C}^{M_\ell \times K_j}$ hold on their columns the respective channel vectors. Note that we absorbed the transmitted signal power p of each user in its channel vectors.

The goal is to recover the cell-edge user signals, $\mathbf{X}_{e_\ell}$, from the received signal $\mathbf{Y}_\ell$. The traditional approach to recover $\mathbf{X}_{e_\ell}$ is to first estimate all user channels via transmitting orthogonal pilots, and then employ the ZF or MMSE detector to decode cell-edge user signals using their estimated channel. This approach usually fails to provide reasonable performance due to the effect of intra-cell interference (transmissions of strong cell-center users), the effect of inter-cell interference (transmissions from users in other cells) and noise. The signals of cell-edge users are consequently received at very low signal to interference plus noise ratio (SINR), which causes high uncertainty in their channel estimates, which in turn seriously degrades their detection performance. One workaround is to use ZF or MMSE followed by successive interference cancellation to decode and subtract the cell-center user signals, thereby mitigating / eliminating the intra-cell interference effect. While this approach can slightly improve the detection performance, the inter-cell interference and channel estimation errors still cause severe performance degradation.

Another potential solution that mitigates the inter-cell interference effect [79], and can indeed enhance the detection of such users, is to use power control and/or scheduling algorithms together with BS cooperation techniques [80]. However, this comes at the expense of throttling the transmission of cell center users, and hence, it also severely degrades the overall system throughput.

In the forthcoming section, we will present a two-stage learning-based approach that leverages BS cooperation to reliably identify cell-edge user signals without knowing their channels, and without resorting to either power control or scheduling.

4.3 Proposed Detector and Identifiability Analysis

The cell-edge users are located far but at roughly equal distances from different BSs. In other words, if we use the distance-power relationship, their received signals are weak but *common* to multiple BSs, i.e., their signals are received at relatively equal power at different BSs. We will show how GCCA can efficiently recover the cell-edge users' signal range space at low SNR, even if they are buried under strong intra- and inter-cell interference. Notice that, owing to the broadcast nature of the wireless medium, all user transmissions are (over)heard, albeit weakly, at all BSs. Hence all user signals are, in principle, common. However, we use the phrases "common" for cell-edge users versus "private" for cell-center users to reflect the power

(im)balance of different users. That is, cell-center users signals are received at very high SNR, e.g., $[20, 30]$ dB, at their serving BS, and very low SNR, eg. $[-10, -30]$ dB at non-serving BSs. On the other hand, cell-edge user signals are received at low but roughly equal SNR, e.g., $[3, 5]$ dB at multiple BSs.

From the geometry of the hexagonal cells shown in Fig. 4.1, one can argue that a user can be common to two or three BSs, i.e., it can be located at relatively equal distance from two or three BSs. For example, a user located on the left corner of the common edge between the top and bottom cells in Fig. 4.1 is common to the BSs in these two cells and the one on the left. Based on this fact, we will design a detector that can recover cell-edge user transmitted signals from the signal received at three BSs. The case of more than three will also be considered in the simulations section.

Remark 6. *It is worth pointing out that we have* only *considered the uniformally hexagonal cells architecture and the roughly-equidistant assumption of cell-edge users for the ease of exposition. We only need the relative delays of the different users to be the same at each of the L=3 BSs (so that the associated views contain the same cell-edge subspace) and the received SNRs for the cell-edge users to be* at least *a few dB above the noise floor. The latter typically holds even when shadowing is considered. Being approximately equidistant from the 3 BSs is one reasonable way to motivate these assumptions, but it is not the only one (e.g., small cells and per-subcarrier processing). Furthermore, while with shadowing coefficients the received SNR at the different views will not be balanced, our approach does not necessarily require the received SNR to be fully balanced at the different views, as we will see in the simulations.*

We will first consider the noiseless case to find the identifiability conditions required to recover the cell-edge user signals. Identifiability is very important as it provides sufficient conditions under which the recovery of the cell-edge user signals via (G)CCA is guaranteed under ideal (noiseless) conditions. Whereas we have derived identifiability conditions in the case of two BSs [14], it turns out that the conditions for three BSs are more relaxed (details will be provided in the next subsection).

4.3.1 Noiseless Case

Let $K_c = \sum_{j=1}^{L} K_{e_j}$ denote the total number of cell-edge users. Assume that all cell-edge users are located around the intersection point of the three hexagonal cells. Thus equation (4.3), with $L = 3$, can be rewritten as

$$\mathbf{Y}_\ell = \mathbf{H}_{\ell p_\ell}\mathbf{X}_{p_\ell}^T + \mathbf{H}_{\ell c}\mathbf{X}_c^T + \sum_{j\neq\ell}^{L} \mathbf{H}_{\ell p_j}\mathbf{X}_{p_j}^T + \mathbf{N}_\ell, \tag{4.4}$$

where the subscripts 'p_ℓ' and 'c' stand for private to the ℓ-th BS and common to all base stations, respectively. The matrices $\mathbf{X}_c \in \mathbb{R}^{N\times Kc}$ and $\mathbf{X}_{p_j} \in \mathbb{R}^{N\times(K_j-K_{e_j})}$ hold the transmitted signals of the cell-edge users and cell-center users in the j-th cell, respectively. Accordingly, $\mathbf{H}_{\ell c} \in \mathbb{C}^{M_\ell\times K_c}$ and $\mathbf{H}_{\ell p_j} \in \mathbb{C}^{M_\ell\times(K_j-K_{e_j})}$ hold on their columns the corresponding channel vectors. Further, we define $\mathbf{E}_\ell := \sum_{j\neq\ell}^{L}\mathbf{H}_{\ell p_j}\mathbf{X}_{jp_j}^T + \mathbf{N}_\ell$ to denote the summation of inter-cell interference and noise at the ℓ-th BS. Thus, (4.4) can be rewritten as

$$\mathbf{Y}_\ell = \mathbf{H}_{\ell p_\ell}\mathbf{X}_{p_\ell}^T + \mathbf{H}_{\ell c}\mathbf{X}_c^T + \mathbf{E}_\ell, \tag{4.5}$$

Our goal now is to recover the cell-edge user signals $\mathbf{X}_c$ given $\{\mathbf{Y}_\ell\}_{l=1}^L$. Recall that the MAXVAR GCCA formulation in the complex domain can be expressed as,

$$\min_{\{\mathbf{Q}_\ell\}_{\ell=1}^L,\mathbf{G}} \sum_{\ell=1}^{L}\|\mathbf{Y}_\ell^H\mathbf{Q}_\ell - \mathbf{G}\|_F^2 \tag{4.6a}$$

$$\text{s.t.} \quad \mathbf{G}^H\mathbf{G} = \mathbf{I}. \tag{4.6b}$$

We will start by showing how the solution of the MAXVAR GCCA formulation (4.6) is related to the column space of the cell-edge user signals, and then we will explain how the original signals can be recovered from the given solution. Upon defining the matrix $\mathbf{V}^{(L)} \in \mathbb{C}^{(L-1)N\times((L-1)K_c+\sum_{\ell=1}^{L}(K-K_{e_\ell}))}$ as follows,

$$\mathbf{V}^{(L)} = \begin{bmatrix} \mathbf{X}_{p1} & -\mathbf{X}_c & \mathbf{X}_{p2} & & \\ \vdots & & \ddots & \ddots & \\ \mathbf{X}_{p1} & & & -\mathbf{X}_c & \mathbf{X}_{pL} \end{bmatrix},$$

we have the following result.

Theorem 2. *In the case where $\mathbf{E}_\ell = 0$, if the matrix $\mathbf{W}_\ell := [\mathbf{H}_{\ell c}, \mathbf{H}_{\ell p_\ell}] \in \mathbb{C}^{M_\ell\times(K_c+K_\ell-K_{e_\ell})}$ and the matrix $\mathbf{V}^{(L)} \in \mathbb{C}^{(L-1)N\times((L-1)K_c+\sum_{\ell=1}^{L}(K-K_{e_\ell}))}$ are full column rank, then the optimal solution $\mathbf{G}^\star$ of problem (4.6) is given by $\mathbf{G}^\star = \mathbf{X}_c\mathbf{F}$, where $\mathbf{F}$ is a $K_c \times K_c$ non-singular matrix.*

Proof. See [24] which offers a comprehensive identifiability analysis of GCCA for general L. □

Remark 7. *The full column rank condition on $\mathbf{W}_\ell$ requires that: i) the number of antennas at each base station is greater than or equal to the number of users assigned to that base station, plus any cell-edge users associated with the other two base stations; and ii) the channel vectors of different users to be linearly independent. The first requirement is supported by massive MIMO technology that aims at equipping the base station with hundreds of antennas [81]. Further, because the user channel vectors can be assumed to be drawn from an absolutely continuous*

distribution (see (4.1)*), the latter condition is satisfied with probability one. The full column rank condition on* $\mathbf{V}$ *is related to the number of samples* N *required to guarantee recovery. For* $L = 3$*, we need: i) the number of samples to be at least equal to the total number of cell-edge users plus one half the total number of cell-center users served by the three BSs; and ii) the transmitted sequences of different users to be linearly independent. For finite alphabets, the two conditions are satisfied with very high probability for modest* N*, as the user transmissions are statistically independent. In addition, it can be easily seen that the requirement on the number of samples* N *becomes less restrictive for* $L > 2$ *compared to the earlier results presented in Chapter 3 for the two-view case.*

Theorem 2 asserts that in an ideal scenario where the effect of inter-cell interference and thermal noise is negligible compared to the intra-cell interference, GCCA successfully recovers the subspace spanned by the cell-edge users signals, under mild conditions. We point out that such a scenario can arise in practice, especially if all cell-center users are close to their serving BS. An interpretation of the statement of Theorem 2 is that if there exist several spatio-temporal signal views that contain very strong but different components (in our case here arising from the transmissions of each group of cell-center users) and very weak but common components (in our case here the received cell-edge user signals), then GCCA recovers the common components range space irrespective of the power of the individual components.

However, in practical deployment scenarios we cannot guarantee the above idealized assumptions. One of the main contributions of this paper is that it offers an analysis of GCCA performance in a realistic scenario with inter-cell interference and noise. This is coming up next.

4.3.2 Noisy Case

We now provide analysis showing how cell-edge users signals can be identified when $\mathbf{E}_\ell \neq 0$. In particular, we show that the signal subspace recovered by identifying the K_c principal eigenvectors of $\mathbf{A}$ is indeed containing the cell-edge user transmitted messages. As shown earlier, this is equivalent to solving the MAXVAR GCCA problem.

Upon defining $K_s = \sum_{\ell=1}^{L} K_\ell$, let us write (4.4) in a more compact form as

$$\mathbf{Y}_\ell = \mathbf{H}_\ell \mathbf{X}^T + \mathbf{N}_\ell, \tag{4.7}$$

where $\mathbf{H}_\ell = [\mathbf{H}_{\ell c}, \mathbf{H}_{\ell p_1}, \cdots, \mathbf{H}_{\ell p_L}] \in \mathbb{C}^{M_\ell \times K_s}$, and $\mathbf{X} = [\mathbf{X}_c, \mathbf{X}_{p_1}, \cdots, \mathbf{X}_{p_L}] \in \mathbb{C}^{N \times Ks}$. Further, we can use (4.1) to factor $\mathbf{H}_\ell = \mathbf{Z}_\ell \mathbf{P}_\ell^{1/2}$, where the columns of $\mathbf{Z}_\ell$ are the channel vectors representing small scale fading between the k-th user and the ℓ-th BS, for $k \in \mathcal{K}_s := \{1, \cdots, K_s\}$. Accordingly, each entry of the diagonal matrix $\mathbf{P}_\ell$ represents the corresponding received signal power that incorporates the transmitted power and the path-loss between each user and the ℓ-th

BS. Thus, (4.7) can be equivalently written as

$$\mathbf{Y}_\ell = \mathbf{Z}_\ell \mathbf{P}_\ell^{1/2} \mathbf{X}^T + \mathbf{N}_\ell, \tag{4.8}$$

Assumption 1 (AS1)**.** *Assume that the matrices* $\mathbf{Z}_\ell$ *and* $\mathbf{C} := \mathbf{X}/\sqrt{N}$ *are approximately orthonormal, i.e.,* $\mathbf{Z}_\ell^H \mathbf{Z}_\ell \approx \mathbf{I}_{M_\ell}$ *for all* $\ell \in [L]$ *and* $\mathbf{C}^H \mathbf{C} \approx \mathbf{I}_{K_s}$.

Remark 8. *Recall that the matrices* $\mathbf{Z}_\ell$ *contain i.i.d entries with zero mean and variance* $1/M_\ell$*, and hence, the approximate orthonormality assumption on* $\mathbf{Z}_\ell$ *requires the number of base station antennas to be greater than the total number of users assigned to all base stations, and large enough for the sample average (inner product of different columns) to be close to the ensemble average (***0***). This requirement on the number of antennas is supported by massive MIMO technology that aims at equipping base stations with hundreds of antennas. On the other hand, the approximate orthonormality of* $\mathbf{C}$ *requires the sequence length to be greater than the total number of users and the columns of* $\mathbf{C}$ *to be linearly independent. For finite alphabets, the latter condition is satisfied with very high probability for modest* N.

Let $\gamma_{k\ell}$ denote the received SNR of the k-th user at the ℓ-th BS. Then, we define $r_{k\ell}$ as

$$r_{k\ell} := \frac{\gamma_{k\ell}}{\gamma_{k\ell} + 1},$$

for all $k \in \mathcal{K}_s$ and $\ell \in [L]$. For any user k, the multi-view correlation measure η_k is defined as $\eta_k := \sum_{\ell=1}^{L} r_{k\ell}$. We will make use of the following assumption on the cell-edge users.

Assumption 2 (AS2)**.** *For any cell-edge user* i *and cell-center user* j, $\eta_i > \eta_j$.

Remark 9. *Empirically, the relation between the average received power* P_r *and the distance is determined by the expression* $P_r \propto d^{-\lambda}$ *where* d *and* λ *denote the distance and the path loss exponent, respectively. The noise power at the receiver is given by* σ^2*. Then, the value of* $r_{k\ell}$ *as function of the distance between the user and the BS (*$d_{k\ell}$*) is given by*

$$r_{k\ell} = \frac{(d_{k\ell})^{-\lambda}}{(d_{k\ell})^{-\lambda} + \sigma^2/c}, \tag{4.9}$$

where c *is constant that depends on the communication medium and the antenna characteristics. This function exhibits a sharp phase transition which means that the ratio* $r_{j\ell}$ *for cell-center users at other cells is almost zero if they are dropped up to certain distance from their serving BS such that their received SNR at their non-serving BSs is a few dBs below zero, while all the ratios* $r_{i\ell}$ *for cell-edge users at their adjacent BSs is close to one if their received SNR at those BSs is a few dBs above zero.*

Our main result is the following:

Proposition 2. *In the presence of inter-cell interference and additive noise, under assumptions (AS1) and (AS2), the optimal solution* $\mathbf{G}^\star$ *of problem* (4.6) *is given by* $\mathbf{G}^\star \approxeq \mathbf{X}_c\mathbf{P}$ *where* $\mathbf{P}$ *is any* $K_c \times K_c$ *non-singular matrix.*

Proof. The proof is relegated to the Appendix A.2. □

We have thus showed that MAXVAR GCCA identifies the *range space* of the cell-edge users signals, under realistic conditions. We will next show how the original signals $\mathbf{X}_c$ can be unraveled from their range space, by exploiting their constellation/modulation structure.

4.3.3 ACMA Stage

Given the subspace $\mathbf{G}^*$, the problem of recovering the user signals $\mathbf{X}_c$ can then be posed as

$$\min_{\mathbf{X}_c,\mathbf{F}} \|\mathbf{G}^\star - \mathbf{X}_c\mathbf{F}\|_F^2 \tag{4.10a}$$

$$\text{s.t.} \quad \mathbf{X}_c(i,j) \in \Omega, \tag{4.10b}$$

where $\mathbf{X}_c(i,j)$ represents the (i,j)-th entry of the matrix $\mathbf{X}_c$, for $i = 1, \cdots, N$ and $j = 1, \cdots, K_s$. Although problem (4.10) is known to be NP-Hard even if $\mathbf{F}$ is known, the Analytical Constant Modulus Algorithm (ACMA) developed by van der Veen [82] provides a good algebraic solution which comes with certain identifiability guarantees. In particular, ACMA transforms (4.10) to a generalized eigenvalue problem. While the obtained solution is subject to both phase and permutation ambiguities, both of them can be resolved in practice by simply matching the preamble of each estimated signal with the identification sequence that is known *a priori* at the serving BS.

It is worth emphasizing that the proposed end-to-end detector of the cell-edge user signals only requires solving two generalized eigenvalue problems. Therefore, the overall computational complexity of our proposed method is dominated by the complexity of solving two generalized eigenvalue problems. This renders our approach favorable for practical implementation.

We also point out that our proposed method works for any modulation scheme and even for analog signals. It is obvious that GCCA (first stage) can identify the common subspace irrespective of the modulation of the cell-edge user signals. The second stage exploits knowledge of the modulation to unravel the constituent signals from their range space. ACMA [82], for instance, deals with constant modulus communication signals such as higher-order PSK, QPSK or even analog phase or frequency-modulated signals. For binary signals (BPSK), the so-called real ACMA (RACMA) [62] can be used to recover the user signals. Furthermore, algorithms such as SIC-ILS [83] or the approach in [24] can be utilized for higher-order QAM – SIC-ILS can also exploit Forward Error Control (FEC) codes to further improve the decoding accuracy.

We assume that the cell-edge users employ BPSK, QPSK, and 8PSK modulation, and hence, as we will see in the simulations, we use either RACMA [62] to recover the BPSK signals or ACMA [82] for QPSK and 8PSK modulation. Recall that the received signals of the cell-edge users are naturally weak, hence these users will typically employ low-order modulation.

4.3.4 Choosing the common subspace dimension

Recall that in our analysis in the Appendix A.2, in order to differentiate between cell-center users and cell-edge users, we assumed that cell-center users are dropped up to certain distance from their serving BS (see Remark 8). However, if all users are randomly dropped throughout the cells, then it is not obvious how to differentiate if a user has to be treated as a cell-center or a cell-edge user. In other words, how to determine the common subspace dimension if the cell-center users are fully scattered within their cell. Note that underestimating the common subspace dimension can naturally lead to a performance degradation as we will see in the simulation section. To overcome such an issue, we propose a GCCA-based algorithm that can accurately estimate the number of cell-edge users (common subspace dimension), and hence, we can classify whether a user is cell-center or cell-edge.

Exploiting the fact that a component that is common to three views should also be common to each pair of the three views, the common subspace dimension can be accurately estimated via checking the mean of correlation coefficients computed from the canonical components of each pair. Recall that $K_c \leq \min\{2M_\ell, N\}$, where K_c is the number of canonical pairs that can be extracted using GCCA. Upon solving problem (4.6) and obtaining the solutions $\{\mathbf{Q}_\ell^\star\}_{\ell=1}^3$, we define the i-th correlation coefficient between views j and ℓ as [24],

$$\rho_{\ell j}^{(n)} = \mathbb{R}\mathrm{e}\{\mathbf{Q}_\ell^H(:,n)\mathbf{Y}_\ell\mathbf{Y}_j^H\mathbf{Q}_j(:,n)\} \tag{4.11}$$

$\forall \ell, j \in [L]$ and $j > \ell$ and $i \in \{1, \cdots, K_c\}$. Afterwards, we compute the i-th average correlation coefficient as $\rho_{\text{avg}}^{(n)} = \frac{1}{3}(\rho_{12}^{(n)} + \rho_{13}^{(n)} + \rho_{23}^{(n)})$, Then, we decide that the n-th canonical components $(\mathbf{Q}_1(:,i), \mathbf{Q}_2(:,i), \mathbf{Q}_3(:,i))$ extract a common signal if $\rho_{\text{avg}}^{(i)}$ is greater than a certain threshold – a reasonable choice of ρ_{th} is 0.5.

It is worth emphasizing that the proposed GCCA approach requires solving a generalized eigenvalue problem which naturally yields $\min\{M_\ell, T\}$ canonical vectors. Once these vectors are obtained, we use our proposed method to estimate the exact number of relevant canonical vectors, i.e., number of cell-edge users K_c, which will be later used to recover the common subspace from the data matrices $\{\mathbf{Y}_\ell\}_{\ell=1}^3$.

4.4 Numerical Results

In this section, we use realistic numerical simulations to assess the performance of the proposed GCCA approach. We consider a scenario with $L = 4$ hexagonal cells, each of radius R = 600 meters. The locations of cell-center users served by each BS are drawn uniformly at random within a distance less than $d = 0.4$R from their serving BS, unless stated otherwise. Cell-edge users, on the other hand, are located around the edges between base stations at distance between 0.95R and 1.05R. Fig. 4.2 shows one simulated scenario where cell-center users and cell-edge users are colored in red and green triangles, respectively. The transmitted power of all users was set to 25 dBm while the transmitted sequence length N was fixed to 800. We assume BPSK modulation for all the users, unless stated otherwise. All results were averaged over 500 Monte Carlo trials. Additive white Gaussian noise was used with variance σ^2 so that the SNR is P_e/σ^2, where P_e is the average received power of cell-edge users. In fact, this enables us to evaluate SNR values required for cell-edge users to achieve a specific BER. The uplink channel between the k-th user in the j-th cell and the ℓ-th BS is modeled as

$$\mathbf{h}_{\ell kj}^H = \sqrt{\frac{1}{M}} \sum_{n=1}^{N_p} \sqrt{\alpha_{\ell kj}^{(n)}} \mathbf{a}_r^H(\phi_k^{(n)}), \tag{4.12}$$

where N_p is the number of paths between the ℓ-th BS and the k-th user in cell j, $\forall\{\ell, j\} \in \mathcal{L}$ and $k \in [K_s]$. To compute the path gain, $\alpha_{\ell kj}^{(n)}$, we use the path-loss model of the urban macro (UMa) scenario from Table $7.4.1-1$ in the 3GPP 38.901 standard, with the carrier frequency set to 2 GHz, $\forall n, \ell, j, k$. Furthermore, all cell-center users are allowed to possibly have a line of sight (LOS) component to their serving BS according to the LOS probability expression for the UMa scenario in Table $7.4.2-1$ in the 3GPP 38.901 standard; however, all cell-edge users have only non-LOS components. The term $\mathbf{a}_r(.)$ is the array response vector at the BS, and $\phi_k^{(n)} \sim \mathcal{U}[-\pi, \pi]$ denotes the azimuth angle of arrival of the n-th path associated with the k-th user. Assuming the BS is equipped with a uniform linear array with omni-directional antenna elements stretched over the vertical direction, then

$$\mathbf{a}_r(\phi) = [1, e^{ikd\cos(\phi)}, \cdots, e^{ikd(M-1)\cos(\phi)}]^T, \tag{4.13}$$

where $i = \sqrt{-1}$, $k = 2\pi/\lambda$, λ is the carrier wavelength and $d = \lambda/2$ is the spacing between antenna elements.

To assess the efficacy of our approach, we implement the following approaches and use them as performance baselines.

- **MMSE / ZF with channel estimation**: the channels of all users are estimated via transmitting sequences of orthogonal pilots with length of 250 each. Then, both the MMSE

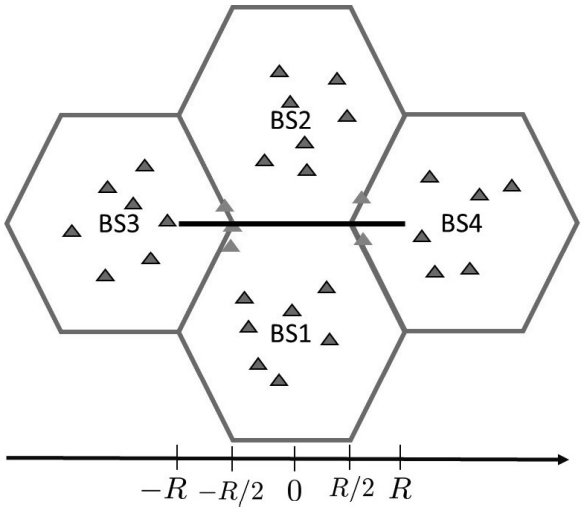


Figure 4.2: Snapshot from the simulated scenario.

and ZF detectors are employed to decode the cell-edge user signals using their estimated channels.

- **MMSE / ZF SIC (R)ACMA with channel estimation**: the channels of the cell-center users associated with each BS are estimated first. Then, we use both the ZF-SIC and MMSE-SIC detectors to decode, and then subtract the re-encoded signals of the cell-center users at their serving BS. Afterwards, we apply (R)ACMA [62, 82] on the residual signal to recover the cell-edge user signals. To guarantee fairness, since we assume BS cooperation, we feed (R)ACMA with the residual signals from all BSs simultaneously.

- **MMSE / ZF SIC (R)ACMA Perfect**: similar to the previous baseline but with perfect knowledge of the cell-center user channels at their serving BS.

- **CCA (R)ACMA combined**: we use CCA to recover the range space of the cell-edge users from the nearest two BSs [14]. Then, we apply (R)ACMA to recover the cell-edge user signals from the resulting subspace.

In the first experiment, we consider a setup with $M_\ell = 12$ antennas and $K_\ell = 8$ single transmit antenna users, $\forall\ \ell \in \mathcal{L}$. Considering the scenario shown in Fig. 4.2, we varied the x-location of one cell-edge user on the black edge between BS1 and BS2 from $x = -\mathrm{R}$ to $x = \mathrm{R}$, while the locations of all other users are kept fixed during the experiment. At each value of x, we report the BER of the proposed GCCA approach using the three closest BSs, GCCA using all BSs, CCA using the closest two BSs and ZF-SIC RACMA with perfect cell-center users channels using the best BS. Note that when the user is located at $x = -\mathrm{R}/2$, its received SNR

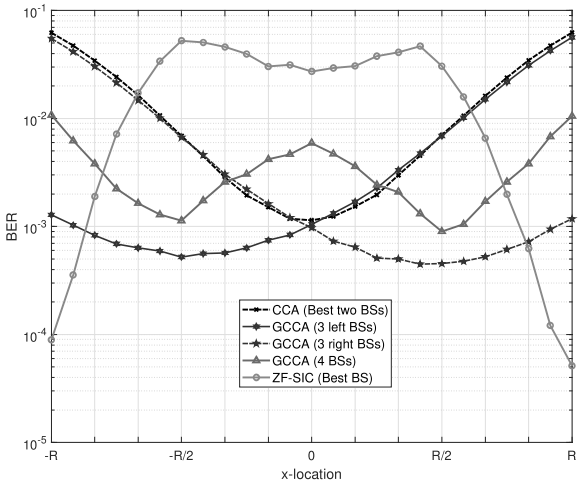


Figure 4.3: BER vs. cell-edge user location: GCCA using 3 closest BSs is always better.

is approximately equal to 3dB at BSs 1,2 and 3, according to our adopted path-loss model. As the user's location shifts towards the center ($x = 0$) of the black edge, its received SNR increases (as path-loss decreases) at BSs 1, 2, and 4 while it decreases at BS 3. When this user passes the center of the edge, the received SNR decreases again at BSs 1 and 2. On the other hand, when the user's location changes towards the very left corner of the black edge ($x = -\mathrm{R}$), its received SNR automatically increases at BS 3, while it decreases at BSs $1, 2, 4$.

As Fig. 4.3 depicts, when the user is located at $x = -\mathrm{R}/2$ (the user is at relatively equal distance from three BSs), the proposed GCCA using the three left BSs (1, 2, and 3) attains the minimum BER compared to GCCA using all BSs, CCA using the two closest BSs and ZF-SIC using the best BS. Similarly, when the user is located at $x = \mathrm{R}/2$, the joint detection using the three BSs 1, 2, and 4 gives the best performance. When the user is close to $x = \pm\mathrm{R}/2$, GCCA using the three nearest BSs attains more than order of magnitude reduction in the BER relative to CCA RACMA and much more relative to ZF-SIC RACMA.

On the other hand, as the location of the user moves towards the center of the cell-edge ($x = 0$), the detection performance of CCA using the two BSs 1, 2 improves gradually until it reaches its best at the origin (minimum path-loss and maximum received SNR), and then it decreases again as shown in Fig. 4.3. This happens as the received SNR of the user's signal at BS 1 and BS 2 becomes higher, and considerable discrepancy becomes evident between the received SNR at BSs 1 and 2 and at BS 3. Note that GCCA using the best three BSs always yields the minimum BER when the user's location is in the interval $[-\mathrm{R}/2, \mathrm{R}/2]$.

When the cell-edge user location becomes close to either BS 3 or BS 4, i.e., $x \approx \pm\mathrm{R}$, Fig. 4.3

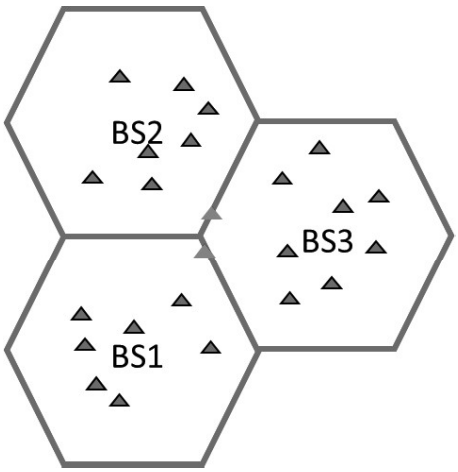


Figure 4.4: Snapshot from the three BSs simulated scenario.

shows that using ZF-SIC RACMA at the nearest BS achieves the best detection performance among all other methods that use joint detection. This can be attributed to the fact that this user is no longer a "common" user - there is a large discrepancy among the received SNR at BS 3 (very high) and at BSs $1, 2$, and 4 (very low), and hence, the power imbalance severely affects the detection performance of the (G)CCA based approaches. This observation suggests that depending on the user's type (center or edge), one should use either the closest BS or the three nearest BSs to detect the user. In other words, if a user is relatively close to any BS, then this user's signal received power is high at this BS and very weak at all other BSs, and hence, it makes sense to decode this user's signal from the nearest BS. However, if a user is close to the edge between cells, then this user is *common* to multiple BSs and jointly detecting such a user from the three closest BSs using GCCA yields the best detection performance.

More interestingly, it turns out that adding more BSs does not always improve the performance. For instance, at $x = \pm \mathrm{R}/2$, while GCCA using four BSs attains a comparable performance relative to GCCA with the three nearest BSs, the latter is considerably better as the user moves towards the center $x = 0$. This is because at $x = 0$, the received SNR is very low at both BS 3 and BS 4 compared to BS 1 and BS 2, and consequently, both views 3 and 4 act as two "noisy" views that naturally degrade the signal recovery of the cell-edge user. Therefore, one can conclude that from the geometry of the hexagonal cells, adding more BSs and feeding their received signals to GCCA to recover the common subspace will further degrade the detection performance of cell-edge users as any additional BS (view) will lead to an additional noisy view that severely affects the cell-edge user's signal recovery. In other words, the more views ($L > 3$) GCCA uses, the more difficult it becomes to reveal common information from all

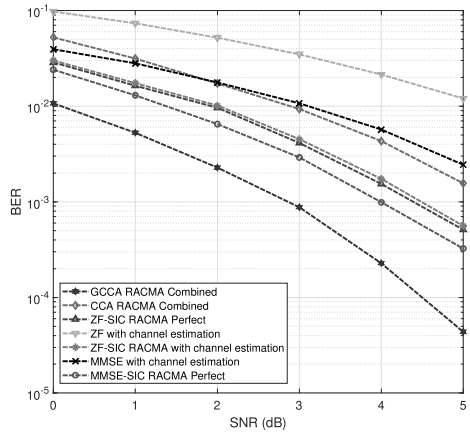


Figure 4.5: BER vs. SNR of cell-edge users.

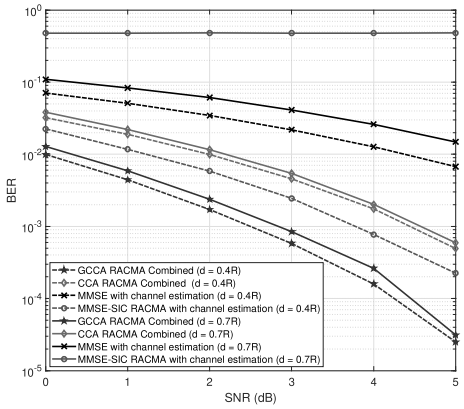


Figure 4.6: BER vs. SNR of cell-edge users, d denotes the distance at which cell-center users are randomly dropped up to.

views simultaneously. Therefore, using the observation that using GCCA with the three closest BSs always yields the minimum BER for cell-edge users, and to further show the effectiveness of our approach under different settings, we will only consider the three BSs scenario, shown in Fig. 4.4, for all subsequent experiments.

We now consider another experiment where we vary the transmitted power of the two cell-edge users from 20 dBm to 25 dBm which corresponds to approximately 0 dB to 5 dB SNR according to the adopted path-loss model, while the transmitted power of all-center users is fixed to 25 dBm. Drawing different cell-center user locations for each Monte-Carlo realization, we compute the average BER among the two cell-edge users as a function of their transmitted power. Fig. 4.5 shows how GCCA provides significant improvement in the BER compared to all other methods. In particular, GCCA achieves an order of magnitude reduction in the BER compared to MMSE-SIC followed by RACMA (the second best method) which jointly detects the cell-edge user signals using the residual signals from the three BSs. Notice that, MMSE-SIC assumes perfect 'oracle' knowledge of the channels of cell-center users at their serving BS. This assumption becomes less realistic when "cell-center" users are fully scattered throughout the cell. Although this gives a big advantage to the MMSE-SIC approach, GCCA still provides considerably better detection performance of cell-edge users. Furthermore, one can easily see how the channel estimation errors severely degrade the detection performance of cell-edge users.

Additionally, we simulate a more realistic scenario where cell-center users are almost fully scattered in their cell. In particular, cell-center users are dropped up to $d = 0.7$R from their

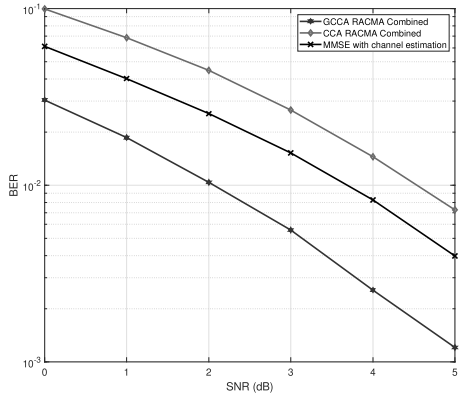


Figure 4.7: BER vs. SNR of cell-edge users for different number of users.

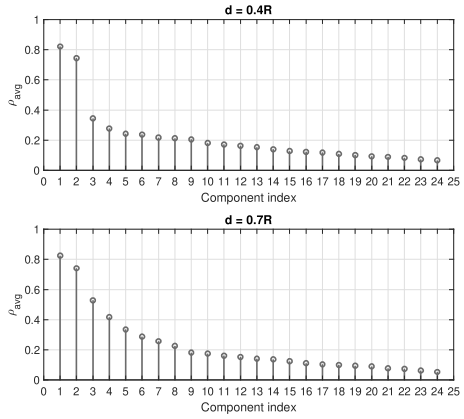


Figure 4.8: Average correlation coefficient of all possible extracted components via GCCA.

serving BS, thereby cell-edge users are experiencing more aggressive inter-cell interference compared to the case where $d = 0.4$R. Note that the MMSE-SIC RACMA is implemented using estimates of the cell-center users' channels instead of assuming perfect knowledge of their channels because that is hard to attain even approximately when cell-center users are fully scattered. Fig. 4.6 depicts the inter-cell interference effect on the detection performance achieved by different methods. It is obvious that the MMSE-SIC RACMA completely fails at $d = 0.7$R compared to $d = 0.4$R. On the other hand, both GCCA and CCA have a slight degradation in their performance, which in turn reflects the efficacy of both methods that principally rely on recovering the subspace of the "equipowered" users. Note that GCCA with three BSs still attains an outstanding detection performance compared to the other methods under this realistic scenario.

We carry out another experiment in a more dense scenario where $K_\ell = 16$ and $M_\ell = 30$. Further, cell-center users are dropped up to $d = 0.8R$. We report the BER of cell-edge users versus the SNR. Although the results in Fig. 4.7 show that the detection performance of all methods significantly degrades compared to the one in the previous experiment where $K_\ell = 8$, our proposed approach still can attain acceptable performance by achieving 10^{-3} BER at 5dB. Notice that doubling the number of users and allowing them to be more scattered naturally leads to greater corruption in the estimated common subspace, and hence, the degradation in the detection performance obtained by the proposed method is expected.

Adding more BSs with K users each might be expected to severely affect the performance which is true in general. However since, in principle, our approach recovers the subspace containing the "equipowered" user signals, adding more users in the far cells (not served by the

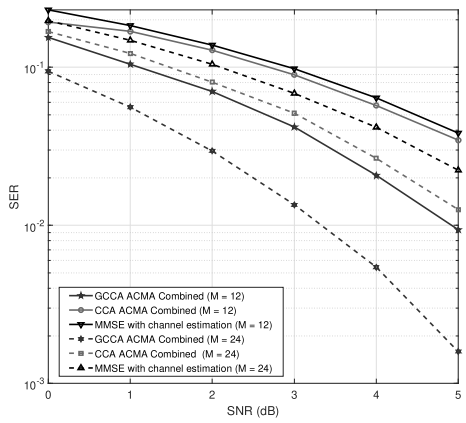


Figure 4.9: SER vs. SNR of cell-edge users with QPSK modulation.

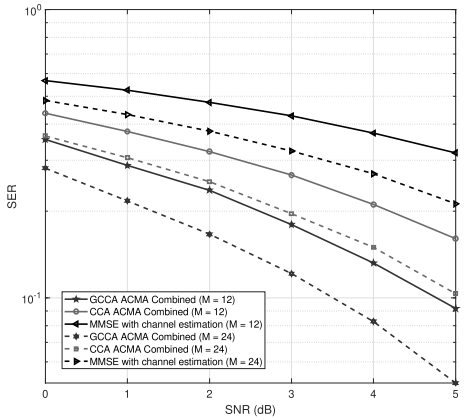


Figure 4.10: SER vs. SNR of cell-edge users with 8PSK modulation.

three closest BSs) can slightly affect the cell-edge detection performance as we observed from our simulations. For instance, looking at Fig. 4.2, increasing the number of users served by BS 4 does not affect the BER obtained when the cell-edge user is located at $x = -\mathrm{R}/2$ if GCCA is used with BSs $1, 2$ and 3. However, increasing the number of cell-center users can only degrade the performance, as shown in Fig. 4.7, when these users are served by any of the BSs used for detecting the cell-edge users via GCCA. This can be attributed to the fact that there is a chance that some users could be included in the common part, and hence, underestimating the dimension of the common subspace can degrade performance.

We now test the proposed algorithm used to detect the number of cell edge users (i.e., the common subspace dimension). We consider a setup with $M_\ell = 12$, $K_\ell = 8$, and SNR ≈ 3 dB. Note that since $2M_\ell < N$, we can find up to $2M_\ell$ canonical components. Fig. 4.8 shows the average correlation coefficient computed at the i-th extracted component, for $i = 1, \cdots, 24$, for two different drop/scatter patterns for the cell-center users. It is obvious that when $d = 0.4\mathrm{R}$, there is a significant gap between the average correlation coefficient of the first two components and the rest of the components. In particular, the average correlation coefficient of the first two components is almost 0.8 while all the rest are less than 0.3. Thus, one can decide that there exist only two cell-edge users in this case. On the other hand, at $d = 0.7\mathrm{R}$, the value of the average correlation coefficient slightly increases for some of the components. For example, the average correlation coefficient of the third component now jumps to 0.52 which means that there is one more user that can be considered as common user. However, considering only the first two components is enough to reliably recover the signals of the two cell-edge users as shown

in Fig. 4.6, where the detection performance was sightly affected by increasing d from 0.4R to 0.7R.

Finally, we evaluate the performance of the proposed method under different modulation schemes. First, we assume a QPSK transmission for all users in a scenario with $K_\ell = 8$ users and two cell-edge users ($K_c = 2$). Further, cell-center users are randomly located up to 0.7R from their serving BS. We report the symbol error rate (SER) vs SNR of cell-edge users for $M_\ell = M = 12, 24$ antennas, $\forall \ell$. As Fig. 4.9 depicts, GCCA followed by ACMA attains the lowest SER compared to both MMSE with channel estimation and CCA with the two closest BSs. It is also clear that, with $M = 12$, the detection performance of all methods is worse for QPSK relative to BPSK in Fig. 4.6, as expected. However, as we double the number of antennas to $M = 24$, a significant improvement can be achieved by our method, which attains approximately 10^{-3} SER at 5 dB SNR.

We also carried out another simulation with the previous setup but with 8PSK modulation instead of QPSK, and with $K_c = 1$. The numerical results in Fig. 4.10 demonstrate the efficacy of our approach with higher-order modulation, in the low SNR region. Note that the detection performance of all methods degrades significantly as we increase the modulation order, which is expected given the SNR range considered. However, for higher order modulation, coding schemes can be employed on top of our proposed approach for improved reliability.

4.5 Summary

In this chapter, we studied the problem of cell-edge user signal detection in the uplink of a multi-cell, multi-user MIMO system, with the aim of designing a detector that can reliably demodulate cell-edge user signals in the presence of strong intra-cell interference from cell-center users, without resorting to power control or/and scheduling algorithms that throttle the cell-center user rates. We proposed a GCCA-based approach that leverages *selective* base station cooperation to reliably identify the common subspace containing cell-edge user signals at low SNR, without even knowing their channels. Then, we used an efficient analytical method ((R)ACMA) that guarantees the identifiability of finite alphabet signals from well-conditioned mixtures to separate the cell-edge user signals from the resulting subspace. The proposed method is appealing for use in dynamic environments because it (i) does not require any knowledge of the channel state information of cell-edge users; (ii) can automatically detect the number of common (cell-edge) user signals regardless of associated channel variations from one coherence interval to the next; (iii) can automatically adjust to varying PSK modulation order (so cell-edge users can even vary their modulation, depending on channel conditions); and (iv) can efficiently deal with synchronization issues.

We presented theoretical results to prove that under an idealized scenario, the proposed GCCA-based approach recovers the subspace containing the cell-edge user signals. Furthermore, we showed through an elegant analysis that under realistic assumptions on the inter-cell interference and the SNR of the cell-edge users, the common subspace recovery is guaranteed via GCCA. Simulations using a *realistic* propagation and system model were carried out to show the superiority of the proposed learning-based method over the prevailing state-of-the-art methods. In particular, our proposed approach attained an order of magnitude reduction in the BER compared to other multi-user detection methods that assume perfect knowledge of the channels of the cell-center users. Furthermore, our experimental results evaluated the cell-edge user detection performance as a function of the number of cooperating BSs, and revealed that using the three closest BSs is always optimal in this regard. This was not obvious *a priori*, as intuition may have suggested that two or even more than three BSs might be preferable in certain cases.

Chapter 5

Seamless Underlay Communication

5.1 Motivation

The rapidly growing demand for wireless connectivity from 5G+ to Internet of Things (IoT) and WiFi-enabled devices has brought renewed interest and impetus behind dynamic spectrum sharing [84–86]. Even with millimeter-wave (mmWave) technology, the propagation loss in the 28 GHz - 300 GHz bands is much higher than in the sub-6 GHz bands [87], making the latter better-suited for various wireless systems. The premium placed on sub-6 GHz bands together with the need to protect scientific uses in the mmWave bands are driving the renewed interest in spectrum sharing and dynamic spectrum access (DSA).

DSA techniques are designed to improve spectrum utilization by allowing secondary unlicensed users to take advantage of ephemeral transmission opportunities in space, time, or frequency [88–90] – a capability often referred to as *cognitive radio*. Currently, there are three widely used DSA techniques for cognitive radio networks (CRN): interweaving, overlay, and underlay [84]. In the interweaving mode, the secondary users search the band for spectrum holes (vacant sub-bands) which represent secondary transmission opportunities. The overlay paradigm requires tight coordination between the primary and secondary users, which complicates implementation. Relative to the interweaving and overlay modalities, underlay spectrum sharing is appealing in terms of its prioritization of the licensed users, practical feasibility, and its relative simplicity – there is no need for continuous spectrum sensing or tight coordination with the primary system.

There is a plethora of works on DSA and cognitive radio, spanning two decades of research

ranging from spectrum sensing [91, 92] and channel gain "cartography" [93] to different spectrum sharing modalities [94–106]. A common assumption in those works is that the signal to interference plus noise ratio (SINR) at the secondary receiver can be made high enough to enable reliable decoding. In practice, this is hard to ensure if the primary transmitter is powerful (e.g., a TV or radio station) while the secondary is power-limited (e.g., a WiFi or IoT device). Furthermore, many of these works are relying on assumptions that are hard to meet in practice – such as the availability of cross-channel knowledge at the secondary users.

Few spectrum underlay works have attempted to circumvent the need for such assumptions. One interesting recent example is [107], where the authors proposed a nice semi-blind beamforming-based underlay spectrum sharing approach, which allows the secondary users to access the spectrum while minimally affecting the primary network performance, without requiring any channel knowledge at the secondary network. However, the proposed method in [107] still requires i) the primary communication to be bidirectional (which does not hold for legacy radio or TV broadcast, or scientific uses); ii) the flow direction of primary traffic to be predictable; iii) effectively time-invariant channels from/to the primary users; and iv) training pilots for designing the beamformer at the secondary receiver. These are still restrictive assumptions. In particular, the reverse transmission of the primary user needs to be synchronized with the forward of the secondary, and vice versa, so the secondary users need to track which node is transmitting in the primary network.

Is it possible to design an underlay strategy that enables reliable decoding at very low SINR and modest SNR at the secondary receiver, without noticeable increase of the noise floor at the primary receiver? Is it possible to do this seamlessly, without any coordination between the primary (legacy / incumbent) and the secondary user?

The answer is, surprisingly, affirmative. We propose a secondary transmission protocol that operates at very low power yet allows reliable secondary communication without requiring any channel knowledge or coordination with the primary system. The key idea is that the secondary user sends its signal twice, each time at very low power. Assuming that the secondary receiver employs at least two receive antennas, the proposed transmission protocol allows the secondary receiver to create two "views" of the signal space that only share the secondary signal – the interference from the primary network is potentially very strong, but different in the two views. Invoking canonical correlation analysis (CCA) on these two views, the secondary receiver can reliably decode its intended signal under very strong interference from the primary user.

Transmitting the same signal twice can be viewed as repetition coding [108], or as elementary direct-sequence spreading [109, 110] with spreading gain equal to two. Our approach is fundamentally different from these classical techniques *in the way that this controlled redundancy is exploited at the receiver* (i.e., on the "decoding" side), where we leverage the unique strengths

of CCA. In particular, we exploit our algebraic interpretation of CCA in Chapter 3 to reliably decode the secondary signal even under strong interference from the primary users.

Our contributions can be summarized as follows:

- We propose a novel secondary underlay framework that enables seamless primary-secondary coexistence – there is no need for coordination between the two. Assuming that the secondary receiver is equipped with two receive antennas and down-conversion chains, simple repetition of the secondary signal coupled with CCA processing at the secondary receiver can recover the secondary transmission even at very low SINR.

- The approach is data-driven and unsupervised in that it directly recovers the secondary information signal (up to complex scaling), without requiring channel state information or primary signal recovery and cancellation. It even works with analog modulation of the primary and/or the secondary signal.

- Time-varying channels for the primary and the secondary user can be naturally accommodated, provided that the channel coherence time is greater than half the secondary transmission frame length (comprising a transmitted packet and its repetition – and the packet length is up to our control and can be fairly short).

- From a computational point of view, what is required is the computation and inversion of small correlation matrices, and then a principal eigenvector computation, which can be done using e.g., the power method. Hence, the approach is attractive for practical implementation.

- The approach is immune to carrier frequency offset, which can be compensated after the secondary symbol sequence is extracted using CCA. Furthermore, exploiting the repetition structure and CCA, we develop a matching synchronization algorithm that identifies the correct timing of the secondary transmission frames even at very low SINR in an unsupervised manner – i.e., without using any pilot symbols, only exploiting the structured redundancy introduced by repetition. These side-benefits are very fortunate, for otherwise synchronization is a very difficult problem at very low SINR without very long pilot sequences for acquisition.

- Last but not least, in order to demonstrate the practical feasibility and merits of our approach we have built and tested a prototype using software defined radios, where both the secondary and primary users were realized using USRP-2920 radios. We conducted multiple experiments to evaluate the performance of the proposed underlay CCA approach under realistic conditions. Our laboratory experiments verified that the proposed approach can reliably recover a secondary user signal that is buried under strong interference from the

primary system (SINR as low as -40 dB), and that it approaches the attainable detection performance in the interference-free regime (where the primary user is idle).

The rest of this chapter is organized as follows. Section 5.2 presents the system model and highlights the major limitations of the prior art in terms of secondary underlay schemes. The proposed secondary transmission protocol is described in Section 5.3, while the proposed detector is presented in Section 5.4. Section 5.5 explains how to resolve synchronization issues at the secondary receiver. Experimental results are provided in Section 5.6, and conclusions are drawn in Section 5.7.

5.2 System and Signal Models

5.2.1 System Model

Consider an underlay cognitive radio network comprising a secondary transmitter (STx) communicating with a secondary receiver (SRx) equipped with $M_s \geq 2$ antennas, in the presence of a primary transmitter (PTx) and primary receiver (PRx) with $M_p \geq 1$ antennas, as shown in Fig. 5.1. Multiple secondary and primary users can also be accommodated as we will explain later. Let $\mathbf{h}_s \in \mathbb{C}^{M_s}$, $\mathbf{h}_{ps} \in \mathbb{C}^{M_s}$, $\mathbf{h}_{sp} \in \mathbb{C}^{M_p}$ and $\mathbf{h}_p \in \mathbb{C}^{M_p}$ be the channel response between the STx and SRx, PTx and SRx, STx and PRx, and PTx and PRx, respectively, defined as

$$\begin{aligned} \mathbf{h}_s &= \sqrt{\sigma_s}\ \mathbf{g}_s, \quad \mathbf{h}_{ps} = \sqrt{\sigma_{ps}}\ \mathbf{g}_{ps}, \\ \mathbf{h}_p &= \sqrt{\sigma_p}\ \mathbf{g}_p, \quad \mathbf{h}_{sp} = \sqrt{\sigma_{sp}}\ \mathbf{g}_{sp}, \end{aligned} \tag{5.1}$$

where $\mathbf{g}_s$, $\mathbf{g}_{ps}$, $\mathbf{g}_{sp}$, and $\mathbf{g}_p$ are the respective small-scale fading vectors, while the terms σ_s, σ_{ps}, σ_{sp}, and σ_p are the corresponding large scale fading coefficients with values dependant on the propagation distance and environment.

Unlike prior works [97–101] that require estimates of the cross channels $\mathbf{h}_{ps}$ and/or $\mathbf{h}_{sp}$ at the secondary receiver and the secondary transmitter, respectively, this paper considers a practical setting where the secondary users have no knowledge about any channel state information in the network.

5.2.2 Signal Model

We assume that the primary users transmission is done over a narrowband channel of bandwidth B Hz. For simplicity of exposition, we assume that both users are employing QPSK modulation, but other types of modulation can be accommodated. The basic approach we propose to recover the secondary signal is modulation-agnostic, and does not assume anything about the primary signal's modulation, which can even be analog.

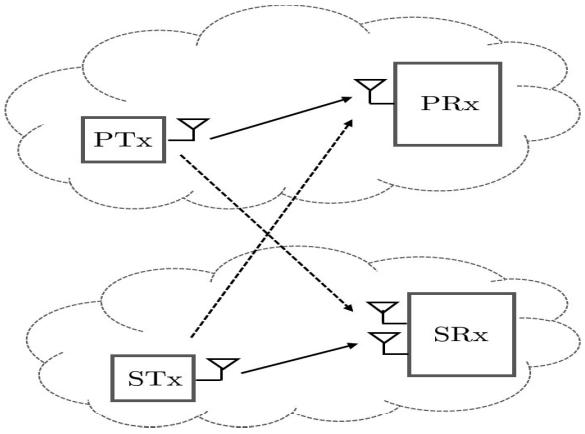


Figure 5.1: System Model

Let $\mathbf{x}_p \in \mathbb{C}^N$ and $\mathbf{x}_s \in \mathbb{C}^N$ denote the digitally-modulated transmitted signal by the primary and secondary user, respectively, where $|\mathbf{x}_p(n)|^2 = 1$ and $|\mathbf{x}_s(n)|^2 = 1$ for $n \in [N] := \{1, \cdots, N\}$. In writing down the discrete-time baseband-equivalent model, we shall assume, for simplicity of exposition, that the primary and secondary signals are synchronized at the symbol level – otherwise writing down the model is cumbersome. However, such an assumption is not required for our approach to work, and we shall later present an algorithm that can lock on the secondary user signal at the SRx. All our laboratory experiments are concerned with this asynchronous setup.

The discrete-time synchronous baseband-equivalent model of the received signal, $\mathbf{Y}_s \in \mathbb{C}^{M_s \times N}$, at the secondary receiver is given by

$$\mathbf{Y}_s = \sqrt{\alpha_s}\mathbf{h}_s\mathbf{x}_s^T + \sqrt{\alpha_p}\mathbf{h}_{ps}\mathbf{x}_p^T + \mathbf{W}_s, \tag{5.2}$$

where α_s and α_p are the transmit power of the STx and PTx, respectively. The term $\mathbf{W}_s \in \mathbb{C}^{M_s \times N}$ contains independent identically (i.i.d) distributed elements with each entry drawn from a complex Gaussian distribution with zero mean and variance σ_s^2. Similarly, the received signal at the primary receiver, $\mathbf{Y}_p \in \mathbb{C}^{M_p \times N}$, is given by

$$\mathbf{Y}_p = \sqrt{\alpha_s}\mathbf{h}_{sp}\mathbf{x}_s^T + \sqrt{\alpha_p}\mathbf{h}_p\mathbf{x}_p^T + \mathbf{W}_p, \tag{5.3}$$

where $\mathbf{W}_p \in \mathbb{C}^{M_p \times N}$ is the noise term at the primary receiver with i.i.d entries drawn from a complex Gaussian distribution with zero mean and variance σ_p^2.

The goal of this work is to show that, in the absence of channel state information at the STx/SRx and without any coordination between the primary and secondary users, seamless secondary underlay communication is possible without affecting the primary network performance. To do this, we will first present a simple secondary transmission protocol together with

a data-driven (unsupervised learning-based) approach that allow i) the STx to transmit its signal at very low power so that it does not affect the detection performance at the PRx, thereby keeping the resulting interference close to the PRx noise floor (the PRx can reliably decode its signal even with one receive antenna), and ii) the SRx to reliably decode its intended signal at significantly low SINR (e.g., -40 dB).

5.3 Secondary Transmission Protocol

In this section, we will present a simple transmission protocol that will assist the secondary transmitter to reliably communicate with its receiver over the same channel occupied by the primary network, and without degrading the primary user's performance.

The secondary transmission scheme is described as follows. If a secondary user desires to transmit in a channel occupied by a primary user, it simply transmits the same sequence twice at very low power – enough to be received above the thermal noise floor at the SRx, but far below what is required to be directly decoded in the face of possibly overwhelming interference by the PTx. The repetition of the secondary user's sequence can happen at the symbol or block level; we assume block-repetition for simplicity of exposition. To do this, we write $\mathbf{x}_s$ as two back-to-back repeated blocks, i.e., $\mathbf{x}_s = [\mathbf{s}^T \mathbf{s}^T]^T$, where $\mathbf{s} \in \mathbb{C}^{N/2}$ is the transmitted QPSK symbols by the secondary user over each block. Partitioning $\mathbf{x}_p = [\mathbf{p}_1^T \mathbf{p}_2^T]^T$ in two blocks for convenience, the received signal at the secondary receiver in (5.2) can be rewritten as

$$\mathbf{Y}_s = \mathbf{H}_s \begin{bmatrix} \mathbf{s} & \mathbf{p}_1 \\ \mathbf{s} & \mathbf{p}_2 \end{bmatrix}^T + \mathbf{W}_s, \tag{5.4}$$

where $\mathbf{H}_s$ is an $M_s \times 2$ matrix that holds on the first column the channel vector containing the channel coefficients between the STx and SRx, $\mathbf{h}_s$, and on the second column the channel from the PTx to the SRx, $\mathbf{h}_{ps}$. Notice that the transmit power terms of both the STx and PTx have been absorbed in the respective channel vectors, for brevity.

As noted earlier, the proposed transmission scheme can be interpreted as repetition coding [108], or equivalently as direct-sequence spreading of the secondary user's transmission with spreading gain equal to two [109]. Treating this situation as CDMA or as an error control problem will not work, because the primary user dominates the received signal, and small spreading / coding gains cannot make up for the large power difference between the secondary and primary user. CDMA performance is known to suffer from the so-called *near-far* problem which is clearly the case for the setup considered herein.

We will next present a low-complexity learning-based approach that allows the SRx to reliably decode its intended signal, $\mathbf{s}$, even if the received SINR is significantly low.

5.4 Secondary Signal Detection via CCA

By exploiting the repetition structure, the SRx can split $\mathbf{Y}_s$ and $\mathbf{W}_s$ into two blocks, $\mathbf{Y}_s = [\mathbf{Y}_1\ \mathbf{Y}_2]$, and $\mathbf{W}_s = [\mathbf{W}_1\ \mathbf{W}_2]$, for which we have

$$\mathbf{Y}_1 = \mathbf{H}_s\ [\mathbf{s}\ \mathbf{p}_1]^T + \mathbf{W}_1, \tag{5.5}$$

$$\mathbf{Y}_2 = \mathbf{H}_s\ [\mathbf{s}\ \mathbf{p}_2]^T + \mathbf{W}_2. \tag{5.6}$$

Now, given the two signal views in (5.5), CCA will be invoked to show that reliable detection of the secondary signal, $\mathbf{s}$, is possible even at low SINR. To see how we can utilize CCA to identify the secondary signal, $\mathbf{s}$, from $\mathbf{Y}_1 \in \mathbb{C}^{M_s \times N/2}$ and $\mathbf{Y}_2 \in \mathbb{C}^{M_s \times N/2}$, we will use the MAX-VAR formulation of CCA assuming a single common component, i.e.,

$$\min_{\mathbf{g},\mathbf{q}_1,\mathbf{q}_2} \sum_{\ell=1}^{2} \|\mathbf{Y}_\ell^T \mathbf{q}_\ell - \mathbf{g}\|_2^2, \tag{5.7a}$$

$$\text{s.t.} \quad \|\mathbf{g}\|_2^2 = 1. \tag{5.7b}$$

Recall that the MAX-VAR formulation aims at finding a direction $\mathbf{g} \in \mathbb{C}^{N/2}$ that is maximally correlated after the linear projections of $\mathbf{Y}_1$ and $\mathbf{Y}_2$ on $\mathbf{q}_1 \in \mathbb{C}^{M_s}$ and $\mathbf{q}_2 \in \mathbb{C}^{M_s}$, respectively. Further, the optimal canonical vectors can be obtained via first solving the following generalized eigenvalue problem to obtain $\mathbf{q}_1^\star$ and $\lambda^\star$,

$$\mathbf{R}_{12}\mathbf{R}_2^{-1}\mathbf{R}_{21}\mathbf{q}_1 = \lambda \mathbf{R}_1 \mathbf{q}_1. \tag{5.8}$$

where $\mathbf{R}_i = \frac{1}{N}\mathbf{Y}_i\mathbf{Y}_i^H$ is the sample auto-covariance of the random vector $\mathbf{y}_i$, and $\mathbf{R}_{ij} := \frac{1}{N}\mathbf{Y}_i\mathbf{Y}_j^H$ is the sample cross-covariance of the two random vectors $\mathbf{y}_i$ and $\mathbf{y}_j$, respectively, for $i, j = 1, 2$ and $i \neq j$. Notice that as explained in Chapter 2, it can be easily verified that the term $\lambda^\star$ represents the square of the correlation coefficient, $\rho(\mathbf{q}_1^\star, \mathbf{q}_2^\star)$, associated with the optimal canonical pair $\mathbf{q}_1^\star$ and $\mathbf{q}_2^\star$, where

$$\rho(\mathbf{q}_1^\star, \mathbf{q}_2^\star) = \mathbb{R}\text{e}\{\mathbf{q}_1^{\star H}\mathbf{Y}_1\mathbf{Y}_2^H\mathbf{q}_2^\star\}. \tag{5.9}$$

In Chapter 3, we have shown that given two multi-antenna signal views that include one shared (common) component and multiple individual ("private", not shared) components in each view, CCA can efficiently extract the common component up to scaling ambiguity no matter how strong the individual components are. One can see from (**??**) and (**??**) that each block (view) is subject to strong interference by the primary user, but, in general, the interference is different in the two blocks – thus there is a unique common subspace, namely (the span of) $\mathbf{s}$ that conveys the secondary transmission. Building upon our theoretical findings, we will next

show that our CCA interpretation applies, and under very mild conditions will recover $\mathbf{s}$ up to scaling , even if $\mathbf{x}_p$ is several orders of magnitude stronger than $\mathbf{x}_s$.

The following theorem, which is a slight modification of the results we presented in Chapter 3, states the conditions for identifying the secondary transmitted signal $\mathbf{s}$ at the SRx.

Theorem 3. *In the noiseless case, if the matrices* $\mathbf{B}_\ell := [\mathbf{s}, \mathbf{p}_\ell] \in \mathbb{C}^{N/2\times 2}$*, for* $\ell \in \{1, 2\}$*, and* $\mathbf{H}_s \in \mathbb{C}^{M_s\times 2}$ *are full column rank , then the optimal solution* $\mathbf{g}^\star$ *of problem* (5.7) *is given by* $\mathbf{g}^\star = \gamma\mathbf{s}$*, where* $\gamma \in \mathbb{C}$*,* $\gamma \neq 0$ *is the scaling ambiguity.*

Proof. The proof is provided in Theorem 1 in Chapter 3. □

Note that the full rank condition on the matrices $\mathbf{B}_\ell$ needs the signals $\mathbf{s}$ and $\mathbf{p}_\ell$ to be linearly independent which is practically always the case for any reasonable "packet" length N, because these signals are drawn from statistically independent sources. On the other hand, the full rank condition on $\mathbf{H}_s$ is in fact the more restrictive one as it requires i) the number of antennas at the SRx to be greater than or equal to the number of co-channel signals (two in our setting) and ii) the channel vectors to be linearly independent. The latter is realistic, these being statistically independent channel vectors from the PTx and the STx to the SRx.

5.4.1 Time-varying channel directions, fading, and intermittent transmissions.

Although the two signal views in (5.5) implicitly assume that the channel is constant across the two secondary repetition blocks, our proposed method in fact can work even if the two channel matrices are different, see Theorem 1 in Chapter 3. Therefore, with block repetition, the coherence time needs to be only greater than one block duration. We will see in the experiments how this feature grants our proposed method robustness against time varying channels.

5.4.2 Interference cancellation

It is worth pointing out that if the primary user signal is order(s) of magnitude stronger and the primary channel remains constant (no intermittent transmissions, no time-division duplex, insignificant channel direction changes) then one can cancel the primary interference by simply projecting the received signal on the minor left singular vector of the matrix $\mathbf{Y}_\ell$, thereby "revealing" the secondary transmission. This can only work when the spatial channels are time-invariant. In practice, the channel gains fluctuate over time, and even if the average secondary signal to interference ratio is low (e.g., -40 dB), there are times when it becomes relatively high (e.g., -20 dB). These fluctuations quickly degrade the subspace estimate, leading to complete failure to detect the secondary signal, as we will see in the laboratory experiments.

5.4.3 Multiple secondary users

Note that our theoretical results dictate that our proposed CCA approach can identify the secondary signal in a network with only one secondary user, and we have argued that finding the secondary user signal is tantamount to solving for a principal eigenvector which can be cheaply computed via the power method. Even with multiple secondary users, our recovery claim holds and receiver complexity is roughly the same, provided that i) each secondary receiver has enough antennas (as many as the maximum number of active users at any given time, see Theorem 1); and ii) there are no persistent and perfectly aligned collisions between any of the secondary users. In other words, no two secondary users transmit their packet pairs at the exact same times. With asynchronous wake-up type devices serving intermittent communication needs, this situation is highly likely.

5.5 Secondary Synchronization

One critical issue that we always face in practice is synchronization. The overall synchronization task comprises time, carrier frequency offset (CFO), and phase synchronization. While effective solutions to these problems are well-established for classical communication modalities, here we are dealing with a secondary signal that is potentially buried under the primary one, which makes secondary time synchronization and CFO acquisition much more challenging.

A standard receiver will naturally lock on the primary user, which means that the secondary signal will present itself with an unknown CFO and unknown start time within the received sequence. Fortunately, the presence of CFO does not destroy the alignment of the two copies of the secondary packet: owing to the temporal shift invariance property of pure complex exponential signals, the second copy is the same as the first except for a complex phase shift. Hence we can proceed with CCA and correct the CFO after recovering the CFO-modulated secondary packet. On the other hand, secondary timing acquisition is a challenge, due to the large power imbalance between the primary and the secondary signal. To deal with this problem, we propose a blind CCA-based algorithm that is practically effective in finding the start time of the secondary packet under such a large power imbalance between the two users.

In practice, the secondary receiver receives a long sequence, $\widetilde{\mathbf{Y}}_s \in \mathbb{R}^{M_s \times \widetilde{N}}$ where $\widetilde{N} > N$. The goal is to find the sample index, k, so that we can extract the desired signal $\mathbf{Y}_s$ from $\widetilde{\mathbf{Y}}_s$, and then use the proposed method in Section 5.4 to decode the secondary user signal.

By exploiting the repetition structure of the transmitted signal, we start with $k = 1$ and construct the two views $\mathbf{Y}_1^{(k)} = \widetilde{\mathbf{Y}}_s(:, k : N/2 + k - 1)$ and $\mathbf{Y}_2^{(k)} = \widetilde{\mathbf{Y}}_s(:, N/2 + k : k + N - 1)$ followed by solving (5.8) to obtain the associated correlation coefficient ρ_k (we use MATLAB notation, i.e., $\mathbf{X}^{(k)} = \mathbf{X}(:, k : N + k - 1)$ contains all the rows of matrix $\mathbf{X}$ and a subset of

columns of $\mathbf{X}$ starting from the k-th column and ending with the $(N+k-1)$-th column). Then, we store ρ_k, set $k = k+1$ and repeat the previous procedure. If we hit the start point of the two copies of the same packet, then CCA of these "views" will yield its maximum correlation coefficient. In other words, the correlation coefficient, ρ_k defined in (5.9), associated with each pair of canonical directions $\mathbf{q}_1^{(k)}$ and $\mathbf{q}_2^{(k)}$ obtained by solving (5.7) at the k-th step, will be at its maximum only when we have all the $N/2$ symbols in both views. This is because the secondary information sequence is uncorrelated, thus even if k is off by one, the two partial sequences will decorrelate. The higher N is, the higher the correlation peak we obtain as we will see in the experiments, but even moderate N, in the order of 128 symbols, can yield very good detection performance. Notice that the procedure utilizes the special frame structure that is designed to enable CCA, but is otherwise agnostic to the specific information sequence that is being sent by the secondary transmitter. In this sense, it is a blind synchronization strategy that leverages the power of CCA to enable reliable timing acquisition at very low SINR. The procedure is summarized as Algorithm 3.

Algorithm 3 Secondary Synchronization

Input: $\widetilde{\mathbf{Y}}_s \in \mathbb{C}^{M_1 \times \widetilde{N}}$,

Initialization: $k = 1$,

while $k \in [\widetilde{N} - N + 1]$ **do**

- Construct $\mathbf{Y}_1^{(k)} = \widetilde{\mathbf{Y}}_s(:, k : N/2 + k - 1)$ and $\mathbf{Y}_2^{(k)} = \widetilde{\mathbf{Y}}_s(:, N/2 + k : k + N - 1)$
- Compute ρ_k after solving (5.7) using $\mathbf{Y}_1^{(k)}$ and $\mathbf{Y}_2^{(k)}$
- Store (k, ρ_k) in a stack
- Set $k := k + 1$

end

Selection: pick the $k^\star := \max_k \rho_k$.

The computational complexity of Algorithm 3 is determined by the complexity of solving a series of CCA problems, which is equivalent to solving for the principal component (canonical pair) of (5.8) a number of times (equal to the search window size). The canonical pair can be cheaply computed via a power iteration. Further, each CCA problem requires inversion of correlation matrices of size $M_s \times M_s$ each – these inverses can be computed analytically since $M_s = 2$. To minimize the search window length, one can start with a coarse estimate for the region with high correlation coefficient and then do a narrow search within a small window size to get the final start time index, as we will see in the experiments. Furthermore, if the secondary transmitter is continuously transmitting, we do not need to run the full Algorithm 3 for each received packet – we only need to do a narrow timing search to compensate for jitter.

Figure 5.2: Experimental Setup.

Figure 5.3: (a) Primary Transmitter. (b) Primary Receiver. (c) Secondary Transmitter. (d) Secondary Receiver.

5.6 Experiments

In this section, we evaluate the performance of the proposed CCA approach for low-power secondary underlay communication in practice (for simplified simulations, see [16]). To do so, we have built a prototype of the proposed CCA underlay scheme using software defined radios (SDR).

5.6.1 Experimental Setup

Both the primary and secondary links are realized using USRP-2920 devices and general-purpose computers. The USRPs are used for radio signal transmission / reception, while the computers are used for baseband signal processing. The experimental layout is shown in Fig. 5.2. We used five USRPs: one for the primary transmitter, one for the primary receiver, one for the secondary transmitter, and two for the secondary receiver, see Fig. 5.2. Each USRP is equipped with a single antenna. The two USRPs of the secondary receiver are connected together with a MIMO cable to synchronize the two receive radio frequency chains, as shown in Fig. 5.3(d).

The locations of the PTx, PRx, STx, SRx are fixed throughout the experiments. The

Parameter	Primary	Secondary
Bandwidth (KHz)	100	100
Carrier frequency (GHz)	1.2	1.2
Modulation	QPSK	QPSK
Sample rate (MSps)	1	1
Maximum transmit power (dBm)	20	-15
Number of antennas	1 Tx, 1 Rx	1 Tx, 2 Rx
Number of symbols	256	128
Oversampling factor	10	10
Number of packets	2000	2000

Table 5.1: Parameter settings for the experiments.

distances between the PTx and PRx, PTx and SRx, STx and PRx, and STx and SRx are 5, 3, 4.5, and 4 meters, respectively. The transmit power of the PTx is set to the maximum possible value, as shown in Table 5.1 unless stated otherwise, while the transmit power of the STx is adjusted for low-power secondary transmission. The sampling rate for both users is set to 1 Mega samples per second (MS/sec), the signal bandwidth is 100 KHz, and the carrier frequency is 1.2 GHz. The PTx uses a block of 256 QPSK symbols and the STx uses repetition over two blocks, each of length 128 QPSK symbols. The parameter settings for our experiments are summarized in Table 5.1.

Signal processing at the transmitters. At each Tx, the constructed block is oversampled by a factor of 10, then the resulting oversampled signal is pulse-shaped using a square-root raised cosine (SRRC) with roll-off factor and amplitude set to 0.5 and 6, respectively. The pulse shaped signal is zero-padded with a number of zeros equal to one third of the packet, yielding a sequence of length 4020 samples. This results in a transmission rate of 128 Kbps for the primary user and 64 Kbps for the secondary user. The zero-padding (used to emulate intermittent packet transmission) is also used at the receiver side to measure the received SNR and SINR, as we will see later. Symbol generation, up-sampling, and pulse shaping are done in MATLAB. Then, the transmit data of each user is fed to GNU radio before being transmitted over the air.

Secondary receiver. We use the proposed CCA algorithm in Section 5.5 to detect both the secondary packet and the start of the 256×2 complex signal. After SRRC matched filtering, down-sampling to the symbol rate, and secondary synchronization, we construct the two signal views by separating the two back-to-back blocks, and then use CCA to recover the secondary signal. After solving the CCA problem (5.7), we average the two soft estimates of $\mathbf{s}$ obtained via $\mathbf{Y}_1^H\mathbf{q}_1$ and $\mathbf{Y}_2^H\mathbf{q}_2$, before hard thresholding.

To benchmark the performance of the proposed CCA approach, we use the following baselines.

SVD without interference: we will use the singular value decomposition (SVD) to estimate the channel direction during a period when the primary user is inactive, i.e., there is no interference from the primary user. To do that, we first exploit the repetition structure to construct the signal $\mathbf{Y} = [\mathbf{Y}_1^T \mathbf{Y}_2^T]^T \in \mathbb{C}^{2M_s \times N/2}$. Then, the secondary user signal can be estimated by projecting the received signal $\mathbf{Y}$ on the left principal vector. Note that our use of the SVD "baseline" without interference (which is more appropriately called an "oracle" method here) is purely to show how well the proposed method works – close to an oracle which operates in a fictitious interference-free environment.

SVD with interference: we will use SVD to project away the interference subspace by projecting on the third principal component of the matrix $\mathbf{Y}$ to estimate the secondary signal. Notice that projecting on the first two components yields the subspace containing the primary user signals, $\mathbf{p}_1$ and $\mathbf{p}_2$.

In order to resolve the scaling ambiguity that is inherent both in the proposed CCA method and the SVD-based baselines, we assume that the first four secondary symbols are known at the SRx. Note that these symbols can be drawn from the packet header that contains the STx identification sequence.

Remark 10. *It is worth noting that for the second baseline (SVD with interference), we use our proposed* blind *method in Section 5.5 to recover the secondary packet start time index at the SRx, thereby giving a big advantage to the SVD based method. The typical synchronization method that would be used with SVD is to allow the STx to transmit a long pilot sequence, long enough to make up for the large power difference between the two users. Then, we would use knowledge of this pilot sequence at the SRx to find the start time index of the secondary signal via cross-correlation / matched filtering. This would seriously reduce the transmission rate of the secondary user relative to our proposed blind method, especially for the setting considered herein where the secondary user is much weaker than the primary. Further, and perhaps worse, such training-based timing recovery requires the SRx to estimate the secondary CFO before (or together with) timing synchronization, which is in another serious complication given the low SINR and moderate SNR of the secondary user.*

Primary receiver. At the PRx, we use energy detection for the primary packet detection. Then, we use primary training symbols to detect the start index of the 256×1 received signal of the primary user. To decode the primary symbols, we use 10 training pilots to estimate the primary channel coefficient and then do the hard detection of the equalized signal.

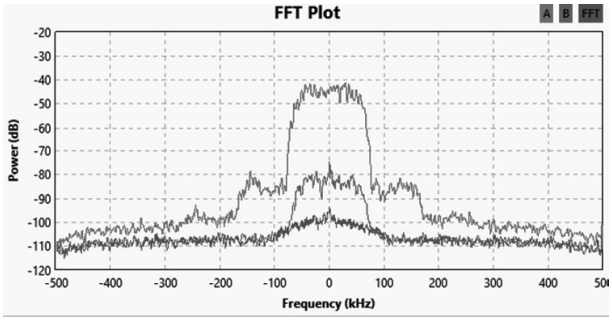


Figure 5.4: GNU radio spectrum analyzer showing 40 dB received power difference between the PTx and STx at the SRx. The received signal of the PTx is shown in red while that of the STx is depicted in purple, and the noise in blue.

Tx power difference (dB)	20	25	30	35	40
SINR (1^{st} antenna)	-17.1213	-20.1632	-27.1965	-29.1996	-32.2388
SINR (2^{nd} antenna)	-15.1248	-18.1909	-25.2433	-30.2522	-31.2015

Table 5.2: Estimated secondary SINR at the SRx over the two receive channels, across the different transmit power imbalance scenarios. The measured average secondary SNR is around 8 dB.

5.6.2 Performance Evaluation

Since we assume digitally-modulated signals for both users, we will use the symbol error rate (SER) as a performance metric (but recall that our method can also work with analog modulation for the primary and the secondary user).

In the first experiment, we test the performance of the proposed approach under different levels of primary interference at the secondary receiver. To do so, we fix the secondary transmit power to -18 dBm. This makes the corresponding measured average received SNR at the SRx equal to approximately 8 dB. We vary the primary transmit power from 0 to 20 dBm in 5 dB steps, thus generating transmit power differences from approximately -20 dB down to a rather extreme -40 dB. To validate the power difference between the two users at the SRx, Fig. 5.4 shows the GNU radio spectrum analyzer at the SRx with the received signal strength level of the PTx and STx in addition to the noise level – the transmit power of the PTx is set to 20 dBm, and Fig. 5.4 shows close to 40 dB power difference between the two users. Furthermore, Fig. 5.5 depicts the squared samples of one of the received packets at the SRx after matched filtering with the SRRC, for primary transmit power set to 15 dBm. It is clear that part of the secondary transmitted packet overlaps with the padded zeros of the primary packet, showcasing the power difference between the two users. Further, the remaining zeros show the low received

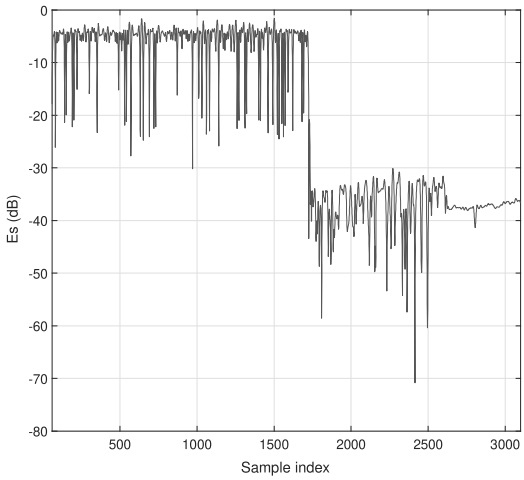
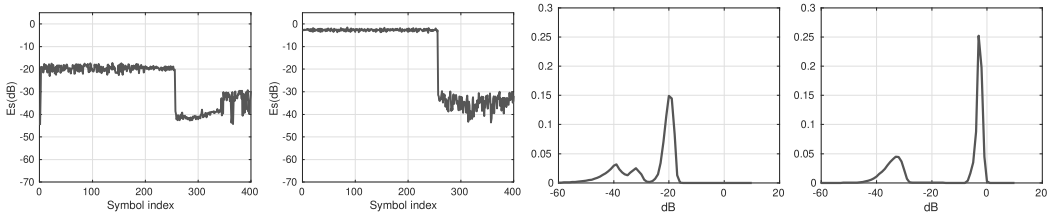


Figure 5.5: Squared samples of one of the received packets after matched filtering with the SRRC. Notice the overlap between part of the secondary packet and the zeros of the primary, and also the low SNR of the secondary.

SNR range of the secondary user.

To compute the received SNR and SINR of the secondary signal at the SRx, we exploit the padded zeros in both the primary and secondary signals to measure the noise power, the secondary signal power, and the primary signal power at the SRx. In particular, we estimate the probability distribution of the symbol energy across 1500 packets, each of length 400 symbols. From the distribution, one can estimate either two peaks or three peaks, depending on the overlap between the secondary (primary) and the zeros of the primary (secondary). For instance, Fig. 5.6(a) clearly shows one of the received packets in one of the channels for the 20 dB transmit power imbalance case. One can clearly see the three different energy levels: one for the (primary, secondary and noise), another for (secondary and noise), and one for noise only. Notice that the first level can also be primary and noise, but since the primary is very strong, treating the first level as (primary and noise) or (primary, secondary and noise) will have negligible impact on the SINR and SNR measurements of the secondary user. Fig. 5.6(c), shows the histogram of the collected data across 1500 packets for the 20 dB transmit power difference, where three distinct peaks are observed. In Fig. 5.6(b), however, one can see a complete overlap between part of the secondary signal and the padded zeros of the primary user for the 40 dB transmit power difference, and hence, only two peaks can be seen in the distribution shown in Fig. 5.6(d).

We use the data collected for the 20 dB transmit power difference to measure the energy levels corresponding to the three observed probability density peaks, see Fig. 5.6(c). We use



(a) (b) (c) (d)

Figure 5.6: Example of the received primary user's packets at the SRx after matched filtering with the SRRC for 20 dB and 40 dB transmit power difference. Plots a) and b) depict the symbol energy of the detected packet, for the two transmit power imbalance scenarios, while (c) and (d) correspond to the estimated probability distribution of the energy (in dB) of the detected symbols for the 20 dB and 40 dB transmit power difference cases, respectively.

these values to solve a system of linear equations (three equations in three unknowns) to compute the received SNR and SINR at the secondary receiver. We repeat the same procedure for the different transmit power difference cases to calculate the associated SINR and SNR values. Note that, since the secondary transmit power is fixed throughout this experiment, we observed approximately the same average energy level (peak value) for either the noise level or the (secondary and noise) level, across all the transmit power difference cases. However, as expected, we observed increase in the estimated energy level that corresponds to the primary, secondary and noise. To confirm this, one can see from Fig. 5.6(b) a complete overlap between part of the secondary signal and the padded zeros of the primary user for the 40 dB transmit power difference case, and hence, only two peaks can be seen in the distribution shown in Fig. 5.6(d). Notice that the energy level associated with the smallest peak (secondary and noise) in Fig. 5.6(d) is roughly equivalent to the energy level associated with the middle peak in Fig. 5.6(c), while one can easily see close to 20 dB increase in the highest peak (primary, secondary and noise) in Fig. 5.6(d) relative to Fig. 5.6(c). The measured SINR values for the different transmit power cases are reported in Table II.

In order to demonstrate the capability of our proposed approach to correctly decode the secondary transmission at very low SINR, we report the SER of the secondary user obtained by our proposed CCA method at five different levels of the (average) transmit power imbalance: from -20 dB to -40 dB (corresponding secondary SINR levels are reported in Table III). Fig. 5.7 depicts SER results obtained by our proposed CCA method, for all five levels of primary interference, and the corresponding SER curve obtained using the SVD-based method at the same SNR *without any interference*. The results are striking: CCA is remarkably insensitive to interference from the primary user. In particular, CCA achieves almost the same performance at power difference levels $(-35, -30, -20, -25)$ dB. On the other hand, at the -40 dB level, the

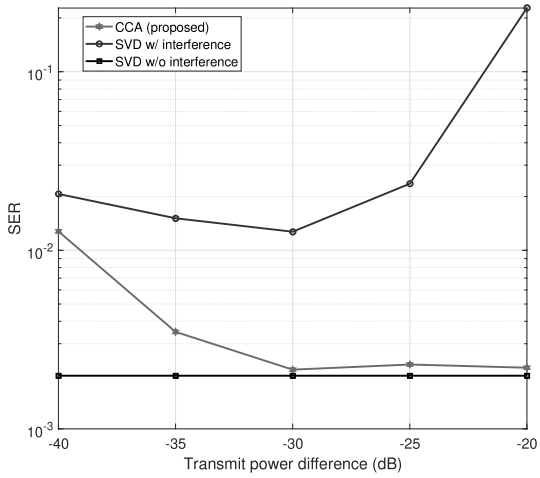


Figure 5.7: Secondary user detection performance at different average SINR levels. The measured average secondary SNR is approximately 8 dB.

CCA performance degrades. This mainly happens due to the limited resolution of the analog to digital converter of our USRP for the wide dynamic range of the input signal – while the average SINR is -32 dB, there are several instances where it drops below -40 dB, and these occasional quantization errors ultimately dominate CCA performance. Despite that, CCA still achieves close to 10^{-2} SER. Finally, one can see that CCA significantly outperforms the SVD method used for interference cancellation, even though the latter is in fact aided by the CCA frame structure to acquire timing – a benefit which it won't have in practice. As shown in Fig. 5.7, SVD performance breaks at 25 dB transmit power difference, where primary subspace estimation becomes very difficult, and hence interference cancellation does not work.

Considering the primary user's performance, we observed that the single-antenna primary receiver is completely insensitive to the secondary interference. Fig. 5.9(a) shows one of the received packets at the PRx (before down sampling), with the primary transmit power set to 0 dBm (minimum primary power in this experiment), while the secondary user is inactive. On the other hand, Fig. 5.9(b) shows one of the received packets at the PRx (before down sampling) when the secondary user is active, where there is approximately 70% overlap between the two users' packets. We observed that in the worst case setting, where the primary user power is fixed to its minimum level (highest interference from the secondary user), the same detection performance can be attained regardless whether the secondary user is active or not. This is due to the fact that the secondary interference is close to the primary's noise floor, as one can see from Fig. 5.9(c) and Fig. 5.9(d), where the two smaller peaks in Fig. 5.9(c) and Fig. 5.9(d)

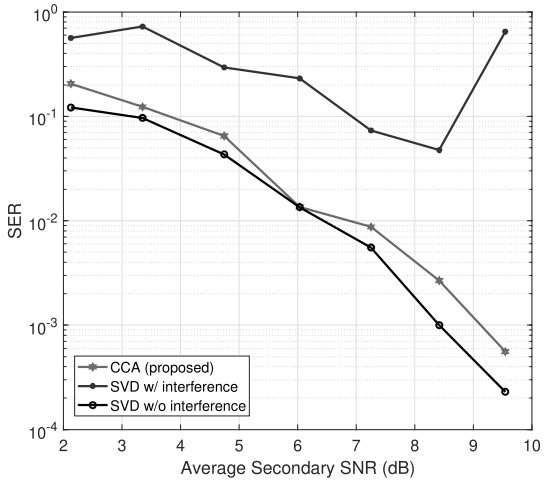
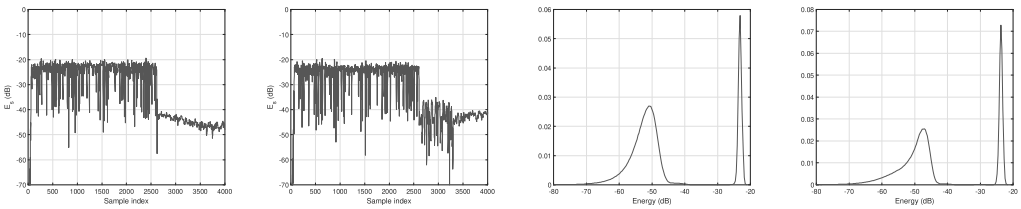


Figure 5.8: Secondary user detection performance at different SNR.



(a) (b) (c) (d)

Figure 5.9: (a) primary packet samples when the STX is inactive. (b) primary packet samples when the STX is active. (c) energy distribution when the STx is inactive. (d) energy distribution when the STx is active.

correspond to the noise level and the secondary plus noise level, respectively. We observed that the SNR of the primary user is 28 dB when the secondary user is inactive, while the primary user's SINR is 25 dB when the secondary user is active.

Next, we consider another experiment to see the performance of the proposed method under different SNR values for the secondary user. To do so, we fixed the primary transmit power to 10 dBm and varied the secondary transmit power from -23 to -17 dBm which corresponds to average SNR values between 2 dB and 10 dB, as observed. At each SNR value, we report the SER of the secondary user. Fig. 5.8 depicts the SER performance of the secondary user versus its SNR. It is obvious how well our proposed method works at very low SNR / SINR values. In particular, our method can achieve 10^{-2} SER at 7 dB and closely approaches what is attained by the interference-free SVD baseline at low SNR values. Further, one can see that the

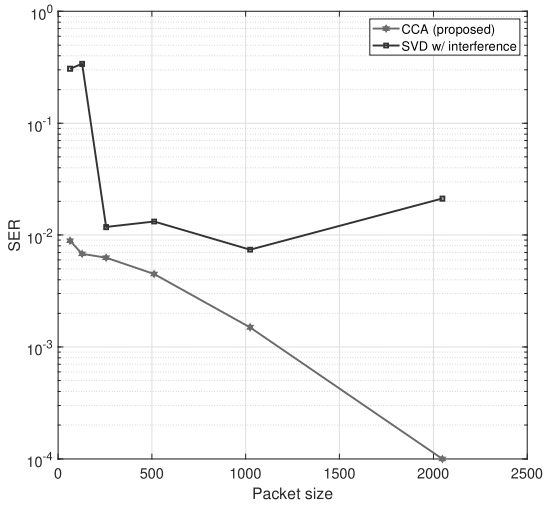


Figure 5.10: Secondary user detection performance for different packet sizes of the secondary user. The packet length of the primary user is fixed to 256 QPSK symbols.

SVD with interference completely fails at both the low SNR and high SNR regions, where in the latter, the secondary user becomes a bit more stronger and then accurate primary subspace estimation becomes more difficult as explained in the previous experiment.

On the other hand, we observed that the secondary user does not affect the primary performance, which remains the same as is attained when the secondary user is inactive. The same SER is observed at the PRx, even at the extreme case where the secondary transmit power is -17 dBm (i.e., the highest interference to the primary).

In our next experiment, we test the performance of the proposed method as a function of secondary packet size. The secondary and primary transmit powers are fixed to -20 dBm and 5 dBm, respectively. The measured average SNR at the secondary receiver is 7 dB. The primary packet length is set to 256 QPSK symbols. Fig. 5.10 shows the SER performance of the proposed approach versus the packet size of the secondary user. We observe a significant improvement in the secondary SER when the secondary packet length increases. This is due to the fact that increasing N renders the transmit sequences closer to being orthogonal and having low auto-correlation sidelobes, which improves the performance of CCA and secondary timing synchronization. Note that our performance analysis in A.1 shows that increasing the packet length yields higher canonical correlation coefficient, and hence a better estimate for the common signal. This suggests that transmitting longer secondary packets provides better secondary detection performance. On the other hand, one can argue that if the channel is fast

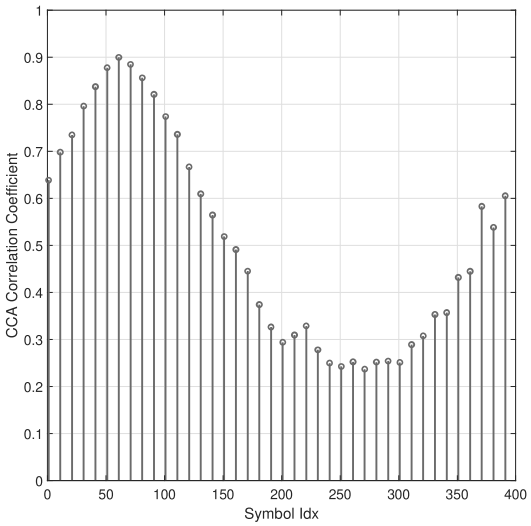


Figure 5.11: Secondary user synchronization using CCA.

time varying, then the higher the packet length, the higher the probability of each block being subject to channel variation, thus violating the presumed mode. Hence, in setting the secondary packet length one has to take into account the coherence time of the channel, in order to choose the optimal packet length for the secondary user.

Finally, we evaluate the performance of the proposed algorithm for finding the start time of the secondary packet. We use the same parameters as the previous experiment but the secondary packet is fixed to 256 symbols. Recall that the received packet length, before down-sampling, is 4020 samples. To find the start of the 256×2 signal, we run Algorithm 3 with a step of 10 symbols on the received signal, which resulted in solving approximately a series of 40 CCA problems. Fig. 5.11 shows that the highest correlation coefficient is attained at symbols index 60. We then performed an additional narrow (fine) search over a window size of 10 symbols centered at the obtained symbol index from the wide search.

5.7 Summary

In this chapter, we proposed a practical low-complexity data-driven spectrum sharing approach for an asynchronous underlay scenario involving a high-power primary user and a low-power secondary link. The proposed method allows the secondary user to reliably communicate over the same channel occupied by the primary, without any coordination, and without any channel

state information. Our proposed solution is based on "repetition coding": the secondary user transmits its signal twice at very low power such that it does not affect the primary user detection performance. Constructing two signal views at the SRx and applying CCA to these views, we showed that the secondary receiver can reliably decode its intended signal at moderate SNR even if it is buried under strong interference from the primary user transmission. We proposed a low-complexity unsupervised based approach that can resolve the crucial low-SINR synchronization issue at the secondary receiver. Laboratory experiments using a custom-built USRP testbed confirmed the efficacy of the proposed method in decoding the secondary signal at very low SINR in real world wireless environments.

The proposed framework can guarantee reliable reception of the secondary underlay signal even under time-varying and intermittent interference from the primary user. Specifically, our theoretical results show that the secondary signal can be identified even if the primary channel is different across the two secondary signal blocks. To the best of our knowledge, this is the first spectrum underlay work that allows a low-power secondary user to occupy the channel with a time varying primary user in a realistic wireless environment, without i) requiring any knowledge about the primary network (waveform, modulation, channel, timing, etc.), ii) coordination between the primary and the secondary system, iii) long pilot sequences for acquisition and channel estimation for the secondary user.

Chapter 6

Summary and Future Research Directions

This dissertation introduced new and widely useful theortical findings of CCA and its multiview generalization GCCA, that allowed for innovative solutions of challenging problems in next-generation cellular systems. In this chapter, we will summarize our main contributions and refer to other possible research directions to pursue.

6.1 Dissertation Summary

In this dissertation, we studied the widely-used statistical learning tool, CCA, in the context of modern wireless communications. Chapter 2 introduced CCA and GGCA by i) presenting their various formulations and the corresponding computational complexities, ii) referring to the wide domain of applications that have been handled by (G)CCA, iii) explaining the plain-vanilla probabilistic interpretation of CCA, and iv) introducing our new algebraic interpretation of CCA as a method that can identify a common (shared) subspace between two or more matrices.

In the subsequent chapters, we showed how effective our algebraic interpretation is, through showing the potential of (G)CCA in handling challenging problems in wireless communications. In Chapter 3, we made two theortical contributions of broader interest. First, we provided an identifiability analysis of broader interest, and can have impact in several other disciplines. First, we showed that CCA can identify the common subspace between two matrices under mild conditions, assuming a linear general model, no matter how strong the individual components are. Second, we derived performance analysis for practical application under non-ideal conditions (notably noise). We showed through those theortical findings that CCA can efficiently handle the

challenging problem of reliably detecting cell-edge users in cellular wireless systems, even when they are buried under intra and inter-cell interference. Further, we developed an algorithmic framework based on CCA to resolve synchronizations issues.

Chapter 4 generalized the cell-edge problem setting considered in Chapter 3 to the more challenging setup comprising multiple BSs. We showed that the best detection performance can be attained when using GCCA on the received signals from three closest BSs, and that adding more BSs (more than three) never helps but degrades the performance. We present a performance analysis which showed that the cell-edge user signal subspace can be accurately estimated via applying GCCA on the received signals from the three closest BSs. We also GCCA based technique that can accurately estimate the common subspace dimension, i.e., the technique can correctly classify whether a user is cell-edge or cell-center.

In Chapter 5, we showed how CCA can be invoked to improve spectrum utilization. We proposed an underlay strategy that allows simultaneous operation of two independent networks without affecting each others. The starting point was to design a simple secondary transmission protocol that operates at very low power, but allows reliable secondary communication without requiring any channel knowledge or coordination with the primary system. We showed that via a repetition coding based scheme employed at the secondary transmitter together with applying CCA at the secondary receiver, the secondary signal can be blindly decoded at very low SINR. The end-to-end approach is practically appealing at it merely requires solving for a principal eigenvector that can be cheaply computed via the power method. We validated our proposed framework in the lab, where we demonstrated that our approach works in real world wireless environment. In particular, we have developed a testbed comprising several USRPs to show that seamless secondary communication is possible in the presence of strong interference from the primary network.

6.2 Future Directions

Based on this dissertation, our future and ongoing research will be focused on the following extensions and applications:

- **Weak target detection in MIMO radar.** The radar technology has witnessed significant progression over the last years. With a variety of applications ranging from air traffic control, security surveillance, military applications, to a plethora of medical applications, modern radar systems are now deployed worldwide. Conventional radar detection and estimation methods usually suffer from scalability issues, and hence, it is desirable to develop new low complexity detection and estimation tools that can efficiently detect and estimate the targets' parameters of interest. In our recent work [31], a CCA-based approach was

developed to reliably detect and accurately estimate the parameter of interests of weak targets. It was shown that by designing two partially overlapping spatial beams and judiciously controlling the degree of their overlap, it is possible to create two views of the signal space that share only the (potential) target of interest at a certain range-DOA cell. Using CCA on these two views, it is possible to detect the sought target, at low SNR and even under strong nearby interference. The promising preliminary results obtained from our simple CCA approach in [31] prompt us to pursue further analytic and experimental development of the CCA approach. In particular, our theortical results in [14, 15] are not directly applicable to the considered beam space model which assumes Vandermonde structure on some of the matrices [31]. Therefore, We seek to deepen the CCA theory through developing a new performance analysis, that shows under which conditions on the angle of arrivals (AoA) and angles of departures (AoD) of the targets, the weak target can be recovered via CCA, thereby discovering under which radar scenarios CCA can work.

Another interesting problem to look at is how to create the two data views to accurately estimate the parameters of interest. Although the two data views were constructed through creating two partially overlapping spatial beams and carefully controlling the overlap degree, it will be more interesting if we create the two views via designing two partially overlapping band-pass filters in the Doppler domain, and then use CCA to estimate the AoA and AoD of the weak target. This way, one can deal with the clutter effect (reflections from ground, oceans, $\cdots$, etc.), which was assumed to be filtered out on [31]. In particular, designing band-pass filter in the Doppler domain can aggressively suppress the clutter term because in practice the Doppler frequency of the clutter is significantly lower than that of the targets. Therefore, we aim at designing new CCA-based method that can solve the problem under more realistic scenarios, and to test it in real experiments as clutter data is publicly available.

- **Unsupervised signal detection in unknown multipath channels.** Multipath effect refers to the reception of multiple independently faded/attenuated and delayed copies of the original transmitted signal. The superposition of these versions of the signal is quite different form the original signal given that destructive and constructive interference might occur. On one hand, owing to the time shift, multipath components introduce inter-symbol interference which together with the co-channel interference represent two major obstacles that can significantly degrade the quality of the received signal and severely impact the reliability of communication. On the other hand, this availability of several copies of the signal results in an increase in the number of degrees of freedom, which if exploited effectively, can enhance the reliability of communication.

Given a mixture of co-channel user signals subject to frequency-selective multipath, sensed through an array of co-located antennas, how can we recover the user signals? In our recent work [38], we have proposed a CCA-based framework for reliably recovering user signals in unknown multipath environment. The method was shown very effective in terms of accurately identifying the user relative delays and reliably decoding the user signals at low SNR. However, the CCA-based approach was developed for a special scenario, where it was assumed that i) different users have distinct relative delay profiles, and ii) each user receives its signal through two paths. It will be interesting to study the more general setting of the problem where the number of paths for the k-th user is $L_k > 2$, where $k \in \{1, \cdots, K\}$. This is in fact renders the problem much more challenging as it introduces more ISI, and more importantly, the number of relative delays in that case will be $\sum_{k=1}^{K} \binom{L_k}{2}$. Thus, a selection scheme needs to be developed to identify which relative delays can be chosen to recover the user signals. Another interesting point is to consider time varying channels and see how CCA works in that case, and developing an analysis that puts conditions on the range of the Doppler frequencies under which can still identify the user signals.

- **Performance analysis for fading channels and connections to multi-terminal information theory.** It will be very interesting to see how well our proposed underlay approach, in Chapter 5, works under fading scenarios. Our results indicate that CCA still works well under random fading, but GCCA with more than two views can certainly help here, together with error control coding. Performance analysis under Rician fading (to better model the primary user) and under time-selective fading models will be of broad interest. There is very limited performance analysis available for CCA, and almost none for GCCA, with the exception of our own prior work that employed a simple i.i.d. Rayleigh model for fading [14,15]. The other key performance analysis question is how well does the proposed approach fare relative to information-theoretic capacity analysis of the two-user multi-terminal interference channel. Note that our scheme does not assume non causal knowledge of interference or even channel state information, and no coordination between the two users. It is interesting to see how well it fares in terms of the pair of rates it can support relative to the attainable rates (capacity region).

References

[1] Intel, *A guide to the internet of things infographic*, 2020. [Online]. Available: https://www.intel.com/content/www/us/en/internet-of-things/infographics/guide-to-iot.html.

[2] D. Gündüz, P. de Kerret, N. D. Sidiropoulos, D. Gesbert, C. R. Murthy, and M. van der Schaar, "Machine learning in the air," *IEEE Journal on Selected Areas in Communications*, vol. 37, no. 10, pp. 2184–2199, Sept. 2019.

[3] J. Jagannath, N. Polosky, A. Jagannath, F. Restuccia, and T. Melodia, "Machine learning for wireless communications in the Internet of Things: A comprehensive survey," *Ad Hoc Networks*, vol. 93, p. 101913, 2019.

[4] H. Sun, X. Chen, Q. Shi, M. Hong, X. Fu, and N. D. Sidiropoulos, "Learning to optimize: Training deep neural networks for wireless resource management," in *Signal Processing Advances in Wireless Communications (SPAWC)*, Japan, July 2017, pp. 247–252.

[5] M. S. Ibrahim, A. S. Zamzam, X. Fu, and N. D. Sidiropoulos, "Learning-based antenna selection for multicasting," in *IEEE 19th International Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*, Kalamata, Greece, Aug. 2018, pp. 1–5.

[6] J. Joung, "Machine learning-based antenna selection in wireless communications," *IEEE Communications Letters*, vol. 20, no. 11, pp. 2241–2244, 2016.

[7] H. Ye, G. Y. Li, and B.-H. Juang, "Power of deep learning for channel estimation and signal detection in ofdm systems," *IEEE Wireless Communications Letters*, vol. 7, no. 1, pp. 114–117, Sept. 2018.

[8] S. Wold, K. Esbensen, and P. Geladi, "Principal component analysis," *Chemometrics and Intelligent Laboratory Systems*, vol. 2, no. 1-3, pp. 37–52, 1987.

[9] A. Hyvärinen and E. Oja, "Independent component analysis: Algorithms and Applications," *Neural Networks*, vol. 13, no. 4-5, pp. 411–430, 2000.

[10] Y. Koren, R. Bell, and C. Volinsky, "Matrix factorization techniques for recommender systems," *Computer*, vol. 42, no. 8, pp. 30–37, 2009.

[11] P. Paatero and U. Tapper, "Positive matrix factorization: A non-negative factor model with optimal utilization of error estimates of data values," *Environmetrics*, vol. 5, no. 2, pp. 111–126, 1994.

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

[13] M. S. Ibrahim and N. D. Sidiropoulos, "Cell-edge interferometry: Reliable detection of unknown cell-edge users via canonical correlation analysis," in *IEEE International Conference on Signal Processing Advances in Wireless Communications (SPAWC)*, Cannes, France, July 2019, pp. 1–5.

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

[15] M. S. Ibrahim and N. D. Sidiropoulos, "Cell-edge detection via selective cooperation and generalized canonical correlation," *IEEE Trans. Wireless Commun.*, May 2021.

[16] M. S. Ibrahim and N. D. Sidiropoulos, "Blind carbon copy on dirty paper: seamless spectrum underlay via canonical correlation analysis," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Toronto, Canada, June 2021.

[17] J. D. Carroll, "Generalization of canonical correlation analysis to three or more sets of variables," in *Proceedings of the 76th annual convention of the American Psychological Association*, vol. 3, 1968, pp. 227–228.

[18] 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, 2004.

[19] J. Rupnik, P. Skraba, J. Shawe-Taylor, and S. Guettes, "A comparison of relaxations of multiset cannonical correlation analysis and applications," *arXiv preprint arXiv:1302.0974*, 2013.

[20] L.-H. Zhang, L.-Z. Liao, and L.-M. Sun, "Towards the global solution of the maximal correlation problem," *Journal of Global Optimization*, vol. 49, no. 1, pp. 91–107, Feb. 2011.

[21] C. I. Kanatsoulis, X. Fu, N. D. Sidiropoulos, and M. Hong, "Structured sumcor multiview canonical correlation analysis for large-scale data," *IEEE Transactions on Signal Processing*, vol. 67, no. 2, pp. 306–319, Jan 2019.

[22] X. Fu, K. Huang, E. E. Papalexakis, H. Song, P. P. Talukdar, N. D. Sidiropoulos, C. Faloutsos, and T. Mitchell, "Efficient and distributed algorithms for large-scale generalized canonical correlations analysis," in *IEEE 16th International Conference on Data Mining (ICDM)*, Dec 2016, pp. 871–876.

[23] X. Fu, K. Huang, M. Hong, N. D. Sidiropoulos, and A. M. So, "Scalable and flexible multiview MAX-VAR canonical correlation analysis," *IEEE Transactions on Signal Processing*, vol. 65, no. 16, pp. 4150–4165, Aug 2017.

[24] M. Sorensen, C. I. Kanatsoulis, and N. D. Sidiropoulos, "Generalized canonical correlation analysis: A subspace intersection approach," *IEEE Transactions on Signal Processing*, pp. 1–1, Feb. 2021.

[25] C. F. Van Loan and G. H. Golub, *Matrix computations*. Baltimore, MD, USA: Johns Hopkins University Press Baltimore, 1983.

[26] H. Ge, I. P. Kirsteins, and X. Wang, "Does canonical correlation analysis provide reliable information on data correlation in array processing?" in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Taiwan, Apr. 2009, pp. 2113–2116.

[27] 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, 2002.

[28] J. Via and I. Santamaria, "Adaptive blind equalization of SIMO systems based on canonical correlation analysis," in *IEEE 6th Workshop on Signal Processing Advances in Wireless Communications*, NY, USA, June 2005, pp. 318–322.

[29] 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, 1994.

[30] Z. Bai, G. Huang, and L. Yang, "A radar anti-jamming technology based on canonical correlation analysis," in *International Conference on Neural Networks and Brain*, vol. 1, China, Oct. 2005, pp. 9–12.

[31] M. S. Ibrahim and N. D. Sidiropoulos, "Weak target detection in MIMO radar via beamspace canonical correlation," in *IEEE 11th Sensor Array and Multichannel Signal Processing Workshop (SAM)*, Hangzhou, China, June 2020, pp. 1–5.

[32] Y.-O. Li, T. Adali, W. Wang, and V. D. Calhoun, "Joint blind source separation by multiset canonical correlation analysis," *IEEE Transactions on Signal Processing*, vol. 57, no. 10, pp. 3918–3929, 2009.

[33] M. Borga and H. Knutsson, "A canonical correlation approach to blind source separation," *Report LiU-IMT-EX-0062 Department of Biomedical Engineering, Linkping University*, 2001.

[34] W. Liu, D. P. Mandic, and A. Cichocki, "Analysis and online realization of the CCA approach for blind source separation," *IEEE Transactions on Neural Networks*, vol. 18, no. 5, pp. 1505–1510, 2007.

[35] 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, 2015.

[36] O. Friman, M. Borga, P. Lundberg, and H. Knutsson, "Exploratory fMRI analysis by autocorrelation maximization," *NeuroImage*, vol. 16, pp. 454–464, June 2002.

[37] N. M. Correa, Y.-O. Li, T. Adali, and V. D. Calhoun, "Canonical correlation analysis for feature-based fusion of biomedical imaging modalities and its application to detection of associative networks in schizophrenia," *IEEE Journal of Selected Topics in Signal Processing*, vol. 2, no. 6, pp. 998–1007, Dec. 2008.

[38] M. S. Ibrahim and N. D. Sidiropoulos, "Delay-locking: Unraveling multiple unknown signals in unknown multipath," in *IEEE 21st International Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*, Atlanta, GA, USA, Aug. 2020, pp. 1–5.

[39] R. Arora and K. Livescu, "Multi-view learning with supervision for transformed bottleneck features," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Italy, May 2014, pp. 2499–2503.

[40] M. W. Browne, "The maximum-likelihood solution in inter-battery factor analysis," *British Journal of Mathematical and Statistical Psychology*, vol. 32, no. 1, pp. 75–86, 1979.

[41] F. R. Bach and M. I. Jordan, "A probabilistic interpretation of canonical correlation analysis," 2005.

[42] S. G. Kiani, G. E. Oien, and D. Gesbert, "Maximizing multicell capacity using distributed power allocation and scheduling," in *IEEE Wireless Communications and Networking Conference (WCNC)*, China, March 2007, pp. 1690–1694.

[43] A. Tolli, M. Codreanu, and M. Juntti, "Cooperative MIMO-OFDM cellular system with soft handover between distributed base station antennas," *IEEE Transactions on Wireless Communications*, vol. 7, no. 4, pp. 1428–1440, 2008.

[44] S. Verdu *et al.*, *Multiuser detection*. Cambridge University Press, 1998.

[45] J. G. Proakis and M. Salehi, *Digital communications*. McGraw-hill New York, 2001, vol. 4.

[46] H. Vikalo and B. Hassibi, "The expected complexity of sphere decoding, part I: Theory, part II: Applications," *IEEE Transactions on Signal Processing*, vol. 53, no. 8, pp. 2819–2834, 2005.

[47] J. Jaldén and B. Ottersten, "On the complexity of sphere decoding in digital communications," *IEEE Transactions on Signal Processing*, vol. 53, no. 4, pp. 1474–1484, 2005.

[48] P. H. Tan and L. K. Rasmussen, "The application of semidefinite programming for detection in CDMA," *IEEE Journal on Selected Areas in Communications*, vol. 19, no. 8, pp. 1442–1449, 2001.

[49] M. Abdi, H. El Nahas, A. Jard, and E. Moulines, "Semidefinite positive relaxation of the maximum-likelihood criterion applied to multiuser detection in a CDMA context," *IEEE Signal Processing Letters*, vol. 9, no. 6, pp. 165–167, 2002.

[50] Z. Luo, W. Ma, A. M. So, Y. Ye, and S. Zhang, "Semidefinite relaxation of quadratic optimization problems," *IEEE Signal Processing Magazine*, vol. 27, no. 3, pp. 20–34, May 2010.

[51] U. Madhow and M. L. Honig, "MMSE interference suppression for direct-sequence spread-spectrum CDMA," *IEEE Transactions on Communications*, vol. 42, no. 12, pp. 3178–3188, 1994.

[52] S. Moshavi, "Multi-user detection for DS-CDMA communications," *IEEE Communications Magazine*, vol. 34, no. 10, pp. 124–136, 1996.

[53] A. Duel-Hallen, "Decorrelating decision-feedback multiuser detector for synchronous code-division multiple-access channel," *IEEE Transactions on Communications*, vol. 41, no. 2, pp. 285–290, 1993.

[54] P. D. Alexander, M. C. Reed, J. A. Asenstorfer, and C. B. Schlegel, "Iterative multiuser interference reduction: Turbo CDMA," *IEEE Transactions on Communications*, vol. 47, no. 7, pp. 1008–1014, 1999.

[55] G. Boudreau, J. Panicker, N. Guo, R. Chang, N. Wang, and S. Vrzic, "Interference coordination and cancellation for 4G networks," *IEEE Communications Magazine*, vol. 47, no. 4, 2009.

[56] N. Himayat, S. Talwar, A. Rao, and R. Soni, "Interference management for 4G cellular standards [WIMAX/LTE UPDATE]," *IEEE Communications Magazine*, vol. 48, no. 8, 2010.

[57] S. Agarwal, S. De, S. Kumar, and H. M. Gupta, "Qos-aware downlink cooperation for cell-edge and handoff users," *IEEE Transactions on Vehicular Technology*, vol. 64, no. 6, pp. 2512–2527, 2014.

[58] C. U. Castellanos, D. L. Villa, C. Rosa, K. I. Pedersen, F. D. Calabrese, P.-H. Michaelsen, and J. Michel, "Performance of uplink fractional power control in utran LTE," in *Vehicular Technology Conference (VTC)*, Calgary, Canada, Sept. 2008, pp. 2517–2521.

[59] H. R. Chayon, K. B. Dimyati, H. Ramiah, and A. W. Reza, "Enhanced quality of service of cell-edge user by extending modified largest weighted delay first algorithm in LTE networks," *Symmetry*, vol. 9, no. 6, p. 81, 2017.

[60] H. Xu and P. Ren, "Joint user scheduling and power control for cell-edge performance improvement in backhaul-constrained network MIMO," in *IEEE 24th Annual International Symposium on Personal, Indoor, and Mobile Radio Communications (PIMRC)*, London, UK, Sept. 2013, pp. 1342–1346.

[61] S. Khattak, *Base Station Cooperation Strategies for Multi-user Detection in Interference Limited Cellular Systems*. Jörg Vogt Verlag, 2008.

[62] A.-J. van der Veen, "Analytical method for blind binary signal separation," *IEEE Transactions on Signal Processing*, vol. 45, no. 4, pp. 1078–1082, 1997.

[63] T. Mayer, H. Jenkac, and J. Hagenauer, "Turbo base-station cooperation for intercell interference cancellation," in *IEEE International Conference on Communications (ICC)*, vol. 11, Turkey, June 2006, pp. 4977–4982.

[64] C. Hoymann, L. Falconetti, and R. Gupta, "Distributed uplink signal processing of cooperating base stations based on IQ sample exchange," in *IEEE International Conference on Communications (ICC)*, Germany, June 2009, pp. 1–5.

[65] E. Björnson, E. G. Larsson, and M. Debbah, "Massive MIMO for maximal spectral efficiency: How many users and pilots should be allocated?" *IEEE Transactions on Wireless Communications*, vol. 15, no. 2, pp. 1293–1308, 2016.

[66] X. Li, E. Bjornson, E. G. Larsson, S. Zhou, and J. Wang, "A multi-cell MMSE detector for massive MIMO systems and new large system analysis," in *IEEE Global Communications Conference (GLOBECOM)*, CA, USA, Dec. 2015, pp. 1–6.

[67] H. Q. Ngo, M. Matthaiou, and E. G. Larsson, "Performance analysis of large scale MU-MIMO with optimal linear receivers," in *Swedish Communication Technologies Workshop (Swe-CTW)*, Sweden, Oct. 2012, pp. 59–64.

[68] M. Wax and T. Kailath, "Detection of signals by information theoretic criteria," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 33, no. 2, pp. 387–392, 1985.

[69] A. J. Paulraj, D. A. Gore, R. U. Nabar, and H. Bolcskei, "An overview of MIMO communications-a key to Gigabit wireless," *Proceedings of the IEEE*, vol. 92, no. 2, pp. 198–218, Nov. 2004.

[70] M. Jiang and L. Hanzo, "Multiuser MIMO-OFDM for next-generation wireless systems," *Proceedings of the IEEE*, vol. 95, no. 7, pp. 1430–1469, Aug. 2007.

[71] X. You, D. Wang, P. Zhu, and B. Sheng, "Cell edge performance of cellular mobile systems," *IEEE J. Sel. Areas Commun.*, vol. 29, no. 6, pp. 1139–1150, June 2011.

[72] M. H. A. Khan, J.-G. Chung, and M. H. Lee, "Downlink performance of cell edge using cooperative BS for multicell cellular network," *EURASIP Journal on Wireless Communications and Networking*, vol. 2016, no. 1, p. 56, Feb. 2016.

[73] J. R. Kettenring, "Canonical analysis of several sets of variables," *Biometrika*, vol. 58, no. 3, pp. 433–451, 1971.

[74] S. Khattak, W. Rave, and G. Fettweis, "Distributed iterative multiuser detection through base station cooperation," *EURASIP Journal on Wireless Communications and Networking*, vol. 2008, p. 17, Feb. 2008.

[75] A. M. Rao, "Reverse link power control for managing inter-cell interference in orthogonal multiple access systems," in *IEEE 66th Vehicular Technology Conference*, MD, USA, Oct. 2007, pp. 1837–1841.

[76] S. Das and H. Viswanathan, "Interference mitigation through intelligent scheduling," in *Proceedings of the Asilomar Conference on Signals and Systems (Asilomar)*, CA, USA, Nov. 2006.

[77] K. Balachandran, J. H. Kang, K. Karakayali, and K. M. Rege, "Nice: A network interference cancellation engine for opportunistic uplink cooperation in wireless networks," *IEEE Trans. Wireless Commun.*, vol. 10, no. 2, pp. 540–549, Dec. 2010.

[78] P. Horst, "Generalized canonical correlations and their applications to experimental data," *Journal of Clinical Psychology*, vol. 17, no. 4, pp. 331–347, 1961.

[79] 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, Oct. 2010.

[80] R. Zhang, "Cooperative multi-cell block diagonalization with per-base-station power constraints," *IEEE J. Sel. Areas Commun.*, vol. 28, no. 9, pp. 1435–1445, Juy 2010.

[81] E. G. Larsson, O. Edfors, F. Tufvesson, and T. L. Marzetta, "Massive MIMO for next generation wireless systems," *IEEE Communications Magazine*, vol. 52, no. 2, pp. 186–195, Feb. 2014.

[82] A. . van der Veen and A. Paulraj, "An analytical constant modulus algorithm," *IEEE Transactions on Signal Processing*, vol. 44, no. 5, pp. 1136–1155, May 1996.

[83] T. Li and N. D. Sidiropoulos, "Blind digital signal separation using successive interference cancellation iterative least squares," *IEEE Trans. Signal Process.*, vol. 48, no. 11, pp. 3146–3152, Nov. 2000.

[84] A. Goldsmith, S. A. Jafar, I. Maric, and S. Srinivasa, "Breaking spectrum gridlock with cognitive radios: An information theoretic perspective," *Proceedings of the IEEE*, vol. 97, no. 5, pp. 894–914, Apr. 2009.

[85] S. Haykin, "Cognitive radio: brain-empowered wireless communications," *IEEE Journal on Selected Areas in Communications*, vol. 23, no. 2, pp. 201–220, Feb. 2005.

[86] A. Ghasemi and E. S. Sousa, "Spectrum sensing in cognitive radio networks: requirements, challenges and design trade-offs," *IEEE Communications Magazine*, vol. 46, no. 4, pp. 32–39, Apr. 2008.

[87] T. S. Rappaport, Y. Xing, G. R. MacCartney, A. F. Molisch, E. Mellios, and J. Zhang, "Overview of millimeter wave communications for fifth-generation (5G) wireless networks—with a focus on propagation models," *IEEE Transactions on Antennas and Propagation*, vol. 65, no. 12, pp. 6213–6230, Aug. 2017.

[88] Q. Zhao and B. M. Sadler, "A survey of dynamic spectrum access," *IEEE Signal Processing Magazine*, vol. 24, no. 3, pp. 79–89, May 2007.

[89] J. M. Peha, "Approaches to spectrum sharing," *IEEE Communications Magazine*, vol. 43, no. 2, pp. 10–12, Feb. 2005.

[90] S. Geirhofer, L. Tong, and B. M. Sadler, "Cognitive radios for dynamic spectrum access-dynamic spectrum access in the time domain: Modeling and exploiting white space," *IEEE Communications Magazine*, vol. 45, no. 5, pp. 66–72, May 2007.

[91] Y. Zeng, Y.-C. Liang, A. T. Hoang, and R. Zhang, "A review on spectrum sensing for cognitive radio: Challenges and solutions," vol. 2010, Jan. 2010.

[92] T. Yucek and H. Arslan, "A survey of spectrum sensing algorithms for cognitive radio applications," *IEEE Communications Surveys Tutorials*, vol. 11, no. 1, pp. 116–130, Mar. 2009.

[93] L. M. Lopez-Ramos, Y. Teganya, B. Beferull-Lozano, and S. J. Kim, "Channel gain cartography via mixture of experts," in *IEEE Global Communications Conference (GLOBECOM)*, Taipei, Taiwan, Dec. 2020, pp. 1–7.

[94] A. G. Marques, L. M. Lopez-Ramos, G. B. Giannakis, and J. Ramos, "Resource allocation for interweave and underlay CRs under probability-of-interference constraints," *IEEE Journal on Selected Areas in Communications*, vol. 30, no. 10, pp. 1922–1933, Nov. 2012.

[95] N. Yi, Y. Ma, and R. Tafazolli, "Underlay cognitive radio with full or partial channel quality information." *International Journal of Navigation and Observation*, vol. 2010, July 2010.

[96] D. Denkovski, V. Rakovic, V. Atanasovski, L. Gavrilovska, and P. Mähönen, "Generic multiuser coordinated beamforming for underlay spectrum sharing," *IEEE Transactions on Communications*, vol. 64, no. 6, pp. 2285–2298, May 2016.

[97] F. A. Khan, K. Tourki, M. Alouini, and K. A. Qaraqe, "Performance analysis of a power limited spectrum sharing system with TAS/MRC," *IEEE Transactions on Signal Processing*, vol. 62, no. 4, pp. 954–967, Jan. 2014.

[98] M. Hanif, H. Yang, and M. Alouini, "Transmit antenna selection for power adaptive underlay cognitive radio with instantaneous interference constraint," *IEEE Transactions on Communications*, vol. 65, no. 6, pp. 2357–2367, Mar. 2017.

[99] R. Sarvendranath and N. B. Mehta, "Exploiting power adaptation with transmit antenna selection for interference-outage constrained underlay spectrum sharing," *IEEE Transactions on Communications*, vol. 68, no. 1, pp. 480–492, Oct. 2020.

[100] R. M. Rao, H. S. Dhillon, V. Marojevic, and J. H. Reed, "Analysis of worst-case interference in underlay radar-massive MIMO spectrum sharing scenarios," in *IEEE Global Communications Conference (GLOBECOM)*, HI, USA, Feb. 2019, pp. 1–6.

[101] S. Dadallage, C. Yi, and J. Cai, "Joint beamforming, power, and channel allocation in multiuser and multichannel underlay MISO cognitive radio networks," *IEEE Transactions on Vehicular Technology*, vol. 65, no. 5, pp. 3349–3359, May 2016.

[102] S. Kusaladharma and C. Tellambura, "Secondary user interference characterization for spatially random underlay networks with massive MIMO and power control," *IEEE Transactions on Vehicular Technology*, vol. 66, no. 9, pp. 7897–7912, Sep. 2017.

[103] V. Nguyen, L. Tran, T. Q. Duong, O. Shin, and R. Farrell, "An efficient precoder design for multiuser MIMO cognitive radio networks with interference constraints," *IEEE Transactions on Vehicular Technology*, vol. 66, no. 5, pp. 3991–4004, May 2017.

[104] S. Kashyap and N. B. Mehta, "SEP-Optimal transmit power policy for peak power and interference outage probability constrained underlay cognitive radios," *IEEE Transactions on Wireless Communications*, vol. 12, no. 12, pp. 6371–6381, Dec. 2013.

[105] A. Ghasemi and E. S. Sousa, "Fundamental limits of spectrum-sharing in fading environments," *IEEE Transactions on Wireless Communications*, vol. 6, no. 2, pp. 649–658, Feb. 2007.

[106] K. Zheng, X. Liu, X. Liu, and Y. Zhu, "Hybrid overlay-underlay cognitive radio networks with energy harvesting," *IEEE Transactions on Communications*, vol. 67, no. 7, pp. 4669–4682, Apr. 2019.

[107] P. K. Sangdeh, H. Pirayesh, A. Quadri, and H. Zeng, "A practical spectrum sharing scheme for cognitive radio networks: Design and experiments," *IEEE/ACM Transactions on Networking*, vol. 28, no. 4, pp. 1818–1831, Aug. 2020.

[108] D. Tse and P. Viswanath, *Fundamentals of Wireless Communication*. USA: Cambridge University Press, 2005.

[109] S. N. Affes, H. Hansen, and P. Mermelstein, "Interference subspace rejection: a framework for multiuser detection in wideband CDMA," *IEEE Journal on Selected Areas in Communications*, vol. 20, no. 2, pp. 287–302, Aug. 2002.

[110] U. Madhow, "Blind adaptive interference suppression for direct-sequence CDMA," *Proceedings of the IEEE*, vol. 86, no. 10, pp. 2049–2069, Oct. 1998.

Appendix A

Proofs and Technical Claims

A.1 Proof of Proposition 1

In order to see how CCA can identify cell-edge user signals in the noisy and inter-cell interference case, let us first rewrite the received signal at the ℓ-th BS as

$$\mathbf{Y}_\ell = \mathbf{H}_{\ell p_\ell}\mathbf{S}_{p_\ell}^T + \mathbf{H}_{\ell c}\mathbf{S}_c^T + \mathbf{H}_{\ell p_j}\mathbf{S}_{p_j}^T + \mathbf{W}_\ell \tag{A.1}$$

where $\ell, j \in \{1, 2\}$ and $j \neq \ell$. Recall that, from (3.1) and (3.2), one can easily see that the channel matrix $\mathbf{H}_{\ell p_\ell}$ in (3.1) can be factored into $\mathbf{Z}_{\ell p_\ell}\mathbf{P}^{1/2}$ where $\mathbf{Z}_{\ell p_\ell} \in \mathbb{C}^{M_\ell \times (K_\ell - K_{e_\ell})}$ holds in its columns the small-scale fading vectors defined in (3.1) while $\mathbf{P}_{\ell p_\ell} \in \mathbb{R}^{K_\ell - K_{e_\ell}}$ is a diagonal matrix whose entries model the received signal power (product of path loss and transmitted signal power) for each of the cell-center users served by the ℓ-th BS, and likewise for $\mathbf{H}_{\ell c}$ and $\mathbf{H}_{\ell p_j}$. Therefore, (A.1) can be equivalently written as

$$\mathbf{Y}_\ell = \mathbf{Z}_{\ell p_\ell}\mathbf{P}_{\ell p_\ell}^{1/2}\mathbf{S}_{p_\ell}^T + \mathbf{Z}_{\ell c}\mathbf{P}_{\ell p_c}^{1/2}\mathbf{S}_c^T + \mathbf{Z}_{\ell p_j}\mathbf{P}_{\ell p_j}^{1/2}\mathbf{S}_{p_j}^T + \mathbf{W}_\ell \tag{A.2}$$

Let us first consider a simple scenario with two cell-center users (one at each BS) and one cell-edge user located at the common edge between the two BSs. We define β_p, β_e and β_f as the received signal power (RSP) of the ℓ-th cell-center user at the ℓ-th BS, the RSP of the cell-edge user at the ℓ-th BS, and the RSP of the j-th cell-center user at the ℓ-th BS, respectively, for $\ell \neq j$. Furthermore, for the sake of simplicity, we assume here that the cell-edge user signal is received with equal power at both BSs, i.e., the cell-edge user is exactly on the edge between the two BSs. Then (A.2) can be expressed as

$$\mathbf{Y}_\ell = \mathbf{Z}_\ell \mathbf{P}_\ell^{1/2}\mathbf{B}^T + \mathbf{W}_\ell \tag{A.3}$$

where $\mathbf{B} = [\mathbf{s}_e, \mathbf{s}_{p_1}, \mathbf{s}_{p_2}]$ holds in its columns the temporal signals of the three users, $\mathbf{P}_1 = \text{Diag}([\beta_e, \beta_p, \beta_f])$, $\mathbf{P}_2 = \text{Diag}([\beta_e, \beta_f, \beta_p])$ and $\mathbf{B} = [\mathbf{s}_e, \mathbf{s}_{p_1}, \mathbf{s}_{p_2}]$, where $\mathbf{D} = \text{Diag}(\mathbf{d})$ is a diagonal matrix with the vector $\mathbf{d}$ on its diagonal. The entries of $\mathbf{Z}_\ell$ represent the small scale fading between each user and the antennas at BS ℓ. $\mathbf{Z}_\ell$ is modeled as i.i.d. circularly symmetric zero-mean Gaussian with variance $1/M_\ell$ (corresponding to a rich scattering scenario).

We will now compute the cross- and auto-correlation matrices $\mathbf{R}_{\mathbf{y}_1\mathbf{y}_2}$, $\mathbf{R}_{\mathbf{y}_\ell\mathbf{y}_\ell}$ as follows. Since the cross correlation matrix, $\mathbf{R}_{\mathbf{y}_1\mathbf{y}_2}$, is given by $\frac{1}{T}\mathbf{Y}_1\mathbf{Y}_2^H$, then it follows that $\mathbf{R}_{\mathbf{y}_1\mathbf{y}_2}$ is given by

$$\begin{aligned}\mathbf{R}_{\mathbf{y}_1\mathbf{y}_2} &= \frac{1}{T}(\mathbf{Z}_1\mathbf{P}_1^{1/2}\mathbf{B}^T + \mathbf{W}_1)(\mathbf{Z}_2\mathbf{P}_2^{1/2}\mathbf{B} + \mathbf{W}_2)^H \\ &= \mathbf{Z}_1\mathbf{P}_{12}\mathbf{H}_2^H\end{aligned} \tag{A.4}$$

where $\mathbf{P}_{12} = (\mathbf{P}_2\mathbf{P}_1)^{1/2}$. Note that, in (A.4), in addition to the assumption that $\frac{1}{T}\mathbf{B}^T\mathbf{B} = \mathbf{I}$, we exploited the fact that, for large T, $\frac{1}{T}\mathbf{W}_\ell\mathbf{W}_j^H \approx 0$ and $\frac{1}{T}\mathbf{B}^T\mathbf{W}_j^H \approx 0$, for $j, \ell \in \{1, 2\}$. Similarly, the auto-correlation matrix of the received signal of the ℓ-th BS can be expressed as

$$\mathbf{R}_{\mathbf{y}_\ell\mathbf{y}_\ell} = \mathbf{Z}_\ell\mathbf{P}_\ell\mathbf{Z}_\ell^H + \sigma^2\mathbf{I} \tag{A.5}$$

Now, we substitute with (A.4) and (A.5) in (2.9) to obtain

$$\begin{aligned}\mathbf{Z}_1\mathbf{P}_{12}\mathbf{Z}_2^H(\mathbf{Z}_2\mathbf{P}_2\mathbf{Z}_2^H + \sigma^2\mathbf{I})^{-1}\mathbf{Z}_2\mathbf{P}_{12}\mathbf{Z}_1^H\mathbf{q}_1 \\ = \lambda^2(\mathbf{Z}_1\mathbf{P}_1\mathbf{Z}_1^H + \sigma^2\mathbf{I})\mathbf{q}_1\end{aligned} \tag{A.6}$$

which can be equivalently written as

$$\begin{aligned}\mathbf{Z}_1\mathbf{\Gamma}_{12}\mathbf{Z}_2^H(\mathbf{Z}_2\mathbf{\Gamma}_2\mathbf{Z}_2^H + \mathbf{I})^{-1}\mathbf{Z}_2\mathbf{\Gamma}_{12}\mathbf{Z}_1^H\mathbf{q}_1 \\ = \lambda^2(\mathbf{Z}_1\mathbf{\Gamma}_1\mathbf{Z}_1^H + \mathbf{I})\mathbf{q}_1\end{aligned} \tag{A.7}$$

where $\mathbf{\Gamma}_1 = \text{Diag}([\gamma_e, \gamma_p, \gamma_f])$, $\mathbf{\Gamma}_2 = \text{Diag}([\gamma_e, \gamma_f, \gamma_p])$ and $\mathbf{\Gamma}_{12} = (\mathbf{\Gamma}_2\mathbf{\Gamma}_1)^{1/2}$, with $\gamma_e = \beta_e/\sigma^2$ be the received SNR of the cell-edge user, $\gamma_p = \beta_p/\sigma^2$ be the received SNR of each cell-center user at its serving BS, and $\gamma_f = \beta_f/\sigma^2$ be the received SNR of each cell-center at the other (non-serving) BS. By left multiplying the two sides of (A.7) by $\mathbf{H}_1^\dagger$, we obtain

$$\begin{aligned}\mathbf{\Gamma}_{12}\mathbf{Z}_2^H(\mathbf{Z}_2\mathbf{\Gamma}_2\mathbf{Z}_2^H + \mathbf{I})^{-1}\mathbf{Z}_2\mathbf{\Gamma}_{12}\mathbf{Z}_1^H\mathbf{q}_1 \\ = \lambda^2(\mathbf{\Gamma}_1\mathbf{Z}_1^H + \mathbf{Z}_1^\dagger)\mathbf{q}_1\end{aligned} \tag{A.8}$$

By substituting with $\mathbf{Z}_1^\dagger = (\mathbf{Z}_1^H\mathbf{Z}_1)^{-1}\mathbf{Z}_1^H$, and by letting $\mathbf{v} = \mathbf{Z}_1^H\mathbf{q}_1$, (A.8) can be expressed as

$$\mathbf{\Gamma}_{12}\mathbf{Z}_2^H(\mathbf{Z}_2\mathbf{\Gamma}_2\mathbf{Z}_2^H + \mathbf{I})^{-1}\mathbf{Z}_2\mathbf{\Gamma}_{12}\mathbf{v} = \lambda^2(\mathbf{\Gamma}_1 + (\mathbf{Z}_1^H\mathbf{Z}_1)^{-1})\mathbf{v} \tag{A.9}$$

By defining the matrix $\mathbf{Z} := \mathbf{Z}_2^H(\mathbf{Z}_2\mathbf{\Gamma}_2\mathbf{Z}_2^H + \mathbf{I})^{-1}\mathbf{Z}_2$, it then follows that $\mathbf{Z}$ can be simplified as

$$\mathbf{Z} = \mathbf{Z}_2^H(\mathbf{Z}_2\mathbf{\Gamma}_2\mathbf{Z}_2^H + \mathbf{I})^{-1}\mathbf{Z}_2 \tag{A.10a}$$

$$= \mathbf{Z}_2^H(\mathbf{Z}_2^\dagger(\mathbf{Z}_2\mathbf{\Gamma}_2\mathbf{Z}_2^H + \mathbf{I}))^\dagger \tag{A.10b}$$

$$= \mathbf{Z}_2^H(\mathbf{Z}_2^H)^\dagger(\mathbf{\Gamma}_2 + (\mathbf{Z}_2^H\mathbf{Z}_2)^{-1})^{-1} \tag{A.10c}$$

$$= (\mathbf{\Gamma}_2 + (\mathbf{Z}_2^H\mathbf{Z}_2)^{-1})^{-1} \tag{A.10d}$$

Note that in (A.10b) and (A.10c), we have exploited the following two properties of the pseudoinverse

P 1. For any square matrix $\mathbf{A}$, if $\mathbf{A}$ is invertible, its pseudoinverse is its inverse, i.e., $\mathbf{A}^\dagger = \mathbf{A}^{-1}$

P 2. $(\mathbf{BA})^\dagger = \mathbf{A}^\dagger\mathbf{B}^\dagger$

By substituting with (A.10d) in (A.9), we obtain

$$\mathbf{\Gamma}_{12}(\mathbf{\Gamma}_2 + (\mathbf{Z}_2^H\mathbf{Z}_2)^{-1})^{-1}\mathbf{\Gamma}_{12}\mathbf{v} = \lambda^2(\mathbf{\Gamma}_1 + (\mathbf{Z}_1^H\mathbf{Z}_1)^{-1})\mathbf{v} \tag{A.11}$$

which can be equivalently expressed as

$$\mathbf{F}\mathbf{v} = \lambda^2\mathbf{v} \tag{A.12}$$

where $\mathbf{F}$:=$(\mathbf{\Gamma}_1+(\mathbf{Z}_1^H\mathbf{Z}_1)^{-1})^{-1}\mathbf{\Gamma}_{12}(\mathbf{\Gamma}_2+(\mathbf{Z}_2^H\mathbf{Z}_2)^{-1})^{-1}\mathbf{\Gamma}_{12}$ is an $K_s \times K_s$ matrix, and $K_s = 3$ for the particular scenario considered here. For ease of exposition, we will assume here that the number of antennas M_ℓ is large enough so that $(\mathbf{Z}_\ell^H\mathbf{Z}_\ell)^{-1}$ is approximately identity. Thus, matrix $\mathbf{F}$ can be expressed as

$$\mathbf{F} := \begin{bmatrix} (\frac{\gamma_e}{\gamma_e+1})^2 & 0 & 0 \\ 0 & \frac{\gamma_f\gamma_c}{(\gamma_f+1)(\gamma_c+1)} & 0 \\ 0 & 0 & \frac{\gamma_f\gamma_c}{(\gamma_f+1)(\gamma_c+1)} \end{bmatrix} \in \mathbb{R}^{K_s\times K_s},$$

If each cell-center user is close to its serving BS, then $\gamma_f << 1$ and $\gamma_c >> 1$. Therefore, the term $\frac{\gamma_f\gamma_c}{(\gamma_f+1)(\gamma_c+1)}$ will be approximately equal to γ_f. Then, it can be easily seen that the maximum eigenvalue of the matrix $\mathbf{F}$ is equal to $(\frac{\gamma_e}{\gamma_e+1})^2$ and the other two eigenvalues will be approximately equal to γ_f. Since the maximum eigenvalue of the matrix $\mathbf{F}$ is nothing but the square of the correlation coefficient that is associated with the vectors $\mathbf{Y}_1^T\mathbf{q}_1$ and $\mathbf{Y}_2^T\mathbf{q}_2$. Then, it turns out that the maximum correlation coefficient is given by

$$\rho_{\max} = \frac{\gamma_e}{\gamma_e + 1} \tag{A.13}$$

Now, we need to compute the eigenvectors $\mathbf{q}_1$ and $\mathbf{q}_2$. Since the maximum eigenvector of the diagonal matrix $\mathbf{F}$ is given by

$$\mathbf{v} = [\pm 1, 0, 0]^T \tag{A.14}$$

the eigenvector $\mathbf{q}_1$ can be obtained by solving the following system of linear equations

$$\mathbf{v} = \mathbf{Z}_1^H \mathbf{q}_1 \tag{A.15}$$

Without loss of generality, we can let $\mathbf{q}_1^* = \mathbf{Z}_1(\mathbf{Z}_1^H\mathbf{Z}_1)^{-1}\mathbf{v}$. The reason is that we can always find two components to the vector $\mathbf{q}_1^*$; one in the subspace spanned by $\mathbf{Z}_1$ and one orthogonal to it, however, the latter will vanish after multiplication with $\mathbf{Z}_1^H$. By substituting with $\mathbf{q}_1^*$ in (2.8), it can be easily proved that the corresponding canonical component of the second view $\mathbf{q}_2^* = \mathbf{Z}_2(\mathbf{Z}_2^H\mathbf{Z}_2)^{-1}\mathbf{v}$. Define $\hat{\mathbf{s}}_{\ell c} := \mathbf{Y}_\ell^H \mathbf{q}_\ell^*$ and substitute with $\mathbf{q}_\ell^*$, we get the following

$$\hat{\mathbf{s}}_{\ell c} = \sqrt{\beta_e} c \mathbf{s}_c + \mathbf{n}_\ell \tag{A.16}$$

where $\mathbf{n}_\ell = \mathbf{W}_\ell^H \mathbf{q}_\ell^* \in \mathbb{C}^T$ and $c = \pm 1$. This means that, in the case of single cell-edge user, the proposed detector can efficiently recover cell-edge user signals at low SNR even in the presence of inter-cell interference.

The generalization to $K_e > 1$ and $K_\ell - K_e > 1$ now follows directly. In that case, the matrix F will have the vector $\mathbf{f} \in \mathbb{R}^{K_s}$ on its diagonal, where

$$\mathbf{f}(j) = \begin{cases} (\frac{\gamma_{e_j}}{\gamma_{e_j}+1})^2, & j \in \{1, \cdots, K_e\} \\ \frac{\gamma_{f_j}\gamma_{p_j}}{(\gamma_{f_j}+1)(\gamma_{p_j}+1)} & j \in \{K_e+1, \cdots, K_s\} \end{cases}$$

Assume that $\gamma_{f_j} << 1, \forall j \in \{K_e+1, \cdots, K_s\}$. Then it can be easily seen that the largest K_e eigen vectors are the first K_e columns of an $K_s \times K_s$ identity matrix. Upon letting $\mathbf{V} = \mathbf{I}(:, 1:K_e)$, the optimal solution $\mathbf{Q}_\ell^* = \mathbf{H}_\ell(\mathbf{Z}_\ell^H\mathbf{Z}_\ell)^{-1}\mathbf{V}\mathbf{M}_\ell$, where $\mathbf{M}_\ell$ is any $K_e \times K_e$ non singular matrix that satisfies the ℓ-th orthonormality constraint in (2.13). Define $\hat{\mathbf{S}}_{\ell c} := \mathbf{Y}_\ell^H \mathbf{Q}_\ell^*$ and substitute in (A.2), we obtain

$$\hat{\mathbf{S}}_{\ell c} = \mathbf{S}_c \mathbf{P}_c^{1/2} \mathbf{M}_\ell + \mathbf{N}_\ell \tag{A.17}$$

where $\mathbf{P}_c = \text{Diag}([\beta_{e_1}, \cdots, \beta_{e_{K_e}}])$, and $\mathbf{N}_\ell = \mathbf{W}_\ell^H \mathbf{Q}_\ell^*$. Note that, after obtaining $\hat{\mathbf{S}}_{\ell c}$, we pass it to RACMA to identify the cell-edge user signals $\mathbf{S}_c$.

A.2 Proof of Proposition 2

In this section, we will show that the principal K_c eigenvectors of the matrix $\mathbf{A} = \sum_{\ell=1}^{L} \mathbf{Y}_\ell^H(\mathbf{Y}_\ell\mathbf{Y}_\ell^H)^{-1}\mathbf{Y}_\ell$ is approximately the column space of the cell-edge user signals. We first write the auto-correlation matrix $\mathbf{Y}_\ell\mathbf{Y}_\ell^H$ as

$$\mathbf{Y}_\ell\mathbf{Y}_\ell^H = \mathbf{Z}_\ell\mathbf{P}_\ell\mathbf{Z}_\ell^H + \sigma^2\mathbf{I}, \tag{A.18}$$

where we have exploited the facts that, at $N > K_s$, $\mathbf{X}^T\mathbf{X}/N \approx \mathbf{I}_{K_s}$ and $\mathbb{E}[\mathbf{N}_\ell\mathbf{N}_\ell^H] = \sigma^2\mathbf{I}_{N_\ell}$. Define the diagonal matrix $\mathbf{\Gamma}_\ell := \mathbf{P}_\ell/\sigma^2 \in \mathbb{R}^{K_s \times K_s}$ that contains the received SNR of each user

at the ℓ-th BS, and $\mathbf{U}_\ell := \mathbf{N}_\ell^H/\sigma \in \mathbb{C}^{N\times M_\ell}$, and the matrix $\mathbf{A}_\ell := \mathbf{Y}_\ell^H(\mathbf{Y}_\ell\mathbf{Y}_\ell^H)^{-1}\mathbf{Y}_\ell$. Then, by direct substitution of (4.8) and (A.18) in the matrix $\mathbf{A}_\ell$, we obtain

$$\begin{aligned}\mathbf{A}_\ell = \mathbf{X}\boldsymbol{\Gamma}_\ell^{1/2}\mathbf{Z}_\ell^H(\mathbf{Z}_\ell\boldsymbol{\Gamma}_\ell\mathbf{Z}_\ell^H+\mathbf{I})^{-1}\mathbf{Z}_\ell\boldsymbol{\Gamma}_\ell^{1/2}\mathbf{X}^T \\ + \mathbf{U}_\ell(\mathbf{Z}_\ell\boldsymbol{\Gamma}_\ell\mathbf{Z}_\ell^H+\mathbf{I})^{-1}\mathbf{U}_\ell^H + \boldsymbol{\delta}_\ell + \boldsymbol{\delta}_\ell^H,\end{aligned} \tag{A.19}$$

where $\boldsymbol{\delta}_\ell := \mathbf{U}_\ell(\mathbf{Z}_\ell\boldsymbol{\Gamma}_\ell\mathbf{Z}_\ell^H+\mathbf{I})^{-1}\mathbf{Z}_\ell\boldsymbol{\Gamma}_\ell^{1/2}\mathbf{X}^T$. By applying the Woodbury matrix identity on the matrix $\mathbf{C}_\ell := (\mathbf{Z}_\ell\boldsymbol{\Gamma}_\ell\mathbf{Z}_\ell^H+\mathbf{I})^{-1}$, we get

$$\mathbf{C}_\ell = (\mathbf{Z}_\ell\boldsymbol{\Gamma}_\ell\mathbf{Z}_\ell^H+\mathbf{I})^{-1} \tag{A.20a}$$
$$= \mathbf{I} - \mathbf{Z}_\ell({\boldsymbol{\Gamma}_\ell}^{-1}+\mathbf{Z}_\ell^H\mathbf{Z}_\ell)^{-1}\mathbf{Z}_\ell^H \tag{A.20b}$$
$$\approxeq \mathbf{I} - \mathbf{Z}_\ell({\boldsymbol{\Gamma}_\ell}^{-1}+\mathbf{I})^{-1}\mathbf{Z}_\ell^H \tag{A.20c}$$
$$= \mathbf{I} - \mathbf{Z}_\ell\mathbf{D}_\ell\mathbf{Z}_\ell^H, \tag{A.20d}$$

where $\mathbf{D}_\ell \approxeq \boldsymbol{\Gamma}_\ell(\boldsymbol{\Gamma}_\ell+\mathbf{I})^{-1}$. It now follows that the first term in (A.19) can be expressed as

$$\mathbf{T}_\ell^{(1)} = \mathbf{X}\boldsymbol{\Gamma}_\ell^{1/2}\mathbf{Z}_\ell^H\mathbf{C}_\ell\mathbf{Z}_\ell\boldsymbol{\Gamma}_\ell^{1/2}\mathbf{X}^T \tag{A.21a}$$
$$\approxeq \mathbf{X}\boldsymbol{\Gamma}_\ell^{1/2}(\mathbf{I}-\mathbf{D}_\ell)\boldsymbol{\Gamma}_\ell^{1/2}\mathbf{X}^T \tag{A.21b}$$
$$= \mathbf{X}\mathbf{D}_\ell\mathbf{X}^T, \tag{A.21c}$$

On the other hand, the second term in (A.19) can be written as

$$\mathbf{T}_\ell^{(2)} = \mathbf{U}_\ell\mathbf{C}_\ell\mathbf{U}_\ell^H \tag{A.22a}$$
$$= \mathbf{U}_\ell\mathbf{U}_\ell^H - \mathbf{U}_\ell\mathbf{Z}_\ell\mathbf{D}_\ell\mathbf{Z}_\ell^H\mathbf{U}_\ell^H, \tag{A.22b}$$

Given that $\mathbf{U}_\ell$ contains i.i.d entries with zero mean and variance $1/N$ while $\mathbf{Z}_\ell$ contains i.i.d entries with zero mean and variance $1/M_\ell$, both $\mathbf{Z}_\ell$ and $\mathbf{U}_\ell$ are uncorrelated, and $\mathbf{D}_\ell \preceq \mathbf{I}_{K_s}$ ($\preceq$ interpreted element-wise), then it follows that the summation in (A.22b) will be dominated by the matrix $\mathbf{U}_\ell\mathbf{U}_\ell^H$. Therefore, $\mathbf{T}_\ell^{(2)}$ can be approximately written as

$$\mathbf{T}_\ell^{(2)} \approxeq \mathbf{U}_\ell\mathbf{U}_\ell^H, \tag{A.23}$$

Then, the expression $\boldsymbol{\delta}_\ell$ can written as

$$\boldsymbol{\delta}_\ell = \mathbf{U}_\ell\mathbf{C}_\ell\mathbf{Z}_\ell\boldsymbol{\Gamma}_\ell^{1/2}\mathbf{X}^T \tag{A.24a}$$
$$\approxeq \mathbf{U}_\ell\mathbf{Z}_\ell(\mathbf{I}-\mathbf{D}_\ell)\boldsymbol{\Gamma}_\ell^{1/2}\mathbf{X}^T \tag{A.24b}$$
$$= \mathbf{U}_\ell\mathbf{Z}_\ell(\boldsymbol{\Gamma}_\ell+\mathbf{I})^{-1}\boldsymbol{\Gamma}_\ell^{1/2}\mathbf{X}^T, \tag{A.24c}$$

By summing (A.23) and (A.24c), we get

$$\mathbf{T}_\ell^{(2)} + \boldsymbol{\delta}_\ell = \mathbf{U}_\ell(\mathbf{U}_\ell^H + \mathbf{Z}_\ell(\boldsymbol{\Gamma}_\ell+\mathbf{I})^{-1}\boldsymbol{\Gamma}_\ell^{1/2}\mathbf{X}^T), \tag{A.25}$$

where the summation on the right hand side of (A.25) is nothing but adding two Gaussian matrices; one with variance $\sigma_1^2 = 1/N$ and the other with variance $\sigma_2^2 = \frac{1}{N}\frac{1}{M_\ell}\sum_{i=1}^{K_s}\frac{\sqrt{\gamma_{i\ell}}}{(\gamma_{i\ell}+1)^2}$, where it can be easily seen that, even for modest M_ℓ, $\sigma_2^2 << \sigma_1^2$. Therefore, the summation in (A.25) will be dominated by $\mathbf{T}_\ell^{(2)}$. Thus, combining (A.21) with (A.23), (A.19) can be written as

$$\mathbf{A}_\ell \approxeq \mathbf{X}\mathbf{D}_\ell\mathbf{X}^T + \mathbf{U}_\ell\mathbf{U}_\ell^H, \tag{A.26}$$

Recall that the optimal solution $\mathbf{G}^*$ of (4.6) is the K_c principal eigenvectors of the following matrix

$$\mathbf{A} = \sum_{\ell=1}^{L}\mathbf{A}_\ell \tag{A.27a}$$

$$= \mathbf{X}\mathbf{D}\mathbf{X}^T + \sum_{\ell=1}^{L}\mathbf{U}_\ell\mathbf{U}_\ell^H, \tag{A.27b}$$

where $\mathbf{D} := \sum_{\ell=1}^{L}\mathbf{D}_\ell \in \mathbb{R}^{K_s\times K_s}$. By defining $\mathbf{V} = [\mathbf{X}, \mathbf{U}_1, \cdots, \mathbf{U}_L] \in \mathbb{C}^{N\times(Ks+\sum_{\ell=1}^{L}M_\ell)}$ and $\mathbf{\Sigma} := \text{Diag}(\mathbf{D}, \mathbf{I}_{M_1}, \cdots, \mathbf{I}_{M_L})$, (A.27) can be equivalently expressed as

$$\mathbf{A} = \mathbf{V}\mathbf{\Sigma}\mathbf{V}^H, \tag{A.28}$$

Since $\mathbf{X}^T\mathbf{X} \approx \mathbf{I}_{K_s}$ and by definition $\mathbf{U}_\ell^H\mathbf{U}_\ell \approx \mathbf{I}_{M_\ell}\forall\ell$, then $\mathbf{V}^H\mathbf{V} \approx \mathbf{I}_{K_s+M_s}$, it can be readily seen that the right hand side of (A.28) is nothing but the eigendecomposition of the matrix $\mathbf{A}$. Recall that the i-th diagonal entry of the matrix $\mathbf{D}$ is given by

$$\mathbf{D}_{(i,i)} = \sum_{\ell=1}^{L} r_{i\ell}, \tag{A.29}$$

From (AS2) and by assuming that the received signal power of the k-th cell-edge user at the ℓ-th BS is few dBs above the noise floor, i.e., $r_{k\ell} > 0.5\ \forall k = 1, \cdots, K_c$ and $\ell = 1, 2, 3$, the eigenspace of the K_c principal components of the matrix $\mathbf{A}$ is given by

$$\mathbf{G}^* = \mathbf{X}_c\mathbf{P}, \tag{A.30}$$

where $\mathbf{P}$ is any $K_c \times K_c$ non-singular matrix.

Appendix B

Acronyms

This appendix contains a table of acronyms and their meaning.

Table B.1: Acronyms

Acronym	Meaning
CCA	Canonical Correlation Analysis
GCCA	Generalized Canonical Correlation Analysis
5G	Fifth Generation
LTE	Long-Term Evolution
MIMO	Multiple-Input Multiple-Output
UE	User Equipment
SDR	Semidefinite Relaxation
BER	Bit Error Rate
SER	Symbol Error Rate
MLD	Maximum Likelihood Detector
ZF	Zero Forcing
MMSE	Minimum Mean Squared Error
SIC	Successive Interference Cancellation
IoT	Internet of Things
DSA	Dynamic Spectrum Access
SNR	Signal to Noise Ratio
SINR	Signal to Interference plus Noise Ratio