# A POMDP FRAMEWORK FOR ANTENNA SELECTION AND USER SCHEDULING IN MULTI-USER MASSIVE MIMO SYSTEMS

by

Sara Sharifi

A thesis submitted to the
School of Graduate and Postdoctoral Studies in partial
fulfillment of the requirements for the degree of

**DOCTOR OF PHILOSOPHY IN**
**ELECTRICAL AND COMPUTER ENGINEERING**

Faculty of Engineering and Applied Science
University of Ontario Institute of Technology (Ontario Tech University)
Oshawa, Ontario, Canada

April 2022

# THESIS EXAMINATION INFORMATION

Submitted by: **Sara Sharifi**

**Doctor of Philosophy in Electrical and Computer Engineering**

**Thesis title**: A POMDP FRAMEWORK FOR ANTENNA SELECTION AND USER SCHEDULING IN MULTI-USER MASSIVE MIMO SYSTEMS

An oral defense of this thesis took place on April 14, 2022 in front of the following examining committee:

**Examining Committee:**

Chair of Examining Committee: Prof. Khalid Elgazzar

Research Supervisor: Prof. Shahram ShahbazPanahi

Research Co-supervisor: Prof. Ali Grami

Examining Committee Member: Prof. Min Dong

Examining Committee Member: Prof. Shahryar Rahnamayan

University Examiner: Prof. Mehran Ebrahimi

External Examiner: Prof. Alagan Anpalagan, Department of Electrical, Computer and Biomedical Engineering, Ryerson University, Toronto, ON, Canada

The examining committee determined that the thesis is acceptable in form and content and that a satisfactory knowledge of the field covered by the thesis was demonstrated by the candidate during an oral examination. A signed copy of the Certificate of Approval is available from the School of Graduate and Postdoctoral Studies.

# ABSTRACT

We use a partially observable Markov decision process (POMDP) framework to design a resource allocation policy for downlink transmit beamforming at a multi-antenna BS that is equipped with a massive number of antennas and only a limited number of RF chains. Considering that channels evolve according to a Markov process and that only partial CSI is available, we use a POMDP framework for antenna selection with the aim to maximize the expected long-term data rate. To avoid the high computational complexity of the value iteration algorithm, we focus on the myopic policy to design a simple yet optimal algorithm. We prove that in the case of a positively correlated two-state Markov channel model, the myopic policy is optimal for antenna selection (for both in massive MISO and MU-MIMO systems) for any number of RF chains. Based on this finding, for general fading channels, we propose to quantize each channel into two levels and apply the myopic policy for antenna selection. Our simulation results show that using this two-level channel quantization for antenna selection results in only a small loss in performance, as compared to the antenna selection technique which use full CSI without quantization. We then utilize a POMDP framework to formulate the joint antenna selection and user scheduling (JASUS) problem for a BS, equipped with a limited number of RF chains that is to serve a large number of single-antenna users in a cell. To do so, we assume that the users are served in a frame, where each frame contains of a finite number of time slots. At the beginning of each frame, given that only partial CSI is available, the BS schedules each user to a time slot, and selects a subset of antennas to serve the scheduled users at that time slot. Considering a positively correlated two-state channel model, we prove the optimality of the myopic policy for our JASUS problem. For Rayleigh fading channels, we devise a low-complexity JASUS algorithm for massive MU-MIMO systems.

**Keywords**: *Antenna selection; joint antenna selection and user scheduling; massive MIMO; partially observable Markov decision process (POMDP); myopic policy*

# AUTHORS DECLARATION

I, Sara Sharifi, hereby declare that this thesis consists of original work of which I have authored. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners I authorize the University of Ontario Institute of Technology (Ontario Tech University) to lend this thesis to other institutions or individuals for the purpose of scholarly research. I further authorize University of Ontario Institute of Technology (Ontario Tech University) to reproduce this thesis by photocopying or by other means, in total or in part, at the request of other institutions or individuals for the purpose of scholarly research. I understand that my thesis will be made electronically available to the public.

________________________________ **Sara Sharifi**

# STATEMENT OF CONTRIBUTIONS

Part of this dissertation (described in Chapters 4) have been published as:

- S. Sharifi, S. ShahbazPanahi, and M. Dong, “A POMDP based antenna selection for massive MIMO communication,” *IEEE Trans.Commun.*, vol. 70, no. 3, pp. 20252041, March 2022.
- S. Sharifi, S. ShahbazPanahi, and M. Dong, “Antenna selection for massive MIMO systems based on POMDP framework,” *ICASSP 2021 - 2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2021, pp. 4450-4454.

I hereby certify that I am the sole author of this thesis and that no part of this thesis has been published or submitted for publication. I have used standard referencing practices to acknowledge ideas, research techniques, or other materials that belong to others. Furthermore, I hereby certify that I am the sole source of the creative works and/or inventive knowledge described in this thesis.

# ACKNOWLEDGMENT

First and foremost, I offer my sincerest gratitude to my supervisor, Professor Shahram ShahbazPanahi. Throughout my PhD studies, he has been an ideal supervisor, skillful mentor, and ultimate role model for me. I cannot express how grateful I am for his invaluable feedback, constructive criticism, and insightful advice for my research. Also, I would like to express my appreciation and gratitude to Professor Ali Grami for his support during my studies.

Further, I would like to thank Professor Min Dong, for her guidance and assistance. This dissertation could not have been completed without her help. My gratitude would not be complete without acknowledging my parents, brothers, sister, and my fiance for their non-stop love, encouragement, help and support. They were the main source of my power to pass through all the ups and downs in this long journey, and I am forever grateful.

# Contents

# List of Tables

# List of Figures

# List of Acronyms

| | |
|---|---|
| **bcu** | bits per channel use |
| **BS** | Base Station |
| **CSI** | Channel State Information |
| **FDD** | Frequency Division Duplex |
| **JASUS** | Joint Antenna Selection and User Scheduling |
| **JAUS** | Joint Antenna and User Selection |
| **ICSI** | Imperfect Channel State Information |
| **MDP** | Markov Decision Process |
| **MIMO** | Multiple Input Multiple Output |
| **MISO** | Multiple Input Single Output |
| **MU-MIMO** | Multi-user, Multiple-input, and Multiple-output |
| **PCSI** | Perfect Channel State Information |
| **POMDP** | Partially Observable Markov Decision Process |
| **QoS** | Quality of Service |
| **RF** | Radio Frequency |
| **SNR** | Signal-to-Noise Ratio |
| **TDD** | Time Division Duplex |

# Chapter 1

# Introduction

## 1.1 Overview

Massive multi-input multi-output (MIMO) systems play an important role in the 5-th generation wireless networks as such systems provide the ability to serve multiple users using the same time and frequency resource blocks. The extensive studies conducted on massive MIMO systems show that deploying large-scale antenna arrays at the base stations (BSs) increases the achievable sum-rate and improves the performance in terms of spectral and energy efficiency [1–4]. However, increasing the number of RF chain, exponentially increases the hardware complexity, cost, and computational complexity [5]. Thus, it may not be practical to have a dedicated RF chain per antenna element in the massive MIMO BSs. To benefit from the advantages offered by using large-scale antenna arrays at the BSs, and at the same time, to overcome hardware complexity, antenna selection techniques have been proposed in the literature [6–9]. Antenna selection is a decision-making technique that selects a subset of available antennas to transmit data at each time slot. By applying the antenna selection techniques, the number of required RF chains can be reduced to the number of selected antennas, leading to reduced RF circuit power consumption, size, price, and hardware complexity [6, 8, 10]. The extensive studies conducted in this area show that when the number of available antennas is more than the number of RF chains, the antenna selection technique can improve data rate and energy efficiency, compared with a system where the number of antennas is as limited as the same number of RF chains [11, 12]. Many studies have aimed at proposing antenna selection algorithms that maximize the sum-rate for massive MIMO systems [13–19]. However, there are still some critical issues that must be addressed. One such is the assumption of the availability of full CSI for designing their algorithms. More technical details are provided in Section 1.2. To overcome the issues and challenges of the existing antenna selection algorithms, we explain our proposed methodology in Section 1.3, for both MISO systems and multi-user MIMO systems. Note that in multi-user massive MIMO schemes, zero-forcing beamforming is often proposed to null the inter-user interference. To fully cancel out the inter-user interference, the number of served users should be less than the number of active antennas (here, the number of active antennas is the same as the number of RF chains). Thus, when the number of users is larger than the number of RF chains, the joint antenna selection and users scheduling (JASUS) needs to be addressed in multi-user massive MIMO systems. Note that there aren't any existing studies that design a JASUS algorithm for massive MIMO systems. However, there are limited existing studies that only addressed the joint antenna and user selection (JAUS) problems in massive MIMO systems. The details of the existing JAUS algorithms are explained in Chapter 2. We also explain why the proposed JAUS algorithms are not applicable for our JASUS problem in Section 1.2.

In this dissertation, considering that the channels evolve according to a finite state Markov process, we first aim to design POMDP-based antenna selection algorithms for both massive MISO, and multi-user massive MIMO systems. We then, assume a scenario where the number of available users is larger than the number of active antennas, meaning that serving all users at the same time results in low quality of service. Thus, to maintain the high quality of service in multi-user massive MIMO systems that the BS (equipped with limited RF chains) serves a large number of users, joint antenna selection and user scheduling (JASUS) is required.

To do so, we assume that users receive data in a frame, which each frame contains a finite number of time slots. At the beginning of each frame, the BS schedules each user to a time slot in a frame and selects a subset of antennas to serve the scheduled users at that time slot. Note that the number of scheduled users at each time slot is bigger than one and less than the number of RF chains. Here, we aim to design a JASUS algorithm in multi-user massive MIMO systems for time-varying channels when only partial CSI is available at the BS.

In the remaining of this chapter, we first present the challenges of the antenna selection and JASUS techniques and elaborate on what motivated us to conduct this research. Later on, for each scenario (i.e., antenna selection in massive MISO systems, antenna selection in multi-user massive MIMO systems, and JASUS in multi-user massive MIMO systems), we define the objective function of the corresponding problem and present the proposed method to solve it.

## 1.2 Challenges and Motivations

For designing the antenna selection algorithms for massive MU-MIMO systems, one common assumption is that the channels over all antennas are *fully observable,* meaning that full channel state information (CSI) can be obtained or estimated. For massive MIMO systems, this assumption is not practical, since it requires either an RF chain for each antenna or switching available RF chains among antennas for training and channel estimation. With limited RF chains, the latter approach adds to the time required for channel estimation and, at the same time, complicates the hardware by adding switching circuitry. Moreover, switching RF chains exacerbates the problem of outdated CSI. Another common assumption in the existing studies of the antenna selection problem is that the channels remain over time slot [20–23]. As a result, in previous studies, the properties of time-varying fading channels have not been completely exploited [24].

Given the above discussion, our motivation is to study the problem of antenna selection for a base station (BS) downlink transmission to a user in a massive MISO (or multiple users in massive MU-MIMO ) system under the assumption that only a finite number of RF chains are available for transmission and CSI acquisition. More specifically, given that at each time slot only partial CSI (that is acquired over the previously selected set of antennas) is available, the BS decides which antennas to select and sends data in the next downlink time slot via transmit beamforming over those selected antennas. Under the assumption that the channel over each antenna evolves according to a Markov chain, this problem can be formulated using a partially observed Markov decision process (POMDP) framework. In the literature, there are extensive studies that utilize a POMDP framework to design resource allocation algorithms (see Chapter 2). However, the designed POMDP-based algorithms for antenna selection can be applied to only a limited number of antennas and RF chains, when channels evolve as a two-state channel model (see Chapter 2 for more details). Here, for time-varying continuous-valued channels, we aim to design a POMDP-based antenna selection algorithm, where the decisions are made based on partial CSI and can be applied to any number of antennas and RF chains. Our objective is to *maximize the expected long-term sum-rate*. Note that in multi-user massive MIMO systems, zero-forcing beamforming is often used to nullify the inter-user interference. To fully cancel out inter-user interference, the number of users should be less than the number of RF chains. Therefore, when a large number of users is available in the cell (i.e., the number of users is more than the number of RF chains), we must address joint antenna selection and user scheduling (JASUS) in multi-user massive MIMO systems. To the best of our knowledge, currently there is no study to design JASUS algorithms for time-varying channels, such that the decisions are made based on partial CSI (for more details see Chapter 2). In this dissertation, motivated by the above explanations, we formulate the JASUS problems using a POMDP framework to design a low-complexity JASUS algorithm that can be applied to actual Rayleigh fading channels, when only partial CSI is available at the BS. In the next section, we define our resource allocation problems (i.e., antenna selection in massive MISO systems, antenna selection in multi-user massive MIMO systems, and JASUS in massive MIMO systems), and explain our proposed methodology to design a low-complexity algorithm for each one of the defined problems.

## 1.3 Objective and Methodology

In this section, we provide an overview on the main objective and proposed methodology of our study presented in each chapter of this dissertation.

### 1.3.1 Antenna Selection in Massive MISO Systems

**Objective**

In Chapter 4, we consider the antenna selection design for the BS downlink transmission to a user in a massive MIMO system under the assumption that *only a finite number of RF chains is available for transmission and CSI acquisition.* We assume the system operates in the time division duplexing (TDD) mode, and therefore, the CSI acquisition for downlink transmission is performed using the uplink channel measurements. Based on the partial CSI and the history of the CSI of other antennas, the BS makes the new antenna selection decision and sends data in the next downlink time slot via transmit beamforming. Given the underlying fading channels are correlated over time and evolve according to a Markov chain, we aim to find the optimal decision for antenna selection at each time slot under partial CSI with the goal of *maximizing the expected long-term MISO data rate.*

**Methodology**

In Chapter 4, we use the POMDP framework to devise an optimal antenna selection policy for the BS transmit beamforming to a single-antenna user. Assuming a TDD system with uplink-downlink channel reciprocity and that the MISO channel state evolves according to a finite-state Markov process, we exploit partial observation of the channel coefficients from uplink CSI training. Thus, at each time slot, the BS uses the obtained partial CSI (the channel coefficients of the previous selected antennas) for antenna selection to *maximize the long-term expected data rate* achieved in the downlink transmission. The solution to this POMDP-based dynamic antenna selection problem can be obtained using the value iteration algorithm. However, the computational complexity of this algorithm is very high for practical implementation, specially for a large state space and/or for a large number of antennas. As such, the myopic policy could offer a computationally attractive solution. While a myopic policy may not always be optimal, we prove rigorously that *if the channels over different BS antennas are independent and evolve according to the same positively correlated two-state Markov process, then the myopic policy is optimal for antenna selection under any number of RF chains.* We obtain this conclusion by showing that the expected long-term data rate is a regular function of the belief vector. To benefit from the optimality of the myopic policy for general fading channels, we propose an antenna selection algorithm such that each channel coefficient is quantized into two levels only for the purpose of antenna selection. We study the impact of the quantization threshold value on the performance of the proposed myopic policy algorithm for antenna selection. We show that, for time-correlated slow fading channels (modeled as a first-order Gauss-Markov process) and for a properly chosen threshold value, the performance of the myopic policy is close to the antenna selection scheme which uses full perfect CSI. Finally, we evaluate the performance of the proposed myopic policy for the imperfect CSI scenario in the presence of channel estimation error.

### 1.3.2 Antenna Selection in Massive MU-MIMO Systems

#### Objective

In Chapter 5, considering the TDD mode, we aim to design an antenna selection policy for a multi-user massive MIMO BS downlink transmission to *multiple single-antenna users* under the assumption that the BS is equipped with a finite number of RF chains. Thus, at each time slot only a subset of antennas is available for data transmission and CSI acquisition from downlink transmission. At each time slot, using the obtained partial CSI from the selected antennas and the history of the CSI of other antennas, the BS selects a new subset of antennas to participate in data transmission in this time slot. Note that here we assume the number of users are less than the number of RF chains. In Chapter 5, we aim to design a real-time decision making algorithm for the antenna selection problem, under obtained partial CSI, such that by selecting the best subset of antennas at each time slot, we can maximize the expected long-term sum-rate.

#### Methodology

In Chapter 5, considering that the underlying fading channels evolve according to a Markov chain, we formulate this problem using a POMDP framework. Here, zero-forcing beamforming is used to null the inter-user interference. In our defined POMDP-based antenna selection problem the reward function is defined as the upper bound of the achievable sum-rate. Furthermore, we assume i.i.d positively correlated two-state channel model. We then first prove the optimality of the myopic policy for our defined antenna selection problem and then propose a novel antenna selection algorithm that can be implemented for the Rayleigh fading channel model. To do so, we propose to quantize the channel gain of each antenna into two levels only for the sake of antenna selection, while all the performance evaluation is based on the non-quantized channel coefficients. In addition to that, we propose an offline learning algorithm to obtain a look-up table for finding the optimal threshold value for channel gain quantization.

### 1.3.3 JASUS in Massive MU-MIMO Systems

**Objective**

In Chapter 6, we study the joint antenna and user scheduling (JASUS) problem for a multi-user massive MIMO system, in which a BS is equipped with a massive number of antennas and a limited number of RF chains to serve *a large number of single-antenna users* in a cell. The number of available users is larger than the number of available RF chains. Here, we assume that the system operates in TDD mode and CSI can be obtained from the uplink measurements. Furthermore, we assume that users receive data in a time frame, where each frame contains of a finite number of time slots. In addition to that, we assume that the channels evolves according to a Markov chain at the beginning of each frame and remains unchanged during the entire time frame. Note that, at each time slot only partial CSI is available due to the limited number of RF chains. In chapter 6, we aim to design a low-complexity JASUS algorithm (that at the beginning of each frame, schedule each user in a time slot, and selected the optimal subset of antennas to serve scheduled users at that time slot) to *maximize the expected long-term sum-rate over frame*.

**Methodology**

In Chapter 6, assuming that channels evolve according to a finite-state Markov process, and only partial CSI is available, we formulate the JASUS problem using a POMDP framework with the main goal of maximizing the expected long-term sum-rate. Here, we assume that the users are served in a frame, and each frame contains a finite number of time slots. According to our proposed algorithm, based on available partial CSI and the history of our past actions and channel observations, at the beginning of each frame, the BS schedules each user to a time slot in a frame to be served and select a subset of antennas to participate in data transmission at each time slot. Note that we assume the state evolves at the beginning of each frame and remains unchanged during the entire frame. We further assume that all users receive data once until the end of the frame such that the number of scheduled users at each time slot is bigger than one and less than the number of RF chains. We show that for positively correlated two-state channel models, the myopic policy provides the optimal solution to our JASUS problem. Furthermore, we use a first-order Gauss Markov channel model and devise a myopic policy algorithm for Rayleigh fading channels that provides a low-complexity suboptimal solution for JASUS problem in multi-user massive MIMO system.

## 1.4 Summary of Contributions

The main contribution of this study is reducing the hardware complexity and cost of the massive MIMO BS in the 5G network by designing a simple yet optimal POMDP-based antenna selection algorithm and POMDP-based JASUS algorithm, that can be easily implemented for large-scale antenna arrays with any given number of antennas and available RF chains. Furthermore, given that at each time slot only a subset of antennas is available for data transmission (partial CSI is available), we aim to design an antenna selection/JASUS policy that provides a high quality of service for the available users. More specifically, given that the underlying fading channels are correlated over time, and only partial CSI is available, at each time slot the BS makes an optimal decision (selecting a subset of antennas in antenna selection policy or selecting a subset of antennas and scheduling users in JASUS policy) to maximize the expected long-term sum rate. To do so, considering that channels evolve as a finite-state Markov process, we use a POMDP framework to formulate our resource allocation problems (antenna selection/JASUS). Here, we use the myopic policy to propose a computationally affordable resource allocation algorithms. In the following sequel, we present the contributions of our studies.

- In the first part of our study, we use a POMDP framework to formulate the antenna selection problem for uplink/downlink massive MISO schemes, where the channel coefficients evolve according to a Markov process. We show that *if the channels over different BS antennas are independent and evolve according to the same positively correlated two-state Markov process, then the myopic policy is optimal for antenna selection under any number of RF chains.* We obtain this conclusion by showing that the expected long-term data rate is a regular function of the belief vector. Note that here, unlike all the previous studies that consider a simple reward function in the POMDP formulation of their resource allocation problems (see Chapter 2), we define the actual data rate as the reward function (which is a complex and realistic function). We then utilize the optimality of the myopic policy to devise an efficient POMDP-based antenna selection technique for time-varying continuous fading channels. To do so, we propose to quantize each channel coefficient into two levels only for the purpose of antenna selection. Interestingly, our simulation results show that using this two-level coarse channel quantization for antenna selection results in a small performance loss that is only within 0.5 (bcu), from the upper bound which can be achieved only by using full (non-quantized) CSI for antenna selection. We also evaluate the affect of the channel estimation on the performance of our antenna selection algorithm.

- We formulate the antennas selection problem for a BS (equipped with a massive number of antennas and a limited number of RF chains) that serves a multiple single-antenna users by utilizing a POMDP framework. In this scenario, we assume that the number of available users is less than the number of RF chains. Here, we use zero-forcing beamforming to eliminate the inter-user interference. In this scenario, an antenna selection policy, which accounts for the channel quality of all users is needed. Here, we assume that channels evolve according to a Markov process. Note that for a positively correlated two-state channel model, the second condition of the optimality of the myopic policy is the regularity of the reward function. However, here the obtained sum rate is not a regular function anymore. In our POMDP formulation, we define the reward function as the upper-bound achievable sum rate, and show that the expected immediate reward function is regular. Thus, *the myopic policy can provide the optimal solution for POMDP-based antenna selection for any number of antennas, any number of RF chains, and any number of users.* We propose a low-complexity myopic policy antenna selection algorithm that can be implemented for Rayleigh fading channels. According to our proposed algorithm, in the selection stage, the channel gain of each antenna is quantized to two levels, allowing us to benefit from the optimality of the myopic policy for i.i.d positively correlated two-state channel models. In addition to that, we propose an offline learning algorithm to obtain a look-up table for finding the optimal threshold value for channel gain quantization. To the best of our knowledge, this is the first study, that propose a low-complexity antenna selection algorithms that its decisions are only relies on partial CSI that can be implemented to the actual Rayleigh fading channels.

- We study the problem of joint antenna selection and user scheduling (JASUS) problem in multi-user massive MIMO systems. In this scenario, we assume that a BS, equipped with a massive number of antennas and a limited number of RF chains, is to serve a large number of single-antenna users in a cell (the number of available users is larger than the number of RF chains). Note that in multi-user MIMO systems, zero-forcing beamforming is often used to eliminate inter-user interference, meaning that at each time slot the number of served users should be equal to or less than the number of RF chains. Thus, when a large number of users is available in a cell (and the BS is equipped with limited RF chains) JASUS must be addressed. We use a POMDP framework to formulate the JASUS problem for a multi-user massive MIMO system, where the number of available users is larger than the number of RF chains. To guarantee that all users will receive data, we assume that users are served in a frame, where each frame contains of a finite number of time slots. In our POMDP formulation, we define the reward function and the objective function as the upper-bound achievable data rate and the expected long term sum-rate over frame, respectively. Assuming that at the beginning of each frame channels evolve according to a positively correlated two-state Markov chain, and remain unchanged during the entire frame, we show that *the myopic policy provides the optimal solution to our JASUS POMDP-based problem.* Furthermore, we model the Rayleigh fading channels as a first-order Gauss Markov channel model and devise a low-complexity myopic policy JASUS algorithm for multi-user massive MIMO systems.

## 1.5 List of Publications

The following publications have either been published or under preparation at the time of writing this dissertation.

1. S. Sharifi, and S. ShahbazPanahi, “A POMDP based antenna selection and user scheduling for massive MU-MIMO communication,” to be submitted to Transactions on Wireless Communications.

2. S. Sharifi, S. ShahbazPanahi, and M. Dong, “A POMDP based antenna selection for massive MU-MIMO communication,” submitted to Transactions on Wireless Communications, 2022.

3. S. Sharifi, S. ShahbazPanahi, and M. Dong, “A POMDP based antenna selection for massive MIMO communication,” *IEEE Trans.Commun.*, vol. 70, no. 3, pp. 20252041, March 2022.

4. S. Sharifi, S. Shahbaz Panahi and M. Dong, “Antenna selection for massive MIMO systems based on POMDP framework,” *ICASSP 2021 - 2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2021, pp. 4450-4454.

## 1.6 Outline of Dissertation

This dissertation is organized as follows. In Chapter 2, we first review the traditional small-scale antenna selection algorithms, and then, provide a brief overview on several existing studies on suboptimal solutions for large antenna selection problems under the assumption of the availability of full CSI. In Section 2.2 of this chapter, we present a survey on studies that existing antenna selection algorithms when only partial CSI is available. In Section 2.3, we overview some other resource allocation algorithms that aim to design a decision making algorithm under the assumption of partial CSI using POMDP framework. And finally, in Section 2.4, we review the limited existing studies on joint antenna selection and user scheduling problem in massive MU-MIMO systems.

In Chapter 3, we first describe the POMDP model and the required tuple to define a POMDP formulation. We then illustrate the policy, the objective function of the stochastic optimization problem in a general POMDP framework. In the final section of this chapter, we present the value iteration algorithm which provides the optimal solution for a POMDP problem.

In Chapter 4, we describe the MISO system model and the POMDP formulation for the antenna selection problem in Sections 4.1 and 4.2, respectively. In Section 4.3, we show the optimality of the myopic policy for a two-state positively correlated channel model. In Section 4.4, we present the myopic policy for antenna selection algorithm for the first-order Gauss-Markov Rayleigh fading channel models. Simulation and performance analysis are presented in Section 4.5.

In Chapter 5, our multi-user massive MIMO system model, problem formulation, and POMDP formulation for the antenna selection are presented in Sections 5.1, 5.2, and 5.3, respectively. In Section 5.4, we provide the proof of the optimality of the myopic policy for our the antenna selection. In Section 5.5, we propose our myopic policy based antenna selection algorithm for first-order Gauss-Markov channels. In Section 5.6, we propose an offline algorithm for obtaining the optimal threshold value for channel quantization. Simulation and evaluation analysis are illustrated in Section 5.7.

In Chapter 6, we describe the multi-user massive MIMO system model, problem formulation and the POMDP formulation (for JASUS problem) in Sections 6.1, 6.2, and 6.3, respectively. In Section 6.4, we show the optimality of the myopic policy for a positively correlated two-state channel model for our POMDP-based JASUS problem. In Section 6.5, we present the myopic policy JASUS algorithm and implement it on the first-order Gauss-Markov Rayleigh fading channel model. Simulation and performance analysis are presented in Section 6.6.

Finally, in Chapter 7 we first conclude this dissertation and then present several ideas and open problems for future work.

## 1.7 Notations

Upper-case and lower-case bold letters are used to represent matrices and vectors, respectively; calligraphic fonts (e.g., $\mathcal{S}$) signify sets; and Sans-serif fonts identify random vectors (e.g,. $\boldsymbol{\mathsf{s}}$) and random scalars (e.g., $\mathsf{s}$). The transpose and the Frobenius norm of a vector/matrix are shown as $(\cdot)^T$, $\|\cdot\|$ respectively; the $\ell_1$ and $\ell_2$ norms of vector $\mathbf{s}$ are denoted as $\|\mathbf{s}\|_1$ and $\|\mathbf{s}\|_2$ respectively; diag($\mathbf{s}$) stands for a diagonal matrix whose diagonal entries are given by vector $\mathbf{s}$. The notation $E\{\cdot\}$ represents the mathematical expectation; $\mathbf{1}_N$ identifies an $N \times 1$ vector with 1 in all elements. The notation $|\mathcal{S}|$ stands for the cardinality of set $\mathcal{S}$.

# Chapter 2

# Literature Review

In this chapter, we first review traditional optimal antenna selection methods in Section 2.1, for MIMO systems (for selection among small number of antennas). We then briefly review some of the existing sub-optimal antenna selection algorithms that are computationally affordable for massive MIMO systems and are designed under the assumption of the availability of the full CSI. Next, in Section 2.2, we review the studies that devised antenna selection algorithms when partial CSI is available. We then offer an overview on the studies that used POMDPs for designing various resource allocation methods in Section 2.3. Finally, in Section 2.4, we review the existing works that studied the joint antenna and user selection (JAUS) problem for multi-user massive MIMO systems.

## 2.1 Traditional Antenna Selection Methods

Antenna selection has been extensively studied in the literature for MIMO systems with a small number of antennas [7, 13, 25]. The optimal antenna selection algorithms for MIMO systems involve an exhaustive search over all possible selections of antennas and finding the antenna subset which maximizes the signal-to-noise-ratio (SNR) or the capacity [8, 26, 27]. Note that exhaustive search methods are not practical for massive MIMO systems due to the high computational complexity of the existing search algorithms. Hence, alternative low-complexity antenna selection algorithms have been sought for massive MIMO systems [14–18]. In [14], the authors developed an iterative antenna selection algorithm which relying on ranking antennas based on their channel gains. In [15], the author proposed to solve massive antenna selection problem using a convex relaxation technique. Considering both single-cell and multi-cell massive MIMO systems, the authors in [16] introduced the so-called trace-based algorithm to reduce the antenna selection problem complexity. In [17], a low-complexity two-step antenna selection algorithm is proposed for a massive MIMO system. The purposed sub-optimal algorithm of [17] consists of a coarse selection of a subset of antennas based on the channel gains followed by a refined selection of antennas from this subset based on the CSI. In [18], the authors used a Monte Carlo tree search algorithm to design a low-complexity suboptimal antenna selection algorithm for massive MIMO systems. In the proposed method, the antenna selection problem is formulated as a decision making problem, where linear regression is used to update the probability of selecting the correct subset of antennas by utilizing the defined features of CSI.

A major issue with the algorithms proposed in [7, 8, 13–18] is that these algorithms rely on the full channel state information (CSI) assumption, meaning that the BS has the knowledge of all antenna channel coefficients. This assumption is not practical for massive MIMO systems. This is due to the fact that this assumption requires a dedicated RF chain per antenna element, which is not possible when the number of RF chains is limited. One may suggest to tackle this issue by switching available RF chains among all available antennas for training purposes and channel estimation. Note however that adding switches increases the hardware complexity and at the same time switching RF chains exacerbates the problem of outdated CSI. In [28] and [19], the authors propose an efficient switching algorithm for transceivers equipped with a massive number of antennas. However, the proposed algorithms still suffer from the fact that switching RF chains exacerbates the problem of outdated CSI. In [19], the authors use the channel capacity as the optimality criterion and design a switching network where each RF chain can be connected to a predefined subset of antennas to reduce the complexity of switching. Although such an algorithm can reduce the required switching RF chains for full CSI acquisition procedures at each symbol time, it provides a suboptimal antenna selection algorithm due to the limited connectivity and access to the individual antennas.

Given the above discussion, it is required to design a low-complexity antenna selection policy which only relies on partial CSI to avoid the need for using switching RF chains for full CSI acquisition.

## 2.2 Antenna Selection Relying on Partial CSI

As we explained in the previous section, obtaining full CSI is not practical for massive MIMO systems with a limited RF chains at the BS. More specifically, to obtain full CSI, switching RF chains is required which in turns adding switches complicate the circuit, and also results in outdate CSI . To tackle this issue, the authors in [29], proposed a Thompson sampling technique that only relies on partial CSI to solve the antenna selection problem for massive MIMO system. However, the authors show that this technique can achieve high data rates for static scenarios (with zero-velocity users), but in the dynamic scenarios, this technique performs only slightly better than the random selection scheme. In [30], considering time-varying channels such that the dynamic of channels evolve according to a positively correlated Gilbert-Elliot model, and only partial CSI is available, the authors formulated the antenna selection problem as a partially observable Markov decision process (POMDP) framework. In [30], the authors defined the problem of selecting only an antenna among available antennas at a receiver as POMDP and show the optimality of myopic policy algorithm with the goal of minimizing the packet error rate (PER). In [31], we utilize the POMDP frame work to design an antenna selection algorithm for massive MISO system when only partial CSI is availbel and channel evolves according to the same positively correlated two-state Markov channel model. In our proposed algorithm, the optimal policy can be obtained for any arbitrary number of antennas and RF chains.

Since POMDP is a powerful framework to design a decision making policy under uncertainly, researchers applied this framework for different applications in communication systems. In the following section, we review some studies that focus on solving the resource allocation problems by using a POMDP framework.

## 2.3 POMDP-based Resource Allocation Methods

By modeling the dynamics of fading channels as either continuous Gauss-Markov models [32, 33], or finite-state Markov chains [34, 35], some existing studies formulate their corresponding control decision policy as an MDP [36] or as a POMDP [30, 37–41], when the feedback to the transmitter provides full CSI or partial CSI. The authors of [36] provide the design of the optimal opportunistic feedback decision policy for transmit beamforming in the frequency division duplexing (FDD) mode and validate that the underlying feedback control problem for transmit beamforming with throughput maximization over Gauss-Markov channels is an MDP problem. In [30, 37–41], the authors formulate the problem of selecting a subset of available channels or antennas using a POMDP framework, when the state evolves as a Markov process and when limited feedback on CSI is available.
For cognitive radio systems, the authors of [37] used the POMDP framework to address the problem of dynamic spectrum sensing and spectrum allocation to a secondary user. Assuming that each of the available channels follows the same two-state Markov process, the authors devise POMDP-based spectrum sensing and spectrum access policies for the user to decide which channel to sense and which channel to access. Using a simplified unit reward for a successful access, the authors establish the optimality of the myopic sensing policy for the case of *two* available channels,

which follow *positively correlated*[1] two-state Gilbert-Elliot channel model. Under this very same model, the studies in [38] and [39] present the proof of optimality of the myopic policy for sensing and selecting one out of an *arbitrary number* of available channels. In [38], a unit reward is assumed upon successful access and in [39], the reward is the number of bits delivered to a secondary user over the selected channel.

Using a POMDP framework for antenna selection has been considered in [30,40]. In [30], focusing on a single-user data downlink transmission, the authors formulate the antenna selection problem at a multi-antenna receiver with a single RF chain as a POMDP problem with the goal of minimizing the packet error rate (PER). Considering perfect CSI for the selected antenna and assuming positively correlated Gilbert-Elliot channel model, the authors consider the simplified unit reward, when the packet is correctly received, otherwise the reward is zero. The authors show that under such a reward function model, the myopic policy is the optimal solution for the considered antenna selection problem. The authors of [40] prove the optimality of the myopic policy for an extended case where a user is allowed to access a subset of available i.i.d. two-state channels. In [40], one unit of reward is collected when each selected channel is indeed in good state. Under a slightly different structure for the reward function, the optimality of the myopic policy may no longer hold [42]. Thus, the structure of the expected long-term reward function plays an important role in the optimality of the myopic policy. The authors of [41] derive sufficient conditions for the expected long-term reward function that guarantee the optimality of the myopic policy for the general POMDP framework. These conditions state that if the expected long-term reward function is a *regular function,* and for positively correlated two-state channel models, the myopic policy is optimal.

Note that, aside from the difference in the applications, our study in this disser-

[1] A two-state Gilbert-Elliot channel model is called positively correlated if the probability of channel changing from bad state to good state is less than that of staying in the good state.

taion (see [30]) differs from [30, 37–39] is that these studies, study the problem of selecting one spectral channel for the positively correlated two-state channel, unlike the previous works in [30,37–39], that only provided the optimality of myopic policy for selecting one channel/antenna out of an arbitrary number of channels/antennas, we show that in our antenna selection problem, for any different numbers of available antennas and RF chains, the defined expected long-term reward satisfies the conditions of the optimality of myopic policy. Although [40] has studied the optimality of myopic policy of selecting a subset of available channels, the defined reward is a simple collecting one unit of reward for selecting good state channels, while our defined collected reward is the actual data rate.

In the following section, we review the studies that consider a more complicated scenario, such that the number of available users that demand data at the same time, is larger than the number of active antennas (number of available RF chains). In this case, to increase the sum-rate, both antenna and user scheduling techniques are required to be applied.

## 2.4 Joint Antenna and User Scheduling Methods

In multi-user MIMO systems, when a large number of users is available to receive data, user scheduling is required to provide and maintain a high quality of service [43,44]. In user scheduling technique, available users can be grouped in finite number of clusters to be served at different frequencies [43–45] or different time slots [46–48]. Note that the mentioned user scheduling algorithms in [43–48] dealt with only user scheduling problem under the assumption that all available antennas at BS participates in data transmission. However, as we explained before, due to cost and computational complexity, antenna selection technique is required in massive MIMO systems. In this dissertation, we aim to design a policy to solve the JASUS problem in multi-user massive MIMO systems. Due to the high computational complexity of solving the JASUS problem, there are only limited number of studies [20–24] that proposed only joint antenna and user selection (JAUS) algorithms. In the following sequel, we explain the critical issues of the proposed JAUS algorithms in [20–24], and explain why it can not be applied to our defined JASUS problem. In [20], the authors proposed to first select the semi-orthogonal users to receive data, and then use an iterative algorithm that starts with all available antennas and ends with the best subset of antennas with the same size of the number of RF chains by deactivating an antenna at each iteration. Note that the main goal of the JAUS problem in [20], is to maximize the sum-rate. In [21], the authors proposed another iterative search algorithm to solve JAUS problem with the goal of maximizing the sum-rate per unit energy consumption. According to the purposed algorithm in [21], at the first, users with high channel gain are selected, and then an iterative search algorithm are designed to deactivate an antenna at each iteration until the number of selected antennas is same as the number of RF chains. Note that, the computational complexity of the proposed algorithms in [20, 21], limits their application to small number of users and RF chains. To address this issue, considering a single-cell scenarios, the authors in [22], proposed low-complexity JAUS algorithms. However, the proposed algorithms achieve low data rate. In [23], the authors extend the system model to a multi-cell scenario and utilized an Adaptive Markov Chain Monte Carlo method to devise a low-complexity algorithm for JAUS problem with the main goal of sum-rate maximization. However, the authors in [24], pointed a few critical issues in [20–23], which are the assumption of static channel model and ignoring fair user scheduling in their proposed JAUS algorithms. To resolve these issues, with considering time-varying fading channels, the authors in [24], devised a suboptimal greedy JAUS algorithm with the goal of maximizing the sum-rate, while it guarantees a minimum average data rate for all available users in a cell. However, in the proposed algorithm in [24], same as other mentioned JAUS algorithm in [20–23], the authors assume that the full CSI is available. Such an assumption is not practical due to the limited number of available RF chains at the BS.

Motivated by the above explanations, in Chapter 6, considering a time division duplexing (TDD) mode, we study JASUS problem for a multi-user massive MIMO system, in which a BS is equipped with massive number of antennas and limited number of RF chains to serve large number of single-antenna users in a cell. We assume that users receive data in a time frame, where each frame contains finite number of time slots. Furthermore, we assume that the channels evolve according to a Markov chain at the beginning of each frame and remains unchanged during the entire frame. Since at each time slot, only partial CSI is available (due to the limited number of RF chains), we formulate the joint antenna and user scheduling problem as a POMDP framework with the main goal of maximizing the expected long-term sum-rate.

# Chapter 3

# Partially Observable Markov Decision Process

Partially observable Markov decision processes (POMDPs) is a generalization of a Markov decision processes (MDPs) when only partial information about the current state is available. One of the common applications of POMDPs were in the control theory [30, 37–41] with the main purpose of modeling the stochastic dynamic of a system as a POMDP to design an optimal decision making policy. Later on, POMDP became a powerful framework as a learning tool under the uncertainty, in the artificial intelligence area [49–52]. POMDP is known as a powerful tool in decision theory because of its well-defined problem formulation. Note that the computational complexity of finding an optimal policy for a POMDP-based problem exponentially increases with the number of state and action space [53, 54].

In this chapter, we first describe the POMDP model, and then we illustrate the value iteration algorithm which obtain the optimal solution of the POMDP problems.

## 3.1 POMDP Model

We represent a POMDP framework by the following tuple

$$(\mathcal{S}, \mathcal{A}, \mathbf{T}, R(\mathbf{s}, \mathbf{a}), \mathcal{O}, \mathbf{O}(\mathbf{o}, \mathbf{a}), \mathbf{b}) \tag{3.1}$$

where $\mathcal{S}$ is the state space, $\mathcal{A}$ is the action space; $\mathbf{T}$ is the state transition probability matrix, $R(\mathbf{s}, \mathbf{a})$ is the reward at state $\mathbf{s}$ when action $\mathbf{a}$ is taken, $\mathcal{O}$ is the observation space, $\mathbf{O}(\mathbf{o}, \mathbf{a})$ is the matrix of conditional probability of observing $\mathbf{o}$ at different states, given action $\mathbf{a}$ is taken, and $\mathbf{b}$ is the belief vector. In the sequel, we elaborate more on these components.

### State space

An environment can be modeled by state space $\mathcal{S}$, which contains the possible states. Although, the number of states can be infinite (states can be continuous), here we focus on finite state space for the sake of simplicity. We can write the state space as

$$\mathcal{S} \triangleq \{\mathbf{s}_1, \mathbf{s}_2, \ldots, \mathbf{s}_{|\mathcal{S}|}\}, \tag{3.2}$$

where $\mathbf{s}_i$, is the $i$-th state in the state space for $i = 1, 2, \ldots, |\mathcal{S}|$. Here,we use $\mathbf{s}_t$ to denote the random state at time $t$, where $\mathbf{s}_t$ can take any state in the state space.

### Action space

Possible actions that an agent can take in an environment are stored in the actio set denoted as $\mathcal{A}$. Here, our action space $\mathcal{A}$ contains finite number of actions that the agent can make based on received partial information about the current state. Roughly speaking, the main goal here is to define a policy that can select the best action in set $\mathcal{A}$, according to the partial observation of the current state to achieve the desired results. We can write the action space as

$$\mathcal{A} = \{\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_{|\mathcal{A}|}\}, \tag{3.3}$$

where, $\mathbf{a}_i$ is the $i$-th action in the action space for $i = 1, 2, \ldots, |\mathcal{A}|$. We use $\mathbf{a}_t$ to denote the random action at time $t$, where it can take any action in the action space.

### Transition matrix

Since POMDP model an environment, with finite number of states, the transition matrix denoted as $\mathbf{T}$, represents the state evolution. Meaning that the state transition can be mathematically described by the transition matrix. Note that, making an action can affect the state transition, and thus action effects should be captured in the transition matrix. However, in this dissertation, the state evolution is independent of the action (*i.e.*, channel variation is independent of the selected antennas). Thus, we can write the transition probability matrix as an $|\mathcal{S}| \times |\mathcal{S}|$ matrix whose $(i, j)$ element, denoted as $T_{ij}$ , is the probability of the state at time $t$ being $\mathbf{s}_j$, given that the state at time $t-1$ is $\mathbf{s}_i$ and action $\mathbf{a}$ is taken.

### Reward function

The reward function denoted as $R(\mathbf{s}, \mathbf{a})$ indicates the earned reward utility when action $\mathbf{a}_t = \mathbf{a}$ is performed in the state $\mathbf{s}_t = \mathbf{s}$. Defining a proper reward function can results in accurate environment modeling, and thus devising an efficient decision making policy.

### Observation space

The partial observation (or a noisy observation) can be obtained from current state after executing an action in set $\mathcal{A}$. The observation space denoted as $\mathcal{O}$ contains all possible observations. We can write the observation space as

$$\mathcal{O} = \{\mathbf{o}_1, \mathbf{o}_2, \ldots, \mathbf{o}_{|\mathcal{O}|}\}, \tag{3.4}$$

where $\mathbf{o}_i$ is the $i$-th possible observation in the observation space for $i = 1, 2, \ldots, |\mathcal{O}|$. We use $\mathbf{o}_t$ to denote the random observation at time $t$. Note that, in this dissertation, considering a finite state space, the observation space is finite as well.

### Observation probability

The conditional observation probability matrix $\mathbf{O}(\mathbf{o},\mathbf{a})$, is an $|\mathcal{S}| \times |\mathcal{S}|$ diagonal matrix and is defined as $\mathbf{O}(\mathbf{o},\mathbf{a}) = \operatorname{diag}\Big(\Pr\Big\{\boldsymbol{\mathsf{o}}_t = \mathbf{o} | \boldsymbol{\mathsf{s}}_t = \mathbf{s}_i, \boldsymbol{\mathsf{a}}_t = \mathbf{a}\Big\}_{i=1}^{|\mathcal{S}|}\Big)$, whose $i$-th diagonal element is the probability of observing $\mathbf{o} \in \mathcal{O}$ at time $t$, given state $\mathbf{s}_i$ and action $\mathbf{a}$ at time $t$.

### Belief vector

The belief vector at time $t$ is defined as $\mathbf{b}_t \triangleq [b_{1,t} \;\; b_{2,t} \;\; \cdots \;\; b_{|\mathcal{S}|,t}]^T$, where $b_{j,t}$ is the probability of the state $\boldsymbol{\mathsf{s}}_t$ at time $t$ being $\mathbf{s}_j \in \mathcal{S}$, given all the action and observation history until time $t$. If we use $\mathcal{H}_{t-1}$ to represent the action and observation history until time $t-1$, we can write

$$b_{j,t} = \Pr\{\boldsymbol{\mathsf{s}}_t = \mathbf{s}_j | \mathcal{H}_{t-1}\}. \tag{3.5}$$

Here, $\mathcal{H}_{t-1}$ represents the action and observation history until time $t-1$, where

$$\mathcal{H}_{t-1} \triangleq \{\mathbf{o}_{t-1}, \mathbf{a}_{t-1}, \mathcal{H}_{t-2}\}. \tag{3.6}$$

We also define the belief space as $\mathcal{B} \triangleq \Big\{\mathbf{b} \in \mathbb{R}^{|\mathcal{S}|} : \mathbf{1}^T\mathbf{b} = 1, \mathbf{b} \succcurlyeq \mathbf{0}\Big\}$. As shown in Appendix A (also, in [31]), using Bayes' rule, we can obtain $\mathbf{b}_t$ from $\mathbf{b}_{t-1}$ as

$$\mathbf{b}_t = \mathbf{g}(\mathbf{o}_{t-1}, \mathbf{a}_{t-1}, \mathbf{b}_{t-1}), \tag{3.7}$$

where we define $\mathbf{g}(\mathbf{o},\mathbf{a},\mathbf{b}) \triangleq \frac{\mathbf{O}(\mathbf{o},\mathbf{a})\mathbf{T}\mathbf{b}}{g(\mathbf{o},\mathbf{a},\mathbf{b})}$, and $g(\mathbf{o},\mathbf{a},\mathbf{b}) \triangleq \mathbf{1}^T\mathbf{O}(\mathbf{o},\mathbf{a})\mathbf{T}\mathbf{b}$, for $\mathbf{o} \in \mathcal{O}$ and $\mathbf{a} \in \mathcal{A}$. It is well-known that $\mathbf{b}_t$ is a sufficient statistic to make decision at time $t$ [55]. Note that instead of given realization observation $\mathbf{o}_{t-1}$, if we consider the observation vector $\boldsymbol{\mathsf{o}}_{t-1}$ which is a random vector, the believe vector in (3.7) also becomes a random vector, denoted by $\boldsymbol{\mathsf{b}}_t$, which is given by

$$\boldsymbol{\mathsf{b}}_t \triangleq \mathbf{g}(\boldsymbol{\mathsf{o}}_{t-1}\ , \mathbf{a}_{t-1}, \mathbf{b}_{t-1}). \tag{3.8}$$

Correspondingly, the $j$-th entry of $\boldsymbol{\mathsf{b}}_t$ is defined as $\mathsf{b}_{j,t} \triangleq \Pr(\boldsymbol{\mathsf{s}}_t = \mathbf{s}_j | \boldsymbol{\mathcal{H}}_{t-1})$, where $\boldsymbol{\mathcal{H}}_{t-1}$ is the collection of all observations and actions as random vectors until time

$t-1$ and is defined as $\boldsymbol{\mathcal{H}}_{t-1} \triangleq \{\mathbf{o}_{t-1}, \mathbf{a}_{t-1}, \boldsymbol{\mathcal{H}}_{t-2}\}$. Note that $\mathcal{H}_{t-1}$ in (3.6) is a realization of $\boldsymbol{\mathcal{H}}_{t-1}$ after observing $\mathbf{o}_{t-1}$.

## 3.2 Policy and Objective Function

Policy $\Pi = \{\pi_0(\cdot), \pi_1(\cdot), \cdots\}$ is a sequence of decision rules $\pi_t(\cdot)$, which at time $t$, maps the belief vector $\mathbf{b}_t$ to the action $\mathbf{a}_t$, that is $\mathbf{a}_t = \pi_t(\mathbf{b}_t)$. The policy is stationary if it consists of a single decision rule used for all time slots. In an infinite horizon POMDP problem, the optimal policy is stationary [55], i.e., $\Pi = \{\pi(\cdot), \pi(\cdot), \cdots\}$.

With the initial belief vector $\mathbf{b}_0$, the objective function denoted as $J_\pi(\mathbf{b}_0)$ for an infinite horizon POMDP frame work is defined as

$$\begin{aligned} J_\pi(\mathbf{b}_0) &= E_{\{\mathbf{s}_t\}}\Big\{\sum_{t=0}^{\infty} R(\mathbf{s}_t, \mathbf{a}_t)\Big|\mathbf{b}_0\Big\} \\ &= E_{\{\mathbf{s}_t\}}\Big\{\sum_{t=0}^{\infty} R(\mathbf{s}_t, \pi(\mathbf{b}_t))\Big|\mathbf{b}_0\Big\}, \end{aligned} \tag{3.9}$$

where $E_{\{\mathbf{s}_t\}}\{\cdot\}$ is the expectation taken with respect to the joint probability distribution of $\{\mathbf{s}_t\}_{t=0}^{+\infty}$, given the initial distribution $\mathbf{b}_0$. Note that as $\mathbf{a}_t = \pi(\mathbf{b}_t)$, the objective function $J_\pi(\mathbf{b}_0)$ is parameterized by the stationary policy $\pi(\cdot)$, and hence, we use subscript $\pi$ to signify $J_\pi(\mathbf{b}_0)$. It is worth mentioning that the random vectors $\mathbf{b}_t$, and $\mathbf{a}_t = \pi(\mathbf{b}_t)$ are functions of random observation vectors $\{\mathbf{o}_{t'}\}_{t'=0}^{t-1}$ and the initial belief vector $\mathbf{b}_0$. Given the POMDP model and the dynamic of the POMDP problem, the main goal is to find the optimal policy as

$$\pi^* = \arg\max_{\pi} J_\pi(\mathbf{b}_0), \text{ for any } \mathbf{b}_0. \tag{3.10}$$

Since $\mathbf{a}_t$ is a function of $\mathbf{b}_t$, which is in turn a function of $\mathbf{o}_{t-1}$, there is a one-to-one correspondence between $\boldsymbol{\mathcal{H}}_t$ and $\{\mathbf{o}_{t'}\}_{t'=0}^{t}$. Hence, we can write

$$\begin{aligned} J_\pi(\mathbf{b}_0) &= E_{\{\mathbf{s}_t\}}\Big\{\sum_{t=0}^{+\infty} R(\mathbf{s}_t, \mathbf{a}_t)\Big|\mathbf{b}_0\Big\} \\ &= \sum_{t=0}^{+\infty} E_{\mathbf{s}_t}\Big\{R(\mathbf{s}_t, \mathbf{a}_t)\Big|\mathbf{b}_0\Big\} \end{aligned}$$

$$
\begin{aligned}
&= \sum_{t=0}^{+\infty} E_{\boldsymbol{\mathcal{H}}_{t-1}} \Big\{ E_{\mathbf{s}_t|\boldsymbol{\mathcal{H}}_{t-1}} \{ R(\mathbf{s}_t, \mathbf{a}_t) | \boldsymbol{\mathcal{H}}_{t-1} \} \Big| \mathbf{b}_0 \Big\} \\
&= \sum_{t=0}^{+\infty} E_{\boldsymbol{\mathcal{H}}_{t-1}} \Big\{ \sum_{j=1}^{|\mathcal{S}|} R(\mathbf{s}_j, \mathbf{a}_t) \Pr(\mathbf{s}_t = \mathbf{s}_j | \boldsymbol{\mathcal{H}}_{t-1}) \Big| \mathbf{b}_0 \Big\} \\
&= \sum_{t=0}^{+\infty} E_{\boldsymbol{\mathcal{H}}_{t-1}} \Big\{ \sum_{j=1}^{|\mathcal{S}|} R(\mathbf{s}_j, \mathbf{a}_t) \mathsf{b}_{j,t} \Big| \mathbf{b}_0 \Big\} \\
&= E_{\{\boldsymbol{\mathcal{H}}_t\}} \Big\{ \sum_{t=0}^{+\infty} \mathbf{r}^T(\mathbf{a}_t) \mathbf{b}_t \Big| \mathbf{b}_0 \Big\}, \qquad (3.11)
\end{aligned}
$$

where $\{\boldsymbol{\mathcal{H}}_t\}$ is the entire history and $\mathbf{r}(\mathbf{a}) = [R(\mathbf{s}_1, \mathbf{a}) \;\; R(\mathbf{s}_2, \mathbf{a}) \;\; \cdots \;\; R(\mathbf{s}_{Q^M}, \mathbf{a})]^T$ is the reward vector of all channel states under action $\mathbf{a}$.

## 3.3 Optimal Policy via Dynamic Programming

Since a POMDP is a continuous belief state MDP, we can straightforwardly write the dynamic programming equation [55] for the infinite horizon continuous-state MDP with dynamics in (3.7) and the objective function in (3.11). The optimal policy $\pi^*(\cdot)$ for the infinite horizon MDP can be obtained by the $K$-horizon dynamic programming recursion when $K \to \infty$. To present this algorithm, we define the value function $V(\mathbf{b})$ as $V(\mathbf{b}) \triangleq \max_{\mathbf{a}\in\mathcal{A}} \mathbf{r}^T(\mathbf{a})\mathbf{b} + \sum_{\mathbf{o}\in\mathcal{O}} V(\mathbf{g}(\mathbf{o}, \mathbf{a}, \mathbf{b}))g(\mathbf{o}, \mathbf{a}, \mathbf{b})$. Then, the following theorem presents Bellman's equations that must be satisfied by the optimal policy.

**Theorem 1.** *(Bellman's equation for an infinite horizon POMDP [55]): Consider an infinite horizon POMDP with the belief state* $\mathbf{b} \in \mathcal{B}$*. The optimal policy* $\pi^*(\cdot)$ *satisfies Bellman's dynamic programming equation as it follows:*

$$
\begin{aligned}
Q(\mathbf{b}, \mathbf{a}) &\triangleq \mathbf{r}^T(\mathbf{a})\mathbf{b} + \sum_{\mathbf{o}\in\mathcal{O}} V(\mathbf{g}(\mathbf{o}, \mathbf{a}, \mathbf{b}))g(\mathbf{o}, \mathbf{a}, \mathbf{b}) \\
\pi^*(\mathbf{b}) &= \arg\max_{\mathbf{a}\in\mathcal{A}} Q(\mathbf{b}, \mathbf{a}), \qquad (3.12) \\
V(\mathbf{b}) &= \max_{\mathbf{a}\in\mathcal{A}} Q(\mathbf{b}, \mathbf{a}), J_{\pi^*}(\mathbf{b}_0) = V(\mathbf{b}_0).
\end{aligned}
$$

The proof of Theorem 1 is provided in [55]. The value iteration algorithm explained in the next subsection provides the solution to Bellman's equation (3.12) by generating a sequence of functions that converges over $\mathcal{B}$ to a unique solution regardless of the initial belief.

### 3.3.1 Value Iteration Algorithm

The value iteration algorithm yields the solution to the Bellman equation regardless of initialization. Let $n$ denote the iteration number and $n = 1, 2, \ldots, K$. The value iteration is a successive approximation algorithm to compute value function $V(\mathbf{b})$ of the Bellman's equation [55]. Presented in Algorithm 1, the value iteration yields the optimal policy $\pi^*(\mathbf{b})$ and optimal expected reward $V(\mathbf{b})$ of the POMDP problem by performing an exhaustive search over all possible actions.

**Algorithm 1** The forward value iteration

```
1: Set $n = 0$ and initialize $V_0(\mathbf{b}) = 0$.
2: Set $n = n + 1$ and compute $V_n(\mathbf{b})$ and $\pi_n^*(\mathbf{b})$ as
       $Q_n(\mathbf{b}, \mathbf{a}) = \mathbf{r}(\mathbf{a})^T\mathbf{b} + \sum_{\mathbf{o}\in\mathcal{O}} V_{n-1}(\mathbf{g}(\mathbf{o}, \mathbf{a}, \mathbf{b}))g(\mathbf{o}, \mathbf{a}, \mathbf{b})$    (3.13)
       $V_n(\mathbf{b}) = \max_{\mathbf{a}\in\mathcal{A}} Q_n(\mathbf{b}, \mathbf{a})$
       $\pi_n^*(\mathbf{b}) = \arg\max_{\mathbf{a}\in\mathcal{A}} Q_n(\mathbf{b}, \mathbf{a})$
3: Stop if $n = K$ , otherwise go to Step 2.
```

Finally, the policy $\pi_K^*(\cdot)$ is used at each time slot $t$ for antenna selection decision in the a real time controller. Since the policy is stationary, only the policy $\pi_K^*(\cdot)$ for very large $K$ needs to be stored for real-time implementations. Several tools exist to solve POMDPs [56, 57]; however, high complexity (SPACE hard) of the optimal solution algorithm restricts its use only to problems with a small number of states.

# Chapter 4

# POMDP-based Antenna Selection Algorithm in Point-to-Point System

## 4.1 System Model

We consider a point-to-point downlink transmission link in a massive MIMO system where a BS, equipped with $M$ antennas and $N$ transmit RF chains, transmits data to a single-antenna user. It is assumed that $M \gg 1$ and $N < M$. The system is slotted and each time slot indexed by $t$, for $t = 0, 1, \ldots$. We assume that the channel between the BS and the user is time-varying and changes over time slots. Since the number of the transmit RF chains is limited, at each time slot, the BS needs to select $N$ out of $M$ antennas for transmission. We assume that the massive MIMO system operates in the TDD mode, and therefore, the CSI acquisition for downlink transmission is performed using the uplink channel measurements. With the selected $N$ RF chains, only the CSI of the corresponding $N$ selected antennas at the current time slot can be measured. We aim to maximize the expected long-term transmission rate by selecting $N$ antennas in each time slot. We assume the channel vector evolves over time as a finite-state Markov process. Since we can only observe $N$ out of $M$ channel coefficients at each time slot, we use a POMDP framework to design our antenna selection policy.

## 4.2 POMDP Formulation

In this section, we formulate our antenna selection problem using the POMDP framework. To do so, we first define the POMDP components of our dynamic antenna selection problem.

**POMDP Components:** As we represent a POMDP framework by the tuple $(\mathcal{S}, \mathcal{A}, \mathbf{T}, R(\mathbf{s}, \mathbf{a}), \mathcal{O}, \mathbf{O}(\mathbf{o}, \mathbf{a}), \mathbf{b})$, in Chapter. 3, we elaborate on these components in our antenna selection problem.

### State space

The state space, denoted by $\mathcal{S}$, is the set of a finite number of states labeled as $\mathbf{s}_j$, each of which takes one of the possible channel vectors denoted by $\tilde{\mathbf{h}}_j$, where[1] $\mathbf{s}_j = \tilde{\mathbf{h}}_j \triangleq [\tilde{h}_{1j} \quad \tilde{h}_{2j} \quad \cdots \quad \tilde{h}_{Mj}]^T$, is one of the possible $M \times 1$ complex vectors of the channel coefficients between the $M$ available antennas at the BS and the user's antenna. We assume that each channel coefficient takes one of the $Q$ possible values, i.e., $\tilde{h}_{ij} \in \{\alpha_1, \alpha_2, \cdots, \alpha_Q\}$, where $\alpha_i \in \mathbb{C}$, for $i = 1, 2, \ldots Q$. The state space has $Q^M$ states and is given by $\mathcal{S} \triangleq \{\tilde{\mathbf{h}}_1, \tilde{\mathbf{h}}_2, \ldots, \tilde{\mathbf{h}}_{Q^M}\}$. The channel state $\mathsf{h}_t$ at time $t$ takes one of the $Q^M$ elements in $\mathcal{S}$. To ease the notation, we use $\mathsf{h}_t$ and $\mathsf{s}_t$ interchangeably to indicate the channel state. As will be explained later and shown in Fig. 4.1, the state is assumed to evolve at the beginning of each time slot.

### Action space

The action in our system model is the decision of selecting $N$ out of $M$ antennas. Hence, there are $L = \binom{M}{N}$ possible actions and the action space is given by $\mathcal{A} \triangleq \{\tilde{\mathbf{a}}_1, \tilde{\mathbf{a}}_2, \cdots, \tilde{\mathbf{a}}_L\}$, where $\tilde{\mathbf{a}}_l \triangleq [a_{l1} \quad a_{l2} \quad \cdots \quad a_{lM}]^T, \quad a_{lj} \in \{0, 1\}$, with $\sum_{j=1}^{M} a_{lj} = N$. The antenna selection decision, denoted by $\mathsf{a}_t$, is made at the beginning of time slot $t$.

[1] Throughout this chapter, to ease the notation, we use $\mathbf{s}_j$ and $\tilde{\mathbf{h}}_j$ interchangeably.

### Transition probability

As we described in section. 3.1, the transition probability matrix here is a $Q^M \times Q^M$ matrix, and we can write $T_{ij} = \Pr(\mathbf{s}_t = \mathbf{s}_j | \mathbf{s}_{t-1} = \mathbf{s}_i)$, for $i = 1, \ldots, Q^M, j = 1, \ldots, Q^M$. It is herein assumed that the transition probability matrix $\mathbf{T}$ is known[2].

### Observation space

In our system model, the observation vector at time $t$, denoted as $\mathbf{o}_t$ is the $M \times 1$ vector, where $M - N$ elements of it are equal to zero, and the other $N$ elements of it are the channel coefficients of the selected antennas at the BS and the user's antenna that is measured via uplink training. The observation space is given as

$$\mathcal{O} \triangleq \{\mathbf{o}_1, \mathbf{o}_2, \ldots, \mathbf{o}_{L'}\} \tag{4.1}$$

where $L' = Q^N \times \binom{M}{M-N}$, and $\mathbf{o}_j$ is one of the $L'$ possible values of the $M \times 1$ channel observations. At time $t$, given the antenna selection vector $\mathbf{a}_t$, we can write

$$\mathbf{o}_t \triangleq \mathrm{diag}(\mathbf{a}_t)\mathbf{s}_t. \tag{4.2}$$

The observation vector $\mathbf{o}_t$ is available at the end of time slot $t$, as shown in Fig. 4.1.

### Observation probability

As the description is provided in section. 3.1, here the conditional observation probability matrix $\mathbf{O}(\mathbf{o}, \mathbf{a})$ is a $Q^M \times Q^M$ diagonal matrix and is defined as $\mathbf{O}(\mathbf{o}, \mathbf{a}) = \mathrm{diag}\Big(\Pr\Big\{\mathbf{o}_t = \mathbf{o} | \mathbf{s}_t = \mathbf{s}_i, \mathbf{a}_t = \mathbf{a}\Big\}_{i=1}^{Q^M}\Big)$.

### Reward

The BS uses transmit beamforming[3] for downlink data transmission. We use the achieved data rate by the antenna selection $\mathbf{a}_t = \mathbf{a}$ at state $\mathbf{s}_t = \mathbf{s}$ as the immediate

[2] We use an offline method to obtain the transition probabilities for slow fading channel models, when the channels are quantized into two states, in the Section on Simulation and Performance Analysis.

[3] Based on our defined MISO system model, where the receiver is a single-antenna user, we are using the maximum ratio transmission (MRT) downlink beamforming. Note that for the single user scenario, MRT is the optimal downlink beamformer [58].

reward $R(\mathbf{s}, \mathbf{a})$ in time slot $t$. Since the system operates in TDD and the channel state is assumed unchanged in a time slot and the uplink and downlink channels are identical. Thus, the immediate reward for this MISO link is given by $R(\mathbf{s}, \mathbf{a}) = \log_2 \left(1 + \frac{P\|\mathrm{diag}(\mathbf{a})\mathbf{s}\|^2}{\sigma^2}\right) = \log_2 \left(1 + \frac{P\|\mathbf{o}\|^2}{\sigma^2}\right)$, where $P$ is the transmit power at the BS power and $\sigma^2$ is the noise power.

### Belief vector

The belief explanation is provided in section. 3.1. The belief vector at time $t$ is defined as $\mathbf{b}_t \triangleq [b_{1,t} \;\; b_{2,t} \;\; \cdots \;\; b_{|\mathcal{S}|,t}]^T$, where $b_{j,t}$ is

$$b_{j,t} = \Pr\{\mathbf{s}_t = \mathbf{s}_j | \mathcal{H}_{t-1}\}, \tag{4.3}$$

where $\mathcal{H}_{t-1}$ represents the action and observation history until time, where

$$\mathcal{H}_{t-1} \triangleq \{\mathbf{o}_{t-1}, \mathbf{a}_{t-1}, \mathcal{H}_{t-2}\}. \tag{4.4}$$

The dynamic of a real-time POMDP controller is explained in the sequel. In this procedure, $\pi_t(\cdot)$ is the decision policy at time $t$ that maps the belief vector $\mathbf{b}_t$ to action $\mathbf{a}_t$, that is $\mathbf{a}_t = \pi_t(\mathbf{b}_t)$. Given channel state $\mathsf{\mathbf{s}}_t$, the BS uses all the information available until time $t-1$ to obtain (update) the belief vector $\mathsf{\mathbf{b}}_t$, and then, makes the antenna selection decision; the resulting transmission accrues as an instantaneous reward $R(\mathsf{\mathbf{s}}_t, \mathsf{\mathbf{a}}_t)$. Fig. 4.1 shows the time-line of the POMDP model in our dynamic antenna selection problem. We assume that the state $\mathsf{\mathbf{s}}_t$ is updated at the beginning of each time slot $t$ (i.e., when the downlink transmission is performed) and remains unchanged during the uplink transmission. The system is initialized with the belief vector $\mathbf{b}_0$. Based on that initial belief, the initial antenna selection $\mathsf{\mathbf{a}}_0$ is made. Downlink transmission is then performed. At the end of the uplink transmission, we receive the observation $\mathsf{\mathbf{o}}_0 = \mathrm{diag}(\mathsf{\mathbf{a}}_0)\mathsf{\mathbf{s}}_0$, which is the $N \times 1$ vector of channel state information corresponding to those selected antennas. The reward is given by $R(\mathsf{\mathbf{s}}_0, \mathsf{\mathbf{a}}_0)$. The belief vector $\mathsf{\mathbf{b}}_1$ is updated as in (3.8). The process described above repeats for the next time slot. Given the observation $\mathsf{\mathbf{o}}_{t-1}$,

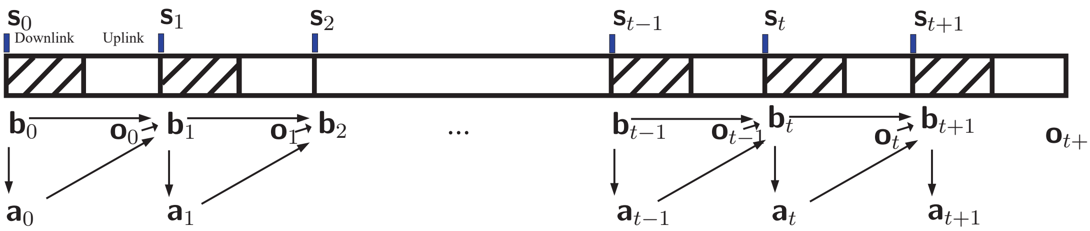


Figure 4.1: An illustration of the antenna selection problem using the POMDP model.

our goal is to design the antenna selection policy such that the expected long-term reward $E\Big\{\sum_{t=0}^{+\infty} R(\mathbf{s}_t, \mathbf{a}_t)\Big\}$, is maximized. Here, the mathematical expectation $E\{\cdot\}$ is taken with respect to random channel states $\{\mathbf{s}_t\}_{t=0}^{+\infty}$. Note that $\mathbf{a}_t$ depends on $\{\mathbf{s}_{t'}\}_{t'=0}^{t-1}$, and hence is random.

### 4.2.1 Policy and Objective Function

According to Section. 3.2, here the policy is stationary such that at time $t$, the policy $\pi$ maps the belief vector $\mathbf{b}_t$ to the action $\mathbf{a}_t$, that is $\mathbf{a}_t = \pi(\mathbf{b}_t)$. With the initial belief vector $\mathbf{b}_0$, the objective function denoted as $J_\pi(\mathbf{b}_0)$ for an infinite horizon POMDP frame work is defined as

$$J_\pi(\mathbf{b}_0) = E_{\{\mathbf{s}_t\}}\Big\{\sum_{t=0}^{\infty} R(\mathbf{s}_t, \mathbf{a}_t)\Big|\mathbf{b}_0\Big\} = E_{\{\mathbf{s}_t\}}\Big\{\sum_{t=0}^{\infty} R(\mathbf{s}_t, \pi(\mathbf{b}_t))\Big|\mathbf{b}_0\Big\}, \tag{4.5}$$

where $E_{\{\mathbf{s}_t\}}\{\cdot\}$ is the expectation taken with respect to the joint probability distribution of $\{\mathbf{s}_t\}_{t=0}^{+\infty}$, given the initial distribution $\mathbf{b}_0$. and the main goal is to find the optimal policy as

$$\pi^* = \arg\max_\pi J_\pi(\mathbf{b}_0), \text{ for any } \mathbf{b}_0. \tag{4.6}$$

According to (3.11), we can write

$$J_\pi(\mathbf{b}_0) = E_{\{\boldsymbol{\mathcal{H}}_t\}}\Big\{\sum_{t=0}^{+\infty} \mathbf{r}^T(\mathbf{a}_t)\mathbf{b}_t\Big|\mathbf{b}_0\Big\}, \tag{4.7}$$

where $\{\boldsymbol{\mathcal{H}}_t\}$ is the entire history and $\mathbf{r}(\mathbf{a}) = [R(\mathbf{s}_1, \mathbf{a}) \;\; R(\mathbf{s}_2, \mathbf{a}) \;\; \cdots \;\; R(\mathbf{s}_{Q^M}, \mathbf{a})]^T$ is the reward vector of all channel states under action $\mathbf{a}$.

### 4.2.2 Optimal Policy via Dynamic Programming

As we explained in Section. 3.3, since a POMDP is a continuous belief state MDP, we can straightforwardly write the dynamic programming equation [55] for the infinite horizon continuous-state MDP . Thus, we can use Bellman's equation to obtain the optimal policy $\pi^*(\cdot)$. To do so, based on our system model, $Q_n(\mathbf{b}, \mathbf{a})$ in (3.13) can be rewritten as

$$Q_n(\mathbf{b}, \mathbf{a}) = \sum_{j=1}^{Q^M} \log_2(1 + \frac{P\|\mathrm{diag}(\mathbf{a})\tilde{\mathbf{h}}_j\|^2}{\sigma^2})b_j + \sum_{\mathbf{o}\in\mathcal{O}} V_{n-1}(\mathbf{g}(\mathbf{o}, \mathbf{a}, \mathbf{b}))g(\mathbf{o}, \mathbf{A}, \mathbf{b}). \quad (4.8)$$

Given (4.8), we can run the value iteration algorithm presented in Algorithm. 1.

So far, we formulated the antenna selection problem as a POMDP problem which can be solved via dynamic programming. Several tools exist for solving POMDPs [56, 57]; however, high complexity (SPACE hard) of the optimal solution algorithm restricts its use only to problems with a small number of states. Since the state dimension in our model is $Q^M$, the optimal solution becomes computationally intractable to obtain as the number of antennas $M$ increases. For massive MIMO systems, we seek a low-complexity suboptimal solution for the selection decision. *Myopic policy*, a greedy solution which maximizes the expected immediate reward[4] ), is a suboptimal solution that is often used to tackle POMDP problems. In the next section, *we rigorously prove that under certain conditions, myopic policy provides the optimal solution to our POMDP-based antenna selection problem.*

## 4.3 Two-State Channels: The Optimality of Myopic Policy

In this section, we consider $Q = 2$, i.e., each channel coefficient $\mathsf{h}_{i,t}$ (i.e., the $i$-th element in $\boldsymbol{\mathsf{h}}_t$) has two possible values denoted as $\alpha$ and $\beta$, where $|\alpha| > |\beta|$. That

[4] Note that $\mathbf{r}^T(\boldsymbol{\mathsf{a}}_t)\boldsymbol{\mathsf{b}}_t = \mathbf{r}^T(\mathbf{a}_t)\mathbf{b}_t$ is the expected immediate reward function at time $t$, given $\boldsymbol{\mathcal{H}}_{t-1} = \mathcal{H}_{t-1}$.

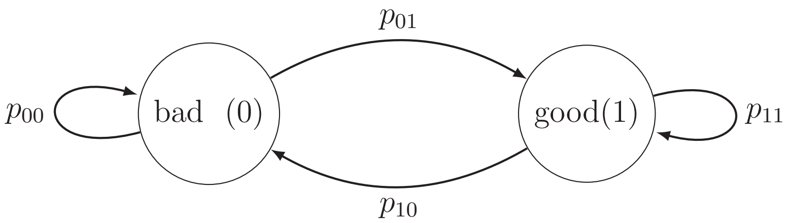


Figure 4.2: A two-state Markov chain model of the channel between each BS antenna and the user device's antenna.

is, we can write[5]

$$\tilde{h}_{ij} \in \{\alpha, \beta\}, \qquad \text{for } j = 1, 2, \dots 2^M \text{ and } i = 1, 2, \dots, M. \tag{4.9}$$

Each channel $\mathsf{h}_{i,t}$ is modeled as a Gilbert-Elliott channel which evolves as a two-state Markov chain with bad (0) and good (1) states over time slots, as shown in Fig. 4.2. The transition probability matrix $\mathbf{P}$ for for each $\mathsf{h}_{i,t}$ is given by

$$\mathbf{P} = \begin{bmatrix} p_{00} & p_{01} \\ p_{10} & p_{11} \end{bmatrix} \tag{4.10}$$

where $p_{ij}$ is the probability of channel changing from state $i$ to state $j$, where $i, j \in \{0, 1\}$. Here, $p_{01}$ is the probability of each channel coefficient changing from the bad channel state to the good channel state and $p_{10}$ is the probability of each channel coefficient changing from good to bad. Also, $p_{00} = 1 - p_{01}$ and $p_{11} = 1 - p_{10}$ is the probability of each channel coefficient, remaining in the bad channel state and good channel state in the next time slot, respectively. We assume that the channel state is positively correlated, i.e, $p_{11} > p_{01}$.

For $j = 1, 2, \dots 2^M$ and $i = 1, 2, \dots, M$, let us define

$$c_{ij} = \begin{cases} 1 & \text{if } \tilde{h}_{ij} = \alpha \\ 0 & \text{if } \tilde{h}_{ij} = \beta \end{cases}. \tag{4.11}$$

Without loss of generality, we redefine state space as $\mathcal{C} = \{\mathbf{c}_j\}_{j=1}^{2^M}$, where $\mathbf{c}_j \triangleq [c_{1j} \;\; c_{2j} \;\; \cdots \;\; c_{Mj}]^T$ is the $j$-th member of $\mathcal{C}$. Consequently, for $\mathsf{s}_t$, we establish an

[5] We will soon see that only the amplitudes of $\alpha$ and $\beta$ are involved in the decision making and the reward and their phases do not have any bearing on the proposed scheme.

equivalent state at time $t$ denoted by $\boldsymbol{\mathsf{c}}_t \triangleq [\mathsf{c}_{1,t} \quad \mathsf{c}_{2,t} \quad \cdots \quad \mathsf{c}_{M,t}]^T \in \mathcal{C}$, where $\mathsf{c}_{i,t}$ is the random variable state of the channel between the the $i$-th BS antenna and the user device antenna. Note that the state $\boldsymbol{\mathsf{s}}_t$ of the channel vector can be uniquely determined from $\boldsymbol{\mathsf{c}}_t$ and vice versa.

The two-state channel model for each $\mathsf{h}_{i,t}$ allows us to simplify the belief formulation, as explained in the sequel. First, we define the conditional probability of the channel between the $i$-th BS antenna and the user at time slot $t$ being in the good state, given the history of all past actions and observations up to time slot $t-1$ as $\omega_{i,t} \triangleq \Pr(\mathsf{c}_{i,t} = 1 | \mathcal{H}_{t-1})$, for $i = 1, 2, \ldots, M$. We also define an equivalent belief vector at time $t$ as $\boldsymbol{\omega}_t \triangleq [\omega_{1,t} \quad \omega_{2,t} \quad \ldots \quad \omega_{M,t}]^T$. Given the antenna selection vector $\mathbf{a}_t$ and the current channel state $\mathbf{c}_t$, the $i$-th entry of the belief vector $\boldsymbol{\omega}_{t+1}$ is updated as

$$\omega_{i,t+1} = \begin{cases} p_{11} & \text{if } a_{i,t} = 1, \;\; c_{i,t} = 1; \\ p_{01} & \text{if } a_{i,t} = 1, \;\; c_{i,t} = 0; \\ \omega_{i,t} p_{11} + (1 - \omega_{i,t}) p_{01} & \text{if } a_{i,t} = 0. \end{cases} , \qquad \text{for } i = 1, \cdots, M. \tag{4.12}$$

We can express the $j$-th entry of the belief vector at time $t$ as

$$b_{j,t} = \Pr(\boldsymbol{\mathsf{s}}_t = \mathbf{s}_j | \mathcal{H}_{t-1}) = \Pr(\boldsymbol{\mathsf{c}}_t = \mathbf{c}_j | \mathcal{H}_{t-1}). \tag{4.13}$$

Since the channels across different antennas are assumed to be statistically independent, we can write

$$\Pr(\boldsymbol{\mathsf{c}}_t = \mathbf{c}_j | \mathcal{H}_{t-1}) \triangleq \prod_{i=1}^{M} \Pr(\mathsf{c}_{i,t} = c_{ij} | \mathcal{H}_{t-1}) = \prod_{i=1}^{M} \hat{f}(\omega_{i,t}, c_{ij}), \tag{4.14}$$

where we define $\hat{f}(\omega, c) = \omega^c (1-\omega)^{1-c}$, and we use the fact that $\Pr(\mathsf{c}_{i,t} = 1 | \mathcal{H}_{t-1}) = \omega_{i,t}$ and $\Pr(\mathsf{c}_{i,t} = 0 | \mathcal{H}_{t-1}) = 1 - \omega_{i,t}$. Based on (4.13) and (4.14), the expected immediate reward function at time $t$, i.e., $\mathbf{r}^T(\mathbf{a}_t)\mathbf{b}_t$, can be written as

$$\bar{R}(\mathbf{a}_t, \boldsymbol{\omega}_t) \triangleq \mathbf{r}^T(\mathbf{a}_t)\mathbf{b}_t = \sum_{j=1}^{|\mathcal{S}|} R(\mathbf{s}_j, \mathbf{a}_t) b_{j,t}$$

$$= \sum_{j=1}^{|\mathcal{C}|} R(\mathbf{s}_j, \mathbf{a}_t)\Pr(\mathbf{c}_t = \mathbf{c}_j|\mathcal{H}_{t-1})$$

$$= \sum_{j=1}^{|\mathcal{C}|} \log_2\left(1 + \frac{P\|\mathbf{o}_{j,t}\|^2}{\sigma^2}\right)\prod_{i=1}^{M}\hat{f}(\omega_{i,t}, c_{ij}), \tag{4.15}$$

where $\mathbf{o}_{j,t}$ is the observation vector at time $t$, if $\mathbf{s}_t = \mathbf{s}_j$ and can be written, using (4.2), as

$$\mathbf{o}_{j,t} = \text{diag}(\mathbf{a}_t)\mathbf{s}_j. \tag{4.16}$$

Here, each entry of $\mathbf{o}_{j,t}$ belongs to the set $\{\alpha, \beta\}$. Note that, if $k$ entries of $\mathbf{o}_{j,t}$ are equal to $\alpha$ and the remaining $N-k$ of non-zero entries of $\mathbf{o}_{j,t}$ are equal to $\beta$, then the corresponding data data rate, denoted as $R_k$, is equal to

$$R_k = \log_2(1 + \frac{P(k|\alpha|^2 + (N-k)|\beta|^2)}{\sigma^2}). \tag{4.17}$$

Given action $\mathbf{a}_t$, the state space $\mathcal{C}$ can be partitioned as

$$\mathcal{C} = \bigcup_{k=0}^{N} \mathcal{C}_k(\mathbf{a}_t), \tag{4.18}$$

where $\mathcal{C}_k(\mathbf{a}_t) = \{\mathbf{c} = [c_1 \quad c_2 \quad \cdots \quad c_M]^T \in \mathcal{C}\big| \ \|\text{diag}(\mathbf{a}_t)\mathbf{c}\|^2 = k\}$. Using (4.18), we can rewrite (4.15) as

$$\begin{aligned}\bar{R}(\mathbf{a}_t, \boldsymbol{\omega}_t) &= \sum_{k=0}^{N}\sum_{\mathbf{c}\in\mathcal{C}_k(\mathbf{a}_t)} R_k \prod_{i=1}^{M}\hat{f}(\omega_{i,t}, c_i) \\ &= \sum_{k=0}^{N} R_k \sum_{\mathbf{c}\in\mathcal{C}_k(\mathbf{a}_t)}\prod_{i=1}^{M}\hat{f}(\omega_{i,t}, c_i) \\ &= \sum_{k=0}^{N} R_k \sum_{\mathbf{c}\in\mathcal{C}_k(\mathbf{a}_t)}\prod_{i\in\mathcal{I}(\mathbf{a}_t)}\hat{f}(\omega_{i,t}, c_i)\prod_{i\in\mathcal{I}(\mathbf{a}_t)^{\perp}}\hat{f}(\omega_{i,t}, c_i),\end{aligned} \tag{4.19}$$

where $\mathcal{I}(\mathbf{a}_t)$ is the index set of the selected antennas while $\mathcal{I}^{\perp}(\mathbf{a}_t)$ is the complement set of $\mathcal{I}(\mathbf{a}_t)$ and contains the indices of the remaining unselected antennas. Note that, $|\mathcal{I}(\mathbf{a}_t)| = N$ and $|\mathcal{I}^{\perp}(\mathbf{a}_t)| = M - N$. Any $\mathbf{c} \in \mathcal{C}_k(\mathbf{a}_t)$ can be split into two sub-vectors $\mathbf{c}' = [c_i]_{i\in\mathcal{I}(\mathbf{a}_t)}$ and $\mathbf{c}'' = [c_i]_{i\in\mathcal{I}^{\perp}(\mathbf{a}_t)}$, where the entries of $\mathbf{c}''$ can be either

0 or 1, that is $\mathbf{c}'' \in \{0,1\}^{M-N}$, while $\mathbf{c}' \in \mathcal{C}_k' \triangleq \{\mathbf{c}' : \mathbf{1}_N^T\mathbf{c}' = k\}$. Therefore, (4.19) can be rewritten as

$$\begin{aligned}
\bar{R}(\mathbf{a}_t, \boldsymbol{\omega}_t) &= \sum_{k=0}^{N} R_k \sum_{\mathbf{c}' \in \mathcal{C}_k'} \sum_{\mathbf{c}'' \in \{0,1\}^{M-N}} \prod_{i \in \mathcal{I}(\mathbf{a}_t)} \hat{f}(\omega_{i,t}, c_i') \prod_{i \in \mathcal{I}^{\perp}(\mathbf{a}_t)} \hat{f}(\omega_{i,t}, c_i'') \\
&= \sum_{k=0}^{N} R_k \left( \sum_{\mathbf{c}' \in \mathcal{C}_k'} \prod_{i \in \mathcal{I}(\mathbf{a}_t)} \hat{f}(\omega_{i,t}, c_i') \right) \underbrace{\left( \sum_{\mathbf{c}'' \in \{0,1\}^{M-N}} \prod_{i \in \mathcal{I}^{\perp}(\mathbf{a}_t)} \hat{f}(\omega_{i,t}, c_i'') \right)}_{=1} \\
&= \sum_{k=0}^{N} R_k \sum_{\mathbf{c}' \in \mathcal{C}_k'} \prod_{i \in \mathcal{I}(\mathbf{a}_t)} \hat{f}(\omega_{i,t}, c_i'), \qquad (4.20)
\end{aligned}$$

where the expression above the bracket is equal to 1 because it is the sum of the probabilities of all possible values of $\mathbf{c}''$ may take. It can be seen from (4.20) that $\bar{R}(\mathbf{a}_t, \boldsymbol{\omega}_t)$ depends only on $[\omega_{i,t}]_{i \in \mathcal{I}(\mathbf{a}_t)}$. Let us define

$$f(\mathbf{x}) \triangleq \sum_{k=0}^{N} R_k \sum_{\mathbf{1}_N^T \mathbf{c}' = k} \prod_{i=1}^{N} \hat{f}(x_i, c_i'). \qquad (4.21)$$

Then for $\mathbf{x} = [\omega_{i,t}]_{i \in \mathcal{I}(\mathbf{a}_t)}$, we can write (4.20) as $f([\omega_{i,t}]_{i \in \mathcal{I}(\mathbf{a}_t)}) \triangleq \bar{R}(\mathbf{a}_t, \boldsymbol{\omega}_t)$. We now rigorously prove that for positively correlated two-state channels, the myopic policy, which maximizes the expected immediate reward (i.e., expected immediate achievable rate) is optimal for the antenna selection problem (4.6), meaning that this policy maximizes the expected long-term reward. To this end, we need to first prove that for positively correlated states i.e., when $p_{01} < p_{11}$, $f(\mathbf{x})$ in (4.21) is regular, as required by the following theorem [41].

**Theorem 2.** *(Optimality of myopic policy [41]): When $p_{01} < p_{11}$, if $f(\mathbf{x})$ is regular, then the myopic policy maximizes the long-term expected reward.*

The definition of a regular function is given as follows.

**Definition 1.** *For $\mathbf{x} = [x_1 \; x_2 \; \cdots \; x_N]^T$, function $f(\mathbf{x})$ is called regular with respect to $\mathbf{x}$, if it satisfies the following three conditions:*

**C1:** $f(\mathbf{x})$ is symmetric, i.e., if, for any $j, l$, $f(\mathbf{x})$ satisfies

$$f([x_1 \ \cdots \ x_j \ \cdots \ x_l \ \cdots \ x_N]^T) = f([x_1 \ \cdots \ x_l \ \cdots x_j \ \cdots \ x_N]^T). \tag{4.22}$$

**C2:** $f(\mathbf{x})$ is decomposable, i.e., if, for $j = 1, ..., N$, $f(\mathbf{x})$ satisfies

$$f([x_1 \ \cdots \ x_j \ \cdots \ x_N]^T) = x_j f([x_1 \ \cdots \ 1 \ \cdots \ x_N]^T) + (1 - x_j) f([x_1 \ \cdots \ 0 \ \cdots \ x_N]^T). \tag{4.23}$$

**C3:** $f(\mathbf{x})$ is monotonically increasing in each entry of $\mathbf{x}$, i.e., if, for any $j$, $x_j > x'_j$, $f(\mathbf{x})$ satisfies

$$f([x_1 \ \cdots \ x_j \ \cdots \ x_M]^T) > f([x_1 \ \cdots \ x'_j \ \cdots \ x_M]^T). \tag{4.24}$$

We now prove that $f(x)$ in (4.21) is regular.

**Lemma 1.** *The function* $f(\mathbf{x})$ *in* (4.21) *is a regular function.*

*Proof.* We proof $f(\mathbf{x})$ satisfies **C1**, **C2**, and **C3** in Definition 1. To prove that $f(\mathbf{x})$ satisfies **C1**, we can write

$$\begin{aligned}
& f([x_1 \ \cdots \ x_j \ \cdots \ x_l \ \cdots \ x_N]^T) - f([x_1 \ \cdots \ x_l \ \cdots \ x_j \ \cdots \ x_N]^T) \\
& = \sum_{k=0}^{N} R_k \sum_{\mathbf{1}_N^T \mathbf{c}' = k} \hat{f}(x_j, c'_j) \hat{f}(x_l, c'_l) \prod_{\substack{i=1 \\ i \neq j,l}}^{N} \hat{f}(x_i, c'_i) - \\
& \sum_{k=0}^{N} R_k \sum_{\mathbf{1}_N^T \mathbf{c}' = k} \hat{f}(x_l, c'_j) \hat{f}(x_j, c'_l) \prod_{\substack{i=1 \\ i \neq j,l}}^{N} \hat{f}(x_i, c'_i) = \\
& \sum_{k=0}^{N} R_k \sum_{\mathbf{1}_N^T \mathbf{c}' = k} \underbrace{\Big( \prod_{\substack{i=1 \\ i \neq j,l}}^{N} \hat{f}(x_i, c'_i) \Big) \big( \hat{f}(x_j, c'_j) \hat{f}(x_l, c'_l) - \hat{f}(x_l, c'_j) \hat{f}(x_j, c'_l) \big),}_{\triangleq l(\mathbf{c}')}
\end{aligned} \tag{4.25}$$

where

$$l(\mathbf{c}') = \begin{cases} \big(x_j x_l - x_l x_j\big) \prod\limits_{\substack{i=1 \\ i\neq j,l}}^{N} \hat{f}(x_i, c_i') = 0, & \text{if } c_j' = 1,\ c_l' = 1, \\ \underbrace{\big(x_j(1-x_l) - x_l(1-x_j)\big)}_{x_j - x_l} \prod\limits_{\substack{i=1 \\ i\neq j,l}}^{N} \hat{f}(x_i, c_i'), & \text{if } c_j' = 1,\ c_l' = 0, \\ \big((1-x_j)(1-x_l) - (1-x_l)(1-x_j)\big) \prod\limits_{\substack{i=1 \\ i\neq j,l}}^{N} \hat{f}(x_i, c_i') = 0, & \text{if } c_j' = 0,\ c_l' = 0, \\ \underbrace{\big((1-x_j)x_l - (1-x_l)x_j\big)}_{x_l - x_j} \prod\limits_{\substack{i=1 \\ i\neq j,l}}^{N} \hat{f}(x_i, c_i'), & \text{if } c_j' = 0,\ c_l' = 1. \end{cases} \tag{4.26}$$

Using the fact that if $\mathbf{1}_N^T[c_1' \ \cdots \ c_l' \ \cdots \ c_j' \ \cdots \ c_N']^T = k$, then we can write $\mathbf{1}_N^T[c_1' \ \cdots \ c_j' \ \cdots \ c_l' \ \cdots \ c_N']^T = k$ in the second and forth cases in (4.26), we can write $\sum\limits_{\mathbf{1}_N^T\mathbf{c}'=k} l(\mathbf{c}') = 0$. Thus, (4.25) is equal to zero, and thus, $f(\mathbf{x})$ is a symmetric function.

To show that $f(\mathbf{x})$ is decomposable as in **C2**, we rewrite (4.21) as

$$f(\mathbf{x}) = \sum_{k=0}^{N} \sum_{\mathbf{1}_N^T\mathbf{c}'=k} R_{\|\mathbf{c}'\|_1} \underbrace{\Big( \prod_{\substack{i=1 \\ i\neq j}}^{N} \hat{f}(x_i, c_i') \Big)}_{\triangleq Q(\mathbf{x}_{-j}, \mathbf{c}'_{-j})} \hat{f}(x_j, c_j'), \tag{4.27}$$

where $\mathbf{c}'_{-j}$ is the same as $\mathbf{c}'$ with the $j$-th entry, $c_j'$, removed and $\mathbf{x}_{-j}$ is similarly defined. Also, since $\|\mathbf{c}'\|_1 = \mathbf{1}_N^T\mathbf{c}' = k$, then $R_{\|\mathbf{c}'\|_1}$ is equivalent to $R_k$ in (4.20). We can further rewrite (4.27) as

$$f(\mathbf{x}) = \sum_{k=0}^{N-1} \sum_{\substack{\mathbf{1}_N^T\mathbf{c}'=k \\ c_j'=0}} R_{\|\mathbf{c}'\|_1} Q(\mathbf{x}_{-j}, \mathbf{c}'_{-j})(1-x_j) + \sum_{k=1}^{N} \sum_{\substack{\mathbf{1}_N^T\mathbf{c}'=k \\ c_j'=1}} R_{\|\mathbf{c}'\|_1} Q(\mathbf{x}_{-j}, \mathbf{c}'_{-j}) x_j. \tag{4.28}$$

Since $\|\mathbf{c}'\|_1 = \|\mathbf{c}'_{-j}\|_1 + c_j'$, we can rewrite (4.28) as

$$f(\mathbf{x}) = \sum_{k=0}^{N-1} \sum_{\substack{\mathbf{1}_N^T\mathbf{c}'=k \\ c_j'=0}} R_{\|\mathbf{c}'_{-j}\|_1} Q(\mathbf{x}_{-j}, \mathbf{c}'_{-j})(1-x_j) + \sum_{k=0}^{N-1} \sum_{\substack{\mathbf{1}_N^T\mathbf{c}'=k+1 \\ c_j'=1}} R_{1+\|\mathbf{c}'_{-j}\|_1} Q(\mathbf{x}_{-j}, \mathbf{c}'_{-j}) x_j$$

$$
\begin{aligned}
&= \sum_{k=0}^{N-1} \Big( \sum_{\mathbf{1}_{N-1}^T \mathbf{c}'_{-j}=k} R_{1+\|\mathbf{c}'_{-j}\|_1} Q(\mathbf{x}_{-j}, \mathbf{c}'_{-j}) - \sum_{\mathbf{1}_{N-1}^T \mathbf{c}'_{-j}=k} R_{\|\mathbf{c}'_{-j}\|_1} Q(\mathbf{x}_{-j}, \mathbf{c}'_{-j}) \Big) x_j \\
&+ \sum_{k=0}^{N-1} \sum_{\mathbf{1}_{N-1}^T \mathbf{c}'_{-j}=k} R_{\|\mathbf{c}'_{-j}\|_1} Q(\mathbf{x}_{-j}, \mathbf{c}'_{-j}) \\
&= \sum_{k=0}^{N-1} \Big( \sum_{\mathbf{1}_{N-1}^T \mathbf{c}'_{-j}=k} Q(\mathbf{x}_{-j}, \mathbf{c}'_{-j}) (R_{1+\|\mathbf{c}'_{-j}\|_1} - R_{\|\mathbf{c}'_{-j}\|_1}) \Big) x_j + \\
& \sum_{k=0}^{N-1} \sum_{\mathbf{1}_{N-1}^T \mathbf{c}'_{-j}=k} R_{\|\mathbf{c}'_{-j}\|_1} Q(\mathbf{x}_{-j}, \mathbf{c}'_{-j}) = \eta_j \ x_j + \theta_j,
\end{aligned}
\tag{4.29}
$$

where $\eta_j \triangleq \sum_{k=0}^{N-1} \sum_{\mathbf{1}_{N}^T \mathbf{c}'_{-j}=k} Q(\mathbf{x}_{-j}, \mathbf{c}'_{-j}) (R_{1+\|\mathbf{c}'_{-j}\|_1} - R_{\|\mathbf{c}'_{-j}\|_1})$ and also we write $\theta_j \triangleq \sum_{k=0}^{N-1} \sum_{\mathbf{1}_{N}^T \mathbf{c}'_{-j}=k} R_{\|\mathbf{c}'_{-j}\|_1} Q(\mathbf{x}_{-j}, \mathbf{c}'_{-j})$. Note that $\eta_j > 0$ as $R_{1+\|\mathbf{c}'_{-j}\|_1} > R_{\|\mathbf{c}'_{-j}\|_1}$ holds true. We can now write

$$
\begin{aligned}
f(\mathbf{x}) &= x_j(\eta_j \ + \theta_j) + (1 - x_j)\theta_j \\
&= x_j f([x_1 \ \cdots \ x_{j-1} \ 1 \ \cdots \ x_N]^T) + (1 - x_j) f([x_1 \ \cdots \ x_{j-1} \ 0 \ \cdots \ x_N).
\end{aligned}
\tag{4.30}
$$

Hence, $f(\mathbf{x})$ is a decomposable function according to **C2**. We now show that $f(\mathbf{x})$ is monotone as in **C3**. To show (4.24), based on (4.29), we can write

$$
f(\mathbf{x}) - f(\mathbf{x}') = \eta_j (x_j - x'_j) > 0 \tag{4.31}
$$

where we use the fact that $\eta_j > 0$ as, by (4.17), $R_{1+\|\mathbf{c}'_{-j}\|_1} > R_{\|\mathbf{c}'_{-j}\|_1}$ holds true. Thus, $f(\mathbf{x})$ is monotone and the proof is complete. Based on the above discussions, $f(x)$ satisfies **C1**, **C2**, and **C3** in Definition 1 and thus is a regular function. ■

Based on Lemma 1 and Theorem 2, we can now show easily that the myopic policy, which is equivalent to selecting those $N$ antennas that have the highest probability of being in good state, is optimal.

**Theorem 3.** *Assume the two-state positively correlated channel model considered in* (4.9). *For the antenna selection problem in* (4.6) *to choose* $N$ *out of* $M$ *antennas, for* *any $N < M$, the myopic policy is optimal. Specifically, the myopic policy amounts to selecting those $N$ antennas which correspond to the $N$ largest entries in the belief vector $\mathbf{b}_t$ at current time slot $t$.*

*Proof.* Lemma 1 holds for any $N < M$. Thus, by Theorem 2, it is straightforward to conclude that, for any $N < M$, the myopic policy is optimal. To see that the myopic policy corresponding to the $N$ largest entries of $\mathbf{b}_t$, we rely on **C3**: According to this condition, given $\mathbf{x}_{-j}$, function $f(\mathbf{x})$ is monotonically increasing in $x_j$. Hence, it immediately follows from (4.20) that selecting the $N$ largest elements in $\mathbf{x}$ (i.e. the $N$ largest $\omega_{i,t}$'s) maximizes the expected immediate reward $f(\mathbf{x})$. The proof is complete. ∎

Note that intuitively, the myopic policy is to select those $N$ antennas that have the highest probabilities (based on the current knowledge) of being in a good state in the next time slot. It is worth mentioning that the authors of [30, 37–39] prove the optimality of myopic policy for the case when *only one* ($N = 1$) resource is to be selected out of the total available resources. Here, we proved that the myopic policy is optimal for our antenna selection problem for any $1 \leq N < M$. Also our results differs from [40], which focuses on the channel selection in multi-channel opportunistic spectrum access for $N > 1$. The difference lies in the fact that [40] assumes one unit of reward for selecting a good-state channel regardless of the quality of the selected channel. In our approach, however, by using the data rate at each time slot, we take into account that reward depends on the quality of the selected channels. For this very same reason, the proof of [40] is not applicable to our problem. And this is exactly where the novelty of our result resides (see [31,59]).

We also point out that, besides its optimality, a significant advantage of the myopic policy is that it incurs minimum computational complexity to determine the selected antennas, as compared to the optimal policy for the general POMDP, where the latter is PSPACE-complete in general [60]. This allows us to apply the myopic policy to the multi-antenna selection problems in massive MIMO systems.

Although the optimality of the myopic policy in Theorem 3 is herein proved only for the two-state channel model (which is rarely the case for the actual fading channels), this policy can be used as a low-complexity method for the antenna selection problem for practical implementation in massive MIMO system. In the next section, we develop an algorithm using the myopic policy for antenna selection over the Rayleigh fading channels for perfect CSI (PCSI) and imperfect (ICSI) scenarios.

## 4.4 Myopic Policy-based Antenna Selection for Fading Channel

We consider the Gauss-Markov model for channel evolution over time slots. This model is widely adopted to represent the dynamics of the Rayleigh fading channels [32, 33]. The first-order Gauss-Markov channel model is given by

$$\mathsf{h}_{i,t} \triangleq \xi \mathsf{h}_{i,t-1} + \sqrt{1-\xi^2}\mathsf{z}_{i,t}, \qquad i = 1, ..., M. \tag{4.32}$$

where $\mathsf{h}_{i,t} \sim \mathcal{CN}(0, \sigma_h^2)$ is the channel between the $i$-th BS antenna and the user at time slot $t$, with $\sigma_h^2$ being the channel variance, the process $\{\mathsf{z}_{i,t}\}$ is the i.i.d. innovation sequence with $\mathsf{z}_{i,t} \sim \mathcal{CN}(0, \sigma_h^2)$, which is independent of the channel $\mathsf{h}_{i,t}$, for $i = 1, ..., M$, and, $\xi \in [0, 1]$ is the fading correlation coefficient. The value of $\xi$ depends on the maximum Doppler frequency [61], with $\xi = 1$ representing a static channel and $\xi = 0$ indicating the channel being i.i.d over $t$.

For the purpose of antenna selection, we quantize the channel coefficients into two values of $\alpha$ and $\beta$, assuming a two-state Markov channel model and then apply the optimal myopic policy for antenna selection decision. That is, the quantized channel coefficient $\mathsf{h}'_{i,t}$ is given by

$$\mathsf{h}'_{i,t} = \begin{cases} \alpha, & \text{if } |\mathsf{h}_{i,t}| \geqslant v, \\ \beta, & \text{if } |\mathsf{h}_{i,t}| < v \ . \end{cases} \tag{4.33}$$

where $v$ is the quantization threshold for the channel amplitude, and $|\alpha| > |\beta|$ must hold. The value of $v$ determines the transition probabilities for the quantized two-state channels, and will dictate the resulting data rate under the myopic policy for antenna selection. Note that the optimal value of $v$, which maximizes the time-averaged expected data rate, depends on the value of the channel correlation coefficient $\xi$. Such dependency, however, can not be expressed in a closed form [62, 63]. For any given value $\xi$, we have to resort to numerical simulations in order to obtain the optimal value of $v$. Given the so-obtained value of $v$, the transition probabilities $p_{11}$ and $p_{01}$ can be obtained empirically.

Based on the quantized channel coefficient described above, we select the best set of $N$ antennas for data transmission at each time slot $t$, by applying the myopic policy summarized as Algorithm 2. Specifically, at each time slot $t$, based on selection decision $\mathbf{a}_t$, the BS obtains vector $\mathbf{o}_t$, i.e., the channel coefficients of the $N$ selected antennas at the end of the time slot. Quantization is then performed for each entry of $\mathbf{o}_t$ according to (4.33). Using the quantized channels, those entries of $\mathbf{c}_t$ which correspond to the selected antennas are obtained. Next, the belief vector $\omega_{i,t+1}$'s are updated using (4.12) . At the beginning of time slot $t+1$, the antenna selection is made by setting we set $a_{i,t+1} = 1$ (i.e., the $i$-th antenna is selected), if $\omega_{i,t+1}$ is among the $N$ largest entries of $\boldsymbol{\omega}_{t+1}$, otherwise $a_{i,t+1} = 0$. The antennas selected by $\mathbf{a}_{t+1}$ is then used for transmission for time slot $t+1$.

**Algorithm 2** The myopic policy based antenna selection for point-to-point systems

```
Inputs: Set the threshold value $v$ based on $\xi$ and $\sigma_h^2$ .
At each time slot $t$:
Input: $\mathbf{o}_t$
  Quantize the elements of $\mathbf{o}_t$ into $\alpha$ and $\beta$ using (4.33) and update the elements of $\mathbf{c}_t$ which correspond to the currently selected antennas based on (4.11).
  Update $\boldsymbol{\omega}_{t+1}$ using (4.12).
  For $i = 1, 2, \cdots, M$, choose the $i$-th entry of $\mathbf{a}_{t+1}$ as
      $a_{i,t+1} = \begin{cases} 1, & \text{if } \omega_{i,t+1} \text{ among the largest } N \text{ entries of } \boldsymbol{\omega}_{t+1}, \\ 0, & \text{otherwise.} \end{cases} \quad (4.34)$
Output: $\mathbf{a}_{t+1}$
```

**Remark 1:** It is worth noting that there are different strategies to schedule users in MIMO systems [46]. Allocating one user per time-slot is one of the common user scheduling schemes [64–66]. There are extensive existing studies where the authors proposed various designs based on one user per time-slot scheduling scheme in MIMO systems [67, 68]. Moreover, the scheme we are considering can be used in point-to-point MIMO systems, where the receiver can have one or more antennas.

We would like to emphasize that the main contribution of this study is to propose a simple yet optimal POMDP antenna selection algorithm that can be easily implemented for large-scale antenna arrays with any given number of antennas and available RF chains. It is worth mentioning that, unlike the studies in [30, 40], where the authors use a simple reward model where one unit of reward is accrued when data is received, in our proposed method, the reward function is the actual rate achieved by MISO beamforming. To the best of our knowledge, under such a realistic and complex reward function, this is the first study that proves the optimality of myopic policy for any given number of antennas and available RF chains. Extending this study to a multi-user scenario involves additional technical difficulties. For a multi-user scheme, the main goal is to find the best $N$ out of $M$ antennas to serve $K$ users such that the expected long term data rate is maximized. In this scenario, a selection policy, which accounts for the channel quality of all users, is needed. Such a policy may not be as simple as the myopic policy because the expected immediate reward function is not a regular function anymore. Finding a proper framework to formulate a POMDP-based antenna selection policy for multi-user cases is a challenging task and needs a full investigation. We consider addressing this challenge in our next step in this line of work that is presented in Chapter. 5.

**Remark 2:** As indicated in the inputs of Algorithm 2, the optimal threshold value depends on the channel correlation over time $\xi$ and variance $\sigma_h^2$. Note that the relationship between the optimal threshold value $v$ and $\sigma_h$ is linear. The reason is that if for given $M$, $N$, $\xi$, and $\mathsf{h}_{i,t} \sim \mathcal{CN}(0, \sigma_h^2)$, the optimal threshold value is $v$, for the same values of $M$, $N$, and $\xi$ but for $\mathsf{h}_{i,t}^{''} \sim \mathcal{CN}(0, \sigma_h^{''2})$, we can scale $\mathsf{h}_{i,t}^{''}$ by $\frac{\sigma_h}{\sigma_h^{''}}$, and thus, assume the channel is $\frac{\sigma_h}{\sigma_h^{''}}\mathsf{h}_{i,t}^{''} \sim \mathcal{CN}(0, \sigma_h^2)$. Such a scaling allows us to use the same value of $v$ as the optimal threshold value for $\frac{\sigma_h}{\sigma_h^{''}}\mathsf{h}_{i,t}^{''}$. In other words, the optimal threshold value for $\mathsf{h}_{i,t}^{''} \sim \mathcal{CN}(0, \sigma_h^{''2})$ is $\frac{\sigma_h^{''}}{\sigma_h}v$, meaning that the threshold $v$ can be scaled accordingly based on $\sigma_h$ of the channel of interest.

**Remark 3:** The computational complexity of the value iteration algorithm is $\mathcal{O}(|\mathcal{S}|^2 \times |\mathcal{A}|)$ per iteration [69]. In our problem, $|\mathcal{S}| = 2^M$ and $|\mathcal{A}| = \binom{M}{N}$. The computational complexity of our proposed myopic antenna selection algorithm resides in updating the elements of the belief vector with the computational complexity $\mathcal{O}(M - N)$ (see (4.12)) Sorting the $M \times 1$ belief vector with computational complexity $\mathcal{O}(M \log N)$ [70]. Thus, the total computational complexity of the myopic policy is only $\mathcal{O}(M \log N)$, which is significantly less than that of the value iteration algorithm.

**Remark 4:** In reality, single-antenna users is a common assumption [67, 68, 71]. However, for scenarios with multi-antenna users, we can use singular value decomposition (SVD) to turn the channel into multiple parallel MISO channels for eigen-beamforming, as typically considered in MIMO systems.. Consider a point-to-point scenario between nodes A and B, where node A uses all of its antennas while node B performs antenna selection. In this case, node A can precode its transmitted data using the conjugate of the left principal singular vector of the channel matrix from node A to node B and receive its data by using left principal singular vector as a receive-beamformer, thereby turning the MIMO channel into a MISO channel. Thus our proposed antenna selection algorithm can be applied to a massive antenna array base station which transmits data to a multi-antenna user.

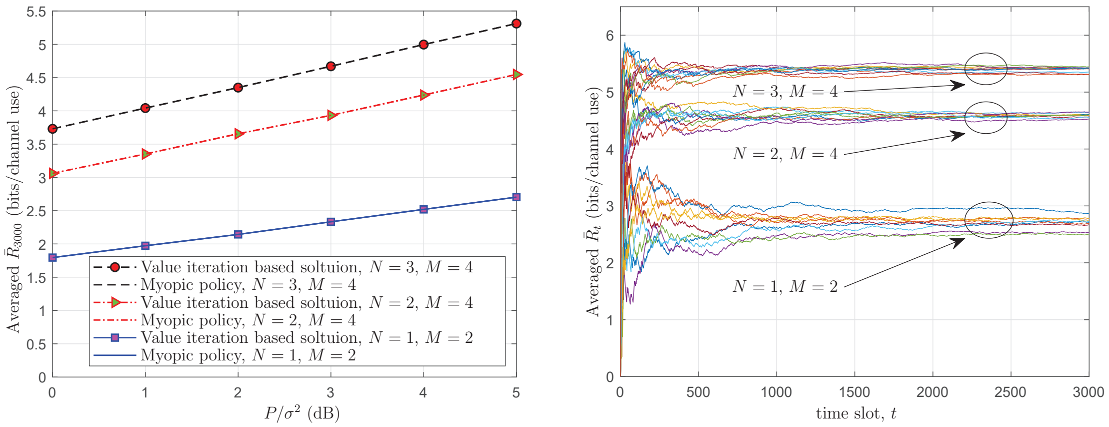
## 4.5 Simulation and Performance Analysis

This section demonstrates the performance of our proposed solutions for the antenna selection problem. For all the simulation runs, we set the transmitter's SNR as $\frac{P}{\sigma^2} = 5$ dB. We study the performance of the optimal POMDP solution for the antenna selection problem in terms of the time-average rate $\bar{R}_t$, defined as $\bar{R}_t \triangleq \frac{1}{t}\sum_{\tau=0}^{t} R(\mathbf{s}_\tau, \mathbf{a}_\tau)$. which is averaged over 100 Monte Carlo runs to obtain its average value.

### 4.5.1 Two-State Channel Model

**POMDP Policy**

In this subsection, using the two-state channel model as an example, we verify the optimality of the myopic policy, as stated in Theorem 1. In particular, we show the convergence behavior of Algorithm 1 for different numbers of antennas $M$ and different numbers of RF chains $N$. Throughout our simulations, we assume that the channel coefficients are i.i.d., and each evolves according to a two-state Markov chain, i.e., $Q = 2$ with $\{\alpha_1, \alpha_2\} = \{\sqrt{0.1}, \sqrt{10}\}$ and the transition probabilities $p_{01} = 0.2$ and $p_{11} = 0.8$. Since channel coefficients are independent, $\mathbf{T}$ can be obtained from the transition probability of the each channel coefficient state. Here, we use the existing POMDP solver software [56] to find the optimal policy of Algorithm 1. In our first series of experiments, we aim to analyze the effect of $M$ and $N$ on the performance of Algorithm 1 for the antenna selection problem, and compare its



(a) (b)

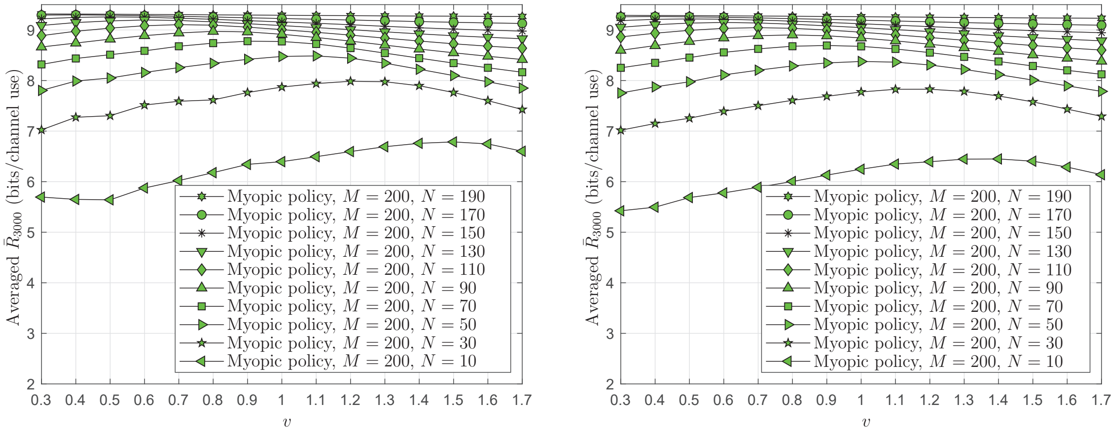
Figure 4.3: For $Q = 2$, $p_{01} = 0.2$, and $p_{11} = 0.8$ (a) time-average rate $\bar{R}_{3000}$ versus $P/\sigma^2$ (b) time-averaged rate $\bar{R}_t$ of 10 Monte Carlo simulation runs for $P/\sigma^2 = 5$ dB .

performance with that of the myopic policy.

Fig. 4.3a shows the average of $\bar{R}_{3000}$ versus $P/\sigma^2$ under both Algorithm 1 and the myopic policy. The results verify that the myopic solution is optimal for the POMDP based antenna selection problem. Fig. 4.3b shows the time trajectory of $\bar{R}_t$ for Algorithm 1, versus time slot $t$, for different pairs of $N$ and $M$ and for 10 simulation runs. As it can be seen from this figure, the POMDP based technique converges in about 1000 time slots.

### 4.5.2 Gauss-Markov Channel Model

In this subsection, considering the Rayleigh fading Gauss-Markov channel model in (4.39) with $\sigma_h^2 = 1$, we apply Algorithm 2 for the antenna selection problem in a massive MIMO system for both perfect CSI and imperfect CSI. Unlike the previous example, the quantized channels are used solely for antenna selection purpose, and the average data rate $\bar{R}_t$ is calculated based on the actual fading channel coefficients. This performance depends on the transition probabilities which in turn depend on the quantization threshold $v$. The optimal value of $v$ depends on $M$, $N$, and $\xi$. Unfortunately, the relationship between optimal $v$ and different values of $M$, $N$,



(a)

(b)

Figure 4.4: Average $\bar{R}_{3000}$ versus $v$ for $P/\sigma^2 = 5$ dB (a) Gauss-Markov channel with $\xi = 0.999$ (b) Gauss-Markov channel with $\xi = 0.99$.

and $\xi$ does not appear to be amenable to a closed-from expression. As such, we resort to numerical simulations to find the optimal values of $v$ for two different scenarios. In the first scenario, considering $M = 200$ and for different values of $N = [10 : 20 : 200]$, we plot the averaged $\bar{R}_{3000}$ versus different values of threshold $v$. The result is shown in Figs. 4.4a and 4.4b for Gauss-Markov channel with $\xi = 0.999$ and $\xi = 0.99$, respectively. These figures show that the optimal value of threshold $v$ depends on both $M$ and $N$. Fig. 4.5 shows the optimal values of $v$ obtained empirically based on Figs. 4.4a and 4.4b for $M = 200$ and $N = [10 : 20 : 200]$ for different values of $\xi$. We observe that the summery of the first scenario results with one extra plot of Gauss-Markov channel with $\xi = 0.97$. In this figure, we plot the optimal value of $v$ for $M = 200$ versus number $N$ of the selected antennas to show that for given $M$, the optimal value of $v$ is insensitive to the channel variation parameter $\xi$, specially for $N > 70$. When $M$ and $N$ values are close to each other, e.g., $M = 200$ and $N = 170$, the optimal threshold value is relatively small ($v = 0.4$). This is due to the fact that when the values of $M$ and $N$ are close, the number of choices for antenna selection is rather limited. Hence, we should be less selective in labeling channels as good by choosing a small $v$. As $N$ becomes small, the number

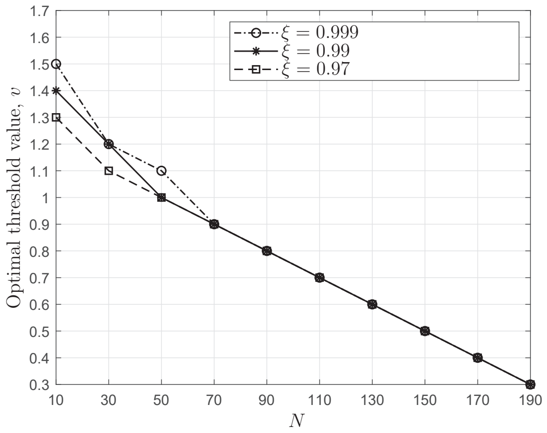


Figure 4.5: Optimal threshold value, $v$ versus $N$ for $M = 200$ and $P/\sigma^2 = 5$ dB.

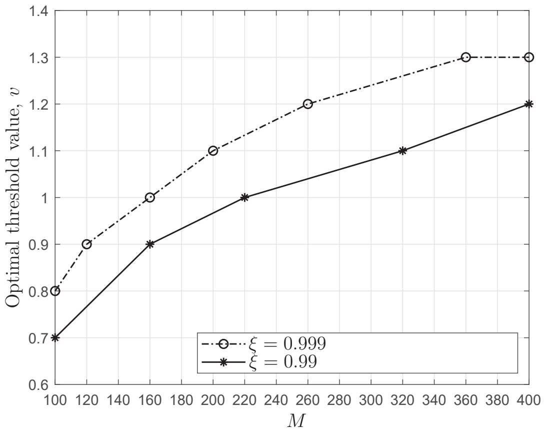
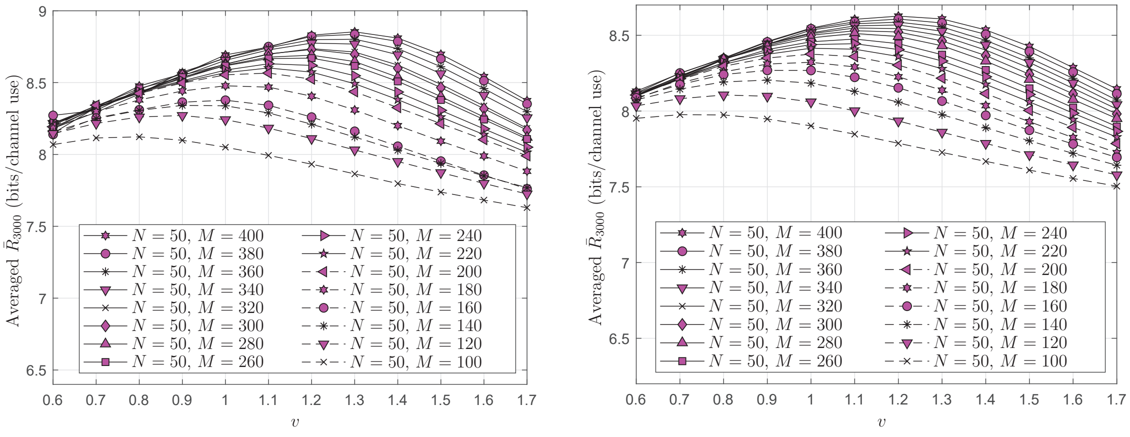


Figure 4.6: Optimal threshold value $v$, versus $M$ for $P/\sigma^2 = 5$ dB.

of possible antenna selections increases, allowing us to set a higher threshold on $v$ to be more opportunistic in selecting good channels,

Fig. 4.6 shows the corresponding optimal threshold values of $v$, obtained from Figs.4.7a and 4.7b, versus $M$. Indeed, fixing $N = 50$ and for different values of $M = [100 : 20 : 400]$, in Figs. 4.7a and 4.7b, we plot $\bar{R}_{3000}$, averaged over 100 Monte Carlo simulation runs, versus different values of threshold $v$, for Gauss-Markov channel models with $\xi = 0.999$ and $\xi = 0.99$, respectively. Fig. 4.6 also shows that for fixed $N$, when $M$ is increased, the optimal threshold $v$ is increased, meaning that we need to be more selective in labeling channels as good, as more choices for antenna selection become available, and vice versa.

In the remainder of our numerical results, for each simulation run, we use the corresponding optimal $v$ for quantization in Algorithm 2. Next, we compare the performance of Algorithm 2 based on the myopic policy, with two other schemes: 1) A random antenna selection policy, and 2) a selection policy (referred to as the full perfect CSI based policy) that uses full perfect CSI of $M$ channels to select $N$ antennas with the $N$ highest channel amplitudes. Considering a BS with $M = 200$ available antennas, in Fig. 4.8a (4.8b), we plot average of $\bar{R}_{3000}$ versus $N$ ($M$) for fixed $M$ ($N$) and for different scenarios of channel variations tabulated in Table 4.1.



(a) (b)

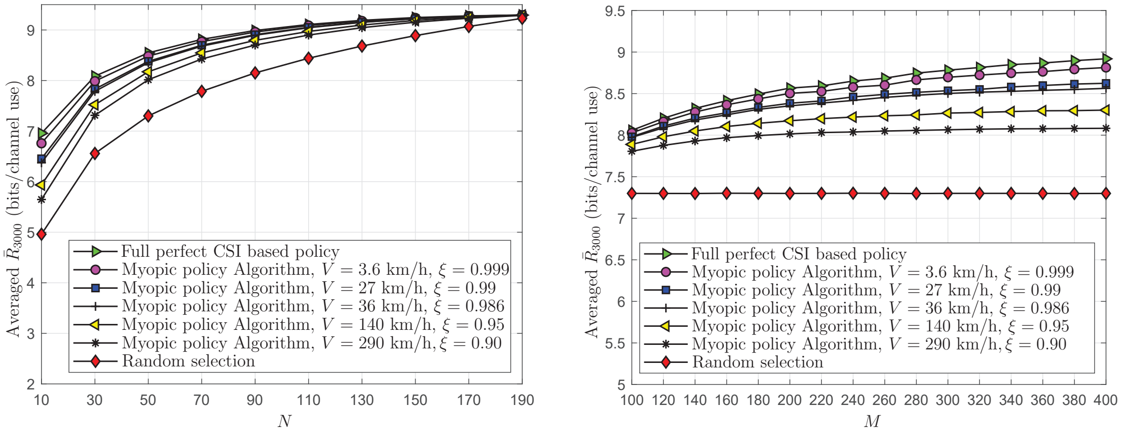
Figure 4.7: Average $\bar{R}_{3000}$ versus $v$ for $P/\sigma^2 = 5$ dB (a) Gauss-Markov channel with $\xi = 0.999$ (b) Gauss-Markov channel with $\xi = 0.99$.

Table 4.1: Values of $\xi$ for different scenarios.

| Scenario | $V$ | $W$ | $T_c$ | $\xi$ |
|---|---|---|---|---|
| WLAN 802.11 operating at 2.4 GHz, pedestrian user | 3.6 km/h | 15 kHz | ∽ 0.3 s | 0.999 |
| LTE network operating at 2.6 GHz, car driving in residential area | 27 km/h | 15 kHz | ∽ 15.3 ms | 0.99 |
| LTE network operating at 2.6 GHz, car driving in residential area | 36 km/h | 15 kHz | ∽ 11.5 ms | 0.986 |
| LTE network operating at 2.6 GHz, high speed car driving in highway | 140 km/h | 15 kHz | ∽ 2.2 ms | 0.95 |
| LTE network operating at 2.6 GHz, high speed train | 290 km/h | 15 kHz | ∽ 1.4 ms | 0.9 |

In this table, the relationship between $\xi$, coherence time $T_{\mathrm{c}}$, and bandwidth, denoted as $W$, is given by $\xi^{T_c W} = \phi$, where $\phi$ is the de-correlation level and is set to 0.1, which is typically determined based on measurements [61]. Table 4.1 shows the values of $\xi$ under different scenarios in a wide range of mobile speed consideration, from pedestrian, vehicle, to high-speed train. For example, the first scenario corresponds to a pedestrian user with velocity $V = 1$ m/s (3.6 km/h). As can be seen in Figs. 4.8a and 4.8b, the performance gap between the myopic policy and the full CSI based policy is at most 1.3 (bcu). In other scenarios, such as when $V = 27$ km/h and $N = 50$, this gap is about 0.35 (bcu) and is at most 0.55 (bcu) for $M = 200$ and $V = 27$ km/h. As can be seen from Fig. 4.8a, for fixed $M$, as $N$ increases, the number of choices for the set of selected antennas decreases. As a result, the gap between the full perfect CSI based policy and the other policies decreases. Furthermore, when the channel variation over time is slow, the performance between Algorithm 2 and the full perfect CSI based scheme is very small (less than 0.2 (bcu)). When the channel variation increases, this performance gap increases only slightly to a maximum of 0.55 (bcu) for a low-speed vehicle. For the high speed vehicle and train scenarios, this gap performance increases to maximum 1.3 (bcu). Fig. 4.8b shows average of $\bar{R}_{3000}$ versus $M$, for $N = 50$ and different scenarios of channel correlation from Table 4.1. This figure also confirms that there is a very small performance gap between the Algorithm 2 and the full perfect CSI based policy. An intuitive explanation behind this small performance gap between these two policies is that in the myopic policy, we use the two-level quantization only for the purpose of antenna selection, while we employ the non-quantized channel to obtain the MISO data rate. This means that for the purpose of antenna selection (which is a binary choice), one only needs coarse knowledge about the channel amplitudes.

To analyze the sensitivity of our proposed myopic policy antenna selection algorithm to uncertainty in our knowledge of the transition probabilities, we consider an estimation error, denoted as $\epsilon$, in the $p_{11}$ and $p_{01}$ and instead of (4.12), we update



(a) (b)

Figure 4.8: Average $\bar{R}_{3000}$ a) versus $N$, for $M = 200$, $P/\sigma^2 = 5$ dB, b) versus $M$, for $N = 50$, $P/\sigma^2 = 5$ dB.

the $i$-th entry of belief vector as

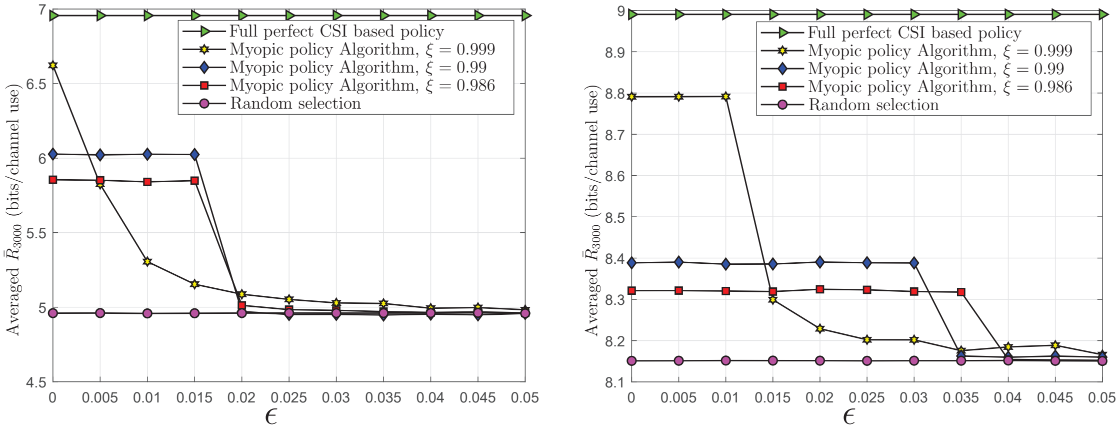
$$\omega_{i,t+1} = \begin{cases} p_{11} + \epsilon, & \text{if } a_{i,t} = 1,\ \ c_{i,t} = 1; \\ p_{01} + \epsilon, & \text{if } a_{i,t} = 1,\ \ c_{i,t} = 0; \\ \omega_{i,t}(p_{11} + \epsilon) + (1 - \omega_{i,t})(p_{01} + \epsilon), & \text{if } a_{i,t} = 0. \end{cases} \tag{4.35}$$

Assuming $M = 200$, we show $\bar{R}_{3000}$ versus $\epsilon$ in Figs. 4.9a and 4.9b for $N = 10$ and $N = 90$, respectively. In these figures, we consider full CSI scheme, our proposed myopic policy algorithm for $\xi = 0.999, 0.99, 0.986$, and the random selection method. As can be seen from these figures, for larger values of $N$, the myopic policy is less sensitive to the value of $\epsilon$. As $\epsilon$ is increased beyond a threshold (whose value increases with $N$), the performance of the myopic policy starts to decrease drastically and approaches to that of the random selection method for large values of $\epsilon$.

### 4.5.3 Imperfect Receiver CSI

We now evaluate the performance of the proposed Algorithm 2, under imperfect CSI available at the receiver side. We define $\breve{\mathsf{h}}_{i,t} \sim \mathcal{CN}(0, \epsilon_h)$ as the $i$-th antenna channel estimation error with variance $\epsilon_h$. The estimated channel of the $i$-th antenna and the corresponding true channel have the following relationship:

$$\mathsf{h}_{i,t} = \hat{\mathsf{h}}_{i,t} + \breve{\mathsf{h}}_{i,t}, \quad i = 1, 2, \cdots, M. \tag{4.36}$$



(a) (b)

Figure 4.9: Average $\bar{R}_{3000}$ a) versus $\epsilon$, for $M = 200$, $N = 10$, $P/\sigma^2 = 5$ dB, b) versus $\epsilon$, for $M = 200$, $N = 90$, $P/\sigma^2 = 5$ dB.

where $\hat{\mathsf{h}}_{i,t}$ and $\breve{\mathsf{h}}_{i,t}$ are statistically independent. To apply Algorithm 2, we need to replace $c_{i,t}$ in (4.12), with its estimate, if the $i$-the antenna is selected, as

$$\hat{c}_{i,t} = \begin{cases} 1 & \text{if } |\hat{h}_{i,t}| \geqslant \upsilon \ . \\ 0 & \text{if } |\hat{h}_{i,t}| < \upsilon \ . \end{cases}, \quad i = 1, 2, \cdots, M. \tag{4.37}$$

where $\hat{h}_{i,t}$ is the observed channel coefficient between the $i$-the antenna and the user.

Based on (4.37), we update the entries of $\boldsymbol{\omega}_{t+1}$ according to (4.12), and Algorithm 2 follows. Due to the imperfect CSI, the reward function is given by

$$\hat{R}(\hat{\mathbf{h}}_t, \mathbf{h}_t, \mathbf{a}_t) \triangleq \log_2\Big(1 + \frac{P| \ \hat{\mathbf{h}}_t^H \operatorname{diag}(\mathbf{a}_t)\mathbf{h}_t \ |}{\sigma^2}\Big), \tag{4.38}$$

where $\hat{\mathbf{h}_t} = [\hat{\mathsf{h}}_{1,t} \ \ \hat{\mathsf{h}}_{2,t} \ \ \cdots \ \ \hat{\mathsf{h}}_{M,t}]^T$ is the estimated CSI vector. In Fig. 4.10, assuming $N = 50$, $M = 200$, and for $\xi = 0.999$ and 0.99, we plot the performance of the following four policies: 1) the full perfect CSI based policy, 2) the full imperfect CSI based policy, which used the full but imperfect CSI, 3) Algorithm. 2, and 4) the random selection policy.

In Algorithm. 2, we choose $\upsilon = 1.1$, for $\xi = 0.999$; and $\upsilon = 1$ when $\xi = 0.99$. These values of $\upsilon$ are optimal for the corresponding scenarios. Note that for the full perfect CSI based policy, we have $\epsilon_h = 0$. As expected, we see that increasing $\epsilon_h$ decreases the rate for our algorithm due to the loss of accuracy of the CSI. The

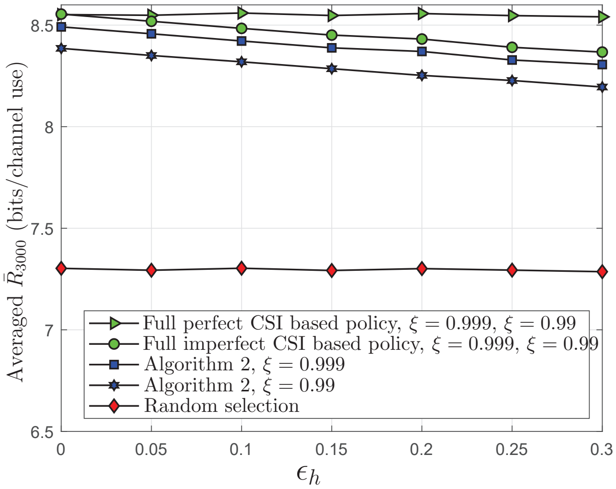


Figure 4.10: Average $\bar{R}_{3000}$ for imperfect CSI versus channel estimation error $\epsilon_h$, for $M = 200$ and $N = 50$.

performance under the random selection policy is not affected by $\epsilon_h$ as the selection does not depend on the CSI. As can be seen from this figure, for $\epsilon_h = 0.3$, Algorithm 2 still performs better than the random selection policy by about 1 bit/channel use, while falls short off of the full perfect CSI based policy by only $0.3-0.4$ bit/channel use.

### 4.5.4 Multi-User Scenario

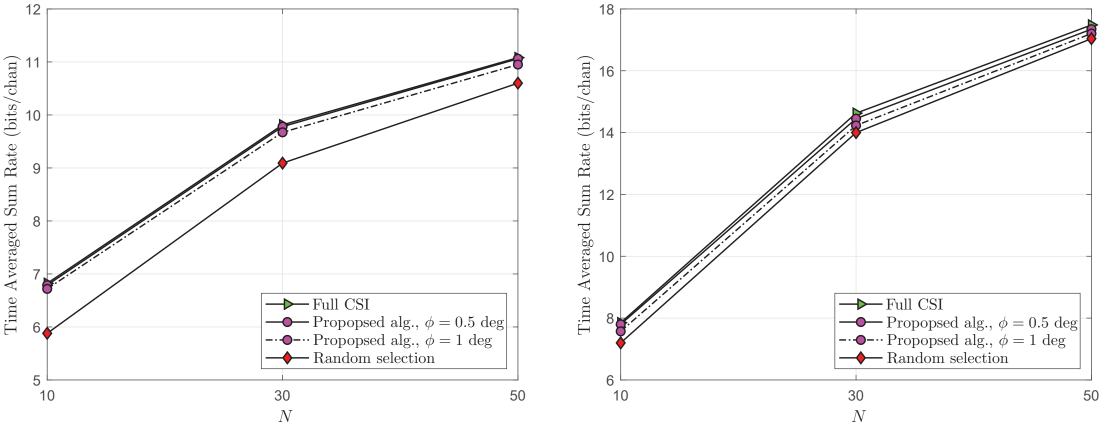
In this subsection, we extend our proposed myopic policy-based antenna selection algorithm to a multi-user scenario and evaluate its performance in this case. To do so, we assume that the BS is equipped with $M$ antennas and $N$ RF chains and serves $K$ users ($K < N$) simultaneously. Here, we assume that the channels between the BS and the users have a direct line-of-sight (LoS) path and follow the Rician fading model. We assume that the LoS components of all $K$ users consists of the common component but differ from each other by a random phase-only component. Such assumptions amount to clustering together those users that are spread in a small geographical area. Indeed, the goal of any clustering technique would be to serve the users that have similar channels. To define the LoS components, we assume a uniform linear antenna array at the BS with inter-element spacing of $\frac{\lambda}{2}$, where $\lambda$ is the carrier wavelength. We denote $\hat{\mathbf{h}}_k = [\hat{h}_{k,1} \ \ \hat{h}_{k,2} \ \ \dots \ \ \hat{h}_{k,M}]^T$ as the LoS portion of the channel vector of the $k$-th user, where $\hat{h}_{k,i}$ is the LoS component of the $i$-th antenna and the $k$-th user, for $i = 1, 2, \dots, M$ and $k = 1, 2, \dots, K$. Given the array geometry, we can write $\hat{h}_{k,i} = \exp(-j(\frac{2\pi d_k}{\lambda} - (i-1)\pi\phi - \psi_{k,i}))$, where $d_k$ is the distance between the $k$-th user and the BS; $\phi = \cos\theta$, where $\theta$ is the angle of the incidence of the common component of the LoS signal with onto the first antenna of the BS; and $\psi_{k,i}$ is a random phase uniformly distributed in the interval $[-\varphi, \varphi]$, where we choose $\varphi$ as 0.5 or 1 degrees. The non-LoS (NLoS) component of the channel of the $i$-th user over all antenna is modelled as $\bar{\mathsf{h}}_{i,t} \sim \mathcal{CN}(\mathbf{0}_{K\times 1}, \mathbf{\Sigma}_h)$, where $\mathbf{\Sigma}_h = E\{\bar{\mathsf{h}}_{i,t}\bar{\mathsf{h}}_{i,t}^H\} = \mathrm{diag}([\sigma^2_{\mathrm{h},k}]_{k=1}^K)$, and $\sigma^2_{\mathrm{h},k}$ is the variance of the NLoS component of channel between the $k$-th user and the $i$-th antenna. We model the time variations of $\mathsf{h}_{i,t}$ using first-order Gauss-Markov channel model as

$$\bar{\mathsf{h}}_{i,t} \triangleq \mathrm{diag}(\boldsymbol{\xi})\bar{\mathsf{h}}_{i,t-1} + \mathrm{diag}(\boldsymbol{\xi}')\mathsf{z}_{i,t}, \qquad i = 1, \dots, M\,. \tag{4.39}$$

Here, $\mathsf{z}_{i,t} \sim \mathcal{CN}(\mathbf{0}_{K\times 1}, \mathbf{\Sigma}_h)$ is independent of the channel vector $\bar{\mathsf{h}}_{i,t}$, for $i = 1, ..., M$ and we define $\boldsymbol{\xi} \triangleq [\xi_1 \quad \xi_2 \quad \cdots \quad \xi_k]$, where $\xi_k \in [0,1]$ is the fading correlation coefficient corresponding to the $k$-th user. Furthermore, we define $\boldsymbol{\xi}'$ as $\boldsymbol{\xi}' = [\sqrt{1-\xi_1^2} \quad \sqrt{1-\xi_2^2} \quad \cdots \quad \sqrt{1-\xi_K^2}]$. The Rician fading channel vector of the $i$-th antenna at time slot $t$, denoted as $\mathsf{h}_{i,t}$, is given as $\mathsf{h}_{i,t} = \hat{h}_i \mathbf{1}_{K\times 1} + \bar{\mathsf{h}}_{i,t}$,for $i = 1, 2, \cdots, M$, where $\mathbf{1}_{K\times 1}$ is a $K \times 1$ vector of all one entries. At time slot $t$, given action $\mathbf{a}_t$, we define the input of Algorithm 2, i.e., the observation vector $\mathsf{o}_t$ as $\mathsf{o}_t = \mathrm{diag}(\mathbf{a}_t)[\frac{\mathbf{1}^T\mathsf{h}_{i,t}}{K}]_{i=1}^M$. That is, the $i$-th entry of $\mathsf{o}_t$ is the average of channel measurements between the $i$-th antenna and all $K$ users. We adopt use zero-forcing beamforming to cancel the inter-user interference, and thus, the time-averaged sum-rate can be written as

$$\hat{R}_t = \frac{1}{t}\sum_{\tau=0}^{t} \log_2 \det \big(\mathbf{I} + \frac{P}{\sigma^2 \|\mathbf{H}_{\mathrm{s},\tau}^H(\mathbf{H}_{\mathrm{s},\tau}\ \mathbf{H}_{\mathrm{s},\tau}^H)^{-1}\|_F^2}\mathbf{I}\big). \tag{4.40}$$

where we define $\mathbf{H}_{\mathrm{s,t}} = [\mathbf{h}_{i,t}]_{i\in\mathcal{I}(\mathbf{a}_t)}$ with $\mathbf{h}_{i,t}$ being a realization $\mathsf{h}_{i,t}$.



(a) Time averaged sum-rate for $K = 2$.

(b) Time averaged sum-rate for $K = 4$.

Figure 4.11: Time averaged sum-rate for $M = 200$, $N = [10 : 30 : 50]$, $\frac{P}{\sigma^2} = 15$ dB, for a) $K = 2$, and b) $K = 4$.

To evaluate the performance of our proposed algorithm for multi-user scenario, we assume $M = 200$, $N = [10 : 30 : 50]$, $\frac{P}{\sigma^2} = 15$ dB, $\theta = 30$ degrees, a carrier frequency of 3 GHz ($\lambda = 0.1$ m), and $\sigma^2_{\mathrm{h},k} = 0.1$ and $\xi_k = 0.999$ for $k = 1, 2, \cdots, K$. The distance $d_k$ is drawn from a uniform distribution with a mean of 2 km and a spread of 0.5 km. In Figs. 5.6a and 5.6b, we show the time-averaged sum-rate versus the number $N$ of RF chains for $K = 2$ and $K = 4$, respectively. As can be seen from these figures, in this multi-user scenario, the performance of our proposed algorithms can be very close to the full CSI based policy, when the LoS components are relatively close to each other. As the difference between the LoS components of different user channels is increased, the performance gap between the proposed policy and the full CSI increase. As these two figures show, this performance gap grows as the number of users increases. Nevertheless, the proposed algorithm performs better than the random selection policy. As this numerical example shows, although not designed for multi-user scenarios, the myopic policy can offer good performance for such scenarios. The extension of this method for multi-user scenarios needs further investigation, and this is exactly what we consider in the next chapter of this dissertation.

# Chapter 5

# POMDP-based Antenna Selection Algorithm in Multi-User MIMO Systems

## 5.1 System Model

We consider a massive MU-MIMO communication system, where a multi-antenna base station (BS), equipped with $M$ antennas and $N$ RF chains ($M \gg N$), aims to communicate with $K$ single-antenna users (see Fig. 5.1). We assume that the number of users is less than the number of the available RF chains ($K< N$). The system is time-slotted and each time slot consists of two phases, namely uplink and downlink. We assume that each channel between the BS and each user evolves over time slots according to a Markov process. As the number of the available RF chains is smaller than the number of BS antennas, only $N$ out of the $M$ antennas can contribute to the data reception/transmission in each time slot. Upon acquiring the channels in the uplink training phase, we select the best $N$ antennas in each time slot with the aim to maximize the expected long-term sum-rate. Note that as we can only observe $N$ out of $M$ available channels during each time slot, the channel state information (CSI) at each time slot is only partially observable. Since each channel evolves according to a Markov process and is only partially observable, a POMDP framework appears to be an appropriate approach to obtain the optimal antenna selection policy to maximize the expected long-term sum-rate.

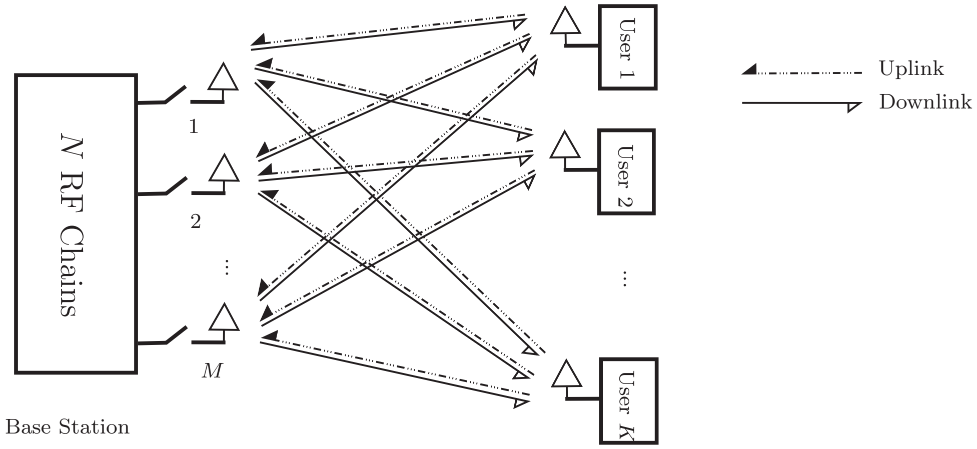


Figure 5.1: Illustration of a Massive MU-MIMO system, where the BS is equipped with $M$ antennas and $N$ RF chains, and communicates with $K$ single-antenna users.

## 5.2 Problem Formulation

To describe the system model, we denote the random channel matrix between the $M$ antennas and the $K$ users at time $t$ as $\mathsf{H}_t \triangleq [\mathsf{h}_{1,t} \quad \mathsf{h}_{2,t} \quad \cdots \quad \mathsf{h}_{M,t}]$, where $\mathsf{h}_{i,t} \in \mathbb{C}^{K\times 1}$ is the random vector of the channel coefficients between all users and the $i$-th antenna at time $t$. At the beginning of time $t$, a subset of $N$ columns of $\mathsf{H}_t$ is selected for transmission during time slot $t$. Let $\mathsf{H}_{\mathrm{s},t} \in \mathbb{C}^{K\times N}$ represent a matrix constructed from such columns of $\mathsf{H}_t$. If $\mathbf{H}_{\mathrm{s}}$ is a realization of $\mathsf{H}_{\mathrm{s},t}$, we can write vector $\mathsf{y} \in \mathbb{C}^{K\times 1}$ of the signals received at all $K$ users as

$$\mathsf{y} = \mathbf{H}_{\mathrm{s}}\mathbf{W}\mathsf{x} + \mathsf{n}, \tag{5.1}$$

where $\mathsf{x} \in \mathbb{C}^{K\times 1}$ is the transmitted signal vector, $\mathsf{n} \in \mathbb{C}^{K\times 1}$ is the noise vector with the noise variance of each element being $\sigma^2$, and $\mathbf{W} \in \mathbb{C}^{N\times K}$ is the precoding (beamforming) matrix. We use zero-forcing beamforming (ZFBF) technique to eliminate the interference among the users. Thus, we can write $\mathbf{W}$ as

$$\mathbf{W} = \mathbf{H}_{\text{s}}^{H}(\mathbf{H}_{\text{s}}\,\mathbf{H}_{\text{s}}^{H})^{-1}\sqrt{\frac{P}{\|\mathbf{H}_{\text{s}}^{H}(\mathbf{H}_{\text{s}}\,\mathbf{H}_{\text{s}}^{H})^{-1}\|_F^2}}\;, \tag{5.2}$$

where $P$ is the total transmit power and $\|\cdot\|_F$ is the Frobenius norm. Based on the singular value decomposition (SVD) of the channel matrix, we can then write

$$\|\mathbf{H}_{\text{s}}^{H}(\mathbf{H}_{\text{s}}\,\mathbf{H}_{\text{s}}^{H})^{-1}\|_F^2 = Tr((\mathbf{H}_{\text{s}}\,\mathbf{H}_{\text{s}}^{H})^{-1}) = \sum_{k=1}^{K}\frac{1}{\lambda_{\text{sk}}^2}, \tag{5.3}$$

where $\lambda_{\text{sk}}$ is the $k$-th singular value of $\mathbf{H}_{\text{s}}$. Using (5.3), we can write the sum-rate of the system as

$$\begin{aligned}\breve{R}(\mathbf{H}_{\text{s}}) &= \log_2\Big(\det\big(\mathbf{I} + \frac{1}{\sigma^2}\mathbf{H}_{\text{s}}\mathbf{W}\mathbf{W}^{H}\mathbf{H}_{\text{s}}^{H}\big)\Big)\\ &= \log_2\Big(\det\big(\mathbf{I} + \frac{P}{\sigma^2\|\mathbf{H}_{\text{s}}^{H}(\mathbf{H}_{\text{s}}\,\mathbf{H}_{\text{s}}^{H})^{-1}\|_F^2}\mathbf{I}\big)\Big) = K\log_2\Big(1 + \frac{P}{\sigma^2\sum_{k=1}^{K}\frac{1}{\lambda_{\text{sk}}^2}}\Big).\end{aligned} \tag{5.4}$$

To improve the tractability of the antenna selection problem, we consider an upper bound of sum-rate expression. Considering the following property

$$\left(\frac{1}{K}\sum_{k=1}^{K}\frac{1}{\lambda_{\text{sk}}^2}\right)^{-1} \leqslant \frac{1}{K}\left(\sum_{k=1}^{K}\lambda_{\text{sk}}^2\right), \tag{5.5}$$

we have

$$K\log_2\Big(1 + \frac{P}{\sigma^2\sum_{k=1}^{K}\frac{1}{\lambda_{\text{sk}}^2}}\Big) \leqslant\; K\log_2\Big(1 + \frac{P(\sum_{k=1}^{K}\lambda_{\text{sk}}^2)}{\sigma^2K^2}\Big) = K\log_2\Big(1 + \frac{P\text{Tr}(\mathbf{H}_{\text{s}}^{H}\mathbf{H}_{\text{s}})}{\sigma^2K^2}\Big). \tag{5.6}$$

where we used the fact that $\sum_{k=1}^{K}\lambda_{\text{sk}}^2 = \text{Tr}(\mathbf{H}_{\text{s}}^{H}\mathbf{H}_{\text{s}})$. Based on (5.6), we perform antenna selection using the sum-rate upper bound, for any selected channel realization $\mathbf{H}_{\text{s}}$, defined as

$$R_{\text{u}}(\mathbf{H}_{\text{s}}) = \; K\log_2\Big(1 + \frac{P\text{Tr}(\mathbf{H}_{\text{s}}^{H}\mathbf{H}_{\text{s}})}{\sigma^2K^2}\Big). \tag{5.7}$$

Note that we can write

$$R_{\mathrm{u}}(\mathbf{H}_{\mathrm{s},t}) = K \log_2 \Big(1 + \frac{P}{\sigma^2 K^2} \sum_{i \in \mathcal{I}} \|\mathbf{h}_{i,t}\|^2\Big). \tag{5.8}$$

where $\mathcal{I}$ is the set of the indices of the selected antennas, that is $\mathbf{H}_{\mathrm{s},t} = \Big[\mathbf{h}_{i,t}\Big]_{i \in \mathcal{I}}$. Specifically, we aim to design an antenna selection policy to maximize the expected long-term sum-rate upper bound, given by

$$E_{\{\mathbf{H}_{\mathrm{s},t}\}}\Big\{\sum_{t=0}^{\infty} R_{\mathrm{u}}(\mathbf{H}_{\mathrm{s},t})\Big\}, \tag{5.9}$$

where the expectation is taken with respect to random selected channel matrices $\{\mathbf{H}_{\mathrm{s},t}\}_{t=0}^{\infty}$. In the next section, using the long-term sum-rate upper bound in (5.9) as the reward function, we formulate the antenna selection problem using a POMDP framework.

## 5.3 POMDP Formulation

In this section, we formulate our antenna selection problem for a massive MU-MIMO system by the presented tuples of a POMDP framework in Section. 3.1, as follows:

**State space**

In our antenna selection problem, at time $t$, the $M \times 1$ state vector $\mathbf{s}_t$ is defined as

$$\mathbf{s}_t = [\|\mathbf{h}_{1,t}\|^2 \ \ \|\mathbf{h}_{2,t}\|^2 \ \ \ldots \ \ \|\mathbf{h}_{M,t}\|^2]^T. \tag{5.10}$$

That is, the $i$th entry of the state vector at time $t$ is the square of the norm of the vector of the channel coefficients between the users and the $i$-th antenna. We model the time-varying state vector as a discrete Markov process. Each element of $\mathbf{s}_t$ is assumed to take one of $Q$ possible levels of $\{\alpha_q\}_{q=1}^{Q}$, that is ,$\|\mathbf{h}_{i,t}\|^2 \in \{\alpha_1, ..., \alpha_Q\}$. As such, $\mathbf{s}_t$ belongs to a state space, denoted as $\mathcal{S}$, which is the finite set of all $Q^M$ possible values of $\mathbf{s}_t$. For $j = 1, 2, \ldots, |\mathcal{S}|$, the $j$-th member of $\mathcal{S}$ is labeled as $\mathbf{s}_j$.

### Action space

An action in our system model is selecting $N$ out of $M$ antennas. Hence, $L \triangleq \binom{M}{N}$ possible actions can be chosen. We define the action space as

$$\mathcal{A} \triangleq \{\tilde{\mathbf{a}}_1, \tilde{\mathbf{a}}_2, \cdots, \tilde{\mathbf{a}}_L\} \tag{5.11}$$

where, for $l = 1, ..., L$, $\tilde{\mathbf{a}}_l$ is a $M \times 1$ vector whose $N$ out of $M$ elements are equal to one and remaining elements are zero, that is

$$\tilde{\mathbf{a}}_l \triangleq [a_{1l} \ \ a_{2l} \ \ \cdots \ \ a_{Ml}]^T, \ \ a_{jl} \in \{0, 1\}, \ \sum_{j=1}^{M} a_{jl} = N. \tag{5.12}$$

At time $t$, our action $\mathbf{a}_t$ is to choose one of the $M \times 1$ selection vectors $\left\{\tilde{\mathbf{a}}_l\right\}_{l=1}^{L}$ from the action space.

### Transition probabilities

The transition probability matrix definition is provided in Section. 3.1. Here, we assume that the transition probability matrix $\mathbf{T}$ is known.

### Observation space

In our system model, the $j$-th element of the observation space, denoted as $\mathbf{o}_j$, can take one of the possible $M \times 1$ observation vectors, which have $N$ non-zero elements corresponding to the observed channel gain of the $N$ selected antennas and the remaining $M - N$ elements are zero. That is, we define the observation space as

$$\mathcal{O} \triangleq \{\mathbf{o}_1, \mathbf{o}_2, \cdots, \mathbf{o}_{L'}\}, \tag{5.13}$$

where $L' = \binom{M}{M-N} Q^N$. With given action $\mathbf{a}_t$, the random observation vector at time $t$, denoted as $\mathbf{o}_t \in \mathcal{O}$, is based on the current state as

$$\mathbf{o}_t \triangleq \mathrm{diag}(\mathbf{a}_t)\mathbf{s}_t. \tag{5.14}$$

### Observation probabilities

The conditional observation probability matrix, denoted by $\mathbf{O}(\mathbf{o}, \mathbf{a})$, is a $Q^M \times Q^M$ diagonal matrix and the definition is provided in Section. 3.1.

### Reward

At time $t$, when action $\mathbf{a}_t$ is taken at state $\mathbf{s}_t$, reward $R(\mathbf{s}_t, \mathbf{a}_t)$ is accrued. The immediate reward function $R(\mathbf{s}_t, \mathbf{a}_t)$ is defined as the sum-rate upper-bound given in (5.8), that is

$$R(\mathbf{s}_t, \mathbf{a}_t) \triangleq R_{\mathrm{u}}(\mathbf{H}_{\mathrm{s},t}) = K \log_2 \Big(1 + \frac{P}{\sigma^2 K^2} \sum_{i \in \mathcal{I}} \|\mathbf{h}_{i,t}\|^2\Big) = K \log_2 \Big(1 + \frac{P}{\sigma^2 K^2} \mathbf{1}^T \mathbf{o}_t\Big) \tag{5.15}$$

where in the third equality, we use the fact that $\sum_{i \in \mathcal{I}} \|\mathbf{h}_{i,t}\|^2 = \mathbf{1}^T \mathbf{o}_t$.

### Belief vector

At time $t$, the belief vector is defined as $\mathbf{b}_t \triangleq [b_{1,t} \quad b_{2,t} \quad \ldots \quad b_{|\mathcal{S}|,t}]^T$, where $b_{j,t}$ is the probability of the state at time $t$, $\mathbf{s}_t$ being equal to $\mathrm{s}_j \in \mathcal{S}$, given all the action and observation history until time $t-1$. If we use $\mathcal{H}_{t-1}$ to represent the action and observation history until time $t-1$, i.e.,

$$\mathcal{H}_{t-1} \triangleq \{\mathbf{o}_{t-1}, \mathbf{a}_{t-1}, \mathcal{H}_{t-2}\}, \tag{5.16}$$

then we can write

$$b_{j,t} \triangleq \Pr\{\mathbf{s}_t = \mathrm{s}_j | \mathcal{H}_{t-1}\}. \tag{5.17}$$

belief vector at time $t$ depends on the observation $\mathbf{o}_{t-1} \in \mathcal{O}$, which is a random vector. Hence, *the random belief vector* $\mathbf{b}_t$ can be defined as

$$\mathbf{b}_t \triangleq \mathbf{g}(\mathbf{o}_{t-1}\ , \mathbf{a}_{t-1}, \mathbf{b}_{t-1}). \tag{5.18}$$

The $j$-th entry of $\mathbf{b}_t$ is defined as

$$\mathsf{b}_{j,t} \triangleq \Pr(\mathbf{s}_t = \mathrm{s}_j | \boldsymbol{\mathcal{H}}_{t-1}), \tag{5.19}$$

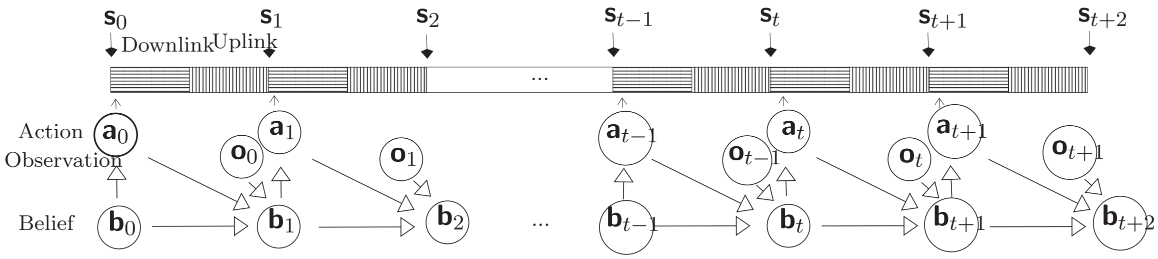


Figure 5.2: Illustration of the antenna selection problem as a POMDP process.

where the *random history* set $\boldsymbol{\mathcal{H}}_{t-1}$ is defined as

$$\boldsymbol{\mathcal{H}}_{t-1} \triangleq \{\mathbf{o}_{t-1}, \mathbf{a}_{t-1}, \boldsymbol{\mathcal{H}}_{t-2}\}. \tag{5.20}$$

Note that $\mathcal{H}_{t-1}$ in (5.16) is a realization of $\boldsymbol{\mathcal{H}}_{t-1}$ after observing $\mathbf{o}_{t-1}$ (see Section. 3.1 for more explanations about belief vector).

The dynamic of a real-time POMDP controller proceeds as shown in Fig. 5.2. Each time slot includes one downlink and one uplink transmission. We assume that channel evolves at the beginning of each time slot and remains unchanged during the entire time slot. At the beginning of the downlink transmission, the BS chooses an action vector which selects $N$ out of $M$ antennas to transmit data. Then via uplink training process, at the end of each time slot, the BS observes the channel coefficients between the $N$ selected antennas and all $K$ available users. At the next time slot $t$, given that the state at time $t$ is $\mathbf{s}_t$, the BS uses all the historical information available until the end of time slot $t-1$, which includes $\mathbf{b}_{t-1}$, and $\mathbf{a}_{t-1}$, obtained at the beginning of time slot $t-1$, and $\mathbf{o}_{t-1}$, obtained at the end of time slot $t-1$, to obtain (update) the belief vector $\mathbf{b}_t$, and then chooses the action vector $\mathbf{a}_t$ at time $t$. Note that $\mathbf{a}_t$ depends on $\{\mathbf{s}_\tau\}_{\tau=0}^{t-1}$, and hence, is random. Our aim is to design a decision policy for actions $\{\mathbf{a}_t\}_{t=0}^{+\infty}$ such that the expected cumulative reward, $E\Big\{\sum_{t=0}^{+\infty} R(\mathbf{s}_t, \mathbf{a}_t)\Big\}$ is maximized. Here, the expectation $E\{\cdot\}$ is taken w.r.t. random states $\{\mathbf{s}_t\}_{t=0}^{+\infty}$.

### 5.3.1 Objective Function

Policy [1] at time $t$, maps the belief vector $\mathbf{b}_t$ to the action $\mathbf{a}_t$, that is $\mathbf{a}_t = \pi(\mathbf{b}_t)$. For the initial belief vector $\mathbf{b}_0$, the objective function denoted as $J_\pi(\mathbf{b}_0)$ for an infinite horizon POMDP framework is defined as

$$J_\pi(\mathbf{b}_0) = E_{\{\mathbf{s}_t\}}\Big\{\sum_{t=0}^{\infty} R(\mathbf{s}_t, \mathbf{a}_t)\Big|\mathbf{b}_0\Big\} = E_{\{\mathbf{s}_t\}}\Big\{\sum_{t=0}^{\infty} R(\mathbf{s}_t, \pi(\mathbf{b}_t))\Big|\mathbf{b}_0\Big\} \tag{5.21}$$

where $\mathbf{s}_t \in \mathcal{S}$, $\mathbf{a}_t = \pi(\mathbf{b}_t) \in \mathcal{A}$, and $E_{\{\mathbf{s}_t\}}\{\cdot\}$ is the expectation w.r.t. random states $\{\mathbf{s}_t\}_{t=0}^{+\infty}$, given the initial belief $\mathbf{b}_0$. Given the defined POMDP model, the main goal is to find the optimal policy as

$$\pi^* = \arg\max_{\pi} J_\pi(\mathbf{b}_0), \text{ for any } \mathbf{b}_0. \tag{5.22}$$

As the random action vector $\mathbf{a}_t$ is a function of $\mathbf{b}_t$, which in turn is a function of $\mathbf{o}_{t-1}$, there is a one-to-one correspondence between $\boldsymbol{\mathcal{H}}_t$ and $\{\mathbf{o}_{t'}\}_{t'=0}^{t}$. Hence, as we explained in Section. 3.3, we can write

$$J_\pi(\mathbf{b}_0) = E_{\{\boldsymbol{\mathcal{H}}_t\}}\Big\{\sum_{t=0}^{+\infty} \mathbf{r}^T(\mathbf{a}_t)\mathbf{b}_t\Big|\mathbf{b}_0\Big\} \tag{5.23}$$

where $\{\boldsymbol{\mathcal{H}}_t\}$ is the whole history, and $\mathbf{r}(\mathbf{a}) \triangleq [R(\mathbf{s}_1, \mathbf{a}) \;\; R(\mathbf{s}_2, \mathbf{a}) \;\; \cdots \;\; R(\mathbf{s}_{Q^M}, \mathbf{a})]^T$ is the reward vector for action $\mathbf{a}$. Note that, given $\boldsymbol{\mathcal{H}}_{t-1} = \mathcal{H}_{t-1}$, we note that $\mathbf{r}^T(\mathbf{a}_t)\mathbf{b}_t$ as the expected immediate reward function.

Since a POMDP is a continuous belief state MDP (see [55]), we can straightforwardly write down the Bellman equation for the infinite-horizon continuous-state MDP with the dynamics of the belief update in (5.18) and the objective function in (5.23) to find the optimal policy. We can use the value iteration algorithm to find the optimal solution. However, the computational complexity of the value iteration algorithm is $\mathcal{O}(|\mathcal{S}|^2 \times |\mathcal{A}|)$ per iteration [69]. In our problem, $|\mathcal{S}| = Q^M$ and $|\mathcal{A}| = \binom{M}{N}$. Thus, the computational complexity of the value iteration algorithm

[1] Here, the policy is a stationary policy (see Section. 3.3).

grows exponentially with increasing the number of antennas. Thus, the value iteration algorithm is computationally intractable for the antenna selection problem in massive MIMO systems. In order to tackle the aforementioned issue, affordable suboptimal solutions, such as myopic policy, are more desirable. In the next section, for the case of two-state channels, we consider the myopic policy to solve our POMDP-based antenna selection problem and prove that this policy is optimal in this special case, for our problem in (5.22).

## 5.4 Two-State Channels: The Optimality of Myopic Policy

In this section, we assume that each element of $\mathbf{s}_t$ takes one of $Q = 2$ possible levels of good and bad, denoted as $\alpha$ and $\beta$, respectively, where $\alpha > \beta$, that is

$$\|\mathbf{h}_{i,t}\|^2 \in \{\alpha, \beta\}, \quad \text{for } i = 1, \dots, M. \tag{5.24}$$

Here, $\|\mathbf{h}_{i,t}\|^2 = \alpha$ $(\beta)$ means the state of the $i$-th antenna is good (bad). Also, we assume that the antennas channel gain evolves according to the same two-state Markov chain at each time slot, as shown in Fig. 4.2. We use $p_{01}$ $(p_{10})$ to denote the probability of changing the state from bad (good) state to good (bad) state. Also, $p_{00} = 1 - p_{01}$ and $p_{11} = 1 - p_{10}$ are the probabilities of $\|\mathbf{h}_{i,t}\|^2$ remaining in the bad and good states in the next time slot, respectively. Here we consider positively correlated model, i.e., $p_{11} \geq p_{01}$, meaning that the probability of $\|\mathbf{h}_{i,t}\|^2$ remaining in good state is higher than the probability of changing the state from bad to good state.

Considering that $s_{ij}$ is the $i$-th element of the $j$-th possible state vector, for $i = 1, \dots, M$ and $j = 1, \dots, |\mathcal{S}|$, we can define the indicator function $I_{[s_{ij}=\alpha]}$ as

$$I_{[s_{ij}=\alpha]} = \begin{cases} 1 & \text{if } s_{ij} = \alpha \\ 0 & \text{if } s_{ij} = \beta \end{cases}, \tag{5.25}$$

to simplify the presentation of the state space. Without loss of generality, we redefine the state space as $\mathcal{C} = \{\mathbf{c}_j\}_{j=1}^{2^M}$, where $\mathbf{c}_j \triangleq [c_{1j} \;\; c_{2j} \;\; \cdots \;\; c_{Mj}]^T$ is the $j$-th member of $\mathcal{C}$, and $c_{ij}$ is given by

$$c_{ij} = I_{[s_{ij}=\alpha]}. \tag{5.26}$$

We also redefine the random vector state at time $t$ as $\boldsymbol{\mathsf{c}}_t \in \mathcal{C}$, as we can write $\boldsymbol{\mathsf{c}}_t \triangleq [\mathsf{c}_{1,t} \;\; \mathsf{c}_{2,t} \;\; \cdots \;\; \mathsf{c}_{M,t}]^T \in \mathcal{C}$, where $\mathsf{c}_{i,t}$ is the random variable state of the $i$-th antenna. Note that the state $\boldsymbol{\mathsf{s}}_t$ of the channel gain vector can be determined from $\boldsymbol{\mathsf{c}}_t$. That is for any $\mathbf{s}_j \in \mathcal{S}$, there is a unique $\mathbf{c}_j \in \mathcal{C}$ and vice versa. The two-state per antenna model allows us to simplify the belief formulation [72], as explained in the sequel. First, for $i = 1, 2, \ldots, M$, we define $\omega_{i,t} \triangleq \Pr(\mathsf{c}_{i,t} = 1 | \mathcal{H}_{t-1})$, which is the conditional probability of $\|\boldsymbol{\mathsf{h}}_{i,t}\|^2$ being in good state given the history of all past actions and observations up to time slot $t-1$ . We also redefine the belief vector at time $t$ as $\boldsymbol{\omega}_t \triangleq [\omega_{1,t} \;\; \omega_{2,t} \;\; \ldots \;\; \omega_{M,t}]^T$. At time slot $t$, based on the antenna selection vector $\mathbf{a}_t$, we can update the $i$-th element of vector $\boldsymbol{\omega}_{t+1}$ as

$$\omega_{i,t+1} = \begin{cases} p_{11} & \text{if } a_{i,t} = 1, \;\; \mathsf{c}_{i,t} = 1; \\ p_{01} & \text{if } a_{i,t} = 1, \;\; \mathsf{c}_{i,t} = 0; \\ \omega_{i,t} p_{11} + (1 - \omega_{i,t}) p_{01} & \text{if } a_{i,t} = 0. \end{cases} , \qquad \text{for } i = 1, \cdots, M. \tag{5.27}$$

We can reexpress the $j$-th entry of the belief vector $\mathbf{b}_t$ as

$$b_{j,t} = \Pr(\boldsymbol{\mathsf{s}}_t = \mathbf{s}_j | \mathcal{H}_{t-1}) = \Pr(\boldsymbol{\mathsf{c}}_t = \mathbf{c}_j | \mathcal{H}_{t-1}). \tag{5.28}$$

Assuming that the channel gains across different antennas are statistically independent, we can write

$$\Pr(\boldsymbol{\mathsf{c}}_t = \mathbf{c}_j | \mathcal{H}_{t-1}) \triangleq \prod_{i=1}^{M} \Pr(\mathsf{c}_{i,t} = c_{ij} | \mathcal{H}_{t-1}) = \prod_{i=1}^{M} \hat{f}(\omega_{i,t}, c_{ij}), \tag{5.29}$$

where we define $\hat{f}(\omega, c) = \omega^c (1-\omega)^{1-c}$, and use the fact that $\Pr(\mathsf{c}_{i,t} = 1 | \mathcal{H}_{t-1}) = \omega_{i,t}$ and $\Pr(\mathsf{c}_{i,t} = 0 | \mathcal{H}_{t-1}) = 1 - \omega_{i,t}$. Based on (5.28) and (5.29), the expected immediate reward function at time $t$, can be written as

$$\begin{aligned}\bar{R}(\mathbf{a}_t, \boldsymbol{\omega}_t) \triangleq \mathbf{r}^T(\mathbf{a}_t)\mathbf{b}_t &= \sum_{j=1}^{|\mathcal{S}|} R(\mathbf{s}_j, \mathbf{a}_t) b_{j,t} = \sum_{j=1}^{|\mathcal{C}|} R(\mathbf{s}_j, \mathbf{a}_t) \Pr(\mathbf{c}_t = \mathbf{c}_j | \mathcal{H}_{t-1}) \\ &= \sum_{j=1}^{|\mathcal{C}|} \underbrace{K \log_2 \Big(1 + \frac{P}{\sigma^2 K^2} \mathbf{1}^T \mathbf{o}_{j,t}\Big)}_{\triangleq R_Z} \prod_{i=1}^{M} \hat{f}(\omega_{i,t}, c_{ij}),\end{aligned} \tag{5.30}$$

where $\mathbf{o}_{j,t}$ is the observation vector at time $t$ if $\mathbf{s}_t = \mathbf{s}_j$ and can be written, using (5.14), as

$$\mathbf{o}_{j,t} = \mathrm{diag}(\mathbf{a}_t)\mathbf{s}_j. \tag{5.31}$$

Here, each entry of $\mathbf{o}_{j,t}$ belongs to the set $\{\alpha, \beta\}$. Note that, if $Z$ entries of $\mathbf{o}_{j,t}$ are equal to $\alpha$ and the remaining $N - Z$ entries[2] of $\mathbf{o}_{j,t}$ are equal to $\beta$, then term defined in (5.30) as $R_Z$ is equal to

$$R_Z = K \log_2 \Big(\ Z(1 + \frac{P\alpha}{\sigma^2 K^2}) + (N - Z)(1 + \frac{P\beta}{\sigma^2 K^2})\Big). \tag{5.32}$$

With $R_Z$ given above, we can follow the steps presented in our earlier work in Section. 4.3 for the single-user scenario to simplify (5.30) [31]. For the sake of completeness, we provide the steps of this simplification in the sequel. Given action $\mathbf{a}_t$, the state space $\mathcal{C}$ can be partitioned as

$$\mathcal{C} = \bigcup_{Z=0}^{N} \mathcal{C}_Z(\mathbf{a}_t), \tag{5.33}$$

where $\mathcal{C}_Z(\mathbf{a}_t) = \{\mathbf{c} = [c_1 \;\; c_2 \;\; \cdots \;\; c_M]^T \in \mathcal{C} \big| \; \|\mathrm{diag}(\mathbf{a}_t)\mathbf{c}\|^2 = Z\}$. We now use (5.33) to rewrite (5.30) as

$$\begin{aligned}\bar{R}(\mathbf{a}_t, \boldsymbol{\omega}_t) &= \sum_{Z=0}^{N} \sum_{\mathbf{c} \in \mathcal{C}_Z(\mathbf{a}_t)} R_Z \prod_{i=1}^{M} \hat{f}(\omega_{i,t}, c_i) \\ &= \sum_{Z=0}^{N} R_Z \sum_{\mathbf{c} \in \mathcal{C}_Z(\mathbf{a}_t)} \prod_{i=1}^{M} \hat{f}(\omega_{i,t}, c_i)\end{aligned}$$

[2] Non-zero elements

$$= \sum_{Z=0}^{N} R_Z \sum_{\mathbf{c}\in\mathcal{C}_Z(\mathbf{a}_t)} \prod_{i\in\mathcal{I}(\mathbf{a}_t)} \hat{f}(\omega_{i,t}, c_i) \prod_{i\in\mathcal{I}(\mathbf{a}_t)^{\perp}} \hat{f}(\omega_{i,t}, c_i), \tag{5.34}$$

where $\mathcal{I}(\mathbf{a}_t)$ is set of the indices of the selected antennas, while $\mathcal{I}^{\perp}(\mathbf{a}_t)$ is the set of the indices of the unselected antennas. As $\mathcal{I}^{\perp}(\mathbf{a}_t)$ and $\mathcal{I}^{\perp}(\mathbf{a}_t)$ are complement to each other, we can write $|\mathcal{I}(\mathbf{a}_t)| = N$ and $|\mathcal{I}^{\perp}(\mathbf{a}_t)| = M - N$. Note that $\mathbf{c} \in \mathcal{C}_Z(\mathbf{a}_t)$ can be split into two sub-vectors $\mathbf{c}' = [c_i]_{i\in\mathcal{I}(\mathbf{a}_t)}$ and $\mathbf{c}'' = [c_i]_{i\in\mathcal{I}^{\perp}(\mathbf{a}_t)}$, where the entries of $\mathbf{c}''$ can be either 0 or 1, that is $\mathbf{c}'' \in \{0,1\}^{M-N}$, while $\mathbf{c}' \in \mathcal{C}'_Z \triangleq \{\mathbf{c}' : \mathbf{1}_N^T\mathbf{c}' = Z\}$. Therefore, we can rewrite (5.34) as

$$\begin{aligned}
\bar{R}(\mathbf{a}_t, \boldsymbol{\omega}_t) &= \sum_{Z=0}^{N} R_Z \sum_{\mathbf{c}'\in\mathcal{C}'_Z} \sum_{\mathbf{c}''\in\{0,1\}^{M-N}} \prod_{i\in\mathcal{I}(\mathbf{a}_t)} \hat{f}(\omega_{i,t}, c'_i) \prod_{i\in\mathcal{I}^{\perp}(\mathbf{a}_t)} \hat{f}(\omega_{i,t}, c''_i) \\
&= \sum_{Z=0}^{N} R_Z \left( \sum_{\mathbf{c}'\in\mathcal{C}'_Z} \prod_{i\in\mathcal{I}(\mathbf{a}_t)} \hat{f}(\omega_{i,t}, c'_i) \right) \underbrace{\left( \sum_{\mathbf{c}''\in\{0,1\}^{M-N}} \prod_{i\in\mathcal{I}^{\perp}(\mathbf{a}_t)} \hat{f}(\omega_{i,t}, c''_i) \right)}_{=1} \\
&= \sum_{Z=0}^{N} R_Z \sum_{\mathbf{c}'\in\mathcal{C}'_Z} \prod_{i\in\mathcal{I}(\mathbf{a}_t)} \hat{f}(\omega_{i,t}, c'_i),
\end{aligned} \tag{5.35}$$

where the summation above the bracket is the sum of the probabilities of all possible values $\mathbf{c}''$ may take and thus is equal to 1. In Section. 4.3, we rigorously prove that under the condition of the positively correlated two-state models, the myopic policy, which maximizes the expected immediate reward in (5.35), is optimal for the antenna selection problem in (5.22). The following theorem can then be used to further simplify the myopic policy.

**Theorem 4.** *For the antenna selection problem defined in* (5.22)*, under the condition of positively correlated two-state channel model, at time $t$, the myopic policy is optimal and amounts to selecting those $N$ out of $M$ antennas, with the largest probability of their channel gains being in good state. In other words, the indices of the $N$ largest entries of $\boldsymbol{\omega}_t$ are the indices of the antennas which have to be selected.*

*Proof.* In Section. 4.3, in lemma 1, we show that the function $f(\mathbf{x})$ is monotonically increasing in $x_j$. Meaning that we can maximize the expected immediate reward $f(\mathbf{x})$ in (5.35) with selecting the $N$ largest entries in $\mathbf{x}$. Thus, it is straightforward to conclude that, for any $N < M$, the myopic policy is selecting $N$ antennas with higher probabilities of being in good state (i.e., selecting $N$ antennas corresponding to the highest $\omega_i$'s value). The proof is now complete. ■

The myopic policy for two-state channel model in Theorem 4 can be applied to the realistic fading channel models as a low-complexity method for the antenna selection problem in massive MIMO systems.. In the next section, we propose to use the myopic policy for the antenna selection problem over Rayleigh fading channels. Given the optimality of the myopic policy for the two-state channel model, we take advantage of this myopic policy by quantizing the channel gain of each antenna into two levels only for the purpose of the antenna selection.

## 5.5 Gauss-Markov Model for Rayleigh Fading Channels

Let us denote the diagonal channel covariance matrix as

$$\boldsymbol{\Sigma}_h = E\{\mathbf{h}_{i,t}\mathbf{h}_{i,t}^H\} = \operatorname{diag}([\sigma_{\mathrm{h},k}^2]_{k=1}^K),$$

where $\sigma_{\mathrm{h},k}^2$ is the large-scale variation of channel between the $k$-th user and the $i$-th antenna. We assume that the channel vector $\mathbf{h}_{i,t} \sim \mathcal{CN}(\mathbf{0}_{K\times 1}, \boldsymbol{\Sigma}_h)$ evolves according to the first-order Gauss-Markov channel model given by

$$\mathbf{h}_{i,t} \triangleq \operatorname{diag}(\boldsymbol{\xi})\mathbf{h}_{i,t-1} + \operatorname{diag}(\boldsymbol{\xi}')\mathbf{z}_{i,t}, \qquad i = 1, ..., M\,. \tag{5.36}$$

In the above model, the i.i.d. innovation sequence $\mathbf{z}_{i,t} \sim \mathcal{CN}(\mathbf{0}_{K\times 1}, \boldsymbol{\Sigma}_h)$ is independent of the channel vector $\mathbf{h}_{i,t}$, for $i = 1, ..., M$. Also, the fading correlation vector is defined $\boldsymbol{\xi} = [\xi_1 \quad \xi_2 \quad \cdots \quad \xi_k]^T$, where $\xi_k \in [0, 1]$ is the fading correlation coefficient corresponding to the $k$-th user. Furthermore, we define $\boldsymbol{\xi}'$ as

$\boldsymbol{\xi}' = [\sqrt{1-\xi_1^2} \ \sqrt{1-\xi_2^2} \ \cdots \ \sqrt{1-\xi_K^2}]^T$. Note that the value of $\xi_k$ depends on the maximum Doppler frequency of the $k$-th user [61], where $\xi_k = 1$ for a static channel model, and $\xi_k = 0$ for a channel evolves independently over time $t$.

For the Gauss-Markov channel model in (5.36), it is shown in [36] that $\|\mathbf{h}_{i,t}\|^2$ asymptotically forms as a Markov process as $K$ becomes large. Thus, we consider that $\|\mathbf{h}_{i,t}\|^2$ evolves based on a Markov process. To benefit from the optimality of myopic policy for two-state channel model, we propose to quantize $\|\mathbf{h}_{i,t}\|^2$ into two values of $\alpha$ and $\beta$ only at the antenna selection stage, such that

$$\mathsf{s}_{i,t} = \begin{cases} \alpha, & \text{if } \|\mathbf{h}_{i,t}\|^2 \geqslant v, \\ \beta, & \text{if } \|\mathbf{h}_{i,t}\|^2 < v \ , \end{cases} \tag{5.37}$$

where $v$ is the quantization threshold, and $\alpha > \beta$ must hold. Then, we apply the myopic policy presented in Theorem 4 to the quantized two-state channel model in the selection stage according to Algorithm 3. At time slot $t$, based on action $\mathbf{a}_t$, $N$ antennas are selected for transmitting data with the users. At the end of uplink transmission, the channel coefficients corresponding to the links between the selected antennas and all $K$ users, $\mathbf{o}_t$ are obtained using (5.14). Quantization is then performed on each element of $\mathbf{o}_t$ to obtain those entries of $\mathbf{c}_t$ that correspond to the selected antennas at time $t$. Next, the elements of the belief vector, i.e., $\omega_{i,t+1}$'s are updated using (5.27). At the beginning of time slot $t+1$, the antenna selection vector is updated as $a_{i,t+1} = 1$ (i.e., the $i$-th antenna is selected), if $\omega_{i,t+1}$ is among the $N$ largest entries of $\boldsymbol{\omega}_{t+1}$, or $a_{i,t+1} = 0$ otherwise. The antennas selected by the action vector $\mathbf{a}_{t+1}$ are used for data transmission with the users at time slot $t+1$.

The computational complexity of our proposed myopic antenna selection algorithm is given as follows. Computing the elements of $\mathbf{o}_t$ as in (5.38) has a computational complexity $\mathcal{O}(NK)$. Updating the elements of the belief vector has a computational complexity $\mathcal{O}(M-N)$ (see (4.12)), and then finding the $N$ largest elements of the $M \times 1$ belief vector has a computational complexity $\mathcal{O}(M \log N)$ [70]. Since we assume that the BS is equipped with a large number of antennas (i.e., $M$

is large), the total computational complexity of the myopic policy-based antenna selection algorithm is $\mathcal{O}(M \log N)$, which is significantly lower than the computational complexity of the value iteration algorithm with the computational complexity $\mathcal{O}(|\mathcal{S}|^2 \times |\mathcal{A}|)$ per iteration [69] (in our problem, $|\mathcal{S}| = 2^M$ and $|\mathcal{A}| = \binom{M}{N}$).

**Algorithm 3** The myopic-policy-based antenna selection for multi-user systems

```
Initialization: Given the channel correlation factor $\boldsymbol{\xi}$ and large-scale channel variations of all users $\{\sigma_{h,k}^2\}_{k=1}^{K}$. set the threshold value $v$.
At each time slot $t$:
Input: $[\mathbf{h}_{i,t}]_{i \in \mathcal{I}(\mathbf{a}_t)}$.
1: Obtain the elements of the observation vector $\mathbf{o}_t = [\mathsf{o}_{i,t}]_{i=1}^{M}$ where
       $\mathsf{o}_{i,t} = \begin{cases} \|\mathbf{h}_{i,t}\|^2, & \text{if } i \in \mathcal{I}(\mathbf{a}_t), \\ 0, & \text{otherwise.} \end{cases} \quad (5.38)$
2: Quantize the elements of $\mathbf{o}_t$ into $\alpha$ and $\beta$ using (5.37) and update the entries of $\mathbf{c}_t$ which correspond to the selected antennas at time $t$ based on (5.25).
3: Update the belief vector $\boldsymbol{\omega}_{t+1}$ using (5.27).
4: For $i = 1, 2, \cdots, M$, choose the $i$-th entry of $\mathbf{a}_{t+1}$ as
       $a_{i,t+1} = \begin{cases} 1, & \text{if } \omega_{i,t+1} \text{ among the first } N \text{ highest entries of } \boldsymbol{\omega}_{t+1}, \\ 0, & \text{otherwise.} \end{cases} \quad (5.39)$
Output: $\mathbf{a}_{t+1}$
```

In our proposed antenna selection method in Algorithm 3, we assume that the optimal threshold value $v$ is known for given channel correlation vector $\boldsymbol{\xi}$ and large-scale channel variations of all users $\{\sigma_{h,k}^2\}_{k=1}^{K}$. Given the value of $v$, the transition probabilities $p_{11}$ and $p_{01}$ can be obtained empirically. Note that, the optimal threshold value $v$ can not be expressed in a closed form [62, 63]. We will show the effect of the optimal threshold value on the performance in our numerical examples. In the next section, we propose an efficient offline method to obtain a lookup table for the optimal value of the threshold $v$ for given $M$, $N$, and $K$ and for any value of the channel correlation vector $\boldsymbol{\xi}$ and $\{\sigma_{h,k}^2\}_{k=1}^{K}$.

## 5.6 Optimal Threshold Value for Channel Quantization

According to Algorithm 3, in the selection stage, $\|\mathbf{h}_{i,t}\|^2$ is quantized into two levels in order to update the entries of $\mathbf{c}_t$. Thus, the performance of the myopic-policy-based antenna selection algorithm depends on the threshold value $v$ used for quantization. In this section we aim to find the optimal threshold value, denoted as $v^*$, which results in the best performance of our proposed myopic policy algorithm. Note that $v^*$ is different for different values of $M$, $N$, $K$ and the large-scale channel variations of all users $\{\sigma^2_{h,k}\}^K_{k=1}$. To find $v^*$ for any given $M$, $N$, $K$ (where $M \gg N > K$), and the given large-scale channel variations of all users $\{\sigma^2_{h,k}\}^K_{k=1}$, we use the MU-MIMO sum-rate in (5.4) to define the time-averaged sum-rate, denoted as $\hat{R}_t$, as

$$\hat{R}_t = \frac{1}{t}\sum_{\tau=0}^{t} \log_2 \Big( \det \big(\mathbf{I} + \frac{P}{\sigma^2 \|\mathbf{H}^H_{\mathrm{s},\tau}(\mathbf{H}_{\mathrm{s},\tau}\ \mathbf{H}^H_{\mathrm{s},\tau})^{-1}\|^2_F}\mathbf{I}\big) \Big). \tag{5.40}$$

We use a search algorithm to find the optimal threshold value $v^*$ as

$$v^* = \arg \max_{0<v\leq L} \hat{R}_T. \tag{5.41}$$

where $L$ is the upper limit of the threshold value and its value is chosen based on the statistics of channel coefficients. More specifically, adopting 3-$\sigma$ rule, we choose $L = \mu' + 3\sigma'_h$, where $\mu'$ and $\sigma'_h$ are the mean and the standard deviation of $\|\mathbf{h}_{i,t}\|^2$. As the $k$-th element of $\mathbf{h}_{i,t}$ is a zero-mean Gaussian random variable with variance $\sigma^2_{h,k}$, the channel gain $\|\mathbf{h}_{i,t}\|^2$ is the sum of gamma random variables with the same shape parameter 0.5 but different the rate parameters $\{\sigma^2_{h,k}\}^K_{k=1}$. Using the derivation of mean and variance of the sum of non-identical gamma variables in [73], we can approximate $\mu'$ and ${\sigma'_h}^2$ as

$$\mu' \approx \frac{1}{2}\sum_{k=1}^{K} \sigma^2_{h,k}, \quad {\sigma'_h}^2 \approx \frac{1}{2}\sum_{k=1}^{K} \sigma^4_{h,k}. \tag{5.42}$$

To avoid the exhaustive search to find $v^*$ in (5.41), we propose a heuristic iterative search method for finding the optimal threshold value as in Algorithm 4. Based on

**Algorithm 4** Optimal threshold value

```
Initialization: Given $\mathbf{\Sigma}_h$ and $K$, obtain the value of $\mu'$ and $\sigma'_h$ using (5.42), and $L = \mu' + 3\sigma'_h$, and choose $v_{\mathrm{C}}^{(1)} = \frac{L}{2}$, $v_{\mathrm{R}}^{(1)} = v_{\mathrm{C}}^{(1)} + \frac{L}{2}$, and $v_{\mathrm{L}}^{(1)} = v_{\mathrm{C}}^{(1)} - \frac{L}{2}$. Set $T = 1000$, stopping threshold $\varepsilon$, and $i = 1$.
Input: $M$, $N$ and $K$
while $\frac{L}{2^{i+1}} > \varepsilon$ do
    Calculate the value of $\hat{R}_T$ for $v_{\mathrm{C}}^{(i)}$, $v_{\mathrm{R}}^{(i)}$, and $v_{\mathrm{L}}^{(i)}$ and find the point which results in the largest value for $\hat{R}_T$. This point is introduced as $v_{\mathrm{C}}^{(i+1)}$.
    Obtain $v_{\mathrm{L}}^{(i+1)} = v_{\mathrm{C}}^{(i+1)} - \frac{L}{2^{i+1}}$, and $v_{\mathrm{R}}^{(i)} = v_{\mathrm{C}}^{(i)} + \frac{L}{2^{i+1}}$.
    $i \leftarrow i + 1$.
end while
Output: $v^* = v_{\mathrm{C}}^{(i+1)}$
```

our proposed search algorithm, at each iteration, three different points are selected as the possible threshold values, namely left-side point, center point, and right-side point, which are denoted as $v_{\mathrm{L}}$, $v_{\mathrm{C}}$, and $v_{\mathrm{R}}$, respectively. Here, we use $i$ to denote the iteration index. In the first iteration ($i = 1$), we choose $v_{\mathrm{C}}^{(1)} = \frac{L}{2}$, $v_{\mathrm{R}}^{(1)} = v_{\mathrm{C}}^{(1)} + \frac{L}{2}$, and $v_{\mathrm{L}}^{(1)} = v_{\mathrm{C}}^{(1)} - \frac{L}{2}$. We then calculate the value of $\hat{R}_T$ for $v_{\mathrm{C}}^{(i)}$, $v_{\mathrm{R}}^{(i)}$, and $v_{\mathrm{L}}^{(i)}$ and find the point which results in the largest value for $\hat{R}_T$. This point is denoted as $v_{\mathrm{C}}^{(i+1)}$, and the following update is performed: $v_{\mathrm{L}}^{(i+1)} = v_{\mathrm{C}}^{(i+1)} - \frac{L}{2^{i+1}}$, and $v_{\mathrm{R}}^{(i+1)} = v_{\mathrm{C}}^{(i+1)} + \frac{L}{2^{i+1}}$. We iterate this procedure until $\frac{L}{2^{i+1}}$ is sufficiently small. Then, we choose $v^* = v_{\mathrm{C}}^{(i+1)}$.

In the next section, we provide the simulation result to evaluate the performance of our proposed myopic-policy-based antenna selection algorithm given that the optimal threshold value is used for the channel gain quantization in the selection stage.

## 5.7 Simulation Result

In this section, first we validate the optimality of myopic policy for positively correlated two-state channel model, and then study the performance of our proposed Algorithm 3, on the actual Rayleigh fading channel.

### 5.7.1 Two-State Channel Model

In this subsection, considering the two-state model in (5.24), we aim to verify the optimality of the myopic policy solution in Theorem 1. In this example, we assume $K = 2$, $\{\alpha, \beta\} = \{4 \times \sqrt{0.1}, 4 \times \sqrt{10}\}$, $p_{01} = 0.2$ and $p_{11} = 0.8$. Here, we use the existing POMDP solver software in [56] use the value iteration method to compute the optimal policy. In our experiment, we aim to show the optimality of the myopic policy when the sum-rate upper-bound in (5.7) is defined as the reward function. We define the time-averaged $\tilde{R}_t$, of the sum-rate upper-bound as

$$\tilde{R}_t = \frac{1}{t} \sum_{\tau=0}^{t} R(\mathbf{s}_\tau, \mathbf{a}_\tau). \tag{5.43}$$

Fig. 5.3 shows $\tilde{R}_{1000}$, in bits per channel use (bcu), versus $\frac{P}{\sigma^2}$, produced by both the myopic-policy-based and the value-iteration-based algorithms, for different values of $M$ and $N$. Note that our POMDP-based antenna selection technique converges in about 1000 time slots, we plot $\tilde{R}_{1000}$, and the results are averaged over 100 Monte Carlo runs (for same users setup). As it can be seen from Fig. 5.3, for same values of $M$, $N$,$K$, and $\frac{p}{\sigma^2}$, the time averaged sun-rate upper-bound values are same for both myopic-policy-based method and value iteration based solution. The results validate the optimality of the myopic policy for the antenna selection problem.

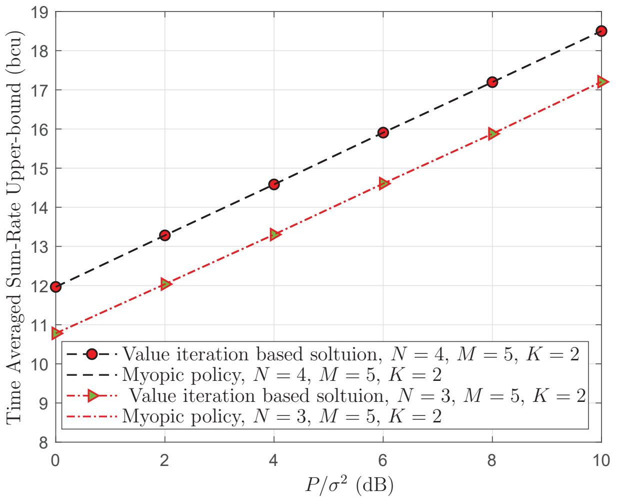


Figure 5.3: The time-averaged sum-rate upper-bound versus $\frac{P}{\sigma^2}$, for the two-state channel model.

### 5.7.2 Gauss-Markov Channel Model

In this subsection, considering the Gauss-Markov channel model in (5.36), we apply our proposed myopic policy (Algorithm 3) to the antenna selection problem in a massive MU-MIMO system. We need to quantize the channel gain $\|\mathbf{h}_{i,t}\|^2$ of the $i$-th antenna at time $t$, based on (5.37), and then use the myopic policy for the antenna selection problem. Note that this two-state quantization is performed only in the selection stage to find the best set of antennas. We use the non-quantized channel coefficients corresponding to the selected antennas for ZF beamforming. Here, we use the time-averaged *achieved* sum-rate $\hat{R}_t$ in (5.40) to evaluate the performance of Algorithm 3. Since the quantization in (5.37) depends on threshold $v$, the transition probabilities depend on $v$. As the transition probabilities have a direct impact on the performance of Algorithm 3, we aim to find the optimal threshold value $v^*$ that results in the best performance. Again, all results are the averaged over 100 Monte Carlo runs.

### Optimal threshold value

In the first part of our simulations, we show the impact of the threshold value $v$ on the performance of our proposed myopic-policy-based antenna selection method, i.e., Algorithm 2. Furthermore, we show that the optimal threshold value $v^*$ depends on $M$, $N$, $K$, and $\{\sigma_{h,k}^2\}_{k=1}^K$. Then, we validate the performance of Algorithm 4 for finding the optimal threshold value $v^*$, by comparing it with that by a numerical search. We first consider five different scenarios with different user channel large-scale variations and fading correlation coefficients. Here, for large-scale variation of a channel, we use the path loss model $\sigma_{hk}^2 = \varrho d_k^{-3}$, where $d_k$ is the distance between the $k$-th user and the BS, the path loss exponent is 3, and the path loss constant $\varrho$ is chosen such that at the cell boundary (i.e., for $d_k = 500\text{m}$), $\sigma_{hk}^2/\sigma^2 = -5$ dB. Furthermore, to obtain the fading correlation coefficient $\xi_k$ of the $k$-th user, we use the Jakes' model [74], that is, for a WLAN 802.11 system operating at carrier frequency $f_c = 2.4$ GHz, $\xi_k$ can be obtained as $\xi_k = J_0(2\pi \frac{V_k f_c}{C f_W})$, where $J_0(\cdot)$ is the Bessel function (order zero), $V_k$ is the $k$-th user's speed, $C = 3 \times 10^9$ m/s is the speed of light, and the bandwidth frequency $f_W = 2.5$ KHz. For example, for a user with speed $V_k = 3.6$ km/h (i.e., pedestrian speed), $V_k = 36$ km/h (moderate vehicular speed), and $V_k = 140$ km/h (high vehicular speed) are $\xi_k = 0.999$ $\xi_k = 0.986$, and $\xi_k = 0.95$, respectively. We consider the following five scenarios:

- Scenario i, low-speed and low-SNR users:
  Four users with $\boldsymbol{\xi} = [0.997 \;\; 0.998 \;\; 0.996 \;\; 0.999]^T$, that are located in varies distances from the BS, where the users' SNR, $\{P\sigma_{hk}^2/\sigma^2\}_{k=1}^4$ uniformly distributed in the interval of $[0, 0.5]$ dB.

- Scenario ii, low-speed and high-SNR users:
  Four users, with $\boldsymbol{\xi} = [0.997 \;\; 0.998 \;\; 0.996 \;\; 0.999]^T$, that are located in varies distances from the BS, where the users' SNR, $\{P\sigma_{hk}^2/\sigma^2\}_{k=1}^4$ uniformly distributed in the interval of $[9.5, 10]$ dB.

- Scenario iii, high-speed and low-SNR users:

  Four users, with $\boldsymbol{\xi} = [0.92 \;\; 0.9 \;\; 0.91 \;\; 0.92]^T$, that are located in varies distances from the BS, where the users' SNR, $\{P\sigma^2_{hk}/\sigma^2\}^4_{k=1}$ uniformly distributed in the interval of $[0, 0.5]$ dB.

- Scenario iv, high-speed and high-SNR users:

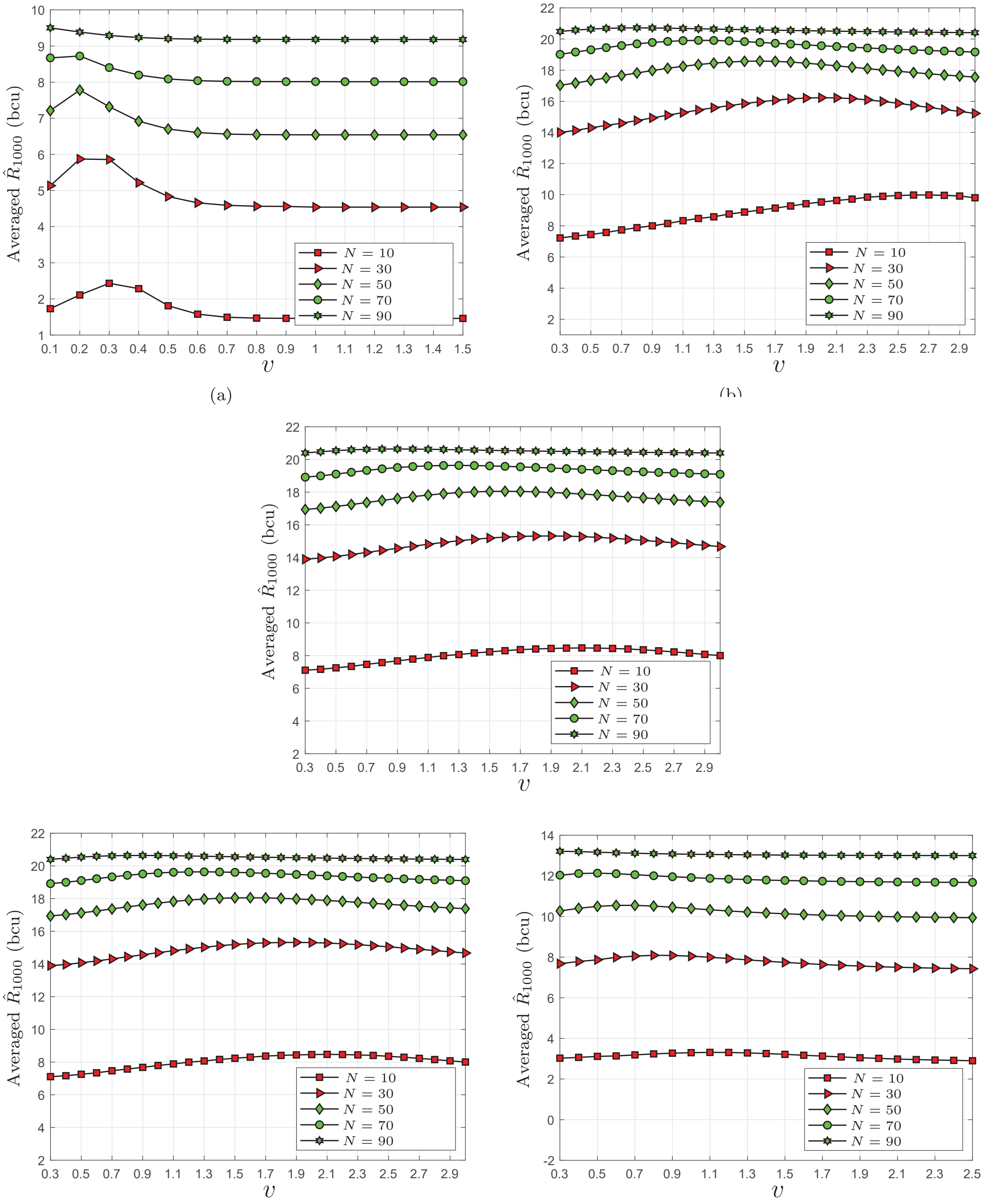
  Four users, with $\boldsymbol{\xi} = [0.92 \;\; 0.9 \;\; 0.91 \;\; 0.92]^T$, that are located in varies distances from the BS, where the users' SNR, $\{P\sigma^2_{hk}/\sigma^2\}^4_{k=1}$ uniformly distributed in the interval of $[9.5, 10]$ dB.

- Scenario v, random speed and random SNR users:

  Four users with $\boldsymbol{\xi} = [0.999 \quad 0.96 \quad 0.97 \quad 0.98]^T$, that are located in varies distances from the BS, where the users' SNR, $\{P\sigma^2_{hk}/\sigma^2\}^4_{k=1}$ uniformly distributed in the interval of $[0, 10]$ dB.

Assuming $K = 4$, $M = 100$, and $N = [10 : 20 : 100]$, we plot averaged sum-rate $\hat{R}_{1000}$ for a range of threshold $v$. The results are shown in Figs 5.4a-5.4e for Scenarios i-v, respectively. These figures show that there is an optimal $v^*$, for these tested scenarios. As can be seen from these figures, for given $M$ and $K$, the impact of the optimal threshold value on the performance of the myopic-policy-based antenna selection algorithm is more noticeable for small values of $N$. This is due to the fact that with decreasing the value of $N$, the choices of selecting $N$ antennas among $M$ available antennas increases, and thus, the performance is more sensitive to labeling channel gains as good state.

Note that in these figures, the accuracy of the exhaustive search approach is limited up to 0.1. It is worth it to mention that in our proposed Algorithm 4, with any changes in $\sigma^2_{h,k}$, we can scale the optimal threshold value to obtain the updated one. To elaborate more on this property, we provide the following illustration. Considering normalized channel vectors ($\sigma^2_{h,k} = 1$ , for $k = 1, 2, \ldots, K$), in Fig. 5.5, we plot the obtained optimal threshold value from Algorithm 4 versus $N$, for



(a)

(b)

(d)

(e)

Figure 5.4: Average $\hat{R}_{1000}$ versus $v$ for $P/\sigma^2 = 5$ dB, $K = 4$, and $M = 100$ (a) Scenario i, (b) Scenario ii, (c) Scenario iii, (d) Scenario iv, and (e) Scenario v.

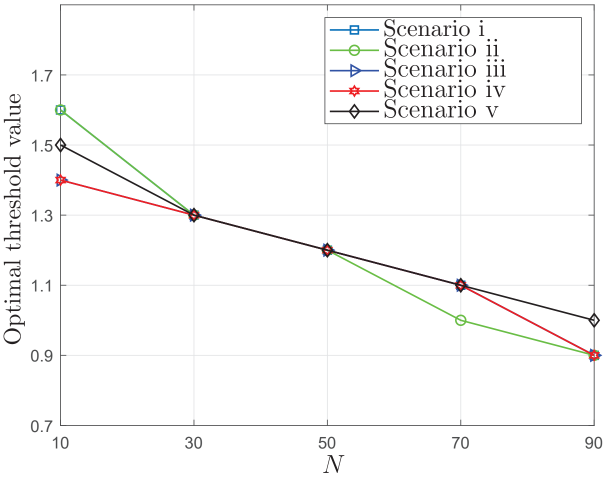


Figure 5.5: Optimal threshold value versus $N$ for Scenarios i-v, for $K = 4$, $M = 100$, and normalized channel vector.

$K = 4$, $M = 100$, and $\boldsymbol{\xi}$ vectors in Scenarios i-v. As can be seen from Fig. 5.5, for given $N$, the optimal threshold values $v^*$ for respective Scenarios i-v are close. This feature allows us to obtain a threshold value for normalized channel vectors, and then scale it to obtain the optimal threshold value corresponding to $\{\sigma^2_{\mathrm{h},k}\}_{k=1}^K$ in each aforemention Scenarios i-v, for the same values of $M$, $N$, and $K$. More specifically, for normalized channel vectors, we first obtain the mean $\mu'$ and the standard derivation $\sigma'_h$ of $\|\mathbf{h}_{i,t}\|^2$ as $\frac{K}{2}$ and $\sqrt{\frac{K}{2}}$, respectively, from (5.42), and then, use these parameters in Algorithm 4 to obtain the optimal threshold value $v^*$. Then, for $\{\sigma^2_{\mathrm{h},k}\}_{k=1}^K$ in each Scenarios i-v, we scale $v^*$ to obtain the optimal threshold value as $(\sigma''_h \frac{(v^* - \frac{K}{2})}{\sqrt{\frac{K}{2}}} + \mu'')$, where $\mu''$ and $\sigma''_h$ are the approximations of the mean and the standard derivation of the channel vectors with $\boldsymbol{\Sigma}_h = \mathrm{diag}([\sigma^2_{\mathrm{h},k}]_{k=1}^K)$ which are obtained from (5.42). So, by obtaining $\mu''$ and $\sigma''_h$ corresponding to each aforemention Scenarios i-v, we can scale the results shown in Fig.5.5 to find the optimal threshold value for each Scenarios i-v. In Table 5.1, for Scenarios i-v, we tabulate the result of obtained optimal threshold value from both exhaustive search and Algorithm 4.

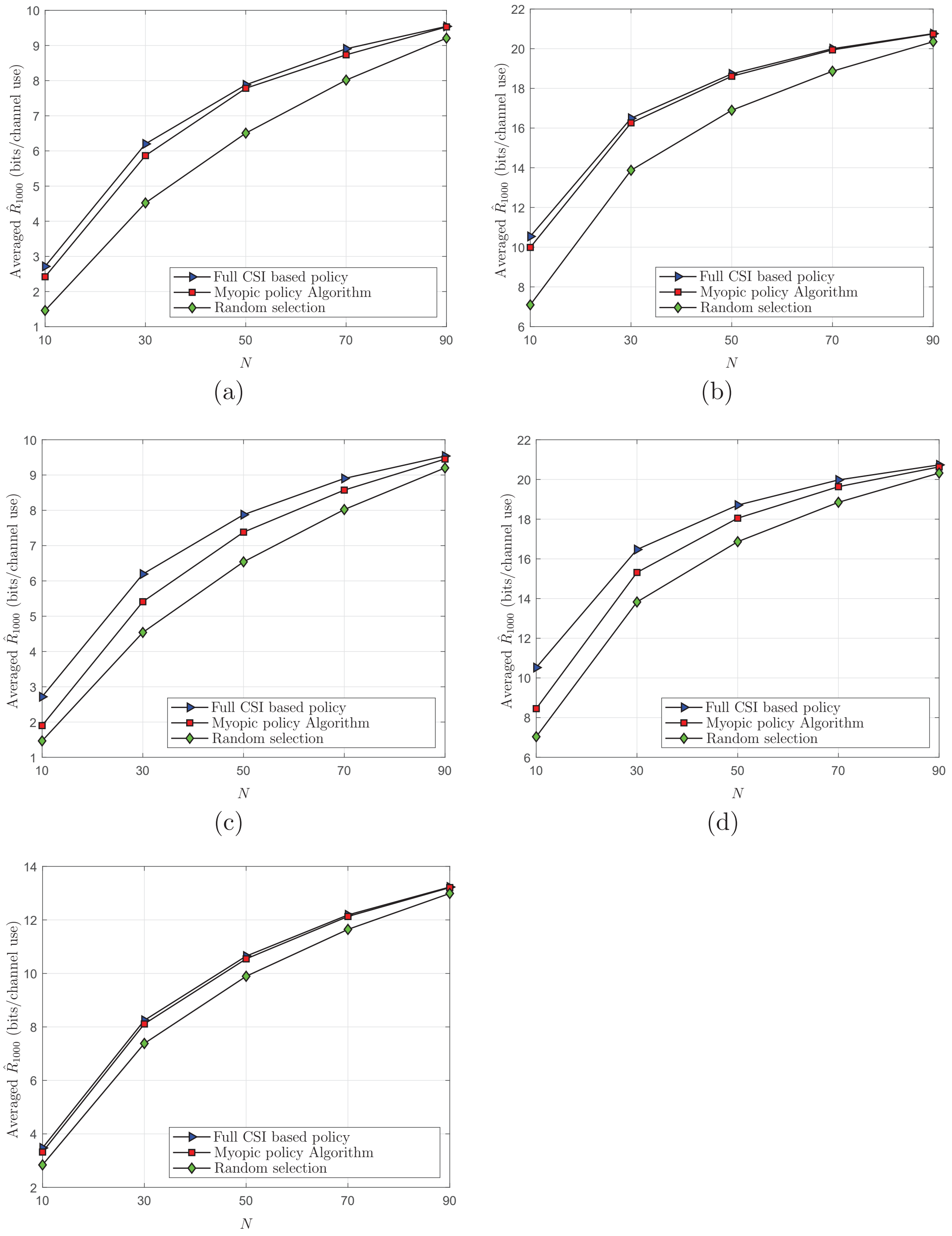
Table 5.1: Values of $v^*$ for different scenarios

| | $N$ | $v^*$, exhaustive search | $v^*$, Algorithm 4 |
|---|---|---|---|
| Scenario i | 10 | 0.3 | 0.335 |
| | 30 | 0.2 | 0.246 |
| | 50 | 0.2 | 0.201 |
| | 70 | 0.2 | 0.156 |
| | 90 | 0.1 | 0.089 |
| Scenario ii | 10 | 2.7 | 2.649 |
| | 30 | 2 | 2.025 |
| | 50 | 1.6 | 1.558 |
| | 70 | 1.2 | 1.246 |
| | 90 | 0.8 | 0.779 |
| Scenario iii | 10 | 0.3 | 0.263 |
| | 30 | 0.2 | 0.239 |
| | 50 | 0.2 | 0.191 |
| | 70 | 0.2 | 0.167 |
| | 90 | 0.1 | 0.119 |
| Scenario iv | 10 | 2.1 | 2.124 |
| | 30 | 1.9 | 1.947 |
| | 50 | 1.6 | 1.598 |
| | 70 | 1.3 | 1.239 |
| | 90 | 0.9 | 0.885 |
| Scenario v | 10 | 1.1 | 1.180 |
| | 30 | 0.8 | 0.853 |
| | 50 | 0.7 | 0.711 |
| | 70 | 0.5 | 0.426 |
| | 90 | 0.3 | 0.284 |

### Performance evaluation of Algorithm 3

In this subsection, using the obtained optimal threshold value $v^*$, we compare the performance of Algorithm 3 with two other schemes: 1) a random antenna selection policy, and 2) a full CSI-based policy, which relies on full CSI of the $M$ antennas to select $N$ antennas with the $N$ highest channel gain. Note that full CSI policy provides an upper bound for the achieved sum-rate. For $K = 4$, and $M = 100$, Figs. 5.6a-5.6e show the average values of $\hat{R}_{1000}$ versus different values of RF chains $N = [10 : 20 : 100]$, for Scenarios i-v, respectively. Figs. 5.6a, and 5.6b show that the performance of Algorithm 3 is within 0.3 bcu from the full CSI based policy.

As can be seen from Figs 5.6c and 5.6d, the performance gap between the myopic policy selection and the full CSI based policy increases. This is expected due to the fact that for high-speed users, $p_{01}$ is significantly higher than that probability for low-speed users. According to (5.27), a higher value of $p_{01}$ lowers the chance of those unselected antennas in a specific time slot, to be selected in the subsequent time slots, as compared with those selected antennas that are in bad state. As such, the algorithm will have less possibility to explore among unselected antennas for the next time slot. However, such exploration (among antennas) possibility increases as $N$ increases, thereby closing the performance gap between the myopic policy and the full CSI based policy. Finally Fig.5.6e shows that in the more realistic Scenario v, where a mixed of low- and high-speed users are to be served, the performance of the myopic policy is very close to that if the full CSI based policy.



(a)

(b)

(c)

(d)

(e)

Figure 5.6: Time averaged sum-rate $\hat{R}_{1000}$ vs $N$ for $K = 4$,and $M = 100$ (a) Scenario i, (b) Scenario ii, (c) Scenario iii, (d) Scenario iv, and (e) Scenario v.

### The impact of increasing the number of users on the performance of Algorithm. 3

To study the impact of increasing number of users on the performance of Algorithm 3, we consider four different scenarios, where in each scenario, all users have the same speed but are located at random distances from the BS. The details of these scenarios are as follows:

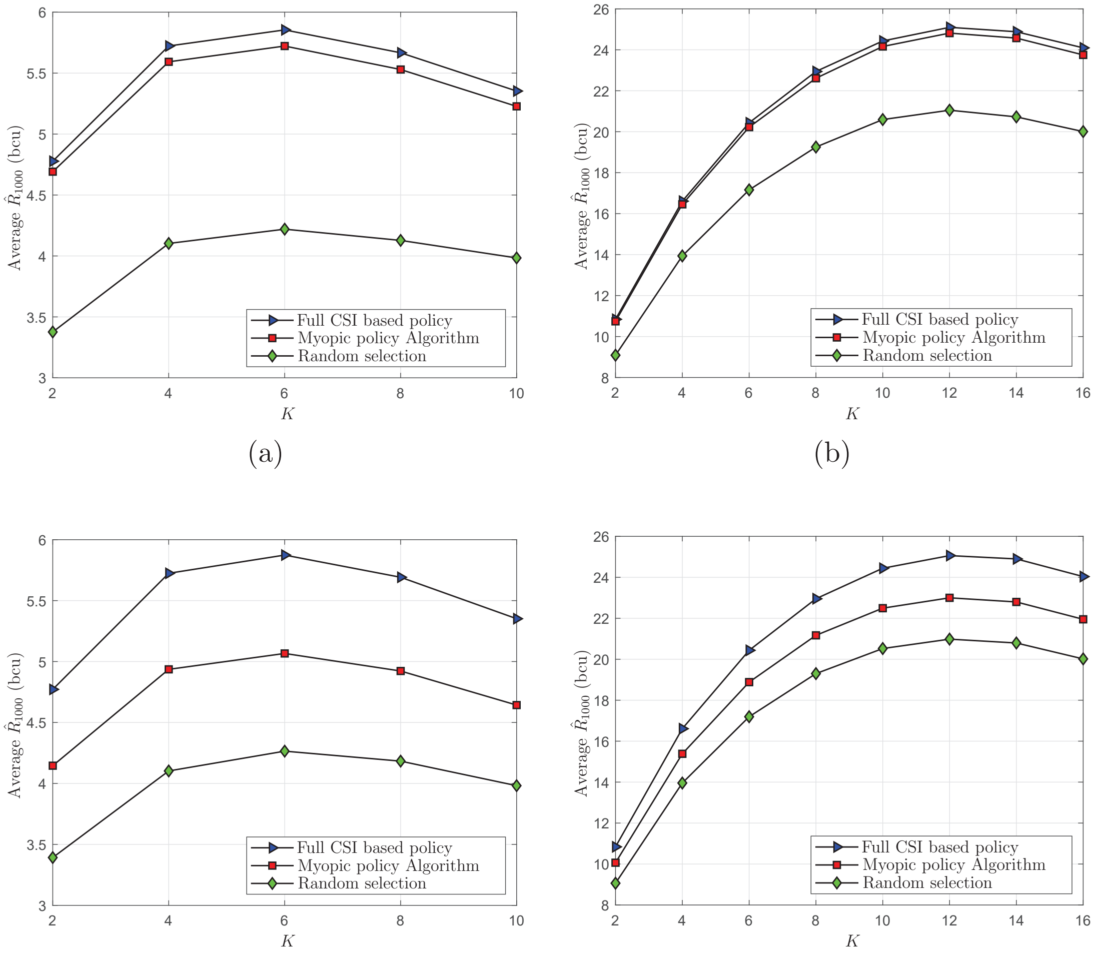
- Scenario vi: Low-speed users, with $\xi_k = 0.999$ for $k = 1, 2, \ldots, K$, are located in varies distances from the BS, where the users' SNR, $\{P\sigma_{hk}^2/\sigma^2\}_{k=1}^4$ uniformly distributed in the interval of $[0, 0.5]$ dB.

- Scenario vii: Low-speed users, with $\xi_k = 0.999$ for $k = 1, 2, \ldots, K$, are located in varies distances from the BS, where the users' SNR ,$\{P\sigma_{hk}^2/\sigma^2\}_{k=1}^4$ uniformly distributed in the interval of $[9.5, 10]$ dB.

- Scenario viii: High-speed users, with $\xi_k = 0.9$ for $k = 1, 2, \ldots, K$, are located in varies distances from the BS, where the users' SNR ,$\{P\sigma_{hk}^2/\sigma^2\}_{k=1}^4$ uniformly distributed in the interval of $[0, 0.5]$ dB.

- Scenario ix: High-speed users, with $\xi_k = 0.9$ for $k = 1, 2, \ldots, K$, are located in varies distances from the BS, where users' SNR, $\{P\sigma_{hk}^2/\sigma^2\}_{k=1}^4$ uniformly distributed in the interval of $[9.5, 10]$ dB.

Considering $M = 100$ and $N = 30$, we first obtain optimal threshold value $v^*$ for different numbers of single-antenna users $K = [2 : 2 : 16]$. We then use the optimal threshold value $v^*$ for each value of $K$ to plot $\hat{R}_{1000}$ versus $K$ in Figs. 5.7a-5.7d for Scenarios vi-ix, respectively. As can be seen from Figs. 5.7a and 5.7b, the performance gap between the myopic policy and the full CSI policy based is relatively small (less than 0.2 bcu use) for low-speed users. This performance gap increases for high-speed users to about less than 0.9 bcu and 1 bcu in Figs. 5.7c and 5.7d, respectively. As can be seen from Fig. 5.7, the gap between the myopic



(a)

(b)

(c)

(d)

Figure 5.7: $\hat{R}_{1000}$ vs $K$ for $N = 30$, and $M = 100$ (a) Scenario vi, (b) Scenario vii, (c) Scenario viii, and (d) Scenario ix.

policy and full CSI/Random selection policy approximately remains unchanged with adding extra users.

Interestingly, it shows from Fig. 5.7 that for given $M$ and $N$, there is an optimal number of users that leads to maximum sum-rate performance. Serving fewer or more users than this optimal number of users leads to performance loss. If the number of users that are to be served is larger than this optimal number of users, user clustering in addition to antenna selection is required. Note that, user scheduling has direct impact on antenna selection techniques and vice versa. Indeed, devising joint antenna selection and user scheduling method is the next step in this line of research that will be presented in the next chapter of this dissertation.

# Chapter 6

# POMDP-based Joint Antenna Selection and User Scheduling Algorithm

## 6.1 System Model

We study a multi-user massive MIMO system where a multi-antenna base station (BS), equipped with $M$ antennas and $N$ RF chains ($M \gg N$), communicates with $U$ single-antenna users in a cell. Here, we assume that the number of available users is larger than the number of RF chains ($U > N$). The system operates in time division duplexing (TDD) mode, and hence, the channel state information (CSI) can be acquired during uplink transmission. Here, we resort to zero forcing (ZF) beamforming to eliminate inter-user interference. To fully cancel out the inter-user interference, at each time slot, the number of scheduled users should be smaller than or equal to the number of active antennas which is the same as the number of available RF chains. Here, to guarantee a fair user scheduling, we assume all users are served in a frame that contains $\lambda$ time slots. We use $\ell$ and $t$ to denote the frame index and time slot index within a frame, respectively. The structure of communication frames is shown in Fig.6.1. As can be seen from this figure, each frame $\ell$, contains $\lambda$ time slots. At each time slot $t$, for $t = 1, 2, \ldots, \lambda$, the BS selects a subset of antennas to serve the users scheduled at that time slot by performing

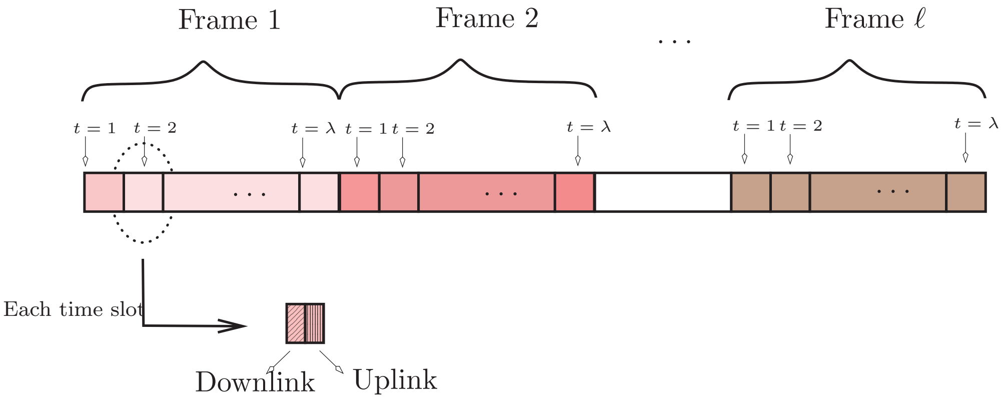


Figure 6.1: An Illustration of a time frame structure

downlink and uplink transmission. Note that we assume channels remain unchanged during the entire frame. Here, a proper user scheduling technique is required to guarantee that all users are served in a time frame, while the number of scheduled users per time slot is larger than one and smaller than or equal to the number of RF chains. More specifically, considering that each frame contains $\lambda$ time slots, at the beginning of each frame, the BS schedules each one of the available users in a time slot to be served. Let us define $\mathcal{K}_{t,l}$ as the set of user indices scheduled at the $t$-th time slot in the $\ell$-th frame, for $t = 1, 2, \ldots, \lambda$ and $\ell = 1, 2, \ldots$. Note that, the number of users scheduled in each time slot should be greater than one and smaller than the number of RF chains, that is, $1 \leq |\mathcal{K}_{t,l}| \leq N$ must hold, for $t = 1, 2, \ldots, \lambda$ and $\ell = 1, 2, \ldots$. Also, at each time slot, only $N$ out of $M$ antennas are selected to participate in data transmission. Here, we consider slow enough fading channels and assume that the channels evolve as a Markov process at the end of each frame. Consequently, as at each time slot, only $N$ antennas are selected to serve a subset of scheduled users, only the CSI between the $N$ selected antennas and the users scheduled in that time slot can be obtained (i.e., partial CSI is available). Our main goal here is to maximize the expected long-term sum-rate over the entire frame by scheduling the best subset of users and finding the best $N$ antennas to participate in data transmission at each time slot. Considering that only partial

CSI is available and the channels' dynamic proceeds as a Markov process, we use POMDP framework to formulate out joint antenna selection and user scheduling (JASUS) problem.

## 6.2 Problem Formulation

To formulate the system model, we denote the random time-varying channel matrix between $M$ antennas and $U$ users in frame $\ell$ as $\mathsf{H}_\ell \triangleq [\mathsf{h}_{1,\ell} \;\; \mathsf{h}_{2,\ell} \;\; \cdots \;\; \mathsf{h}_{M,\ell}]$, where $\mathsf{h}_{i,\ell} = [\mathsf{h}_{ui,\ell}]_{u=1}^{U}$ is the $U \times 1$ random between the $i$-th antenna and all $U$ users at the $\ell$-th frame, where $\mathsf{h}_{ui,\ell}$ is the channel coefficient between the $u$-th user and the $i$-th antenna, for $i = 1, 2, \dots, M$ and $u = 1, 2, \dots, U$. Note that we assume channels evolve at the beginning of each frame and remain unchanged during the entire frame. As defined earlier, $\boldsymbol{\mathcal{K}}_{t,\ell}$ is a random set containing the indices of the users scheduled at the $t$-th time slot of the $\ell$-th frame. We use $\hat{\mathsf{H}}_{t,\ell} \in \mathbb{C}^{|\boldsymbol{\mathcal{K}}_{t,\ell}| \times N}$ to denote the channel matrix between the $N$ selected antennas and $|\boldsymbol{\mathcal{K}}_{t,\ell}|$ scheduled users at the $t$-th time slot of the $\ell$-th frame. Note that $\hat{\mathsf{H}}_{t,\ell}$ is a $|\boldsymbol{\mathcal{K}}_{t,\ell}| \times N$ sub-matrix of $\mathsf{H}_\ell$. If $\hat{\mathbf{H}}$ and $\mathcal{K}$ are any realizations of $\hat{\mathsf{H}}_{t,\ell}$ and $\boldsymbol{\mathcal{K}}_{t,\ell}$, respectively, we can write the vector $\mathbf{y} \in \mathbb{C}^{|\mathcal{K}| \times 1}$ of the corresponding signals received at the users in set $\mathcal{K}$ as

$$\mathbf{y} = \hat{\mathbf{H}}\mathbf{W}\mathbf{x} + \mathbf{n}, \tag{6.1}$$

where $\mathbf{x} \in \mathbb{C}^{|\mathcal{K}| \times 1}$ and $\mathbf{n} \in \mathbb{C}^{|\mathcal{K}| \times 1}$ are the transmitted signal vector and the noise vector, respectively, and $\mathbf{W}$ is the $N \times |\mathcal{K}|$ pre-coding (ZF beamforming) matrix. We can obtain $\mathbf{W}$ as

$$\mathbf{W} = \hat{\mathbf{H}}^H(\hat{\mathbf{H}}\,\hat{\mathbf{H}}^H)^{-1}\sqrt{\frac{P}{\|\hat{\mathbf{H}}^H(\hat{\mathbf{H}}\,\hat{\mathbf{H}}^H)^{-1}\|_F^2}}\,, \tag{6.2}$$

where $P$ is the total power and $\|\cdot\|_F$ stands for the Frobenius norm. We can write

$$\|\hat{\mathbf{H}}^H(\hat{\mathbf{H}}\,\hat{\mathbf{H}}^H)^{-1}\|_F^2 = Tr((\hat{\mathbf{H}}\,\hat{\mathbf{H}}^H)^{-1}) = \sum_{k=1}^{|\mathcal{K}|} \frac{1}{\lambda_k^2}, \tag{6.3}$$

where $\lambda_k$ is the $k$-th singular value of $\hat{\mathbf{H}}$. We use (6.3) to write sum-rate in the $t$-th time slot as

$$R(\hat{\mathbf{H}}) = \log_2\Big(\det\big(\mathbf{I} + \frac{P}{\sigma^2\|\hat{\mathbf{H}}^H(\hat{\mathbf{H}}\,\hat{\mathbf{H}}^H)^{-1}\|_F^2}\mathbf{I}\big)\Big) = |\mathcal{K}|\log_2\Big(1+\frac{P}{\sigma^2\sum_{k=1}^{|\mathcal{K}_t|}\frac{1}{\lambda_k^2}}\Big), \tag{6.4}$$

where $P$ is the transmit power at BS and $\sigma^2$ is the noise variance. To design a tractable solution for the joint antenna selection and user scheduling problem, we use an upper bound for the sum-rate in (5.4) in our POMDP formulation. To this end, considering the fact that $\Big(\frac{1}{|\mathcal{K}|}\sum_{k=1}^{|\mathcal{K}_t|}\frac{1}{\lambda_k^2}\Big)^{-1} \leqslant \frac{1}{|\mathcal{K}|}\Big(\sum_{k=1}^{|\mathcal{K}|}\lambda_k^2\Big) = \frac{1}{|\mathcal{K}|}Tr(\hat{\mathbf{H}}^H\hat{\mathbf{H}})$ holds true, we introduce the $t$-th time slot sum-rate upper-bound $\hat{R}_{\mathrm{u}}(\hat{\mathbf{H}})$, for any given channel realization $\hat{\mathbf{H}}$, as

$$\hat{R}_{\mathrm{u}}(\hat{\mathbf{H}}) = \ |\mathcal{K}|\log_2\Big(1+\frac{P(Tr(\hat{\mathbf{H}}^H\hat{\mathbf{H}}))}{\sigma^2|\mathcal{K}|^2}\Big). \tag{6.5}$$

Let us define the random selected antennas set at the $t$-th time slot in the $\ell$-th frame as $\boldsymbol{\mathcal{I}}_{t,\ell}$ and use (6.5) to write the *random* sum-rate upper-bound of the $t$-th time slot of the $\ell$-th frame as

$$\hat{R}_{\mathrm{u}}(\hat{\mathsf{H}}_{t,\ell}) = |\boldsymbol{\mathcal{K}}_{t,\ell}|\log_2\Big(1+\frac{P}{\sigma^2|\boldsymbol{\mathcal{K}}_{t,\ell}|^2}\sum_{i\in\boldsymbol{\mathcal{I}}_{t,\ell}}\sum_{u\in\boldsymbol{\mathcal{K}}_{t,\ell}}|\mathsf{h}_{ui,\ell}|^2\Big). \tag{6.6}$$

Furthermore, we define $\hat{\boldsymbol{\mathcal{H}}}_\ell = \{\hat{\mathsf{H}}_{t,\ell}\}_{t=1}^{\lambda}$ and write the *random* upper-bound sum-rate over the entire frame as

$$R_{\mathrm{u}}(\hat{\boldsymbol{\mathcal{H}}}_\ell) = \frac{1}{\lambda}\sum_{t=1}^{\lambda}|\boldsymbol{\mathcal{K}}_{t,\ell}|\log_2\Big(1+\frac{P}{\sigma^2|\boldsymbol{\mathcal{K}}_{t,\ell}|^2}\sum_{i\in\boldsymbol{\mathcal{I}}_{t,\ell}}\sum_{u\in\boldsymbol{\mathcal{K}}_{t,\ell}}|\mathsf{h}_{ui,\ell}|^2\Big). \tag{6.7}$$

Here, we aim to design an optimal joint antenna selection and user scheduling policy to maximize the expected long-term sum-rate upper-bound over frame, given by

$$E_{\{\hat{\boldsymbol{\mathcal{H}}}_\ell\}}\Big\{\sum_{\ell=0}^{\infty}R_{\mathrm{u}}(\hat{\boldsymbol{\mathcal{H}}}_\ell)\Big\}. \tag{6.8}$$

In the next section, we formulate the joint antenna selection and user scheduling problem using a POMDP framework with the aim to maximize the expected long-term sum-rate upper-bound, defined in (6.8).

## 6.3 POMDP Formulation

In this section, we first define the POMDP components, and then, formulate the JASUS problem using a POMDP framework. A POMDP formulation is represented by the following components:

$$(\mathcal{S}, \mathcal{A}, \mathbf{T}, \mathcal{R}(\mathbf{S}, \mathbf{A}), \mathcal{O}, P_{\mathbf{O}}(\mathbf{O}, \mathbf{A}), \mathbf{b}) \tag{6.9}$$

where the state space denoted as $\mathcal{S}$ is the set of all possible states; the action space denoted as $\mathcal{A}$ is the set of all possible actions; $\mathbf{T}$ is a $|\mathcal{S}| \times |\mathcal{S}|$ matrix of state transition probabilities; $R(\mathbf{S}, \mathbf{A})$ is the reward at state $\mathbf{S}$ when action $\mathbf{A} \in \mathcal{A}$ is taken; the observation state denoted as $\mathcal{O}$ is a set of all possible observations; $P_{\mathbf{O}}(\mathbf{O}, \mathbf{A})$ is a diagonal matrix whose diagonal elements are the probabilities of observing $\mathbf{O}$ at different states, when action $\mathbf{A}$ is taken; and $\mathbf{b}$ is the belief vector of probability distribution over all possible states in $\mathcal{S}$.

Given the above definitions, we now formulate the JASUS problem as a POMDP problem.

### State space

The state space $\mathcal{S}$ is the set of finite number of states labeled as $\mathbf{S}_j$. The $j$-th point of the state space, $\mathbf{S}_j$, takes one of the possible antennas channel coefficients matrices, such that

$$\mathbf{S}_j = \tilde{\mathbf{H}}_j \triangleq [\tilde{\mathbf{h}}_{1,j} \;\; \tilde{\mathbf{h}}_{2,j} \;\; \cdots \;\; \tilde{\mathbf{h}}_{M,j}], \tag{6.10}$$

where $\tilde{\mathbf{h}}_{i,j} = [\tilde{h}_{ui,j}]_{u=1}^{U}$, for $i = 1, 2, \ldots, M$, and $\tilde{h}_{ui,j}$ is the $j$-th possible value that square absolute value of the channel between $i$-th antenna and $u$-th user can take. Indeed, we assume that each $\tilde{h}_{ui,j}$ is quantized to $Q$ levels, i.e., $\tilde{h}_{ui,j} \in \{\alpha_1, ..., \alpha_Q\}$, where $\alpha_q$ is the $q$-th quantization level. As a result, we can write the state space as

$$\mathcal{S} \triangleq \{\mathbf{S}_1, \mathbf{S}_2, \ldots, \mathbf{S}_{Q^{UM}}\}. \tag{6.11}$$

The *random time-varying* state of the system at the $\ell$-th frame. denoted as $\mathsf{S}_\ell = [|\mathsf{h}_{ui,\ell}|^2]_{\substack{i=1,2,\ldots,M\\u=1,2,\ldots,U}}$, can take one of the $Q^{UM}$ elements in $\mathcal{S}$.

### Action space

The action space $\mathcal{A}$ is defined as the set of all possible actions. For the JASUS problem, action consists of scheduling users and selecting antennas at each time slot. For ease of notation, let us break down the action of our joint antenna selection and user scheduling problem to two sperate actions: the antenna selection action and user scheduling action.

We denote $\tilde{\mathcal{A}}$ as the set that contains all possible antenna selection action. The antenna selection action amounts to selecting $N$ out of $M$ antennas for each time slot $t$ in a frame, for $t = 1, 2, \ldots, \lambda$. Thus, the number of possible actions for antenna selection is $|\tilde{\mathcal{A}}| = \binom{M}{N}^\lambda$. We represent the $j$-th possible antenna selection action as an $M \times \lambda$ matrix denoted as $\tilde{\mathbf{A}}_j = [\tilde{\mathbf{a}}_{1,j} \;\; \tilde{\mathbf{a}}_{2,j} \;\; \ldots \;\; \tilde{\mathbf{a}}_{\lambda,j}]$, where

$$\tilde{\mathbf{a}}_{t,j} = [\tilde{a}_{1t,j} \;\; \tilde{a}_{2t,j} \;\; \ldots \;\; \tilde{a}_{M,tj}]^T, \;\; \tilde{a}_{it,j} \in \{0,1\}, \quad \text{for } t = 1,\ 2,\ \ldots, \lambda. \tag{6.12}$$

Here, for $t = 1, 2, \ldots, \lambda$ and $i = 1, 2, \ldots, M$; $\tilde{a}_{it,j} = 1$ means that the $i$-th antenna at the $t$-th time slot is selected to participate in data exchange, otherwise $\tilde{a}_{it,j} = 0$. Note that only $N$ elements of $\tilde{\mathbf{a}}_{tj}$ are equal to one and the remaining elements are zero. The random antenna selection action matrix at the $\ell$-th frame denoted as $\mathsf{A}'_\ell = [\mathsf{a}'_{t,\ell}]_{t=1}^{\lambda}$ (where $\mathsf{a}'_{t,\ell}$ is the random antenna selection vector at time slot $t$ in the $\ell$-th frame) can take one of the possible action matrices in $\tilde{\mathcal{A}}$.

We now define the user scheduling action, which is assigning each user to one of the $\lambda$ time slots at the beginning of each frame. To do so, we denote the $n$-th possible $U \times 1$ user scheduling vector as $\hat{\mathbf{a}}_n = [\hat{a}_{1,n} \;\; \hat{a}_{2,n} \;\; \ldots \hat{a}_{U,n}]^T$, where $\hat{a}_{u,n}$ being equal to $t \in \{1, 2, \ldots, \lambda\}$ means that the $u$-th user is scheduled in the $t$-th time slot in a frame. Note that at each time slot, at least one user and at most $N$ users can be scheduled. Based on these limitations, the action space for all possible user scheduling vectors can be written as

$$\bar{\mathcal{A}} = \{\hat{\mathbf{a}}_n \mid 1 \leq \sum_{u=1}^{U} \mathbb{I}(\hat{a}_{u,n} = t) \leq N, \text{ for } t = 1, 2, \ldots, \lambda\} \tag{6.13}$$

where $\mathbb{I}(\hat{a}_{u,n} = t)$ is an indicator function defined as

$$\mathbb{I}(\hat{a}_{u,n} = t) = \begin{cases} 1, & \text{if } \hat{a}_{u,n} = t \\ 0, & \text{otherwise.} \end{cases} \tag{6.14}$$

Note that, the user scheduling vector $\hat{\mathbf{a}}_n$ is equivalent to a $U \times \lambda$ matrix denoted as $\bar{\mathbf{A}}_n = [\bar{\mathbf{a}}_{1,n} \;\; \bar{\mathbf{a}}_{2,n} \;\; \ldots \;\; \bar{\mathbf{a}}_{\lambda,n}]$, where

$$\bar{\mathbf{a}}_{tn} \;= [\bar{a}_{1t,n} \;\; \bar{a}_{2t,n} \;\; \ldots \;\; \bar{a}_{Ut,n}]^T, \;\; \bar{a}_{ut,n} \in \{0, 1\}, \quad \text{for } t = 1,\; 2,\; \ldots, \lambda. \tag{6.15}$$

Here, $\bar{a}_{ut,n} = 1$, if the $u$-th user is scheduled at the $t$-th time slot, for $t = 1, 2, \ldots, \lambda$, and $u = 1, 2, \ldots, U$. The random user scheduling action at the $\ell$-th frame denoted as $\mathsf{A}''_\ell = [\mathsf{a}''_{t,\ell}]_{t=1}^{\lambda}$ (where $\mathsf{a}''_{t,\ell}$ is the random user scheduling action vector at time slot $t$ in the $\ell$-th frame), takes one of the possible action vector $\bar{\mathbf{A}}_n$, for $n = 1, 2, \ldots, |\bar{\mathcal{A}}|$.

Here, we denote the $z$-th possible JASUS action as $\mathbf{A}'_z$, where $\mathbf{A}^{'}_z = \begin{bmatrix} \tilde{\mathbf{A}}_j \\ \bar{\mathbf{A}}_n \end{bmatrix}$, for $j = 1, 2, \ldots, |\tilde{\mathcal{A}}|$ and $n = 1, 2, \ldots, |\bar{\mathcal{A}}|$. Thus, the total number of possible JASUS actions is $Z = \binom{M}{N}^{\lambda} \times |\bar{\mathcal{A}}|$ and we can write our action space as

$$\mathcal{A} \triangleq \{\mathbf{A}'_1, \mathbf{A}'_2, \ldots, \mathbf{A}'_Z\}. \tag{6.16}$$

We denote $\mathbf{A}_\ell \in \mathcal{A}$ as the JASUS action matrix taking at the $\ell$-th frame and $\mathsf{A}_\ell \in \mathcal{A}$ as the random JASUS action matrix at the $\ell$-th frame, for $\ell = 1, 2, \ldots \infty$.

**Transition probability**

The transition probability matrix, denoted by $\mathbf{T}$, is a $Q^{UM} \times Q^{UM}$ matrix whose $(i, j)$-th element, denoted as $T_{ij}$ , is the probability of the state at frame $\ell$ being equal to $\mathbf{S}_j$, given that the state at frame $\ell - 1$ is $\mathbf{S}_i$. Note that in our problem, the selected action does not impact the channel variations, and thus, the transition probability matrix is independent of the taken action. The $(i, j)$-th element of the transition probability matrix $T_{ij}$ can be written as

$$T_{ij} = \Pr(\mathbf{S}_\ell = \mathbf{S}_j | \mathbf{S}_{\ell-1} = \mathbf{S}_i). \tag{6.17}$$

Here, we assume that the transition probability matrix $\mathbf{T}$ is known.

### Observation space

The observation space denoted by $\mathcal{O}$ is the set of all possible observation matrices. We denote the $q$-th possible observation matrix as $\mathbf{O}_q$, where

$$\mathbf{O}_q \; = \sum_{t=1}^{\lambda} \text{diag}(\bar{\mathbf{a}}_{t,n}) \, \mathbf{S}_j \, \text{diag}(\tilde{\mathbf{a}}_{t,p}), \tag{6.18}$$

for $j = 1, 2, \ldots, |\mathcal{S}|$, $p = 1, 2, \ldots, |\tilde{\mathcal{A}}|$, and $n = 1, 2, \ldots, |\bar{\mathcal{A}}|$. Let us denote $\bar{\mathbf{O}}_{t,\ell}$, as the random observation matrix at time slot $t$ in the $\ell$-th frame such that

$$\bar{\mathbf{O}}_{t,\ell} = [\bar{\mathsf{o}}_{uit.\ell}]_{\substack{i=1,2,\ldots,M \\ u=1,2,\ldots,U}} \tag{6.19}$$

where $\bar{\mathsf{o}}_{uit.\ell}$ can take one of the possible quantized channel coefficient level, $\alpha_q$ for $q = 1, 2, \ldots, Q$, if the $i$-th antenna is selected to serve the $u$-th user at time slot $t$ in the $\ell$-th frame; otherwise, the value of $\mathsf{o}_{uit,\ell} = 0$, for $t = 1, 2, \ldots, \lambda$, $i = 1, 2, \ldots M$, and $u = 1, 2, \ldots U$. We can obtain $\bar{\mathbf{O}}_{t,\ell}$ as

$$\bar{\mathbf{O}}_{t,\ell} = \text{diag}(\mathbf{a}''_{t,\ell}) \, \mathbf{S}_\ell \, \text{diag}(\mathbf{a}'_{t,\ell}), \tag{6.20}$$

where $\mathbf{a}'_{t,\ell}$ is the first $M$ rows of the $t$-th column of $\mathbf{A}_\ell$, while $\mathbf{a}''_{t,\ell}$ is the last $U$ rows of the $t$-th column of $\mathbf{A}_\ell$. Thus, the random observation matrix of the entire $\ell$-th frame denoted as $\mathbf{O}_\ell$, can be obtained as

$$\mathbf{O}_\ell = \sum_{t=1}^{\lambda} \bar{\mathbf{O}}_{t,\ell}, \tag{6.21}$$

where $\mathbf{O}_\ell \in \mathcal{O}$.

### Observation probability

We denote the conditional observation probability matrix as $P_{\mathbf{O}}(\mathbf{O}, \mathbf{A})$, that is a $|\mathcal{S}| \times |\mathcal{S}|$ diagonal matrix. We define $P_{\mathbf{O}}(\mathbf{O}, \mathbf{A})$ as

$$P_{\mathbf{O}}(\mathbf{O}, \mathbf{A}) = \text{diag}\Big(P_r\Big\{\mathbf{O}_\ell = \mathbf{O} | \mathbf{S}_\ell = \mathbf{S}_j, \mathbf{A}_\ell = \mathbf{A}\Big\}_{j=1}^{Q^{UM}}\Big), \tag{6.22}$$

whose the $j$-th diagonal element is the probability of observing $\mathbf{O} \in \mathcal{O}$ at frame $\ell$, when action $\mathbf{A} \in \mathcal{A}$ is taken and the state at the $\ell$-th frame is $\mathbf{S}_j$.

### Reward

The function $R(\mathbf{S}, \mathbf{A})$ represents the immediate reward function for taking action $\mathbf{A}_\ell = \mathbf{A}$ at state $\mathbf{S}_\ell = \mathbf{S}$. Note that here, we assume that the state remains unchanged during the entire frame. We use the defined total sum-rate upper-bound in (6.7) to define the reward function for taking action $\mathbf{A}_\ell = \mathbf{A}$ at state $\mathbf{S}_\ell = \mathbf{S}$ as

$$R(\mathbf{S}, \mathbf{A}) \triangleq \frac{1}{\lambda} \sum_{t=1}^{\lambda} |\mathcal{K}(\mathbf{a}''_t)| \log_2 \Big(1 + \frac{P(Tr(\bar{\mathbf{O}}_t^H \bar{\mathbf{O}}_t)}{\sigma^2 |\mathcal{K}(\mathbf{a}''_t)|^2}\Big), \tag{6.23}$$

where $\mathbf{a}'_t$ is the first $M$ rows of the $t$-th column of $\mathbf{A}$, while $\mathbf{a}''_t$ is the last $U$ rows of the $t$-th column of $\mathbf{A}$, and $\mathcal{K}(\mathbf{a}''_t)$ is the set of the indices of the users scheduled at the $t$-th time slot. Also, for given action $\mathbf{a}''{}_t$, $\bar{\mathbf{O}}_t = \text{diag}(\mathbf{a}''_t)\ \mathbf{S}\ \text{diag}(\mathbf{a}'_t)$. We can rewrite (6.23) as

$$R(\mathbf{S}, \mathbf{A}) = \frac{1}{\lambda} \sum_{t=1}^{\lambda} |\mathcal{K}(\mathbf{a}''_t)| \log_2 \Big(1 + \frac{P}{\sigma^2 |\mathcal{K}(\mathbf{a}''_t)|^2} \sum_{i \in \mathcal{I}(\bar{\mathbf{a}}_t)} \sum_{u \in \mathcal{K}(\bar{\mathbf{a}}_t)} |h_{ui}|^2\Big), \tag{6.24}$$

where $h_{ui}$ is the channel coefficient between the $i$-th antenna and $u$-th user. Here, $|\mathcal{K}(\mathbf{a}''_t)| = \|\mathbf{a}''{}_t\|_1$, is the $\ell$-1 norm of vector $\bar{\mathbf{a}}_t$ that shows the total number of scheduled users at time slot $t$. Also, $\mathcal{I}(\mathbf{a}'_t)$ is the set of the indices of all selected antennas at the $t$-th time slot, when action $\mathbf{a}'_t$ is taken.

### Belief vector

At the $\ell$-th frame, the belief vector is defined as $\mathbf{b}_\ell \triangleq [b_{1,\ell}\ \ b_{2,\ell}\ \ \ldots\ \ b_{Q^{UM},\ell}]^T$, where $b_{j,\ell}$ is the probability of the state at frame $\ell$ being equal to $\mathbf{S}_j \in \mathcal{S}$, given all the previous actions and obtained observations until frame $\ell - 1$. Let us use $\mathcal{H}_{\ell-1}$ to denote the action and observation history until frame $\ell - 1$, which can be written as

$$\mathcal{H}_{\ell-1} \triangleq \{\mathbf{O}_{\ell-1}, \mathbf{A}_{\ell-1}, \mathcal{H}_{\ell-2}\}, \tag{6.25}$$

then we can write the $j$-th element of the belief vector as

$$b_{j,\ell} = \Pr\{\mathbf{S}_\ell = \mathbf{S}_j | \mathcal{H}_{\ell-1}\}. \tag{6.26}$$

We also define the belief space as $\mathcal{B} \triangleq \left\{ \mathbf{b} \in \mathbb{R}^{|\mathcal{S}|} : \mathbf{1}^T\mathbf{b} = 1, \mathbf{b} \succcurlyeq \mathbf{0} \right\}$. Note that $\mathbf{b}_\ell$ is a sufficient statistic for $\mathcal{H}_{\ell-1}$ [55], and as we showed in [31], we can obtain $\mathbf{b}_\ell$ from $\mathbf{b}_{\ell-1}$ as

$$\mathbf{b}_\ell = \mathbf{g}(\mathbf{O}_{\ell-1}, \mathbf{A}_{\ell-1}, \mathbf{b}_{\ell-1}), \tag{6.27}$$

where

$$\mathbf{g}(\mathbf{O}, \mathbf{A}, \mathbf{b}) \triangleq \frac{P_{\boldsymbol{\mathsf{O}}}(\mathbf{O}, \mathbf{A})\mathbf{T}\mathbf{b}}{g(\mathbf{O}, \mathbf{A}, \mathbf{b})}, \tag{6.28}$$

$$g(\mathbf{O}, \mathbf{A}, \mathbf{b}) = \mathbf{1}^T P_{\boldsymbol{\mathsf{O}}}(\mathbf{O}, \mathbf{A})\mathbf{T}\mathbf{b}, \tag{6.29}$$

for $\mathbf{O} \in \mathcal{O}$ and $\mathbf{A} \in \mathcal{A}$. For a given action matrix and belief vector at frame $\ell - 1$, the belief vector at frame $\ell$ depends on the frame observation matrix $\boldsymbol{\mathsf{O}} \in \mathcal{O}$ which is a random matrix, and thus, the believe vector in (6.27) also becomes a random vector. We denote the random belief vector as $\boldsymbol{\mathsf{b}}_\ell$ which can be written as

$$\boldsymbol{\mathsf{b}}_\ell \triangleq \mathbf{g}(\boldsymbol{\mathsf{O}}_{\ell-1}\ , \mathbf{A}_{\ell-1}, \mathbf{b}_{\ell-1}). \tag{6.30}$$

Also, we use $\boldsymbol{\mathcal{H}}_{\ell-1}$ to represent the collection of all past observations and actions as random vectors until frame $\ell - 1$ which is given by

$$\boldsymbol{\mathcal{H}}_{\ell-1} \triangleq \{\boldsymbol{\mathsf{O}}_{\ell-1}, \boldsymbol{\mathsf{A}}_{\ell-1}, \boldsymbol{\mathcal{H}}_{\ell-2}\}. \tag{6.31}$$

Note that, the observation history $\mathcal{H}_{\ell-1}$ defined in (6.25) is a realization of $\boldsymbol{\mathcal{H}}_{\ell-1}$ after observing $\mathbf{O}_{\ell-1} \in \mathcal{O}$.

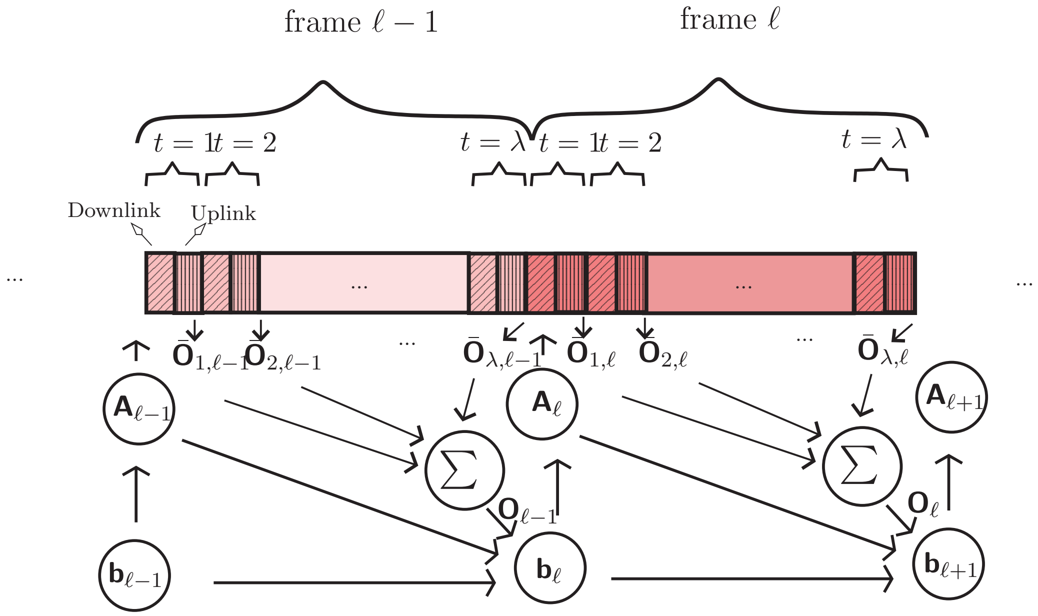


Figure 6.2: Illustration of the JASUS problem as a POMDP process.

Given the above defined POMDP-based JASUS problem, we describe the dynamic of a real-time POMDP controller in the sequel. We assume that each frame includes $\lambda$ time slots and each time slot includes one downlink and uplink transmission. We assume that channels evolve at the beginning of each frame and remain unchanged during the entire frame. Note that here the system mode is TDD, and thus, at each time slot, the channel coefficients between the selected antennas and scheduled users can be obtained at the end of the uplink transmission. At the beginning of each frame, the BS chooses the JASUS action matrix, meaning that the BS schedules all available $U$ users in $\lambda$ time slots and determines the selected $N$ out of $M$ antennas to serve scheduled users for each time slot in that frame. According to our defined user scheduling restrictions, at least one user and a maximum of $N$ users can be scheduled in each time slot in a frame. In addition to that, for designing a fair scheduling method, we guarantee that all $U$ available users receive data only once in a frame, meaning that each user will be scheduled in one time slot in a frame. At frame $\ell-1$, at time slot $t$, the observation matrix $\bar{\mathbf{O}}_{t,\ell-1}$ can be obtained from the uplink transmission for $t = 1, 2, \dots, \lambda$. At the end of the frame $\ell - 1$, $\mathbf{O}_{\ell-1}$ can be obtained by using (6.21). Given that action $\mathbf{A}_{\ell-1}$ was taken, the BS uses obtained $\mathbf{O}_{\ell-1}$ and all the available history information until frame $\ell - 1$ (which is represented in $\mathbf{b}_{\ell-1}$) to obtain (update) the belief vector $\mathbf{b}_\ell$, and then, chooses an action matrix at frame $\ell$. The main goal here is to take optimal actions $\{\mathbf{A}_\ell\}_{\ell=0}^{+\infty}$ such that the expected cumulative reward $E\Big\{ \sum_{\ell=0}^{+\infty} R(\mathbf{S}_\ell, \mathbf{A}_\ell)\Big\}$, is maximized. Here, $E\{\cdot\}$ is the expectation taken with respect to random states $\{\mathbf{S}_\ell\}_{\ell=0}^{+\infty}$.

### 6.3.1 Objective Function

At each frame $\ell$, given the updated belief vector $\mathbf{b}_\ell$, the BS selects an action that is $\mathbf{A}_\ell = \pi_\ell(\mathbf{b}_\ell)$, where $\pi_\ell(\cdot)$ is used to denote policy for frame $\ell = 1, 2, \dots, \infty$. Note that since our problem is an infinite horizon POMDP problem, our policy is stationary meaning that a single decision-making rule $\pi(\cdot)$, can be used for all time frames [55].

Considering above definition of policy, we denote the objective function as $J_\pi(\mathbf{b}_0)$, where $\mathbf{b}_0$ is the initial belief vector, and we define $J_\pi(\mathbf{b}_0)$ as

$$J_\pi(\mathbf{b}_0) = E_{\{\mathbf{S}_\ell\}}\Big\{ \sum_{\ell=0}^{\infty} R(\mathbf{S}_\ell, \mathbf{A}_\ell)\Big|\mathbf{b}_0\Big\} = E_{\{\mathbf{S}_\ell\}}\Big\{ \sum_{\ell=0}^{\infty} R(\mathbf{S}_\ell, \pi(\mathbf{b}_\ell))\Big|\mathbf{b}_0\Big\}, \tag{6.32}$$

where $\mathbf{S}_\ell \in \mathcal{S}$, $\mathbf{A}_\ell = \pi(\mathbf{b}_\ell) \in \mathcal{A}$ and $E_{\{\mathbf{S}_\ell\}}\{\cdot\}$ is the expectation which is taken with respect to the states probability distribution $\{\mathbf{S}_\ell\}_{\ell=0}^{+\infty}$, given initial belief vector $\mathbf{b}_0$, for $\ell = 0, 1, 2, \dots, \infty$. Note that, the notation $J_\pi(\mathbf{b}_0)$ presents the objective function under policy $\pi(\cdot)$. The main goal here is to find the optimal policy such that

$$\pi^* = \arg\max_{\pi} J_\pi(\mathbf{b}_0), \text{ for any } \mathbf{b}_0. \tag{6.33}$$

Considering the fact that random action matrix $\mathbf{A}_\ell$ is a function of the belief vector $\mathbf{b}_\ell$, and $\mathbf{b}_\ell$ is a function of $\mathbf{O}_{\ell-1}$, as shown in [31](see Section. 3.3), we can rewrite $J_\pi(\mathbf{b}_0)$ as

$$J_\pi(\mathbf{b}_0) = \sum_{\ell=0}^{+\infty} E_{\boldsymbol{\mathcal{H}}_{\ell-1}}\Big\{ \sum_{j=1}^{|\mathcal{S}|} R(\mathbf{S}_j, \mathbf{A}_\ell)\mathsf{b}_{j,\ell}\Big|\mathbf{b}_0\Big\} = E_{\{\boldsymbol{\mathcal{H}}_\ell\}}\Big\{ \sum_{\ell=0}^{+\infty} \mathbf{r}^T(\mathbf{A}_\ell)\mathbf{b}_\ell\Big|\mathbf{b}_0\Big\}, \tag{6.34}$$

where $\{\boldsymbol{\mathcal{H}}_{\ell-1}\}$ is the whole history until frame $\ell - 1$, and the vector $\mathbf{r}(\mathbf{A}) \triangleq [R(\mathbf{S}_1, \mathbf{A}) \;\; R(\mathbf{S}_2, \mathbf{A}) \;\; \cdots \;\; R(\mathbf{S}_{Q^{UM}}, \mathbf{A})]^T$ is the corresponding reward vector for taking action $\mathbf{A}$. Considering $\mathcal{H}_{\ell-1}$ as a realization of $\boldsymbol{\mathcal{H}}_{\ell-1}$, we define the function $\mathbf{r}^T(\boldsymbol{\mathsf{A}}_\ell)\mathbf{b}_\ell$ as the expected immediate reward function at the $\ell$-th frame.

Considering the fact that a POMDP problem is a continuous belief state MDP (see [55]), we can use the value iteration algorithm to find the optimal policy. However, the high computational complexity of the large-scale JASUS problem, prohibits us to use such an algorithm. Indeed, the computational complexity of the value iteration algorithm is $\mathcal{O}(|\mathcal{S}|^2 \times |\mathcal{A}|)$ per iteration [69], which is computationally intractable for large state space and action space. Thus, a suboptimal policy which is computationally affordable, is more attractive for the massive JASUS problem. In the next section, we propose to use myopic policy due to its low computational complexity and show that under certain conditions, the myopic policy provides the optimal solution for our defined POMDP-based JASUS problem.

## 6.4 Myopic Policy for Two-state Channel

In this section, we consider that the elements of states are quantized to two levels (i.e., $Q = 2$) denoted as $\alpha$ and $\beta$ to present good and bad state, respectively. Assuming $\alpha > \beta$, we write

$$\tilde{h}_{ui,j} \in \{\alpha, \beta\},$$
$$\text{for } j = 1,\ 2,\ \ldots,\ 2^{UM}, i = 1,\ 2,\ \ldots,\ M, \text{ and for } u = 1,\ 2,\ \ldots,\ U. \tag{6.35}$$

Here, we assume that the channel state is modeled as a two-state Markov chain with bad (0) and good (1) states over frames, as shown in Fig. 4.2, where $p_{ij}$ is the probability of state transition from $i$ to $j$, for $i, j \in \{0, 1\}$. The two-state Markov chain model is called *positively correlated*, if $p_{11} > p_{01}$, meaning that $\tilde{h}_{ui,j}$ with higher probability remains in good state compared to changing from bad state to good state.

Considering a two-level quantization for $\tilde{h}_{ui,j}$, we redefine the state space as $\mathcal{C} = \{\mathbf{C}_j\}_{j=1}^{2^{UM}}$, where the $j$-th element of the state space $\mathbf{C}_j$ is a $U \times M$ matrix, such that $\mathbf{C}_j = [\mathbf{c}_{1,j} \quad \mathbf{c}_{2,j} \quad \dots \quad \mathbf{c}_{M,j}]$, where $\mathbf{c}_{i,j} = [c_{1i,j} \quad c_{2i,j} \quad \dots \quad c_{Ui,j}]^T$, for $i = 1, 2, \dots, M$. Thus, for $j = 1, 2, \dots 2^{UM}$ , $i = 1, 2, \dots, M$, and $u = 1, 2, \dots, U$, we can obtain $c_{ui,j}$ as

$$c_{ui,j} = \begin{cases} 1 & \text{if } \tilde{h}_{ui,j} = \alpha \\ 0 & \text{if } \tilde{h}_{ui,j} = \beta \end{cases}. \tag{6.36}$$

Furthermore, here we denote $\mathsf{c}_{ui,l}$ as the random channel state of the link between the $i$-th BS antenna and the $u$-th user in the $\ell$-th frame. Using the redefined state space, we can simplify the belief update formulation. To do so, first we define the conditional probability of $|\mathsf{h}_{ui,\ell}|^2$ (the square absolute value of the channel between the $i$-th BS antenna and the $u$-th user at frame $\ell$) being in the good state, given the history up to frame $\ell - 1$ as $\omega_{ui,\ell} \triangleq \Pr(\mathsf{c}_{ui,\ell} = 1|\mathcal{H}_{\ell-1})$, for $i = 1, 2, \dots, M$, and $u = 1, 2, \dots, U$. Thus, at the $\ell$-th frame, the belief matrix can be defined as

$$\mathbf{\Omega}_\ell = [\omega_{ui,\ell}]_{\substack{i=1,2,\dots,M \\ u=1,2,\dots,U}}.$$

More specifically, since channels are independent, and as for each $\mathbf{\mathsf{S}}_\ell$ there is a corresponding $\mathbf{\mathsf{C}}_\ell$, at the $\ell$-th frame, the $j$-th element of the belief vector can be obtained as

$$\begin{aligned} b_{j,\ell} = \Pr(\mathbf{\mathsf{C}}_\ell = \mathbf{C}_j|\mathcal{H}_{\ell-1}) &\triangleq \prod_{u=1}^{U}\prod_{i=1}^{M} \Pr(\mathsf{c}_{ui,\ell} = c_{ui,j}|\mathcal{H}_{\ell-1}) \\ &= \prod_{u=1}^{U}\prod_{i=1}^{M} \omega_{ui,\ell}^{\mathsf{c}_{ui,j}} (1 - \omega_{ui,\ell})^{1-\mathsf{c}_{ui,j}}. \end{aligned} \tag{6.37}$$

Thus, to update the belief vector, we can update the probability of channel being in good state, $\omega_{ui,\ell}$. According to our proposed user scheduling restrictions, all users will be scheduled in only one time slot in a frame. Therefore, at each time slot $t$, only the CSI of the scheduled users and the selected antennas at that time slot can be observed. At frame $\ell$, for given action $\mathbf{A}_\ell$, the antenna selection vector $\mathbf{a}'_{t,\ell} = [a'_{it,\ell}]_{i=1}^M$ is the first $M$ rows of the $t$-th column of $\mathbf{A}_\ell$, while the user scheduling vector $\mathbf{a}''_{t,\ell} = [a''_{ut,\ell}]_{i=1}^U$ is the last $U$ rows of the $t$-th column of $\mathbf{A}_\ell$, for $t = 1, 2, \ldots, \lambda$. At time slot $t$, in the $\ell$-th frame, the $i$-th antenna and $u$-th user are selected to participate in data transmission, if $a'_{it,\ell} = 1$, and $a''_{ut,\ell} = 1$, respectively; otherwise $a'_{it,\ell} = 0$, and $a''_{ut,\ell} = 0$, for $i = 1, 2, \ldots, M$, and $u = 1, 2, \ldots, U$. ?

Based on our action matrix $\mathbf{A}_\ell$ and observed CSI corresponding to that action, for the next time frame $\ell + 1$, we can update the probability of being in good state, $\omega_{ui,\ell+1}$ according to Algorithm 5. At each frame $\ell$, and at each time slot $t$, observe $|\mathsf{h}_{ui,\ell}|^2$, if the $i$-th antenna is selected (i.e., if $a'_{it,\ell} = 1$, for $i = 1, 2, \ldots, M$) and $u$-th user is scheduled to be served (i.e., if $a''_{ut,\ell} = 1$, for $u = 1, 2, \ldots, U$) at that time slot. If the observed state is good (i.e., $c_{ui,\ell} = 1$), update as $\omega_{ui,\ell+1} = p_{11}$, and if the observed state is bad (i.e., $c_{ui,\ell} = 0$), update as $\omega_{ui,\ell+1} = p_{01}$. The probability of being in good state of the remaining channel links (non-observed channel links) can be updated as $\omega_{ui,\ell+1} = \omega_{ui,\ell} p_{11} + (1 - \omega_{ui,\ell}) p_{01}$.

**Algorithm 5** Updating the probability of being in a good state

```
for $t = 1 : \lambda$ do
    for $u = 1 : U$ do
        for $i = 1 : M$ do
            $$\omega_{ui,\ell+1} = \begin{cases} p_{11} & \text{if } a'_{it,\ell} = 1,\ a''_{ut,\ell} = 1, \text{ and } \ c_{ui,\ell} = 1; \\ p_{01} & \text{if } a'_{it,\ell} = 1,\ a''_{ut,\ell} = 1, \text{ and } c_{ui,\ell} = 0; \\ \omega_{ui,\ell} p_{11} + (1 - \omega_{ui,\ell}) p_{01} & \text{Otherwise} \end{cases} \qquad (6.38)$$
        end for
    end for
end for
```

We now use the redefined state space and belief vector to rewrite the expected immediate reward function. To do so, considering that $\Pr(\mathsf{c}_{ui,\ell} = 1 | \mathcal{H}_{\ell-1}) = \omega_{ui,\ell}$ and $\Pr(\mathsf{c}_{ui,\ell} = 0 | \mathcal{H}_{\ell-1}) = 1 - \omega_{ui,\ell}$. we define $\hat{f}(\omega, c) = \omega^c (1 - \omega)^{1-c}$, and rewrite

(6.37) as

$$b_{j,\ell} = \prod_{u=1}^{U} \prod_{i=1}^{M} \hat{f}(\omega_{ui,\ell}, c_{ui,j}). \tag{6.39}$$

Thus, using (6.39), we can obtain the elements of the belief vector $\mathbf{b}_\ell$ based on the values of $\mathbf{\Omega}_\ell = [\omega_{ui,\ell}]_{\substack{i=1,2,\ldots,M \\ u=1,2,\ldots,U}}$, and for given action $\mathbf{A}_\ell$ at frame $\ell$, we can rewrite the expected immediate reward function $\mathbf{r}^T(\mathbf{A}_\ell)\mathbf{b}_\ell$ as

$$\begin{aligned} &\bar{R}(\mathbf{A}_\ell, \mathbf{\Omega}_\ell) \triangleq \mathbf{r}^T(\mathbf{A}_\ell)\mathbf{b}_\ell = \sum_{j=1}^{|\mathcal{S}|} R(\mathbf{S}_j, \mathbf{A}_\ell) b_{j,t} = \sum_{j=1}^{|\mathcal{C}|} R(\mathbf{S}_j, \mathbf{A}_\ell) \Pr(\mathbf{C}_\ell = \mathbf{C}_j | \mathcal{H}_{\ell-1}) \\ &= \sum_{j=1}^{|\mathcal{C}|} \frac{1}{\lambda} \sum_{t=1}^{\lambda} |\mathcal{K}(\mathbf{a}''_{t,\ell})| \log_2 \Big(1 + \frac{P}{\sigma^2 |\mathcal{K}(\mathbf{a}''_{t,\ell})|^2} \sum_{i \in \mathcal{I}(\mathbf{a}'_{t,\ell})} \sum_{u \in \mathcal{K}(\mathbf{a}''_{t,\ell})} \tilde{h}_{ui,j} \Big) \prod_{u=1}^{U} \prod_{i=1}^{M} \hat{f}(\omega_{ui,\ell}, c_{ui,j}). \end{aligned} \tag{6.40}$$

where $\mathbf{a}'_{t,\ell}$ is the first $M$ rows of the $t$-th column of $\mathbf{A}_\ell$, while $\mathbf{a}''_{t,\ell}$ is the last $U$ rows of the $t$-th column of $\mathbf{A}_\ell$. Also, note that here $\tilde{h}_{ui,j} \in \{\alpha, \beta\}$. Let us define $K_{t,\ell} \triangleq |\mathcal{K}(\mathbf{a}''_{t,\ell})|$ as the total number of users scheduled at the $t$-th time slot in the $\ell$-th frame. For the $j$-th possible state matrix, and for given $K_{t,\ell}$, at time slot $t$ in the $\ell$-th frame, only $(N \times K_{t,\ell})$ elements of the state matrix corresponding to the $N$ selected antennas and $K_{t,\ell}$ scheduled users, can be observed. Considering $\tilde{h}_{ui,j} \in \{\alpha, \beta\}$, at the $t$-th time slot of the $\ell$-th frame, we denote $Z'_{jt,\ell}$ as the total number of elements in the set $\{\tilde{h}_{ui,j}, i \in \mathcal{I}(\mathbf{a}'_{t,\ell}), u \in \mathcal{K}(\mathbf{a}''_{t,\ell})\}$ that are equal to $\alpha$, for $j = 1, 2, \ldots, |\mathcal{C}|$. Thus, the remaining $(NK_{t,\ell}) - Z'_{jt,\ell}$ elements of this set are equal to $\beta$. We can then write

$$\begin{aligned} &|\mathcal{K}(\mathbf{a}''_{t,\ell})| \log_2 \Big(1 + \frac{P}{\sigma^2 |\mathcal{K}(\mathbf{a}''_{t,\ell})|^2} \sum_{i \in \mathcal{I}(\mathbf{a}'_{t,\ell})} \sum_{u \in \mathcal{K}(\mathbf{a}''_{t,\ell})} \tilde{h}_{ui,j} \Big) = \\ &\underbrace{K_{t,\ell} \log_2 \Big( \; Z'_{jt,\ell}(1 + \frac{P\alpha}{\sigma^2 K_{t,\ell}^2}) + ((NK_{t,\ell}) - Z'_{jt,\ell})(1 + \frac{P\beta}{\sigma^2 K_{t,\ell}^2}) \Big)}_{\triangleq R_{Z'_{jt,\ell}}}. \end{aligned} \tag{6.41}$$

Note that for given antenna selection vector $\mathbf{a}'_{t,\ell}$, and user scheduling vector $\mathbf{a}''_{t,\ell}$, at time slot $t$ in the $\ell$-th frame, for any possible state matrix in the state space that have same value of $Z'_{jt,\ell}$, their obtained $R_{Z'_{jt,\ell}}$ values are equal. Thus, for given antenna selection vector $\mathbf{a}'_{t,\ell}$, and user scheduling vector $\mathbf{a}''_{t,\ell}$, we can partition the state space $\mathcal{C}$ as

$$\mathcal{C} = \bigcup_{z=0}^{NK_{t,\ell}} \mathcal{C}_z(\mathbf{a}''_{t,\ell}, \mathbf{a}'_{t,\ell}), \tag{6.42}$$

where $\mathcal{C}_z(\mathbf{a}''_{t,\ell}, \mathbf{a}'_{t,\ell}) = \{\mathbf{C} = [\mathbf{c}_1 \;\; \mathbf{c}_2 \;\; \cdots \;\; \mathbf{c}_M] \in \mathcal{C} \mid \|\text{diag}(\mathbf{a}''_{t,\ell})\ \mathbf{C}\ \text{diag}(\mathbf{a}'_{t,\ell})\|^2 = z\}$. We now use (6.42) to rewrite (6.40) as

$$\begin{aligned}
\bar{R}(\mathbf{A}_\ell, \mathbf{\Omega}_\ell) &= \frac{1}{\lambda}\sum_{t=1}^{\lambda}\sum_{z=0}^{NK_{t,\ell}} \sum_{\mathbf{C}\in\mathcal{C}_z(\mathbf{a}''_{t,\ell},\mathbf{a}'_{t,\ell})} R_z \prod_{u=1}^{U}\prod_{i=1}^{M} \hat{f}(\omega_{ui,\ell}, c_{ui}) \\
&= \frac{1}{\lambda}\sum_{t=1}^{\lambda}\sum_{z=0}^{NK_{t,\ell}} R_z \sum_{\mathbf{C}\in\mathcal{C}_z(\mathbf{a}''_{t,\ell},\mathbf{a}'_{t,\ell})} \prod_{u=1}^{U}\prod_{i=1}^{M} \hat{f}(\omega_{ui,\ell}, c_{ui}) \\
&= \frac{1}{\lambda}\sum_{t=1}^{\lambda}\sum_{z=0}^{NK_{t,\ell}} R_z \sum_{\mathbf{C}\in\mathcal{C}_z(\mathbf{a}''_{t,\ell},\mathbf{a}'_{t,\ell})} \prod_{u\in\mathcal{K}(\mathbf{a}''_{t,\ell})}\prod_{i\in\mathcal{I}(\mathbf{a}'_{t,\ell})} \hat{f}(\omega_{ui,\ell}, c_{ui}) \prod_{u\in\mathcal{K}^{\perp}(\mathbf{a}''_{t,\ell})}\prod_{i\in\mathcal{I}^{\perp}(\mathbf{a}'_{t,\ell})} \hat{f}(\omega_{ui,\ell}, c_{ui}),
\end{aligned} \tag{6.43}$$

where, $R_z$ can be obtained from (6.41). Also, $\mathcal{I}^{\perp}(\mathbf{a}'_{t,\ell})$ and $\mathcal{K}^{\perp}(\mathbf{a}''_{t,\ell})$ are the set of the indices of the non-selected antennas and non-scheduled users at time slot $t$ in the $\ell$-th frame. Since $\mathcal{I}^{\perp}(\mathbf{a}'_{t,\ell})$ is the complement set of $\mathcal{I}(\mathbf{a}'_{t,\ell})$, we can write $|\mathcal{I}(\mathbf{a}'_{t,\ell})| = N$ and $|\mathcal{I}^{\perp}(\mathbf{a}'_{t,\ell})| = M - N$. Also, for the same reason we can write $|\mathcal{K}(\mathbf{a}''_{t,\ell})| = K_{t,\ell}$ and $|\mathcal{K}^{\perp}(\mathbf{a}''_{t,\ell})| = U - K_{t,\ell}$. Here, we can split $\mathbf{C} \in \mathcal{C}_z(\mathbf{a}''_{t,\ell}, \mathbf{a}'_{t,\ell})$ into two sub-matrixes $\mathbf{C}' = [c'_{ui}]_{\substack{i=1,2,\ldots,M\\u=1,2,\ldots,U}}$ and $\mathbf{C}'' = [c''_{ui}]_{\substack{i=1,2,\ldots,M\\u=1,2,\ldots,U}}$, such that

$$\begin{aligned}
c'_{ui} &= \begin{cases} c_{ui} & \text{if } i \in \mathcal{I}(\mathbf{a}'_{t,\ell}) \ \&\ u \in \mathcal{K}(\mathbf{a}''_{t,\ell}) \\ 0 & \text{Otherwise} \end{cases}, \quad \text{and} \\
c''_{ui} &= \begin{cases} c_{ui} & \text{if } i \in \mathcal{I}^{\perp}(\mathbf{a}'_{t,\ell}) \ \&\ u \in \mathcal{K}^{\perp}(\mathbf{a}''_{t,\ell}) \\ 0 & \text{Otherwise} \end{cases}
\end{aligned} \tag{6.44}$$

Note that the entries of $\mathbf{C}''$ can be either 0 or 1, while $\mathbf{C}' \in \mathcal{C}'_z(\mathbf{a}''_{t,\ell}, \mathbf{a}'_{t,\ell}) \triangleq \{\mathbf{C}' : \|\mathbf{C}'\|^2 = z\}$. Therefore, we can rewrite (6.43) as

$$\bar{R}(\mathbf{A}_\ell, \mathbf{\Omega}_\ell) = \frac{1}{\lambda}\sum_{t=1}^{\lambda}\sum_{z=0}^{NK_{t,\ell}} R_z \times$$

$$\sum_{\mathbf{C}'\in\mathcal{C}'_z(\mathbf{a}''_{t,\ell},\mathbf{a}'_{t,\ell})} \sum_{\mathbf{C}''} \prod_{u\in\mathcal{K}(\mathbf{a}''_{t,\ell})} \prod_{i\in\mathcal{I}(\mathbf{a}'_{t,\ell})} \hat{f}(\omega_{ui,\ell}, c'_{ui}) \prod_{u\in\mathcal{K}^{\perp}(\mathbf{a}''_{t,\ell})} \prod_{i\in\mathcal{I}^{\perp}(\mathbf{a}'_{t,\ell})} \hat{f}(\omega_{ui,\ell}, c''_{ui}),$$

$$= \frac{1}{\lambda}\sum_{t=1}^{\lambda}\sum_{z=0}^{NK_{t,\ell}} R_z \left( \sum_{\mathbf{C}'\in\mathcal{C}'_z(\mathbf{a}''_{t,\ell},\mathbf{a}'_{t,\ell})} \prod_{u\in\mathcal{K}(\mathbf{a}''_{t,\ell})} \prod_{i\in\mathcal{I}(\mathbf{a}'_{t,\ell})} \hat{f}(\omega_{ui,\ell}, c'_{ui}) \right)$$

$$\underbrace{\left( \sum_{\mathbf{C}''} \prod_{u\in\mathcal{K}^{\perp}(\mathbf{a}''_{t,\ell})} \prod_{i\in\mathcal{I}^{\perp}(\mathbf{a}'_{t,\ell})} \hat{f}(\omega_{ui,\ell}, c''_{ui}) \right)}_{=1}$$

$$= \frac{1}{\lambda}\sum_{t=1}^{\lambda}\sum_{z=0}^{NK_{t,\ell}} R_z \sum_{\mathbf{C}'\in\mathcal{C}'_z(\mathbf{a}''_{t,\ell},\mathbf{a}'_{t,\ell})} \prod_{u\in\mathcal{K}(\mathbf{a}''_{t,\ell})} \prod_{i\in\mathcal{I}(\mathbf{a}'_{t,\ell})} \hat{f}(\omega_{ui,\ell}, c'_{ui}), \tag{6.45}$$

where the second parentheses is the summation of the probabilities of all possible values that the elements of $\mathbf{C}''$ may take and thus it is equal to 1.

Let us denote $\tilde{R}(\mathbf{a}''_{t,\ell}, \mathbf{a}'_{t,\ell}, \mathbf{\Omega}_\ell)$ as the obtained expected immediate reward function at the $t$-th time slot in the $\ell$-th frame, when taking actions $\mathbf{a}'_{t,\ell}$ and $\mathbf{a}''_{t,\ell}$. Considering that channels remain unchanged during the entire frame, we can write

$$\tilde{R}(\mathbf{a}''_{t,\ell}, \mathbf{a}'_{t,\ell}, \mathbf{\Omega}_\ell) = \sum_{z=0}^{NK_{t,\ell}} R_z \sum_{\mathbf{C}'\in\mathcal{C}'_z(\mathbf{a}''_{t,\ell},\mathbf{a}'_{t,\ell})} \prod_{u\in\mathcal{K}(\mathbf{a}''_{t,\ell})} \prod_{i\in\mathcal{I}(\mathbf{a}'_{t,\ell})} \hat{f}(\omega_{ui,\ell}, c'_{ui}). \tag{6.46}$$

Note that

$$\bar{R}(\mathbf{A}_\ell, \mathbf{\Omega}_\ell) = \frac{1}{\lambda}\sum_{t=1}^{\lambda} \tilde{R}(\mathbf{a}''_{t,\ell}, \mathbf{a}'_{t,\ell}, \mathbf{\Omega}_\ell) \tag{6.47}$$

As we showed in (6.45), at time slot $t$, the expected immediate reward function depends only on $(\{\omega_{ui,\ell}\}, i \in \mathcal{I}(\mathbf{a}'_{t,\ell}), u \in \mathcal{K}(\mathbf{a}''_{t,\ell}))$. We define

$$f(\mathbf{x}) \triangleq \sum_{z=0}^{NK_{t,\ell}} R_z \sum_{\|\mathbf{C}'\|^2=z} \prod_{i=1}^{N}\prod_{u=1}^{K_{t,\ell}} \hat{f}(x_{ui}, c'_{ui}). \tag{6.48}$$

Then, for given vector $\mathbf{x} = \text{vec}(\{\omega_{ui,\ell}\}, i \in \mathcal{I}(\mathbf{a}'_{t,\ell}), u \in \mathcal{K}(\mathbf{a}''_{t,\ell}))$, where $\text{vec}(\{\cdot\})$ generates a vector with entries of the given set elements, we can write (6.46) as $f(\text{vec}(\{\omega_{ui,\ell}\}, i \in \mathcal{I}(\mathbf{a}'_{t,\ell}), u \in \mathcal{K}(\mathbf{a}''_{t,\ell}))) = \tilde{R}(\mathbf{a}''_{t,\ell}, \mathbf{a}'_{t,\ell}, \mathbf{\Omega}_\ell)$. Here, we aim to show that $f(\mathbf{x})$ in (6.48) is a regular function with respect to $\mathbf{x} = \text{vec}(\{\omega_{ui,\ell}\}, i \in \mathcal{I}(\mathbf{a}'_{t,\ell}), u \in \mathcal{K}(\mathbf{a}''_{t,\ell}))$. We define a regular function in the below definition.

**Definition 2.** *$f(\mathbf{x})$ is a regular function with respect to $\mathbf{x} = [x_1\ x_2\ \cdots\ x_{KN}]^T$, if $f(\mathbf{x})$ satisfies structural properties **C1**, **C2**, and **C3** which are described in Definition 1 in Section. 4.3.*

According to the regular function definition, the function $f(\mathbf{x})$ is regular if its symmetric, decomposable and monotone. The authors in [41] showed that, *for positively correlated two-state model, if the expected immediate reward function is a regular function, myopic policy is the optimal solution for the POMDP-based problem.* More specifically, a policy which maximizes the expected immediate reward function, results in maximizing the expected long-term reward function as well.

**Lemma 2.** *Function $f(\mathbf{x})$ in* (6.48) *is regular.*

*Proof.* We show that **C1**, **C2**, and **C3** described in Definition. 1 hold true for $f(\mathbf{x})$ in (6.48). To do so, we first start to prove that $f(\mathbf{x})$ is symmetric and satisfies **C1**, and for any $j, l$ if $f([x_1\ \cdots\ x_j\ \cdots\ x_l\ \cdots\ x_{KN}]^T) - f([x_1\ \cdots\ x_l\ \cdots x_j\ \cdots\ x_{KN}]^T) = 0$, $f(\mathbf{x})$ is symmetric. Thus, we write

$$
\begin{aligned}
&f([x_{11}\ \cdots\ x_{pj}\ \cdots\ x_{ql}\ \cdots\ x_{KN}]^T) - f([x_{11}\ \cdots\ x_{ql}\ \cdots\ \ x_{pj}\ \cdots\ x_{KN}]^T)\\
&= \sum_{z=0}^{NK_{t,\ell}} R_{Z_{t,\ell}} \sum_{\|\mathbf{C}'\|^2=z} \hat{f}(x_{pj}, c'_{pj})\hat{f}(x_{ql}, c'_{ql}) \prod_{\substack{i=1\\ i\neq j,l}}^{N} \prod_{\substack{u=1\\ u\neq q,p}}^{K_t} \hat{f}(x_{ui}, c'_{ui}) -\\
&\sum_{z=0}^{NK_{t,\ell}} R_z \sum_{\|\mathbf{C}'\|^2=z} \hat{f}(x_{ql}, c'_{pj})\hat{f}(x_{pj}, c'_{ql}) \prod_{\substack{i=1\\ i\neq j,l}}^{N} \prod_{\substack{u=1\\ u\neq p,q}}^{K_{t,\ell}} \hat{f}(x_{ui}, c'_{ui})\\
&= \sum_{z=0}^{NK_{t,\ell}} R_z \sum_{\|\mathbf{C}'\|^2=z} \underbrace{\Big(\prod_{\substack{i=1\\ i\neq j,l}}^{N} \prod_{\substack{u=1\\ u\neq p,q}}^{K_{t,\ell}} \hat{f}(x_{ui}, c'_{ui})\Big)\big(\hat{f}(x_{pj}, c'_{pj})\hat{f}(x_{ql}, c'_{ql}) - \hat{f}(x_{ql}, c'_{pj})\hat{f}(x_{pj}, c'_{ql})\big)}_{\triangleq l(\mathbf{C}')},
\end{aligned}
\tag{6.49}
$$

where

$$
l(\mathbf{C}') = \tag{6.50}
$$

$$\begin{cases} \big(x_{pj}x_{ql} - x_{ql}x_{pj}\big) \displaystyle\prod_{\substack{i=1\\ i\neq j,l}}^{N} \prod_{\substack{u=1\\ u\neq p,q}}^{K_{t,\ell}} \hat{f}(x_{ui}, c'_{ui}) = 0, & \text{if } c'_{pj} = 1,\ c'_{ql} = 1, \\ \underbrace{\big(x_{pj}(1 - x_{ql}) - x_{ql}(1 - x_{pj})\big)}_{x_{pj}-x_{ql}} \displaystyle\prod_{\substack{i=1\\ i\neq j,l}}^{N} \prod_{\substack{u=1\\ u\neq p,q}}^{K_{t,\ell}} \hat{f}(x_{ui}, c'_{ui}), & \text{if } c'_{pj} = 1,\ c'_{ql} = 0, \\ \big((1 - x_{pj})(1 - x_{ql}) - (1 - x_{ql})(1 - x_{pj})\big) \displaystyle\prod_{\substack{i=1\\ i\neq j,l}}^{N} \prod_{\substack{u=1\\ u\neq p,q}}^{K_{t,\ell}} \hat{f}(x_{ui}, c'_{ui}) = 0, & \text{if } c'_{pj} = 0,\ c'_{ql} = 0, \\ \underbrace{\big((1 - x_{pj})x_{ql} - (1 - x_{ql})x_{pj}\big)}_{x_{ql}-x_{pj}} \displaystyle\prod_{\substack{i=1\\ i\neq j,l}}^{N} \prod_{\substack{u=1\\ u\neq p,q}}^{K_{t,\ell}} \hat{f}(x_{ui}, c'_{ui}), & \text{if } c'_{pj} = 0,\ c'_{ql} = 1. \end{cases} \tag{6.51}$$

Using the fact that we define $Z_{t,\ell}$ as the number of channels to be in good state, we can write if $\mathbf{1}^T[c'_{11} \ \cdots \ c'_{ql} \ \cdots \ c'_{pj} \ \cdots \ c'_{KN}]^T = z$, then we can say that $\mathbf{1}^T[c'_{11} \ \cdots \ c'_{pj} \ \cdots \ c'_{ql} \ \cdots \ c'_{KN}]^T = z$. Meaning that $\sum_{\|\mathbf{C}'\|^2=z} l(\mathbf{C}') = 0$. Thus, (6.49) is equal to zero, and $f(\mathbf{x})$ is a symmetric function. We now show that **C2** property holds true for $f(\mathbf{x})$, and thus, $f(\mathbf{x})$ is decomposable. To do so, we rewrite (6.48) as

$$f(\mathbf{x}) = \sum_{z=0}^{NK_{t,\ell}} R_z \sum_{\|\mathbf{C}'\|^2=z} R_{\|\mathbf{C}'\|^2} \underbrace{\Big( \prod_{\substack{i=1\\ i\neq j}}^{N} \prod_{\substack{u=1\\ u\neq p}}^{K_{t,\ell}} \hat{f}(x_{ui}, c'_{ui}) \Big)}_{\triangleq Q(\mathbf{x}_{-pj}, \mathbf{C}'_{-pj})} \hat{f}(x_{pj}, c'_{pj}), \tag{6.52}$$

where $\mathbf{C}'_{-pj}$ is the same matrix as $\mathbf{C}'$ where the element $c'_{pj}$, is removed. Also, with similar definition, $\mathbf{x}_{-pj}$ is the same vector $\mathbf{x}$ where the element $x_{pj}$ is removed. Note that $R_{\|\mathbf{C}'\|^2}$ is equivalent to $R_z$. We can rewrite (6.52) as

$$\begin{aligned} f(\mathbf{x}) = &\sum_{z=0}^{NK_{t,\ell}-1} \sum_{\substack{\|\mathbf{C}'\|^2=z\\ c'_{pj}=0}} R_{\|\mathbf{C}'\|^2} Q(\mathbf{x}_{-pj}, \mathbf{C}'_{-pj})(1 - x_{pj}) + \\ &\sum_{z=1}^{NK_{t,\ell}} \sum_{\substack{\|\mathbf{C}'\|^2=z\\ c'_{pj}=1}} R_{\|\mathbf{C}'\|^2} Q(\mathbf{x}_{-pj}, \mathbf{C}'_{-pj}) x_{pj}. \end{aligned} \tag{6.53}$$

Since $\|\mathbf{C}'\|^2 = \|\mathbf{C}'_{-pj}\|^2 + c'_{pj}$, we can now rewrite (6.53) as

$$f(\mathbf{x}) = \sum_{z=0}^{NK_{t,\ell}-1} \sum_{\substack{\|\mathbf{C}'\|^2=z \\ c'_{pj}=0}} R_{\|\mathbf{C}'_{-pj}\|_1} Q(\mathbf{x}_{-pj}, \mathbf{C}'_{-pj})(1 - x_{pj}) +$$

$$\sum_{z=0}^{NK_{t,\ell}-1} \sum_{\substack{\|\mathbf{C}'\|^2=z+1 \\ c'_{pj}=1}} R_{1+\|\mathbf{C}'_{-pj}\|^2} Q(\mathbf{x}_{-pj}, \mathbf{C}'_{-pj}) x_{pj}.$$

$$= \sum_{z=0}^{NK_{t,\ell}-1} \Bigg( \sum_{\|\mathbf{C}'_{-pj}\|^2=z} R_{1+\|\mathbf{C}'_{-pj}\|^2} Q(\mathbf{x}_{-pj}, \mathbf{C}'_{-pj}) - \sum_{\|\mathbf{C}'_{-pj}\|^2=z} R_{\|\mathbf{C}'_{-pj}\|^2} Q(\mathbf{x}_{-pj}, \mathbf{C}'_{-pj}) \Bigg) x_{pj}$$

$$+ \sum_{z=0}^{NK_{t,\ell}-1} \sum_{\|\mathbf{C}'_{-pj}\|^2=z} R_{\|\mathbf{C}'_{-pj}\|^2} Q(\mathbf{x}_{-pj}, \mathbf{C}'_{-pj}) =$$

$$\sum_{z=0}^{NK_{t,\ell}-1} \Bigg( \sum_{\|\mathbf{C}'_{-pj}\|^2=z} Q(\mathbf{x}_{-pj}, \mathbf{C}'_{-pj})(R_{1+\|\mathbf{C}'_{-pj}\|_1} - R_{\|\mathbf{C}'_{-pj}\|_1}) \Bigg) x_{pj} +$$

$$\sum_{Z_{t,\ell}=0}^{NK_{t,\ell}-1} \sum_{\|\mathbf{C}'_{-pj}\|^2=z} R_{\|\mathbf{C}'_{-pj}\|^2} Q(\mathbf{x}_{-pj}, \mathbf{C}'_{-pj}) = \eta_{pj}\ x_{pj} + \theta_{pj}, \tag{6.54}$$

where $\eta_{pj} \triangleq \sum_{z=0}^{NK_{t,\ell}-1} \Bigg( \sum_{\|\mathbf{C}'_{-pj}\|^2=z} Q(\mathbf{x}_{-pj}, \mathbf{C}'_{-pj})(R_{1+\|\mathbf{C}'_{-pj}\|^2} - R_{\|\mathbf{C}'_{-pj}\|^2}) \Bigg)$ and $\theta_{pj} \triangleq \sum_{z=0}^{NK_{t,\ell}-1} \sum_{\|\mathbf{C}'_{-pj}\|^2=z} R_{\|\mathbf{C}'_{-pj}\|^2} Q(\mathbf{x}_{-pj}, \mathbf{C}'_{-pj})$ holds true. Thus, we can write

$$\begin{aligned} f(\mathbf{x}) &= x_{pj}(\eta_{pj}\ + \theta_{pj}) + (1 - x_{pj})\theta_{pj} \\ &= x_{pj} f([x_{11}\ \cdots\ x_{pj-1}\ 1\ \cdots\ x_{NK}]^T) + \\ &(1 - x_{ju'}) f([x_{11}\ \cdots\ x_{pj-1}\ 0\ \cdots\ x_{NK}]^T). \end{aligned} \tag{6.55}$$

Thus, according to (6.55) and **C2**, $f(\mathbf{x})$ is a decomposable function. Using the decomposable property of the $f(\mathbf{x})$, we show that $f(\mathbf{x})$ is a monotone function as it holds true for **C3**. To do so, using (6.54), we can write

$$f(\mathbf{x}) - f(\mathbf{x}') = \eta_{pj}(x_{pj} - x'_{pj}) \tag{6.56}$$

Note that since $R_{1+\|\mathbf{C}'_{-pj}\|^2} > R_{\|\mathbf{C}'_{-pj}\|^2}$, it is true that in (6.54), $\eta_{pj} > 0$. Therefore, for $x_{pj} > x'_{pj}$, we can write

$$f(\mathbf{x}) - f(\mathbf{x}') = \eta_{pj}(x_{pj} - x'_{pj}) > 0 \tag{6.57}$$

Thus, $f(\mathbf{x})$ is monotone and the proof is complete. Based on the above discussions, $f(\mathbf{x})$ holds true for all the three properties, **C1**, **C2**, and **C3** in Definition. 1, and thus $f(\mathbf{x})$ is a regular function. ■

We proved that for given $K_{t,\ell} = |\mathcal{K}(\mathbf{a}''_{t,\ell})|$ , the function $\tilde{R}(\mathbf{a}''_{t,\ell}, \mathbf{a}'_{t,\ell}, \mathbf{\Omega}_\ell)$ in (6.46) which is the $t$-th time slot expected immediate reward function, is a regular function. However, to show that myopic policy is optimal, we need to prove that the entire frame expected immediate reward function defined in (6.47) is a regular function.

**Lemma 3.** *The expected immediate reward function in* (6.47) *is regular.*

*Proof.* For given number of users scheduled at each time slot $t$, (i.e., $|\mathcal{K}(\mathbf{a}''_{t,\ell})| = K_{t,\ell}$, for $t = 1, 2, \dots, \lambda$), we showed that $\tilde{R}(\mathbf{a}''_{t,\ell}, \mathbf{a}'_{t,\ell}, \mathbf{\Omega}_\ell)$ is a regular function with respect to $\text{vec}(\{\omega_{ui,\ell}\}, i \in \mathcal{I}(\mathbf{a}'_{t,\ell}), u \in \mathcal{K}(\mathbf{a}''_{t,\ell}))$. More specifically, at each time slot $t$, the expected immediate reward function only depends on the probability of being in good state of the channel links between the selected antennas and the users scheduled at that time slot (i.e., $\{\omega_{ui,\ell}\}, i \in \mathcal{I}(\mathbf{a}'_{t,\ell}), u \in \mathcal{K}(\mathbf{a}''_{t,\ell})$). Thus, for given number of users at each time slot, according to (6.47), maximizing individual time slots' expected immediate reward function, is equivalent to maximizing the entire frame expected immediate reward function. Proof is complete. ■

We use the following theorem to further simplify the myopic policy for our JASUS problem.

**Theorem 5.** *For given number of scheduled users at each time slot $t$ (i.e., $K_{t,\ell}$), under the assumption of positively correlated two-state channel model, the myopic policy is optimal and amounts to select $NK_{t,\ell}$ channels with highest probability of* *being in good state (i.e., $\{\omega_{ui,\ell}\}, i \in \mathcal{I}(\mathbf{a}'_{t,\ell}), u \in \mathcal{K}(\mathbf{a}''_{t,\ell})$). More specifically, the indices of $N$ antennas and the indices of $K_{t,\ell}$ users that have largest entries of $\mathbf{\Omega}_\ell$, are the indices of selected antennas and users scheduled at time slot $t$, in the $\ell$-th frame.*

*Proof.* In Lemma 2, we showed that the $t$-th time slot expected immediate reward function $\tilde{R}(\mathbf{a}''_{t,\ell}, \mathbf{a}'_{t,\ell}, \mathbf{\Omega}_\ell)$, is a monotone function with respect to $\text{vec}(\{\omega_{ui,\ell}\}, i \in \mathcal{I}(\mathbf{a}'_{t,\ell}), u \in \mathcal{K}(\mathbf{a}''_{t,\ell}))$. Meaning that $\tilde{R}(\mathbf{a}''_{t,\ell}, \mathbf{a}'_{t,\ell}, \mathbf{\Omega}_\ell)$ is monotonically increasing with $(\{\omega_{ui,\ell}\}, i \in \mathcal{I}(\mathbf{a}'_{t,\ell}), u \in \mathcal{K}(\mathbf{a}''_{t,\ell}))$. Thus, we can maximize $\tilde{R}(\mathbf{a}''_{t,\ell}, \mathbf{a}'_{t,\ell}, \mathbf{\Omega}_\ell)$ with selecting $NK_{t,\ell}$ channels with highest probability of being in good state (i.e., highest $(\{\omega_{ui,\ell}\}, i \in \mathcal{I}(\mathbf{a}'_{t,\ell}), u \in \mathcal{K}(\mathbf{a}''_{t,\ell}))$ values). ■

For given $K_{t,\ell}$, we denote the optimal set of $N$ antenna indices that are selected at time slot $t$ of frame $l$ as $\mathcal{I}^*_{t,\ell}$, and represent the optimal set of $K_{t,\ell}$ user indices that are scheduled at time slot $t$ of frame $l$ as $\mathcal{K}^*_{t,\ell}$, for $t = 1, 2, \ldots, \lambda$ and $\ell = 1, 2, \ldots, \infty$. Noting that $|\mathcal{K}^*_{t,\ell}| = K_{t,\ell}$ and $\mathcal{I}^*_{t,\ell} = N$ hold true, we can then write, for any $i' \notin \mathcal{I}^*_{t,\ell}$ or $u' \notin \mathcal{K}^*_{t,\ell}$,

$$\omega_{u'i',\ell} \leq \text{minimum}\{\omega_{ui,\ell}, i \in \mathcal{I}^*_{t,\ell}, u \in \mathcal{K}^*_{t,\ell}\} \tag{6.58}$$

holds true, for $t = 1, 2, \ldots, \lambda$ and $\ell = 1, 2, \ldots, \infty$.

So far, we have shown that for given number of users (i.e., $K_{t,\ell}$) at each time slot $t$ in a frame, for positively correlated two-state channel model, myopic policy is the optimal solution, meaning that it maximizes the expected long-term reward function. However, finding the optimal number of users scheduled at each time slot $t$ amounts to solving another optimization problem, as explain in the sequel. Here, we denote a $\lambda \times 1$ vector that contains the number of users scheduled at each time slot in the $\ell$-th frame as $\mathbf{k}_\ell = [K_{1,\ell} \;\; K_{2,\ell} \;\; \cdots \;\; K_{\lambda,\ell}]^T$, whose $t$-th element is equal to the total number of users scheduled at time slot $t$ in the $\ell$-th frame, for $t = 1, 2, \ldots, \lambda$ and $\ell = 1, 2, \ldots, \infty$. Note that at the beginning of each frame, the BS schedules all available $U$ users to $\lambda$ time slots such that the number of users per time slot should be more than one and less than the number of RF chains. Let us denote $\mathcal{U}$ as the set of all possible $\mathbf{k}_\ell$ vectors, when $U$ users are available to be scheduled in $\lambda$ time slots in a frame. We can write

$$\mathcal{U} = \Big\{ \mathbf{k} = [k_1 \;\; k_2 \;\; \cdots \;\; k_\lambda]^T \;\Big|\; \mathbf{1}^T\mathbf{k} = U \;, \text{ while } 1 \leq k_t \leq N \;, \text{for } t = 1, 2, \ldots, \lambda \Big\} \,. \tag{6.59}$$

Here, we denote $\mathbf{k}_\ell^*$ as the optimal vector that contains the optimal number of users scheduled at time slot $t$ in the $\ell$-th frame. Note that for any possible $\mathbf{k}_\ell \in \mathcal{U}$, at each time slot $t$, based on $K_{t,\ell}$, we can obtain $\mathcal{I}_{t,\ell}^*$, and $\mathcal{K}_{t,\ell}^*$ according to (6.58) to find the corresponding action vectors $\mathbf{a}'_{t,\ell} = [a'_{it,\ell}]_{i=1}^M$ and $\mathbf{a}''_{t,\ell} = [a''_{ut,\ell}]_{u=1}^U$, such that for $t = 1, 2, \ldots, \lambda$,

$$a'_{it,\ell} = \begin{cases} 1 & \text{if } i \in \mathcal{I}_{t,\ell}^* \\ 0 & \text{if } i \notin \mathcal{I}_{t,\ell}^* \end{cases} , \tag{6.60}$$

$$a''_{ut,\ell} = \begin{cases} 1 & \text{if } u \in \mathcal{K}_{t,\ell}^* \\ 0 & \text{if } u \notin \mathcal{K}_{t,\ell}^* \end{cases} . \tag{6.61}$$

where $a'_{it,\ell}$ is the $i$-th element of $\mathbf{a}'_{t,\ell}$ and $a''_{ut,\ell}$ is the $u$-th element of $\mathbf{a}''_{t,\ell}$. Note that $a'_{it,\ell} = 1$, means that $i$-th antenna is selected to serve the users scheduled at time slot $t$, and $a''_{ut,\ell} = 1$, means that $u$-th user is scheduled at the $t$-th time slot in the $\ell$-th frame to be served. Otherwise, $a'_{it,\ell} = 0$, and $a''_{ut,\ell} = 0$. For give vector $\mathbf{k}_\ell = [K_{1,\ell} \;\; K_{2,\ell} \;\; \cdots \;\; K_{\lambda,\ell}]^T \in \mathcal{U}$, the corresponding action matrix $\mathbf{A}_\ell = \begin{bmatrix} \mathbf{A}'_\ell \\ \mathbf{A}''_\ell \end{bmatrix}$, with $\mathbf{A}'_\ell = [\mathbf{a}'_{t,\ell}]_{t=1}^\lambda$ and $\mathbf{A}''_\ell = [\mathbf{a}''_{t,\ell}]_{t=1}^\lambda$ can be determined from (6.58), (6.60), and (6.61). Then the optimal vector $\mathbf{k}_\ell^*$ can be obtained as

$$\mathbf{k}_\ell^* = \arg \max_{\forall \mathbf{k}_\ell \in \mathcal{U}} \Big( \bar{R}(\mathbf{A}_\ell, \mathbf{\Omega}_\ell) \Big), \tag{6.62}$$

where for obtained $\mathbf{A}_\ell$, we can obtain the expected immediate reward function $\bar{R}(\mathbf{A}_\ell, \mathbf{\Omega}_\ell)$ form (6.47).

To summarize, we proved in this section that for given optimal $\mathbf{k}_\ell^*$, and for positively correlated two-state channel model, myopic policy provides the optimal solution to our JASUS problem. In the next section, motivated by the optimality of the myopic policy, we aim to design a JASUS algorithm that can be applied to Rayleigh fading channels.

## 6.5 Gauss-Markov Model for Rayleigh Fading Channels

Here, we aim to devise a myopic policy-based algorithm for our JASUS problem that can be implemented for Rayleigh fading channels. To do so, we assume that channels evolve according to the first-order Gauss-Markov channel model. We use $\mathbf{\Sigma}_h = E\{\mathsf{h}_{i,t}\mathsf{h}_{i,t}^H\} = \mathrm{diag}([\sigma_{\mathrm{h},u}^2]_{u=1}^U)$ to denote the diagonal channel covariance matrix, where $\sigma_{\mathrm{h},u}^2$ large-scale variation of channel between the $u$-th user and the $i$-th antenna. We then write the channel vector $\mathsf{h}_{i,t} \sim \mathcal{CN}(\mathbf{0}_{U\times 1}, \mathbf{\Sigma}_h)$ as

$$\mathsf{h}_{i,t} \triangleq \mathrm{diag}(\boldsymbol{\xi})\mathsf{h}_{i,t-1} + \mathrm{diag}(\boldsymbol{\xi}')\mathsf{z}_{i,t}, \qquad i = 1, ..., M\,. \tag{6.63}$$

where $\mathsf{z}_{i,t} \sim \mathcal{CN}(\mathbf{0}_{U\times 1}, \mathbf{\Sigma}_h)$ is the i.i.d. innovation sequence which is independent of the channel vector $\mathsf{h}_{i,t}$, for $i = 1, 2, \ldots, M$, $\boldsymbol{\xi} = [\xi_1 \;\; \xi_2 \;\; \cdots \;\; \xi_U]^T$ is the fading correlation vector, with $\xi_u \in [0, 1]$ being the fading correlation coefficient corresponding to the $u$-th user, and $\boldsymbol{\xi}' = [\sqrt{1-\xi_1^2} \;\; \sqrt{1-\xi_2^2} \;\; \cdots \;\; \sqrt{1-\xi_U^2}]^T$. Note that we can obtain the value of $\xi_u$ according to the maximum Doppler frequency [61].

We aim to quantize the square absolute value of channel coefficients (i.e., $|\mathsf{h}_{ui,\ell}|^2$, for $i = 1, 2, \ldots, M$, and $u = 1, 2, \ldots, U$) to two levels, good (1) and bad (0), only in the selection stage, to benefit from the optimality of myopic policy for positively correlated two-state channel model, such that

$$\mathsf{c}_{ui,\ell} = \begin{cases} 1, & \text{if } |\mathsf{h}_{ui,\ell}|^2 \geqslant v, \\ 0, & \text{if } |\mathsf{h}_{ui,\ell}|^2 < v\ , \end{cases} \tag{6.64}$$

where $v$ is the quantization threshold value. Using (6.64) to obtain a two-state channel model, we propose Algorithm. 6, thereby applying the myopic policy for our JASUS problem. According to our proposed myopic policy algorithm, based on action $\mathbf{A}_{\ell-1}$, at the end of frame $\ell-1$, the observation matrix of the entire frame $\mathsf{O}_{\ell-1}$ can be obtained. Note that $\mathsf{O}_{\ell-1} = \sum_{t=1}^{\lambda} \bar{\mathsf{O}}_{t,\ell-1}$, where the $t$-th time slot observation matrix $\bar{\mathsf{O}}_{t,\ell-1}$ can be obtained at the end of the uplink transmission of that time slot. At the beginning of the $\ell$-th frame, given $\mathsf{O}_{\ell-1}$, we propose to use (6.64) to quantize non-zero elements of matrix $\mathsf{O}_{\ell-1}$ (i.e., the observed channel links between selected antennas and users scheduled at each time slot $t = 1, 2, \ldots, \lambda$ ). We then update the elements of the belief matrix $\mathbf{\Omega}_{\ell}$ of the next frame using Algorithm 5. Next for any possible $\mathbf{k}_{\ell} \in \mathcal{U}$, based on $K_{t,\ell}$ (i.e., the $t$-th element of $\mathbf{k}_{\ell}$), we can obtain $\mathcal{I}^{*}_{t,\ell}$, and $\mathcal{K}^{*}_{t,\ell}$ according to (6.58), to find the corresponding action vectors, $\mathbf{a}'_{t,\ell}$ and $\mathbf{a}''_{t,\ell}$ based on (6.60) and (6.61), respectively, for $t = 1, 2, \ldots, \lambda$. Thus, for any possible $\mathbf{k}_{\ell} \in \mathcal{U}$, there is a corresponding action matrix $\mathbf{A}_{\ell}$ obtained from (6.58), (6.60), and (6.61). We use $\Upsilon_{\ell}(\cdot)$ to denote a function that maps $\mathbf{k}_{\ell}$ to its corresponding action matrix $\mathbf{A}_{\ell}$ at the $\ell$-th frame, such that $\mathbf{A}_{\ell} = \Upsilon_{\ell}(\mathbf{k}_{\ell})$. We then use (6.62) to find $\mathbf{k}^{*}_{\ell}$, and select its corresponding action matrix $\mathbf{A}_{\ell} = \Upsilon_{\ell}(\mathbf{k}^{*}_{\ell})$, as the $\ell$-th frame JASUS action matrix.

The computational complexity of our proposed myopic-based JASUS algorithm resides in updating the elements of the belief matrix with the computational complexity $\mathcal{O}(UM-UN)$ (see Algorithm.5), and then, in finding the optimal number of users scheduled at each time slot and its corresponding set of selected antenna indices and scheduled users indices with computational complexity $\mathcal{O}(UM\lambda \log UN)$ [70]. Since in our defined system model we assume that the BS is equipped with massive number of antennas ($M$ is a large number), the computational complexity of the myopic policy-based JASUS algorithm is $\mathcal{O}(UM\lambda \log UN)$, which is significantly lower than the computational complexity of the value iteration algorithm of $\mathcal{O}(|\mathcal{S}|^{2} \times |\mathcal{A}|)$ per iteration [69].

**Algorithm 6** The myopic policy based antenna selection

```
Initialization: Obtain $\mathcal{U}$ from (6.59). Given the channel correlation factor vector $\boldsymbol{\xi}$ and $\mathbf{\Sigma}_h$, set the threshold value $v$.
At each frame $\ell$:
Input: $\mathbf{O}_{\ell-1}$.
Quantize the non-zero elements of $\mathbf{O}_{\ell-1}$ according to (6.64).
Update the elements of $\mathbf{\Omega}_\ell$ using Algorithm.5.
for any possible $\mathbf{k}_\ell \in \mathcal{U}$ do
    for $t = 1 : \lambda$ do
        Obtain $\mathcal{I}^*_{t,\ell}, \mathcal{K}^*_{t,\ell}$, according to $K_{t,\ell}$, using (6.58).
        Obtain the elements of $\mathbf{a}'_{t,\ell}$ and $\mathbf{a}''_{t,\ell}$ from (6.60) and (6.61), respectively.
    end for
    Obtain the corresponding action matrix $\mathbf{A}_\ell = \begin{bmatrix} \mathbf{A'}_\ell \\ \mathbf{A''}_\ell \end{bmatrix}$.
    Save $\mathbf{k}_\ell$ and its corresponding obtained $\mathbf{A}_\ell$ in a mapping table, $\mathbf{A}_\ell = \Upsilon_\ell(\mathbf{k}_\ell)$.
end for
Obtain $\mathbf{k}^*_\ell$ from (6.62).
Output: $\mathbf{A}_\ell = \Upsilon_\ell(\mathbf{k}^*_\ell)$.
```

## 6.6 Simulation Results

In this section, considering that channels evolve according to a first-order Gauss Markov model presented in Section 6.5, we aim to evaluate the performance of our proposed myopic policy-based JASUS presented as Algorithm 6, for a multi-user massive MIMO system. We evaluate the performance of our proposed algorithm by using the non-quantized channel coefficients to obtain time-average sum-rate up to time frame $\ell$, denoted as $\hat{R}_\ell$, given by

$$\hat{R}_\ell = \frac{1}{\ell\lambda}\sum_{\tau=0}^{\ell}\sum_{t=1}^{t=\lambda}\log_2\Big(\det\big(\mathbf{I}+\frac{P}{\sigma^2\|\hat{\mathbf{H}}^H_{\mathrm{t},\tau}(\hat{\mathbf{H}}_{\mathrm{t},\tau}\ \hat{\mathbf{H}}^H_{\mathrm{t},\tau})^{-1}\|_F^2}\mathbf{I}\big)\Big). \tag{6.65}$$

where $\hat{\mathbf{H}}_{\mathrm{t},\ell}$ is the channel matrix between the $N$ selected antennas and $K_{t,\ell}$ users scheduled at time slot $t$ in the $\ell$-th frame. Here, we compare our results with two other polices namely, a random selection policy and a full CSI-based policy. In the random selection policy, we randomly schedule each of the $U$ available users in time slot $t$, for $t = 1, 2, \ldots, \lambda$, and select $N$ antennas randomly to transmit data at each time slot $t$ in a frame. In the full CSI-based policy, considering that at each frame $\ell$, full CSI is available, an exhaustive search is carried out to find the best subset of users to schedule in time slot $t$, for $t = 1, 2, \ldots, \lambda$, and the best subset of antennas for data transmission. Note that the presented results are the mean of $\hat{R}_\ell$ over 100 Monte Carlo runs. Furthermore, for finding the optimal threshold value (denoted as $v^*$) for channel quantization in Algorithm 6, we use a low-complexity search algorithm, proposed in Section. 5.6.

### 6.6.1 Evaluating the Performance of Algorithm 6:

In the first part of our simulations, to evaluate the performance Algorithm 6, we define five scenarios, where in each one of them, the available users have different speeds (i.e., different values of $\xi_u$ and are located at different distances from the BS (i.e., different values of $\sigma^2_{h,u}$). More specifically, we use Jakes' model presented in [74] to obtain the fading correlation coefficient of the $u$-th user according to its speed, for $u = 1, 2, \ldots, U$, as we explain in the sequel. Considering a WLAN 802.11 system which is operating at the carrier frequency $f_c = 2.4$ GHz, according to the Jakes' model, we can obtain $\xi_u = J_0(2\pi \frac{V_u f_c}{C f_W})$, where $V_u$ is the speed of the $u$-th user, $C = 3 \times 10^9$ m/s is speed of wave propagation, and $f_W = 2.5$ KHz is the communication bandwidth. For example, for a pedestrian user (i.e., $V_u = 3.6$ km/h ), a user with typical driving speed in residential areas (i.e., $V_u = 36$ km/h), and a high speed user, such as user in a car driving on a highway (i.e., $V_k = 140$ km/h), the obtained fading correlation coefficient are $\xi_u = 0.999$ $\xi_u = 0.986$, and $\xi_u = 0.95$, respectively. Furthermore, denoting the $u$-th user distance from the BS as $d_u$, we use the simple path loss model $\sigma^2_{h,u} = \varrho d_u^{-3}$ to define the SNR range for each user, where the path loss constant $\varrho$ is chosen such that for the $u$-th user at the cell boundary (i.e., for $d_u = 500$m), the value of $\frac{P\sigma^2_{h,u}}{\sigma^2}$ is 0 dB. Given the above explanations we now define five scenarios as listed below.

- Scenario i, low-speed and low-SNR users: This scenario involves 12 users, each of which has a fading correlation coefficient uniformly distributed in the interval $[0.996, 0.999]$ Also, the users are located at random distances from the BS such that $\{\frac{P\sigma^2_{h,u}}{\sigma^2}\}_{u=1}^{12}$ are in the range of $[0, 0.5]$ dB.

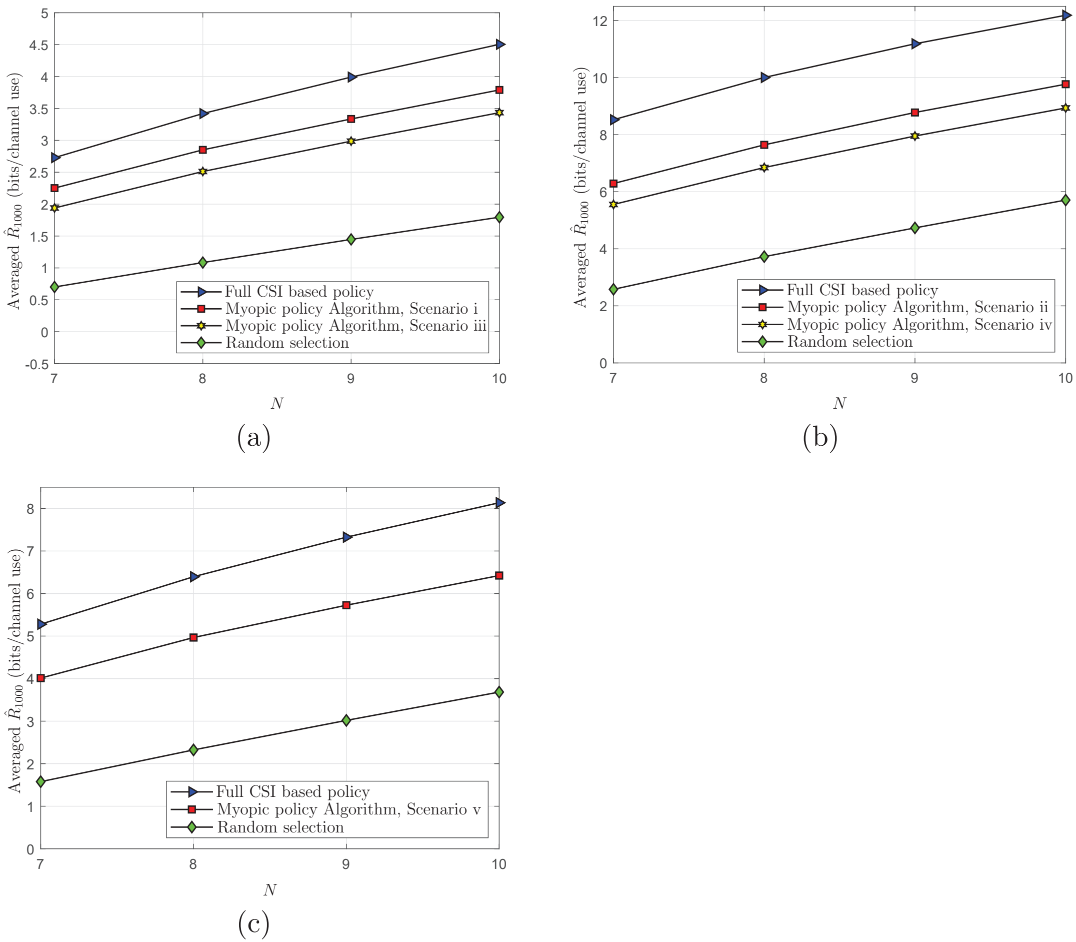
- Scenario ii, low-speed and high-SNR users: This scenario involves 12 users, each of which has a fading correlation coefficient uniformly distributed in the interval $[0.996, 0.999]$. Also, the users are located at random distances from the BS such that $\{\frac{P\sigma^2_{h,u}}{\sigma^2}\}_{u=1}^{12}$ are in the range of $[9.5, 10]$ dB.

- Scenario iii, high-speed and low-SNR users: This scenario involves 12 users, each of which has a fading correlation coefficient uniformly distributed in the interval $[0.95, 0.96]$. Also, the users are located at random distances from the BS such that $\{\frac{P\sigma^2_{h,u}}{\sigma^2}\}_{u=1}^{12}$ are in the range of $[0, 0.5]$ dB.

- Scenario iv, high-speed and high-SNR users: This scenario involves 12 users, each of which has a fading correlation coefficient uniformly distributed in the interval $[0.95, 0.96]$. Also, the users are located at random distances from the BS such that $\{\frac{P\sigma_{h,u}^2}{\sigma^2}\}_{u=1}^{12}$ are in the range of $[9.5, 10]$ dB.

- Scenario v, random speed and random SNR users: This scenario involves 12 users, each of which has a fading correlation coefficient uniformly distributed in the interval $[0.95, 0.999]$. Also, the users are located at random distances from the BS such that $\{\frac{P\sigma_{h,u}^2}{\sigma^2}\}_{u=1}^{12}$ are in the range of $[0, 10]$ dB.

Considering $\lambda = 2$ time slots per super-fame, and assuming that the BS is equipped with $M = 126$ antennas to serve $U = 12$ single-antenna users, in Figs. 6.3, we plot $\hat{R}_{1000}$ versus different number of RF chains $N = [7, 8,\ 9,\ 10]$, for aforementioned Scenarios i, ii, iii, iv, and v. We plot Figs. 6.3a, 6.3b, and 6.3c, to show $\hat{R}_{1000}$ versus different $N$ for low-SNR range (Scenarios i, and iii), high-SNR range (Scenarios ii, and iv) and random SNR range (Scenario v), respectively. As can be seen from Figs. 6.3a, and 6.3b, for low-speed users, the performance gap between the full CSI-based policy and the myopic policy for $N = 7$ are less than 0.5 and 2 bit per channel use (bcu), respectively. However, for the high-speed users, the performance gap is larger. For example, for $N = 7$, Figs. 6.3a, and 6.3b show that the performance of our proposed Algorithm 6 is about 0.8 (bcu) and 2.5 (bcu) lower that that of the full CSI-based policy, respectively. Since with increasing the speed of users, the value of $p_{01}$ increases (and hence the probability of switching channel state increases), increasing this performance gap for high speed users is expected. More specifically, higher value of $p_{01}$ results in less possibility for searching among channel links between unselected antennas and non-scheduled users for the subsequent time frames (see Algorithm 5). This in turns results in lower performance compared to scenarios with low-speed users. Finally in Fig. 6.3c, we show the performance of the more realistic Scenario v, which involves users with different speed ranges that be located at any distance from the BS in a cell. As can be seen from this figure, the performance gap between the myopic policy-based selection and random selection policy is about 1.3 (bcu) for $N = 7$, and 1.4 (bcu) for $N = 10$.



(a)

(b)

(c)

Figure 6.3: Time averaged sum-rate $\hat{R}_{1000}$ vs $N$ for $U = 12$, $M = 128$, and $N = [7 : 1 : 10]$ (a) Scenarios i, and iii (b) Scenarios ii, and iv, and (c) Scenario v.

### 6.6.2 The Impact of Increasing the Number of Users on the Performance of Algorithm 6:

We now aim to analyze the performance of our proposed Algorithm 6 for fixed values of $M$ and $N$, but for different number of available users $U$. To do so, below we define four different scenarios:

- Scenario vi: This scenario involves $U$ low-speed users (pedestrians users with $V_k = 3.6$ km/h ) with the fading correlation coefficient $\xi_u = 0.999$, for $u =$

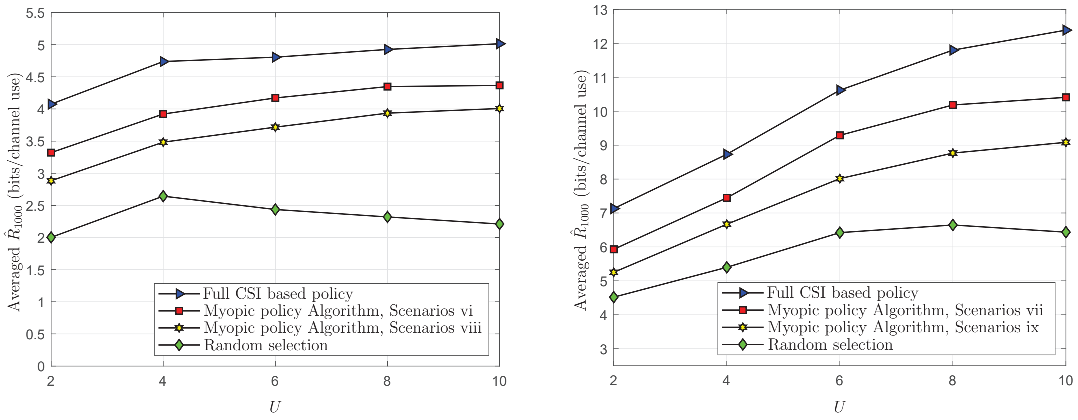
$1, 2, \ldots, U$. Also, the users are located at random distances from the BS such that $\{\frac{P\sigma^2_{h,u}}{\sigma^2}\}_{u=1}^U$ are in the range of $[0, 0.5]$ dB.

- Scenario vii: This scenario involves $U$ low-speed users (pedestrians with $V_k = 3.6$ km/h) with the fading correlation coefficient $\xi_u = 0.999$, for $u = 1, 2, \ldots, U$. Also, the users are located at random distances from the BS such that $\{\frac{P\sigma^2_{h,u}}{\sigma^2}\}_{u=1}^U$ are in the range of $[9.5, 10]$ dB.

- Scenario viii: This scenario involves $U$ high-speed users (users in a car driving on a highway with $V_k = 140$ km/h) with the fading correlation coefficient $\xi_u = 0.95$, for $u = 1, 2, \ldots, U$. Also, the users are located at random distances from the BS such that $\{\frac{P\sigma^2_{h,u}}{\sigma^2}\}_{u=1}^U$ are in the range of $[0, 0.5]$ dB.

- Scenario ix: This scenario involves $U$ high-speed users (users in cars driving on a highway with $V_k = 140$ km/h), with the fading correlation coefficient $\xi_u = 0.95$ for $u = 1, 2, \ldots, U$. Also, the users are located at random distances from the BS such that $\{\frac{P\sigma^2_{h,u}}{\sigma^2}\}_{u=1}^U$ are in the range of $[9.5, 10]$ dB.

Here, assuming that the BS is equipped with $M = 128$ antennas and $N = 10$ RF chains, and that each frame consists of two time slots ($\lambda = 2$), we plot the average of $\hat{R}_{1000}$ over 100 Monte Carlo runs, versus different number of users $U$ in Fig. 6.4. Fig. 6.4a presents the results of Scenarios vi, viii, and Fig 6.4b presents the results of Scenarios vii, and ix. As can be seen from these figures, and as we expect (see Section. 6.6.1), the performance gap between the myopic policy algorithm and the full CSI based policy is lower for low-speed users compared to that high-speed users in the same SNR range. For instance, in Fig 6.4a, for $U = 10$, this performance gap is less than 0.6 (bcu), and 1 (bcu), for low-speed users and high-speed users, respectively. One can also see in Figs 6.4a and 6.4b that with increasing number of users, the gap between the full CSI based policy and and the random selection increases significantly. However, with increasing the number of users, in these figure, the performance gap between the myopic policy algorithm and the full CSI policy approximately remains unchanged in Fig 6.4a, and only slightly increases in Fig 6.4b. For instance, in Fig 6.4b, for $U = 2$, the gap between the full CSI based policy and random selection is 2.5 (bcu) and for $U = 10$, this gap increases to 6 (bcu). In contrast, in Fig 6.4b, for $U = 2$, the performance gap between the full CSI policy and the myopic policy is 1 (bcu) and 1.7 (bcu) for low-speed and high-speed users, respectively, and for $U = 10$, this gap increases to about 1.6 (bcu) and 3 (bcu), respectively.



(a) (b)

Figure 6.4: Time averaged sum-rate $\hat{R}_{1000}$ vs $U$ for $M = 128$, $N = 10$ and $U = [2 : 2 : 10]$ (a) Scenarios vi, and viii, and (b) Scenarios vii, ix.

### 6.6.3 The Importance of User Scheduling

In this section, we aim to analyze the results of user scheduling performance in our JASUS problem. To do so, considering a BS equipped with $M = 100$ antennas and $N = 30$ RF chains, we plot the average of $\hat{R}_{1000}$ over 100 Monte Carlo runs for different number of users in Figs. 6.5a and 6.5b for Scenarios vi and viii, respectively, for two different cases: 1) each time frame only contains one time slot, i.e., $\lambda = 1$ (meaning that the BS serves all the users in one time slot), and 2) each time frame contains two time slots, i.e., $\lambda = 2$. The main goal of this comparison is to show the benefit of user scheduling when large number of users are available. As can be seen from Fig. 6.5, for $\lambda = 1$, with increasing the number of users, after $U = 6$, the sum-rate drops significantly. However, for $\lambda = 2$, after $U = 6$, the sum-rate increases with increasing the number of users. Thus, from Fig. 6.5, it appears that for different numbers of users, there is an optimal number of time slots per frame. Designing a time frame with optimal number of time slots is not in the scope of this dissertation, but can be considered as an extension of this line of work.

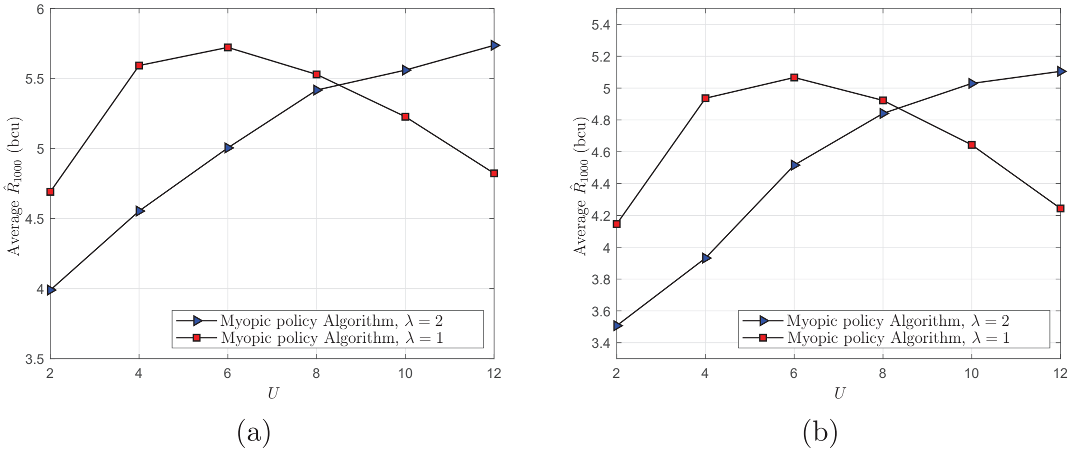


Figure 6.5: Time averaged sum-rate $\hat{R}_{1000}$ vs $U$ for $N = 30$, $M = 100$, and for different values of $U$(a) Scenario vi, and (b) Scenario viii.

# Chapter 7

# Conclusion and Future Work

In this chapter, we first provide the conclusion section, and then we outline the possible future work.

## 7.1 Conclusion

In Chapter 4 of this dissertation, we formulated the antenna selection problem at the BS, equipped with $M$ antennas and $N$ RF chains ($M \gg N$), for downlink transmissions using a POMDP framework. In a TDD system, we assumed the channel state evolves according to a finite-state Markov process and remains unchanged during each time slot, which consists of the uplink and downlink transmission (i.e., channel reciprocity holds). Given the partial CSI, to maximize the expected long-term downlink data rate, the value iteration algorithm can be used to extract the optimal policy. However, this algorithm has high computational complexity, and thus a simple myopic policy could offer an attractive alternative solution. We prove that, for a positively correlated two-state channel model, the myopic policy is optimal for selecting any $N$ out of $M$ antennas. Based on this result, for general fading channels, we proposed the channels be quantized into two levels and apply the myopic policy for antenna selection. Although in the antenna selection problem, only partial quantized CSI is available, our simulation results show that the performance of our proposed algorithm is within 0.5 (bcu) from the full CSI based policy (the upper bound data rate) for antenna selection.

In Chapter 5, we utilized a POMDP framework to formulate the antenna selection problem for a BS equipped with a massive number of antennas and a limited number of RF chains in a massive MU-MIMO system. Using ZF beamforming, we defined the sum-rate upper bound as the reward function and prove that for i.i.d positively correlated two-state channel model, the expected long-term reward function is a regular function. Thus, myopic policy provides the optimal solution to our antenna selection problem. Furthermore, we proposed a low-complexity antenna selection algorithm which can be implemented to Rayleigh fading channel model. According to our proposed algorithm, given an optimal threshold value, to benefit from the optimality of myopic policy for two-state channel model, we quantized the channels' gain into two levels only in the selection stage. To obtain the optimal threshold value for channel gain quantization, we proposed an efficient offline algorithm, which results in high achievable performance in our simulation results. Considering users with random speeds and SNR ranges, our results show that the proposed myopic policy algorithm is within 0.3 (bcu) from the full CSI policy which is the upper-bound in our simulation results.

In Chapter 6, we used a POMDP framework to formulate the joint antenna selection an user scheduling (JASUS) problem for a large-scale antenna BS with $M$ antennas and $N$ RF chains ($M \gg N$), that transmits data to $U$ single-antenna users in a cellular system. Here, we assumed that the number of users is larger than the number of RF chains ($U > N$), and we used zero-forcing (ZF) beamforming to eliminate inter-user interference. Thus, to fully cancel out the inter-user interference, the number of served users is limited as the number of RF chains at each time slot. To grantee that all users receive data, we assumed that users are served in a frame that contains of a finite number of time slots. At the beginning of each frame, the BS schedules users to different time slots in a frame and then selects a subset of antennas to serve the scheduled users at each time slot by performing downlink and uplink transmission. Note that the number of scheduled users is smaller than or equal to the number of RF chains. Here, we assumed that channels evolve according to a same Markov chain at the beginning of each frame and remain unchanged during the entire frame. We showed that for a positively correlated two-state channel model, the myopic policy provides the optimal solution to our JASUS problem. We then proposed a low-complexity JASUS algorithm that can be implemented to Rayleigh fading channels. Considering time-varying Rayleigh fading channels, our designed low-complexity JASUS algorithm can make a real-time decisions based on only available partial CSI.

## 7.2 Possible Future Work

This research can be extended in several directions as explained below.

- **Switching Cost in designing the Antenna Selection Algorithm**

  In this dissertaion, for designing the antenna selection algorithm, we assumed that the constraint is selecting N out of M antennas at each time slot. In the problem formulation, we can consider the cost of switching RF chains in the massive MIMO systems as another constraint when designing an antenna selection algorithm. More specifically, one can formulate the antenna selection problem for massive MU-MIMO systems as a POMDP framework with considering the switching cost as a constraint in the defined optimization problem.

- **Antenna Selection/JASUS in Multi-user Massive MIMO Systems When System Operates in FDD Mode**

  The analytical results in this dissertation are derived under the assumption that the perfect CSI is available. Assuming that the system operates in TDD mode, we can acquire CSI via traditional training procedures at the end of uplink transmission (using this assumption is a common practice). However, as future work, one can investigate the effect of the channel estimation error on the performance of our proposed POMDP-based algorithms for both antenna selection and JASUS. Note that, the same problem exists in FDD mode. As in FDD mode, different frequency bands are used for the downlink and uplink transmission, channel estimation is required (to obtain the partial CSI corresponding to the previously selected antennas' channel coefficient), before applying our proposed POMDP-based resource allocation algorithms.

- **Finding the Optimal Number of Time-slots per Frame in JASUS**

  In Chapter 6, we assumed that the number of time-slots in a frame is given. As can be seen in Fig. 6.5, finding the optimal number of time-slots per frame can improve the performance of our proposed JASUS algorithm. More specifically, as can be concluded from Fig. 6.5 (for the given scenario), when there are less than six number of users, one time-slot per frame results in higher time-averaged rate compared to two time-slots per frame. However, when there are more than six users, two time-slots per frame provides higher time-averaged rate compared to one time-slot per frame. Therefore, obtaining the optimal number of time-slots per frame for different scenarios can be an interesting problem for a future work.

- **Antenna Selection and JASUS in Cellular Systems**

  Considering a cellular system where at each cell there is a BS with its corresponding unique frequency band, the BS can use our proposed antenna selection and JASUS algorithm to provide a high quality of service for the available users. However, due to the limited amount of spectrum, for large areas (especially when cells are small), reusing the same frequency in adjacent cells could be a desired feature. Thus, a proper POMDP formulation is required to formulate the antennas selection/JASUS in cellular systems. In this case, due to the interference, the myopic policy may not provide the optimal solution. Therefore, another extension of this dissertation can be looking for other suboptimal POMDP solutions that provide an efficient and low-complexity antenna selection/JASUS algorithm for cellular systems.

# Appendix A

# Proof of Equation (3.7)

We are interested to present the sufficient statistic of the belief vector proof provided by [75] and demonstrate the updating belief vector formula. Therefore, by substitution (3.6) into (3.5), we can write

$$\begin{aligned}
b_{j,t} &= P_r\{\mathsf{\mathbf{s}}_t = \mathrm{s}_j | \mathsf{\mathbf{a}}_{t-1} = \mathbf{a}_{t-1}, \mathsf{\mathbf{o}}_t = \mathrm{o}_t, \mathcal{H}_{t-1}\} \\
&= \frac{P_r\{\mathsf{\mathbf{s}}_t = \mathrm{s}_j, \mathsf{\mathbf{o}}_t = \mathrm{o}_t | \mathsf{\mathbf{a}}_{t-1} = \mathbf{a}_{t-1}, \mathcal{H}_{t-1}\}}{P_r\{\mathsf{\mathbf{o}}_t = \mathrm{o}_t | \mathsf{\mathbf{a}}_{t-1} = \mathbf{a}_{t-1}, \mathcal{H}_{t-1}\}} \\
&= \sum_i P_r\{\mathsf{\mathbf{s}}_{t-1} = \mathrm{s}_i | \mathsf{\mathbf{a}}_{t-1} = \mathbf{a}_{t-1}, \mathcal{H}_{t-1}\} P_r\{\mathsf{\mathbf{s}}_t = \mathrm{s}_j | \mathsf{\mathbf{s}}_{t-1} = \mathrm{s}_i, \mathsf{\mathbf{a}}_{t-1} = \mathbf{a}_{t-1}, \mathcal{H}_{t-1}\} \times \\
&\quad P_r\{\mathsf{\mathbf{o}}_t = \mathrm{o}_t | \mathsf{\mathbf{s}}_t = \mathrm{s}_j, \mathsf{\mathbf{s}}_{t-1} = \mathrm{s}_i, \mathsf{\mathbf{a}}_{t-1} = \mathbf{a}_{t-1}, \mathcal{H}_{t-1}\} \Big/ P_r\{\mathsf{\mathbf{o}}_t = \mathrm{o}_t | \mathsf{\mathbf{a}}_{t-1} = \mathbf{a}_{t-1}, \mathcal{H}_{t-1}\}
\end{aligned} \tag{A.1}$$

where, the first probability in the numerator is independent of $\mathsf{\mathbf{a}}_{t-1}$, and thus we can write $P_r\{\mathsf{\mathbf{s}}_{t-1} = \mathrm{s}_i | \mathsf{\mathbf{a}}_{t-1} = \mathbf{a}_{t-1}, \mathcal{H}_{t-1}\} = b_{i,t-1}$. The second term in numerator is state transition probability, which is independent of action, and thus we can write $P_r\{\mathsf{\mathbf{s}}_t = \mathrm{s}_j | \mathsf{\mathbf{s}}_{t-1} = \mathrm{s}_i, \mathsf{\mathbf{a}}_{t-1} = \mathbf{a}_{t-1}, \mathcal{H}_{t-1}\} = P_r(\mathsf{\mathbf{s}}_t = \mathrm{s}_j | \mathsf{\mathbf{s}}_{t-1} = \mathrm{s}_i)$; and the third term in numerator is the observation probability at time slot $t$, which only depends on the $t$-th time slot state and the previous action $\mathbf{a}_{t-1}$. Note that the denominator in the equation is summed over all $j$. Hence, we can write

$$b_{j,t} = \frac{\sum_i b_{i,t-1} P_r(\mathsf{\mathbf{s}}_t = \mathrm{s}_j | \mathsf{\mathbf{s}}_{t-1} = \mathrm{s}_i) P_r(\mathsf{\mathbf{o}}_t = \mathrm{o}_t | \mathsf{\mathbf{s}}_t = \mathrm{s}_j, \mathsf{\mathbf{a}}_{t-1} = \mathbf{a}_{t-1})}{\sum_{i,j} b_{i,t-1} P_r(\mathsf{\mathbf{s}}_t = \mathrm{s}_j | \mathsf{\mathbf{s}}_{t-1} = \mathrm{s}_i) P_r(\mathsf{\mathbf{o}}_t = \mathrm{o}_t | \mathsf{\mathbf{s}}_t = \mathrm{s}_j, \mathsf{\mathbf{a}}_{t-1} = \mathbf{a}_{t-1})}. \tag{A.2}$$

Therefore, the updating belief vector can be written as $\mathbf{b}_t = \frac{\mathbf{O}(\mathbf{o}_t,\mathbf{a}_{t-1})\mathbf{T}\mathbf{b}_{t-1}}{\mathbf{1}^T\mathbf{O}(\mathbf{o}_t,\mathbf{a}_{t-1})\mathbf{T}\mathbf{b}_{t-1}}$. The proof is complete.

# Bibliography

[1] T. L. Marzetta, "Massive MIMO: An introduction," *Bell Labs Technical Journal*, vol. 20, pp. 11–22, 2015.

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

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

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

[5] A. Liu and V. Lau, "Phase only RF precoding for massive MIMO systems with limited RF chains," *IEEE Trans. Signal Process.*, vol. 62, no. 17, pp. 4505–4515, 2014.

[6] S. Sanayei and A. Nosratinia, "Antenna selection in MIMO systems," *IEEE Commun. Mag.*, vol. 42, pp. 68–73, Oct 2004.

[7] M. Gharavi-Alkhansari and A. B. Gershman, "Fast antenna subset selection in MIMO systems," *IEEE Trans. Signal Process.*, vol. 52, pp. 339–347, Feb 2004.

[8] A. F. Molisch and M. Z. Win, "MIMO systems with antenna selection," *IEEE Microw. Mag.*, vol. 5, pp. 46–56, March 2004.

[9] G. Brante, I. Stupia, R. D. Souza, and L. Vandendorpe, “Outage probability and energy efficiency of cooperative MIMO with antenna selection,” *IEEE Trans. Wireless Commun.*, vol. 12, pp. 5896–5907, Nov 2013.

[10] A. F. Molisch, M. Z. Win, Yang-Seok Choi, and J. H. Winters, “Capacity of MIMO systems with antenna selection,” *IEEE Trans. Wireless Commun.*, vol. 4, pp. 1759–1772, July 2005.

[11] A. Gorokhov, D. Gore, and A. Paulraj, “Receive antenna selection for MIMO spatial multiplexing: theory and algorithms,” *IEEE Trans. Signal Process.*, vol. 51, no. 11, pp. 2796–2807, Nov 2003.

[12] M. T. Kakitani, G. Brante, R. D. Souza, and M. A. Imran, “Energy efficiency of transmit diversity systems under a realistic power consumption model,” *IEEE Commun. Lett.*, vol. 17, no. 1, pp. 119–122, Jan 2013.

[13] R. W. Heath, S. Sandhu, and A. Paulraj, “Antenna selection for spatial multiplexing systems with linear receivers,” *IEEE Commun. Lett.*, vol. 5, pp. 142–144, April 2001.

[14] T.-W. Ban and B. C. Jung, “A practical antenna selection technique in multiuser massive MIMO networks,” *IEICE transactions on communications*, vol. 96, no. 11, pp. 2901–2905, Nov 2013.

[15] X. Gao, O. Edfors, F. Tufvesson, and E. G. Larsson, “Massive MIMO in real propagation environments: Do all antennas contribute equally?” *IEEE Trans. Commun.*, vol. 63, no. 11, pp. 3917–3928, Nov 2015.

[16] M. Hanif, H.-C. Yang, G. Boudreau, E. Sich, and H. Seyedmehdi, “Antenna subset selection for massive MIMO systems: A trace-based sequential approach for sum rate maximization,” *Journal of Communications and Networks*, vol. 20, no. 2, pp. 144–155, April 2018.

[17] J. Joung and S. Sun, "Two-step transmit antenna selection algorithms for massive MIMO," in *2016 IEEE International Conference on Communications (ICC)*, 2016, pp. 1–6.

[18] J. Chen, S. Chen, Y. Qi, and S. Fu, "Intelligent massive MIMO antenna selection using Monte Carlo tree search," *IEEE Trans. Signal Process.*, vol. 67, pp. 5380–5390, Oct 2019.

[19] R. Mndez-Rial, C. Rusu, N. Gonzlez-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, Jan 2016.

[20] M. Benmimoune, E. Driouch, W. Ajib, and D. Massicotte, "Joint transmit antenna selection and user scheduling for massive MIMO systems," in *2015 IEEE Wireless Communications and Networking Conference (WCNC)*, 2015, pp. 381–386.

[21] A. Salh, N. S. M. Shah, L. Audah, Q. Abdullah, W. A. Jabbar, and M. Mohamad, "Energy-efficient power allocation and joint user association in multiuser-downlink massive MIMO system," *IEEE Access*, vol. 8, pp. 1314–1326, 2020.

[22] G. Xu, A. Liu, W. Jiang, H. Xiang, and W. Luo, "Joint user scheduling and antenna selection in distributed massive MIMO systems with limited backhaul capacity," *IEEE China Commun.*, vol. 11, no. 5, pp. 17–30, 2014.

[23] S. Maimaiti, G. Chuai, W. Gao, K. Zhang, X. Liu, and Z. Si, "A low-complexity algorithm for the joint antenna selection and user scheduling in multi-cell multi-user downlink massive MIMO systems," *EURASIP Journal on Wireless Communications and Networking*, vol. 2019, no. 1, pp. 1–14, 2019.

[24] Y.-X. Zhu, D.-Y. Kim, and J.-W. Lee, “Joint antenna and user scheduling in the massive MIMO system over time-varying fading channels,” *IEEE Access*, vol. 9, pp. 92 431–92 445, 2021.

[25] B. H. Wang, H. T. Hui, and M. S. Leong, “Global and fast receiver antenna selection for MIMO systems,” *IEEE Trans. Commun.*, vol. 58, no. 9, pp. 2505–2510, 2010.

[26] C. Jiang and L. J. Cimini, “Antenna selection for energy-efficient MIMO transmission,” *IEEE Wireless Commun. Lett.*, vol. 1, no. 6, pp. 577–580, 2012.

[27] D. Gore, R. Nabar, and A. Paulraj, “Selecting an optimal set of transmit antennas for a low rank matrix channel,” in *2000 IEEE International Conference on Acoustics, Speech, and Signal Processing. Proceedings (Cat. No.00CH37100)*, vol. 5, 2000, pp. 2785–2788 vol.5.

[28] A. Garcia-Rodriguez, C. Masouros, and P. Rulikowski, “Reduced switching connectivity for large scale antenna selection,” *IEEE Trans. Commun.*, vol. 65, pp. 2250–2263, May 2017.

[29] Z. Kuai and S. Wang, “Thompson sampling-based antenna selection with partial CSI for TDD massive MIMO systems,” *IEEE Trans. Commun.*, vol. 68, no. 12, pp. 7533–7546, 2020.

[30] S. Padmanabhan, R. G. Stephen, C. R. Murthy, and M. Coupechoux, “Training-based antenna selection for PER minimization: A POMDP approach,” *IEEE Trans. Commun.*, vol. 63, pp. 3247–3260, Sept 2015.

[31] S. Sharifi, S. S. Panahi, and M. Dong, “A POMDP based antenna selection for massive MIMO communication,” *IEEE Trans. Commun.*, vol. 70, no. 3, pp. 2025–2041, March 2022.

[32] C. C. Tan and N. C. Beaulieu, "On first-order Markov modeling for the Rayleigh fading channel," *IEEE Trans. Commun.*, vol. 48, pp. 2032–2040, May 2000.

[33] H. S. Wang and P. PChang, "On verifying the first-order Markovian assumption for a Rayleigh fading channel model," *IEEE Trans. Veh. Technol.*, vol. 45, pp. 353–357, May 1996.

[34] Q. Zhang and S. A. Kassam, "Finite-state Markov model for rayleigh fading channels," *IEEE Trans. Commun.*, vol. 47, pp. 1688–1692, Nov 1999.

[35] P. Sadeghi, R. A. Kennedy, P. B. Rapajic, and R. Shams, "Finite-state Markov modeling of fading channels - a survey of principles and applications," *IEEE Signal Process. Mag.*, vol. 25, pp. 57–80, Sept 2008.

[36] S. Sun, M. Dong, and B. Liang, "On stochastic feedback control for multi-antenna beamforming: Formulation and low-complexity algorithms," *IEEE Trans. Wireless Commun.*, vol. 13, pp. 4731–4745, Sept 2014.

[37] Q. Zhao, B. Krishnamachari, and K. Liu, "On myopic sensing for multi-channel opportunistic access: structure, optimality, and performance," *IEEE Trans. Wireless Commun.*, vol. 7, pp. 5431–5440, Dec 2008.

[38] S. H. A. Ahmad, M. Liu, T. Javidi, Q. Zhao, and B. Krishnamachari, "Optimality of myopic sensing in multichannel opportunistic access," *IEEE Trans. Inf. Theory*, vol. 55, pp. 4040–4050, Sept 2009.

[39] Q. Zhao, L. Tong, A. Swami, and Y. Chen, "Decentralized cognitive MAC for opportunistic spectrum access in ad hoc networks: A POMDP framework," *IEEE J. Sel. Areas Commun.*, vol. 25, pp. 589–600, April 2007.

[40] A. Sahand Haji Ali and L. Mingyan, "Multi-channel opportunistic access: A case of restless bandits with multiple plays," in *2009 47th Annual Allerton*

*Conference on Communication, Control, and Computing (Allerton)*. IEEE, Oct 2009, pp. 1361–1368.

[41] K. Wang, Q. Liu, and L. Chen, “Optimality of greedy policy for a class of standard reward function of restless multi-armed bandit problem,” *IET Signal Processing*, vol. 6, pp. 584–593, Aug 2012.

[42] K. Wang, L. Chen, K. A. Agha, and Q. Liu, “On optimality of myopic policy in opportunistic spectrum access: The case of sensing multiple channels and accessing one channel,” *IEEE Wireless Commun. Lett.*, vol. 1, pp. 452–455, July 2012.

[43] Z. Shen, R. Chen, J. Andrews, R. Heath, and B. 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.

[44] J. Nam, A. Adhikary, J.-Y. Ahn, and G. Caire, “Joint spatial division and multiplexing: Opportunistic beamforming, user grouping and simplified downlink scheduling,” *IEEE J. Sel. Topics Signal Process.*, vol. 8, no. 5, pp. 876–890, 2014.

[45] Y. Xu, G. Yue, and S. Mao, “User grouping for massive MIMO in FDD systems: New design methods and analysis,” *IEEE Access*, vol. 2, pp. 947–959, 2014.

[46] E. Castaeda, A. Silva, A. Gameiro, and M. Kountouris, “An overview on resource allocation techniques for multi-user MIMO systems,” *IEEE Communications Surveys Tutorials*, vol. 19, no. 1, pp. 239–284, 2017.

[47] H. Liu, H. Gao, S. Yang, and T. Lv, “Low-complexity downlink user selection for massive MIMO systems,” *IEEE Syst. J.*, vol. 11, no. 2, pp. 1072–1083, 2017.

[48] G. Dimic and N. Sidiropoulos, “On downlink beamforming with greedy user selection: performance analysis and a simple new algorithm,” *IEEE Trans. Signal Process.*, vol. 53, no. 10, pp. 3857–3868, 2005.

[49] C. C. Bennett and K. Hauser, “Artificial intelligence framework for simulating clinical decision-making: A Markov decision process approach,” *Artificial intelligence in medicine*, vol. 57, no. 1, pp. 9–19, 2013.

[50] F. Doshi-Velez, J. Pineau, and N. Roy, “Reinforcement learning with limited reinforcement: Using bayes risk for active learning in POMDPs,” *Artificial Intelligence*, vol. 187, pp. 115–132, 2012.

[51] M. T. Spaan and N. Vlassis, “Perseus: Randomized point-based value iteration for POMDPs,” *Journal of artificial intelligence research*, vol. 24, pp. 195–220, 2005.

[52] L. P. Kaelbling, M. L. Littman, and A. R. Cassandra, “Planning and acting in partially observable stochastic domains,” *Artificial intelligence*, vol. 101, pp. 99–134, May 1998.

[53] C.-S. Chow and J. N. Tsitsiklis, “The complexity of dynamic programming,” *Journal of complexity*, vol. 5, no. 4, pp. 466–488, 1989.

[54] N. L. Zhang and W. Zhang, “Speeding up the convergence of value iteration in partially observable Markov decision processes,” *Journal of Artificial Intelligence Research*, vol. 14, pp. 29–51, 2001.

[55] V. Krishnamurthy, *Partially observed Markov decision processes*. Cambridge University Press, 2016.

[56] R. C. Manthony, *POMDP Solver Software*, http://www.pomdp.org/code/.

[57] W. Erwin, *SolvePOMDP*, https://www.erwinwalraven.nl/solvepomdp/.

[58] T. K. Y. Lo, “Maximum ratio transmission,” *IEEE Trans. Commun.*, vol. 47, no. 10, pp. 1458–1461, 1999.

[59] S. Sharifi, S. Shahbaz Panahi, and M. Dong, “Antenna selection for massive MIMO systems based on POMDP framework,” in *ICASSP 2021 - 2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2021, pp. 4450–4454.

[60] C. H. Papadimitriou and J. N. Tsitsiklis, “The complexity of Markov decision processes,” *Mathematics of operations research*, vol. 12, pp. 441–450, Aug 1987.

[61] I. Abou-Faycal, M. Medard, and U. Madhow, “Binary adaptive coded pilot symbol assisted modulation over Rayleigh fading channels without feedback,” *IEEE Trans. Commun.*, vol. 53, pp. 1036–1046, June 2005.

[62] P. Sadeghi and P. Rapajic, “Capacity analysis for finite-state Markov mapping of flat-fading channels,” *IEEE Trans. Commun.*, vol. 53, pp. 833–840, May 2005.

[63] W. Turin and R. Van Nobelen, “Hidden Markov modeling of flat fading channels,” *IEEE J. Sel. Areas Commun.*, vol. 16, pp. 1809–1817, Dec 1998.

[64] E. Björnson and E. Jorswieck, *Optimal resource allocation in coordinated multi-cell systems.* Now Publishers Inc, 2013.

[65] A. Asadi and V. Mancuso, “A survey on opportunistic scheduling in wireless communications,” *IEEE Communications Surveys Tutorials*, vol. 15, no. 4, pp. 1671–1688, 2013.

[66] X. Wang, G. B. Giannakis, and A. G. Marques, “A unified approach to QoS-guaranteed scheduling for channel-adaptive wireless networks,” *Proceedings of the IEEE*, vol. 95, no. 12, pp. 2410–2431, 2007.

[67] E. A. Jorswieck, E. G. Larsson, and D. Danev, "Complete characterization of the Pareto boundary for the MISO interference channel," *IEEE Trans. Signal Process.*, vol. 56, no. 10, pp. 5292–5296, 2008.

[68] J. Jianwei Huang, R. A. Berry, and M. L. Honig, "Distributed interference compensation for wireless networks," *IEEE J. Sel. Areas Commun.*, vol. 24, no. 5, pp. 1074–1084, 2006.

[69] A. Condon, "The complexity of stochastic games," *Information and Computation*, vol. 96, no. 2, pp. 203–224, 1992.

[70] M. Wilkes, "The art of computer programming, volume 3, sorting and searching," *The Computer Journal*, vol. 17, no. 4, pp. 324–324, 1974.

[71] Q. Li, X. Yu, M. Xie, N. Li, and X. Dang, "Performance analysis of uplink massive spatial modulation MIMO systems in transmit-correlated rayleigh channels," *IEEE China Commun.*, vol. 18, no. 2, pp. 27–39, 2021.

[72] K. Wang and L. Chen, "On optimality of myopic policy for restless multi-armed bandit problem: An axiomatic approach," *IEEE Trans. Signal Process.*, vol. 60, pp. 300–309, Jan 2011.

[73] H. Murakami, "Approximations to the distribution of sum of independent non-identically gamma random variables," *Mathematical Sciences*, vol. 9, no. 4, pp. 205–213, 2015.

[74] G. Caire, N. Jindal, M. Kobayashi, and N. Ravindran, "Multiuser MIMO achievable rates with downlink training and channel state feedback," *IEEE Trans. Inf. Theory*, vol. 56, no. 6, pp. 2845–2866, 2010.

[75] R. D. Smallwood and E. J. Sondik, "The optimal control of partially observable Markov processes over a finite horizon," *Operations research*, vol. 21, pp. 1071–1088, Oct 1973.