April 2023

On the Performance Enhancement of Beamspace MIMO and Non-orthogonal Multiple Access for Future Cellular Networks

Sinasi Cetinkaya
University of South Florida

On the Performance Enhancement of Beamspace MIMO and Non-orthogonal Multiple Access for Future Cellular Networks

by

Sinasi Cetinkaya

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
Department of Electrical Engineering
College of Engineering
University of South Florida

Co-Major Professor: Hüseyin Arslan, Ph.D.
Co-Major Professor: Gokhan Mumcu, Ph.D.
Ismail Uysal, Ph.D.
Nasir Ghani, Ph.D.
Alessio Gaspar, Ph.D.
Ali Imran, Ph.D.

Date of Approval:
April 12, 2023

Keywords: Millimeter Wave, Beam Selection, Precoding, Lens Antenna Subarrays

Dedication

To my beloved wife, Büşra, for her endless support, patience, and love; to my renunciant father & enduring mother & loving siblings, for everything; to my wonderful children, Eylül Asel & Ali Aras, for letting me experience the love that people can freely die for.

Acknowledgments

All journeys eventually end. The most challenging but rewarding journey that I have experienced was my Ph.D. Achieving success at the end of this journey requires health, patience, endurance, ceaseless effort, and self-belief. Thus, on this first page of my dissertation, I would like to present my sincere praises to Allah the Almighty for giving me all I need.

I would like to express my special thanks and gratitude to the Republic of Türkiye Ministry of National Education for honoring me with a prestigious scholarship (1416, YLSY) and giving me the opportunity to pursue my graduate studies abroad. None of this would have been possible without their financial support. I also would like to thank the Turkish Consulate General in New York, Office of the Education Attache, for always helping me with any problem I faced throughout this journey.

Behind every great man, there is a great woman. Therefore, I would like to give all compliments to my beloved and wonderful wife, Büşra Can Çetinkaya, for keeping me strong with her endless support, effort, patience, and sacrifice. This dissertation would never have been completed without her. The greatest gifts that this life has given me are my children, Eylül Asel Çetinkaya and Ali Aras Çetinkaya. I also thank my wife for taking care of them while I was very busy with my research. Thus, she always deserves appreciation from my dear children and me.

My beautiful mother, Döndü Çetinkaya, and my father, Hurşit Çetinkaya, you will always have my respect and appreciation to you. Knowing that I have your support and unconditional love made this possible. Thank you for always being my compass in this journey and showing me the right path to achieve my goals. I also would like to thank my brothers, Kadir Çetinkaya and Barış Çetinkaya, and my lovely sister, Elife Arslan, for always being with me in any condition and supporting me mentally despite the distance between us.

There is always someone who touches your life. Hence, I would like to present my warm greetings to all my friends who shared their experiences and taught me. My deepest appreciations go to Dr. Hasan Farooq, and Dr. Ahmad Asghar. They always encouraged me, shared their priceless experience in the field, and opened the path for my journey. Thus, they have an inevitable role in the completion of this dissertation. I would also like to thank my friends from the Artificial Intelligence for Networks (AI4Networks) group, Dr. Umair Sajid Hashmi, Dr. Azar Taufique, and Arsalan Darbandi. I can never thank Dr. Cüneyt Tamer, Abdulhamit Yıldırım, Bilal Ekdi, and Bahri Türk enough for their friendship and emotional support. During this journey, I understood that brotherhood is not always blood. Thence, special greetings go to Ahmet Topçuoğlu, Ömer Faruk Fırat, and Dr. İsmail Uluer for being my brothers. I always survived the hardship thanks to them and cannot repay their support.

I cannot forget my colleagues in the Wireless Communication and Signal Processing (WCSP) group. It was an honor to be in this research group. Since the day I joined the team, I have always felt the support of my friends Dr. Ali Fatih Demir, Dr. Berker Pekoz, Dr. Mohammed Hafiz, and Dr. Murat Karabacak. I am always grateful to earn experience with them. Their background and success inspired me to enhance my knowledge in the field. I also owe special thanks to Dr. Cenk Albayrak, Mehmet Mert Şahin, and Mehmet Yazgan for their valuable brotherhood and academic support.

I would like to express my deep gratitude to my advisor, Dr. Hüseyin Arslan, for his patient guidance, advice, time, and enthusiastic encouragement. This dissertation would not be possible without him. Special thanks to Dr. Nasir Ghani and Dr. Ismail Uysal for their emotional support. Thanks for the valuable time you spared to listen to me whenever I needed. I would like to thank Dr. Alessio Gaspar and Dr. Gokhan Mumcu for being on my committee and proving valuable feedback. Special gratitudes go to Dr. Ali Imran, my previous advisor, for letting me list him as one of my committee members and for never stopping his guidance.

Table of Contents

List of Tables

List of Figures

Abstract

The ever-growing demand for higher data rates and greater data capacity at lower cost has led the mobile cellular industry to investigate new physical layer techniques and possible utilization of unused spectrums at higher frequencies for next-generation cellular networks. Thus, exploitation of the millimeter-wave (mmWave) spectrum and non-orthogonal multiple-access (NOMA) have been envisioned as the most promising enablers in meeting capacity demand. Due to the smaller wavelengths offered in mmWave frequencies, it is possible to deploy many antennas into a relatively smaller physical space in mmWave frequencies. This property leads to a promising integration between mmWave and massive multiple-input multiple-output (M-MIMO) architecture to surmount the severe free-space pathloss thanks to the high directional beamforming gain. MmWave M-MIMO also offers significantly improved spectral efficiency by allowing simultaneous transmission of multiple data streams and utilizing the abundant and large bandwidth. However, conventional digital precoding causes excessive power consumption and hardware cost when directly adopted for mmWave M-MIMO since each antenna element necessitates its own radio-frequency (RF) chain. This problem can be addressed by beamspace MIMO (B-MIMO) because it can reduce the required RF chain by taking advantage of inherent sparsity in mmWave channels and applying a proper beam selection. On the other hand, NOMA enhances spectral efficiency by multiplexing multiple users' signals in the power domain using the same time and frequency resources, where the detection of multiple users' signals is performed by successive interference cancellation (SIC).

The research has concentrated on the beam selection problem, precoding design in B-MIMO, and spectral/energy efficiency enhancement in mmWave M-MIMO and NOMA. Specifically, the dissertation addresses the following:

First, we investigate the complexity reduction of the existing beam selection algorithms with incremental QR precoder (I-QR-P) and decremental QR precoder (D-QR-P). The proposed two-stage and three-stage algorithms reduce the complexity of D-QR-P and I-QR-P, respectively. Both aim to lower complexity by decreasing the candidate beam size by eliminating the beams with no contribution to any user and using matrix perturbation theory to update QR decompositions.

Second, we propose a hybrid precoding algorithm for the lens antenna subarray (LAS)-MIMO architecture in mmWave to control the LAS design efficiently. The precoding problem is formulated as a sparse reconstruction problem due to the inherent sparsity of the mmWave channel. The proposed algorithm is an iterative process developed jointly using artificial bee colony (ABC) optimization with orthogonal matching pursuit (OMP) algorithms. In each iteration, the algorithm randomly selects the switches for each lens using ABC and then uses OMP to approximate optimal unconstrained precoders.

Third, we investigate the spectral efficiency and energy efficiency tradeoff in downlink NOMA with the consideration of the quality of service (QoS) requirements. The non-convex multi-objective optimization problems are solved using population-based multi-objective evolutionary algorithms (MOEAs).

Finally, we propose an algorithm for the user-cell association problem in M-MIMO ultra-dense heterogeneous networks (UDHNs)., where the spectral and energy efficiency tradeoff is addressed. To this end, we formulate a convex multi-objective optimization problem and convert it into a single objective optimization problem where a priority is assigned for the spectral efficiency and energy efficiency with a weighting factor. The problem aims to maximize the weighted sum of spectral efficiency and energy efficiency. As a solution, Lagrange duality analysis is performed, and a distributed game theoretical user-cell association (GTUCA) algorithm is developed, considering the fairness among users.

Chapter 1: Introduction

Despite millimeter-wave (mmWave) spectrum utilization and massive multiple input multiple output (M-MIMO) being the key technologies to meet the ever-growing capacity demands for the next-generation cellular networks, some signal processing challenges arise due to the implementation of large-scale antenna arrays and mmWave channel characteristics such as detection algorithms requiring large matrix inversion, training overhead for channel estimation, and increased hardware cost and power consumption to fully achieve the spatial multiplexing with precoding. However, M-MIMO can also provide some level of simplification in signal processing due to the asymptotic orthogonality obtained as the number of antenna elements increases [1].

MmWave frequencies suffer from high atmospheric absorption, severe free-space pathloss at the first meter of the propagation, and poor penetration through water and concrete walls [2, 3, 4]. However, the smaller wavelength of mmWave frequencies makes accommodating many antenna elements in physically smaller areas possible. Hence, this characteristic enables a fruitful integration between M-MIMO and mmWave spectrum, compensating for the severe free-space pathloss thanks to the high directional beamforming gain [2]. MmWave M-MIMO improves spectral efficiency by introducing high spatial degrees of freedom (DoF) [1], allowing multiple data streams [5], and exploiting the larger bandwidth [6].

Unfortunately, it is not practical to implement mmWave M-MIMO entirely digitally since each antenna element requires one dedicated radio-frequency (RF) chain leading to prohibitively high power consumption and unaffordable hardware cost, even though it ensures full capacity and flexibility. On the other hand, severe performance loss is obtained when mmWave M-MIMO deploys a fully analog beamforming structure due to the constant

amplitude constraints led by the phase shifters with a few RF chains during the realization of the analog beamforming [7]. It is also not practical to be extended to multi-user communication systems [8]. In order to overcome the issues in fully analog and digital structures, hybrid precoding is considered a feasible approach [9, 10]. The hybrid precoding consists of a small-size digital precoder with a few RF chains to eliminate the interference and a large-size analog beamformer having many phase shifters to enhance the antenna array gain. Since each RF chain is connected to all the antenna elements through a phase shifter, hybrid precoding reduces the required RF chains without significant performance degradation. MmWave M-MIMO with hybrid precoding suffers also from high signal processing complexity and power consumption when considerably many phase shifters are used. Therefore, beamspace MIMO (B-MIMO) proposed in [11] reduces the RF chain requirement by exploiting the inherent sparsity in the mmWave channels. In B-MIMO, the angular domain (i.e., beamspace) representation of the spatial channel is performed by employing a discrete lens array (DLA) at the base station to explore the channel sparsity [11]. Hence, a reduction in radio-frequency (RF) chains required is achieved without compromising the system performance by performing beam selection [6, 11].

1.1 Overview of Precoding Techniques to Enable mmWave Communication

1.1.1 Digital Precoding

Digital precoding is a powerful technique widely applied in conventional MIMO systems operating at low frequencies to enhance transmission quality and provide a more reliable connection where each antenna element is connected to an RF chain leading to high flexibility to control the phase and amplitude of the transmitted signal [8]. Realizing beamforming with digital precoders also brings additional advantages, such as interference cancelation and high spatial multiplexing gain. It is also possible for both single-user and multi-user scenarios to realize the digital precoding by linear or nonlinear signal processing techniques.

1.1.1.1 *Single-user Digital Precoding*

Let Fig. 1.1 represent a single-user M-MIMO architecture where an N_r-antenna user is served by a base station deploying N_t antennas through simultaneously transmitted N_s data streams. Since each antenna requires its own dedicated RF chain in digital precoding, the number of transmit RF chains should be $N_{RF} = N_t$, where $N_s \leq N_{RF} \leq N_t$. Accordingly, the received signal vector $\mathbf{y} \in \mathbb{C}^{N_r \times 1}$ is expressed as

$$\mathbf{y} = \sqrt{\rho}\mathbf{H}^H\mathbf{D}\mathbf{s} + \mathbf{n}, \tag{1.1}$$

where ρ is the average received signal power, $\mathbf{H} \in \mathbb{C}^{N_r \times N_t}$ is the channel matrix, and $\mathbf{D} \in \mathbb{C}^{N_t \times N_r}$ is the digital precoder matrix where the total transmit power constraint is normalized such that $||\mathbf{D}||_F^2 = \text{tr}(\mathbf{D}^H\mathbf{D}) = N_s$. The transmitted data $\mathbf{s} \in \mathbb{C}^{N_r \times 1}$ has the normalized power of $\mathbb{E}[\mathbf{s^H s}] = \frac{1}{N_r}\mathbf{I}_{N_r}$ [1, 8]. Additionally, $\mathbf{n} \sim \mathcal{CN}(0, \sigma^2)$ stands for the additive white Gaussian noise (AWGN) with zero mean and variance σ^2.

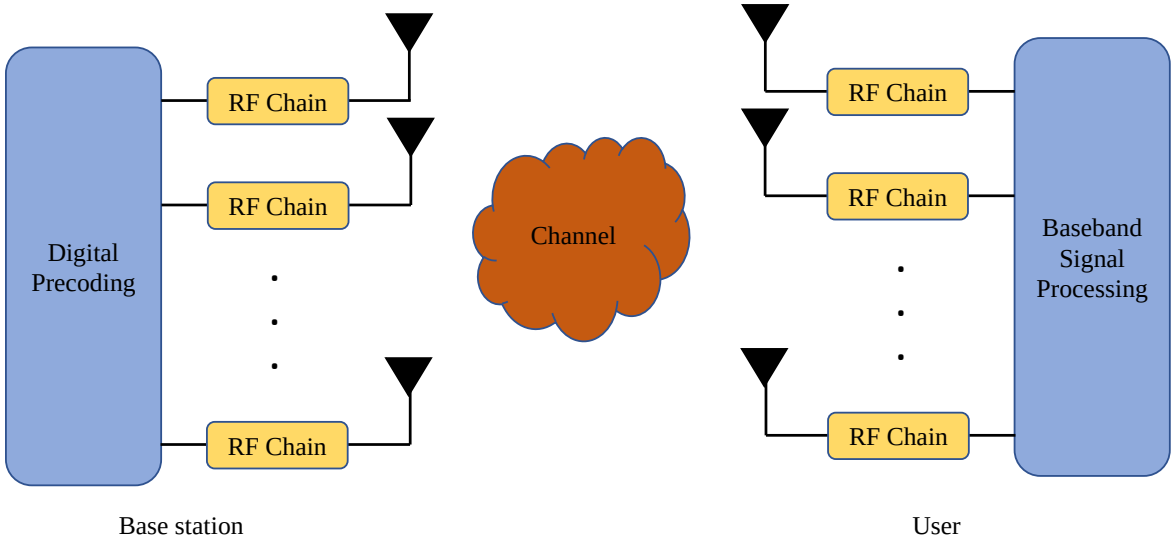


Figure 1.1: Single-user M-MIMO architecture with digital precoding

- Matches filter (MF) precoding [12]: It is a precoding technique used to enhance the SNR on the user side. However, it also carries a significant downside; the transmission

of multiple data streams may result in severe interference among the different data streams. The digital MF precoder can be obtained by $\mathbf{D} = \sqrt{\frac{N_r}{\mathbf{tr}(\mathbf{F}\mathbf{F}^H)}}\mathbf{F}$, where $\mathbf{F} = \mathbf{H}^H$ [13].

- Zero-forcing (ZF) precoding [12]: This precoding technique aims to maximize the received signal power at the receiver by properly designing the precoding matrix to cancel out the interference entirely as long as the channel is noise-free. Mathematically, the digital precoding matrix can be obtained by the inverse of the channel matrix such that $\mathbf{D} = \sqrt{\frac{N_r}{\mathbf{tr}(\mathbf{F}\mathbf{F}^H)}}\mathbf{F}$, where $\mathbf{F} = \mathbf{H}^H(\mathbf{H}\mathbf{H}^H)^{-1}$. However, any noise present negatively affects the system performance when ZF is used to precode the transmitted signal since it also amplifies the noise power.

- Regularized ZF (RZF) precoding [14, 15]: The effect of noise on ZF precoding can be compensated by RZF precoding technique, which utilizes a regularization term to the inverse of the channel matrix to minimize the noise amplification effect. Therefore, the RZF precoder can mathematically be computed as $\mathbf{D} = \sqrt{\frac{N_r}{\mathbf{tr}(\mathbf{F}\mathbf{F}^H)}}\mathbf{F}$, where $\mathbf{F} = \mathbf{H}^H(\mathbf{H}\mathbf{H}^H + \frac{\sigma_n^2 N_t}{\rho}\mathbf{I}_{N_r})^{-1}$.

1.1.1.2 Multi-user Digital Precoding

Let Fig. 1.2 represent a multi-user M-MIMO architecture where an K N_r-antenna users are simultaneously served by a base station with N_t antennas. For this scenario, the received signal vector $\mathbf{y_k}$ of the k-th user is expressed as [8]

$$\mathbf{y_k} = \mathbf{H}_k \sum_{n=1}^{K} \mathbf{D}_n \mathbf{s}_n + \mathbf{n}_k, \tag{1.2}$$

where $\mathbf{s}_n \in \mathbb{C}^{N_r \times 1}$ is the original signal with normalized power, $\mathbf{D} = [\mathbf{D}_1, \mathbf{D}_2, \ldots, \mathbf{D}_K] \in \mathbb{C}^{N_t \times K N_r}$ stands for the entire digital precoder matrix and $\mathbf{D}_k \in \mathbb{C}^{N_t \times N_r}$ is the digital precoder for the k-th user, where the total transmit power constraint is normalized such that $||\mathbf{D_k}||_F =$

N_r [8]. Additionally, $\mathbf{n}_k \sim \mathcal{CN}(0, \sigma^2)$ stands for the additive white Gaussian noise (AWGN) with zero mean and variance σ^2.

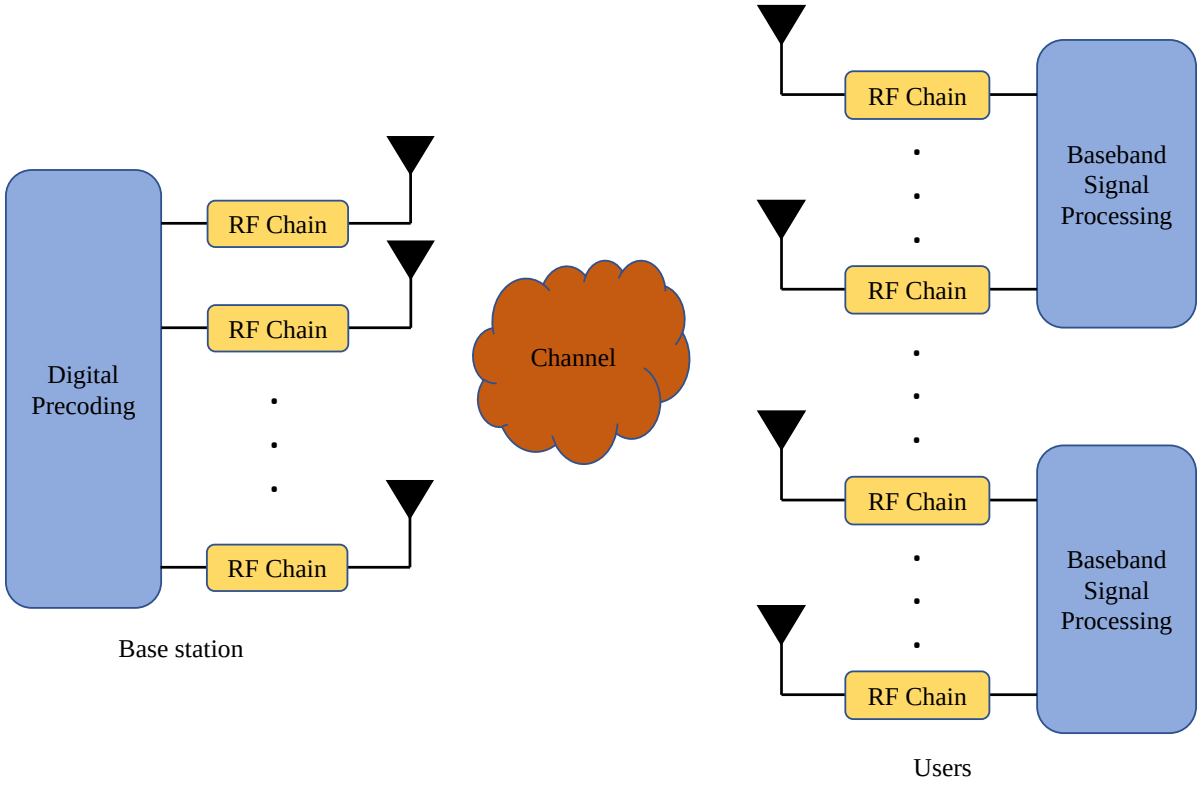


Figure 1.2: Multi-user M-MIMO architecture with digital precoding

- Block Diagonalization (BD) precoding [16]: It is obvious from (1.2) that multi-user M-MIMO communication is vulnerable to multi-user interference. Since $\mathbf{H}_k\mathbf{D}_n\mathbf{s}_n$ for $n \neq k$ refers to the interferences to the k-th user, BD precoding cancels out the multi-user interference by forcing $\mathbf{D}_n\mathbf{s}_n = 0$. The main idea of BD precoding is to convert the MIMO channel into parallel subchannels by applying singular value decomposition (SVD), each of which represents the channel response between a user and the base station. Let the SVD of the channel matrix of the k-th user be $\hat{\mathbf{H}}_k = \hat{\mathbf{U}}_k\hat{\mathbf{\Sigma}}_k\hat{\mathbf{V}}_k^H = \hat{\mathbf{U}}_k\hat{\mathbf{\Sigma}}_k[\hat{\mathbf{V}}_k^1\ \hat{\mathbf{V}}_k^2]^H$, where $\hat{\mathbf{V}}_k^1$ and $\hat{\mathbf{V}}_k^2$ denotes the semi-unitary vectors containing nonzero singular values and zero singular values of $\hat{\mathbf{H}}_k$, respectively. Hence, the MIMO channel

can be converted into parallel subchannels when the digital precoder $\mathbf{D}_n = \hat{\mathbf{V}}_k^1(:, 1 : N_r)$ [17], eliminating the multi-user interference.

In conclusion, mmWave M-MIMO employs a large number of antenna arrays, and each antenna requires a dedicated RF chain for connection. This makes it impractical for the next generation wireless networks due to its excessive hardware cost, signal processing complexity, and power consumption.

1.1.2 Analog Beamforming

Analog beamforming enables point-to-point communication for mmWave M-MIMO systems. The typical implementation of analog beamforming includes a single RF chain and many analog phase shifters to transmit a single data stream. Thence, exploiting the phase shifters enhances the antenna array gain and SNR since the technique adjusts the phase of transmitted signals [1, 8].

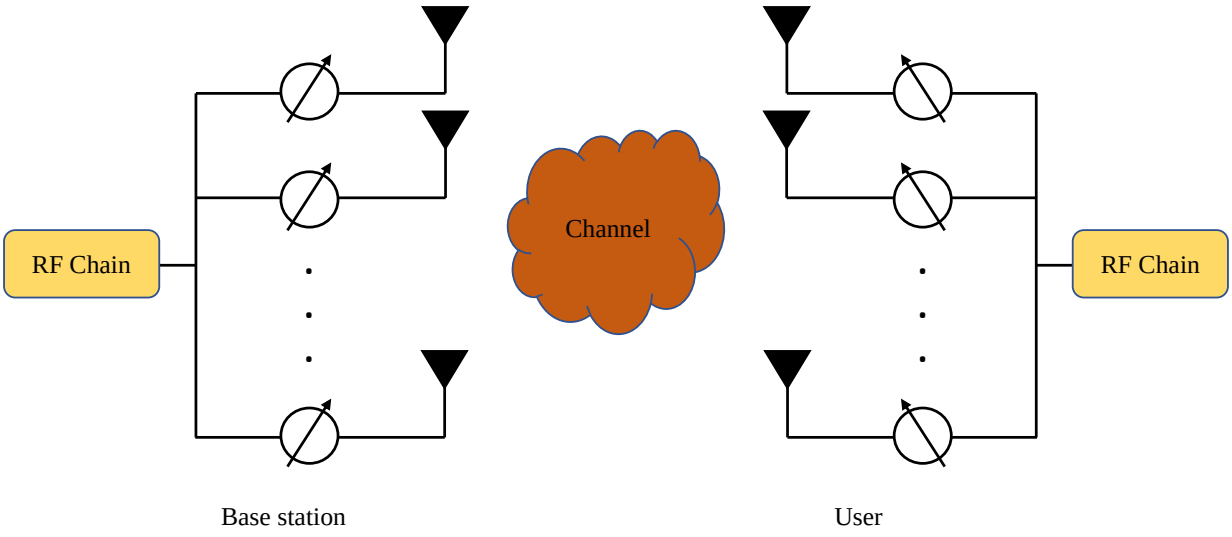


Figure 1.3: Single-user M-MIMO architecture with analog beamforming

Fig. 1.3 represents a typical massive MIMO architecture with analog beamforming. Since the aim is to maximize the SNR, the analog beamforming can be designed for a massive MIMO system, where a base station deploying N_t antennas communicates with an

N_r-antenna user via a single data stream, as solving the optimization problem given as [8]

$$(\mathbf{f}^*, \mathbf{w}^*) = \text{argmax}\, |\mathbf{w}^H \mathbf{H} \mathbf{f}|^2, \quad \text{s.t.,} \quad \mathbf{f}(i) = \sqrt{\frac{1}{N_t}} e^{j\psi_i}, \quad \mathbf{w}(i) = \sqrt{\frac{1}{N_r}} e^{j\phi_i}, \tag{1.3}$$

where $\mathbf{f} \in \mathbb{C}^{N_t \times 1}$ and $\mathbf{w} \in \mathbb{C}^{N_r \times 1}$ stand for beamforming vectors for the base station and user, respectively. Note that the notation (*) denotes the optimal solution. In problem (1.3), direct solution with SVD decomposition does not provide an optimal solution when N_t and N_r is not sufficiently large. However, it is not the case in M-MIMO since N_t and N_r are sufficiently large and the solution converges to the optimal solution obtained with the SVD of the channel. Therefore, one can say that there is a relationship between the optimal precoders and array response vectors. Accordingly, a near-optimal solution can be realized when $\mathbf{f} = \mathbf{a}_t(\phi_k^t, \theta_k^t)$ and $\mathbf{w} = \mathbf{a}_r(\phi_k^r, \theta_k^r)$, where k is the maximum steering direction [8].

The major drawback of the analog beamforming is that it can only adjust the signal phase and not the amplitude, leading to degradation of the spectral efficiency performance. Furthermore, as only a single RF chain is employed, the system can activate a single beam causing only a user to be efficiently served at a given time. Therefore, it is not suitable for the next-generation cellular networks where many applications require multiple beams to be activated at a given time to meet the data rate requirements [15].

1.1.3 Hybrid Precoding

The aforementioned limitations makes both analog beamforming and digital precoding impractical for M-MIMO systems. Therefore, several hybrid precoding techniques are proposed [9, 18, 19, 20, 21, 22, 23, 24] for efficiently transmitting a large amount of data. The hybrid precoding is realized by cascading a digital precoder in the baseband and an analog beamformer between the RF chains and antenna arrays. It combines the advantages of analog beamforming and digital precoding to optimize power consumption, hardware complexity, and spectral efficiency. In hybrid precoding, the purpose of the analog beamformer

is to improve the beamforming gain due to the accurate directional beams obtained through manipulating the phases of the transmitted signal. In contrast, the digital precoder performs additional digital signal processing to enable multiple data transmissions and mitigate the interference among them. A widely adopted fully connected massive MIMO architecture applying hybrid precoding is given in Fig. 1.4.

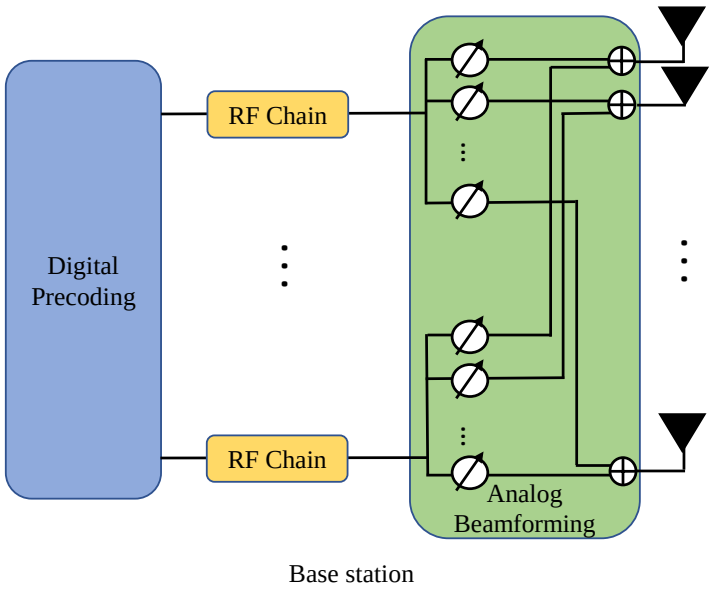


Figure 1.4: Fully connected single-user M-MIMO architecture with hybrid precoding

Note that it is possible to have hybrid precoding with single-user [24], multi-user [20], and sub-connected [18] M-MIMO systems.

Although hybrid precoding provides many advantages, it is still not practical for mmWave M-MIMO systems since using an enormous number of phase shifters leads to considerable hardware complexity, signal processing complexity, and power consumption. Thus, advanced antenna designs such as beamspace MIMO (B-MIMO) architecture [11], and lens antenna subarray (LAS)-MIMO architecture [25, 26] are proposed to reduce signal processing complexity and RF chain cost without significant performance degradation [27].

1.1.4 Beamspace MIMO Architecture

Due to the limitations of the digital precoding at mmWave frequencies, beamspace MIMO is proposed [11] since it significantly reduces the RF chain requirement that digital precoding suffers from by exploiting the inherent sparsity in mmWave channels. Utilizing a discrete lens antenna array (DLA) converts the spatial domain channel into the angular domain channel [11] to explore the channel sparsity. Accordingly, a significant decrement in the required RF chains is achieved due to the selection of dominant beams only. B-MIMO given in Fig. 1.5 is capable of creating narrow beams even with a few RF chains after the beam selection, reducing the inter-beam interferences [11]. Hence, reduced power consumption is achieved without compromising the system performance by performing beam selection [6, 11, 28]. However, each RF chain in B-MIMO can only serve one user at the same-time frequency resources. Therefore, the number of users that can be served is limited by the number of RF chains.

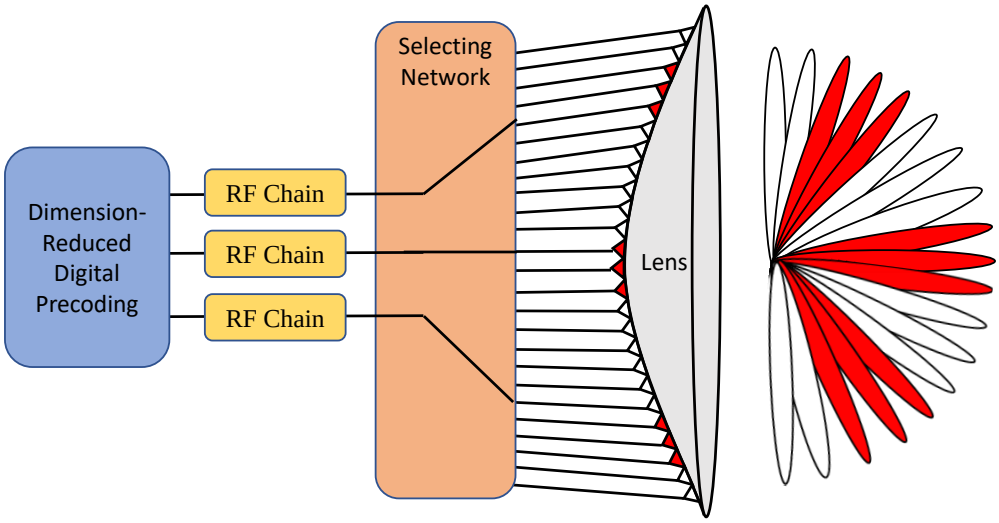


Figure 1.5: Beamspace MIMO architecture

1.2 Overview of Non-orthogonal Multiple Access (NOMA)

This subsection provides an overview of non-orthogonal multiple access (NOMA) by comparing its fundamental principles, advantages, and disadvantages with orthogonal multiple access (OMA) schemes.

The multiple access techniques of wireless communication systems have evolved considerably over the past few decades. Conventional OMA schemes have been adopted from first-generation (1G) to 4G mobile systems to allocate orthogonal wireless resources to multiple users. Specifically, 1G utilizes frequency division multiple access (FDMA), where each user is assigned a unique frequency resource to transmit its data. On the other hand, time division multiple access (TDMA), code division multiple access (CDMA), and orthogonal frequency division multiple access (OFDMA) are employed in 2G, 3G, and 4G, respectively. TDMA allocates an exclusive time slot to each user over the same frequency resource, while CDMA allows multiple users to transmit their data simultaneously over the same-time frequency resources using orthogonal codes. OFDMA intelligently combines FDMA and TDMA schemes to enable multiple access by allocating subsets of time-frequency resources to different users [29]. In theory, OMA schemes do not cause inter-user interference since all resources are orthogonal. Thus, low-complexity receivers can easily distinguish different users' signals [29, 30]. However, there are some drawbacks to OMA schemes. First, it restricts the number of users transmitting their signal since the orthogonal resources are not unlimited. Therefore, it cannot meet the massive number of connections required by massive machine-type communication (mMTC) devices [31]. Second, it cannot always attain the maximum sum-rate capacity in downlink due to the lack of channel knowledge [12]. Finally, it suffers from increased latency and scheduling overhead due to the necessity of dynamic scheduling in the uplink [32].

Recent investigations of NOMA as an alternative to OMA schemes have been conducted to counteract the aforementioned limitations. Non-orthogonal resource allocation in NOMA enables more users than the number of orthogonal resources to transmit their signals simul-

taneously and enhances the capacity [29]. However, it increases the receiver complexity due to advanced inter-user interference cancellation requirements. NOMA can also reduce the scheduling overhead by applying for comprehensive (sensing) random access, which enjoys grant-free in the uplink [33].

NOMA was first commercially introduced in Long Term Evolution-Advanced (LTE-A) under the name of multi-user superposition transmission (MUST) to enhance the sum rate [34]. MUST can be realized in the power domain. In the power domain NOMA (PD-NOMA), the transmitter directly superimposes the signals generated by different users depending on their channel condition. The power allocation is inversely proportional to the users' channel condition. The user with a poor channel condition (i.e., weak user) is allocated more power than the user with a rich channel condition (i.e., strong user). Note that the strong user experiences a high SNR. Thus, it can extract its own signals by first detecting the weak user's signal with high power and subtracting it from the received signal, leading to no additional processing for the weak user [29, 35]. This process is called successive interference cancellation (SIC), which is only required for the strong user. Since the weak user utilizes the classical receiver algorithms, it provides backward compatibility. However, the performance of PD-NOMA is highly affected by some practical factors, such as mobility, imperfect CSI, multi-user power allocation, and SIC error propagation [29].

1.3 Dissertation Outline

1.3.1 Beamspace MIMO Systems with Reduced Beam Selection Complexity

B-MIMO systems with proper beam selection promise to lower the required RF chains overhead with no noticeable degradation in performance. Most of the existing beam selection schemes are not practical due to the computational cost arising from iterative search or alternating optimization. Hence, this study examines the complexity reduction of the existing beam selection with incremental QR precoder (I-QR-P) and decremental QR precoder (D-QR-P). The proposed two-stage and three-stage algorithms reduce the complexity of D-

QR-P and I-QR-P, respectively. Both aim to lower complexity by decreasing the candidate beam size by eliminating the beams with no contribution to any user and using matrix perturbation theory to update QR decompositions. Numerical results reveal that the proposed algorithms considerably reduce the complexity while maintaining a similar sum-rate with baseline algorithms.

1.3.2 Heuristic Inspired Precoding for mmWave MIMO Systems with Lens Antenna Subarrays

A traditional array (TA)-MIMO architecture in mmWave with hybrid beamforming suffers from high power consumption and hardware overhead. Therefore, LAS-MIMO architecture has been recently proposed as a promising technology for a power-efficient system and reducing hardware cost and complexity. Additionally, the LAS-MIMO can offer spectral efficiency performance close to TA-MIMO and higher than single-lens antenna array (SLA)-MIMO. In this study, we propose a hybrid precoding algorithm for the LAS-MIMO in mmWave to efficiently control the LAS design. The precoding problem is formulated as a sparse reconstruction problem due to the sparse behavior of the mmWave channel. The proposed algorithm is an iterative process developed jointly using artificial bee colony (ABC) optimization with orthogonal matching pursuit (OMP) algorithms. In each iteration, the algorithm first selects the switches for each lens randomly using ABC and then uses OMP to approximate optimal unconstrained precoders. This process continues until achieving maximum spectral efficiency. The simulation results show that LAS has around a 30% increase in spectral efficiency compared to SLA while providing a significant gain in energy efficiency for single RF chain and multi RF chain scenarios.

1.3.3 Energy and Spectral Efficiency Tradeoff in NOMA: Multi-Objective Evolutionary Approaches

Non-orthogonal multiple access (NOMA) deployment in future wireless networks has been recently considered a promising radio access technology to enhance spectral efficiency. However, gain in spectral efficiency comes always with the cost of energy efficiency. In this study, we investigate the spectral efficiency and energy efficiency tradeoff in downlink NOMA with the consideration of quality of service (QoS) requirements based on three population-based multi-objective evolutionary algorithms (MOEAs): multi-objective particle swarm optimization (MOPSO), non-dominated sorting genetic algorithm-II (NSGA-II) and strength Pareto evolutionary algorithm-2 (SPEA2). The tradeoff is optimized and Pareto optimal solutions are obtained through MOEAs. The effectiveness of the algorithms is evaluated based on the hypervolume metric and the capability of solving multi-objective optimization problems. Simulation results reveal that SPEA2 outperforms NSGA-II and MOPSO. Furthermore, NSGA-II is the loser among all algorithms in terms of finding Pareto optimal results.

1.3.4 A Distributed User-Cell Association for Spectral and Energy Efficiency Tradeoff in Massive MIMO UDHNs

Massive MIMO enabled ultra-dense heterogeneous networks (UDHNs) have been considered as the indispensable and emerging approach to meet the demand of growing data traffic for next-generation networks. Although deploying many antennas causes high circuit power consumption in massive MIMO UDHNs, there is always a tradeoff between energy efficiency and spectral efficiency. Therefore, an energy- and spectral-efficient user-cell association will become crucial and challenging in massive MIMO UDHNs. This study addresses a user-cell association problem for the spectral and energy efficiency tradeoff. To this end, we formulate a convex multi-objective optimization problem and convert it into a single objective optimization problem where a priority is assigned for the energy efficiency and spectral efficiency with a weighting factor which means the problem can be adjusted whether the priority

is on energy efficiency or spectral efficiency. The problem aims to maximize the weighted sum of energy efficiency and spectral efficiency. As a solution, Lagrange duality analysis is performed, and a distributed game theoretical user-cell association (GTUCA) algorithm is developed, considering the fairness among users. The results confirm that the proposed algorithm outperforms the baseline algorithm, namely maximum rate-based cell selection, regarding energy efficiency and spectral efficiency when the weighting factor is set properly.

Chapter 2: Beamspace MIMO Systems with Reduced Beam Selection Complexity

Utilization of the millimeter-wave (mmWave) spectrum is presumed to be a key enabler for emerging next-generation wireless communication networks [36]. Since mmWave frequencies offer small wavelengths, it is possible to pack many antennas into small physical areas. This property enables a promising marriage between mmWave frequencies and massive multiple-input multiple-output (M-MIMO), thereby overcoming the severe free-space path-loss due to high directional beamforming gain [2]. MmWave M-MIMO also enhances spectral efficiency by allowing multiple data streams [5] and utilizing its larger bandwidth [6]. However, excessive power consumption and hardware cost are its drawbacks since each antenna entails its own RF chain. Beamspace MIMO (B-MIMO) proposed in [11] has, therefore, received much consideration to reduce the RF chain requirement by taking advantage of the inherent sparsity in the mmWave channels. In B-MIMO, the angular domain (i.e., beamspace) representation of the spatial channel is performed by employing a discrete lens array (DLA) at the base station to explore the channel sparsity [11]. Hence, a reduction in radio-frequency (RF) chains required is achieved without compromising the system performance by performing beam selection [11, 6].

Magnitude maximization beam selection (MM-BS) [11] assigns a beam proving the maximum received signal power to each user. Nevertheless, it suffers from high multi-user interference and RF redundancy; however, these limitations were handled by interference-aware beam selection (IA-BS) [6]. Several signal-to-interference-plus-noise ratio (SINR) and sum-rate maximization-based iterative beam selection algorithms were investigated in [37, 38, 39], while zero-forcing (ZF) precoding was employed to eliminate the multi-user interference. By

enabling multiple-beam group selection in [40], rate-loss was mitigated by creating a reliable channel cluster for each user. Recently, beam selection and precoding schemes were studied for a new B-MIMO architecture with lens antenna subarrays (LASs) [27]. Several studies [41, 42] on beam selection for wideband B-MIMO were also conducted to overcome the beam squint occurring in mmWave.

Due to the limited power at the base stations and to avoid latency, the complexity of several beam selection methods has been investigated. In [43], low-complexity beam selection methods were proposed based on the graph theory and heuristic greedy algorithm. QR decomposition of the beamspace channel was inspected in [44], and an iterative beam selection algorithm and a precoder for eliminating multi-user interference were proposed. Along with its outstanding system performance over the existing algorithms, it is not practical due to its high complexity. In [45], the complexity of the conventional ZF precoder and QR precoder [44] were probed and reduced by updating the factorization or decomposition results using matrix perturbation theory instead of performing from scratch again.

This study proposes two complexity-reduced beam selection algorithms. The main contributions are summarized as follows:

- The proposed two-stage and three-stage beam selection algorithms remarkably decrease the complexity of decremental QR (D-QR) precoder (D-QR-P) [45] and incremental QR (I-QR) precoder (I-QR-P) [45], respectively.

- Both enjoy considerably higher sum-rate than the existing algorithms proposed in [6, 11, 37].

- The sum-rate performance of D-QR-P is high; however, it suffers from high complexity as the number of antennas increases. The two-stage beam selection obtains the complexity reduction with almost identical sum-rate performance.

- The three-stage beam selection algorithm significantly reduces complexity while compensating for the sum-rate performance loss at low signal-to-noise ratio (SNR)s that I-QR-P suffers from.

- Both utilize the matrix perturbation to update QR decomposition and aim to decrease the beam size by removing the beams with no contribution to any user from the beam set.

The following notations are used in this Chapter[1]: $\mathbf{A}$, $\mathbf{a}$, a, $\mathcal{A}$ denote a matrix, a vector, a scalar, and a set, respectively. $\mathbf{A}^H$, $\mathbf{A}^T$, $\mathbf{A}^{-1}$ are Hermitian, transpose, and inverse of $\mathbf{A}$ respectively. diag($\mathbf{a}$) is a diagonal matrix with $\mathbf{a}$ on its diagonal. $\mathbf{I}$ is the identity matrix, and $\mathbb{C}^{M\times N}$ is the space of $M \times N$ complex-valued matrices, $\mathbb{E}[\cdot]$ is the expectation operator, and Card($\mathcal{A}$) is the cardinality of $\mathcal{A}$.

2.1 System Model

This study considers a B-MIMO architecture in a downlink mmWave scenario. For a conventional M-MIMO architecture where a base station consists of N antennas modeled as a uniform linear array (ULA) to serve K single-antenna users, the received signal vector $\mathbf{y} \in \mathbb{C}^{K\times 1}$ is expressed as

$$\mathbf{y} = \mathbf{H}^H\mathbf{P}\mathbf{s} + \mathbf{n}, \tag{2.1}$$

where $\mathbf{H} = [\mathbf{h}_1, \mathbf{h}_2, \ldots, \mathbf{h}_K] \in \mathbb{C}^{N\times K}$ stands for the channel matrix where $\mathbf{h}_k \in \mathbb{C}^{N\times 1}$ denotes the channel vector between k-th user and the base station. The normalized transmitted signal vector is defined by $\mathbf{s} \in \mathbb{C}^{K\times 1}$ fulfilling $\mathbb{E}[\mathbf{s}\mathbf{s}^H] = \mathbf{I}_K$ and $\mathbf{P} \in \mathbb{C}^{N\times K}$ is the precoding matrix designed to cancel the multi-user interference. Additionally, $\mathbf{n} \sim \mathcal{CN}(0, \sigma^2\mathbf{I}_K)$ is additive white Gaussian noise (AWGN).

[1]Part of this chapter was published in [28]. Permission is included in Appendix A.

This study considers the Saleh-Valenzuela mmWave channel model [9], which models $\mathbf{h}_k$ as [6]

$$\mathbf{h}_k = \sum_{l=0}^{L} \alpha_k^{(l)} \mathbf{a}\left(\varphi_k^{(l)}\right). \tag{2.2}$$

In (3.3), the complex gain and the spatial direction of the k-th user for the l-th path are stated by $\alpha_k^{(l)}$ and $\varphi_k^{(l)}$, respectively. Note that $l = 0$ refers to the line-of-sight (LoS) component, while $l = 1, 2, ..., L$ represents the non-line-of-sight (NLoS) components. For the N-element typical ULA, the array steering vector of the l-th path is stated as [6] $\mathbf{a}(\varphi) = \frac{1}{\sqrt{N}}\left[e^{-j2\pi\varphi b}\right]_{b\in\mathcal{I}(N)} \in \mathbb{C}^{N\times 1}$, where $\mathcal{I}(N) = \{p - (N-1)/2, p = 0, 1, ..., N-1\}$ is a symetric set of indices centered around zero. Moreover, $\varphi = \frac{d}{\lambda}\sin(\theta)$, where θ, $d = \lambda/2$, and λ denote the physical direction, the antenna spacing, and the carrier signal wavelength, respectively.

The mmWave channel is inherently sparse since the LoS components of the channel strongly dominate the NLoS components [11]. The use of a DLA at the base station converts the spatial channel (3.3) into the beamspace channel as it behaves like a spatial discrete Fourier transformer represented by matrix $\mathbf{U} = \left[\mathbf{a}\left(\bar{\varphi}_1\right), \mathbf{a}\left(\bar{\varphi}_2\right), \cdots, \mathbf{a}\left(\bar{\varphi}_N\right)\right]^H \in \mathbb{C}^{N\times N}$ that consists of the array steering vectors corresponding to N predefined orthogonal directions covering the entire angular space [6, 45], where $\bar{\varphi_n} = \frac{1}{N}(n - \frac{N+1}{2})$ for $n = 1, 2, \ldots, N$ stands for the predefined spatial directions. Ultimately, the beamspace channel[2] is $\mathbf{H}_b = \mathbf{UH}$, and the corresponding received signal vector $\mathbf{y}_b$ is given as

$$\mathbf{y}_b = {\mathbf{H}_b}^H \mathbf{Ps} + \mathbf{n}. \tag{2.3}$$

The beamspace channel can be represented with a considerably small number of precisely chosen beams without compromising the system performance due to intrinsic sparsity. As a result of the beam selection process, the dimension-reduced M-MIMO, so-called B-MIMO, is obtained as in (2.4), where the dimension-reduced beamspace channel is $\tilde{\mathbf{H}}_\mathbf{r} = \mathbf{H}_\mathbf{b}(\mathcal{B}, :)$,

[2] We assume that channel state information (CSI) is known by the base station where channel estimation can be performed as in [46].

in which $\mathcal{B}$ denotes the set containing the chosen beam indexes, and $\mathbf{P}_r$ stands for the corresponding dimension-reduced precoding matrix [6].

$$\mathbf{y}_b \approx \tilde{\mathbf{H}}_r^H \mathbf{P}_r \mathbf{s} + \mathbf{n}. \tag{2.4}$$

Note that using DLA and performing beam selection reduces the required RF chains while preserving the narrow beamwidth [37]. As a result, the adopted B-MIMO architecture is suitable for mmWave systems due to low hardware complexity and high antenna gain properties, even with fewer RF chains [37].

Additionally, the base station communicates with every user in set $\mathcal{K} = \{1, 2, \ldots, K\}$ via *only one data stream* to assure the spatial multiplexing gain. Thus, the number of data streams and RF chains are set to $N_s = K$ and $N_{RF} = K$, respectively.

2.1.1 QR Decomposition of Dimension-Reduced B-MIMO

Let $\tilde{\mathbf{H}}_r$ be decomposed into a unitary matrix of $\mathbf{Q} \in \mathbb{C}^{K \times K}$ and an upper triangular matrix of $\mathbf{R} \in \mathbb{C}^{K \times K}$, such that $\tilde{\mathbf{H}}_r = \mathbf{QR}$ [47]. Hence, (2.4) becomes $\mathbf{y}_b \approx \mathbf{R}^H \mathbf{s} + \mathbf{n}$ when the precoder is $\mathbf{P}_r = \mathbf{Q}$. Thus, the k-th user receives [44]

$$\tilde{y}_k = \tilde{r}_{kk} s_k + I_k + n_k, \tag{2.5}$$

where $\tilde{r}_{kk}$ equals to the k-th element of diag($\mathbf{R}$), and the interference $I_k = \sum_{k>j} \tilde{r}_{kj} s_j$ can be eliminated for all users by diagonalizing $\mathbf{R}^H$. Let the precoder be $\mathbf{P}_r = \mathbf{QG}$, where $\mathbf{G} \in \mathbb{C}^{K \times K}$ is the Given rotations such that the diagonal elements of $\mathbf{R}^H \mathbf{G} \in \mathbb{C}^{K \times K}$ and $\mathbf{R}^H$ are same. Accordingly, the sum-rate is expressed as in (2.6), where ρ and $\gamma = \rho/\sigma^2$ stand for the signal power and SNR, respectively [44, 45].

$$R_{sum} = \sum_k \log_2 \left(1 + \frac{\gamma}{K} \tilde{r}_{kk}^2\right) \quad \text{bit/s/Hz}. \tag{2.6}$$

2.1.2 QR Decomposition Update

Matrix perturbation theory allows QR decomposition of a matrix to be updated easily instead of recomputing from scratch when the matrix undergone a modification. Suppose we have the decomposition of $\mathbf{H}_b = \mathbf{QR}$, and let $\mathbf{H}_b^{(\pm n)} = \mathbf{H}_b - \mathbf{u}\mathbf{z}^T$ represent the modified matrix after inserting or eliminating the n-th row (i.e., $\mathbf{z}^T$), where $\mathbf{u} = \mp\mathbf{e}_n$. Note that QR decomposition is called I-QR decomposition when a new row inserted to a matrix, while it is called D-QR decomposition when a row is deleted from a matrix. The upper-Hessenberg matrix of $\mathbf{H}_b^{(\pm n)}$ can be expressed as $\mathbf{H}_b^{(\pm n)} = \mathbf{Q}[\mathbf{R} + \mathbf{w}\mathbf{z}^T]$, where $\mathbf{w} = \mathbf{Q}^H\mathbf{u}$. Denoting $\mathbf{J}_m$ is a Given rotation acting in planes m and $m+1$, where $m = 1, 2, \ldots, (N-1)$, the series of rotations is obtained by $\mathbf{J}_1^T \ldots \mathbf{J}_{N-1}^T\mathbf{w} = \mp||w||_2\mathbf{e}_1$, where $\mathbf{e}_1 = (1, 0, \ldots, 0)$ represents the unit vector. Assuming that same rotations are applied to $\mathbf{R}$, we acquire an upper-Hessenber matrix of $\mathbf{H}_0 = \mathbf{J}_1^T \ldots \mathbf{J}_{N-1}^T\mathbf{R}$. Ultimately, $\mathbf{H}_1 = \mathbf{J}_1^T \ldots \mathbf{J}_{N-1}^T[\mathbf{R} + \mathbf{w}\mathbf{z}^T] = \mathbf{H}_0 \mp ||w||_2\mathbf{e}_1\mathbf{z}^T$. To update $\mathbf{R}^{(\pm n)}$, $(N-1)$ $\mathbf{G}_m$ Given rotations are applied to $\mathbf{H}_1$ [47] such that

$$\mathbf{R}^{(\pm n)} = \mathbf{G}_{N-1}^T\mathbf{G}_{N-2}^T \ldots \mathbf{G}_1^T\mathbf{H}_1 \tag{2.7}$$

is an upper triangular matrix. Then, $\mathbf{Q}$ can be updated as [47]

$$\mathbf{Q}^{(\pm n)} = \mathbf{Q}\mathbf{J}_{N-1} \ldots \mathbf{J}_1\mathbf{G}_1 \ldots \mathbf{G}_{N-1}. \tag{2.8}$$

Note that, we omit the values of rotation matrices $\mathbf{J}_m$ and $\mathbf{G}_m$; however, details are available in [47].

2.2 Proposed Beam Selection Algorithms

This section revisits the I-QR-P and D-QR-P given in [45], and investigates the complexity of beam selection for further reduction.

2.2.1 Proposed Two Stage Beam Selection with D-QR-P

Conventional QR precoder (C-QR-P) [44] for beam selection is an iterative process where beams with a minimum contribution to system performance (i.e., sum-rate) are discarded, causing unaffordable computational complexity due to the required QR decomposition from scratch in each iteration such that the number of required iterations in the outer loop is $(N-K)$ in the i-th iteration, where $i = 0, 1, \ldots, (N-K-1)$. However, the main complexity arises from the inner loop which contains $(N-i)$ QR decomposition operation to eliminate the beam with the least contribution to the system sum-rate from the beam set. Since the complexity to compute the QR decomposition from scratch is $\mathcal{O}(2(N-i)K^2)$, the total complexity of C-QR-P is $\sum_{i=0}^{N-K-1}(N-i)\mathcal{O}(2(N-i)K^2) = \mathcal{O}((2K^2N^3 - 2K^5)/3)$ [44].

To overcome the complexity problem of C-QR-P, D-QR-P [45] was proposed by utilizing matrix perturbation theory. Note that D-QR-P updates $\mathbf{R}$ and $\mathbf{Q}$ using (2.7) and (2.8), respectively while D-QR-P regenerates from scratch when a row is deleted. The QR update process can be executed in $\mathcal{O}(4K(N-i) + 4K^2)$ in the i-th iteration. Thus, the beam selection with D-QR-P requires the complexity of $\sum_{i=0}^{N-K-1}(N-i)\mathcal{O}(4K(N-i) + 4K^2) = \mathcal{O}((4KN^3 + 6K^2N^2 - 10K^4)/3)$ [45]. Although it significantly reduces the complexity while providing almost similar sum-rate performance with [44], it is still not practical, especially when $N \geq K$ is large [45]. Therefore, we propose Algorithm 2.1 which consists of the following two stages.

2.2.1.1 Identify M Strongest Beams for All K Users

This stage aims to reduce the number of QR updates in the inner loop of the D-QR-P by decreasing the number of candidate beams. Let's consider the following two definitions for lucidity.

Definition 1: Let $b_{k,m}$ represent the m-th strongest beam of the k-th user, the strongest beam $b_{k,1} \in \mathcal{D}$ contains the most of the channel power, and it is the first element of the

Algorithm 2.1 Two Stage Beam Selection with D-QR-P

Input: $\mathbf{H}_b$, $\mathcal{D}$, $\mathcal{K}$, $\mathcal{G} = \emptyset$, $\mathcal{B} = \emptyset$, $M = 0$
Output: $\tilde{\mathbf{H}}_r$
Stage 1: Identify M strongest beams for all users
while $\text{Card}(\mathcal{G}) < K$ **do**
  $M = M + 1$,
  for $l = 1 : K$ **do**
    $\mathbf{t} = \text{sort}(|\mathbf{H}_b(:, l)|)$, and $\mathcal{G} = \mathcal{G} \cup \{\mathcal{D}(\mathbf{t}(1 : M))\}$,
  end
end
Stage 2: Beam Selection with D-QR-P
$\mathbf{A} = \mathbf{H}_b(\mathcal{G}, :)$, and $\mathbf{A} = \mathbf{QR}$,
for $j = 0 : \text{Card}(\mathcal{G}) - K - 1$ **do**
  for $k = 1 : \text{Card}(\mathcal{G}) - j$ **do**
    $\mathbf{u} = \mathbf{e}_k$, and $\mathbf{z} = \mathbf{A}(k, :)$,
    Update $\mathbf{R}^{(-k)}$, $\mathbf{Q}^{(-k)}$, and obtain $R_{sum}^{(k)}$ using (2.6),
  end
  $b_j = \underset{k}{\text{argmax}}\{R_{sum}^{(k)}\}$, $\mathcal{G} = \mathcal{G} \backslash \{b_i\}$, $\mathcal{B} = \mathcal{B} \cup b_j$,
  $\mathbf{A} = \mathbf{A}(\mathcal{G}, :)$, $\mathbf{R} = \mathbf{R}^{(-b_j)}$, and $\mathbf{Q} = \mathbf{Q}^{(-b_j)}$,
end
$\tilde{\mathbf{H}}_r = \mathbf{H}_b(\mathcal{B}, :)$

sorted $|\mathbf{H}_b(:, k)|$ in descending order. Then, $\mathcal{G}_k^* = \{b_{k,m}\}_{m=1}^{M} \in \mathcal{D}$ includes the indices for the M strongest beams, where $\mathcal{D} = \{1, 2, \ldots, N\}$ is the set containing all beams available.

Definition 2: Users sharing identical beams are called interfering user (IU)s, while a user is defined as non-interfering user (NIU) if its beam is not selected by any other users. The sets representing the IUs and NIUs are defined by $\mathcal{K}_{IU}$ and $\mathcal{K}_{NIU}$, respectively.

In this stage, the algorithm first identifies $\mathcal{G}_k^*$ for all K users. Then, the candidate beam set is $\mathcal{G} = \mathcal{G}_1^* \cup \mathcal{G}_2^* \cup \cdots \cup \mathcal{G}_K^*$, where $\mathcal{G} \subset \mathcal{D}$ and $\text{Card}(\mathcal{G}) \leq MK$. Note that this stage decides the value of M to provide enough beam diversity for all K users so that they can be served by K best-unshared beams simultaneously. An example of how the value of M is decided is shown in Fig. 2.1. In Fig. 2.1 (a) $\mathcal{G} = \{1, 3, 8\}$ when $M = 1$. Since $\text{Card}(\mathcal{G}) < K$, there are not enough beams for all K users. Therefore, the algorithm tries the case of $M = 2$ in Fig. 2.1 (b), where $\mathcal{G} = \{1, 3, 7, 8\}$ and still $\text{Card}(\mathcal{G}) < K$. Thence, the case of $M = 3$ is

tried as in Fig. 2.1 (c), where $\mathcal{G} = \{1, 2, 3, 4, 7, 8\}$ and $\text{Card}(\mathcal{G}) > K$. There is now enough beam diversity to provide an unshared beam for all K users. Since beam 5 and 6 have no contribution to any users, they are removed from the candidate beam set as in Fig. 2.1 (d).

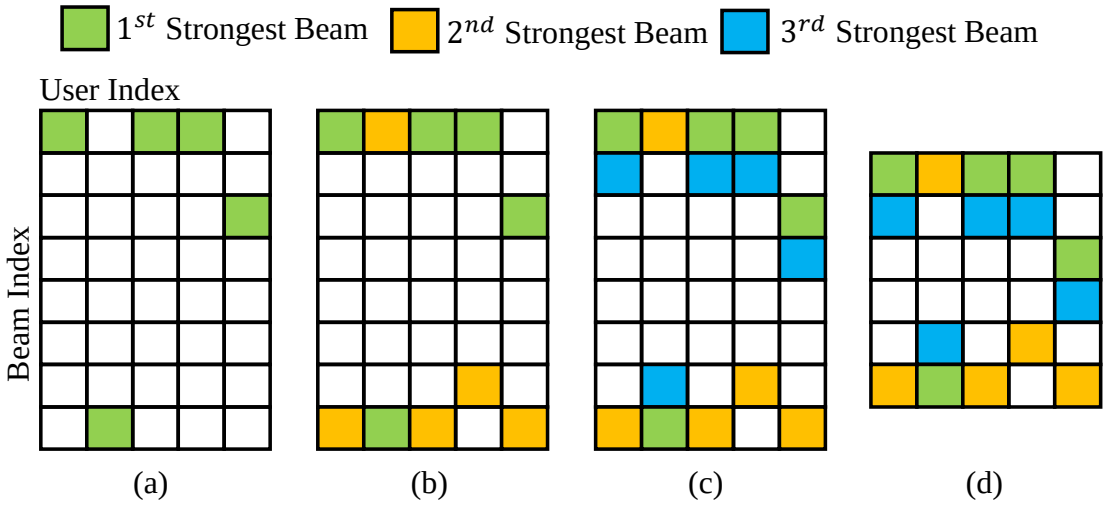


Figure 2.1: An example of showing the decision of M (a) $M = 1$, (b) $M = 2$, (c) $M = 3$, (d) Candidate beam set.

If this stage decides that $M = 1$, there are no IUs in the network. In this case, Algorithm 2.1 selects the K strongest beams as in *Definition 1*, resulting in a near-optimal solution, already proven in [6]. Additionally, multi-user interference is eliminated by the precoder $\mathbf{P}_r = \mathbf{Q}$.

2.2.1.2 Beam Selection with D-QR-P

If $M > 1$, Algorithm 2.1 performs the beam selection process summarized in stage 2. Similar to the D-QR-P in [45], the beam that contributes the least to the sum-rate performance is discarded in each iteration. However, the number of $\mathbf{Q}$ and $\mathbf{R}$ updates in the inner loop is reduced from $(N - i)$ to $(\text{Card}(\mathcal{G}) - i)$. Thence, the overall complexity of Algorithm 2.1 is $\mathcal{O}(2\text{Card}(\mathcal{G})K^2) + \mathcal{O}((4K(\text{Card}(\mathcal{G}))^3 + 6K^2(\text{Card}(\mathcal{G}))^2 - 10K^4)/3)$.

2.2.2 Proposed Three Stage Beam Selection with I-QR-P

We propose Algorithm 2.2 to reduce the complexity of I-QR-P [45]. It consists of the following three stages.

2.2.2.1 Beam Selection for NIUs

The complexity of the I-QR-P arises from the search process to identify the first beam that contributes the most to the system sum-rate since it requires complex QR decomposition for all N beams. To avoid this process, Algorithm 2.2 groups users as IUs and NIUs inspired by [6] in this stage. To do this, it first identifies the strongest beam set $\mathcal{B} = \{b_1, b_2, \ldots, b_K\}$ for all K user as in *Definition 1.* Since the probability of having IUs, $P = 1 - \frac{N!}{N^K(N-K)!}$, is considerably high in spite of N being large [6], it then removes the repeatedly selected beams from $\mathcal{B}$ and defines the non-interfering beam set as $\mathcal{V} \subset \mathcal{D}$. Since the beams in set $\mathcal{V}$ contain the most of the channel power and cause considerably low interference to others, the algorithm directly assigns these beams to the NIUs. Note that no beams are yet selected for the IUs in this stage.

2.2.2.2 Identify M Strongest Beams for the IUs

After NIUs are directly assigned with the beams in $\mathcal{V}$, we have a beamspace channel matrix $\mathbf{A} = \mathbf{H}_b(\mathcal{V}, :) \in \mathbb{C}^{\mathrm{Card}(\mathcal{V}) \times K}$ for the NIUs and candidate beam set for IUs are updated as $\mathcal{D} = \mathcal{D} \backslash \mathcal{V}$, where $\mathrm{Card}(\mathcal{D}) = N - \mathrm{Card}(\mathcal{V})$. This stage aims to reduce the size of the candidate beams for the IUs. Let $\mathbf{B} = \mathbf{H}_b(\mathcal{D}, \mathcal{K}_{IU}) \in \mathbb{C}^{\mathrm{Card}(\mathcal{D}) \times \mathrm{Card}(\mathcal{K}_{\mathcal{IU}})}$ represent the beamspace channel for the IUs. Following the same process presented in Section 2.2.1.1, Algorithm 2.2 decides the value of M and acquires a new beam set for IUs as $\mathcal{G} \subset \mathcal{D}$, where $\mathrm{Card}(\mathcal{G}) \leq M\mathrm{Card}(\mathcal{K}_{IU})$.

Algorithm 2.2 Three Stage Beam Selection with I-QR-P

Input: $\mathbf{H}_b$, $\mathcal{D}$, $\mathcal{K}$, $\mathcal{K}_{IU} = \mathcal{K}_{NIU} = \mathcal{G} = \mathcal{B} = \emptyset$, $M = 0$
Output: $\tilde{\mathbf{H}}_r$
Stage 1: Beam Selection for NIUs
for $k = 1 : K$ **do**
  $b_k = \text{argmax}_{b_k} |\mathbf{H}_b(:, k)|$, and $\mathcal{B} = \mathcal{B} \cup \{b_k\}$,
end
Set $\mathcal{V} = \text{unique}(\mathcal{B})$,
for $i = 1 : \text{Card}(\mathcal{V})$ **do**
  $j = \text{find}(\mathcal{B} == \mathcal{V}(j))$,
  if $\text{Card}(j) > 1$ **then**
    $\mathcal{V}(\mathcal{V} == \mathcal{V}(j)) = \emptyset$, and $\mathcal{K}_{IU} = \mathcal{K}_{IU} \cup \{j\}$,
  end
end
$\mathcal{B} = \mathcal{V}$, $\mathcal{K}_{NIU} = \mathcal{K} \backslash \mathcal{K}_{IU}$, $\mathcal{D} = \mathcal{D} \backslash \mathcal{B}$, $\mathbf{A} = \mathbf{H}_b(\mathcal{B}, :)$, and $\mathbf{B} = \mathbf{H}_b(\mathcal{D}, \mathcal{K}_{IU})$,
Stage 2: Identify M strongest beams for the IUs
Follow same steps in **Algorithm 2.1** to obtain $\mathcal{G}$
Stage 3: Beam Selection with I-QR-P for IUs
$\mathbf{A} = \mathbf{QR}$,
for $i = 1 : \text{Card}(\mathcal{K}_{IN})$ **do**
  for $j = 1 : \text{Card}(\mathcal{G})$ **do**
    $\mathbf{u} = \mathbf{e}_j$, and $\mathbf{z} = \mathbf{B}(j, :)$,
    Update $\mathbf{R}^{(+j)}$, $\mathbf{Q}^{(+j)}$, obtain $R_{sum}^{(j)}$ using (2.9),
  end
  $b_i = \underset{j}{\text{argmax}}\{R_{sum}^{(j)}\}$, $\mathcal{B} = \mathcal{B} \cup \{b_i\}$, $\mathcal{G} = \mathcal{G} \backslash \{b_i\}$,
  $\mathbf{Q} = \mathbf{Q}^{b_i}$, and $\mathbf{R} = \mathbf{R}^{b_i}$,
end
$\tilde{\mathbf{H}}_r = \mathbf{H}_b(\mathcal{B}, :)$

2.2.2.3 Beam Selection for the IUs with I-QR-P

Algorithm 2.2 overcomes the computational complexity of I-QR-P [45] mentioned in Section 2.2.2.1 since it does not include this step. Instead, it first decomposes the channel matrix $\mathbf{A} = \mathbf{QR}$ obtained for the NIUs in stage 1, then keeps adding a new row to $\mathbf{A}$ (i.e., a new beam) from $\mathcal{G}$ iteratively until all IUs have an unshared beam, and $\mathbf{R}$ and $\mathbf{Q}$ are updated using (2.7) and (2.8), respectively. Since the number of selected beams K' is less

than K with I-QR-P beam selection, (2.6) is modified as [45]

$$R_{sum} = \sum_{k} \log_2 \left(1 + \frac{\gamma}{K'} \tilde{r}_{kk}^2\right) \quad \text{bit/s/Hz}, \tag{2.9}$$

where $\tilde{r}_{kk}$ is the k-th element of diag($\mathbf{R}(1:K', 1:K')$).

For the IUs, this stage requires Card($\mathcal{K}_{IN}$) iterations and Card($\mathcal{G}$) QR decomposition updates in the i-th iteration. Thus, the total complexity is $\mathcal{O}(2\text{Card}(\mathcal{V})\text{Card}(\mathcal{K}_{NIU})^2) + \mathcal{O}((3\text{Card}(\mathcal{G})^2\text{Card}(\mathcal{K}_{IN})^2 + 2\text{Card}(\mathcal{G})\text{Card}(\mathcal{K}_{IN})^3)/3)$.

2.3 Performance Evaluation

This section presents the results for the proposed beam selection algorithms. We have gauged their performance against benchmark algorithms, D-QR-P, I-QR-P [45], IA-BS [6], MM-BS [11], and maximizing SINR [37], for a fair comparison. Note that, we do not include the results for C-QR-P[3].

We generate a mmWave B-MIMO, where the base station is a ULA with $N = 256$ serving $K = 16$ single antenna users, having random distribution, simultaneously. The channel consists of one LoS component defined by $\alpha_k^{(0)} \sim \mathcal{CN}(0, 1)$ and two NLoS components given as $\alpha_k^{(1,2)} \sim \mathcal{CN}(0, 10^{-2})$ where the spatial direction φ is uniformly distributed over $[-\pi/2, \pi/2]$. Additionally, all results are produced on a computer with a 16 GB RAM and 3.4 GHz Intel i7-6700 CPU, and averaged over 500 channel realizations.

The achievable sum-rate performance of the algorithms is compared in Fig. 2.2. It is evident from the plot that the proposed algorithms outperform IA-BS, MM-BS, and maximizing SINR while they perform almost identically with I-QR-P and D-QR-P at high SNRs. However, I-QR-P suffers a slight performance loss at low SNRs, which can be compensated by Algorithm 2.2. This is because I-QR-P starts with the QR decomposition of only one row (i.e., a beam), which provides limited information about the sparse channel, and adds

[3] It was already proven in [45] that I-QR-P and D-QR-P provide almost similar sum-rate with the C-QR-P but with considerably less complexity.

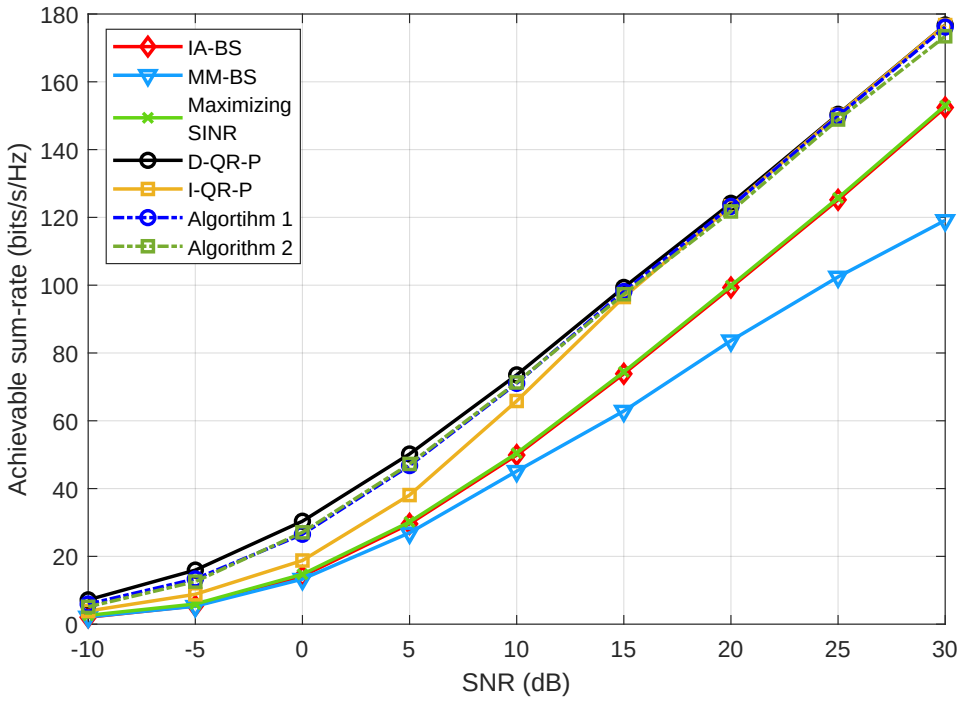


Figure 2.2: Achievable sum-rate vs SNR, where $N = 256$ and $K = 16$.

new beams iteratively. In contrast, Algorithm 2.2 starts with the QR decomposition of the channel matrix acquired for the NIUs, which delivers more information about the sparse channel than I-QR-P at the initial step and keeps adding new beams.

The effect of beam diversity is evaluated in Fig. 2.3. Since Algorithm 2.1 and Algorithm 2.2 choose M adaptively to provide enough diversity, as mentioned in Fig. 2.1, they outperform the other fixed cases. The case of $M = 2$ has poor performance, especially for Algorithm 2.2, due to not having enough diversity.

The increment in beam resolution increases the sum-rate for the proposed algorithms, as shown in Fig. 2.4 where N varies from 128 to 1024. However, depending on the sparsity, we obtain a performance gap between them. When high sparsity (i.e., $K \ll N$) exists, this performance gap decreases, so that Algorithm 2.1 is superior to Algorithm 2.2 in case of low sparsity.

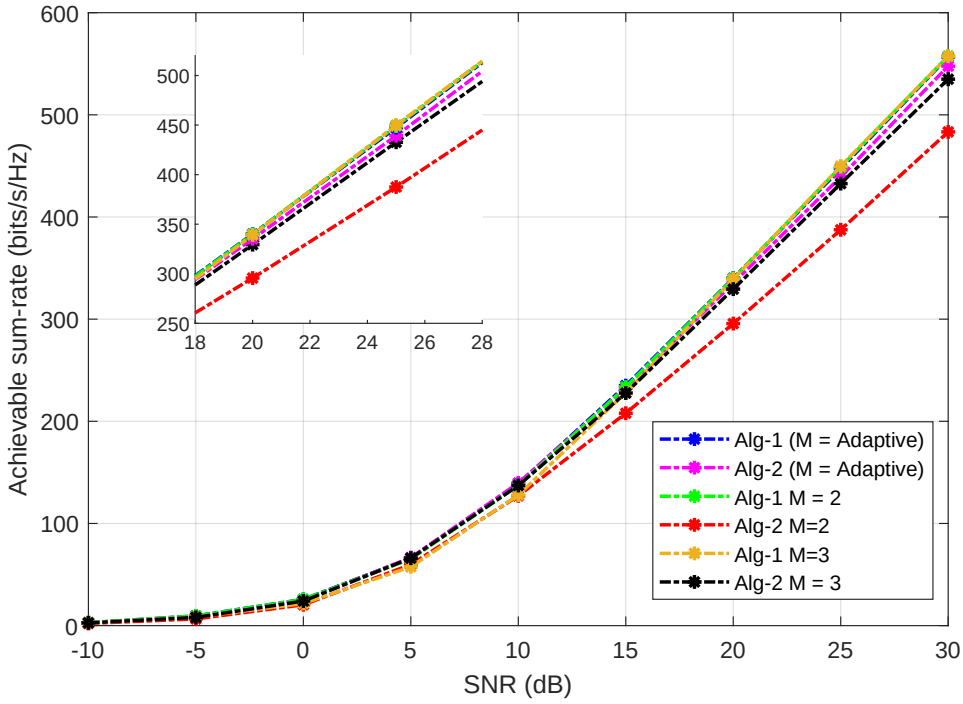


Figure 2.3: Effect of M on sum-rate, where $N = 256$ and $K = 64$.

Next, we compare the average running times in Fig. 2.5 and Fig. 2.6. Note that both figures do not provide the result for the complexity evaluation of D-QR-P since it was already stated in [45] that I-QR-P enjoys much less complexity than D-QR-P.

In Fig. 2.5, we set $K = N/16$, while N varies from 128 to 1024. Both Algorithm 2.1 and Algorithm 2.2 perform considerably faster beam selection than the I-QR-P. Note that Algorithm 2.2 outperforms other algorithms and its speed-up factor gets more prominent as the number of antenna increases since it decreases the search size significantly in stage 2 and selects beams for the IUs from the set $\mathcal{G}$ in stage 3. A low complexity beam selection for the NIUs is already performed in stage 1. However, the baseline I-QR-P selects beams for all K users from the beam set $\mathcal{D}$, leading to increased complexity. Fig. 2.2 and Fig. 2.5 reveal that reducing the size of candidate beams is critical in speeding up the beam selection while maintaining almost the same sum-rate performance. In other words, if a beam does not contribute to any user, there is no need to consider it in the selection process.

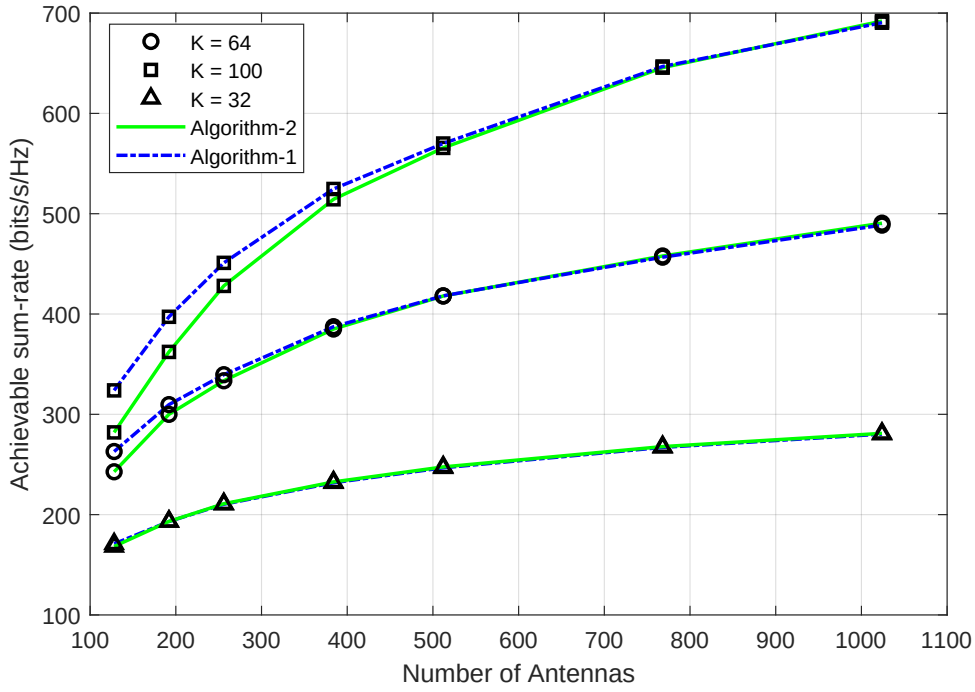


Figure 2.4: Effect of the sparsity on sum-rate.

The effectiveness of the proposed algorithms for a sparse environment (i.e., $K \ll N$) is evaluated in Fig. 2.6. In this case, Algorithm 2.1 and Algorithm 2.2 still select beams faster than the benchmark. The average run time reduces for Algorithm 2.2 as the sparsity increases since the probability of choosing the same strongest beam, $P = 1 - \frac{N!}{N^K(N-K)!}$, in stage 1 decreases as N grows when $K = 64$ is fixed. For example, $P \approx 99\%$ when $N = 256$, $P \approx 93\%$ when $N = 768$, and $P \approx 82\%$ when $N = 1024$. Since the decrease in P and increase in beam diversity occur as N grows, it is more likely to have fewer IUs in stage 1. Thus, most users directly select their strongest beams in stage 1 because NIUs outnumber IUs. Consequently, the number of iterations decreases to select beams for IUs in stage 3 since the main complexity arising from the QR update is avoided. The behavior of Algorithm 2.2 is different in Fig. 2.5 and Fig. 2.6. This is because the simulation setup in Fig. 2.5, where $K = N/16$), causes an increase in P. For example, $P \approx 38\%$ when $N = 256$, $P \approx 78\%$ when $N = 768$, and $P \approx 87\%$ when $N = 1024$. Therefore, there will be more IUs with this setup as N grows, increasing complexity in stage 3.

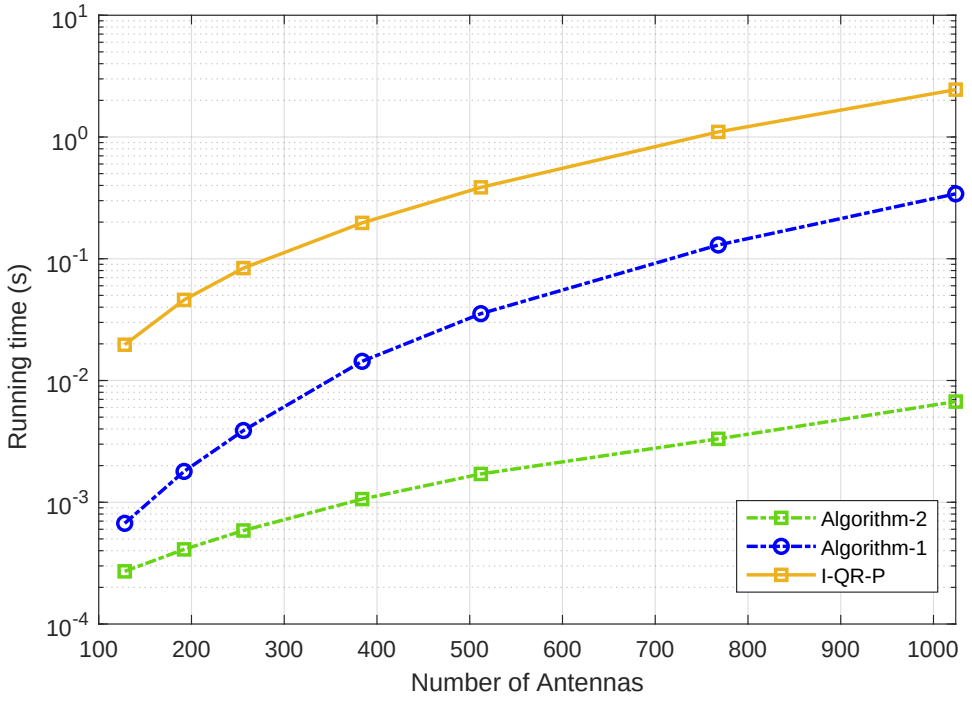


Figure 2.5: Averaged running time vs number of antennas N, where $K = N/16$.

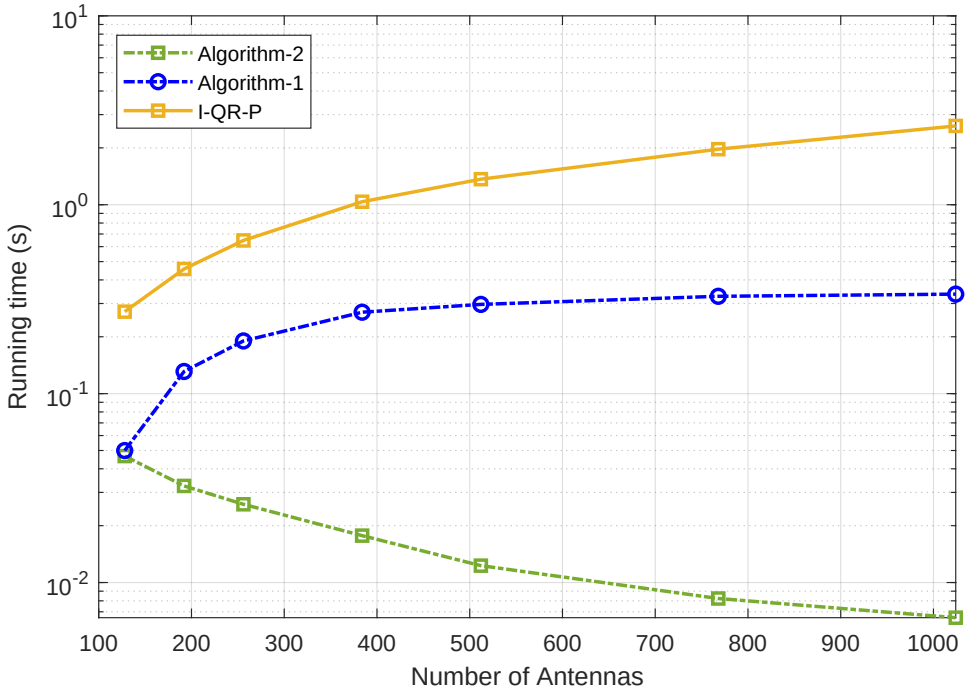


Figure 2.6: Averaged running time vs number of antennas N with fixed $K = 64$.

Chapter 3: Heuristic Inspired Precoding for Millimeter-Wave MIMO Systems with Lens Antenna Subarrays

Next-generation wireless communications are presumed to meet the demand for higher spectral efficiency (bits/s/Hz) and handle exponentially growing traffic volume [48]. To this end, less-congested millimeter-wave (mmWave) spectrum utilization is considered a promising solution to meet higher spectral efficiency requirements and deal with enormous traffic demand [36]. A forte of mmWave frequencies is the ability to pack a large number of antenna elements into small physical areas due to smaller wavelengths. Hence, mmWave facilitates the use of massive multiple-input multiple-output (MIMO) which can overcome the severe free-space path loss due to high directional beamforming gain [2]. In addition, it is possible to enhance spectral efficiency with massive MIMO by allowing multiple data streams with proper precoding techniques [5]. Typically, precoding in traditional-array (TA)–MIMO is performed digitally where each antenna element requires a dedicated radio-frequency (RF) chain resulting in huge cost and power consumption [9]. Therefore, the use of mmWave in MIMO systems makes hybrid analog and digital precoding preferable [18, 49, 50], which is performed by cascading a digital precoder in the baseband and an analog network between the RF chains and antenna elements. Hence, beam gain and interference management can be achieved simultaneously.

The analog network typically consists of phase shifters with combiners [19] or switches [21]. In massive MIMO systems, the use of an enormous number of phase shifters causes considerable hardware complexity along with signal processing complexity and power consumption, while the use of switches results in a significant performance loss. Accordingly, a promising research line is introduced by utilizing advanced antenna designs, such as single-

lens antenna array (SLA) [11] and lens antenna subarray (LAS) [25, 26] to reduce signal processing complexity and RF chain cost without notable performance degradation. Due to the SLA-MIMO architecture's limitation including beamforming precoding/combining in a multipath channel and the large lens size that leads to high insertion loss and lack of scalability, the work in [26] presents an energy and spectral-efficient LAS-MIMO architecture. In the LAS-MIMO architecture shown in Fig. 3.1, each M antennas out of N antennas are connected to L small-sized lenses while the lenses are associated with a phase shifter network to control all the lenses together. For a specific lens, a simple switching network consisting of a single-pole multiple-throw (SPMT) switch controls the antenna elements. When the number of lenses is $L = 1$, the system falls to SLA-MIMO. On the other hand, the system performance is similar to the TA-MIMO when $L = N$. The LAS-MIMO provides better energy efficiency (EE) than the TA-MIMO with the expense of reducing the spectral efficiency as L decreases. Hence, an appropriate precoding design is essential to enhance the spectral efficiency.

Despite the attractive design of lens-aided MIMO systems, precoding design, and beam selection problems remain an open issue, especially in LAS-MIMO, and they are not yet investigated, to the best of our knowledge. Therefore, in this study, we take into account this problem. The contributions of this study are summarized as follows:

- We propose a hybrid precoding algorithm for the LAS-MIMO based on artificial bee colony (ABC) and orthogonal matching pursuit (OMP) algorithms. It solves a non-convex optimization problem iteratively by exploiting the sparse characteristics of the mmWave channel.

- The spectral efficiency and energy efficiency of the LAS architecture are investigated for a single user scenario with a single RF chain and multiple RF chains. As well, the energy efficiency performance is evaluated when different switch types (SP2T or SP4T) are utilized.

The following notations are used in this Chapter [4]: $\mathbf{A}$, $\mathbf{a}$, a denote a matrix, a vector, and a scalar variable, respectively. $\|\mathbf{A}\|_F$ denotes $\mathbf{A}$'s Frobenius norm. $\mathbf{A}^*$, $\mathbf{A}^T$, $\mathbf{A}^{-1}$ are $\mathbf{A}$'s conjugate, transpose, and inverse respectively. diag($\mathbf{a}$) is a diagonal matrix with $\mathbf{a}$ on its diagonal. $\mathbf{I}$ is the identity matrix, and $\mathbb{C}^{M\times N}$ denotes the space of $M \times N$ complex-valued matrices. $\mathcal{CN}(\mu, \sigma^2)$ is a complex Gaussian random vector with mean μ and covariance σ^2. j is the imaginary unit of complex numbers with $j^2 = -1$.

3.1 System Model

This section introduces the radio environment, spectral efficiency, and power consumption model of the LAS-MIMO architecture.

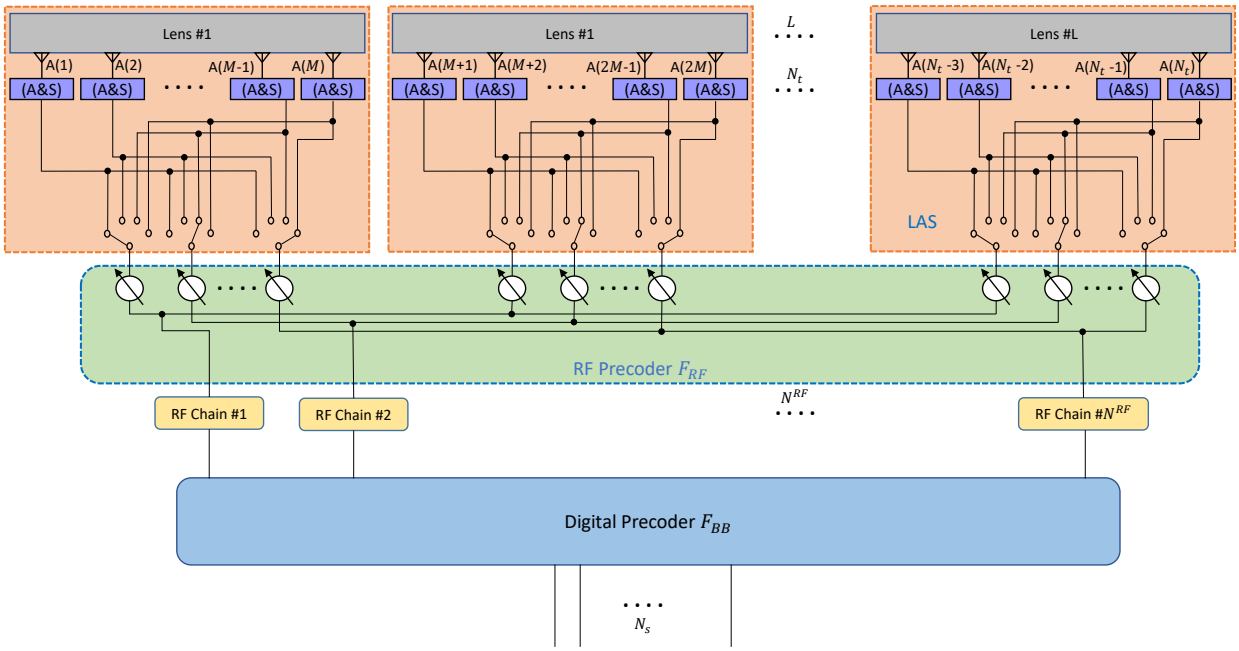


Figure 3.1: LAS-MIMO architecture with hybrid precoding.

3.1.1 Radio Environment and Parameters

We consider a single-user LAS-MIMO operating at mmWave frequencies where the transmitter employs N_t antennas connected to L_t lenses to transmit N_s data streams to a receiver

[4]Part of this chapter was published in [27]. Permission is included in Appendix A.

equipped with N_r phased antenna array. Multi-stream communication is enabled by employing the transmitter with N_t^{RF} RF chains where $N_s \leq N_t^{\mathrm{RF}} \leq N_t$. According to hybrid precoding presented in Fig. 3.1, the received signal is given as

$$\mathbf{r} = \sqrt{\rho}\mathbf{H}\mathbf{F}\mathbf{s} + \mathbf{n}, \tag{3.1}$$

where $\mathbf{r} \in \mathbb{C}^{N_r \times 1}$ is the received signal, ρ is the average received signal power, $\mathbf{H} \in \mathbb{C}^{N_r \times N_t}$ is the channel matrix, and $\mathbf{F} \in \mathbb{C}^{N_t \times N_s}$ is the precoder matrix. The transmitted data $\mathbf{s} \in \mathbb{C}^{N_s \times 1}$ has the normalized power of $\mathbb{E}[\mathbf{s}\mathbf{s}^*] = \mathbf{I}_{N_s}$. Additionally, the additive white Gaussian noise (AWGN) with zero mean and variance σ^2 is modeled as $\mathbf{n} \sim \mathcal{CN}(0, \sigma^2)$. The precoding matrix is expressed as $\mathbf{F} = \mathbf{F}_{\mathrm{Lens}}^{\mathbf{m}}\mathbf{F}_{\mathrm{RF}}\mathbf{F}_{\mathrm{BB}}$ where $\mathbf{F}_{\mathrm{Lens}}^{\mathbf{m}} \in \mathbb{C}^{N_t \times L_t}$ is the lens antenna effect for a given $\mathbf{m} \in \mathbb{C}^{1 \times L_t}$ vector containing the selected antenna indexes [26], $\mathbf{F}_{\mathrm{RF}} \in \mathbb{C}^{L_t \times N_t^{\mathrm{RF}}}$ is the analog beamformer obtained by the phase shifters, and $\mathbf{F}_{\mathrm{BB}} \in \mathbb{C}^{N_t^{\mathrm{RF}} \times N_s}$ is the digital baseband precoder where the total transmit power constraint is normalized such that $||\mathbf{F}_{\mathrm{Lens}}^{\mathbf{m}}\mathbf{F}_{\mathrm{RF}}\mathbf{F}_{\mathrm{BB}}||_F^2 = N_s$ [9]. Due to the subarray structure of the LAS-MIMO, no contribution among subarrays is obtained [26]. Therefore, each column of the $\mathbf{F}_{\mathrm{Lens}}^{\mathbf{m}}$ matrix contains zeros and $\mathbf{v}^{\mathbf{m}}$ vector representing the selected beam from each lens [26], and is given as

$$\mathbf{F}_{\mathrm{Lens}}^{\mathbf{m}} = \begin{bmatrix} \mathbf{v}^{\mathbf{m(1)}} & \mathbf{0}_{M\times 1} & \dots & \mathbf{0}_{M\times 1} \\ \mathbf{0}_{M\times 1} & \mathbf{v}^{\mathbf{m(2)}} & \mathbf{0}_{M\times 1} & \\ \vdots & & \ddots & \\ \mathbf{0}_{M\times 1} & \dots & \mathbf{0}_{M\times 1} & \mathbf{v}^{\mathbf{m(L_t)}} \end{bmatrix}_{N_t \times L_t}, \tag{3.2}$$

where M is the number of antennas placed under each lens and $\mathbf{v}^{\mathbf{m}} = [e^{-jkd\sin(\theta_m)\Omega}]_{\Omega \in \mathcal{I}(M)}$, where $\mathcal{I}(M) = \{q - (M-1)/2, q = 0, 1, \dots, M-1\}$, $k = \frac{2\pi}{\lambda}$, d denotes the antenna element spacing, λ is signal wavelength and $\theta_m = \frac{\pi}{4} - \frac{\pi(m-1)}{2(M-1)}$ is the metric showing the direction of the radiating beam from LAS for a given antenna element $m = 1, 2, ..., M$ chosen by the switch network [26].

We adopt a narrowband clustered channel model for the channel matrix $\mathbf{H}$, based on the Saleh-Valenzuela model, in order to capture the characteristics of the mmWave MIMO channel precisely [9, 51, 52, 53] which is expressed as [54]

$$\mathbf{H} = \sqrt{\frac{N_t N_r}{N_{cl} N_{ray}}} \sum_{i=1}^{N_{cl}} \sum_{k=1}^{N_{ray}} \gamma_{i,k} \mathbf{a_r}(\phi_{i,k}) \mathbf{a_t}^*(\theta_{i,k}), \tag{3.3}$$

where $\gamma_{i,k}$ stands for the complex gain of the k^{th} ray in the i^{th} scattering cluster. The angle of departure (AoD) and angle of arrival (AoA) for the k^{th} ray in the i^{th} scattering cluster are defined by $\theta_{i,k}$ and $\phi_{i,k}$, respectively. The received and transmitted array response vectors are represented by $\mathbf{a_r}(\phi_{i,k})$ and $\mathbf{a_t}(\theta_{i,k})$, respectively. Considering an N-element uniform linear array (ULA), the array response vector for a given $\psi \in \{\theta, \phi\}$ can be stated as $\mathbf{a}(\psi) = \frac{1}{\sqrt{N}} \left[1, e^{jkd\sin(\psi)}, \ldots, e^{j(N-1)kd\sin(\psi)}\right]^T$ [55].

3.1.2 Spectral Efficiency and Power Consumption of LAS Hybrid MIMO Architecture

In this study, we assume that the transmitter is equipped with LAS and the receiver is equipped with no LAS. Considering the ULA, the combiner can be stated as $\mathbf{W} = \mathbf{W}_{\text{RF}}\mathbf{W}_{\text{BB}}$ [56] where $\mathbf{W}_{\text{RF}} \in \mathbb{C}^{N_r \times N_r^{\text{RF}}}$ is the analog combiner obtained by the phase shifters, and $\mathbf{W}_{\text{BB}} \in \mathbb{C}^{N_r^{\text{RF}} \times N_s}$ is the digital baseband combiner. Additionally, the number of RF chains at the receiver is defined by N_r^{RF}.

Assuming that the base station can obtain perfect channel state information (CSI), the average spectral efficiency for the LAS-MIMO can be deriven from [9] and [26] and given as

$$R = \log_2 \left(\mathbf{I}_{N_s} + \frac{\rho}{N_s} \mathbf{R_n}^{-1} \mathbf{W}^* \mathbf{H} \mathbf{F} \mathbf{F}^* \mathbf{H}^* \mathbf{W} \right), \tag{3.4}$$

where $\mathbf{R_n} = \sigma^2 \mathbf{W}^* \mathbf{W}$ stands for the noise variance matrix.

The LAS-MIMO architecture aims to reduce power consumption and hardware complexity by connecting each RF chain to each lens through one phase shifter and one switch. In contrast, the hybrid TA-MIMO requires all the RF chains to be connected to each antenna

element through a phase shifter. Therefore, an accurate power consumption model for the LAS-MIMO transmitter is derived in [26] as $P_t^{LAS} = \frac{P_{T_x}^{LAS}}{(\eta_{PA}\eta_{SW})} + N_t^{\mathrm{RF}}(LP_{PS} + LN_{SW}P_{SW} + P_{RF})$, where $P_{T_x}^{LAS}$, P_{PS}, P_{SW}, and P_{RF} stand for the transmit power consumption in the LAS system, the power consumption of a phase shifter, a switch, and an RF chain, respectively. Additionally, η_{PA} and $\eta_{\mathrm{SW}} = 10^{-\zeta \mathrm{IL_{SW}}/10}$ stand for the efficiency of the transmitted amplifiers, and the efficiency of the switches, respectively. ζ is the number of series switches needed to be placed under each lens and $\mathrm{IL_{SW}}$ is the insertion loss for a switch.

Another performance metric that needs to be defined is energy efficiency which is defined as the number of bits that can be transmitted per unit of energy [57] and expressed as $EE = R/P_t^{LAS}$ (bps/Hz/W) for the LAS architecture.

3.2 Problem Formulation

For simplicity and tractability of the optimization problem, we consider only designing hybrid precoders $\mathbf{F}_{\mathrm{Lens}}^{\mathbf{m}}\mathbf{F}_{\mathrm{RF}}\mathbf{F}_{\mathrm{BB}}$ since joint optimization of hybrid precoders and combiners $\mathbf{W}_{\mathrm{RF}}\mathbf{W}_{\mathrm{BB}}$ are unlikely due to the non-convex constraints caused by phase shifters and switches [9]. Note that, the hybrid combiner design will be investigated as future work. Since our goal is designing only the hybrid precoders, the equation (3.4) needs to be rewritten as

$$R = \log_2\left(\mathbf{I}_{N_s} + \frac{\rho}{N_s\sigma^2}\mathbf{H}\mathbf{F}\mathbf{F}^*\mathbf{H}^*\right), \tag{3.5}$$

Then, the optimization problem is formulated as

$$(\mathbf{F}_{\mathrm{Lens}}^{\mathrm{opt}}, \mathbf{F}_{\mathrm{RF}}^{\mathrm{opt}}, \mathbf{F}_{\mathrm{BB}}^{\mathrm{opt}}) = \underset{\mathbf{F}_{\mathrm{Lens}}^{\mathbf{m}}, \mathbf{F}_{\mathrm{RF}}, \mathbf{F}_{\mathrm{BB}}}{\operatorname{argmax}} \; R, \tag{3.6a}$$

$$\text{s.t.} \quad \mathbf{F}_{\mathrm{Lens}}^{\mathbf{m}} \in \mathcal{F}_{\mathrm{Lens}}, \tag{3.6b}$$

$$\mathbf{F}_{\mathrm{RF}} \in \mathcal{F}_{\mathrm{RF}}, \tag{3.6c}$$

$$||\mathbf{F}_{\mathrm{Lens}}^{\mathbf{m}}\mathbf{F}_{\mathrm{RF}}\mathbf{F}_{\mathrm{BB}}||_F^2 = N_s. \tag{3.6d}$$

where $\mathcal{F}_{\text{Lens}}$ and $\mathcal{F}_{\text{RF}}$ stand for the set containing all the feasible lens antenna effects and the set of feasible RF precoders, respectively.

The problem given in (3.6) is a challenging optimization problem due to its non-convex amplitude constraints in (3.6b) and (3.6c). Although no optimal solution methodology exists for the problem (3.6) [9], an approximation is proposed in [9] to provide a near-optimal solution and proved that the maximization problem (3.6) is equivalent to the minimization problem of the distance between optimal unconstrained singular value decomposition (SVD) based precoder $\mathbf{F}_{\text{opt}}$ and practical hybrid precoder. Thus, the problem (3.6) can be rewritten as

$$(\mathbf{F}_{\text{Lens}}^{\text{opt}}, \mathbf{F}_{\text{RF}}^{\text{opt}}, \mathbf{F}_{\text{BB}}^{\text{opt}}) = \underset{\mathbf{F}_{\text{Lens}}^{\mathbf{m}}, \mathbf{F}_{\text{RF}}, \mathbf{F}_{\text{BB}}}{\text{argmin}} ||\mathbf{F}_{\text{opt}} - \mathbf{F}_{\text{A}}\mathbf{F}_{\text{BB}}||_F, \tag{3.7a}$$

$$\text{s.t.} \quad \text{(3.6b) to (3.6d)}, \tag{3.7b}$$

where $\mathbf{F}_{\text{A}} = \mathbf{F}_{\text{Lens}}^{\mathbf{m}}\mathbf{F}_{\text{RF}}$ stands for the total analog precoder which is the matrix multiplication of lens antenna effect and RF precoder. There is a relationship between the analog part of the precoder $\mathbf{F}_{\text{A}}$ and the transmit antenna array response vector $\mathbf{a_t}(\theta_{i,k})$ where the sparse-scattering structure of mmWave can be exploited to represent $\mathbf{F}_{\text{A}}$ as a function of $\mathbf{a_t}(\theta_{i,k})$ [9]. Considering that, equation (3.7) can be modified as

$$(\mathbf{F}_{\text{Lens}}^{\text{opt}}, \mathbf{F}_{\text{RF}}^{\text{opt}}, \mathbf{F}_{\text{BB}}^{\text{opt}}) = \underset{\mathbf{F}_{\text{Lens}}^{\mathbf{m}}, \mathbf{F}_{\text{RF}}, \mathbf{F}_{\text{BB}}}{\text{argmin}} ||\mathbf{F}_{\text{opt}} - \mathbf{F}_{\text{A}}\mathbf{F}_{\text{BB}}||_F, \tag{3.8a}$$

$$\text{s.t.} \quad {\mathbf{F}_{\text{A}}}^{(i)} \in \{\mathbf{a_t}(\theta_{i,k}), \forall i, k\}, \tag{3.8b}$$

$$\mathbf{F}_{\text{A}} = \mathbf{F}_{\text{Lens}}^{\mathbf{m}}\mathbf{F}_{\text{RF}}, \tag{3.8c}$$

$$||\mathbf{F}_{\text{A}}\mathbf{F}_{\text{BB}}||_F^2 = N_s. \tag{3.8d}$$

The precoding design for the LAS-MIMO requires a switch selection step to find the best beam selected from each lens since the lens antenna effect $\mathbf{F}_{\text{Lens}}^{\mathbf{m}}$ depends on the position of the activated switches. Therefore, the proposed precoding design first selects the $\mathbf{m}$ vector containing the selected beam indexes then find the $\mathbf{F}_{\text{RF}}$ and $\mathbf{F}_{\text{BB}}$ accordingly. This iterative process continues until finding the optimum precoders. Hence, $\mathbf{F}_{\text{Lens}}^{\mathbf{m}}$ can be omitted for a given $\mathbf{m}$ in each iteration, while $\mathbf{F}_{\text{A}}{}^{(i)}$ in (3.8b) can be embedded into the optimization problem due to the direct relationship between $\mathbf{F}_{\text{A}}{}^{(i)}$ and $\mathbf{a_t}(\theta_{i,k})$ [9]. Hence, the optimization problem becomes

$$(\tilde{\mathbf{F}}_{\text{RF}}^{\text{opt}}, \tilde{\mathbf{F}}_{\text{BB}}^{\text{opt}}) = \underset{\tilde{\mathbf{F}}_{\text{RF}}, \tilde{\mathbf{F}}_{\text{BB}}}{\text{argmin}} ||\mathbf{F}_{\text{opt}} - \mathbf{A_t}\tilde{\mathbf{F}}_{\text{BB}}||_F, \tag{3.9a}$$

$$\text{s.t.} \quad ||\text{diag}(\tilde{\mathbf{F}}_{\text{BB}}\tilde{\mathbf{F}}_{\text{BB}}^{*})||_0 = N_{\text{t}}^{\text{RF}}, \tag{3.9b}$$

$$\mathbf{F}_{\text{Lens}}^{\mathbf{m}}\tilde{\mathbf{F}}_{\text{RF}} = \mathbf{A_t}, \tag{3.9c}$$

$$||\mathbf{A_t}\tilde{\mathbf{F}}_{\text{BB}}||_F^2 = N_s, \tag{3.9d}$$

where $\mathbf{A_t} = [\mathbf{a_t}(\theta_{1,1}), \ldots, \mathbf{a_t}(\theta_{N_{cl},N_{ray}})] \in \mathbb{C}^{N_t \times N_{cl}N_{ray}}$ stands for the array response vector which is also the auxiliary variable obtained from $\mathbf{F}_{\text{A}}$ while the auxiliary variables for $\mathbf{F}_{\text{BB}}$ and $\mathbf{F}_{\text{RF}}$ are given as $\tilde{\mathbf{F}}_{\text{BB}}$ and $\tilde{\mathbf{F}}_{\text{RF}}$, respectively [9].

3.3 Solution of the Problem

This section proposes a hybrid beamforming algorithm to solve the NP-hard and non-convex problem (3.9). Consequently, we propose a swarm-based heuristic algorithm, namely ABC-aided spatially sparse precoding. It deploys both ABC and OMP by exploiting the sparse scattering characteristics of the mmWave channel. It is also possible to consider other existing swarm-based optimization tools (i.e., particle swarm optimization (PSO) [58] and ant colony optimization (ACO) [59]). However, they are likely to fall into a local minimum or

optimum solution region and be stuck there [60] if a problem has non-convexity properties. Therefore, ABC is more suitable since finding the global optimum solution rather than the local optimum solution is its strength [60, 61]. Another reason for selecting ABC is that it can be easily implemented in real-time applications due to its minimum parameter requirements for tuning and its fast convergence ability [62].

ABC is inspired by the food search behavior of the honey bees and proposed by Karaboga [63] in 2005. In a bee swarm, food sources define the possible solutions and the nectar amount of a food source represents the quality (fitness) of the food source. The number of food sources is equal to half of the population. The algorithm consists of four phases: initialization phase, employed bees phase, onlooker bees phase, and scout bees phase. The proposed solution of problem (3.9) is presented in Algorithm 1.

1) Initialization Phase: We randomly initialize the food sources (selected antenna indexes) $\mathbf{m}_i$'s such that $i \in \{1, \ldots, S\}$ where S is the population size. Since the antenna selection is an integer programming problem, the initial solutions have to be integer values where $1 \leq \mathbf{m}_i \leq M$ and can be produced by

$$\mathbf{m}_{ij} = round(\mathbf{m}_j^{\max} + rand(0,1) \times (\mathbf{m}_j^{\max} - \mathbf{m}_j^{\min})), \tag{3.10}$$

where $j \in \{1, 2, \ldots, L_t\}$, $\mathbf{m}_j^{\min} = 1$, and $\mathbf{m}_j^{\max} = M$. Then, the corresponding $\mathbf{F}_{\text{Lens}}^{\mathbf{m}_i}$ is found using (3.2). $\mathbf{F}_{\text{Lens}}^{\mathbf{m}_i}$, and randomly generated $\mathbf{F}_{\text{RF}}$ and $\mathbf{F}_{\text{BB}}$ are used to calculate the spectral efficiency using (3.5) to find the best solution $\mathbf{m}^{\mathbf{best}}$ providing the highest spectral efficiency at this time.

2) Employed Bee Phase: The bees look for new possible solutions providing better results than the results kept in their memory. The possible solutions in the neighborhood are given as

$$\mathbf{v}_{ij} = round(\mathbf{m}_{ij} + \alpha_{ij} \times (\mathbf{m}_{ij} - \mathbf{m}_{kj})), \tag{3.11}$$

where i and $k \in \{1, \ldots, S\}$ are randomly chosen indexes and $k \neq i$. $\alpha_{ij} \in [-1, 1]$ is a control parameter and responsible of keeping the newly produced solutions around $\mathbf{m}_{ij}$. After the search procedure is completed, $\mathbf{F}_{\text{Lens}}^{\mathbf{m}_i}$ is calculated using (3.2). Then, $\mathbf{F}_{\text{Lens}}^{\mathbf{m}_i}$ is sent to Algorithm 2 to calculate $\mathbf{F}_{\text{RF}}$ and $\mathbf{F}_{\text{BB}}$ guaranteeing the objective function given in (3.9). Accordingly, a greedy selection is applied between $\mathbf{v}_i$ and $\mathbf{m}_i$ using (3.5) to find the better solutions. After the selection, $\mathbf{m}_i$ is updated and the fitness function of using $\mathbf{m}_i$ are calculated as

$$\mathbf{F}_i = \begin{cases} \frac{1}{1+g(\mathbf{m}_i)}, & g\left(\mathbf{m}_i\right) \geq 0 \\ 1 + \text{abs}\left(g(\mathbf{m}_i)\right), & \text{otherwise} \end{cases}, \tag{3.12}$$

where $g(\mathbf{m}_i)$ is the objective function of the updated solution vector.

3) Onlooker Bee Phase: According to the solution vector and their fitness values shared by the employed bees, onlooker bees select their solution based on a probabilistic model which uses the fitness function given as $p_i = \mathbf{F}_i / \sum_i^S \mathbf{F}_i$. After new solutions are selected, the onlooker bees update their position using (3.11) and fitness function using (3.12) accordingly.

4) Scout Bee Phase: The bees replace the abandoned solutions, not improved for a particular number of trials, with new randomly generated possible solutions using (3.10). Bees memorize all these steps and share them, and the algorithm runs until it reaches the maximum number of iterations.

3.4 Simulation Results

This section illustrates the spectral efficiency and energy efficiency performance of the proposed algorithm for the hybrid LAS-MIMO in mmWave. The TA-MIMO and SLA-MIMO are chosen as the baseline architectures for fair performance comparison. The power consumption model is evaluated utilizing switch types of SP2T and SP4T to compare the energy efficiency performance. In the simulation, the channel parameters are set to $N_{cl} = 6$, $N_{ray} = 8$, $f_c = 38$ GHz, and 500 MHz bandwidth. Furthermore, AoAs and AoDs are uniformly distributed over $[-\frac{\pi}{4}, \frac{\pi}{4}]$ and $[-\frac{\pi}{2}, \frac{\pi}{2}]$, respectively [64]. It is assumed that we have

Table 3.1: Simulation Parameters

Parameters	**Value**
Power consumption of a phase shifter (P_{PS})	30 mW [21]
Power consumption of an SP2T switch (P_{SP2T})	10 mW [26]
Power consumption of an SP4T switch (P_{SP4T})	20 mW [26]
P_{RF}	220 mW [26]
η_{PA}	0.2 [26]
$\mathrm{IL}_{\mathrm{SW}}$	1 dB [26]

a downlink MIMO scenario where the precoding is designed using the proposed algorithm and the combiner is designed using the algorithm in [22], while the CSI is assumed to be perfectly known. The results are averaged over 500 channel realizations. The rest of simulation parameters are listed in Table 3.1.

Fig. 3.2 shows the spectral efficiency of 64×16 LAS-MIMO and TA-MIMO architectures with a single RF chain. The proposed algorithm provides better performance in the LAS-MIMO ($L = 4, 8, 16$) than in the SLA-MIMO ($L = 1$) as the number of lenses increases in the array. More precisely, the spectral efficiency is almost enhanced by 16%, 24% and 35% for $L = 4$, $L = 8$ and $L = 16$, respectively at $SNR = 5$ dB. In the simulation, we inspired and modified the spatially sparse precoding algorithm in [9], which performs very close to the optimal unconstrained SVD precoding, to present the simulation results for the TA-MIMO. The results show that TA-MIMO outperforms all LAS-MIMO scenarios due to its high precoding capability.

For the same system configurations, the energy efficiency analysis is shown in Fig. 3.3 where the proposed algorithm provides better performance in the LAS-MIMO than the TA-MIMO and SLA-MIMO as the number of lenses increases. Using SP4T switches in the switching network shows some enhancement in the performance rather than using SP2T switches. In particular, $L = 16$ outperforms all others while SLA ($L = 1$) has the worst performance among others when SP2T switch type is used to implement LAS-MIMO. On the other hand, $L = 4$ becomes the winner in energy efficiency due to the reduced number of switches and switch insertion loss when SP4T switch type is utilized. Although we obtain

a high precoding gain with the ABC-OMP algorithm, it is essential to note that the LAS requires a careful design to make a fair decision between spectral efficiency and energy efficiency trade-off.

Multi-RF chain scenarios are evaluated in Fig. 3.4 and Fig. 3.5 showing spectral efficiency and energy efficiency, respectively. The proposed algorithm provides satisfactory results even with $N_t = 64$, $N_r = 16$, and $N_t^{\mathrm{RF}} = 8$ scenario. Since the precoding capability increases when the number of lenses increases, $L = 16$ still outperforms all the other LAS architectures with 32% more spectral efficiency at $SNR = 5$ dB than SLA. On the other hand, Fig. 3.5 shows that LAS-MIMO with $L = 4$ provides the best performance in energy efficiency when SP2T switches are utilized. However, the performance can further be enhanced with SP4T switches. Specifically, using SP4T switch instead of SP2T switch provides 9% more energy efficiency at $SNR = 5$ dB. For $L = 16$ and $L = 8$ architectures, the reason for having worse performance than the SLA architecture is that the number of required phase shifters and switches increases as N_t^{RF} increases, causing higher power consumption.

The spectral efficiency and energy efficiency analysis of the proposed algorithm indicates that the LAS-MIMO with $L = 16$ and SP2T switches provide the optimum performance compared to other systems (SLA and TA-MIMO) when ${N_t}^{\mathrm{RF}} = 1$. In multi-RF scenarios, LAS-MIMO offers the best energy efficiency and spectral efficiency when $L = 4$, and $L = 16$, respectively. Thus, one can design the system depending on the energy efficiency and spectral efficiency trade-off to provide optimum performance based on the system requirements.

Finally, the convergence property of the proposed algorithm is presented in Fig. 3.6 when $SNR = 5$ dB. The algorithm is run for 100 iterations for the same channel configuration, and the convergence rate is averaged. We can see that it quickly converges after almost 21 iterations when $L = 16$ and 27 iterations when $L = 8$.

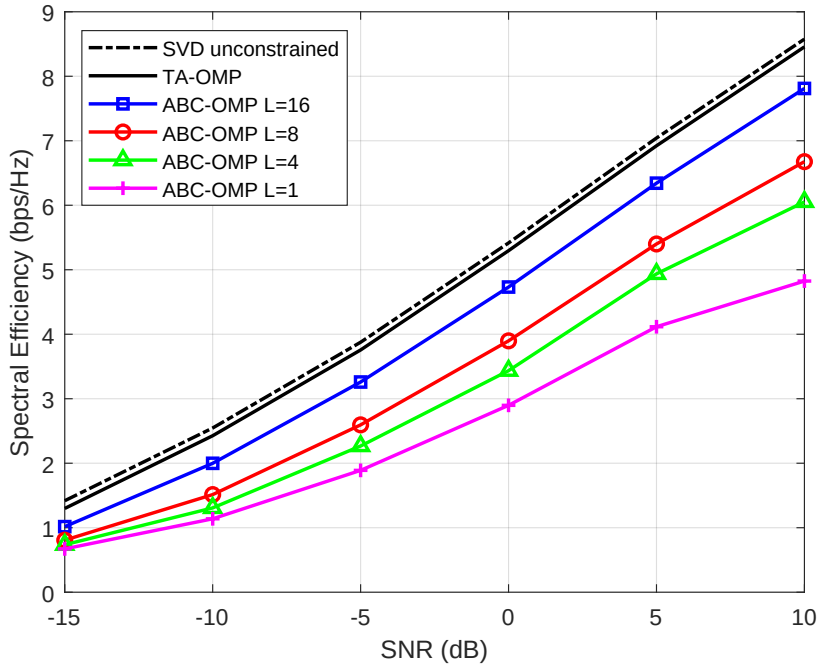


Figure 3.2: Spectral efficiency vs SNR for a single RF chain.

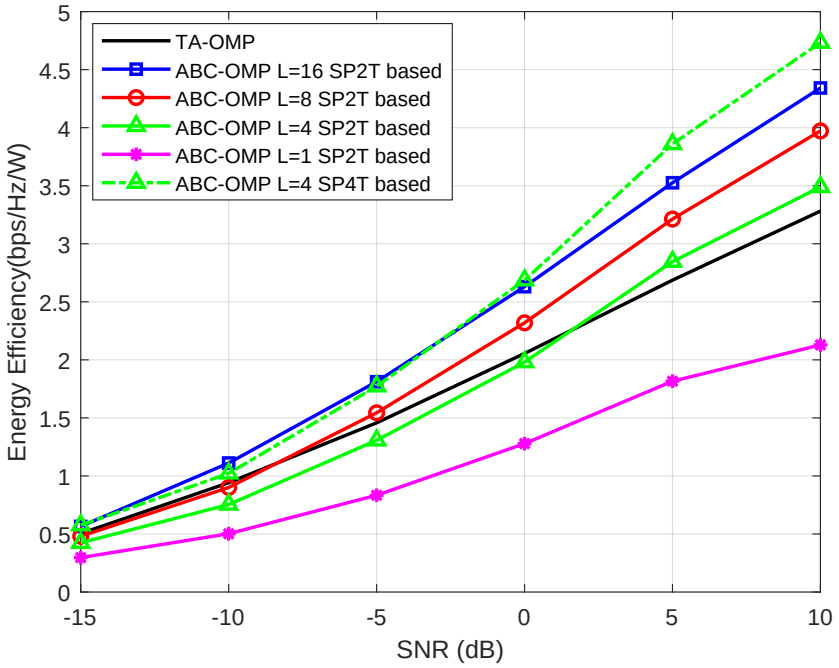


Figure 3.3: Energy efficiency vs SNR for a single RF chain.

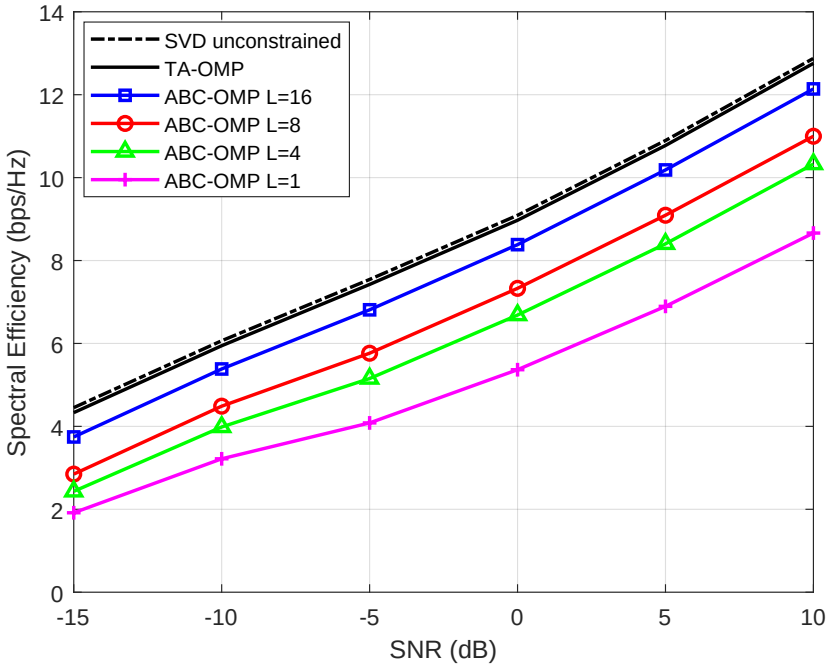


Figure 3.4: Spectral efficiency vs SNR for multiple RF chains.

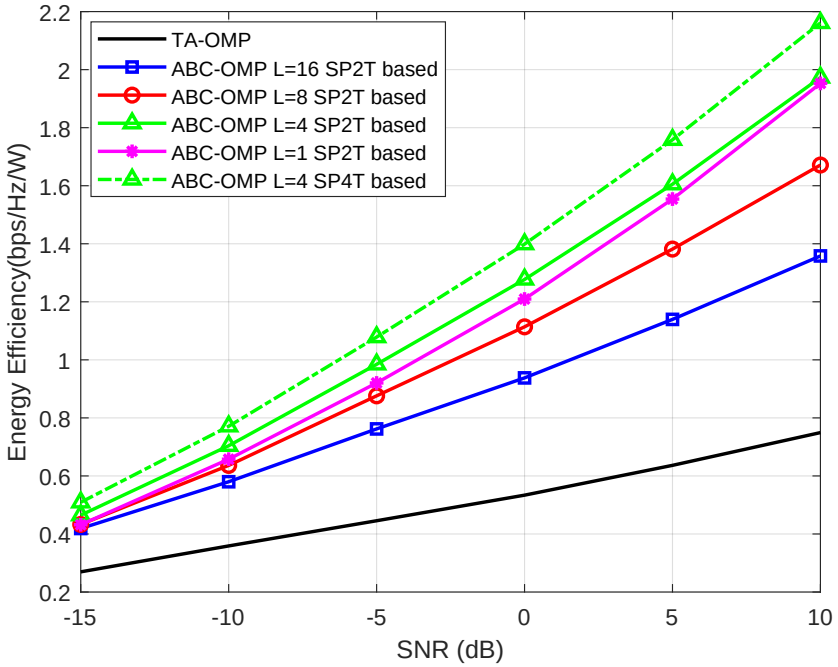


Figure 3.5: Energy efficiency vs SNR for multiple RF chains.

Algorithm 3.1 ABC Aided Hybrid Sparse Precoding

Input: $\mathbf{F}_{\text{opt}}$, and $\mathbf{A_t}$
ABC parameters: $S = 100$, $i_{max} = 500$
Output: $\mathbf{F}^{\mathbf{m}}_{\text{Lens}}$, $\mathbf{F}_{\text{RF}}$, $\mathbf{F}_{\text{BB}}$
Randomly generate S solutions $\mathbf{m}_i$ using (3.10),
Calculate $\mathbf{F}^{\mathbf{m}_i}_{\text{Lens}}$ using (3.2),
Randomly generate $\mathbf{F}_{\text{RF}}$ and $\mathbf{F}_{\text{BB}}$,
Evaluate the function using (3.5), and select $\mathbf{m}^{\mathbf{best}}$,
while $i < i_{max}$ **do**
    Phase-1: Employed Bee Phase
    for $s = 1 : S$ **do**
        Produce a new solutions $\mathbf{v}_i$ using (3.11) and calculate corresponding $\mathbf{F}^{\mathbf{v}_i}_{\text{Lens}}$ as in (3.2),
        Calculate $\mathbf{F}_{\text{RF}}$ and $\mathbf{F}_{\text{BB}}$ using **Algorithm 2**,
        Evaluate the function using (3.5), and apply greedy selection between $\mathbf{v}_i$ and $\mathbf{m}_i$,
        Update $\mathbf{m}_i$ and find fitness function using (3.12),
    end
    Phase-2: Onlooker Bee Phase
    for $s = 1 : S$ **do**
        Find the selection probabilities p_i,
        Use p_i to generate new solutions $\mathbf{v}_i$ from $\mathbf{m}_i$
        Select a food source vector $\mathbf{m}^{\text{curr}}$ according to p_s value,
        Follow same steps from step-9 to step-11.
    end
    Phase-3: Scout Bee Phase
    for $s = 1 : S$ **do**
        Identify the abandoned solutions not improved after a predetermined number of trials,
        Replace them with new randomly generated solutions using (3.10),
        Store the best solution ever found,
    end
end
return $\mathbf{F}^{\mathbf{m}}_{\text{Lens}}$, $\mathbf{F}_{\text{RF}}$, $\mathbf{F}_{\text{BB}}$.

Algorithm 3.2 Spatially Sparse Precoding

Input: $\mathbf{A_t}$, $\mathbf{F}_{\text{opt}}$, and $\mathbf{F}^{\mathbf{m}}_{\text{Lens}}$

Output: $\mathbf{F}_{\text{RF}}$, and $\mathbf{F}_{\text{BB}}$

$\mathbf{F}_{\text{A}}$ = Empty Matrix, and $\mathbf{F_{res}} = \mathbf{F}_{\text{opt}}$,

for $i = 1 : N_t^{RF}$ **do**

$\mathbf{\Psi} = \mathbf{A_t}^{*}\mathbf{F_{res}}$,

$k = \text{argmax}_{l=1,\ldots,N_{cl}N_{ray}}(\mathbf{\Psi}\mathbf{\Psi}^{*})_{l,l}$,

$\mathbf{F}_{\text{A}} = [\mathbf{F}_{\text{A}}|\mathbf{A}_{\mathbf{t}}^{(\mathbf{k})}]$,

$\mathbf{F}_{\text{RF}} = (\mathbf{F}^{\mathbf{m}}_{\text{Lens}}{}^{*}\mathbf{F}^{\mathbf{m}}_{\text{Lens}})^{-1}\mathbf{F}^{\mathbf{m}}_{\text{Lens}}{}^{*}\mathbf{F}_{\text{A}}$,

$\mathbf{F}_{\text{BB}} = (\mathbf{F}_{\text{A}}{}^{*}\mathbf{F}_{\text{A}})^{-1}\mathbf{F}_{\text{A}}{}^{*}\mathbf{F}_{\text{opt}}$,

$\mathbf{F_{res}} = \frac{\mathbf{F}_{\text{opt}}-\mathbf{F}^{\mathbf{m}}_{\text{Lens}}\mathbf{F}_{\text{RF}}\mathbf{F}_{\text{BB}}}{||\mathbf{F}_{\text{opt}}-\mathbf{F}^{\mathbf{m}}_{\text{Lens}}\mathbf{F}_{\text{RF}}\mathbf{F}_{\text{BB}}||_F}$,

end

$\mathbf{F}_{\text{BB}} = \sqrt{N_s}\frac{\mathbf{F}_{\text{BB}}}{||\mathbf{F}^{\mathbf{m}}_{\text{Lens}}\mathbf{F}_{\text{RF}}\mathbf{F}_{\text{BB}}||_F}$,

return $\mathbf{F}_{\text{RF}}$, $\mathbf{F}_{\text{BB}}$

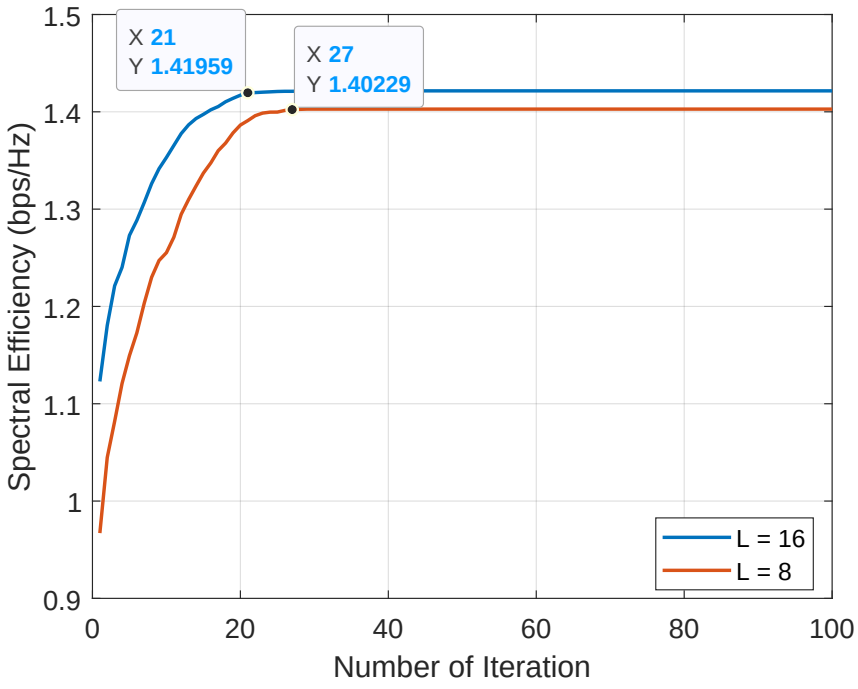


Figure 3.6: Convergence rate of ABC-OMP algorithm.

Chapter 4: Energy and Spectral Efficiency Tradeoff in NOMA: Multi-Objective Evolutionary Approaches

Next-generation wireless networks are presumed to meet the demand of users for gigabit data rates. To this end, non-orthogonal multiple access (NOMA) was proposed [65] as a candidate radio access technology to deal with this ambitious high data rate requirement. Since it enhances the spectral efficiency by multiplexing multiple users with distinct channel conditions at the same time and frequency resource with different power levels and it applies successive interference cancellation (SIC) at the receiver side in order to mitigate multi-user interference [65].

The exponential growth in traffic volume along with high data rate requirement leads to electromagnetic pollution due to high energy consumption by communication technology infrastructures [66]. Although high spectral efficiency can be achieved by NOMA, more power allocation to the user with poor channel gain creates an energy inefficiency [67]. Therefore, there is a need to study the spectral and energy efficiency tradeoff in NOMA.

To date, many outstanding contributions for power allocation to maximize the spectral efficiency [68, 69, 70] have been proposed for NOMA systems. Furthermore, many solution methodologies [71, 72, 73, 74], was developed for the energy efficiency maximization problem in NOMA. However, all of them focus on either maximizing the spectral efficiency or energy efficiency and do not consider the spectral and energy efficiency tradeoff.

In the literature, there are a limited number of proposals [66], [67], [75], and [76] have been investigated for the spectral and energy efficiency tradeoff in NOMA systems. In [66], the authors developed an improved particle swarm optimization (PSO) method having faster convergence than traditional PSO to optimize the spectral and energy efficiency tradeoff in

a downlink NOMA system where quasi-concave energy efficiency function was maximized with the consideration of quality of service (QoS) requirements. Reference [67] considered a multi-objective optimization problem (MOP) for the spectral and energy efficiency tradeoff in a downlink NOMA system. The authors tried to maximize the spectral efficiency and minimize the total power consumption simultaneously by using a weighted sum approach having the flexibility to adjust the weight according to the need. In [75], the authors proposed a single-objective optimization problem (SOP) in a downlink NOMA system where the energy efficiency was maximized under the minimum QoS constraints. Finally, a power allocation problem algorithm was developed by [76] to optimize the spectral and energy efficiency tradeoff with fairness consideration in downlink NOMA systems.The MOP was converted into a SOP through a weighted sum approach and solved by the Lagrangian method.

Contrary to the aforementioned proposals for the spectral and energy efficiency tradeoff optimization, to the best of our knowledge, this study is the first attempt to solve the multi-objective spectral and energy efficiency optimization problem in NOMA systems and find the Pareto optimal solution set through multi-objective evolutionary algorithms (MOEAs). The contributions of this study presented in this Chapter [5] can be represented as follows:

- In this study, we investigate the spectral and energy efficiency tradeoff problem in a downlink NOMA. We consider a MOP approach where multiple objectives are solved simultaneously to obtain a Pareto optimal set. Having a set of solutions provides an advantage since one possible solution can be chosen freely from the set freely according to the priority of the energy efficiency or spectral efficiency while a single solution is offered in [67], and [76] according to the predefined weight.

- We first formulate the problem to maximize both the energy efficiency and spectral efficiency and then create the second problem to maximize spectral efficiency and minimize power consumption. Both problems consider the same set of constraints, i.e., QoS requirements, power allocation coefficients, and power threshold for perfect SIC.

[5] Part of this chapter was published in [35]. Permission is included in Appendix A.

Additionally, the problems with constraints are first converted into an unconstrained problem with penalty function method [77] and then solved through MOEAs.

- The performance of three commonly known MOEAs (non-dominated sorting genetic algorithm-II (NSGA-II), multi-objective particle swarm optimization (MOPSO), and strength Pareto evolutionary algorithm-2 (SPEA2)) are compared in terms of Pareto solution set quality. The comparison of the Pareto front quality is performed by a metric called hypervolume indicator as in [78]. The results indicate that SPEA2 outperforms all others in terms of Pareto front quality. However, SPEA2 does not perform well in terms of computational time while MOPSO is the best.

4.1 System Model and Problem Formulation

4.1.1 Radio Environment and Parameters

In this study, we consider a downlink NOMA network with a single base station (BS) having a single antenna element located at the center of a hexagon. The notation $k \in \mathbb{K} = \{1, 2, \ldots.., 2K\}$ is used to index randomly distributed users with a single antenna element in the coverage area. The total bandwidth of B is equally shared among M pairs where each pair consists of two users with the bandwidth of B_s for SIC simplicity. The paired users share the same resource block (RB) while each pair is orthogonal to each other. For each pair, a user with high channel gain is treated as a strong user denoted by U_s while the weak user denoted by U_w is the one who suffers from poorer channel gain. We borrow the suboptimal user pairing algorithm proposed in [79] where the user with the highest channel gain and the user with the poorest channel gain creates one NOMA pair, while the second NOMA pair consists of having the user with the second-highest channel gain and the user with the second-poorest channel gain, and so on. Since this methodology enhances the throughput of the users with poor channel gain. In addition, fairness is assumed for all users by keeping their minimum data rate requirements over a threshold defined by R_{min}.

According to the main principle of NOMA, it is assumed that the BS with perfect knowledge of channel state information (CSI) broadcasts the superposition coded symbols for all users. For each pair, s_w^m and s_s^m denote modulation symbols of U_w and U_s, respectively and the superposition coded symbol for a given pair $m \in \mathbb{M} = \{1, 2, \ldots.., K\}$ is given as [76]

$$x = \sqrt{\alpha_s^m P_m} s_s^m + \sqrt{(1 - \alpha_s^m) P_m} s_w^m, \tag{4.1}$$

where the notation $0 \leq \alpha_s^m \leq 0.5$ expresses the power allocation coefficient of U_s in pair m, which guarantees that more power is allocated to U_w to assure a reasonable signal-to-interference-plus-noise ratio (SINR). P_m is the total transmit power allocated to pair m.

It is assumed that the channel model considers the effect of both large-scale fading denoted by g_k and small-scale fading (Rayleigh) denoted by β_k. The large-scale fading consists of both pathloss and shadowing between the BS and user k. We assume that g_k is constant across all RBs. However, β_k changes across different slots but they are constant within a slot. The channel coefficient h_k is defined as a function of both large-scale fading and Rayleigh fading and given as $h_k = \beta_k . {g_k}^{-1}$. The received signal at an arbitrary user k is expressed as

$$y_k = h_k x + n_k, \tag{4.2}$$

where the additive white Gaussian noise (AWGN) with zero mean and variance σ^2 is modeled as $n_k \sim \mathcal{CN}(\mu, \sigma^2)$.

Without loss of generality, the channel gain of users are sorted as $|h_1|^2 \geq |h_2|^2 \geq \ldots \geq |h_{2K}|^2$. In addition, the condition that needs to be fulfilled is expressed as $|h_s^m|^2 \geq |h_w^m|^2$ for a given pair m. In a given pair m, s_w^m is decodable due to its higher power level. Thus, s_w^m is first decoded and removed from s_s^m by U_s, then U_s decodes s_s^m. Therefore, U_s is able to receive data without inter-user interference with the assumption of perfect SIC. However, U_w is not able to decode s_s^m and remove it from s_w^m due to low power assigned to U_s. For this reason, U_w treats s_s^m as noise and decodes its own signal s_w^m. In other words, U_w does

not perform SIC. For a given pair m, users' achievable rates are given as

$$R_s^m = B_s \log\left(1 + \frac{\alpha_s^m h_s^m P_m}{N_0}\right), \tag{4.3}$$

$$R_w^m = B_s \log\left(1 + \frac{(1-\alpha_s^m) h_w^m P_m}{N_0 + \alpha_s^m h_w^m P_m}\right). \tag{4.4}$$

In (4.3) and (4.4), N_0 is the system noise power and given by $N_0 = N_F k_b T_K B_s$, where N_F, k_b, T_k are noise figure, Boltzmann's constant, and temperature in degree Kelvin, respectively.

4.1.2 Spectral Efficiency and Energy Efficiency

As opposed to most of the existing works related to spectral and energy efficiency tradeoff for NOMA [66], [76], and [75] where energy efficiency metric is defined as a function of bandwidth and it is measured in bit/Joule/Hz instead of bit/Joule, we define energy efficiency metric as the number of bits that can be transmitted per unit of energy in order to avoid misleading. In other words, we define energy efficiency as the sum throughput divided by consumed power as in [57]. Measuring energy efficiency in bit/Joule/Hz, where the transmit power is bandwidth-normalized while the noise power is proportionally dependent to the bandwidth, is meaningless [57]. The energy efficiency and spectral efficiency metrics are given respectively as

$$f_{EE} = \frac{\sum_{m=1}^{M} (R_w^m + R_s^m)}{f_{PC}}, \tag{4.5}$$

$$f_{SE} = \frac{\sum_{k=1}^{2K} R_k}{B}. \tag{4.6}$$

In (4.5), $f_{PC} = P_c + \sum_{m=1}^{M} P_m$ is the power consumption metric where P_c represents the constant circuit power consumption and $M = K$.

4.1.3 Problem Formulation

This subsection focuses on the formulation of the MOP that aims to optimize spectral efficiency and energy efficiency tradeoff subject to some constraints given as follows:

$$\sum_{m=1}^{M} P_m \leq P_t, \qquad \forall m \in \mathbb{M}, \tag{4.7a}$$

$$R_k \geq R_{min}, \qquad \forall k \in \mathbb{K}, \tag{4.7b}$$

$$P_m \gamma_s^m (1 - 2\alpha_s^m) \geq P_{Thr}, \qquad \forall m \in \mathbb{M}, \tag{4.7c}$$

$$0 < \alpha_s^m \leq 0.5, \qquad \forall m \in \mathbb{M}. \tag{4.7d}$$

Accordingly, (4.7a) limits the maximum transmit power and guarantees that it cannot exceed P_t. Constraint (4.7b) assures that each user will be able to achieve at least the minimum required data rate where R_k stands for the achievable rate for an arbitrary user k . Moreover, (4.10c) indicates that the perfect SIC can be obtained if there is enough power difference between users in each pair [79] and α_s^m in (4.7d) shows that more power is allocated to the user with poor channel gain for a proper SINR in a given pair m. Additionally, γ_s^m in (4.10c) denotes the bandwidth-normalized channel gain of the U_s and is expressed by $\gamma_s^m = h_s^m / B_s$ and P_{Thr} stands for the minimum power difference for SIC.

We consider different MOPs in order to optimize the spectral and energy efficiency tradeoff. Thus, different objectives namely, spectral efficiency, energy efficiency, and power consumption, are taken into account for the problem formulation.

1) Optimization Problem 1: We create the first problem to maximize two conflicting objectives, namely energy efficiency and spectral efficiency, in order to exemplify the tradeoff. As opposed to spectral efficiency maximization with the requirement of excess power consumption, taking both spectral efficiency and energy efficiency into consideration in the formulation assures that no excess power consumption required to optimize the tradeoff [80].

The problem is notated as SE-EE and given as:

$$\begin{aligned} &\max_{\boldsymbol{\alpha_s^m}, \boldsymbol{P_m}} \ \boldsymbol{f_1} = [f_{EE}, f_{SE}], \\ &\textbf{s.t.} \ 4.7\text{a to } 4.7\text{d}, \end{aligned} \tag{4.8}$$

where $\boldsymbol{f_1}$ is a vector containing objective functions. Additionally, $\boldsymbol{\alpha_s^m}$ and $\boldsymbol{P_m}$ are also vectors containing power allocation coefficients of U_s's for all pairs and power allocated to all pairs, respectively.

2) Optimization Problem 2: The second problem aims to provide users maximum throughput while consuming minimum power. Thus, it maximizes spectral efficiency but minimizes power consumption. The problem is named as SE-PC and given as:

$$\begin{aligned} &\max_{\boldsymbol{\alpha_s^m}, \boldsymbol{P_m}} \ \boldsymbol{f_2} = [f_{SE}, -f_{PC}], \\ &\textbf{s.t.} \ 4.7\text{a to } 4.7\text{d}, \end{aligned} \tag{4.9}$$

where $\boldsymbol{f_2}$ is a vector having the objective functions regarding the second optimization problem.

4.2 Solution Methodologies: Evolutionary Algorithms

Optimization of MOPs is an intrinsically very challenging process due to the simultaneous optimization of multiple conflicting objectives [81]. Therefore, evolutionary algorithms are well-suited to find a set of solutions, called Pareto optimal solution, between conflicting objectives. For this reason, we first give the general formulation of MOP with inequality

constraints as [77]:

$$\textbf{min} \quad f(\mathbf{X}) = [f_1(\mathbf{X}), f_2(\mathbf{X}), \ldots., f_n(\mathbf{X})], \tag{4.10a}$$

$$\textbf{s.t.} \quad g_j(\mathbf{X}) \leq 0, \ j = 1, 2, \ldots., N, \tag{4.10b}$$

$$h_l(\mathbf{X}) = 0, \quad l = 1, 2, \ldots., L, \tag{4.10c}$$

where n, N, and L stand for the number of objective functions, the number of inequality constraints, and the number of equality constraints, respectively.

For the solution of such a problem, we adopt the penalty function method from [77] to transform a constrained problem into an unconstrained problem as:

$$\textbf{min } F(\mathbf{X}) = f(\mathbf{X}) + \tau_l \sum_{l=1}^{L} h_l^2(\mathbf{X}) + \tau_g \sum_{j=1}^{N} \psi_j(g_j(\mathbf{X})), \tag{4.11}$$

where

$$\psi_j(g_j(\mathbf{X})) = \begin{cases} 0, & \text{if } g_j(\mathbf{X}) \leq 0 \\ \exp(g_j(\mathbf{X})), & \text{(otherwise)} \end{cases}. \tag{4.12}$$

Additionally, τ_g and τ_l are positive penalty constants for inequality and equality constraints, respectively. Note that the problem (4.8) and (4.9) do not consider any equality constraints. Thus, we can first eliminate the equality constraint terms from (4.11) and then define a general formulation for (4.8) and (4.9) as:

$$\textbf{min } \boldsymbol{F_i} = -\boldsymbol{f_i} + \tau_g \sum_{j=1}^{4} \psi_j(g_j(\mathbf{X})), \quad i = 1, 2. \tag{4.13}$$

The following subsection introduces the MOEAs for the solution of the spectral and energy efficiency tradeoff problems given in (4.13).

4.2.1 Non-dominated Sorting Genetic Algorithm-II (NSGA-II)

NSGA-II is a population-based evolutionary algorithm, which provides a Pareto optimal set of solutions, based on a non-dominated sorting technique, and genetic operations like selection, crossover, and mutation [82].

The adaptation process to the environment of each individual in a population creates the basic idea behind NSGA-II. Each individual obtains the value of objective functions, which are (4.8), and (4.9) in our case, to be able to perform adaptation and pass to the next generation. The possible set of solutions for the spectral and energy efficiency tradeoff in each generation is denoted by multiple chromosomes consisting of multiple genes representing the possible decision variables.

NSGA-II is one of the best evolutionary algorithms to deal with MOPs due to its faster sorting process, elitism, utilization of less number of parameters, and preservation of diversity [82]. The working flow of NSGA-II is derived from [82] and [83] and given in Algorithm 4.1[6].

4.2.2 Strength Pareto Evolutionary Algorithm-2 (SPEA2)

SPEA is the first version of SPEA2 proposed by Zitzler and Thiele [84] as a population-based MOEA. The main idea behind the algorithm is that the non-dominated solutions obtained in each iteration are stored in an external archive having not a fixed size along with its own population in order to let all individuals in the population observe their scaler fitness value with Pareto dominance method. Additionally, the clustering technique is used in SPEA to limit the number of individuals stored in the external archive. This varying archive size was the limitation faced in SPEA and was overcome with SPEA2 [85] by keeping the external size fixed. Moreover, it utilizes a method called truncation to remove dominated solutions from the archive. More accuracy is provided by a nearest neighbor density estima-

[6] Some concepts and the formulas required for the calculation process regarding the MOEAs are omitted and can be found in the references provided.

tion methodology. Algorithm 4.2[6] represents the working flow of the SPEA2 derived from [85].

Algorithm 4.1 NSGA-II for the SE-EE Optimization

Input:
Design variables: $\boldsymbol{\alpha_s^m}$ and $\boldsymbol{P_m}$,
NSGA-II parameters: S, p_c, p_m, μ, and i_{max}.
Output:
Solution: Particle position $\boldsymbol{X} = [\boldsymbol{\alpha_s^m}, \boldsymbol{P_m}]$.
Set all input parameters except $\boldsymbol{X}$,
Set initial non-dominated population **Pop** $= \emptyset$,
for $s = 1 : S$ **do**
  Initialize $\boldsymbol{X}$ randomly for each individual in **Pop**,
  Evaluate the initial fitness value of $\boldsymbol{X}$ as in (4.13),
end
for $i = 1 : i_{max}$ **do**
  Apply crossover to **Pop** and obtain **Popc**,
  Apply mutation to **Pop** and obtain **Popm**,
  Set **Pop** $\leftarrow$ **Pop** $\cup$ **Popc** $\cup$ **Popm** to maintain elitism,
  Apply non-dominated sorting to **Pop** to find the non-dominated fronts **R**,
  Compute crowding distance of **Pop** in all **R**,
  Sort **Pop** in ascending order and update **R** accordingly,
  Set the next generation parents **Pop** $\leftarrow$ **Pop(R(1))**,
end

4.2.3 Multi-Objective Particle Swarm Optimization (MOPSO)

The social behavior of a swarm creates the fundamental principles of PSO proposed by Kennedy and Eberhart [58] where each particle in a swarm is defined by its velocity and position. The search process is performed by each particle in the swarm. During the search process, each particle has its own memory to remember the best position and share the information discovered with other particles in the swarm to update their position and velocity accordingly. This process continues until there is no change in the position of all the particles. MOPSO is an extended version of PSO to solve MOPs simultaneously [86]. The non-dominated solution set is stored in a population set according to the Pareto preference

Algorithm 4.2 SPEA2 for the SE-EE Optimization

Input:
Design variables: $\boldsymbol{\alpha_s^m}$ and $\boldsymbol{P_m}$,
SPEA2 parameters: S, S_r, p_c, p_m, and i_{max}.
Output:
Solution: $\boldsymbol{X} = [\boldsymbol{\alpha_s^m}, \boldsymbol{P_m}]$.
Set all input parameters except $\boldsymbol{X}$,
Initialize external archive $\boldsymbol{EA} = \emptyset$,
for $i = 1 : S$ **do**
  Generate initial $\boldsymbol{X}$ randomly,
  Evaluate the initial fitness values for $\boldsymbol{X}$ as in (4.13),
end
while $i < i_{max}$ **do**
  Evaluate the fitness function with $\boldsymbol{X}$ for each individual in **Pop** and $\boldsymbol{EA}$,
  Copy non-dominated individuals from **Pop** and $\boldsymbol{EA}$ to $\boldsymbol{EA}$,
  if $lenght(\boldsymbol{EA}) \geq S_r$ **then**
    Apply truncation to $\boldsymbol{EA}$ to remove elements,
  else
    Send non-dominated individuals from **Pop** to $\boldsymbol{EA}$,
  end
  Apply binary tournament selection to select individuals of lower front,
  Perform crossover and mutation to create next generation.
end

which decides the domination of the possible solution sets. The working flow of the MOPSO derived from [86] is presented in Algorithm 4.3[6].

4.3 Simulation Results

In this section, the performance of NSGA-II, MOPSO, and SPEA2 are evaluated for spectral and energy efficiency tradeoff in a downlink NOMA scenario consisting of a single BS located at the center of the hexagon area. The simulation is first run with 8 users and then the number of users is set to 24 to show which algorithm is the winner among others to handle more design variables. The parameters for MOEAs are given in Table 4.1.

To guarantee the perfect SIC, we introduce the constraint (4.10c) which assures that there is enough power difference between users in a given pair to let U_s perfectly decode the signal

Table 4.1: Parameters for MOEAs

Parameters	**NSGA-II**	**MOPSO**	**SPEA2**
Max iteration (i_{max})	200	200	200
Population size (S)	100	100	100
Archive size (S_r)	-	100	100
Personal learning coeff. (l_p)	-	1.25	-
Global learning coeff. (l_g)	-	2.25	-
Inertia weight (w)	-	0.6	-
Mutation rate (μ)	0.02	0.1	-
Mutation percentage (p_m)	0.4	-	0.7
Crossover percentage (p_c)	0.7	-	0.3

of U_w. Additionally, assuming the minimum data rate requirements for all users provides the QoS of all users. The simulation parameters and the pathloss model are borrowed from 3GPP standards [87] and represented in Table 4.2.

In Fig. 4.1, we compare the performance of NSGA-II, MOPSO, and SPEA2 for the spectral and energy efficiency tradeoff in a downlink NOMA scenario by maximizing both energy efficiency and spectral efficiency as in (4.8). The number of users is set to 8 during the simulation and Pareto optimal set (non-dominated solutions) is obtained. It is seen that SEPA2 dominates NSGA-II and MOPSO while NSGA-II has the worst performance among all MOEAs. Although all MOEAs can visualize Pareto optimal solution sets for the spectral and energy efficiency tradeoff in Fig. 4.1, NSGA-II provides a broader range of non-dominated solutions (wider decision space for the tradeoff) than others.

Fig. 4.2 shows the Pareto optimal solution for the problem (4.9) where SPEA2 outperforms the other algorithms. However, this statement cannot be seen clearly in Fig. 4.2 since SPEA2 and MOPSO look similar. In such cases, we use the hypervolume metric **H**, which stands for the volume of the objective space having non-dominated solutions[7], to compare the MOEAs [78]. The quality of Pareto optimal set for all MOEAs is evaluated by the hypervolume indicator and presented in Table 4.3. According to Table 4.3, the best performance is observed by SPEA2 having the smallest hypervolume indicator value and the second-smallest

[7]The details of **H** are omitted and can be calculated as in [78].

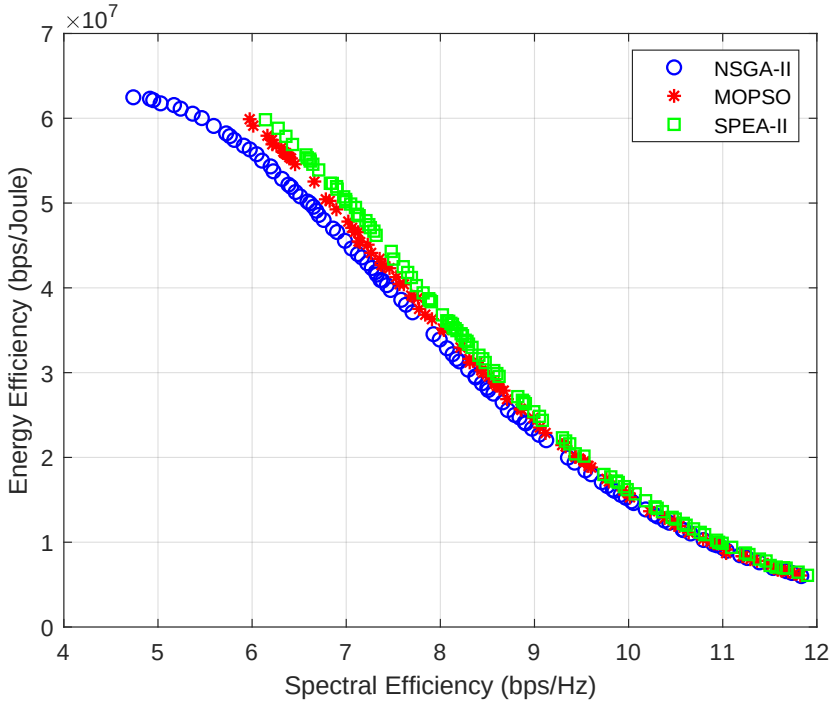


Figure 4.1: Pareto optimal set of problem (4.8) with 8 users

hypervolume indicator value is obtained by MOPSO. We want the hypervolume indicator to be small since it shows the difference between the reference point and the non-dominated solution set. Additionally, Table 4.3 holds the elapsed time for each MOEA. MOPSO is the fastest among all algorithms in terms of running time while SPEA-II is the slowest.

Performance comparison of the problem (4.8) and (4.9) are represented in Fig. 4.3. We show the results for only SPEA2 since it is the best performer. Apparently, maximizing spectral efficiency and energy efficiency slightly offers a better spectral and energy efficiency tradeoff than minimizing power consumption and maximizing SE. Since, unlike the problem (4.8), the problem (4.9) ignores the maximization of energy efficiency and concentrates maximizing spectral efficiency with as low power consumption as.

Finally, we run the simulation for 24 users to show which algorithm performs best when the number of variables increases. Surprisingly, SPEA2 is still the winner as in Fig. 4.4.

Table 4.2: Simulation Parameters

Parameters	**Value**
Bandwidth	20 MHz
Carrier frequency	2 GHz
Num. of users	8 and 24
N_F, T_K, k_b	7 dB, 290^o K, 1.38x10^{-23} J/ K
LOS pathloss	$22log(d) + 34.02 + X_\sigma$, $\sigma = 4$ dB
NLOS pathloss	$39.1log(d) + 19.56 + X_\sigma$, $\sigma = 6$ dB
Probability of LOS	$\min(18/d, 1)(1 - e^{(-\frac{d}{63})}) + e^{(-\frac{d}{63})}$
P_t, and P_c	46 dBm, and 30 dBm, respectively
R_{min}, and P_{Thr}	100 kbps, and 10 dBm, respectively
User and BS heights	1.5 m and 25 m, respectively
Cell radius	200 m
Num. of realizations	20
τ_g	1

Table 4.3: Performance Evaluation of MOEAs

Performance Metric	**NSGA-II**	**MOPSO**	**SPEA2**
H for Problem (4.8)	1.9498e+08	1.8692e+08	**1.7934e+08**
H for Problem (4.9)	2.9696e+08	2.7391e+08	**2.7130e+08**
Elapsed time (sec)	145.96	**91.76**	235.85

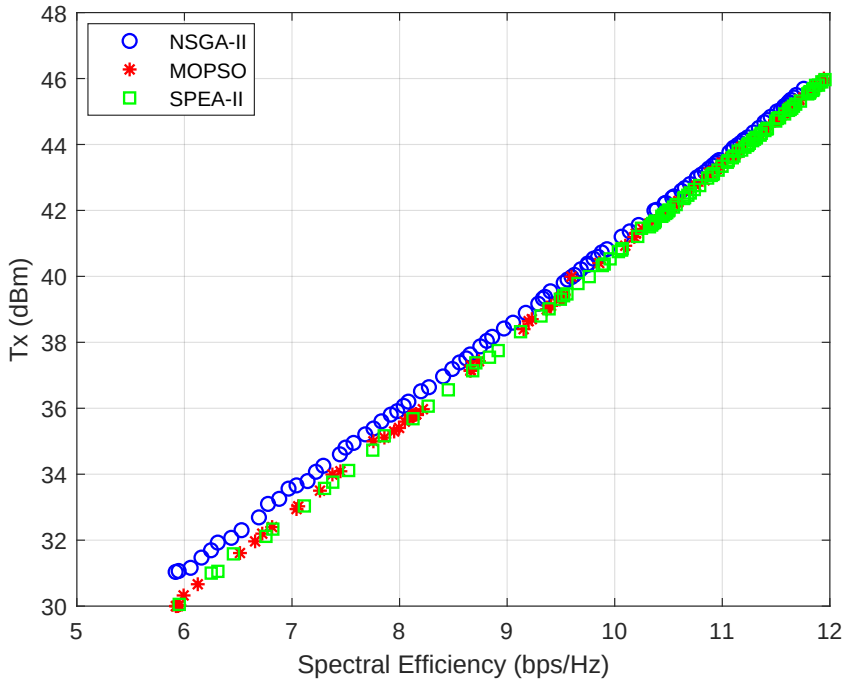


Figure 4.2: Pareto optimal set of problem (4.9) with 8 users

Algorithm 4.3 MOPSO for the SE-EE Optimization

Input:
Design variables: $\boldsymbol{\alpha_s^m}$ and $\boldsymbol{P_m}$,
SPEA2 parameters: S, S_r, w, w_d, l_p, l_g, μ, i_{max}.
Output:
Solution: Particle position $\boldsymbol{X} = [\boldsymbol{\alpha_s^m}, \boldsymbol{P_m}]$.
Set all input parameters except $\boldsymbol{X}$,
Initialize external archive $\boldsymbol{EA} = \emptyset$,
for $i = 1 : S$ **do**
  Generate initial $\boldsymbol{X}$, and its velocity $\boldsymbol{V}$ randomly,
  Evaluate the initial fitness values for $\boldsymbol{X}$ as in (4.13),
  Update personal best **Pbest** and global best **Gbest**,
end
Check domination and send leaders to $\boldsymbol{EA}$
while $i < i_{max}$ **do**
  for $s = 1 : S$ **do**
    Select a leader,
    Update $\boldsymbol{V}$,
    Update $\boldsymbol{X}$,
    Evaluate the fitness function according to (4.13),
    Apply mutation to $\boldsymbol{X}$ for new position $\hat{\boldsymbol{X}}$,
    if $\hat{\boldsymbol{X}}$ *dominates* $\boldsymbol{X}$ **then**
      $\boldsymbol{X} \leftarrow \hat{\boldsymbol{X}}$,
    else
      $\boldsymbol{X} \leftarrow \boldsymbol{X}$,
    end
    Update ***Pbest***
  end
  Update ***Gbest***
  Send non-dominated particles to $\boldsymbol{EA}$,
  Check for domination of new particles in $\boldsymbol{EA}$,
  Keep only non-dominated particles in $\boldsymbol{EA}$.
end

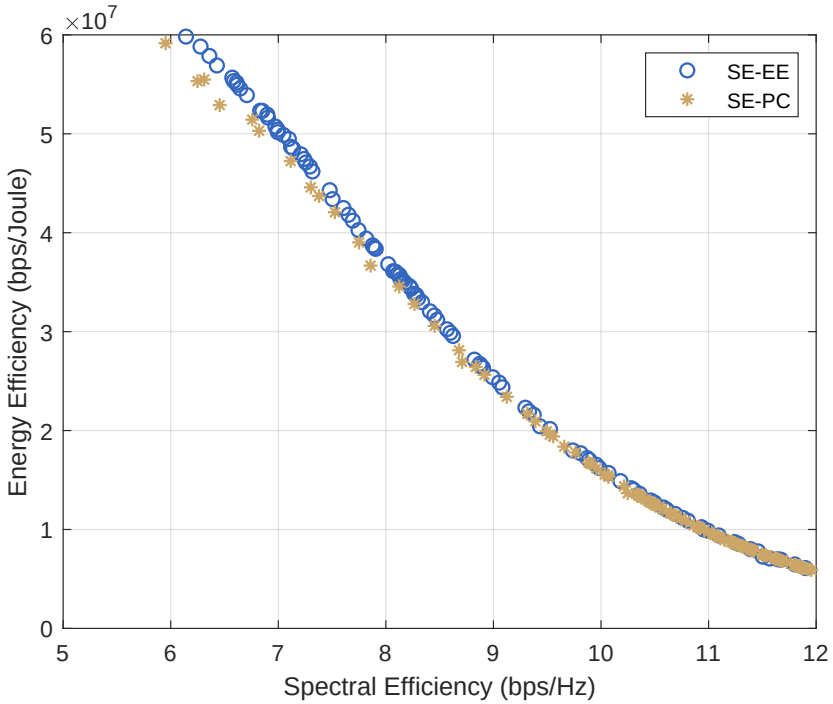


Figure 4.3: Comparison of problem (4.8) and problem (4.9)

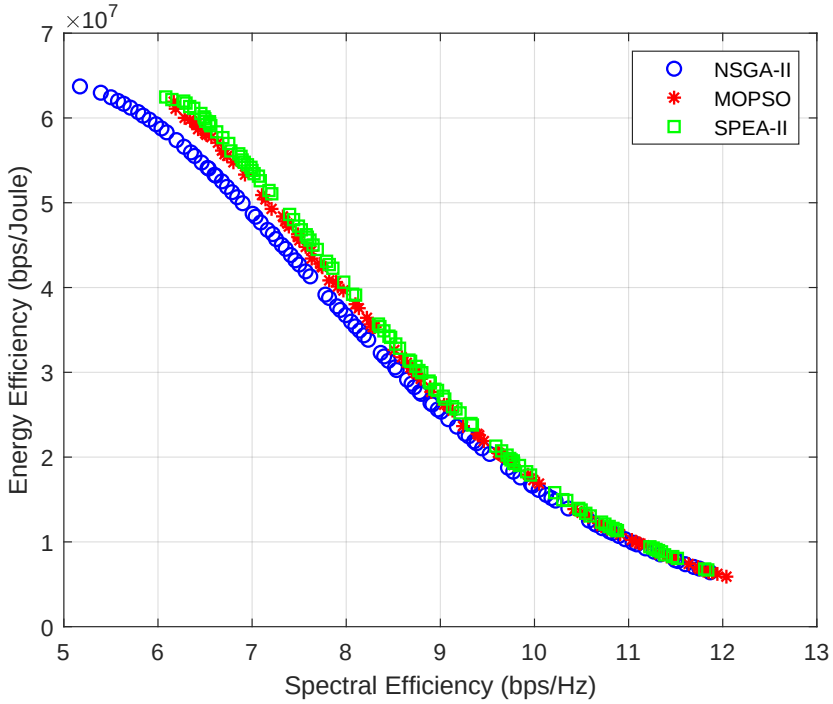


Figure 4.4: Pareto optimal set of problem (4.8) with 24 users

Chapter 5: A Distributed User-Cell Association for Spectral and Energy Efficiency Tradeoff in Massive MIMO UDHNs

The demand of users for gigabit data rates attracts researchers to focus on more academic research regarding the evolution of the fifth generation (5G). Higher spectral efficiency is one of the key aspects of the 5G design target to achieve orders of more capacity. To this end, cell densification, also known as heterogeneous networks (HetNets), is presented as one of the key enablers to achieving higher spectral efficiency due to dense spectrum reuse across the geographical area. In HetNets, dense small base stations (SBSs) with low transmit power are underlaid over macro base stations (MBSs) with high transmit power. Network densification has already been exercised in the third-generation (3G) and fourth-generation (4G) networks where SBSs aim to offload the traffic from MBSs.

Deployment of massive multiple-input and multiple-output (mMIMO) significantly boosts the spectral efficiency due to the independent data streams transmitted over a large number of antenna elements. Furthermore, mMIMO increases power gain by increasing the received signal power. Therefore, a desired quality-of-service (QoS) can be provided with less transmit power [88, 89]. Additionally, due to the channel hardening effect in massive MIMO systems, the effect of small-scale channel coefficients and channel estimation error is ignorable [90].

User association turns out to be more challenging with the deployment of mMIMO and HetNets. Although network densification promises high spectral efficiency, it brings its own challenges due to the installation of base stations having diverse transmit powers, such as load balancing problems, since users tend to associate with MBSs instead of SBSs due to their higher transmit power. On the other hand, the enormous number of antennas causes high

circuit power consumption in mMIMO systems [91]. Therefore, a user association algorithm for mMIMO can create an energy efficiency problem.

Many outstanding contributions for cell association [92, 93, 94] have been proposed, especially for single antenna HetNets. User association problem is standardized by Third Generation Partnership Project (3GPP) Long Term Evaluation (LTE) and performed based on Reference Signal Received Power (RSRP) and Reference Signal Received Quality (RSRQ), both are sent by users, [95, 96] which causes load imbalance between tiers due to different transmit power of BSs. Thus, users tend to associate with MBSs due to high transmit power. To avoid this drawback, a *biasing* method is proposed within 3GPP to keep the load balanced across tiers by scaling the RSRP for SBSs with an artificial and bias value[88, 97].

A limited number of proposals [91, 98, 99, 100, 101] have been investigated for mMIMO-enabled HetNets. Reference [91] investigated an energy-efficient, low complex, distributed, and fair user association where sum logarithmic energy efficiency maximization was performed with users' QoS constraints. The convex problem was solved via Lagrange duality analysis. In [98], authors developed a convex centralized methodology to maximize the sum instantaneous rate of all users and a decentralized methodology to maximize the instantaneous rate of each individual user with the consideration of fairness among users, and both problems are solved by Lagrange duality. Surprisingly, both strategies provide almost similar results where the decentralized approach has low complexity and does not require a control unit, while the centralized strategy requires high computational complexity and a control unit. In [99], the author proposed an algorithm for log-concave energy efficient utility function maximization with the help of Lagrange duality, where users' QoS requirements for the overall system were considered. A distance-based user association strategy was proposed in [100] where each user is associated with the nearest BS, and the effect of dense SBS deployment was analyzed for the energy efficiency maximization. Dense SBSs deployment and nearest base station association strategy enhance the energy efficiency. The authors in [101]

developed a fair strategy for joint maximization of energy efficiency and power allocation to solve user association problems in mMIMO HetNets.

Although most of the studies above do not consider the spectral and energy efficiency tradeoff, there are limited number of proposals [102, 103, 104] to overcome the tradeoff. In [102], the authors considered a resource allocation for the spectral and energy efficiency tradeoff by proposing a multi-objective optimization problem to optimize the subscribers' power allocation of orthogonal frequency-division multiplexing (OFDM) with imperfect channel state information (CSI). The authors in [103] investigated the spectral and energy efficiency tradeoff by employing an adjustable utility function for interference-aware power coordination. Reference [104] investigated the spectral and energy efficiency tradeoff with the consideration of fairness for mMIMO HetNets. To this end, a multi-objective optimization problem is first formulated as a mixed-integer non-convex problem and then converted into a single-objective optimization problem. The contributions of this Chapter[8] can be listed as follows:

- We perform a game theoretical algorithm inspired from [98]. Contrary to the methodology investigated in [98], which aims to maximize only the throughput of each individual user in a selfish manner, we propose an algorithm to optimize the spectral and energy efficiency tradeoff.

- Contrary to reference [104], which does not guarantee a Pareto-optimal solution due to the non-concavity of the rate function, our approach guarantees a Pareto-optimal solution for the spectral and energy efficiency tradeoff since both utilities for the spectral efficiency and energy efficiency are defined as convex functions.

- The problem is first formulated as a multi-objective problem and then converted into a single-objective optimization problem by incorporating a weighting factor to optimize the tradeoff.

[8]Part of this chapter was published in [105]. Permission is included in Appendix A.

- We compare the effect of the weighting factor on performance metrics by assigning different values between 0 and 1. Based on the observations, we deduce that there is always a spectral and energy efficiency tradeoff. Thus, we propose a flexible and adjustable distributed algorithm. Therefore, selection of the weighting factor is significantly crucial.

- We also evaluate the effect of transmit power of MBSs based on the simulation results and conclude that the proposed algorithm can provide energy efficiency and spectral efficiency gain for mMIMO ultra-dense heterogeneous networks (UDHNs) with an appropriate selection of weighting factor and less transmit power.

5.1 System Model

5.1.1 Radio Environment and Parameters

In this study, we focus on a two-tier downlink system with MBSs and SBSs distributed across a 2-dimensional plane where both of them share the same frequency band, and SBSs are assumed to be deployed to boost the capacity of MBSs in its coverage area. The notation $j \in \mathbb{J} = \{MBS_1, ..., MBS_M\} U \{SBS_1, ..., SBS_S\}$ and $k \in \mathbb{K} = \{1, 2, ..., K\}$ is used to index the base stations (combination of M MBSs and S SBSs) and users respectively. The distribution of the SBSs is assumed to be randomly located in the hot-zones where users have more density, and SBSs are deployed in the coverage area of the MBS within each cell with a certain distance. The distribution of the users is modeled as random and in a non-homogeneous manner with more density in the hot-zones created around the SBSs than in the MBSs. Proportional fairness (PF) is assumed among users to provide more resource blocks (RBs) to the users having good channel conditions.

Large-scale antenna arrays are deployed for both MBSs and SBSs where the notation for the number of antennas and the number of downlink data streams per time slot at any base station are assigned as A_j and S_j, respectively, where $1 < S_j < A_j$. We assume

a reciprocity-based channel estimation where the base stations estimate the channel using the pilots received from the associated users in uplink in time division multiplexing (TDD) mode. In addition, we assume an intra-cell interference free network due to linear zero-forcing beamforming (LZFBF) used in both MBS and SBS tiers [106].

This study assumes a transmission based on orthogonal frequency division multiple access (OFDMA) to let each user perform its transmission over contiguous time-frequency slots [107, 106]. The channel model considers the effect of both large-scale and small-scale fading (Rayleigh). However, channel hardening in massive MIMO makes negligible the effect of small-scale fading [108]. The large-scale fading consists of both pathloss and shadowing and is denoted by $g_{k,j}$ between a base station j and user k. We assume that $g_{k,j}$ is not changing across all RBs.

5.1.2 Modeling Instantaneous Rate and SINR

This study assumes that no user will suffer with zero throughput and all the RBs are used (heavy load condition) in a given time slot t. In addition, we define the load of an arbitrary base station j as the number of users $|K_j|$ connected to it in a given time slot t.

When a user k is served by an arbitrary base station j, the instantaneous rate $R_{k,j}$ given as in (5.1) which is borrowed from [109].

$$R_{k,j} = log_2(1 + \gamma_{k,j}), \tag{5.1}$$

where the notation $\gamma_{k,j}$ expresses the signal to interference plus noise ratio (SINR) with LZFBF [110] when a user k is served by an arbitrary base station j. Assuming that the base station has the perfect knowledge of the channel state information (CSI), $\gamma_{k,j}$ is given by

$$\gamma_{k,j} = (\frac{A_j - S_j + 1}{S_j}) \frac{g_{k,j} P_j}{No + \sum_{l \in \mathbb{J}, l \neq j} g_{k,l} P_l}. \tag{5.2}$$

In (5.2), P_j is the transmit power of the base station j, and the system noise power is given by $N_0 = N_F k_b T_K B$.

Finally, the achievable downlink data rate for user k, known as user throughput, is expressed as

$$r_k = \sum_{j \in \mathbb{J}} \omega_{k,j} R_{k,j}, \tag{5.3}$$

where $\omega_{k,j} \in [0, 1]$ is the number of required RBs for a user k when served by an arbitrary base station j. As can be seen from (5.3), we assume more than one base station can serve a user in the network.

5.1.3 Power Consumption Model for Massive MIMO

This subsection focuses on modeling the power consumption model for massive MIMO systems. With the deployment of a large number of antenna elements, the power consumption model dramatically changes and differs from the conventional single antenna base stations where it is a direct function of the radiated transmit power of the base station [111]. This concept does not fit massive MIMO systems since when the number of antenna elements is considered as $A \to \infty$, energy efficiency goes to infinity [91].

Power consumption for massive MIMO systems has been widely and clearly studied in [112]. We borrow the power consumption model $\mathcal{P}_j$ derived in [112] and [111] given by

$$\mathcal{P}_j = P_j/\eta_j + \sum\nolimits_{m=0}^{3} C_{m,0} S^m + \sum\nolimits_{m=0}^{2} C_{m,1} S^m A_j, \tag{5.4}$$

where A_j and S^m stand for the effect of the number of antenna elements and the number of active users, respectively. In addition, $C_{m,0}$ and $C_{m,1}$ are the coefficients related to the coding and decoding, channel estimation, precoding, power consumption of transceiver chains, and architectural cost derived in [112] and η_j is defined as the base station power amplifier efficiency.

5.1.4 Energy Efficiency Model

In conventional cellular networks, user-cell association is performed based on RSRP (or RSRQ) levels where the user is served by the base station, which offers the highest received power. However, this approach does not consider the energy efficiency (bits/Hz/Joule) of a user k when associated with the base station j, although it is a crucial performance metric for UDHNs. A large number of antenna elements in massive MIMO systems and ultra-dense deployment of the base stations increase the power consumption due to growing circuit power consumption [99]. Therefore, an energy-efficient user-cell association is needed for green communications.

The energy efficiency of a user k is defined as a function of both user throughput and power consumption and expressed as

$$EE_k = \sum_{j \in \mathbb{J}} \frac{\omega_{k,j} R_{k,j}}{\mathcal{P}_j}. \tag{5.5}$$

Equation (5.5) emphasizes that higher energy efficiency means lower power consumption. However, there is always a tradeoff between spectral efficiency and energy efficiency. Therefore, we propose a multi-objective user-cell association scheme that aims to optimize this tradeoff with a weighting factor.

5.2 Distributed User-Cell Association Scheme

This section presents the problem formulation and the solution methodology for the multi-objective optimization problem.

5.2.1 Problem Formulation

This subsection focuses on the formulation of a distributed user-cell association problem that aims to optimize the spectral and energy efficiency tradeoff. To this end, this study proposes a methodology where each user tries to maximize its weighted spectral efficiency

and energy efficiency in a selfish manner. We assume PF among users to allocate more RBs to the users who have stronger downlink channels. Therefore, we define a logarithmic user-centric utility function to perform a multi-objective optimization problem. The multi-objective optimization problem is given in (5.6), where the constraint (5.6c) limits the usage of RBs used to be within the downlink data streams S_j. Finally, the constraint (5.6d) guarantees that all users will have positive throughput and required number of RBs cannot exceed 1.

$$\max_{\omega} \sum_{k \in \mathbb{K}_j} \log(r_{k,j}), \tag{5.6a}$$

$$\max_{\omega} \sum_{k \in \mathbb{K}_j} \log(EE_{k,j}), \tag{5.6b}$$

$$\text{s.t.} \quad \sum_{k \in \mathcal{K}_j} \omega_{k,j} \leq S_j, \tag{5.6c}$$

$$0 \leq \omega_{k,j} \leq 1,\ r_{k,j} \geq 0, \quad \forall k \in \mathcal{K}_j. \tag{5.6d}$$

In order to solve problem (5.6), we convert it into a single-objective optimization problem using the weighting function method [113]. Such that the new objective function can be written as

$$\max_{\omega} \sum_{k \in \mathbb{K}_j} \lambda \log(r_{k,j}) + \sum_{k \in \mathbb{K}_j} (1-\lambda) \log(EE_{k,j}), \tag{5.7}$$

$$\text{s.t.} \quad \text{(5.6c) and (5.6d)},$$

where trading between the impact of energy efficiency and spectral efficiency can be performed thanks to $\lambda \in [0, 1]$ introduced as a weighting factor in the objective function (5.7). Note that, with the $\lambda = 1$, the user-cell association simply becomes a throughput maximization problem. On the other hand, when $\lambda = 0$, it is basically an energy efficiency maximization problem.

Due to the characteristics of logarithmic functions, it is possible to rewrite the objective function given in (5.7) as

$$\max_{\omega} \sum_{k \in \mathbb{K}_j} \log(\frac{\omega_{k,j} R_{k,j}}{\mathcal{P}_j^{(1-\lambda)}}), \tag{5.8}$$

$$\text{s.t.} \quad \text{(5.6c) and (5.6d).}$$

5.2.2 Lagrangian Dual Analysis and KKT Conditions

Lagrangian duality analysis of the problem (5.8) can be performed as in [98] to acquire an optimal solution and can be expressed as

$$L(\boldsymbol{\omega_j}, \boldsymbol{\mu}) = \sum_{k \in \mathbb{K}_j} \log(\frac{\omega_{k,j} R_{k,j}}{\mathcal{P}_j^{(1-\lambda)}}) - \mu(\sum_{k} \alpha_{k,j} - S_j), \tag{5.9}$$

where $\boldsymbol{\omega} \geq 0$ is the primal variable and $\boldsymbol{\mu}$ is the non-negative Lagrange multiplier related to the objective function (5.8) and the constraint (5.6c) respectively.

The partial derivative of (5.9) with respect to $\omega_{k,j}$ gives the sufficient Karush-Kuhn-Tucker (KKT) conditions for the optimal solution of (5.8) and the partial derivative takes the form of

$$\frac{\partial L}{\partial \omega_{k,j}} = \frac{1}{\omega_{k,j}} - \mu \leq 0, \tag{5.10}$$

$$\omega_{k,j} \leq \frac{1}{\mu}, \tag{5.11}$$

$$\sum_{k \in \mathcal{K}_j} \omega_{k,j} = S_j. \tag{5.12}$$

The optimal solution requires strict equality in (5.11) for positive component μ. Additionally, the constraint (5.6c) must hold strict equality as in (5.12) in a heavy loaded network where all the RBs are used in the optimal point. Therefore, an expression for the optimal

$\boldsymbol{\omega_j}$ can be derived as

$$\omega_{k,j} = \begin{cases} 1, & \text{for } |K_j| < S_j \\ \frac{1}{\mu}, & \text{for } |K_j| \geq S_j \end{cases} \tag{5.13}$$

where μ can be expressed in terms of the downlink data stream S_j of an arbitrary base station j and the number of users $|K_j|$ associated with it from the (5.11) and (5.13) as

5.2.3 Game Theoretical Solution

After making the Lagrangian duality analysis and analyzing KKT conditions, it is obvious that (5.8) is maximized when each user maximizes its weighted sum of energy efficiency and spectral efficiency individually based on the condition given in (5.14) by considering the overall scenario and the load information of the base stations. To this end, we propose and game theoretical methodology to find the Pareto-optimal solution to the problem (5.8).

We consider a fully loaded scenario where each base station serves at least S_j number of users. The algorithm performs a game theory where a user (Player) tends to change its unique association configuration (Action) selfishly until there is no base station yielding better solution to joint spectral and energy efficiency optimization (Payoff function). In other words, all users keep changing their association until they reach *Nash equilibrium* (NE). The NE condition is given as

$$\frac{S_{j_k} R_{k,j_k}}{|K_{j_k}| \mathcal{P}_{j_k}^{(1-\lambda)}} > \frac{S_j R_{k,j}}{(|K_j|+1) \mathcal{P}_j^{(1-\lambda)}}, \quad \forall k \in \mathbb{K}, \quad \forall j, j_k \in \mathbb{J}, \quad j_k \neq j, \tag{5.14}$$

where j_k stands for the current serving base station of user k and j is the prospective base station that can serve if it can offer a better spectral and energy efficiency tradeoff.

Algorithm 5.1 presents the user-centric approach for the user-cell association. In Algorithm 1, if user k can obtain a better payoff function from an arbitrary base station j, user k changes its association configuration from base station j_k to j in step (8). If step (8) holds for the current iteration i, migration count m_{i_k} increases (step 10), the load of the prospective

base station $|K_j|$ increases (step 11), the load of the current serving BS $|K_{j_k}|$ decreases (step 12), and the algorithm updates the association of user k in (step 13). Additionally, i_{max}, p, and λ stand for the maximum number of iteration, switching probability, and weighting factor, respectively.

Algorithm 5.1 Game theoretical user-cell association (GTUCA)

Set initial parameters:
i_{max}, p, and λ,
set initial values of $i = 0$ and $\boldsymbol{m} = 0$,
while $i < i_{max}$ **do**
  for $k = 1 : K$ **do**
    for $j = 1 : J$ **do**
      The base station j calculates power consumption P_j using (5.4),
      if $\frac{S_j R_{k,j}}{(|K_j|+1)\mathcal{P}_j^{(1-\lambda)}} > \frac{S_{j_k} R_{k,j_k}}{|K_{j_k}|\mathcal{P}_{j_k}^{(1-\lambda)}}$ **then**
        if $rand < p^{(m_{i_k}+1)}$ **then**
          $m_{i_k} \leftarrow m_{i_k} + 1$,
          $|K_j| \leftarrow |K_j| + 1$,
          $|K_{j_k}| \leftarrow |K_{j_k}| - 1$,
          $j_k \leftarrow j$,
        end
      end
    end
  end
end

5.3 Numerical Evaluation

5.3.1 Network Model

In this section, we perform a simulation to evaluate the proposed GTUCA algorithm in a downlink UDHN consisting of 7 MBSs having 100 antennas each with 46 dBm transmit power, and randomly deployed three hot-zones having more user density than MBSs are considered. Within each hot-zone, there are 4 randomly deployed SBSs having 10 antennas each with 35 dBm transmit power in a hexagonal region. The MBSs are placed in the center of each hexagon area.

The coefficients regarding the power consumption are borrowed from [112] and given in Table 5.1, including detailed information about the simulation parameters. We assume no intra-cell interference in this study. Thus, we assign a set of 10 pilots shared by MBSs, and a different set of 2 pilots for channel estimation. Additionally, the pathloss models and simulation parameters for both MBS and SBS deployments are taken from existing 3GPP standards [87].

5.3.2 User Throughput and Energy Efficiency

The maximum rate-based user-cell association scheme is chosen as the baseline algorithm, where a user is associated with a base station providing the highest rate, to evaluate the performance of the proposed algorithm. In Fig. 5.1 and Fig. 5.2, we compare the geometric mean of the energy efficiency and the spectral efficiency, respectively, and obtain the performance of the proposed algorithm with different weighting factors, i.e., $\lambda = 0, 0.3, 0.5, 0.7, 1$. It is evident from the figures that the best performance in terms of energy efficiency is obtained by the proposed approach when we set the $\lambda = 0$. However, spectral efficiency will perform worst when $\lambda = 0$. We can deduce that smaller λ causes less spectral efficiency but higher energy efficiency since reducing λ allows more users to be associated with SBSs with low transmit power leading to an increase in energy efficiency.

The selection of λ is highly critical in optimizing the spectral and energy efficiency tradeoff. As can be seen from Fig. 5.2, when $\lambda = 0.5$, the baseline and the proposed algorithm perform almost similar in terms of spectral efficiency. However, the proposed algorithm provides better energy efficiency than the baseline given in Fig. 5.1. To obtain a better spectral and energy efficiency tradeoff, the range of the weighting factor should be specified as $0.5 < \lambda < 1$.

Fig. 5.3 and Fig. 5.4 illustrate that the proposed algorithm provides better energy and spectral efficiency than the baseline for most of the realization when the transmission power of MBSs relatively reduced from 46 dBm to 43 dBm with the same number of antennas and

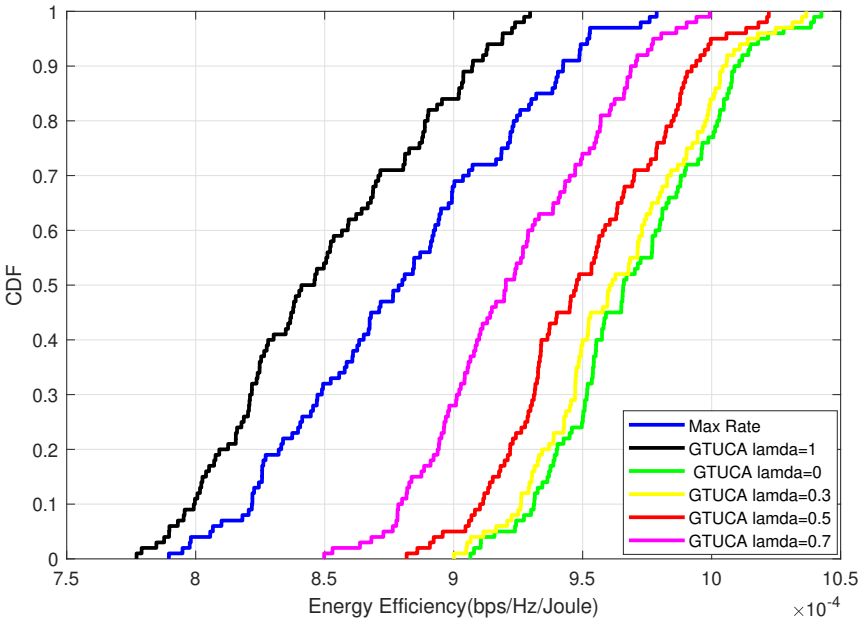


Figure 5.1: Geometric mean of the energy efficiency.

$\lambda = 0.6$. We observe higher spectral efficiency since the interference from MBSs is reduced and higher energy efficiency because users are more likely to associate with SBSs due to the lower transmit power of MBSs.

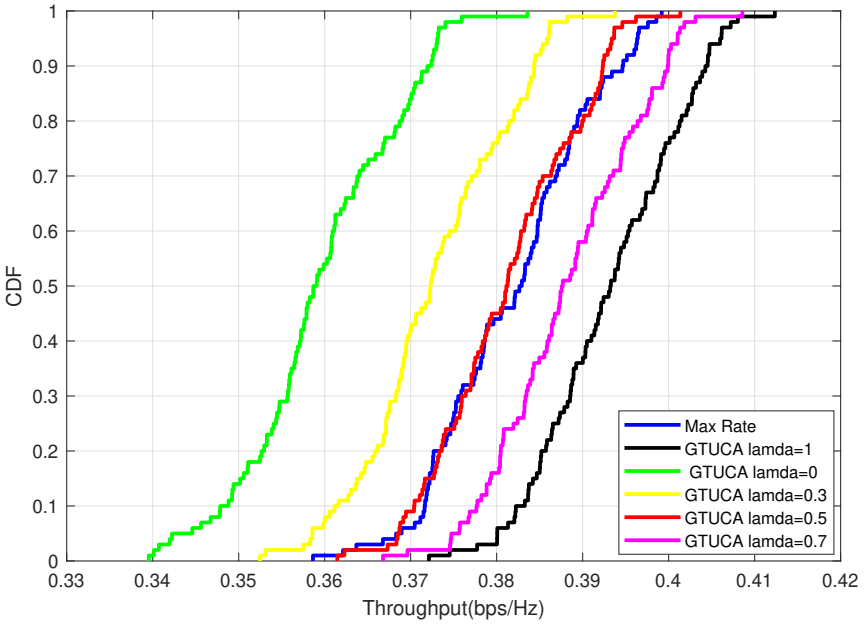


Figure 5.2: Geometric mean of the spectral efficiency.

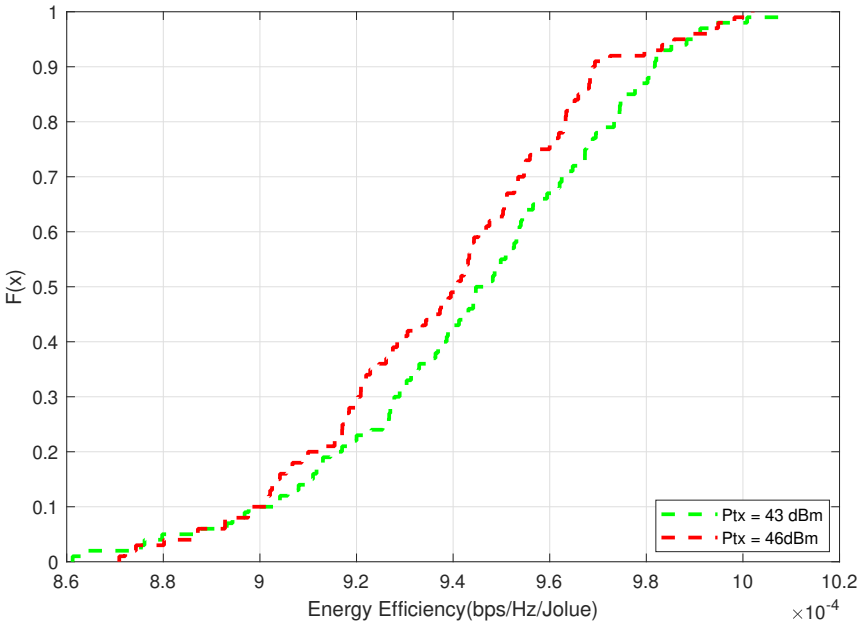


Figure 5.3: Energy efficiency gain.

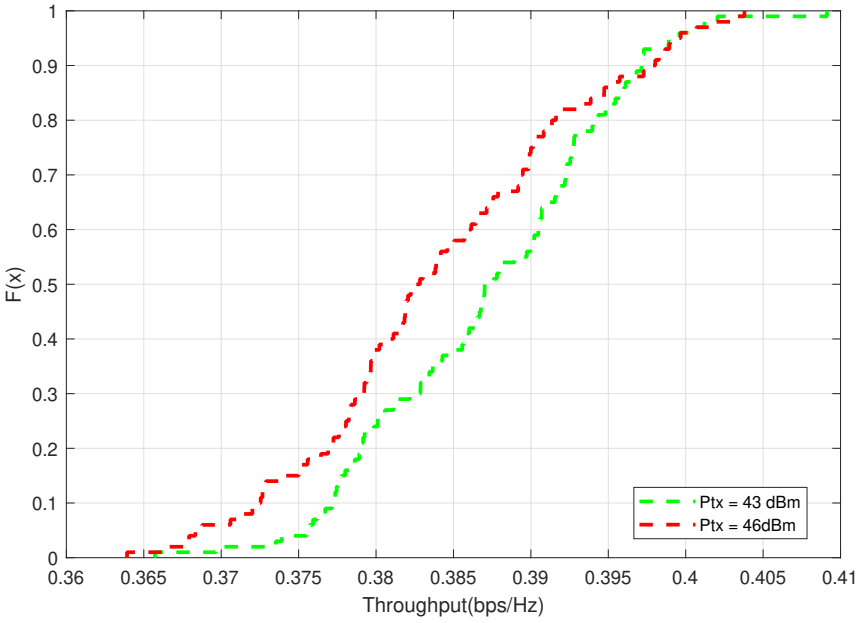


Figure 5.4: Throughput gain.

Table 5.1: Simulation Parameters

Parameters	**Value**
Bandwidth	20 MHz
Carrier frequency	2 GHz
Number of users	1400
Q , T	3, 7
N_F, T_K, k_b	7 dB, 290^o K, 1.38x10^{-23} J/ K
Two-slope LOS path loss model of MBS	$22log(d) + 34.02 + X_\sigma$, $d < d1$ $40log(d) - 11.02 + X_\sigma$,$d1 < d < d2$ $\sigma = 4$ dB, $d1 = 320$m, $d2 = 5000$m
NLOS pathloss model of MBS	$39.1log(d) + 19.56 + X_\sigma$ $\sigma = 6$ dB
Two-slope LOS path loss model of SBS	$22log(d) + 34.02 + X_\sigma$, $d < 120$m $40log(d) - 3.36 + X_\sigma$,$120 < d < 5000$m $\sigma = 3$ dB
NLOS pathloss model of SBS	$36.7log(d) + 30.53 + X_\sigma$ $\sigma = 4$ dB
$P_{LOS}(d)$ for SBS	$\min(18/d, 1)(1 - e^{(-\frac{d}{36})}) + e^{(-\frac{d}{36})}$
$P_{LOS}(d)$ for MBS	$\min(18/d, 1)(1 - e^{(-\frac{d}{63})}) + e^{(-\frac{d}{63})}$
Transmit power of MBS and FBS respectively	46 dBm, 35 dBm,
A_j for MBS, FBS	100, 10
S_j for MBS, FBS	10, 2
User height	1.5 m
BS height	FBS: 10 m, MBS: 25m
FBS radius	40 m
Number of realizations	100
Power consumption coefficients	$C_{0,0} = 4, C_{1,0} = 4.8, C_{2,0} = 0,$ $C_{3,0} = 2.08 \times 10^{-}8, C_{0,1} = 1,$ $C_{1,1} = 9.5 \times 10^{-}8, C_{2,1} = 6.25x10^{-}8$

Chapter 6: Concluding Remarks

In chapter 2, we have proposed a two-stage beam selection algorithm to reduce the complexity of existing decremental QR precoding (D-QR-P) based beam selection algorithm and a three-stage beam selection algorithm for incremental QR precoder (I-QR-P) based beam selection algorithm. The benchmark algorithms select beams from all available beams during the beam selection process, although most do not contribute to users. Nevertheless, the proposed algorithms first identify the most contributing beams to narrow the candidate beam set and perform the beam selection afterward. Combining this strategy with matrix perturbation theory reduces the beam selection complexity significantly. The results validate that the proposed algorithms provide almost identical sum-rate performance with the baseline algorithms, and both can be adopted at low and high signal-to-noise ratios (SNR)s. Note that I-QR-P is only suitable for application at medium and high SNRs due to the loss observed in low SNRs. Thence, the three-stage method is more attractive since it can compensate for this loss along with its lowest complexity. Additionally, the practical implementation of proposed methods for wideband mmWave scenarios will be investigated with multi-antenna users as a future study.

In chapter 3, we have proposed a hybrid precoding scheme for millimeter-wave (mmWave) lens antenna subarray (LAS)-MIMO architectures that uses the heuristic artificial bee colony (ABC) and orthogonal matching pursuit (OMP) algorithms. Thus, it is called ABC-aided spatially sparse precoding. The proposed precoding algorithm first selects the antennas that need to be activated for each lens and calculates the corresponding lens antenna effect. This information is then used in OMP to find the analog and digital precoders where the precoding problem is formulated as a sparse reconstruction problem due to the sparse behavior of

the mmWave channel. ABC runs until it finds the best precoding components providing the highest SE. The simulation results show that it can achieve near-optimal performance in terms of spectral efficiency as the number of lenses increases in the LAS system while outperforming the traditional array (TA)-MIMO in terms of energy efficiency for single and multiple radio frequency (RF) chains. Additionally, using different switch types may further improve energy efficiency while maintaining the same spectral efficiency. The future scope of this study can be proposing an algorithm that can handle joint precoding and combining for a multi-user scenario in a wideband mmWave spectrum.

In chapter 4, we have investigated on solving the spectral and energy efficiency tradeoff formulated as a multi-objective optimization problem (MOP) in a downlink non-orthogonal multiple access (NOMA) with the help of population-based multi-objective evolutionary algorithms (MOEA)s, namely non-dominated sorting genetic algorithm-II (NSGA-II), multi-objective particle swarm optimization (MOPSO), and strength Pareto evolutionary algorithm-2 (SPEA2), where Pareto optimal solutions are attained by the simultaneous optimization of objectives. Additionally, the performance comparison has been observed through the hypervolume indicator and running time of the algorithms. The results demonstrated that SPEA2 works best for the spectral and energy efficiency tradeoff in a downlink NOMA. However, it is the slowest among MOEAs while MOPSO is the fastest. Moreover, we tried to understand which problem formulation is the best for the spectral and energy efficiency tradeoff optimization and concluded that maximizing energy efficiency and spectral efficiency simultaneously provides better Pareto front quality than maximizing spectral efficiency and minimizing power consumption. The future scope of this study can be the implementation of other popular MOEAs in the existence of imperfect SIC and finding the best one.

In chapter 5, we have proposed a user-centric and distributed user-cell association algorithm based on game theory for ultra-dense heterogeneous networks (UDHNs) consisting of massive macro base stations (MBSs) and small base stations (SBSs). We formulate a multi-objective optimization problem where each user aims to optimize the spectral and energy

efficiency tradeoff in a selfish and fair way. After we formulate the problem, it is converted into a single objective optimization problem by the weighting function method. Numerical results indicate that the proposed algorithm exhibits better performance than the baseline algorithm with an appropriate weighting factor. Finally, the fully decentralized implementation and the low complexity of the proposed algorithm make it attractive for future massive MIMO-enabled UDHNs.

References

[1] J. Tan, X. Wei, S. Bi, M. Cui, and L. Dai, "Massive MIMO for 5G and beyond," *Flexible and Cognitive Radio Access Technologies for 5G and Beyond*, p. 197, 2020.

[2] J. Huang, C.-X. Wang, R. Feng, J. Sun, W. Zhang, and Y. Yang, "Multi-frequency mmWave massive MIMO channel measurements and characterization for 5G wireless communication systems," *IEEE J. Sel. Areas Commun.*, vol. 35, no. 7, pp. 1591–1605, 2017.

[3] A. I. Sulyman, A. T. Nassar, M. K. Samimi, G. R. MacCartney, T. S. Rappaport, and A. Alsanie, "Radio propagation path loss models for 5G cellular networks in the 28 GHz and 38 GHz millimeter-wave bands," *IEEE Commun. Mag.*, vol. 52, no. 9, pp. 78–86, 2014.

[4] M. Elkashlan, T. Q. Duong, and H.-H. Chen, "Millimeter-wave communications for 5G: Fundamentals: Part I [Guest Editorial]," *IEEE Commun. Mag.*, vol. 52, no. 9, pp. 52–54, 2014.

[5] D. J. Love and R. W. Heath, "Multimode precoding for MIMO wireless systems," *IEEE Trans. Signal Process.*, vol. 53, no. 10, pp. 3674–3687, 2005.

[6] X. Gao, L. Dai, Z. Chen, Z. Wang, and Z. Zhang, "Near-optimal beam selection for beamspace mmWave massive MIMO systems," *IEEE Commun. Lett.*, vol. 20, no. 5, pp. 1054–1057, 2016.

[7] S. Hur, T. Kim, D. J. Love, J. V. Krogmeier, T. A. Thomas, and A. Ghosh, "Millimeter wave beamforming for wireless backhaul and access in small cell networks," *IEEE Trans. Commun.*, vol. 61, no. 10, pp. 4391–4403, 2013.

[8] X. Gao, L. Dai, Z. Gao, T. Xie, and Z. Wang, "Precoding for mmWave massive MIMO," in *mmWave Massive MIMO*. Elsevier, 2017, pp. 79–111.

[9] O. E. Ayach, S. Rajagopal, S. Abu-Surra, Z. Pi, and R. W. Heath, "Spatially sparse precoding in millimeter wave MIMO systems," *IEEE Trans. Wireless Commun.*, vol. 13, no. 3, pp. 1499–1513, 2014.

[10] J. A. Zhang, X. Huang, V. Dyadyuk, and Y. J. Guo, "Massive hybrid antenna array for millimeter-wave cellular communications," *IEEE Wireless Commun.*, vol. 22, no. 1, pp. 79–87, 2015.

[11] J. Brady, N. Behdad, and A. M. Sayeed, "Beamspace MIMO for millimeter-wave communications: System architecture, modeling, analysis, and measurements," *IEEE Trans. Antennas Propag.*, vol. 61, no. 7, pp. 3814–3827, 2013.

[12] D. Tse and P. Viswanath, "Capacity of wireless channels," *Fundamentals of wireless communication*, 2005.

[13] X. Gao, L. Dai, S. Han, I. Chih-Lin, and X. Wang, "Reliable beamspace channel estimation for millimeter-wave massive MIMO systems with lens antenna array," *IEEE Trans. Wireless Commun.*, vol. 16, no. 9, pp. 6010–6021, 2017.

[14] D. A. Schmidt, M. Joham, and W. Utschick, "Minimum mean square error vector precoding," *European Transactions on Telecommunications*, vol. 19, no. 3, pp. 219–231, 2008.

[15] A. Liza and H. Arslan, "Beamforming and beam management in 5G and beyond," *Flexible and Cognitive Radio Access Technologies for 5G and Beyond*, 2020.

[16] Z. Shen, R. Chen, J. G. Andrews, R. W. Heath, and B. L. Evans, "Low complexity user selection algorithms for multiuser MIMO systems with block diagonalization," *IEEE Trans. Signal Process.*, vol. 54, no. 9, pp. 3658–3663, 2006.

[17] T. Xie, L. Dai, X. Gao, M. Z. Shakir, and J. Li, "Geometric mean decomposition based hybrid precoding for millimeter-wave massive MIMO," *China Communications*, vol. 15, no. 5, pp. 229–238, 2018.

[18] X. Gao, L. Dai, S. Han, I. Chih-Lin, and R. W. Heath, "Energy-efficient hybrid analog and digital precoding for mmWave MIMO systems with large antenna arrays," *IEEE J. Sel. Areas Commun.*, vol. 34, no. 4, pp. 998–1009, 2016.

[19] F. Sohrabi and W. Yu, "Hybrid digital and analog beamforming design for large-scale antenna arrays," *IEEE J. Sel. Topics Signal Process.*, vol. 10, no. 3, pp. 501–513, 2016.

[20] A. Alkhateeb, G. Leus, and R. W. Heath, "Limited feedback hybrid precoding for multi-user millimeter wave systems," *IEEE Trans. Wireless Commun.*, vol. 14, no. 11, pp. 6481–6494, 2015.

[21] R. Mendez-Rial, C. Rusu, N. González-Prelcic, A. Alkhateeb, and R. W. Heath, "Hybrid MIMO architectures for millimeter wave communications: Phase shifters or switches?" *IEEE Access*, vol. 4, pp. 247–267, 2016.

[22] D. H. Nguyen, L. B. Le, T. Le-Ngoc, and R. W. Heath, "Hybrid MMSE precoding and combining designs for mmWave multiuser systems," *IEEE Access*, vol. 5, pp. 19 167–19 181, 2017.

[23] C. Chen, Y. Dong, X. Cheng, and L. Yang, "Low-resolution PSs based hybrid precoding for multiuser communication systems," *IEEE Trans. Veh. Technol.*, vol. 67, no. 7, pp. 6037–6047, 2018.

[24] D. Zhang, P. Pan, R. You, and H. Wang, "SVD-based low-complexity hybrid precoding for millimeter-wave MIMO systems," *IEEE Commun. Lett.*, vol. 22, no. 10, pp. 2176–2179, 2018.

[25] M. Karabacak, G. Mumcu, and H. Arslan, "Hybrid MIMO architecture using lens arrays," Jul. 14 2020, US Patent 10,714,836.

[26] M. Karabacak, H. Arslan, and G. Mumcu, "Lens antenna subarrays in mmWave hybrid MIMO systems," *IEEE Access*, vol. 8, pp. 216 634–216 644, 2020.

[27] S. Cetinkaya, L. Afeef, G. Mumcu, and H. Arslan, "Heuristic inspired precoding for millimeter-wave MIMO systems with lens antenna subarrays," in *Proc. IEEE 95th Veh. Technol. Conf. (VTC2022-Spring)*, 2022, pp. 1–6.

[28] S. Cetinkaya and H. Arslan, "Beamspace MIMO systems with reduced beam selection complexity," *IEEE Commun. Lett. ,early access*, 2023.

[29] L. Dai, B. Wang, Z. Ding, Z. Wang, S. Chen, and L. Hanzo, "A survey of non-orthogonal multiple access for 5G," *IEEE Commun. Surveys Tuts.*, vol. 20, no. 3, pp. 2294–2323, 2018.

[30] J. Umehara, Y. Kishiyama, and K. Higuchi, "Enhancing user fairness in non-orthogonal access with successive interference cancellation for cellular downlink," in *Proc. IEEE Int. Conf. Commun. Syst. (IEEE ICCS)*, 2012, pp. 324–328.

[31] Z. Yuan, G. Yu, W. Li, Y. Yuan, X. Wang, and J. Xu, "Multi-user shared access for internet of things," in *Proc. IEEE 83rd Veh. Technol. Conf. (VTC Spring)*. IEEE, 2016, pp. 1–5.

[32] T. A. Levanen, J. Pirskanen, T. Koskela, J. Talvitie, and M. Valkama, "Radio interface evolution towards 5G and enhanced local area communications," *IEEE Access*, vol. 2, pp. 1005–1029, 2014.

[33] G. Wunder, P. Jung, and C. Wang, "Compressive random access for post-LTE systems," in *Proc. IEEE Int. Conf. Commun. Workshops*. IEEE, 2014, pp. 539–544.

[34] H. Lee, S. Kim, and J.-H. Lim, "Multiuser superposition transmission (must) for lte-a systems," in *Proc. IEEE Int. Conf. Commun. (ICC)*. IEEE, 2016, pp. 1–6.

[35] S. Cetinkaya and H. Arslan, "Energy and spectral efficiency tradeoff in NOMA: Multi-objective evolutionary approaches," in *Proc. IEEE Int. Conf. Commun. Workshops*, 2020, pp. 1–6.

[36] M. M. Lodro, N. Majeed, A. A. Khuwaja, A. H. Sodhro, and S. Greedy, "Statistical channel modelling of 5G mmWave MIMO wireless communication," in *Proc. Int. Conf. Comput., Math. Eng. Technol. (iCoMET)*, 2018, pp. 1–5.

[37] P. V. Amadori and C. Masouros, "Low RF-complexity millimeter-wave beamspace-MIMO systems by beam selection," *IEEE Trans. Commun.*, vol. 63, no. 6, pp. 2212–2223, 2015.

[38] I. Orikumhi, J. Kang, H. Jwa, J.-H. Na, and S. Kim, "SINR maximization beam selection for mmWave beamspace MIMO systems," *IEEE Access*, vol. 8, pp. 185 688–185 697, 2020.

[39] D. Wang, W. Zhang, and Q. Zhu, "Heuristic search inspired beam selection algorithms for mmWave MU-MIMO system with discrete lens array," *IEEE Access*, vol. 9, pp. 61 324–61 333, 2021.

[40] M. A. L. Sarker, M. F. Kader, and D. S. Han, "Rate-loss mitigation for a millimeter-wave beamspace MIMO lens antenna array system using a hybrid beam selection scheme," *IEEE Syst. J.*, vol. 14, no. 3, pp. 3582–3585, 2020.

[41] W. Shen, X. Bu, X. Gao, C. Xing, and L. Hanzo, "Beamspace precoding and beam selection for wideband millimeter-wave MIMO relying on lens antenna arrays," *IEEE Trans. Signal Process.*, vol. 67, no. 24, pp. 6301–6313, 2019.

[42] C. Feng, W. Shen, and J. An, "Beam selection for wideband millimeter wave MIMO relying on lens antenna arrays," *IEEE Commun. Lett.*, vol. 23, no. 10, pp. 1875–1878, 2019.

[43] R. Pal, A. K. Chaitanya, and K. V. Srinivas, "Low-complexity beam selection algorithms for millimeter wave beamspace MIMO systems," *IEEE Commun. Lett.*, vol. 23, no. 4, pp. 768–771, 2019.

[44] R. Pal, K. V. Srinivas, and A. K. Chaitanya, "A beam selection algorithm for millimeter-wave multi-user MIMO systems," *IEEE Commun. Lett.*, vol. 22, no. 4, pp. 852–855, 2018.

[45] Q. Zhang, X. Li, B.-Y. Wu, L. Cheng, and Y. Gao, "On the complexity reduction of beam selection algorithms for beamspace MIMO systems," *IEEE Wireless Commun. Lett.*, vol. 10, no. 7, pp. 1439–1443, 2021.

[46] M. Nazzal, M. A. Aygül, A. Görçin, and H. Arslan, "Dictionary learning-based beamspace channel estimation in millimeter-wave massive MIMO systems with a lens antenna array," in *Proc. 15th Int. Wireless Commun. Mobile Comput. Conf. (IWCMC)*, 2019, pp. 20–25.

[47] G. H. Golub and C. F. V. Loan, *Matrix Computations, 3rd Ed.* Baltimore, MD, USA: The John Hopkins University Press, 1996.

[48] V. K. Trivedi, K. Ramadan, P. Kumar, M. I. Dessouky, and F. E. Abd El-Samie, "Enhanced OFDM-NOMA for next generation wireless communication: a study of PAPR reduction and sensitivity to CFO and estimation errors," *AEU-International Journal of Electronics and Communications*, vol. 102, pp. 9–24, 2019.

[49] I. Ahmed, H. Khammari, A. Shahid, A. Musa, K. S. Kim, E. De Poorter, and I. Moerman, "A survey on hybrid beamforming techniques in 5G: Architecture and system model perspectives," *IEEE Commun. Surveys Tuts.*, vol. 20, no. 4, pp. 3060–3097, 2018.

[50] Z. Pi and F. Khan, "An introduction to millimeter-wave mobile broadband systems," *IEEE Commun. Mag.*, vol. 49, no. 6, pp. 101–107, 2011.

[51] A. Forenza, D. J. Love, and R. W. Heath, "Simplified spatial correlation models for clustered MIMO channels with different array configurations," *IEEE Trans. Veh. Technol.*, vol. 56, no. 4, pp. 1924–1934, 2007.

[52] H. Xu, V. Kukshya, and T. S. Rappaport, "Spatial and temporal characteristics of 60-GHz indoor channels," *IEEE J. Sel. Areas Commun.*, vol. 20, no. 3, pp. 620–630, 2002.

[53] A. M. Sayeed, "Deconstructing multiantenna fading channels," *IEEE Trans. Signal Process.*, vol. 50, no. 10, pp. 2563–2579, 2002.

[54] R. W. Heath, N. Gonzalez-Prelcic, S. Rangan, W. Roh, and A. M. Sayeed, "An overview of signal processing techniques for millimeter wave MIMO systems," *IEEE J. Sel. Topics Signal Process.*, vol. 10, no. 3, pp. 436–453, 2016.

[55] S. Han, I. Chih-Lin, Z. Xu, and C. Rowell, "Large-scale antenna systems with hybrid analog and digital beamforming for millimeter wave 5G," *IEEE Commun. Mag.*, vol. 53, no. 1, pp. 186–194, 2015.

[56] A. M. Elbir, "CNN-based precoder and combiner design in mmWave MIMO systems," *IEEE Commun. Lett.*, vol. 23, no. 7, pp. 1240–1243, 2019.

[57] E. Björnson, J. Hoydis, L. Sanguinetti *et al.*, "Massive MIMO networks: Spectral, energy, and hardware efficiency," *Foundations and Trends® in Signal Processing*, vol. 11, no. 3-4, pp. 154–655, 2017.

[58] R. Eberhart and J. Kennedy, "Particle swarm optimization," in *Proceedings of the IEEE international conference on neural networks*, vol. 4. Citeseer, 1995, pp. 1942–1948.

[59] M. Dorigo, V. Maniezzo, and A. Colorni, "The ant system: An autocatalytic optimizing process," 1991.

[60] D. Karaboga and B. Akay, "A comparative study of artificial bee colony algorithm," *Appl. Math. Comput.*, vol. 214, no. 1, pp. 108–132, 2009.

[61] Z. Xiao, H. Dong, L. Bai, D. O. Wu, and X.-G. Xia, "Unmanned aerial vehicle base station (UAV-BS) deployment with millimeter-wave beamforming," *IEEE Internet Thing J.*, vol. 7, no. 2, pp. 1336–1349, 2019.

[62] S. Khalid, W. B. Abbas, H. S. Kim, and M. T. Niaz, "Evolutionary algorithm based capacity maximization of 5G/B5G hybrid pre-coding systems," *Sensors*, vol. 20, no. 18, p. 5338, 2020.

[63] D. Karaboga, "An idea based on honey bee swarm for numerical optimization," Citeseer, Tech. Rep., 2005.

[64] H. Li, M. Li, Q. Liu, and A. L. Swindlehurst, "Dynamic hybrid beamforming with low-resolution PSs for wideband mmWave MIMO-OFDM systems," *IEEE J. Sel. Areas Commun.*, vol. 38, no. 9, pp. 2168–2181, 2020.

[65] Y. Saito, Y. Kishiyama, A. Benjebbour, T. Nakamura, A. Li, and K. Higuchi, "Non-orthogonal multiple access (NOMA) for cellular future radio access," in *Proc. IEEE 77th Veh. Technol. Conf. (VTC Spring)*. IEEE, 2013, pp. 1–5.

[66] H. Xiao, Y. Wang, Q. Cheng, and Y. Wang, "An improved PSO-based power allocation algorithm for the optimal EE and SE tradeoff in downlink NOMA systems," in *Proc. IEEE 29th Annu. Int. Symp. Pers. Indoor Mobile Radio Commun. (PIMRC)*. IEEE, 2018, pp. 1–5.

[67] D. Ni, L. Hao, X. Qian, and Q. T. Tran, "Energy-spectral efficiency tradeoff of downlink NOMA system with fairness consideration," in *Proc. IEEE 87th Veh. Technol. Conf. (VTC Spring)*. IEEE, 2018, pp. 1–5.

[68] J. Choi, "Power allocation for max-sum rate and max-min rate proportional fairness in NOMA," *IEEE Commun. Lett.*, vol. 20, no. 10, pp. 2055–2058, 2016.

[69] T. Manglayev, R. C. Kizilirmak, and Y. H. Kho, "Optimum power allocation for non-orthogonal multiple access (NOMA)," in *Proc. IEEE 10th Int. Conf. Appl. Inf. Commun. Technologies*. IEEE, 2016, pp. 1–4.

[70] P. Sindhu, A. H. KM *et al.*, "A novel low complexity power allocation algorithm for downlink NOMA networks," in *Proc. IEEE Recent Adv. Intell. Comput. Syst. (RAICS)*. IEEE, 2018, pp. 36–40.

[71] N. Glei and R. B. Chibani, "Power allocation for energy-efficient downlink NOMA systems," in *n Proc. 19th Int. Conf. Sci. Techn. Autom. Control Comput. Eng. (STA)*. IEEE, 2019, pp. 611–613.

[72] F. Fang, H. Zhang, J. Cheng, and V. C. Leung, "Energy-efficient resource allocation for downlink non-orthogonal multiple access network," *IEEE Trans. Commun.*, vol. 64, no. 9, pp. 3722–3732, 2016.

[73] Y. Zhang, H.-M. Wang, T.-X. Zheng, and Q. Yang, "Energy-efficient transmission design in non-orthogonal multiple access," *IEEE Trans. Veh. Technol.*, vol. 66, no. 3, pp. 2852–2857, 2016.

[74] J. Wang, H. Xu, L. Fan, B. Zhu, and A. Zhou, "Energy-efficient joint power and bandwidth allocation for NOMA systems," *IEEE Commun. Lett.*, vol. 22, no. 4, pp. 780–783, 2018.

[75] Q. Liu, F. Tan, T. Lv, and H. Gao, "Energy efficiency and spectral-efficiency tradeoff in downlink NOMA systems," in *Proc. IEEE Int. Conf. Commun. Workshops.* IEEE, 2017, pp. 247–252.

[76] Y. Zuo, X. Zhu, Y. Jiang, Z. Wei, H. Zeng, and T. Wang, "Energy efficiency and spectral efficiency tradeoff for multicarrier NOMA systems with user fairness," in *Proc. IEEE/CIC Int. Conf. Commun. China (ICCC).* IEEE, 2018, pp. 666–670.

[77] W. Y. Yang, W. Cao, T.-S. Chung, and J. Morris, *Applied numerical methods using MATLAB.* John Wiley & Sons, 2005.

[78] S. K. Goudos, P. D. Diamantoulakis, and G. K. Karagiannidis, "Multi-objective optimization in 5G wireless networks with massive MIMO," *IEEE Commun. Lett.*, vol. 22, no. 11, pp. 2346–2349, 2018.

[79] M. S. Ali, H. Tabassum, and E. Hossain, "Dynamic user clustering and power allocation for uplink and downlink non-orthogonal multiple access (NOMA) systems," *IEEE access*, vol. 4, pp. 6325–6343, 2016.

[80] M. Masoudi, H. Zaefarani, A. Mohammadi, and C. Cavdar, "Energy and spectrum efficient resource allocation in two-tier networks: A multiobjective approach," in *Proc. IEEE Wireless Commun. Netw. Conf. (WCNC).* IEEE, 2017, pp. 1–6.

[81] K. Deb, "Multi-objective optimisation using evolutionary algorithms: an introduction," in *Multi-objective evolutionary optimisation for product design and manufacturing.* Springer, 2011, pp. 3–34.

[82] K. Deb, A. Pratap, S. Agarwal, and T. Meyarivan, "A fast and elitist multiobjective genetic algorithm: NSGA-II," *IEEE Trans. Evol. Comput.*, vol. 6, no. 2, pp. 182–197, 2002.

[83] A. Golchha and S. G. Qureshi, "Non-dominated sorting genetic algorithm-II–A succinct survey," *Int. J. Comput. Sci. Inf. Technol*, vol. 6, no. 1, pp. 252–255, 2015.

[84] E. Zitzler and L. Thiele, "An evolutionary algorithm for multiobjective optimization: The strength pareto approach," *TIK-report*, vol. 43, 1998.

[85] E. Zitzler, M. Laumanns, and L. Thiele, "SPEA2: Improving the strength Pareto evolutionary algorithm," *TIK-report*, vol. 103, 2001.

[86] J. Moore and R. Chapman, "Application of particle swarm to multiobjective optimization," *Department of Computer Science and Software Engineering, Auburn University*, vol. 32, 1999.

[87] M. Series, "Guidelines for evaluation of radio interface technologies for IMT-advanced," *Report ITU*, vol. 638, 2009.

[88] D. Liu, L. Wang, Y. Chen, M. Elkashlan, K.-K. Wong, R. Schober, and L. Hanzo, "User association in 5G networks: A survey and an outlook," *IEEE Commun. Surveys Tuts.*, vol. 18, no. 2, pp. 1018–1044, 2016.

[89] H. Q. Ngo, E. G. Larsson, and T. L. Marzetta, "Energy and spectral efficiency of very large multiuser MIMO systems," *IEEE Trans. Commun.*, vol. 61, no. 4, pp. 1436–1449, 2013.

[90] B. M. Hochwald, T. L. Marzetta, and V. Tarokh, "Multiple-antenna channel hardening and its implications for rate feedback and scheduling," *IEEE Trans. Inf. Theory*, vol. 50, no. 9, pp. 1893–1909, 2004.

[91] D. Liu, L. Wang, Y. Chen, T. Zhang, K. K. Chai, and M. Elkashlan, "Distributed energy efficient fair user association in massive MIMO enabled HetNets," *IEEE Commun. Lett.*, vol. 19, no. 10, pp. 1770–1773, 2015.

[92] S. Corroy, L. Falconetti, and R. Mathar, "Dynamic cell association for downlink sum rate maximization in multi-cell heterogeneous networks," in *Proc. IEEE Int. Conf. Commun. Workshops*. IEEE, 2012, pp. 2457–2461.

[93] K. Shen and W. Yu, "Distributed pricing-based user association for downlink heterogeneous cellular networks," *IEEE J. Sel. Areas Commun.*, vol. 32, no. 6, pp. 1100–1113, 2014.

[94] H. Pervaiz, L. Musavian, and Q. Ni, "Joint user association and energy-efficient resource allocation with minimum-rate constraints in two-tier HetNets," in *Proc. IEEE 24th Annu. Int. Symp. Pers. Indoor Mobile Radio Commun. (PIMRC)*. IEEE, 2013, pp. 1634–1639.

[95] 3rd Generation Partnership Project (3GPP), "3GPP TR#36.872," 2013.

[96] H. Ramazanali, A. Mesodiakaki, A. Vinel, and C. Verikoukis, "Survey of user association in 5G HetNets," in *Proc. 8th IEEE Latin-Amer. Conf. Commun. (LATINCOM)*. IEEE, 2016, pp. 1–6.

[97] J. G. Andrews, S. Singh, Q. Ye, X. Lin, and H. Dhillon, "An overview of load balancing in HetNets: Old myths and open problems," *arXiv preprint arXiv:1307.7779*, 2013.

[98] D. Bethanabhotla, O. Y. Bursalioglu, H. C. Papadopoulos, and G. Caire, "Optimal user-cell association for massive MIMO wireless networks," *IEEE Trans. Wireless Commun.*, vol. 15, no. 3, pp. 1835–1850, 2016.

[99] I. S. Baqer, "Energy efficient cell association in two-tier cellular networks," in *2018 International Conference on Advance of Sustainable Engineering and its Application (ICASEA)*. IEEE, 2018, pp. 79–84.

[100] N. I. Devi, R. Mahapatra, V. K. Reddy, and A. Kohli, "Deployment efficiency vs energy efficiency in dense wireless environment using massive MIMO," in *2016 International Conference on Emerging Trends in Communication Technologies (ETCT)*. IEEE, 2016, pp. 1–5.

[101] Y. Lin, Y. Wang, C. Li, Y. Huang, and L. Yang, "Energy efficient joint user association and power allocation design in massive MIMO empowered dense HetNets," in *Proc. IEEE 84th Veh. Technol. Conf. (VTC-Fall)*. IEEE, 2016, pp. 1–6.

[102] O. Amin, E. Bedeer, M. H. Ahmed, and O. A. Dobre, "Energy efficiency–spectral efficiency tradeoff: A multiobjective optimization approach," *IEEE Trans. Veh. Technol.*, vol. 65, no. 4, pp. 1975–1981, 2015.

[103] C. Yang, J. Li, X. Jiang, and A. Anpalagan, "Interference-aware spectral-and-energy efficiency tradeoff in heterogeneous networks," in *Proc. IEEE Wireless Commun. Netw. Conf. (WCNC)*. IEEE, 2015, pp. 819–824.

[104] Y. Hao, Q. Ni, H. Li, and S. Hou, "Energy and spectral efficiency tradeoff with user association and power coordination in massive MIMO enabled HetNets," *IEEE Commun. Lett.*, vol. 20, no. 10, pp. 2091–2094, 2016.

[105] S. Cetinkaya and H. Arslan, "A distributed user-cell association for spectral and energy efficiency tradeoff in massive MIMO UDHNs," in *Proc. IEEE 30th IEEE Annu. Int. Symp. Pers. Indoor Mobile Radio Commun. (PIMRC)*. IEEE, 2019, pp. 1–6.

[106] S. Cetinkaya, U. S. Hashmi, and A. Imran, "What user-cell association algorithms will perform best in mmWave massive MIMO ultra-dense HetNets?" in *Proc. IEEE 28th Annu. Int. Symp. Pers. Indoor Mobile Radio Commun. (PIMRC)*. IEEE, 2017, pp. 1–7.

[107] A. Molisch, *Wireless Communications*, ser. Wiley - IEEE. Wiley, 2010. [Online]. Available: https://books.google.com/books?id=vASyH5-jfMYC

[108] B. M. Hochwald, T. L. Marzetta, and V. Tarokh, "Multi-antenna channel hardening and its implications for rate feedback and scheduling," *IEEE Trans. Inform. Theory*, vol. 50, no. 9, pp. 1893–1909, 2004.

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

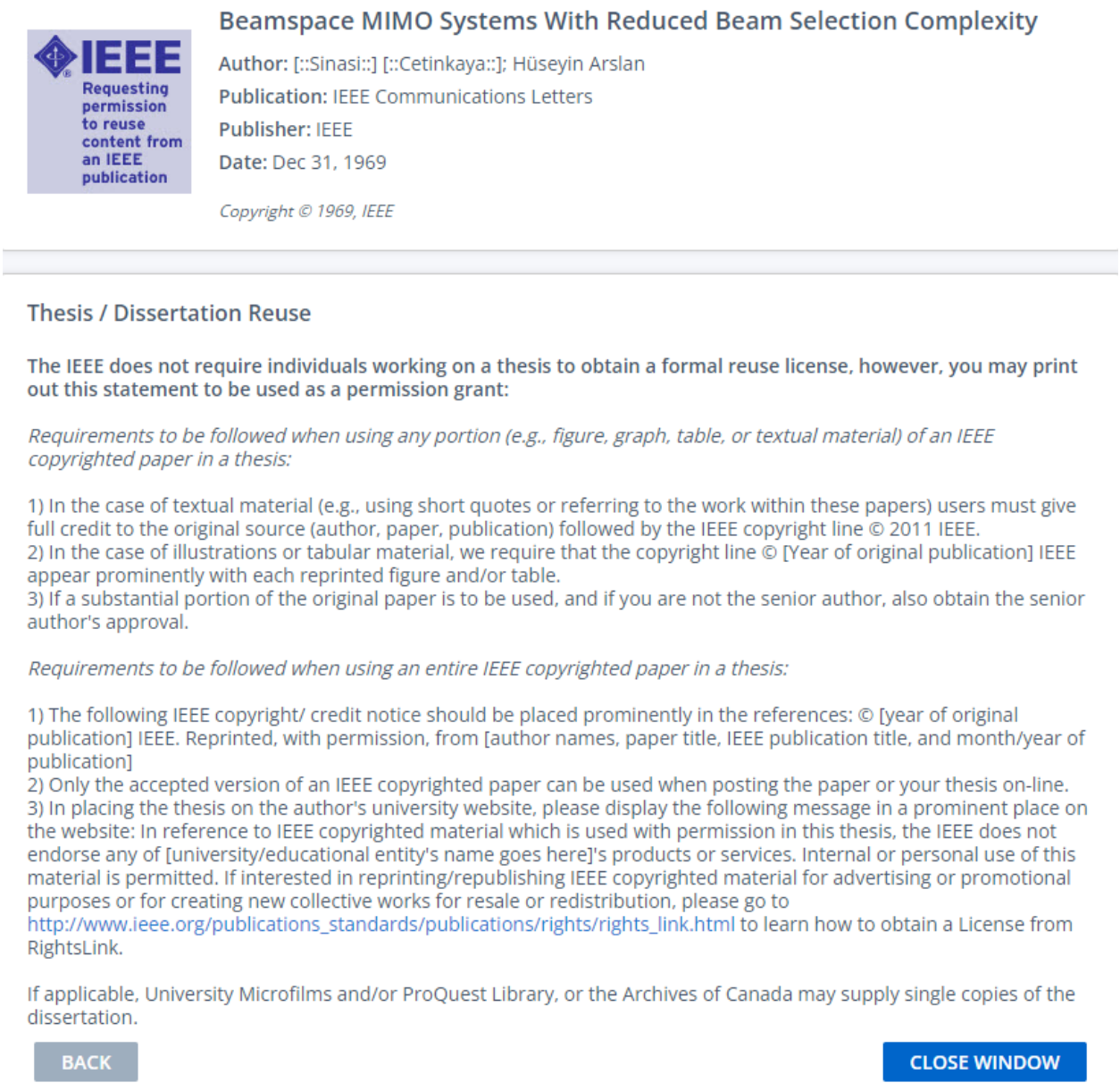
[110] D. Bethanabhotla, O. Y. Bursalioglu, H. C. Papadopoulos, and G. Caire, "User association and load balancing for cellular massive mimo." in *ITA*, 2014, pp. 1–10.

[111] G. Auer, V. Giannini, C. Desset, I. Godor, P. Skillermark, M. Olsson, M. A. Imran, D. Sabella, M. J. Gonzalez, O. Blume *et al.*, "How much energy is needed to run a wireless network?" *IEEE Wireless Commun.*, vol. 18, no. 5, 2011.

[112] E. Björnson, L. Sanguinetti, J. Hoydis, and M. Debbah, "Designing multi-user MIMO for energy efficiency: When is massive MIMO the answer?" in *Proc. IEEE Wireless Commun. Netw. Conf. (WCNC)*. IEEE, 2014, pp. 242–247.

[113] S. S. Rao, *Engineering optimization: theory and practice*. John Wiley & Sons, 2009.

Appendix A: Copyright Permissions



IEEE
Requesting permission to reuse content from an IEEE publication

Beamspace MIMO Systems With Reduced Beam Selection Complexity

Author: [::Sinasi::] [::Cetinkaya::]; Hüseyin Arslan
Publication: IEEE Communications Letters
Publisher: IEEE
Date: Dec 31, 1969

Copyright © 1969, IEEE

Thesis / Dissertation Reuse

The IEEE does not require individuals working on a thesis to obtain a formal reuse license, however, you may print out this statement to be used as a permission grant:

Requirements to be followed when using any portion (e.g., figure, graph, table, or textual material) of an IEEE copyrighted paper in a thesis:

1) In the case of textual material (e.g., using short quotes or referring to the work within these papers) users must give full credit to the original source (author, paper, publication) followed by the IEEE copyright line © 2011 IEEE.
2) In the case of illustrations or tabular material, we require that the copyright line © [Year of original publication] IEEE appear prominently with each reprinted figure and/or table.
3) If a substantial portion of the original paper is to be used, and if you are not the senior author, also obtain the senior author's approval.

Requirements to be followed when using an entire IEEE copyrighted paper in a thesis:

1) The following IEEE copyright/ credit notice should be placed prominently in the references: © [year of original publication] IEEE. Reprinted, with permission, from [author names, paper title, IEEE publication title, and month/year of publication]
2) Only the accepted version of an IEEE copyrighted paper can be used when posting the paper or your thesis on-line.
3) In placing the thesis on the author's university website, please display the following message in a prominent place on the website: In reference to IEEE copyrighted material which is used with permission in this thesis, the IEEE does not endorse any of [university/educational entity's name goes here]'s products or services. Internal or personal use of this material is permitted. If interested in reprinting/republishing IEEE copyrighted material for advertising or promotional purposes or for creating new collective works for resale or redistribution, please go to http://www.ieee.org/publications_standards/publications/rights/rights_link.html to learn how to obtain a License from RightsLink.

If applicable, University Microfilms and/or ProQuest Library, or the Archives of Canada may supply single copies of the dissertation.

BACK CLOSE WINDOW

The permission above is for the reproduction of material in Chapter 2.

Heuristic Inspired Precoding for Millimeter-Wave MIMO Systems with Lens Antenna Subarrays

Conference Proceedings: 2022 IEEE 95th Vehicular Technology Conference: (VTC2022-Spring)

Author: [::Sinasi::] [::Cetinkaya::]; Liza Afeef; Gokhan Mumcu; Huseyin Arslan

Publisher: IEEE

Date: 19-22 June 2022

Thesis / Dissertation Reuse

The IEEE does not require individuals working on a thesis to obtain a formal reuse license, however, you may print out this statement to be used as a permission grant:

Requirements to be followed when using any portion (e.g., figure, graph, table, or textual material) of an IEEE copyrighted paper in a thesis:

1) In the case of textual material (e.g., using short quotes or referring to the work within these papers) users must give full credit to the original source (author, paper, publication) followed by the IEEE copyright line © 2011 IEEE.
2) In the case of illustrations or tabular material, we require that the copyright line © [Year of original publication] IEEE appear prominently with each reprinted figure and/or table.
3) If a substantial portion of the original paper is to be used, and if you are not the senior author, also obtain the senior author's approval.

Requirements to be followed when using an entire IEEE copyrighted paper in a thesis:

1) The following IEEE copyright/ credit notice should be placed prominently in the references: © [year of original publication] IEEE. Reprinted, with permission, from [author names, paper title, IEEE publication title, and month/year of publication]
2) Only the accepted version of an IEEE copyrighted paper can be used when posting the paper or your thesis on-line.
3) In placing the thesis on the author's university website, please display the following message in a prominent place on the website: In reference to IEEE copyrighted material which is used with permission in this thesis, the IEEE does not endorse any of [university/educational entity's name goes here]'s products or services. Internal or personal use of this material is permitted. If interested in reprinting/republishing IEEE copyrighted material for advertising or promotional purposes or for creating new collective works for resale or redistribution, please go to http://www.ieee.org/publications_standards/publications/rights/rights_link.html to learn how to obtain a License from RightsLink.

If applicable, University Microfilms and/or ProQuest Library, or the Archives of Canada may supply single copies of the dissertation.

BACK CLOSE WINDOW

The permission above is for the reproduction of material in Chapter 3.

Energy and Spectral Efficiency Tradeoff in NOMA: Multi-Objective Evolutionary Approaches

Conference Proceedings:
2020 IEEE International Conference on Communications Workshops (ICC Workshops)

Author: [::Sinasi::] [::Cetinkaya::]; Huseyin Arslan

Publisher: IEEE

Date: 7-11 June 2020

Thesis / Dissertation Reuse

The IEEE does not require individuals working on a thesis to obtain a formal reuse license, however, you may print out this statement to be used as a permission grant:

Requirements to be followed when using any portion (e.g., figure, graph, table, or textual material) of an IEEE copyrighted paper in a thesis:

1) In the case of textual material (e.g., using short quotes or referring to the work within these papers) users must give full credit to the original source (author, paper, publication) followed by the IEEE copyright line © 2011 IEEE.
2) In the case of illustrations or tabular material, we require that the copyright line © [Year of original publication] IEEE appear prominently with each reprinted figure and/or table.
3) If a substantial portion of the original paper is to be used, and if you are not the senior author, also obtain the senior author's approval.

Requirements to be followed when using an entire IEEE copyrighted paper in a thesis:

1) The following IEEE copyright/ credit notice should be placed prominently in the references: © [year of original publication] IEEE. Reprinted, with permission, from [author names, paper title, IEEE publication title, and month/year of publication]
2) Only the accepted version of an IEEE copyrighted paper can be used when posting the paper or your thesis on-line.
3) In placing the thesis on the author's university website, please display the following message in a prominent place on the website: In reference to IEEE copyrighted material which is used with permission in this thesis, the IEEE does not endorse any of [university/educational entity's name goes here]'s products or services. Internal or personal use of this material is permitted. If interested in reprinting/republishing IEEE copyrighted material for advertising or promotional purposes or for creating new collective works for resale or redistribution, please go to http://www.ieee.org/publications_standards/publications/rights/rights_link.html to learn how to obtain a License from RightsLink.

If applicable, University Microfilms and/or ProQuest Library, or the Archives of Canada may supply single copies of the dissertation.

BACK CLOSE WINDOW

The permission above is for the reproduction of material in Chapter 4.

A Distributed User-Cell Association for Spectral and Energy Efficiency Tradeoff in Massive MIMO UDHNs

Conference Proceedings:
2019 IEEE 30th Annual International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC)

Author: [::Sinasi::] [::Cetinkaya::]; Huseyin Arslan

Publisher: IEEE

Date: 8-11 Sept. 2019

Thesis / Dissertation Reuse

The IEEE does not require individuals working on a thesis to obtain a formal reuse license, however, you may print out this statement to be used as a permission grant:

Requirements to be followed when using any portion (e.g., figure, graph, table, or textual material) of an IEEE copyrighted paper in a thesis:

1) In the case of textual material (e.g., using short quotes or referring to the work within these papers) users must give full credit to the original source (author, paper, publication) followed by the IEEE copyright line © 2011 IEEE.
2) In the case of illustrations or tabular material, we require that the copyright line © [Year of original publication] IEEE appear prominently with each reprinted figure and/or table.
3) If a substantial portion of the original paper is to be used, and if you are not the senior author, also obtain the senior author's approval.

Requirements to be followed when using an entire IEEE copyrighted paper in a thesis:

1) The following IEEE copyright/ credit notice should be placed prominently in the references: © [year of original publication] IEEE. Reprinted, with permission, from [author names, paper title, IEEE publication title, and month/year of publication]
2) Only the accepted version of an IEEE copyrighted paper can be used when posting the paper or your thesis on-line.
3) In placing the thesis on the author's university website, please display the following message in a prominent place on the website: In reference to IEEE copyrighted material which is used with permission in this thesis, the IEEE does not endorse any of [university/educational entity's name goes here]'s products or services. Internal or personal use of this material is permitted. If interested in reprinting/republishing IEEE copyrighted material for advertising or promotional purposes or for creating new collective works for resale or redistribution, please go to http://www.ieee.org/publications_standards/publications/rights/rights_link.html to learn how to obtain a License from RightsLink.

If applicable, University Microfilms and/or ProQuest Library, or the Archives of Canada may supply single copies of the dissertation.

BACK CLOSE WINDOW

The permission above is for the reproduction of material in Chapter 5.

About the Author

Sinasi Cetinkaya received his B.S. degree in electrical and electronics engineering from Ondokuz Mayis University, Samsun, Turkey, in 2013 and his M.S. degree in telecommunication engineering from the University of Oklahoma, Tulsa, in 2017. He is currently pursuing his Ph.D. degree as a member of the Wireless Communication and Signal Processing (WCSP) Group in the Department of Electrical Engineering, University of South Florida, Tampa. His research interests are in the fields of wireless communication and channel modeling, multiple-input multiple-output (MIMO) beamforming, millimeter-wave communication, lens antenna arrays, non-orthogonal multiple access (NOMA), and user-cell association and resource allocation. He is a graduate student member of IEEE.