Optimal Transmission over Decentralized Gaussian MAC with Large Users

Project report submitted in partial fulfillment of the requirement for the degree of

BACHELOR OF TECHNOLOGY

IN

ELECTRONICS AND COMMUNICATION ENGINEERING

By

Smriti Thakur (151069)

Akhil Rana (151049)

Shivangi (151091)

UNDER THE GUIDANCE OF

Dr. Neeru Sharma

DEPARTMENT OF ELECTRONICS & COMMUNICATION ENGG.

JAYPEE UNIVERSITY OF INFORMATION TECHNOLOGY, WAKNAGHAT

MAY 2019

Declaration

We hereby declare that the project work entitled **"Optimal Transmission over Decentralized Gaussian MAC with Large Users"** submitted to the **Department of Electronics and Communication, Jaypee University of Information Technology, Waknaghat India,** is a record of an original work carried out under the supervision of Dr. Neeru Sharma. The places where words or ideas from others have been utilized or included, we have adequately cited and referenced the original sources. This project work is submitted as a part of partial fulfilment of award of the degree of Bachelor of Technology under Jaypee University of Information Technology.

Name and Signature of the students:

....
Smriti Thakur (151069)

....
Akhil Rana (151049)

....
Shivangi (151091)

May, 2019
Place: JUIT Waknaghat, India

CERTIFICATE

This is to certify that the work titled **"Optimal Transmission over Decentralized Gaussian MAC with Large Users"** submitted by **Smriti Thakur, Akhil Rana** and **Shivangi** in partial fulfilment for the award of degree of Bachelor of Technology in Electronics and Communication Engineering of Jaypee University of Information Technology, Solan has been carried out under my supervision.

....

Signature of Supervisor

Name of Supervisor- Dr. Neeru Sharma

Designation- Assistant Professor

Date: May, 2019

Place: JUIT Waknaghat, India

Acknowledgement

We take this opportunity to express our gratitude to our supervisor, Dr. Neeru Sharma, for her insightful advice, suggestions, guidance, and support in the successful completion of this project and also for her constant encouragement and information throughout our Bachelors program. We also thank the Head of Department, Prof. M. J. Nigam, for providing the computational facilities throughout the project work. We want to convey our thanks to the teaching and non-teaching staff of the Department of Electronics and Communication Engineering for their help and support. We also acknowledge feedback and comments received on this project work from the Ph.D. student, Kamal Singh, Department of Electrical Engg., at IIT Bombay, India.

Smriti Thakur (151069)

Akhil Rana (151049)

Shivangi (151091)

Abstract

In this project, we investigate a recently proposed alternating maximization (AM) method to numerically solve the optimal decentralized powers and ergodic capacity of wireless Gaussian MACs. This method is suited well for MAC when there are small number of transmitters. For moderate to large MACs, the numerical solutions are possible but at very high expense of computational cost. We seek to investigate this AM algorithm for the decentralized MAC to improve its computational efficiency. In this direction, we notice a key observation that central limit theorem can be invoked for the MAC with large users to solve the partial optimization quickly for the identical-users MAC settings. We also investigate the possibility of increasing the computational efficiency of the AM algorithm by making use of the fast Fourier transform (FFT) algorithms.

Contents

List of Figures

Chapter 1

Introduction

1.1 Overview

Since the birth of radio communications in early twentieth century, point-to-point communication systems were extensively studied, designed and improved in the decades to come. The main challenges at the time were towards developing communication technologies albeit without the full knowledge of the fundamental principles of the underlying theory. Claude E. Shannon, in his paper “A mathematical theory of communication” published in the year 1948, laid down the mathematical foundations governing the fundamental aspects of communication, such as the maximum possible rate of communication in a given channel [1]. However, much of the discoveries were not immediately useful as it seems now that there was not much clarity about the utility of these results among the communication theory experts as well as engineering community. This was further aggravated by the lack of technological innovations at the time. Technological breakthroughs, particularly in IC design and DSPs in the 1970s further led by reinvigorating interest in the practical aspects of information theory in system design, have resulted in complete overhaul in the development of communication technologies for both wired as well as wireless systems. Armed with the knowledge of fundamental limits of performance, the system design goals now align with optimizing the system performance under constrained resources such as power and bandwidth as well as design and implementation issues.

With growing demand for mobile telephony services further by increasing consumption in data based services at an explosive pace, particularly in in the last two decades, it has been well identified that multi-user communication scenarios are inevitable. Consequently, efficient allocation of resources is going to be a very challenging task keeping in mind the scarce and expensive resources namely bandwidth and energy.

Unlike communication between a single transmitter and receiver pair (point-to-point) where optimal communication schemes are explicitly known for most of the popular channel models such as additive white Gaussian noise (AWGN), fading AWGN etc., finding the optimal schemes for the multi-user channels is generally considered a difficult task. The complexity arises due to the inherent nature of the multi-user communication: multiple users try to compete and cooperate to enhance the overall performance and it is not always possible to determine the best possible strategy of competition and cooperation in order to optimize the performance. Furthermore, multi-user communication is limited by the interference generated by the users. To fully exploit the multi-user channel, it is crucial to examine the nature of the interference itself. This, except for a few special cases, is not a trivial task. Some interesting results have been published for the point-to-point communication systems such as in [2] [3] [4] [5]. Unlike point-to-point communication where there is only one sender and a single receiver, in a multi-user communication system, multiple senders try to utilize the resource as efficiently as possible.

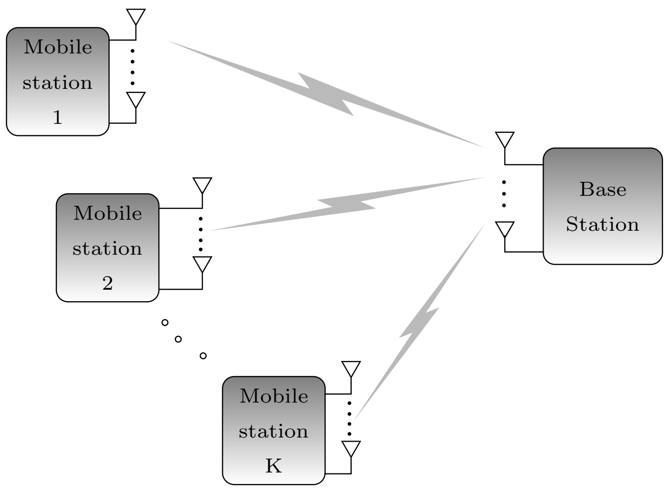

Figure 1.1: A typical wireless cellular uplink. (Courtesy of Singh, K. [6]).

A well known example is the uplink of a cellular wireless radio where multiple mobile users communicate via the shared wireless channel to a common base station, see Figure 1.1. The shared channel is accessed by the users in a pre-planned manner. For example, in the GSM cellular standard, the multiple users access the uplink in different time slots allocated by the base-station. This multiple access is known in the literature by the term *time division multiple access* or simply TDMA. Alternatively, the shared channel bandwidth could be divided among the users, which is known by the term *frequency division multiple access* or FDMA. Another popular multiple access, known as code division multiple access or CDMA, allows the multiple users to simultaneously communicate to the base-station utilizing the full bandwidth. This is made possible by orthogonalizing the senders by assigning unique orthogonal codes to the users. More multi-user communication examples can be such as wireless terminals connected to the access point in a Wireless local area network, or multiple sensors in an ad-hoc network trying to send their data to a common handler etc. Such many-to-one communication situation is modelled in network information theory under the umbrella term “Multiple Access Channel” [6, Chapter 23]. In this project, motivated by these examples, we try to evaluate the optimal communication schemes for a variant of multiple access channel (MAC) suitable for wireless environment.

1.2 Multiple Access Channel

MAC is a multi-user channel model where multiple senders communicate to a common receiver via a shared channel medium. Consider a two sender discrete memoryless MAC as in Figure 1.2, where X_1^n and X_2^n are codeword symbols transmitted for independent messages M_1 and M_2 respectively. The channel is represented by the the transition probability matrix $p(y|x_1, x_2)$. The task of the receiver is to produce an estimate of the transmitted messages M_1 and M_2, denoted by ($\widehat{M}_1$ and $\widehat{M}_2$), using the received codeword Y^n. Formally, a codebook $(n, 2^{nR_1}, 2^{nR_2})$ for a multiple access channel consists of

- An ensemble of message sets $M_i \in \{1, \cdots, 2^{nR_i}\}$, $i = 1, 2$.
- A set of encoders (one for each message) that assigns a codeword $X_i^n(M_i)$, $i = 1, 2$.
- A decoder function that assigns an estimate $(\widehat{M}_1(Y^n), \widehat{M}_2(Y^n))$.

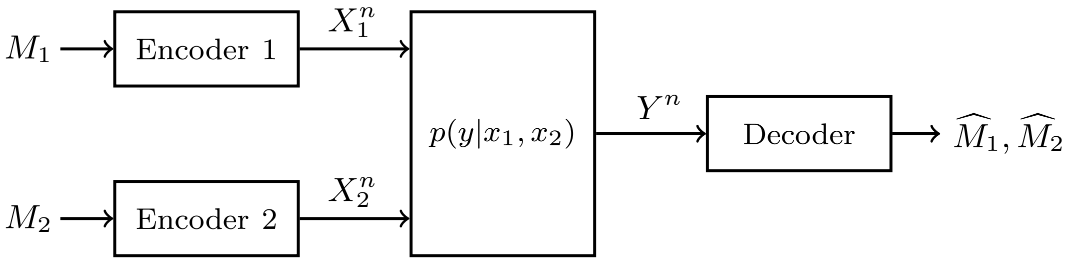

Figure 1.2: Two-User Discrete Memoryless Multiple access channel

"Decoding error" occurs whenever $(\widehat{M}_1(Y^n), \widehat{M}_2(Y^n)) \neq (M_1, M_2)$. By reliable communication, we mean that, on the average, there is no decoding error.

Definition 1. *The average probability of error $P_e^{(n)}$ is defined as*

$$P_e^{(n)} = \Pr((\widehat{M}_1(Y^n), \widehat{M}_2(Y^n)) \neq (M_1, M_2)). \tag{1.1}$$

Definition 2. *A rate pair (R_1, R_2) is said to achievable if there exists a sequence of codes $(n, 2^{nR_1}, 2^{nR_2})$, indexed by n, for which the average probability of error $P_e^{(n)}$ approaches zero as block code length n increases unboundedly i.e. $\lim_{n\to\infty} P_e^{(n)} = 0$.*

The collection of all achievable rate-pairs is termed as the *capacity region* of the multiple access channel. Ahlswede [7] and Liao [8] characterized the capacity region of the discrete memoryless multiple access channel which is shown next for the MAC with two senders.

Theorem 3. *With fixed input distribution $p(x_1)p(x_2)$, the capacity region of a discrete-memoryless multiple access channel denoted by the transition matrix $p(y|x_1,x_2)$ is given as*

$$\begin{aligned} R_1 &\leq I(X_1;Y|X_2), \\ R_2 &\leq I(X_2;Y|X_1), \\ R_1 + R_2 &\leq I(X_1, X_2;Y). \end{aligned} \tag{1.2}$$

In general, when the input distributions are not fixed, the capacity region of a discrete memoryless multiple access channel is given by the *convex hull* of the union of the instantiated capacity regions over all $p(x_1)p(x_2)$ distributions. The convex hull operation corresponds to *time-sharing* between "rate tuples" in different rate regions. As previously mentioned, due to independence of the x_1 and x_2 inputs, the input distribution has the product form as $p(x_1)p(x_2)$. The input distribution $p(x_1)p(x_2)$ being non-convex, search for the optimal input distribution of a multiple access channel is in general a hard problem [9] compared to single user channel where the optimization is a convex program and thus is a lot simpler to work with [10] [11] (see also [12]). For the multiple access channel with correlated sources, coding theorems were given by Slepian and Wolf [13] for a special class of correlations and by Cover, Al Gamal, and Salehi [14] for arbitrarily correlated sources.

1.2.1 Preliminaries: Gaussian multiple access channel

The main focus of this project is Gaussian multiple access channel with independent sources or users. A Gaussian multiple access channel is a multiple access channel for which the channel law $p(y|x_1,x_2)$ is Gaussian. For a Gaussian MAC with scalars X_1 and X_2 as inputs and Y output in the Figure 1.3, the channel output is given as

$$Y = h_1 \cdot X_1 + h_2 \cdot X_2 + Z,$$

where h_1 and h_2 are fixed (can be varying) multiplicative path gains from the users towards the receiver, P_1 and P_2 are the fixed transmit powers of the respective users and Z is the unity-variance AWGN noise.

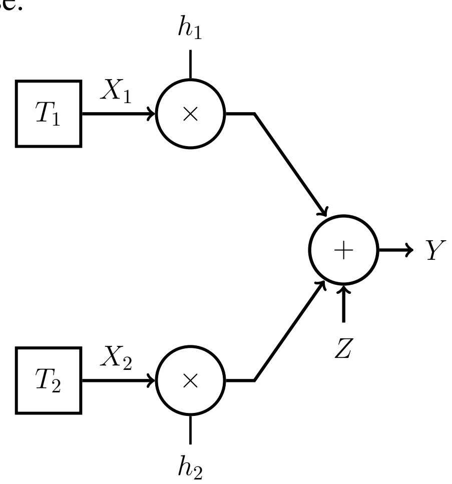

Figure 1.3: Two-User Gaussian MAC

For scalar Gaussian MAC, it turns out that independent Gaussian distributions for the inputs maximize the channel capacity. Furthermore, the union and the convex-hull become superfluous. Application of the Theorem 3 to the scalar Gaussian MAC gives its capacity region in an elegant form as stated in the following theorem [15] [16].

Theorem 4. *The capacity region of Gaussian MAC is the set of* (R_1, R_2) *pairs such that*

$$\begin{aligned} R_1 &\leq \log(1+h_1^2 P_1), \\ R_2 &\leq \log(1+h_2^2 P_2), \\ R_1 + R_2 &\leq \log(1+h_1^2 P_1 + h_2^2 P_2), \end{aligned} \tag{1.3}$$

Achievability of the capacity region: The mutual information between the input X_1 and ouput Y given X_2, denoted by $I(X_1;Y|X_2)$, is computed as

$$\begin{aligned} I(X_1;Y|X_2) &= h(Y|X_2) - h(Y|X_1,X_2), \\ &= h(h_1 \cdot X_1 + h_2 \cdot X_2 + Z|X_2) - h(h_1 \cdot X_1 + h_2 \cdot X_2 + Z|X_1,X_2), \\ &= h(h_1 \cdot X_1 + Z) - h(Z), \\ &\leq \log\left(2\pi e(1+h_1^2 P_1)\right) - \log\left(2\pi e\right), \\ &= \log(1 + h_1^2 \cdot P_1), \\ &\triangleq C(h_1^2 P_1), \end{aligned} \tag{1.4}$$

where $C(\text{SNR}) \triangleq \log(1+\text{SNR})$ is the well known capacity formula for an AWGN channel with received signal power to noise ratio given as SNR. The inequality above is due to the fact that for a given variance (power) of input X_1, Gaussian distribution maximizes entropy. With similar arguments, the other two mutual information quantities in (1.2) can be shown bounded by respective quantities in the RHS of (1.3) which can be achieved by choosing $X_1 \sim \mathcal{N}(0, P_1)$ and $X_2 \sim \mathcal{N}(0, P_2)$ independent distributions. The region bounded by these inequalities is a pentagon as shown in the Figure 1.4.

To show that this is indeed the capacity region, it is required to prove that the all rate-pairs inside the bounds in (1.3) are achievable. The rates choice at the boundary point A

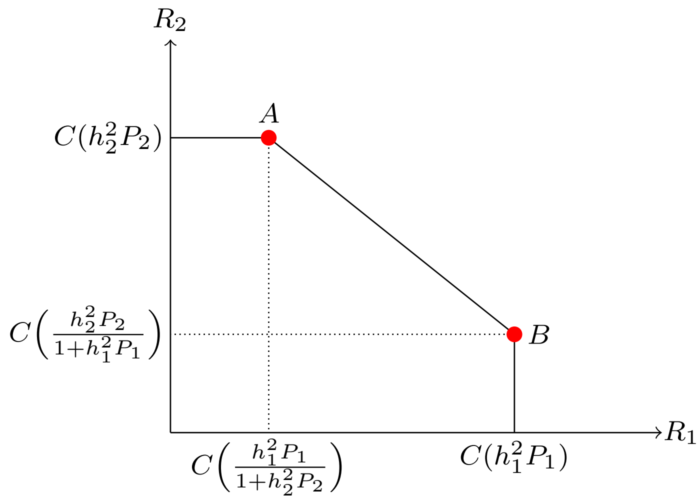

Figure 1.4: Capacity region of a Two-User Gaussian Multiple access channel

given as $(R_1, R_2) = \left(C(h_2^2 P_2), C\left(\frac{h_1^2 P_1}{1+h_2^2 P_2}\right)\right)$ is achievable: the receiver decodes the information of user-1 treating the Gaussian signal of user-2 as interference alongwith the Gaussian noise. In doing so, the best rate possible for user-1 is equal to $C\left(\frac{h_1^2 P_1}{1+h_2^2 P_2}\right)$. The receiver then exactly reconstructs the user-1 signal and subtracts it from the overall received signal. The resulting signal is thus cleaned off user-1 signal previously acting as interference for the user-2 signal and hence, the only interference to the user-2 signal is due to Gaussian noise Z present in the channel. Thus, the maximum possible rate for user-2 is its single-user bound $C(h_2^2 P_2)$. Similarly, the rate choice at the point B can be achieved with the roles of decoding reversed. This idea of successive cancellation and decoding the individual user codewords at the corner points of the capacity region is due to Bergmans and Cover [15] and to Wyner [16]. Furthermore, *successive interference cancellation* (SIC) decoding requires only single-user decoders. To complete the proof, all the rate-pairs on the line segment AB can be achieved by proper time-sharing of the rate-tuples at the corner points A and B. It is also possible to perform joint decoding albeit with a complex multi-user decoding at the receiver. For the more general case of Gaussian MAC where link between for each sender and the receiver is characterized by fixed multipaths, the capacity region is completely characterized in [3] where power allocation is performed in the frequency domain.

In the interest of the reader, a brief note on significant research in the *asynchronous* multiple access channel follows. The capacity region theorem for the MAC channel presented above is based on the assumption that the codewords of the senders are frame-synchronized i.e. the block lengths of the codes used by the different users are identical and the beginnings of these codewords are always in unison. Furthermore, the rates chosen using the time sharing principle requires time synchronization among the users such that a sender transmits only when the other sender has completed transmitting its codeword. For the asynchronous memoryless MAC, the lack of frame synchronization results in the removal of the convex hull or time-sharing operation in the capacity region as reported first by [17] and later independently by [18]. However, when the delay mismatch in the codeword frames of the users is bounded, it is shown [19] that convex closure is still possible using a generalized time sharing. Asynchronous MAC with memory with and without frame synchronization was studied by Verdú in [20]. Later, an alternative multiple access based on successive decoding for Gaussian MAC called *rate-splitting* is devised in [21] to reduce the implementation complexity and frame synchronization dependence of successive cancellation and time-sharing approaches mentioned earlier. For additional reading in this area, [22] [23] [24] are interesting sources.

In this project, we focus on a special MAC, namely the wireless Gaussian MAC where communication from each sender to the receiver is perturbed by the random or time-varying behaviour of the channel or medium. This random nature of the wireless medium is commonly known by the term *fading* in the literature. Availability of the fading or channel information at the receiver and/or transmitter generally enhances the throughput performance of the Gaussian MAC [25] [26]. We investigate a *decentralized* coherent Gaussian MAC, where channel information is available only at the respective sender. The objective is to find the optimal decentralized power laws and the ergodic sum-capacity of this MAC for the case when there are moderate to large number of active senders.

The organization of this project report is as follows. Chapter 2 introduces the decentralized wireless fading MAC model, followed by defining the utility of interest and introducing the optimization problem in Section 2.1. The AM method to numerically solve the optimal decentralized powers is presented in Section 2.2. Further, the details of the AM approach are covered. In Section 2.3, we propose the transform based method developed to solve the decentralized powers faster using the fast Fourier transform (FFT) algorithms. The special case of MAC with large users is presented in Section 2.4, detailing the simplifications possible using central limit theorem (CLT). In Section 2.5, we present the numerical computations for decentralized identical users MAC, namely the optimal power controls and ergodic sum-capacity. Chapter 3 concludes this project report with summarizing the results and possible future directions.

Chapter 2

Wireless MAC with Distributed CSI

Before we delve into the specific details of the decentralized wireless MAC, it is important to briefly discuss the meaning and impact of *fading* on communication or signalling aspect in a wireless environment. In a typical communication over a wireless medium, the received signal strength fluctuates randomly. This manifestation of the random behaviour of a typical non *line of sight* wireless channel in the received signal strength is commonly referred to as *fading*. There are many definitions on channel information in the literature. Here, we will refer the instantaneous exact value of the fading as channel state information (CSI). Knowledge of CSI at the receiver and/or at the transmitters can be exploited to leverage the opportunistic nature of the wireless medium. That is, the transmitters may cooperate rather than merely compete, to increase the overall performance. In general, with some degree of channel information, the communication strategy can be improved to utilize the resources (power and bandwidth) efficiently; as a coarse example, consider the situation when the channel is good for one user and is bad for the other, it seems advantageous for the better user to utilize the communication resources while the weaker user waits while conserving its resources hoping for better channel states in the future. It is indeed true that both the users may get benefit in the long run while not losing much in the short run. For example, in a centralized MAC where all the transmitters and the receiver knows the CSI at all instants, the optimal communication strategy is *opportunistic* TDMA where at any instant only the transmitter with the best channel state transmits [25]. However, if the data to be transmitted is delay-sensitive such as digital voice, it is imperative to design schemes that allow *guaranteed* short-term rates while conceding some long-term benefits.

In this project, we investigate a wireless fast-fading MAC where the multiple users have individual CSI while the receiver knows the full CSI at all instants. We will refer to this MAC as *decentralized MAC*. The decentralized MAC is a suitable model for situations where acquiring full CSI at all the transmitters is not practically viable but individual CSI at the respective transmitters can be made possible [27]. This can possibly be true in time-division duplex (TDD) setting where the time scale of fading variations are comparable to the feedback delay in sending the full CSI from the receiver to all the transmitters [6]. Another possibility could be simply the cost of feedback mechanism for providing full CSI at the transmitters is too high.

This decentralized MAC model was first introduced and investigated in [27], concluding that the optimal solutions cannot be determined analytically. As an alternative, asymptotically good power law is proposed for the decentralized power controls in [27], which are significantly improved in recent research by [28] [29]. Very recently, a numerical algorithm is proposed to compute the ergodic sum-capacity and the optimal power controls of the decentralized MAC [30]. This method is based on the principle of alternating maximization (AM). [30] suggests that the computational complexity of the algorithm *severely* restricts its usage for MAC with moderate to large number of users. This is due to the nature of the computational approach of the proposed algorithm. In this project, we investigate whether *computational efficiency* of the AM algorithm can be improved for the Gaussian MAC with moderate to large number of users. Precisely, we explore the AM algorithm structure to take advantage, if possible. Furthermore, we look at the optimization problem for intuitions that can possibly lead to some simplifications that are otherwise missing in the current numerical scheme.

2.1 System Model

As already mentioned in Chapter 1, MAC is a many-to-one channel that models the communication between multiple senders and a common receiver, see Figure 2.1. To formalize the model, consider a discrete-time wireless fading MAC with K transmitters. The symbol

X_i represent the signal transmitted by the user i, which in turn, experiences independent flat fading denoted by complex-valued multiplying coefficient H_i. Notice that for simplicity of exposition, we have dropped the time dependence of the transmitted signal, fading values, noise etc. Thus, at the receiver, the symbol is given as

$$Y = \sum_{i=1}^{K} H_i X_i + Z,$$

where Z is the additive white Gaussian noise (AWGN) present in the receiver.

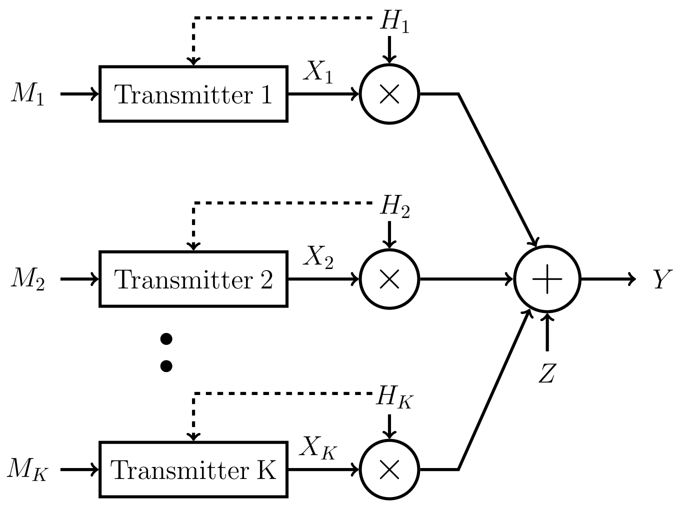

Figure 2.1: Decentralized/Distributed gaussian MAC. (Courtesy of Singh, K. [6])

The symbols X_i, $1 \leq i \leq K$, represent the MAC channel inputs and are statistically independent of each other. Further, the fadings H_i, $1 \leq i \leq K$, are also statistically independent across all users and vary $i.i.d.$ (independently and identically distributed) in time.

The transmitters adapt their rate and power allocation according to the available CSI. We consider a channel state information (CSI) model where transmitter i only knows its respective current channel state h_i at all instants. The full CSI, represented as $(h_1, \cdots, h_K)$, is available at each instant to the receiver. In a fast fading model such as ours, it is sufficient to perform only power control to achieve ergodic capacity while keeping individual communication rates fixed at the transmitters [26]. The transmitter i, observing channel state h_i, expends $P_i(h_i) \triangleq \mathbb{E}\ [|X_i|^2|h_i]$ as transmit power under the long-term average power constraint $\mathbb{E}\ [P_i(h_i)] \leq P_i^{avg}$. For simplicity, we shall drop the dependence of $P_i(\cdot)$ on h_i wherever no confusion arises and denote the i-th user power control by P_i.

Definition 5. *The* ergodic sum-capacity *of the decentralized MAC is given by [6]*

$$\begin{aligned} C_{sum} &= \max_{P_1,\ldots,P_K} \mathbb{E}\left[\log\Big(1+\sum_{i=1}^{K}|H_i|^2 P_i(H_i)\Big)\right], \qquad (2.1) \\ s.t. &\qquad \mathbb{E}\ P_i(H_i) \leq P_i^{avg}, 1 \leq i \leq K. \end{aligned}$$

The receiver in our model is coherent i.e. full CSI is known the receiver at all instants. Hence, only the fading magnitudes are necessary and the above definition of ergodic sum-capacity thus gets simplified as

$$C_{sum} = \max_{P_1,\ldots,P_K} \mathbb{E}\left[\log\Big(1+\sum_{i=1}^{K} V_i P_i(V_i)\Big)\right], \qquad (2.2)$$

where $V_i := |H_i|^2$ and $\mathbb{E}\ [P_i(V_i)] \leq P_i^{avg}$, $1 \leq i \leq K$. It is straightforward to notice that the optimization in (2.2) is a convex program; the objective function is concave and the average power constraints are convex. Thus, KKT conditions are both necessary and sufficient for solving the optimal solutions. Let us write the unconstrained objective or utility function as

$$\mathbb{E}\ \left[\log\Big(1+\sum_{i=1}^{K} v_i P_i(v_i)\Big)\right] - \sum_{i=1}^{K}\lambda_i \cdot \mathbb{E}\ [P_i(v_i)],$$

where the constant parameter λ_i is the Lagrange multiplier for the average power constraint $\mathbb{E}\ [P_i(V_i)] \leq P_i^{avg}$, $1 \leq i \leq K$. For the evaluation of the optimal power control P_i^* whenever non-zero, we can set the derivative of the utility function w.r.t. P_i, $1 \leq i \leq K$ to zero. Thus, optimal power $P_i^*(v_i)$, $1 \leq i \leq K$ are the solutions of

$$v_i \cdot \mathbb{E}\ \left[\frac{1}{1+v_i P_i(v_i)+\sum_{j\neq i} v_j P_j(v_j)}\right] = \frac{\lambda_i}{v_i}, 1 \leq i \leq K, \qquad (2.3)$$

where v_i in (2.3) is fixed and the expectation is over $(V_1,\cdots,V_K)$ except V_i. Extracting or solving (2.3) for $P_i(v_i)$ does not seem feasible by applying the available variational methods, also suggested in [27] [29]. It thus becomes imperative to try numerical methods to solve the optimization [30].

2.2 AM Algorithm for Decentralized Gaussian MAC

Now we focus on the main topic of this project i.e. investigation of the alternating maximization (AM) algorithm for the optimal decentralized powers proposed in [30]. The working of this AM algorithm is briefly explained next, refer [30] for exact details.

The joint optimization in (2.2) gets simplified into partial optimizations due to convex nature of (2.2), and the partial optimization is then solved numerically utilizing the monotone structure of the optimal powers [30]. For ease of exposition, we reconsider the sum-rate maximization problem in (2.2) albeit for a "two-user" Gaussian fading MAC as follows:

$$\begin{aligned} \max_{P_1, P_2} \quad & \mathbb{E}\left[\log(1 + v_1 P_1(v_1) + v_2 P_2(v_2))\right], \\ \text{subj.to} \quad & \mathbb{E}\left[P_i(v_i)\right] \le P_i^{\text{avg}}, \ i = 1, 2. \end{aligned} \tag{2.4}$$

Employing alternative maximization (AM) technique to find the optimal decentralized powers in (2.4) as shown in [30], the steps involved for our two user MAC are as follows

---

Algorithm **Optimal powers for Decentralized MAC**

```
Initialization:   P_2^0(v_2), n = 1.

Repeat

        Compute   P_1^n = max_{P_1} E [log(1 + v_1 P_1(v_1) + v_2 P_2^{n-1}(v_2))]

        Compute   P_2^n = max_{P_2} E [log(1 + v_1 P_1^n(v_1) + v_2 P_2(v_2))]

until convergence
```

It is straightforward to notice that key step in the AM algorithm is solving the partial optimization as follows

$$P_i^* = \max_{P_i} \mathbb{E}\left[\log\left(1 + v_i P_i(v_i) + v_j P_j(v_j)\right)\right],\ i = 1, 2 \quad \text{such that } i \neq j.$$

To solve the partial maximization under the constraint $[P_i(v_i)] \leq P_i^{\text{avg}}$, in [30], it is shown that the monotonicity structure of the optimal power can be exploited. Further, it is shown that

$$\lim_{n\to\infty} \mathbb{E}\left[\log\left(1 + v_1 P_1^n(v_1) + v_2 P_2^{n-1}(v_2)\right)\right] = C_{sum},$$

i.e. the ergodic sum-rate in the AM algorithm converges to the ergodic sum-capacity C_{sum}. For more details on convergence and optimality of the AM algorithm, please refer [30].

2.2.1 Identical Users MAC

For convenience sake, we henceforth consider *identical-users*[1] MAC with $(K+1)$ transmitters. As mentioned in [27], for this special MAC configuration, all the optimal power schemes are identical due to concavity of the logarithmic utility. In the interest of the reader and completeness, we present the proof of this property below.

Lemma 6. *For identical-users Gaussian MAC, all optimal powers are identical i.e.*

$$P_1^* = P_2^* = \cdots = P_{K+1}^* = P^* say.$$

Proof. W.l.o.g we shall focus on P_1 and P_2 i.e. the power schemes of user 1 and user 2 respectively. Further, for convenience, we denote the sum of the remaining terms $\sum_{i>2} V_i P_i(V_i)$ by G. The average sum-rate achievable for these power schemes set is computed as

$$\begin{aligned} R &= \mathbb{E}\left[\log(1 + V_1 P_1(V_1) + V_2 P_2(V_2) + G)\right], \\ &= \mathbb{E}\left[\log(1 + V_1 P_2(V_1) + V_2 P_1(V_2) + G)\right], \\ &= \tfrac{1}{2} \cdot \mathbb{E}\left[\log(1 + V_1 P_1(V_1) + V_2 P_2(V_2) + G)\right] \\ &\qquad\qquad + \tfrac{1}{2} \cdot \mathbb{E}\left[\log(1 + V_1 P_2(V_1) + V_2 P_1(V_2) + G)\right], \end{aligned}$$

[1] All users have identical channel statistics and average power constraints

$$\leq \mathbb{E} \log\left(1 + V_1 \frac{P_1(V_1)+P_2(V_1)}{2} + V_2 \frac{P_2(V_2)+P_1(V_2)}{2} + G\right),$$

where the second equality follows from the fact that both the channels V_1 and V_2 are identically distributed, and the last inequality results from the concavity property of the logarithm function. The last inequality suggests that the sum-rate can always be improved by choosing the average of the two power functions as the common power control scheme. Thus, in the optimal case, $P_1^* = P_2^*$ holds. To complete the proof for the remaining powers, these arguments can be extended to the remaining terms in G in an iterative manner. ■

For the $(K+1)$ identical-users MAC, with the help of Lemma 6, the optimization in the AM algorithm simplifies to

$$\begin{aligned} &\max_{P} \ \mathbb{E}\left[\log(1 + v_1P(v_1) + v_2P(v_2) + \cdots + v_{K+1}P(v_{K+1}))\right], \\ &\text{subj. to} \qquad\qquad \mathbb{E}\left[P(v)\right] \leq P_{\text{avg}}, \end{aligned} \tag{2.5}$$

where P_{avg} is the average power constraint common to all the transmitters. Suppose $P^*(v)$ denotes the optimal solution of (2.5) i.e.

$$P^* = \arg\max_{P} \ \mathbb{E}\left[\log(1 + v_1P(v_1) + v_2P(v_2) + \cdots + v_{K+1}P(v_{K+1}))\right]. \tag{2.6}$$

More importantly, the partial optimization step in the AM algorithm for the *identical-users* MAC configuration is given by

$$\hat{P}^* = \arg\max_{P} \ \mathbb{E}\left[\log\left(1 + vP(v) + \sum_{j=1}^{K} v_j\widehat{P}(v_j)\right)\right], \tag{2.7}$$

where $\widehat{P}$ is the initialized or given power scheme chosen exactly identical for all the remaining transmitters. (2.7) being analytically difficult, numerical techniques can be employed to solve the partial optimization in (2.9) for the partially optimal $\hat{P}^*$ for each run of the AM algorithm. In the limit as the number of runs of iterative AM algorithm goes to infinity, the partially optimal $\hat{P}^*(v)$ approaches the globally optimal $P^*(v)$ [30].

Let us denote the identical fading distribution by $\Psi(\cdot)$. Similar to the joint optimization described previously in (2.1), the partial maximization in (2.7) subjected to the constraint $\mathbb{E}\,[P(v)] \leq P_{avg}$ is a convex program and thus, convex optimization techniques can be readily applied. Thus, the necessary and sufficient KKT condition on the partially optimal power $\hat{P}^*(v)$ can be obtained by setting the derivative of the utility function to zero. Thus we have [30]

$$\int_{v_1} \cdots \int_{v_K} \frac{d\Psi(v_1,\ldots,v_K)}{1+vP(v)+\sum_{j=1}^{K} v_j\widehat{P}(v_j)} = \frac{\lambda}{v}. \tag{2.8}$$

The partial optimization in (2.7) is solved numerically by utilizing the monotonically increasing behaviour of the $\hat{P}^*(v)$ with increasing v, see [30] for exact details. For the interest of the reader and completeness, we restate the monotone structural property of the partially optimal power $\hat{P}^*(v)$ and the proof follows in a slightly tweaked manner compared to [28], [30], as shown below:

Theorem 7. *$\hat{P}^*(v)$ of (2.7), whenever non-zero, always satisfies* $\dfrac{d\hat{P}^*(v)}{dv} > 0$.

Proof. KKT condition for the partial optimal power $\hat{P}^*(v)$ in (2.7) subjected to $\mathbb{E}\,[P(v)] \leq P_{\text{avg}}$ is

$$\int \frac{d\Psi(v_1,\ldots,v_K)}{1+vP(v)+\sum_{j=1}^{K} v_j\widehat{P}(v_j)} = \frac{\lambda}{v}, \text{ whenever } P(v) > 0. \tag{2.9}$$

Rewriting this as

$$v\int \frac{dF_Z(z)}{1+vP(v)+z} = \lambda \;\text{ whenever }\; P(v) > 0, \tag{2.10}$$

where we replace $\sum_{j=1}^{K} v_j\widehat{P}(v_j)$ by Z for notational convenience and its probability distribution is denoted by $F_Z(z)$. Let us rewrite the above condition more compactly as

$$\mathbb{E}\left[\frac{v}{1+vP(v)+z}\right] = \lambda. \tag{2.11}$$

Please note that the expectation operation above is with respect to the random variable Z. Taking derivative of (2.11) with respect to v on both sides, we get

$$
\begin{aligned}
& \frac{d}{dv}\left(\mathbb{E}\left[\frac{v}{1+vP(v)+z}\right]\right) &=& \; 0, \\
\Rightarrow \quad & \mathbb{E}\left[\frac{d}{dv}\left(\frac{v}{1+vP(v)+z}\right)\right] &=& \; 0, \\
\Rightarrow \quad & \mathbb{E}\left[\frac{(1+vP(v)+z)\cdot 1 - v\frac{d}{dv}vP(v)}{(1+vP(v)+z)^2}\right] &=& \; 0, \\
\Rightarrow \quad & \mathbb{E}\left[\frac{(1+vP(v)+z) - v(vP'(v)+P(v)\cdot 1)}{(1+vP(v)+z)^2}\right] &=& \; 0, \\
\Rightarrow \quad & \mathbb{E}\left[\frac{1+z-v^2P'(v)}{(1+vP(v)+z)^2}\right] &=& \; 0, \\
\Rightarrow \quad & \mathbb{E}\left[\frac{1+z}{(1+vP(v)+z)^2}\right] &=& \; v^2P'(v)\cdot\mathbb{E}\left[\frac{1}{(1+vP(v)+z)^2}\right].
\end{aligned} \tag{2.12}
$$

Notice that the LHS of (2.12) above is strictly positive since the argument of the expectation operator is strictly positive as $z \geq 0$ always and also the expectation on the RHS is also strictly positive. Thus, $P'(v)$ must be strictly positive for the condition (2.12) to hold true. Hence the partially optimal power $\hat{P}^*(v)$ satisfies $\frac{d\hat{P}^*(v)}{dv} > 0$. ■

Recall (2.10)

$$
\int \frac{dF_Z(z)}{1+vP(v)+z} = \frac{\lambda}{v}, \tag{2.13}
$$

and that $Z := \sum_{j=1}^{K} v_j\widehat{P}(v_j)$ is sum of K i.i.d. random variables $V_j\widehat{P}(V_j)$, $1 \leq j \leq K$. Since the optimal $P(v)$ is monotone (increasing) in v, the LHS of (2.13) denoted by $t(vP(v))$ has one-to-one correspondence to the RHS. Thus

$$
\begin{aligned}
t(vP(v)) &= \frac{\lambda}{v}, \\
\Rightarrow \quad P(v) &= \frac{1}{v}\, t^{-1}\left(\frac{\lambda}{v}\right),
\end{aligned} \tag{2.14}
$$

(2.14) can be used to solve for the optimal $P(v)$ (see Corollary 6 in [30]). However, this requires computation of the integral in the LHS of (2.13) which in turn necessitates determination of the distribution $F_Z(z)$ or $f_Z(z)$. Clearly, if there exists some simplification about solving (2.13), it can be exploited to reduce the computation of the optimal power. This aspect is explored in further sections.

2.3 Fast Alternating Maximization (AM) using FFT

In this section, we explore the possibility of ease of computation of the optimal powers using (2.13) using transform method combined with fast Fourier transform (FFT) algorithm. By transform method, we mean the characteristic function approach to compute the probability distribution of sum of i.i.d. random variables, this will become clearer in the discussion to follow. We begin with describing the implementation of the AM method using FFT algorithm.

Let $W_i = V_i\widehat{P}(V_i)$ and $Z = \sum_{i=1}^{K} W_i$ i.e. Z is sum of *i.i.d.* random variables denoted by W_i's. The probability distribution of Z, being sum of K i.i.d. random variables, can be found using the K-fold convolution of the distribution of W or alternatively by taking the inverse transform of the K-th power of the characteristic function of W. We compute the distribution of W (general) using transform approach to obtain the distribution of Z as it is more efficient for large number of samples (will be shown later). Since the optimal power is monotonically increasing and $P(0) = 0$, we can safely argue that the power allocation begins only after $V \geq v_0$ for some $v_0 > 0$, thus implying that

$$\begin{aligned} W &= 0 \quad \text{whenever } V < v_0 \\ \text{and} \quad W &> 0 \quad \text{whenever } V \geq v_0. \end{aligned}$$

During the initialization step in the AM algorithm, the power scheme $\widehat{P}(\cdot)$ is chosen to be monotonically increasing for $V \geq v_0$. For example, the initialized $\widehat{P}(v)$ can be standard or modified waterfilling scheme [28]. From then onwards, the currently computed partial optimal $P(v)$ (set as $\widehat{P}(v)$ in the next iteration of the AM maximization) will also be monotonically increasing [30, Remark 4]. This monotone property guarantees one-to-one mapping between V and W and thus, the inverse $W^{-1}(\dot{)}$ mapping exists for all choices of $W > 0$. Given the cumulative distribution function (CDF) $\Psi_V(v)$ (or probability density function (PDF) $f_V(v)$) of V, the CDF for the random variable W is as found to be

$$F_W(w) = \begin{cases} 0, \; w < 0, \\ \Psi_V(v_0), \; w = 0, \\ \Psi_V(W^{-1}(w)) - \Psi_V(v_0), \; w > 0. \end{cases}$$

Here, we have assumed a regularity condition on W namely differentiability. This can be easily achieved choosing a suitable power control on V. Differentiating the CDF $F_W(w)$ gives

$$f_W(w) = \Psi_V(v_0) \cdot \delta(w) + f_V(W^{-1}(w)) \frac{d\,(W^{-1}(w))}{dw} \mathbb{1}_{\{w>0\}}.$$

Notice that monotone (increasing) property of the function $W(\cdot)$ and hence of $W^{-1}(\cdot)$ function implies $\dfrac{d\,(W^{-1}(w))}{dw}$ is positive for all $w > 0$.

Let $\Phi_Z(\omega)$ and $\Phi_W(\omega)$ denote the characteristic functions of the random variables Z and W respectively. The characteristic function of W is

$$\begin{aligned} \Phi_W(\omega) &= \int_0^\infty e^{j\omega w} f_W(w)\,dw, \\ &= \Psi_V(v_0) + \int_{0+}^\infty e^{j\omega w} f_W(w)\,dw, \\ &= \Psi_V(v_0) + \int_{0+}^\infty e^{j\omega w} f_V(W^{-1}(w)) \frac{d\,(W^{-1}(w))}{dw}\,dw. \end{aligned}$$

Similary, the characteristic function of Z follows as

$$\begin{aligned} \Phi_Z(\omega) &= \int_0^\infty e^{j\omega z} f_Z(z)\,dz, \\ &= \int_{w_1=0}^\infty \cdots \int_{w_K=0}^\infty e^{j\omega \sum_{i=1}^K w_i} f_W(w_1)dw_1 \cdots f_W(w_K)dw_K, \\ &= \left\{ \int_w e^{j\omega w} f_W(w)\,dw \right\}^K, \\ &= \{\Phi_W(\omega)\}^K. \end{aligned} \tag{2.15}$$

Using the characteristic function $\Phi_Z(\omega)$, the PDF of the random variable Z is obtained by taking inverse transform as following:

$$f_Z(z) \;=\; \frac{1}{2\pi}\int_{-\infty}^{\infty}\Phi_Z(\omega)e^{-j\omega z}\,d\omega\,.$$

Finally, with $dF_Z(z) = f_Z(z)\,dz$ in (2.13), the KKT condition simplifies as

$$\int_0^{\infty}\frac{f_Z(z)\,dz}{1+\nu P(\nu)+z} \;=\; \frac{\lambda}{\nu}. \tag{2.16}$$

2.3.1 Continuous to Discrete Representation

It is obvious that we need to compute the distribution $f_Z(z)$ efficiently particularly for the MAC with moderate to large number of users, if possible. This is exactly the bottleneck of the proposed AM algorithm in [30], *severely* restricting the usage to a few number of MAC users.

In this direction, we desire to make use of the fast Fourier transform algorithms to speed up the computations involved. We apply discretization to transform continuous attributes into discrete ones wherever necessary. This is explained next in detail.

Recall the computation of characteristic function of W

$$\begin{aligned}\Phi_W(\omega) \;&=\; \int_0^{\infty} e^{j\omega w} f_W(w)\,dw\,,\\ &=\; \Psi_V(v_0) + \int_{0+}^{\infty} e^{j\omega w} f_W(w)\,dw\,,\end{aligned} \tag{2.17}$$

To allow numerical computation of the integral form in (2.17) we substitute the integration variable w by $n\Delta$ i.e.

$$\begin{aligned}\Phi_W(\omega) \;&\overset{(a)}{=}\; \Psi_V(v_0) + \int_{0+}^{\infty} e^{j\omega w} f_W(w)\,dw\,,\\ &\overset{(b)}{\approx}\; \Psi_V(v_0) + \Delta\cdot\sum_{n=1}^{\infty} e^{j\omega n\Delta} f_W(n\Delta),\\ &\overset{(c)}{\approx}\; \Psi_V(v_0) + \Delta\cdot\sum_{n=1}^{N} e^{j\omega n\Delta} f_W(n\Delta).\end{aligned} \tag{2.18}$$

- The justification for the approximation in (b) is straightforward that the step size Δ in the variable w can be chosen small enough such that the difference between the actual integral in the equality a and summation term in the equality b can be made as small as desired. However, the actual choice of Δ is decided by the trade-off between the accuracy and the computational load of the modified version of AM algorithm (to be shown later).

- The justification for the "finite" upper limit of the summation N in the approximation in (b) is the fact that most of the practical fadings observed in the wireless communication have negligible probability distribution for very large fading values. As an example, the PDF of the normalized Rayleigh fading X is given as $f_X(x) = e^{-x}$. Clearly, for this fading, the probability distribution for the very large fade values is negligible. Since, with each run of the AM algorithm, the distribution $f_W(w)$ varies and so does the appropriate choice of N with each run of the algorithm. One way to resolve this issue is to choose N large enough but finite to take care of this reasonable justification in the third equality above. Other possibility could be choosing N by hit and trial procedure.

To compute N sample points of $\Phi_W(\omega)$, we substitute $\omega = k\omega_0$ in (2.18) where $\omega_0 = \frac{2\pi}{N\Delta}$. That is, ω_0 is the step-size in the ω variable. Thus, for $0 \le k \le N-1$, we get

$$\Phi_W[k] := \Phi_W(k\omega_0) = \Psi_V(v_0) + \Delta \cdot \sum_{n=1}^{N} e^{j\frac{2\pi}{N}kn} f_W[n],$$

where $f_W[n] := f_W(n\Delta)$. Notice that

$$\Phi_W[k] = \Phi_W[k+N],$$

i.e. $\Phi_W[k]$ periodic in k over the period N since $e^{j\frac{2\pi}{N}kn} = e^{j\frac{2\pi}{N}(k+N)n}$. Further notice that

$$\begin{aligned}\Phi_W[k] &= \Psi_V(v_0) + \Delta \cdot \sum_{n=1}^{N} e^{j\frac{2\pi}{N}kn} f_W[n], \\ &= \Psi_V(v_0) + \Delta \cdot \left(\sum_{n=1}^{N-1} f_W[n] \cdot e^{j\frac{2\pi}{N}kn} + f_W[N] \cdot e^{j\frac{2\pi}{N}Nk} \right),\end{aligned}$$

$$= \Psi_V(v_0) + \Delta \cdot \sum_{n=0}^{N-1} \widehat{f}_W[n] \cdot e^{j\frac{2\pi}{N}kn}, \; 0 \le k \le N-1,$$

where

$$\widehat{f}_W[n] = \begin{cases} f_W[n] & \text{when} \quad n = 1, \cdots, N-1. \\ f_W[N] & \text{when} \quad n = 0. \end{cases} \tag{2.19}$$

Thus

$$\Phi_W[k] = \Psi_V(v_0) + \Delta \cdot DFT(k : \widehat{f}_W[n], \; 0 \le n \le N-1). \tag{2.20}$$

where, for convenience, we have defined $DFT(k : X[n], 0 \le n \le N-1) := \sum_{n=0}^{N-1} X[n] \cdot e^{j\frac{2\pi}{N}kn}$ i.e. the sample-wise discrete Fourier transform operation, a standard analysis tool in the digital signal processing. For convenience of notation, we shall drop the index k in the above definition of $DFT(\cdot : \cdots)$ and denote the the sampled characteristic function in (2.20) as

$$\Phi_W[k] = \Psi_V(v_0) + \Delta \cdot DFT(\widehat{f}_W[n], \; 0 \le n \le N-1). \tag{2.21}$$

On the other side, the sequence $f_W[n]$ can be computed from the $\Phi_W[k]$ sequence using the inverse DFT relation as following

$$\widehat{f}_W[n] = \frac{1}{N\Delta} \sum_{k=0}^{N-1} \Big(\Phi_W[k] - \Psi_V(v_0) \Big) e^{-j\frac{2\pi}{N}kn}, \; 0 \le n \le N-1.$$

- ***Computing PDF of W*:** For each value of $w = n\Delta$ (greater than 0), we compute $f_W(n\Delta)$ as following:

$$f_W(n\Delta) = f_V(W^{-1}(n\Delta)) \left. \frac{d\,(W^{-1}(w))}{dw} \right|_{w=n\Delta},$$

$$= f_V(W^{-1}(n\Delta)) \lim_{\Delta \to 0} \left. \frac{W^{-1}(w+\Delta) - W^{-1}(w)}{(w+\Delta) - w} \right|_{w=n\Delta},$$

$$\stackrel{(d)}{\approx} \; f_V(W^{-1}(n\Delta)) \, \frac{W^{-1}((n+1)\Delta) - W^{-1}(n\Delta)}{\Delta}, \; 1 \leq n \leq N, \tag{2.22}$$

where the approximation in the equality (d) above is possible to the desired accuracy by choosing the appropriate step size Δ in the w variable.

- ***Computing the characteristic function of*** Z: Since

$$\begin{aligned} \Phi_Z(\omega) &= \Phi_W(\omega)^K, \; \forall\, \omega, \\ \Rightarrow \quad \Phi_Z(k\omega_0) &= \Phi_W(k\omega_0)^K, \; 0 \leq k \leq N-1, \\ \Rightarrow \text{ i.e. } \; \Phi_Z[k] &= \Phi_W[k]^K, \; 0 \leq k \leq N-1. \end{aligned} \tag{2.23}$$

However, to compute the $\Phi_Z[k]$ sequence using (2.23), it is crucial to understand an important omission we made while writing (2.21). To explain this, recall that

$$Z = \sum_{i=1}^{K} W_i,$$

where we have already discretized and assumed the range of W large enough denoted by $[\Delta, N\Delta]$. It is obvious that the range of Z will be approximately K times greater than that of W. Precisely, the range of Z is $[K\Delta, NK\Delta]$ with samples spaced apart by Δ. To be able to use the DFT transform to compute the distribution function of Z, notice the exact number of sample points in $Z[k]$ sequence is

$$\widehat{N} := NK - K + 1.$$

Thus, it is required to perform $\widehat{N}$-point inverse DFT of the $\Phi_W[k]^K$ sequence rather than N-point inverse DFT. To achieve $\widehat{N}$-point inverse DFT of the $\Phi_W[k]^K$, it is suffice to do zero-padding of the $f_W[n]$ sequence at the end to extend the sequence length to reach $\widehat{N}$ and then compute $\widehat{N}$-point DFT of $f_W[n]$. Therefore, (2.23) is corrected as

$$\text{i.e. } \; \Phi_Z[k] = \Phi_W[k]^K, \; 0 \leq k \leq \widehat{N}-1, \tag{2.24}$$

where $\Phi_W[k]$ sequence is a $\widehat{N}$-point sequence computed as

$$\Phi_W[k] \;=\; \Psi_V(v_0) \,+\, \Delta \cdot DFT(k : \widehat{f}_W[n],\; 0 \leq n \leq \widehat{N}-1). \tag{2.25}$$

- ***Numerical Computation of PDF of Z*:** In our problem setup, Z is a continuous random variable. Further we need to find its distribution $f_Z(z)$ numerically from the corresponding characteristic function, which is numerically computed in the previous step for a finite number of equally distant points. Consider the relation between the characteristic function and distribution function of Z in the continuous domain:

$$\begin{aligned} f_Z(z) \;&=\; \frac{1}{2\pi}\int_{-\infty}^{\infty} \Phi_Z(\omega) e^{-j\omega z}\, d\omega\,, \\ &\approx\; \frac{\delta_\omega}{2\pi} \sum_{k\,=\,-M}^{M-1} \Phi_Z(k\delta_\omega)\, e^{-jk\delta_\omega z}, \end{aligned} \tag{2.26}$$

where the continuous variable ω is discretized with samples spaced δ_ω apart and the sum neglects the integral term outside the interval $[-M\delta_\omega,\, M\delta_\omega)$.

To be able to use the $\Phi_Z[k]$ sequence computed already, we set $\delta_\omega \;=\; \dfrac{2\pi}{\widehat{N}\Delta} \;=\; \omega_0$ and $2M \;=\; \widehat{N}$. Notice that the truncated random variable Z lies in the range $[K\Delta,\, NK\Delta]$. With the substitution $z \;=\; n\Delta$ in (2.26), we get

$$f_Z(n\Delta) \;\approx\; \frac{1}{\widehat{N}\Delta} \sum_{k\,=\,-\widehat{N}/2}^{\widehat{N}/2-1} \Phi_Z(k\omega_0) e^{-j\frac{2\pi}{\widehat{N}}kn},\;\; K \,\leq\, n \,\leq\, NK,$$

$$\text{i.e.}\quad f_Z[n] \;\approx\; \frac{1}{\widehat{N}\Delta} \sum_{k\,=\,-\widehat{N}/2}^{\widehat{N}/2-1} \Phi_Z[k] e^{-j\frac{2\pi}{\widehat{N}}kn},\;\; K \,\leq\, n \,\leq\, NK,$$

where $f_Z[n] \;:=\; f_Z(n\Delta)$.

Recall that $\Phi_Z[k] \;=\; \Phi_W[k]^K$ and $\widehat{N}$-point $\Phi_W[k]$ sequence is periodic in k with period $\widehat{N}$ and therefore, the $\Phi_Z[k]$ sequence is also periodic with period $\widehat{N}$ i.e.

$$\Phi_Z[k] \;=\; \Phi_Z[k+\widehat{N}],$$

and $e^{-j\frac{2\pi}{\widehat{N}}kn} = e^{-j\frac{2\pi}{\widehat{N}}(k+\widehat{N})n}$ also holds. Thus, for $K \leq n \leq NK$, rewriting

$$\begin{aligned}
f_Z[n] &\approx \frac{1}{\widehat{N}\Delta}\left(\sum_{k=0}^{\widehat{N}/2-1}\Phi_Z[k]e^{-j\frac{2\pi}{N}kn} + \sum_{k=-\widehat{N}/2}^{-1}\Phi_Z[k]e^{-j\frac{2\pi}{N}kn}\right), \\
&= \frac{1}{\widehat{N}\Delta}\left(\sum_{k=0}^{\widehat{N}/2-1}\Phi_Z[k]e^{-j\frac{2\pi}{N}kn} + \sum_{k=-\widehat{N}/2}^{-1}\Phi_Z[k+\widehat{N}]e^{-j\frac{2\pi}{N}kn}\right), \\
&= \frac{1}{\widehat{N}\Delta}\left(\sum_{k=0}^{\widehat{N}/2-1}\Phi_Z[k]e^{-j\frac{2\pi}{N}kn} + \sum_{k=\widehat{N}/2}^{\widehat{N}-1}\Phi_Z[k]e^{-j\frac{2\pi}{N}kn}\right), \\
&:= \frac{1}{\widehat{N}\Delta}\sum_{k=0}^{\widehat{N}-1}\Phi_Z[k]e^{-j\frac{2\pi}{N}kn}, \ K \leq n \leq NK. \qquad (2.27)
\end{aligned}$$

where

$$\widehat{\Phi}_Z[k] = \begin{cases} \Phi_Z[k] & \text{when} \quad 0 \leq k \leq \widehat{N}/2-1. \\ \\ \Phi_Z[k-\widehat{N}] & \text{when} \quad \widehat{N}/2 \leq k \leq \widehat{N}-1. \end{cases} \qquad (2.28)$$

It is obvious that

$$f_Z[n] \approx \frac{1}{\Delta}\cdot IDFT(n:\Phi_Z[k], 0 \leq k \leq \widehat{N}-1),$$

where, for convenience, we have defined the inverse of the point-wise discrete Fourier transform as $IDFT(n:X[k], 0 \leq k \leq N-1) := \sum_{k=0}^{N-1} x[k]\cdot e^{-j\frac{2\pi}{N}kn}$.

Remark 8. *By choosing $\widehat{N}$ to be the nearest power of 2, we can compute the characteristic function of the random variable W as well as distribution function of Z very efficiently by implementing the* DFT *and inverse* DFT *operations using the FFT algorithms. For example, assuming sequence length N is some exponential of 2, direct computation of discrete Fourier transform of a N-point sequence involves N^2 complex multiplications, while computing the same using fast Fourier transform (FFT) algorithm requires only $N\log N$ complex multiplications. Clearly, the computational gain is increasing with N, e.g. for $N = 1024$ length sequence, FFT algorithm is approximately 100 times faster than the direct DFT method.*

Remark 9. *It is important to remember that to compute the PDF of Z from the PDF of W, we need to do zero-padding to $f_W(n\Delta)$ (see Section 7.6 in [31])*

Finally, we summarize the required computational steps as follows:

1. Compute the samples of characteristic function $\Phi_X(\omega)$ as follows

$$\Phi_X[k] \;=\; \Psi_V(v_0) + \Delta \cdot \mathrm{DFT}(\,\widehat{f}_X[n]\,,\; 0 \,\le\, n \,\le\, N-1\,).$$

2. Compute the samples of characteristic function $\Phi_Y(\omega)$ as follows

$$\Phi_Y[k] \;=\; \Phi_X[k]^K,\; 0 \,\le\, k \,\le\, N-1\,.$$

3. Compute inverse transform: $\mathscr{F}^{-1}\,\Phi_Y(\omega)$ to obtain $\Psi_Y(y)$ as follows

$$f_Y[n] \quad\approx\; \frac{1}{\Delta} \cdot \mathrm{IDFT}(\Phi_Y[k],\; 0 \,\le\, k \,\le\, N-1)\,.$$

We now combine the fast Fourier transform (FFT) based digital processing presented in the section above with the alternating maximization (AM) method proposed in [30] for faster computation of the decentralized MAC utilities such as optimal transmitter power control and sum-capacity.

Algorithm `FFT-AM algorithm for Decentralized MAC`

Initialization: Initialize $\widehat{P}(\cdot)$ observing average power constraint (say with standard water-filling). Initialize KKT multiplier λ, error tolerance ε and increment step size δ for the λ parameter and Δ as the discretization interval for the variable $v\widehat{P}(v)$. Choose a large N such that N is some power of 2.

Repeat

1. Compute the distribution $f_Z(z)$ numerically as follows:

- 1(a): Compute the N-length sequence $f_W[n]$ by sampling the $f_W(w)$ given as

$$f_W(w) = \Psi_V(v_0)\cdot\delta(w) + f_V(W^{-1}(w))\frac{d\,(W^{-1}(w))}{dw}\,\mathbb{1}_{\{w>0\}}.$$

This requires computation of $W^{-1}(w)$ function.

- 1(b): Compute the N-length sequence $\Phi_W[k]$ as

$$\Phi_W[k] = \Psi_V(v_0) + \Delta\cdot\mathrm{FFT}(\widehat{f_W}[n],\ 0\le n\le N-1),$$

for $0\le k\le N-1$, where $\widehat{f_W}[n]$ is a rearranged version of $f_W[n]$, see (2.19).

- 1(c): Compute the N-length sequence $\Phi_Z[k]$ as

$$\Phi_Z[k] = \Phi_W[k]^K,\ 0\le k\le N-1.$$

- 1(d): Compute $f_Z[n] \approx \frac{1}{\Delta}\cdot\mathrm{IFFT}(\Phi_Z[k],\ 0\le k\le N-1)$, for $0\le n\le N-1$.

2. Compute the partial optimal:

$$\widehat{P}^* = \arg\max_P\ \mathbb{E}\left[\log(1+vP(v)+Z)\right],$$

using the discrete KKT condition (see (2.16)):

$$\Delta\cdot\sum_{n=0}^{N-1}\frac{f_Z[n]}{1+v\widehat{P}^*(v)+n\Delta} = \frac{\lambda}{v},$$

where $Z=\sum_i v_i\widehat{P}(v_i)$.

3. Find $\bar{P}_{\mathrm{avg}} = \int\widehat{P}^*(v)\,d\Psi(v)$.

4. If $\begin{cases} \left(P_{\text{avg}} - \bar{P}_{\text{avg}}\right) > \varepsilon, \text{ then } \lambda = \lambda - \delta; \texttt{ goto step 1} \\ \left(P_{\text{avg}} - \bar{P}_{\text{avg}}\right) < -\varepsilon, \texttt{ then } \lambda = \lambda + \delta; \texttt{ goto step 1} \end{cases}$

Until the sum rate converges.

2.3.2 AM algorithm: Issues & Resolutions

We have identified an issue in the proposed AM algorithm in [30] that can severely affect its convergence particularly in the case when there are large number of MAC users and or when the average powers at the transmitter side are huge. To explain this, we recall step for the computation of the partial optimal in the AM algorithm i.e. (2.16) as follows:

$$\int_0^\infty \frac{f_Z(z)\,dz}{1 + vP(v) + z} = \frac{\lambda}{v}.$$

Since the partial optimal $P(v)$ function is increasing in the argument, it is easy to deduce that the LHS of the above equation decreases as v increases. To solve this equation numerically for $P(v)$, it is important, in the algorithm, to know beforehand the maximum that the function $vP(v)$ can take. Too small the choice of $vP(v)$ in the algorithm will result in sub-optimal solution. At the other extreme, choosing maximum of $vP(v)$ too large will result in excess of computational requirements. Thus, a good estimate of max. of $vP(v)$ is necessary. This will depend upon both the distribution of v as well as the max. of the optimal partial $P(v)$.

For popular or typical fading conditions, the distribution of the fading magnitude is negligible for large values, e.g. such as Rayleigh fading where $f_V(v) = e^{-v}$ becomes negligible for $v > 12$ etc. In our numerical computations to be presented in the next section, we have assumed Rayleigh distribution with v ranging between 0 to 20 (approx.).

Estimating the maximum of $P(v)$ safely is a bit tedious task. One way is to do hit and trial approach. To avoid this, we have identified a useful condition. It is not difficult to see from the KKT condition that the optimal partial $P(v)$ satisfies $\lim_{v\to\infty} P(v) \geq \frac{1}{\lambda}\cdot$ This lower bound on the maximum of the optimal $P(v)$ has turned out to be very useful in our numerical computations. We have used this lower bound together with v maximum as large as 20 to determine the the maximum of the $vP(v)$.

2.4 Identical Users MAC with Large Users

In our decentralized *identical users* fading MAC, the total receiver signal power, denoted by Z, is sum of the i.i.d. received signal powers of each of the individual transmitters denoted by W_i, $1 \leq i \leq K$. Recall $Z = \sum_i W_i$ where $W_i = v_i P_i(v_i)$. Invoking Central Limit Theorem (CLT) for the decentralized MAC with large number of users, we assert that Z is a Gaussian random variable with say μ_Z and σ_Z^2 as mean and variance parameters of the distribution respectively i.e. $Z \sim \mathcal{N}(\mu_Z, \sigma_Z^2)$. For reference, we state the CLT theorem next.

Theorem 10. *Given a set of i.i.d. random variables, denoted as $W_1, W_2, , \cdots, W_K$, with mean $E[W_i] = \mu_W < \infty$ and finite non-zero σ_W^2, then the random variable*

$$\widehat{Z}_K = \frac{(W_1 + W_2 + \cdots + W_K) - K \cdot \mu_W}{\sqrt{K} \cdot \sigma_W} \tag{2.29}$$

converges in distribution to the standard Gaussian random variable as K approaches infinity. In other words

$$\lim_{K \to \infty} \text{Prob}(\widehat{Z}_K \leq z) = \Phi(z), \forall z \in \mathbb{R}, \tag{2.30}$$

where $\Phi(z) \triangleq \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} e^{-\frac{u^2}{2}} du$ is the standard normal CDF.

Remark 11. *Generally, the distribution of the sum of i.i.d. random variables begins to approach the Gaussian shape ("bell curve") even for a relatively small number K. Thus, once the distribution of the sum is close enough (to the desired accuracy) to the Gaussian for a certain K say K_0, Theorem 10 suggests that it is safe to assume that the distribution of the sum for $K \geq K_0$ i.i.d. variables is Gaussian with sufficient accuracy.*

Under this condition, computation of the distribution of Z gets a lot easier as it requires only the determination of two parameters, namely the mean μ_Z and the variance σ_Z^2. Recall the PDF of W as following

$$f_X(w) = \Psi_V(v_0) \cdot \delta(w) + f_V(W^{-1}(w)) \frac{d\,(W^{-1}(w))}{dw} \cdot \mathbb{1}_{\{W>0\}}. \tag{2.31}$$

We denote the second component of $f_W(w)$ in (2.31) as $\widehat{f}_W(w)$. Thus

$$\begin{aligned}
\mu_W &= \mathbb{E}\,[W], \\
&= \int_0^\infty w f_W(w)\,dw, \\
&= \int_0^\infty w(\Psi_V(v_0)\delta(w) + \widehat{f}_W(w))\,dw, \\
&= \int_{0+}^\infty w\widehat{f}_W(w)\,dw, \\
&\approx \Delta \cdot \sum_{n=1}^{N} \widehat{f}_W(n\Delta) \cdot n\Delta, \\
&= \Delta^2 \cdot \sum_{n=1}^{N} \widehat{f}_W(n\Delta) \cdot n.
\end{aligned} \tag{2.32}$$

The numerical computation of σ_W^2 is as following

$$\begin{aligned}
\sigma_W^2 &= \mathbb{E}\,[W^2] - \mu_W^2, \\
&= \int_0^\infty w^2 f_W(w)\,dw - \mu_W^2, \\
&= \int_0^\infty w^2(\Psi_V(v_0)\delta(w) + \widehat{f}_W(w))\,dw - \mu_W^2, \\
&= \int_{0+}^\infty w^2\widehat{f}_W(w)\,dw - \mu_W^2, \\
&\approx \Delta \cdot \sum_{n=1}^{N} \widehat{f}_W(n\Delta) \cdot n^2\Delta^2 - \mu_W^2, \\
&= \Delta^3 \cdot \sum_{n=1}^{N} \widehat{f}_W(n\Delta) \cdot n^2 - \mu_W^2.
\end{aligned} \tag{2.33}$$

The mean and variance parameters of the random variable Z are computed numerically as shown below. Firstly, the mean is computed as

$$\begin{aligned}
\mu_Z &= \mathbb{E}\,[Z] \\
&= \mathbb{E}\,[\sum_{i=1}^{K} W_i] \\
&= K\,\mu_W \\
&= K\,\Delta^2 \cdot \sum_{n=1}^{N} \widehat{f}_W(n\Delta) \cdot n\,.
\end{aligned} \tag{2.34}$$

Similarly, the variance is determined as

$$\begin{aligned}
\sigma_Z^2 &= \mathbb{E}\,[(Z-\mu_Z)^2] \\
&= K\,\mathbb{E}\,[(W-\mu_W)^2] \\
&= K\,\sigma_W^2 \\
&= K\,\Delta^3 \cdot \sum_{n=1}^{N} \widehat{f}_W(n\Delta) \cdot n^2 - K\,\mu_W^2\,.
\end{aligned} \tag{2.35}$$

Recall the KKT condition (2.13) for the partial optimal power scheme:

$$\int \frac{dF_Z(z)}{1 + vP(v) + z} = \frac{\lambda}{v},\ \forall\, v > v_0\,. \tag{2.36}$$

For large MAC, choosing the Gaussian distribution in the above condition, we get

$$\frac{1}{\sqrt{2\pi\sigma_Z^2}} \int_{-\infty}^{\infty} \frac{e^{-\frac{(z-\mu_Z)^2}{2\sigma_Z^2}}\ dz}{1 + vP(v) + z} = \frac{\lambda}{v},\ \forall\, v > v_0\,. \tag{2.37}$$

Together with the property (2.14) and that LHS in (2.37) is a single variable integral, and the mean and variance parameters determined from (2.34) and (2.35) respectively, (2.37) can be easily solved numerically for the partially optimal power $P(v)$.

2.5 Numerical Results

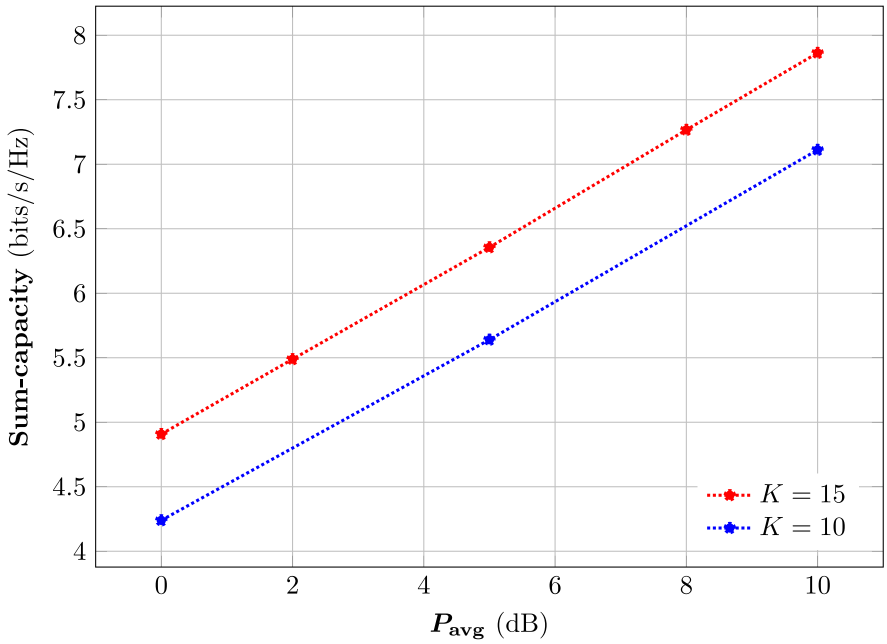

Figure 2.2: Ergodic Sum-capacity of $K = 10, 15$ decentralized *identical users* MAC with fading on all the links as Rayleigh distributed.

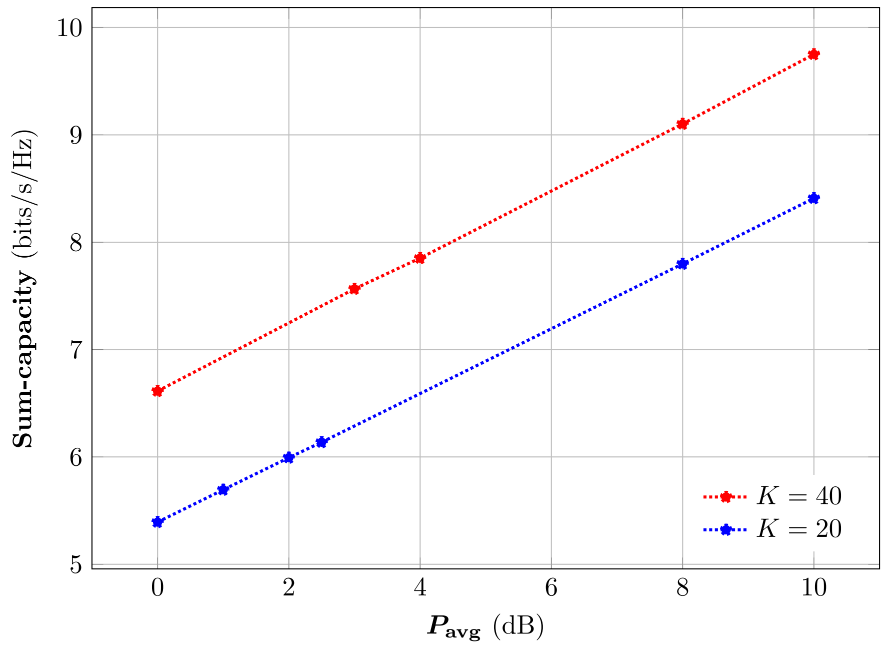

Figure 2.3: Ergodic Sum-capacity of $K = 20, 40$ decentralized *identical users* MAC with fading on all the links as Rayleigh distributed.

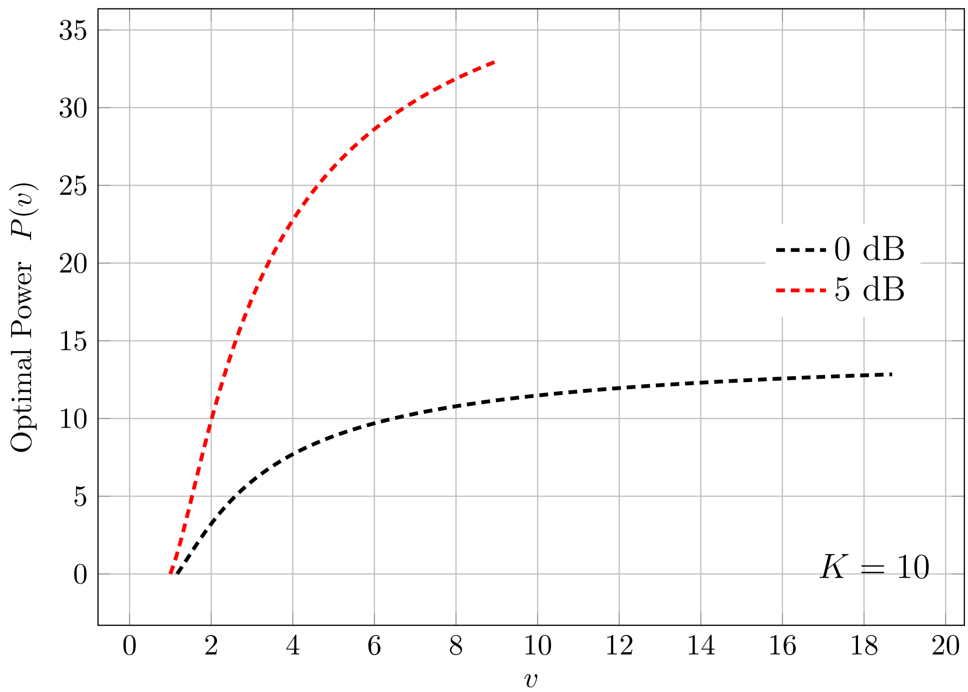

Figure 2.4: Optimal power schemes for $K = 10$ decentralized *identical users* MAC with fading on all the links as Rayleigh distributed.

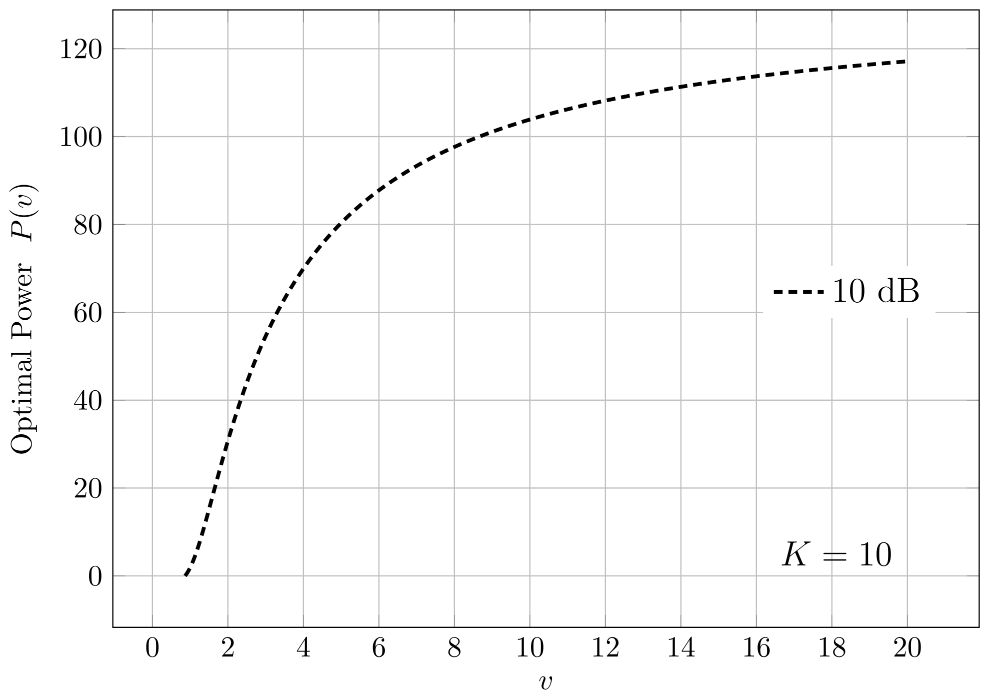

Figure 2.5: Optimal power schemes for $K = 10$ decentralized *identical users* MAC with fading on all the links as Rayleigh distributed.

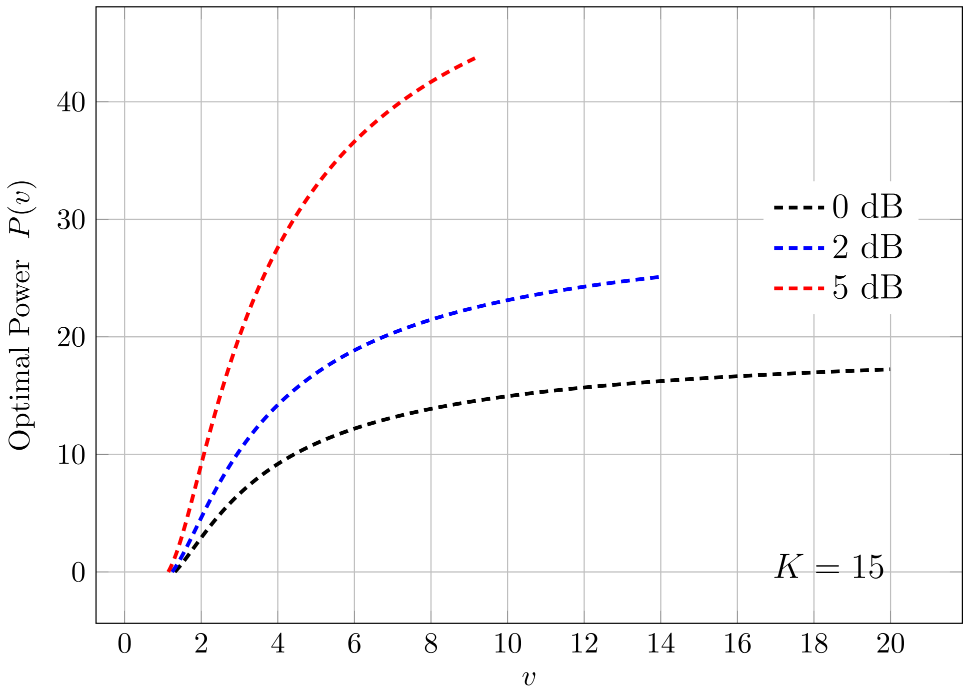

Figure 2.6: Optimal power schemes for $K = 15$ decentralized *identical users* MAC with fading on all the links as Rayleigh distributed.

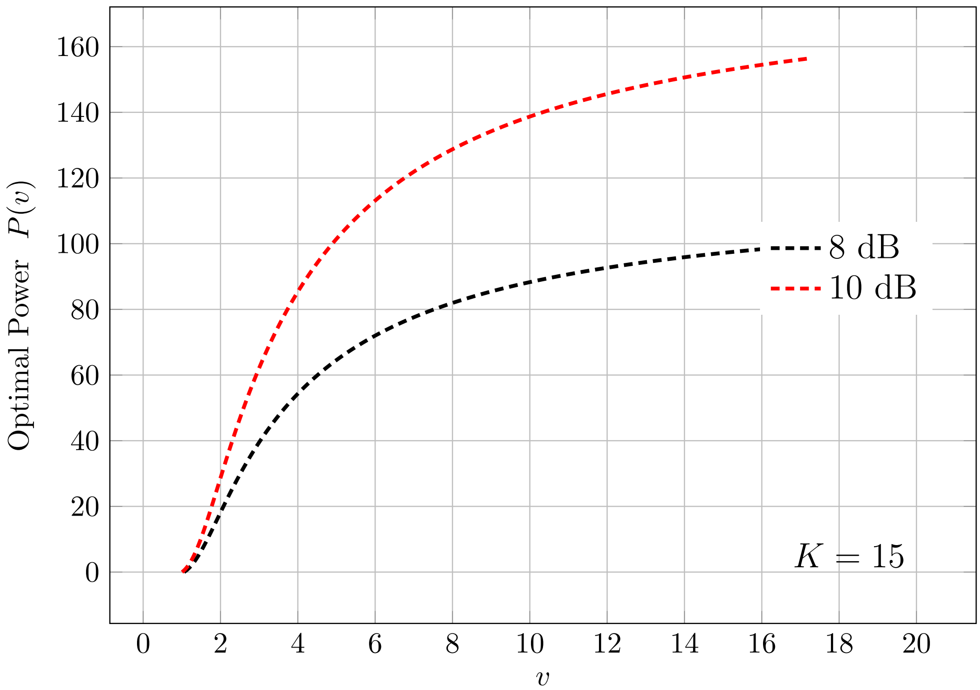

Figure 2.7: Optimal power schemes for $K = 15$ decentralized *identical users* MAC with fading on all the links as Rayleigh distributed.

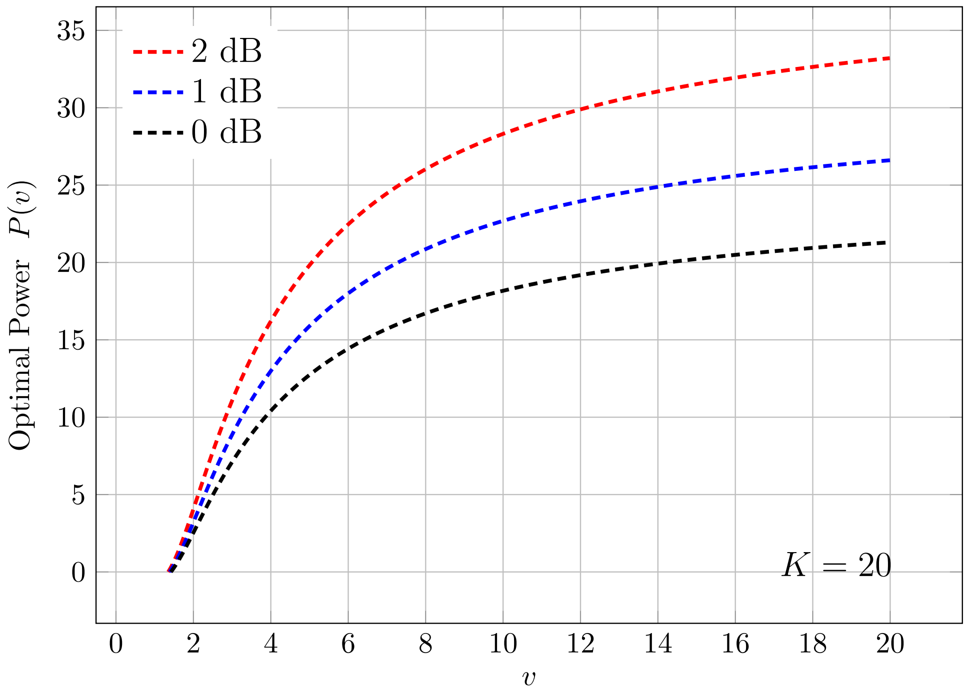

Figure 2.8: Optimal power schemes for $K = 20$ decentralized *identical users* MAC with fading on all the links as Rayleigh distributed.

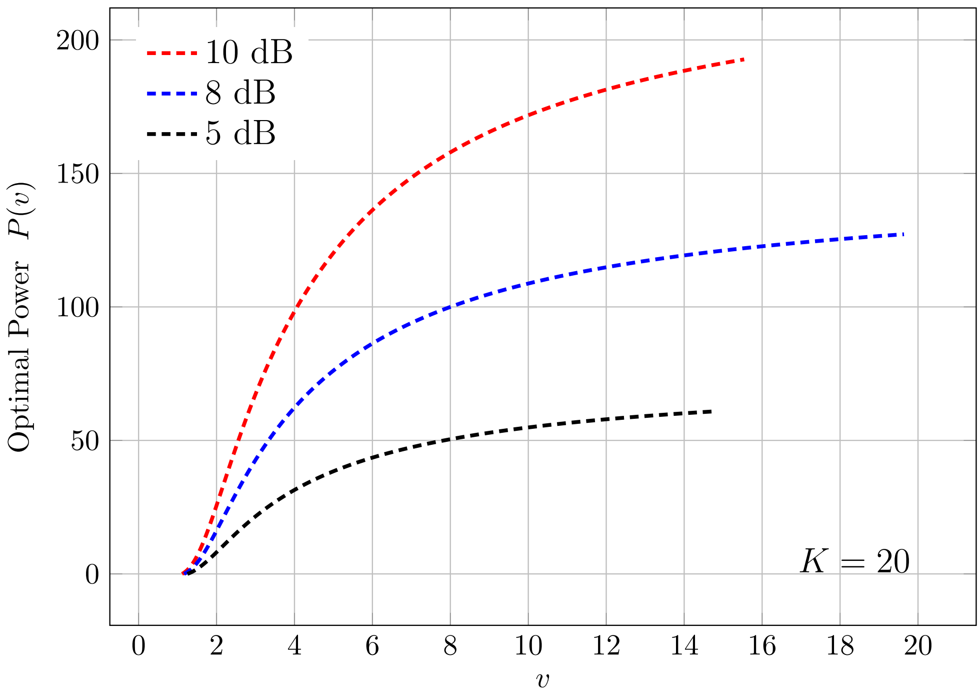

Figure 2.9: Optimal power schemes for $K = 20$ decentralized *identical users* MAC with fading on all the links as Rayleigh distributed.

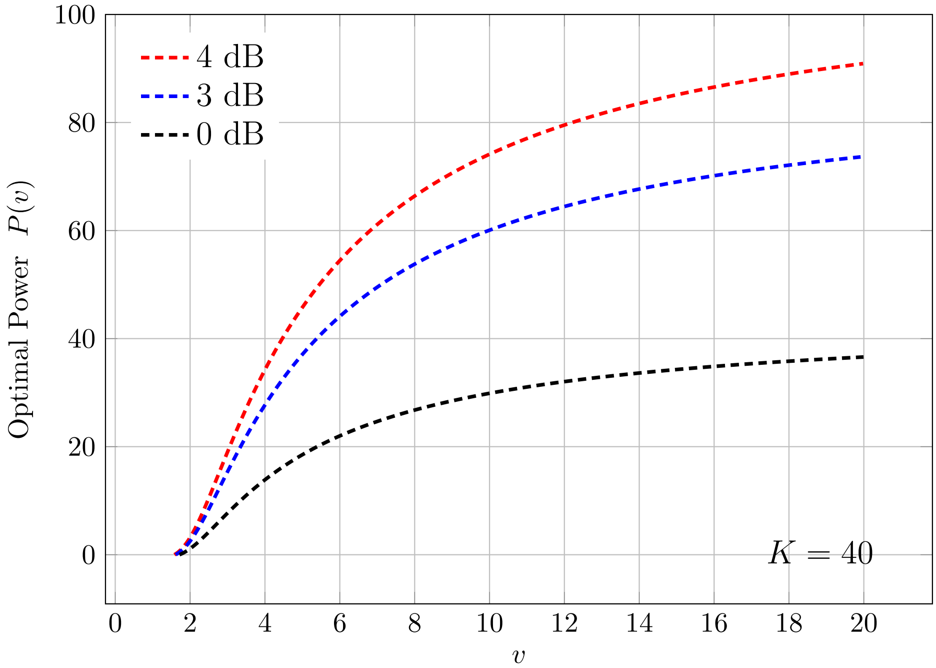

Figure 2.10: Optimal power schemes for $K = 40$ decentralized *identical users* MAC with fading on all the links as Rayleigh distributed.

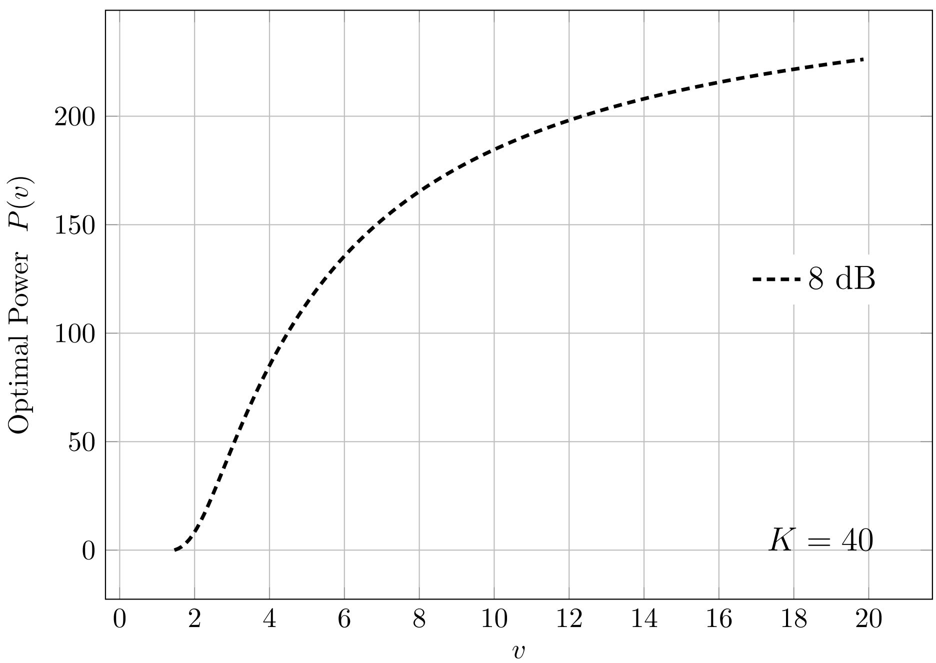

Figure 2.11: Optimal power schemes for $K = 40$ decentralized *identical users* MAC with fading on all the links as Rayleigh distributed.

	P_{avg} (dB)	C_{sum} (bits/s/Hz)
$K = 10$	0	4.239702
	5	5.639849
	10	7.111351
$K = 15$	0	4.908040
	2	5.488721
	5	6.355934
	8	7.266379
	10	7.862811
$K = 20$	0	5.392953
	1	5.694281
	2	5.993126
	2.5	6.137255
	8	7.798280
	10	8.408797
$K = 40$	0	6.613407
	3	7.565446
	4	7.851965
	8	9.102039
	10	9.750037

Table 2.1: Summary table for the Sum-capacity computations presented in the Figure 2.2 and Figure 2.3.

Chapter 3

Conclusion

We investigated the performance of the AM algorithm proposed for the optimal decentralized powers in Gaussian fading MAC [30]. Firstly, we identified the bottleneck in the algorithm; computation of the distribution of the received SNR is tedious and inefficient. Secondly, to enhance the performance in terms of required computations, we utilized the fast Fourier transform (FFT) algorithm to compute the probability distributions involved in the partial optimization step in the AM algorithm.

For the special case of *identical* users fading MAC with large number of users, we approximated the probability distribution of the total received SNR by the Gaussian. This is done in line with the Central Limit Theorem (CLT). This approximation is proven to be excellent for large users and verified by comparing the actual and approximated distributions for some large MAC users. This revelation have enabled us to compute the ergodic sum-capacity of large identical users MAC of size as large as up to 40 users so far in the presence of IID Rayleigh fadings. These numerical results are extremely significant given that prior to our work, the optimal sum-rates for the same MAC configuration are known only for up to 4 MAC users, see Figure 2(b) in [30]. In [30], the convergence of the proposed AM algorithm to the desired accuracy is based on the comparing the actual and estimated average tranmitter power. This seems as an another bottleneck in the proposed AM algorithm. Future work will include finding a more efficient criteria for convergence of this algorithm.

Although general version of the CLT suggests the Gaussian approximation even for the 'large' *non-identical* users MAC, it seems difficult to determine how the accuracy of the

Gaussian approximation of the probability distribution for large MAC vary as there are a lot of known and unknown factors involved unlike the identical users MAC. Nevertheless, the decentralized ‘large’ *non-identical* users MAC is an equally important model in some cases.

References

[1] C. E. Shannon, “A mathematical theory of communication,” *Bell system technical journal*, vol. 27, no. 3, pp. 379–423, 1948.

[2] A. J. Goldsmith and P. P. Varaiya, “Capacity of fading channels with channel side information,” *IEEE transactions on information theory*, vol. 43, no. 6, pp. 1986–1992, 1997.

[3] R. S. Cheng and S. Verdú, “Gaussian multiaccess channels with ISI: Capacity region and multiuser water-filling,” *IEEE Transactions on Information Theory*, vol. 39, no. 3, pp. 773–785, 1993.

[4] E. Telatar, “Capacity of multi-antenna gaussian channels,” *European transactions on telecommunications*, vol. 10, no. 6, pp. 585–595, 1999.

[5] N. David, P. Viswanath, and L. Zheng, “Diversity-multiplexing tradeoff in multiple-access channels.” *IEEE Trans. Information Theory*, vol. 50, no. 9, pp. 1859–1874, 2004.

[6] A. El Gamal and Y.-H. Kim, *Network information theory*. Cambridge university press, 2011.

[7] R. Ahlswede, “Multi-way communication channels,” in *Second International Symposium on Information Theory: Tsahkadsor, Armenia, USSR, Sept. 2-8, 1971*, 1973.

[8] H. Liao, “Multiple access channels,” University of Hawaii, Tech. Rep., 1972.

[9] R. Gallager, "A perspective on multiaccess channels," *IEEE Transactions on Information Theory*, vol. 31, no. 2, pp. 124–142, March 1985.

[10] S. Arimoto, "An algorithm for computing the capacity of arbitrary discrete memoryless channels," *IEEE Transactions on Information Theory*, vol. 18, no. 1, pp. 14–20, January 1972.

[11] R. Blahut, "Computation of channel capacity and rate-distortion functions," *IEEE transactions on Information Theory*, vol. 18, no. 4, pp. 460–473, 1972.

[12] S. Boyd and L. Vandenberghe, *Convex optimization*. Cambridge university press, 2004.

[13] D. Slepian and J. K. Wolf, "A coding theorem for multiple access channels with correlated sources," *Bell System Technical Journal*, vol. 52, no. 7, pp. 1037–1076, 1973.

[14] T. Cover, A. E. Gamal, and M. Salehi, "Multiple access channels with arbitrarily correlated sources," *IEEE Transactions on Information theory*, vol. 26, no. 6, pp. 648–657, 1980.

[15] P. Bergmans and T. Cover, "Cooperative broadcasting," *IEEE Transactions on Information Theory*, vol. 20, no. 3, pp. 317–324, 1974.

[16] A. Wyner, "Recent results in the shannon theory," *IEEE Transactions on information Theory*, vol. 20, no. 1, pp. 2–10, 1974.

[17] G. S. Poltyrev, "Coding in an asynchronous multiple-access channel," *Problemy Peredachi Informatsii*, vol. 19, no. 3, pp. 12–21, 1983.

[18] J. Hui and P. Humblet, "The capacity region of the totally asynchronous multiple-access channel," *IEEE Transactions on Information Theory*, vol. 31, no. 2, pp. 207–216, 1985.

[19] T. M. Cover, R. J. McEliece, and E. C. Posner, "Asynchronous multiple-access channel capacity," *IEEE Transactions on Information Theory*, vol. 27, no. 4, pp. 409–413, 1981.

[20] S. Verdu, "Multiple-access channels with memory with and without frame synchronism," *IEEE Transactions on Information Theory*, vol. 35, no. 3, pp. 605–619, 1989.

[21] B. Rimoldi and R. Urbanke, "A rate-splitting approach to the gaussian multiple-access channel," *IEEE Transactions on Information Theory*, vol. 42, no. 2, pp. 364–375, 1996.

[22] S. Verdu, "The capacity region of the symbol-asynchronous gaussian multiple-access channel," *IEEE Transactions on Information Theory*, vol. 35, no. 4, pp. 733–751, 1989.

[23] R. S. Cheng and S. Verdu, "The effect of asynchronism on the total capacity of gaussian multiple-access channels," *IEEE transactions on information theory*, vol. 38, no. 1, pp. 2–13, 1992.

[24] S. Verdu, "Minimum probability of error for asynchronous gaussian multiple-access channels," *IEEE transactions on information theory*, vol. 32, no. 1, pp. 85–96, 1986.

[25] R. Knopp and P. A. Humblet, "Information capacity and power control in single-cell multiuser communications," in *Communications, 1995. ICC'95 Seattle,'Gateway to Globalization', 1995 IEEE International Conference on*, vol. 1. IEEE, 1995, pp. 331–335.

[26] A. Das and P. Narayan, "Capacities of time-varying multiple-access channels with side information," *IEEE Transactions on Information Theory*, vol. 48, no. 1, pp. 4–25, 2002.

[27] S. Shamai and E. Telatar, "Some information theoretic aspects of decentralized power control in multiple access fading channels," in *2013 IEEE International Symposium on Information Theory*, 1999, pp. 23–23.

[28] K. Singh and S. R. B. Pillai, "Decentralized power adaptation in ergodic fading multiple access channels," in *Communications (NCC), 2015 Twenty First National Conference on*. IEEE, 2015, pp. 1–6.

[29] N. Mital, K. Singh, and S. R. B. Pillai, "On the ergodic sum-capacity of decentralized multiple access channels," *IEEE Communications Letters*, vol. 20, no. 5, pp. 854–857, 2016.

[30] K. Singh, “Optimal power control in decentralized gaussian multiple access channels,” *IEEE Communications Letters*, vol. 22, no. 9, pp. 1938–1941, 2018.

[31] A. Leon-Garcia, *Probability and Random Processes for Electrical Engineering*. Pearson Education India, 1994.