

Power Allocation for Compute-and-Forward over Fading Channels

Lanwei Zhang, *Student Member, IEEE*, Jamie Evans, *Senior Member, IEEE*,
and Jingge Zhu, *Member, IEEE*.

Abstract

Compute-and-forward (CF) is a relaying strategy which allows the relay to decode a linear combination of the transmitted messages. This work studies the optimal power allocation problem for the CF scheme in fast fading channels for maximizing the symmetric computation rate, which is a non-convex optimization problem with no simple analytical or numerical solutions. In the first part of the paper, we investigate the problem when there are finitely many channel states (discrete case). We establish several important properties of the optimal solutions and show that if all users share the same power allocation policy (symmetric policy), the optimal solution takes the form of a water-filling type when the power constraint exceeds a certain threshold. However, if asymmetric policies are allowed, the optimal solution does not take this form for any power constraint. We propose a low-complexity order-based algorithm for both scenarios and compare its performance with baseline algorithms. In the second part of the paper, we state relevant results when the channel coefficients are modelled as continuous random variables (continuous case) and propose a similar low-complexity iterative algorithm for the symmetric policy scenario. Numerical results are provided for both discrete and continuous cases. It is shown that in general our proposed algorithm finds good suboptimal solutions with low complexity, and for some examples considered, finds an exact optimal solution.

Index Terms

Compute-and-forward, Fading channel, Power allocation, non-convex optimization.

I. INTRODUCTION

In a wireless network, the signal sent by a transmitter is not only received by the intended receiver but also by surrounding nodes due to the broadcast nature of the medium. In the same way, the received signal at a node will include both the desired message and the unwanted interference signal. Traditionally, orthogonal transmission schemes are applied to avoid interference

so that the receiver only receive the transmitted signal from the desired transmitter. However, these methods will suffer from a diminishing rate, especially when many nodes want to send messages simultaneously [1].

If the transmitters and the receivers cannot communicate with each other directly, some intermediate nodes can be used as relays to help their communication. Conventional relay strategies include amplify-and-forward and decode-and-forward where either a simple analog operation is used, or the messages from different transmitters are treated "separately" in a decoding process. The discovery of network coding [2] shows that the traditional relay strategies are not sufficient to achieve optimal throughput. To address the same problem over noisy channels, physical-layer network coding (PNC) was proposed in [3]. It is a non-orthogonal approach which exploits the physical nature of the electromagnetic waves. The relays will recover a set of functions of the signals from multiple transmitters and then learn the desired messages from the functions.

Compute-and-forward (CF) is a linear PNC scheme proposed in [4]. It allows the relay to decode an integer linear combination of the messages from multiple transmitters, and either forward it to the next receiver or combine it with other linear combinations to recover individual messages. To achieve this, nested lattice codes are used to guarantee that any integer linear combination of the codewords is still a codeword. Due to the favourable algebraic property of lattice codes, the idea of CF has been applied to different AWGN network models and results in good performances that cannot be achieved otherwise [4]–[10]. Due to the difference between the channel coefficients and the integer coefficients of the linear combinations, the achievable computation rate to decode the linear combination will have a loss. One can allocate different transmission powers to different transmitters in order to match the channel coefficients to the integers if the channel state information (CSI) is known to the transmitters. When considering fading channels, where the channel coefficients are modelled as random variables (RV), one should further decide judiciously whether the transmission should happen and how much power should be allocated for different channel states to maximize the average transmission rate. However, since the rate expression of the CF scheme is non-smooth and non-convex, the resulting optimization problem is thus non-convex. In this paper, we will explore the characteristics of the optimal solutions to this non-convex problem and discuss several low-complexity algorithms.

A. Related work

There is a rich collection of work on physical-layer network coding and the compute-and-forward scheme in many network scenarios. It has been shown in [3] that for a two-user two-way relay network, the PNC scheme has nearly twice the throughput of the orthogonal scheme. It is achieved by decoding the sum of the transmitted signals at the relay and sending out the sum. Since both users know their own messages, they can learn the message of the other user using the sum.

By using the CF scheme, the relay will recover the integer linear combination of the transmitted signals. After collecting enough independent combinations, the receiver can recover all the individual messages. Following this idea, the integer forcing linear receiver has been proposed in [9]. This linear receiver has low complexity and significantly outperforms conventional linear architectures such as zero forcing. It can also apply to the multiple-input multiple-output (MIMO) channel with no coding across transmit antennas while still achieving the optimal diversity-multiplexing tradeoff. The original CF scheme is expanded in [5], [6], which allows unequal rates among the transmitters. In both papers, the authors show that their proposed low-complexity scheme can achieve the channel capacity for the Gaussian multiple-access channels (MAC) given high signal-to-noise ratio (SNR). Apart from the Gaussian MAC, CF is also applied to the random access Gaussian channel with a relatively large number of active users [8]. It is shown that the energy-per-bit required by each user is significantly smaller than the popular solutions such as treating interference as noise. Moreover, the reverse-CF scheme has been proposed for the Gaussian broadcast channels as the duality of the MAC [7], [10]. In this scheme, a precoded process is required at the base station. In [11], a CF-based joint coding scheme for Gaussian fast fading multiple access channels with CSIR only has been proposed and its ergodic capacity achievability has been discussed.

To fully exploit the advantage of the CF scheme, the coefficients of the decoded linear combinations should be determined carefully. For example, to recover the messages with the same amount of linear combinations as the amount of users, we should decide all the combinations are linearly independent. The integer coefficients should also be adapted to the channel state. Optimizing these coefficients is a Shortest Lattice Vector (SLV) problem. In [12], the authors design the coefficients by Fincke-Pohst-based candidate set searching algorithm and network coding system matrix constructing algorithm, with which the transmission rate of the multi-

source multi-relay system is maximized. It has also been shown in [13] that this SLV problem from the CF scheme can be solved in low polynomial complexity and an explicit deterministic algorithm that is guaranteed to find the optimal solution has been proposed.

Another aspect of this CF scheme is to understand the power allocation among the different users in the network. As discussed in the previous subsection, if CSI is available at the transmitters, allocating the transmitted powers can match the channel state with the integer coefficients, which will give a better rate. To this end, the authors in [14] propose an iterative method to optimize the achievable rate over the integer coefficients and the power allocation alternatively in a multi-user multi-relay network with the CF scheme. However, in the power allocation phase, the Lagrange method is applied, which is a sub-optimal approach for the non-convex problem. We will see the gap between the optimal approach and this approach in the later sections. Another network model is considered in [15], which includes two users, a single relay and one receiver. The relay helps the transmission from the users to the receiver while there are direct links between both users and the receiver. The CF scheme is applied at both the relay and the receiver. Then the receiver will learn a linear combination directly from both users and receive another linear combination from the relay. The power allocation problem over the users is then addressed. The authors formulate the problem by approximated geometric programming, which is sub-optimal due to the approximation. Another power allocation problem involving CF discussed in [16] applies a time division protocol in a relay network, where only the first phase applies the CF scheme. The optimization problem focuses on the time fraction and the power allocation among the relays.

Among all the CF power allocation papers, the optimal solutions are not discussed in detail. More importantly, none of the existing papers studies the CF model with multiple transmitters and one receiver in a fading scenario, which is the basic building block in any CF-based schemes. Under the traditional orthogonal scheme, the classical water-filling algorithm provides an optimal power allocation policy over fading channels [17]. However, as we will show later, more complex power allocation policies are required to optimize the computation rate under the CF scheme. In particular, the optimization problem is in general a non-convex problem with a non-smooth objective function. As a result, most off-the-shelf optimization algorithms (for example, `fmincon` in MATLAB) are strictly sub-optimal. The problem does not admit a simple optimal solution either numerically or analytically, which might explain why there is no prior work on this particular problem although the CF scheme has received a lot of attention in

the field.

B. Paper contributions

In this paper, we consider a multiple access fading channel, where the receiver requires a linear combination of the messages with predetermined equation coefficients. To the best of our knowledge, this is the first paper that discusses the power allocation problem for CF over both users and fading channels. We formulate the power allocation problem under the original CF scheme where users are equipped with lattice codes of the same rate. In general, the problem is nontrivial to solve due to its non-convexity. The main contributions of this paper are as follows.

- For the case with finitely many channel states, we establish several important properties of the optimal solutions of this non-convex problem.
- A simplified power allocation scheme where all users share the same allocation policy (called symmetric policy) is studied. It is shown that when the power constraint exceeds a certain threshold, the optimal symmetric policy takes the form of a water-filling type solution. We propose a low complexity order-based algorithm to calculate the power allocation scheme for all values of power constraints.
- We extend the proposed algorithm to calculate a general power allocation policy where different users can have different power allocation policies.
- We state relevant results for the problem when the channel coefficients are modelled as continuous random variables (continuous case). We propose a similar low-complexity iterative algorithm for the symmetric policy for the continuous case.
- Numerical results are provided for both discrete case and continuous case. The performance of the proposed algorithms is compared with several baseline algorithms, including an off-the-shelf algorithm (`fmincon` in MATLAB). In general, our proposed algorithm finds good suboptimal solutions with low complexity, and for some examples, it finds an exact optimal solution.

II. PROBLEM FORMULATION

We consider an L -user Gaussian multiple access channel (MAC) with real-valued channel coefficients h_1, \dots, h_L and the additive noise at the receiver is assumed to have zero mean and unit variance.

A. CF computation rate

Let $\mathbf{w}_1, \dots, \mathbf{w}_L$ drawn from the vector space \mathbb{F}_p^k represent the messages from L users, where p is a primal number. The encoded signals to be transmitted are denoted as $\mathbf{x}_1, \dots, \mathbf{x}_L$ correspondingly, where $\mathbf{x}_l \in \mathbb{R}^n, l = 1, \dots, L$, and n is the codeword length. The transmission powers per dimension are P_1, \dots, P_L for each user respectively. The receiver wants to recover a linear combination over the finite field, i.e., $\bigoplus_{l=1}^L q_l \mathbf{w}_l$, where q_l are the coefficients taking values in \mathbb{F}_p . It has been shown in [4] that with nested lattice codes and lattice decoding, this linear combination can be recovered by decoding an integer linear combination of the transmitted signals, i.e., $\sum_{l=1}^L a_l \mathbf{x}_l$, where $q_l = g(a_l \bmod p)$, and g is the function mapping the elements in $\{0, 1, \dots, p-1\}$ to the corresponding elements in \mathbb{F}_p . Based on the original CF scheme in [4], a generalized scheme in [5] introduced an extension to the lattice coding scheme, characterized by parameters $\boldsymbol{\beta} = (\beta_1, \dots, \beta_L)^T$ to allow unequal computation rates for different users. In this case, the achievable computation rate of each user $l \in \{1, \dots, L\}$ with channel coefficients $\mathbf{h} = (h_1, \dots, h_L)^T$ and decoding coefficients $\mathbf{a} = (a_1, \dots, a_L)^T$ is given by

$$R_{comp}^{(l)}(\mathbf{h}, \mathbf{a}, \mathbf{p}, \boldsymbol{\beta}) = \frac{1}{2} \log_2^+ \left(\beta_l^2 \left(\|\tilde{\mathbf{a}}\|^2 - \frac{((\sqrt{\mathbf{p}} \circ \mathbf{h})^T \tilde{\mathbf{a}})^2}{1 + \|\sqrt{\mathbf{p}} \circ \mathbf{h}\|^2} \right)^{-1} \right) \text{ bpcu}, \quad (1)$$

where \circ is the Hadamard product, $\tilde{\mathbf{a}} = \mathbf{a} \circ \boldsymbol{\beta}$, $\sqrt{\mathbf{p}} = (\sqrt{P_1}, \dots, \sqrt{P_L})^T$, $f^+(\cdot) = \max(0, f(\cdot))$ and bpcu is short for bits per channel use. For brevity, we will assume all rate expressions have the same unit bpcu and use $\log(\cdot)$ to stand for $\log_2(\cdot)$ in the rest of the paper unless otherwise specified.

B. Computation rate in the fading channel

In this work, we are interested in understanding the above computation rates in a fading scenario. To start with a simpler problem and keep the fairness of the users in the transmission, we will assume $\beta_l = 1$ for $l = 1, \dots, L$. Then every user has the same achievable computation rate given by

$$R_{comp}(\mathbf{h}, \mathbf{a}, \mathbf{p}) = \frac{1}{2} \log^+ \left(\left(\|\mathbf{a}\|^2 - \frac{((\sqrt{\mathbf{p}} \circ \mathbf{h})^T \mathbf{a})^2}{1 + \|\sqrt{\mathbf{p}} \circ \mathbf{h}\|^2} \right)^{-1} \right). \quad (2)$$

Following a standard model for wireless channels (see, e.g. [18, Chapter 23]), we assume a fast fading channels where the coherence time is smaller than the codeword length and we also assume the transmitters and the receiver have the channel state information (CSI). It is assumed that the channel coefficients h_l 's are independent for each user l . Moreover, we assume

the decoding coefficient \mathbf{a} is pre-determined and known to all the transmitters and the receiver. This means the decoding purpose is a given integer linear combination. We will then study the achievable computation rate defined in Definition 1 with a power allocation policy denoted by $\mathbf{p}(\mathbf{h}) = (P_1(\mathbf{h}), \dots, P_L(\mathbf{h}))^T$.

Definition 1. For a fading MAC with channel gain \mathbf{h} where \mathbf{h} is a random vector, the achievable computation rate with a power allocation policy $\mathbf{p}(\mathbf{h})$ and decoding coefficient \mathbf{a} is given by

$$R_{comp}^{fading}(\mathbf{h}, \mathbf{a}, \mathbf{p}(\mathbf{h})) = \mathbb{E}_{\mathbf{h}} \left[\frac{1}{2} \log^+ \left(\left(\|\mathbf{a}\|^2 - \frac{((\sqrt{\mathbf{p}(\mathbf{h})} \circ \mathbf{h})^T \mathbf{a})^2}{1 + \|\sqrt{\mathbf{p}(\mathbf{h})} \circ \mathbf{h}\|^2} \right)^{-1} \right) \right]. \quad (3)$$

The basic idea to achieve rate (3) is to construct different codebooks for different channel states. Similar to the process in [18, Chapter 7.4], we construct independent codebooks for each sub-channels, where the rate of the codebooks are matched to the computation rate of each sub-channel. The messages are split and encoded by the corresponding codebooks and the codewords are sent depending on the sub-channels. The decoder will decompose the received sequence according to the sub-channels and do the lattice decoding. Based on the argument in [4], with any message rate no larger than the achievable computation rate of each sub-channel, the error probability to decode each sub-sequence becomes arbitrarily small when the sequence length grows large enough. Thus, the total error probability tends to zero as the sequence length approaches infinity. In addition, with CSIT, the transmitters can also allocate the transmission power according to the sub-channels. The users are allowed to have their own power allocation policy while they are assumed to have the same average transmitted power constraint \bar{P} , i.e., $\mathbb{E}_{\mathbf{h}}[P_l(\mathbf{h})] \leq \bar{P}, \forall l \in \{1, \dots, L\}$. Then the problem of maximizing the achievable computation rate for the fading MAC can be formulated as follows.

$$\begin{aligned} \max_{\mathbf{p}(\mathbf{h})} \quad & R_{comp}^{fading}, \\ \text{s.t.} \quad & \mathbb{E}_{\mathbf{h}}[P_l(\mathbf{h})] \leq \bar{P}, \quad l = 1, \dots, L, \\ & P_l(\mathbf{h}) \geq 0, \quad l = 1, \dots, L. \end{aligned} \quad (4)$$

To fully solve this problem, we need to find L power allocation policies, one for each user. The power policies cannot be simply decoupled due to the structure of the optimization problem. Moreover, the expression in (3) is non-convex and non-smooth in the power policy $\mathbf{p}(\mathbf{h})$, which makes the optimization problem even more difficult if we consider its optimal solution. In the next section, we will first look at the discrete case where \mathbf{h} will take a finite number of values.

Remark 1. We note that the formulated power allocation problem may not fully exploit the CF scheme with CSIT and the limitations are summarized in the following points.

- We only consider equal computation rate in this case. Besides the reason that this assumption makes the problem tractable, it is a more sensible choice from a fairness point of view. If we allow different computation rates and say, maximize the sum computation rate, it often happens that assigning non-zero rate to only one user (hence effectively no computation involved) is optimal, making the problem non-interesting.
- We assume that the coefficient a is given. It is possible to extend the problem where the coefficient \mathbf{a} can be chosen (optimized) by the receiver with additional optimization procedures. This is not explored in the current paper.
- The strategy is based on separate coding for each sub-channel and does not consider jointly coding on all channel uses, for example, constructing codebooks across all sub-channels (see [11]).

However, we will see later in this paper, the problem formulated with equal rate expression and only power optimisation is already a non-trivial problem. Especially, we focus on analysing the optimal solution of the CF power allocation problem, which is not thoroughly studied in the previous literature. The derived results for this simplified case provide important insights on the properties of the optimal solutions and connect the classical water-filling algorithm to this non-convex problem. We believe these results are instructive for a more general problem formulation, and importantly, the proposed order-based algorithms can be used as a framework or building block when studying more general problems.

III. DISCRETE CASE

We first assume h_1, \dots, h_L are discretely distributed and independent real random variables. The channel h_i , $i = 1, \dots, L$, takes value h_{ij} with probability p_{ij} , $j = 1, \dots, l_i$. There are in total $M = \prod_{i=1}^L l_i$ channel states denoted as $\mathbf{h}_1, \dots, \mathbf{h}_M \in \mathbb{R}^L$, and their probability f_m can be calculated by multiplying the corresponding p_{ij} 's. The target optimization problem is given by

$$\begin{aligned}
 DP1 : \quad & \max_{\mathbf{p}} \quad \sum_{m=1}^M f_m R_{\text{comp}}(\mathbf{h}_m, \mathbf{a}, \mathbf{p}_m), \\
 & \text{s.t.} \quad \sum_{m=1}^M f_m P_{lm} \leq \bar{P}, \quad l = 1, \dots, L, \\
 & \quad \quad P_{lm} \geq 0, \quad l = 1, \dots, L, \quad m = 1, \dots, M,
 \end{aligned} \tag{5}$$

where $\mathbf{P} = (\mathbf{p}_1, \dots, \mathbf{p}_M)$, $\mathbf{p}_m = (P_{1m}, \dots, P_{Lm})^T$, $m = 1, \dots, M$, P_{lm} is the transmission power for user l at channel state m and R_{comp} is defined in (2). We further define $R_m(\mathbf{p}_m) = \frac{1}{2} \log \left(\left(\|\mathbf{a}\|^2 - \frac{((\sqrt{\mathbf{p}_m} \circ \mathbf{h}_m)^T \mathbf{a})^2}{1 + \|\sqrt{\mathbf{p}_m} \circ \mathbf{h}_m\|^2} \right)^{-1} \right)$, so that $R_{comp}(\mathbf{h}_m, \mathbf{a}, \mathbf{p}_m) = R_m^+(\mathbf{p}_m)$, $\forall m$.

A. Theoretical analysis of the optimization problem

1) **General properties of the optimal solution:** The problem $DP1$ is a non-convex problem with a non-smooth objective function, which does not admit a simple solution either analytically or numerically. The following definition and auxiliary results help us to understand the properties of the optimal solutions. Note that all norms in the paper refer to the ℓ_2 norm if it is a vector (instead of a function). We call a power allocation policy a feasible policy if it satisfies the power constraints in (5).

Definition 2. (Active set) *The active set $\mathcal{A}_{\mathbf{P}}$ of a feasible solution \mathbf{P} of the problem $DP1$ is the set of channel states in which at least one user transmits with positive power, i.e., $\mathcal{A}_{\mathbf{P}} = \{m \in \{1, \dots, M\} \mid \|\mathbf{p}_m\| > 0\}$.*

Definition 3. (Optimal active set) *Given an optimal solution \mathbf{P}^* , we use $\mathcal{A}_{\mathbf{P}^*}$ to denote the optimal active set associated to this solution. For simplicity we also use \mathcal{A}^* to denote an optimal active set with the understanding that there is an associated optimal solution.*

Remark 2. (The uniqueness of the optimal solution)

In some cases, the global optimal solution as well as the optimal active set is not unique. One example is provided as follows. Assume both users have two possible channel coefficients 0.5 and 1 with equal probability. There are totally four possible channel states, namely $\mathbf{h}_1 = (0.5, 0.5)^T$, $\mathbf{h}_2 = (0.5, 1)^T$, $\mathbf{h}_3 = (1, 0.5)^T$, $\mathbf{h}_4 = (1, 1)^T$ with equal probability 0.25. We set the computational coefficients $\mathbf{a} = (1, 1)^T$ and power constraint $\bar{P} = 2$. The optimal active set(s) will among all the $2^4 = 16$ possible active sets. It can be verified that when the active set $\mathcal{A} = \{2, 4\}$ and $\{3, 4\}$, the power policies give the same optimal average rate (0.4102). Their power policies are given by $\mathbf{P} = \begin{pmatrix} 0 & 3.3896 & 0 & 4.6105 \\ 0 & 2.5853 & 0 & 5.4145 \end{pmatrix}$ and $\begin{pmatrix} 0 & 0 & 2.5853 & 5.4145 \\ 0 & 0 & 3.3896 & 4.6105 \end{pmatrix}$, respectively. Thus, $\{2, 4\}$ and $\{3, 4\}$ are both the optimal active sets in this case, and the optimal solution is not unique.

Given an active set $\mathcal{A} = \mathcal{A}_{\mathbf{P}}$ for some feasible power allocation policy \mathbf{P} , the channel states

outside \mathcal{A} (with zero power) have no contribution to the average rate. However, the channel states inside \mathcal{A} (with positive power) do not guarantee a positive rate because the function $R_m(\mathbf{p})$ can be negative given some nonzero \mathbf{p} . In other words, some channel states may consume certain positive power without contributing to the average rate if the power policy is not constructed carefully. Note that this may happen for some feasible power allocation policies while not possible for an optimal policy, as shown in the following lemma.

Lemma 1. *If \mathcal{A}^* is an optimal active set of DP1 and \mathbf{P}^* is the corresponding solution, then all the channel states in \mathcal{A}^* have positive rate, i.e., $R_m(\mathbf{p}_m^*) > 0, \forall m \in \mathcal{A}^*$.*

Proof. Assume there is an optimal active set \mathcal{A}^* with the corresponding optimal power allocation policy \mathbf{P}^* such that it includes a channel state i with $R_i(\mathbf{p}_i^*) \leq 0$. Since $i \in \mathcal{A}^*$, we have $\|\mathbf{p}_i^*\| > 0$. Denote the average rate in this case as R_{avg} . We now show that there exists another power allocation policy that gives a higher average rate. First we set the power allocation policy of channel i to be zero, i.e., $\mathbf{p}'_i = \mathbf{0}$. Then we re-allocate \mathbf{p}_i^* to another channel state $j \in \mathcal{A}^*$ with $R_j(\mathbf{p}_j^*) > 0$. We construct its new power allocation policy \mathbf{p}'_j such that $\mathbf{p}'_j = \alpha \mathbf{p}_j^*$ with some $\alpha > 1$. Notice that such an α exists as long as $(\alpha - 1)\mathbf{p}_j^* \leq \mathbf{p}_i^*$ in an element-wise sense, which keep the power constraint satisfied. Thus,

$$\begin{aligned} R_j(\mathbf{p}'_j) &= \frac{1}{2} \log \left(\left(\|\mathbf{a}\|^2 - \frac{((\sqrt{\mathbf{p}'_j} \circ \mathbf{h}_j)^T \mathbf{a})^2}{1 + \|\sqrt{\mathbf{p}'_j} \circ \mathbf{h}_j\|^2} \right)^{-1} \right) \\ &= \frac{1}{2} \log \left(\left(\|\mathbf{a}\|^2 - \frac{\alpha((\sqrt{\mathbf{p}_j^*} \circ \mathbf{h}_j)^T \mathbf{a})^2}{1 + \alpha\|\sqrt{\mathbf{p}_j^*} \circ \mathbf{h}_j\|^2} \right)^{-1} \right) \\ &\stackrel{(a)}{>} \frac{1}{2} \log \left(\left(\|\mathbf{a}\|^2 - \frac{((\sqrt{\mathbf{p}_j^*} \circ \mathbf{h}_j)^T \mathbf{a})^2}{1 + \|\sqrt{\mathbf{p}_j^*} \circ \mathbf{h}_j\|^2} \right)^{-1} \right) = R_j(\mathbf{p}_j^*) \end{aligned}$$

The inequality (a) holds since $\frac{\alpha x}{1+\alpha y} > \frac{x}{1+y}$ for $x, y > 0$ and $\alpha > 1$. Denote the average rate in this case as R'_{avg} . Comparing with the previous rate R_{avg} , we have

$$R'_{avg} - R_{avg} = \sum_{m=i,j} f_m(R_m^+(\mathbf{p}'_m) - R_m^+(\mathbf{p}_m^*)) > 0$$

This shows \mathbf{P}^* is not optimal hence Lemma 1 is proved. \square

Given Lemma 1, the next Lemma shows that DP1 can be simplified if the optimal active set is known.

Lemma 2. *Given an optimal active set \mathcal{A}^* , the corresponding optimal power allocation policy \mathbf{P}^* satisfies $\mathbf{p}_m^* = \mathbf{0}$ for all $m \notin \mathcal{A}^*$. Furthermore, the optimal solution \mathbf{p}_m^* with $m \in \mathcal{A}^*$ is the solution to the following optimization problem with $\mathcal{S} = \mathcal{A}^*$.*

$$\begin{aligned}
 DP2 : \quad & \max_{\{\mathbf{p}_m\}_{m \in \mathcal{S}}} \sum_{m \in \mathcal{S}} f_m R_m(\mathbf{p}_m), \\
 & s.t. \quad \sum_{m \in \mathcal{S}} f_m P_{lm} \leq \bar{P}, \quad l = 1, \dots, L, \\
 & \quad \quad P_{lm} \geq 0, \quad l = 1, \dots, L, \quad m \in \mathcal{S},
 \end{aligned} \tag{6}$$

where \mathcal{S} is the support set of $DP2$, and $\mathcal{S} \subseteq \{1, \dots, M\}$.

Proof. If we know an optimal active set, from Lemma 1, $R_m(\mathbf{p}_m^*) > 0$, hence $R_m^+(\mathbf{p}_m^*) = R_m(\mathbf{p}_m^*)$ for $m \in \mathcal{A}^*$. Thus, we can release the $\max(0, \cdot)$ function for those channel m . Note that for other channel state $n \notin \mathcal{A}^*$, we have $\mathbf{p}_n^* = \mathbf{0}$ and $R_n^+(\mathbf{p}_n^*) = 0$. Therefore we can further exclude those channels in the optimization problem, which leads to $DP2$. \square

Compared to the target problem $DP1$, the problem $DP2$ has a smooth objective function which is in general easier to deal with. If we can identify the optimal active set \mathcal{A}^* , we can solve $DP1$ by solving $DP2$. However, it turns out that finding the optimal set \mathcal{A}^* is a highly non-trivial problem. In order to motivate the idea of how to determine the optimal active set, we will first look at a special case in the next section.

2) **Guarantee for the symmetric power allocation policy:** In this subsection, we consider a special case where all users share the same power allocation policy, i.e., $\mathbf{p}_m = P_m \times \mathbf{1}_L$ for $m = 1, \dots, M$, where $\mathbf{1}_L$ denotes a L -dimensional vector with all entries equaling one and P_m is the common allocated power for all users at channel state m . This simplification might also be of practical interest because the power allocation policy is easier to compute. In this case, the function $R_m(\mathbf{p}_m)$ is given by $R_m(\mathbf{p}_m) = R_m(P_m) = \frac{1}{2} \log \frac{1 + P_m \|\mathbf{h}_m\|^2}{\|\mathbf{a}\|^2 + P_m (\|\mathbf{h}_m\|^2 \|\mathbf{a}\|^2 - (\mathbf{h}_m^T \mathbf{a})^2)}$, $m = 1, \dots, M$. The target problem $DP1$ will be simplified to

$$\begin{aligned}
 DP1^{(s)} : \quad & \max_{P_1, \dots, P_M} \sum_{m=1}^M f_m R_m^+(P_m), \\
 & s.t. \quad \sum_{m=1}^M f_m P_m \leq \bar{P}, \\
 & \quad \quad P_m \geq 0, \quad m = 1, \dots, M.
 \end{aligned} \tag{7}$$

If the optimal active set is known, we have the following problem corresponding to the problem $DP2$ in Lemma 2 by setting \mathcal{S} to be the optimal active set.

$$\begin{aligned}
DP2^{(s)} : \quad & \max_{\{P_m\}_{m \in \mathcal{S}}} \sum_{m \in \mathcal{S}} f_m R_m(P_m), \\
& \text{s.t.} \quad \sum_{m \in \mathcal{S}} f_m P_m \leq \bar{P}, \\
& P_m \geq 0, \quad l = 1, \dots, L.
\end{aligned} \tag{8}$$

Note that the above problem is convex and satisfies Slater's condition. Thus, the optimal solution can be given by Karush–Kuhn–Tucker (KKT) conditions [19]. We present its optimal solution in an analytical form in Lemma 3. Note that we say two nonzero vectors $\mathbf{x}_1, \mathbf{x}_2$ are collinear if they satisfy $\mathbf{x}_1 = c\mathbf{x}_2$ for some real number c .

Lemma 3. *The optimal solutions P_m^* of the problem $DP2^{(s)}$ is given by*

$$P_m^* = (P_m^{KKT}(\lambda^*))^+, m \in \mathcal{S} \tag{9}$$

with some $\lambda^* > 0$ satisfying the equation

$$\sum_{m \in \mathcal{S}} f_m P_m^* = \bar{P}. \tag{10}$$

P_m^{KKT} is defined as

$$P_m^{KKT}(\lambda) = \begin{cases} \frac{1}{\lambda} - \frac{1}{\|\mathbf{h}_m\|^2}, & \mathbf{h}_m \text{ is collinear with } \mathbf{a} \\ \frac{-b_m + \sqrt{b_m^2 - 4d_m c_m(\lambda)}}{2d_m}, & \mathbf{h}_m \text{ is not collinear with } \mathbf{a} \end{cases} \tag{11}$$

with $d_m = \|\mathbf{h}_m\|^2(\|\mathbf{h}_m\|^2\|\mathbf{a}\|^2 - (\mathbf{h}_m^T \mathbf{a})^2)$, $b_m = 2\|\mathbf{h}_m\|^2\|\mathbf{a}\|^2 - (\mathbf{h}_m^T \mathbf{a})^2$, $c_m(\lambda) = \|\mathbf{a}\|^2 - \frac{(\mathbf{h}_m^T \mathbf{a})^2}{\lambda}$.

Proof. We will first show the problem $DP2^{(s)}$ is a convex problem. Firstly, the constraint space is an intersection of halfspaces, thus convex. Secondly, the second derivative of $R_m(P_m)$ is given by

$$R_m''(P_m) = -\frac{(\mathbf{h}_m^T \mathbf{a})^2(\|\mathbf{h}_m\|^2\|\mathbf{a}\|^2 + (1 + 2P_m\|\mathbf{h}_m\|^2)(\|\mathbf{h}_m\|^2\|\mathbf{a}\|^2 - (\mathbf{h}_m^T \mathbf{a})^2))}{(\|\mathbf{a}\|^2 + P_m\|\mathbf{h}_m\|^2)^2(1 + P_m\|\mathbf{h}_m\|^2)^2} < 0, \tag{12}$$

for any $P_m \geq 0$, which implies $R_m(P_m)$ is strictly concave. Here we use Cauchy–Schwarz inequality to show that $\|\mathbf{h}_m\|^2\|\mathbf{a}\|^2 - (\mathbf{h}_m^T \mathbf{a})^2 \geq 0$. Then the objective function of $DP2^{(s)}$ is strictly concave since it is a linear combination of $R_m(P_m)$ with all positive coefficients f_m , for $m \in \mathcal{S}$. Therefore, $DP2^{(s)}$ is a convex optimization problem.

The Lagrangian function of the objective of the problem $DP2^{(s)}$ is given by

$$L(\mathbf{P}, \lambda) = \left(\sum_{m \in \mathcal{S}} -f_m R_m(P_m) \right) + \lambda \left(\sum_{m \in \mathcal{S}} f_m P_m - \bar{P} \right) \quad (13)$$

with complementary constraint $P_m \geq 0$, where $\lambda \geq 0$ is the Lagrange multiplier. By setting $\frac{\partial L(\mathbf{P}, \lambda)}{\partial P_m} = 0$, we obtain

$$-f_m R'_m(P_m) + \lambda f_m = 0. \quad (14)$$

Calculating $R'_m(P_m)$ and plugging it into the above equation, we have

$$\frac{(\mathbf{h}_m^T \mathbf{a})^2}{\|\mathbf{h}_m\|^2 (\|\mathbf{h}_m\|^2 \|\mathbf{a}\|^2 - (\mathbf{h}_m^T \mathbf{a})^2) P_m^2 + (2\|\mathbf{h}_m\|^2 \|\mathbf{a}\|^2 - (\mathbf{h}_m^T \mathbf{a})^2) P_m + \|\mathbf{a}\|^2} = 2\lambda. \quad (15)$$

Since the left hand side (LHS) of (15) is nonzero for nonzero \mathbf{h}_m , it implies that $\lambda > 0$. Here we replace 2λ with λ without loss of generality. The equation (15) can be further simplified as

$$\underbrace{\|\mathbf{h}_m\|^2 (\|\mathbf{h}_m\|^2 \|\mathbf{a}\|^2 - (\mathbf{h}_m^T \mathbf{a})^2)}_{:=d_m} P_m^2 + \underbrace{(2\|\mathbf{h}_m\|^2 \|\mathbf{a}\|^2 - (\mathbf{h}_m^T \mathbf{a})^2)}_{:=b_m} P_m + \underbrace{\|\mathbf{a}\|^2 - \frac{(\mathbf{h}_m^T \mathbf{a})^2}{\lambda}}_{:=c_m(\lambda)} = 0. \quad (16)$$

Recall that the channel coefficients are real, thus d_m, b_m and $c_m(\lambda)$ are all real. When \mathbf{h} and \mathbf{a} are collinear, we have $d_m = 0$. In this case the solution of (16) is given by

$$P_m = -\frac{c_m(\lambda)}{b_m} = \frac{1}{\lambda} - \frac{1}{\|\mathbf{h}_m\|^2}. \quad (17)$$

When \mathbf{h} and \mathbf{a} are not collinear, the solution of (16) is given by

$$P_m = \frac{-b_m \pm \sqrt{b_m^2 - 4d_m c_m(\lambda)}}{2d_m}. \quad (18)$$

Since $d_m, b_m > 0$, the solution with minus sign will be negative in any case, thus should be discarded. Recalling the complementary constraint $P_m \geq 0$, and based on the fact that $R_m(P_m)$ is increasing with P_m , we have the solution form as shown in Lemma 3. \square

It can be observed that when \mathbf{h} and \mathbf{a} are collinear, the solution actually reduces to the classical water-filling solution [17]. Since $c_m(\lambda)$ is monotonically increasing with λ , we note that $P_m^{KK^T}(\lambda)$ is always a monotonically decreasing function of λ . The result in Lemma 3 is reminiscent of the well-known water-filling algorithm, where λ determines the water level. The water level will decide which channel states are active. Therefore, in practice we can apply bisection algorithm to determine the optimal λ as shown in Algorithm 1.

As mentioned in the previous subsection, finding the optimal active set \mathcal{A}^* is a highly non-trivial problem. In this simplified case, we study the property of \mathcal{A}^* by introducing the following two definitions.

Algorithm 1 Optimal algorithm for $DP2^{(s)}$

- 1: Define $q(\lambda) = \sum_{m \in \mathcal{S}} f_m(P_m^{KKT}(\lambda))^+$ \triangleright Note that the power constraint is $q(\lambda) \leq \bar{P}$
 - 2: Choose a small enough λ_l (e.g. $\lambda_l < 10^{-3}$) such that $q(\lambda_l) > \bar{P}$
 - 3: Choose a large enough λ_u (e.g. $\lambda_u > 10^3$) such that $q(\lambda_u) < \bar{P}$
 - 4: $\lambda \leftarrow \frac{\lambda_l + \lambda_u}{2}$
 - 5: **while** $\bar{P} - q(\lambda) > \epsilon$ **||** $\bar{P} - q(\lambda) < 0$ **do** $\triangleright \epsilon$ is a small number, e.g. 10^{-3}
 - 6: **if** $\bar{P} - q(\lambda) > 0$ **then**
 - 7: $\lambda_u \leftarrow \lambda$
 - 8: **else**
 - 9: $\lambda_l \leftarrow \lambda$
 - 10: **end if**
 - 11: $\lambda \leftarrow \frac{\lambda_l + \lambda_u}{2}$
 - 12: **end while**
 - 13: $P_m^* = (P_m^{KKT}(\lambda))^+, \forall m \in \mathcal{S}$
-

Definition 4. (Good set) The good set \mathcal{G} is the subset of $\{1, \dots, M\}$ which include all m with $\|\mathbf{h}_m\|^2 > \|\mathbf{h}_m\|^2 \|\mathbf{a}\|^2 - (\mathbf{h}_m^T \mathbf{a})^2$, i.e., $\mathcal{G} = \left\{ m \in \{1, \dots, M\} \mid \|\mathbf{h}_m\|^2 > \|\mathbf{h}_m\|^2 \|\mathbf{a}\|^2 - (\mathbf{h}_m^T \mathbf{a})^2 \right\}$.

Definition 5. (Bad set) The bad set \mathcal{B} is the complement of the good set in $\{1, \dots, M\}$, i.e., $\mathcal{B} = \{1, \dots, M\} \setminus \mathcal{G}$.

The following theorem will present the relationship between \mathcal{A}^* and these two sets.

Theorem 1. Let \mathcal{A}^* denote an optimal active set for the symmetric power allocation policy. It always holds that $\mathcal{A}^* \cap \mathcal{B} = \emptyset$ and $\mathcal{A}^* \subseteq \mathcal{G}$.

Proof. Let P_1^*, \dots, P_M^* denote the optimal power allocations corresponding to \mathcal{A}^* . Assume \mathcal{A}^* includes some channel state $k \in \mathcal{B}$. By the definition of the bad set, $\|\mathbf{h}_k\|^2 \leq \|\mathbf{h}_k\|^2 \|\mathbf{a}\|^2 - (\mathbf{h}_k^T \mathbf{a})^2$. Since \mathbf{a} is a nonzero integer vector, we have $\|\mathbf{a}\| \geq 1$. Then we have $1 + P_k \|\mathbf{h}_k\|^2 \leq \|\mathbf{a}\|^2 + P_k (\|\mathbf{h}_k\|^2 \|\mathbf{a}\|^2 - (\mathbf{h}_k^T \mathbf{a})^2), \forall P_k \geq 0$. Recalling the expression of $R_m(P_m)$, we have $R_k(P_k) \leq 0, \forall P_k \geq 0$, especially $R_k(P_k^*) \leq 0$. However, from Lemma 1, any channel state in the optimal active set will return a positive rate. Thus, a contradiction happens. Therefore \mathcal{A}^* should not include any channel state in the bad set. By the relationship between the good set and the bad set, the optimal active set is a subset of the good set. \square

Furthermore, the following result states that when the power constraint \bar{P} is large enough, the optimal active set \mathcal{A}^* is equal to the good set \mathcal{G} .

Theorem 2. *The optimal active set of the problem $DP1^{(s)}$ is the good set, i.e., $\mathcal{A}^* = \mathcal{G}$ when the power constraint \bar{P} is larger than $\bar{P}_o = \sum_{m \in \mathcal{G}} f_m P_m^{KKT}(\lambda_o)$ with*

$$\lambda_o = \min \left\{ \lambda \mid R_m(P_m^{KKT}(\lambda)) = R'_m(P_m^{KKT}(\lambda)) P_m^{KKT}(\lambda), m \in \mathcal{G} \right\},$$

where $R_m(\cdot)$ is given in (7) and $P_m^{KKT}(\lambda)$ is given by (11).

Proof. Assume $\mathcal{A}^* \neq \mathcal{G}$ when $\bar{P} > \bar{P}_o$. From Theorem 1 $\mathcal{A}^* \subseteq \mathcal{G}$, then $\mathcal{A}^* \subset \mathcal{G}$. Denote $\mathcal{A}^* = \mathcal{G} \setminus \mathcal{K}$, where $\mathcal{K} \subset \mathcal{G}$ is some non-empty set. From Lemma 2 and Lemma 3, since \mathcal{A}^* is given, the optimal solution of the problem $DP1^{(s)}$ can be given by Algorithm 1 with $\mathcal{S} = \mathcal{A}^*$. Denote the optimal Lagrangian multiplier as λ_k . Then the nonzero power allocation is given by $P_m^{KKT}(\lambda_k)$, $m \in \mathcal{A}^*$, and the optimal average rate is $\sum_{m \in \mathcal{A}^*} f_m R_m(P_m^{KKT}(\lambda_k))$. Here, we omit the $\max(0, \cdot)$ function in the rate expression from Lemma 1. In addition, the power constraint gives $\sum_{m \in \mathcal{A}^*} f_m P_m^{KKT}(\lambda_k) = \bar{P}$.

We can also derive a non-optimal solution by applying Algorithm 1 with $\mathcal{S} = \mathcal{G}$. Denote the Lagrangian multiplier of this solution as λ_g . Then the power allocation is given by $P_m^{KKT}(\lambda_g)$ for $m \in \mathcal{G}$ and zero for other channel states. The corresponding average rate is $\sum_{m \in \mathcal{G}} f_m R_m^+(P_m^{KKT}(\lambda_g))$, and the power constraint gives $\sum_{m \in \mathcal{G}} f_m P_m^{KKT}(\lambda_g) = \bar{P}$.

From the optimality of \mathcal{A}^* , we have $\sum_{m \in \mathcal{A}^*} f_m R_m(P_m^{KKT}(\lambda_k)) \geq \sum_{m \in \mathcal{G}} f_m R_m^+(P_m^{KKT}(\lambda_g))$. Since $\max(0, x) \geq x, \forall x$, we have $\sum_{m \in \mathcal{A}^*} f_m R_m(P_m^{KKT}(\lambda_k)) \geq \sum_{m \in \mathcal{G}} f_m R_m(P_m^{KKT}(\lambda_g))$. By moving terms, the following is observed.

$$\sum_{m \in \mathcal{A}^*} f_m (R_m(P_m^{KKT}(\lambda_k)) - R_m(P_m^{KKT}(\lambda_g))) \geq \sum_{m \in \mathcal{K}} f_m R_m(P_m^{KKT}(\lambda_g)). \quad (19)$$

From the power constraint, $\sum_{m \in \mathcal{A}^*} f_m P_m^{KKT}(\lambda_k) = \sum_{m \in \mathcal{G}} f_m P_m^{KKT}(\lambda_g)$. As $\mathcal{A}^* \subset \mathcal{G}$, $P_m^{KKT}(\lambda_k) \geq P_m^{KKT}(\lambda_g), \forall m \in \mathcal{A}^*$. This is because $P_m^{KKT}(\lambda)$ is monotonically increasing with λ . By moving terms of the power equation, we have

$$\sum_{m \in \mathcal{A}^*} f_m (P_m^{KKT}(\lambda_k) - P_m^{KKT}(\lambda_g)) = \sum_{m \in \mathcal{K}} f_m P_m^{KKT}(\lambda_g) \quad (20)$$

As $R_m(P_m)$ is concave, it holds that $R_m(x_2) - R_m(x_1) \leq R'_m(x_1)(x_2 - x_1), \forall x_2 \geq x_1 \geq 0$. Then $R_m(P_m^{KKT}(\lambda_k)) - R_m(P_m^{KKT}(\lambda_g)) \leq R'_m(P_m^{KKT}(\lambda_g))(P_m^{KKT}(\lambda_k) - P_m^{KKT}(\lambda_g)), \forall m \in \mathcal{A}^*$. Thus, combining (19), we have

$$\sum_{m \in \mathcal{A}^*} f_m R'_m(P_m^{KKT}(\lambda_g))(P_m^{KKT}(\lambda_k) - P_m^{KKT}(\lambda_g)) \geq \sum_{m \in \mathcal{K}} f_m R_m(P_m^{KKT}(\lambda_g)). \quad (21)$$

Note that $R'_m(P_m^{KKKT}(\lambda_g)) = \lambda_g, \forall m \in \mathcal{G}$, then it can be lifted out of the summation on the left side of (21). Then using (20), we can rewrite (21) as

$$\lambda_g \sum_{m \in \mathcal{K}} f_m P_m^{KKKT}(\lambda_g) \geq \sum_{m \in \mathcal{K}} f_m R_m(P_m^{KKKT}(\lambda_g)). \quad (22)$$

Placing λ_g into the summation of the left side results in

$$\sum_{m \in \mathcal{K}} f_m (R'_m(P_m^{KKKT}(\lambda_g)) P_m^{KKKT}(\lambda_g) - R_m(P_m^{KKKT}(\lambda_g))) \geq 0. \quad (23)$$

This indicates that there exists at least one channel state $k \in \mathcal{K}$ such that

$$R'_k(P_k^{KKKT}(\lambda_g)) P_k^{KKKT}(\lambda_g) - R_k(P_k^{KKKT}(\lambda_g)) \geq 0. \quad (24)$$

Next we will find out a contradiction to the above statement. Recall that $R_m(x)$ is an increasing concave function for $x \geq 0$ and $m \in \mathcal{G}$, i.e., $R'_m(x) > 0$, $R''_m(x) < 0$. Thus, the function $l_m(x) = R_m(x) - R'_m(x)x$ is an increasing function of x since $l'_m(x) = -R''_m(x) > 0$. In addition, $l_m(0) = R_m(0) = -\frac{1}{2}$ and $\lim_{x \rightarrow \infty} l_m(x) = \lim_{x \rightarrow \infty} R_m(x) > 0$. So there exists only one $x_m > 0$ such that $l_m(x_m) = 0$ for each $m \in \mathcal{G}$. Since $P_m^{KKKT}(\lambda)$ is a monotonically decreasing function of λ , there exists only one $\lambda_m > 0$ such that $l_m(P_m^{KKKT}(\lambda_m)) = 0$ for each $m \in \mathcal{G}$, and $l_m(P_m^{KKKT}(\lambda)) > l_m(P_m^{KKKT}(\lambda_m)) = 0, \forall 0 < \lambda < \lambda_m$. When $\bar{P} > \bar{P}_o$, from the expression of \bar{P}_o , we have $\lambda_g < \lambda_o \leq \lambda_m$. Thus, $l_m(P_m^{KKKT}(\lambda_g)) > 0$ i.e.,

$$R_m(P_m^{KKKT}(\lambda_g)) - R'_m(P_m^{KKKT}(\lambda_g)) P_m^{KKKT}(\lambda_g) > 0, \forall m \in \mathcal{G}. \quad (25)$$

This leads to a contradiction of (24). Thus, the assumption $\mathcal{A}^* \neq \mathcal{G}$ breaks. Therefore we can conclude that $\mathcal{A}^* = \mathcal{G}$ when $\bar{P} > \bar{P}_o$. \square

Theorem 2 shows that when $\bar{P} > \bar{P}_o$, the target problem in the symmetric policy case can be solved efficiently with Lemma 3 by setting $\mathcal{A}^* = \mathcal{G}$. In the following, we consider the low to medium power regime where the optimal active set could be a strict subset of \mathcal{G} . Moreover, we will discuss how to extend those algorithms to the asymmetric policy case previously formulated in DP1.

3) **Further discussion on the asymmetric power allocation policy scenario:** We restate the objective function $R_m(\mathbf{p}_m)$ when we have a general (asymmetric) power allocation policy.

$$R_m(\mathbf{p}_m) = \frac{1}{2} \log \frac{1 + \|\sqrt{\mathbf{p}_m} \circ \mathbf{h}_m\|^2}{\|\mathbf{a}\|^2 + (\|\sqrt{\mathbf{p}_m} \circ \mathbf{h}_m\|^2 \|\mathbf{a}\|^2 - ((\sqrt{\mathbf{p}_m} \circ \mathbf{h}_m)^T \mathbf{a})^2)}. \quad (26)$$

First notice that unlike the objective function for the symmetric policy, this objective function is non-convex in general. Hence the solution to the KKT conditions cannot guarantee a global

optimal solution. Intuitively, the optimal power allocation tries to align the channel coefficients \mathbf{h}_m to the equation coefficients \mathbf{a}_m as close as possible to minimise the term $\|\sqrt{\mathbf{p}_m} \circ \mathbf{h}_m\|^2 \|\mathbf{a}\|^2 - ((\sqrt{\mathbf{p}_m} \circ \mathbf{h}_m)^T \mathbf{a})^2$, hence achieve a larger rate. Particularly, the channel states in the Bad Set can have positive rate with asymmetric power allocation. This is different from the symmetric power policy scenario, and Theorem 1 will not apply to the asymmetric policy scenario in general. If the power constraint is large enough, we naturally expect that all the channel states will be active. However, since we do not have the analytical expressions for the optimal solution in this asymmetric policy scenario, it is difficult to give a threshold like we did in Theorem 2 for the symmetric policy scenario. In the sequel, we will investigate how to determine the active set for a general asymmetric policy, and use off-the-shelf numerical methods to solve the associated optimization problem in *DP2*.

B. Power allocation algorithms for the discrete case

We will first introduce some benchmark algorithms and then propose our iterative algorithm to identify the optimal active set for a general power constraint. We present the algorithms for both symmetric power allocation policy case (special case) and the asymmetric policy case (general case) on the left and the right side of the table, respectively.

The first benchmark algorithm Algorithm A0 is the trivial power allocation scheme that applies equal power to all users at all channel states.

Algorithm A0 Constant power

$$DP1^{(s)}:$$

$$1: P_m = \bar{P}, \forall m \in \mathcal{G}.$$

$$DP1:$$

$$P_{lm} = \bar{P}, \forall l \in \{1, \dots, L\}, m \in \{1, \dots, M\}.$$

We consider Algorithm A1, which directly sets all the possible channel state to be active. Lemma 3 provided a more explicit characterization of the solution for the symmetric policy. The exact expression is different from the classical water-filling solution. But in order to keep the notations simple, we still call this algorithm “water-filling algorithm”.

Here, we initially choose the Good Set as the starter in the special case according to Theorem 1. Since Algorithm 1 only solves the modified problem $DP2^{(s)}$ (or $DP2$ for the general case) but not the original problem, the solution $\{P_m^{(1)}\}_{m \in \mathcal{S}}$ (or $\{\mathbf{p}_m^{(1)}\}_{m \in \mathcal{S}}$ for the general case) in the first iteration may allocate positive power to channel states under which the rate is zero, due

Algorithm A1 Water-filling algorithm

| | |
|--|--|
| $DP1^{(s)}$: <ol style="list-style-type: none"> 1: Set $\mathcal{S}^{(1)} = \mathcal{G}$, $\{P_m^{(1)}\}_{m=1}^M = \{P_m^{(2)}\}_{m=1}^M = \mathbf{0}$; 2: Run Algorithm 1 with $\mathcal{S} = \mathcal{S}^{(1)}$; 3: Update solution $\{P_m^{(1)}\}_{m \in \mathcal{S}^{(1)}}$; 4: Update $\mathcal{S}^{(2)} = \{m \in \mathcal{G} R_m(P_m^{(1)}) > 0\}$; 5: Run Algorithm 1 with $\mathcal{S} = \mathcal{S}^{(2)}$; 6: Update solution $\{P_m^{(2)}\}_{m \in \mathcal{S}^{(2)}}$. | $DP1$: <p>Set $\mathcal{S}^{(1)} = \{1, \dots, M\}$, $\mathbf{p}^{(1)} = \mathbf{p}^{(2)} = \mathbf{0}$;</p> <p>Numerically solve $DP2$;</p> <p>Update solution $\{\mathbf{p}_m^{(1)}\}_{m \in \mathcal{S}^{(1)}}$;</p> <p>Update $\mathcal{S}^{(2)} = \{m R_m(\mathbf{p}_m^{(1)}) > 0\}$;</p> <p>Numerically solve $DP2$;</p> <p>Update solution $\{\mathbf{p}_m^{(2)}\}_{m \in \mathcal{S}^{(2)}}$.</p> |
|--|--|

to the $\max(0, \cdot)$ function in the original objective function. Therefore, it is natural to exclude the channel states with zero rate and do another iteration with the remaining channel states. This is reflected in the 3rd and the 4th line in Algorithm A1. We also point out that after the second iteration, every channel state will be associated with a positive rate. As shown in Theorem 2, this algorithm is able to solve $DP1^{(s)}$ optimally when the power constraint is large enough, but in general we will show later with numerical examples, this two-step algorithm is not guaranteed to find the optimal active set. This is because it fails to capture enough characteristics of the objective function as the $\max(0, \cdot)$ function is omitted. To this end, our next algorithm (Algorithm A2) attempts to take the channel state information into account. Specifically, we will include channel states in the active set by constructing a certain order on different channel states.

The idea of Algorithm A2 is to sort the channel states in a certain order, which indicates the goodness of the channels, and solve the simplified problem $DP2^{(s)}$ (or $DP2$ in the general case) with the ordered channel set shrunk by one channel state per iteration. Specifically, for each iteration, we successively eliminate the worst channel in the set \mathcal{S} , until the sum rate starts to decrease. Therefore, the performance of this algorithm highly depends on the choice of the ordering methods. In fact, the water-filling algorithm for the symmetric power policy case also has an implicit order of the channel states. The KKT solution shows the power allocation grows monotonically with $1/\lambda$. Since λ is a common parameter for all channel states, the channels will become active (with positive power allocated) gradually in a certain order as $1/\lambda$ increases. Thus, the order criterion of the water-filling algorithm is just the order in which channels become active as $1/\lambda$ increases. This particular ordering does give an optimal solution in some cases but is in general suboptimal, as we will show in later examples. To this end, we propose two other ordering method with explicit characterizations.

Algorithm A2 Proposed iterative algorithm

| | |
|--|---|
| <p><i>DP1</i>^(s):</p> <ol style="list-style-type: none"> 1: Set $\mathcal{S} = \mathcal{G}_s$, where \mathcal{G}_s is the ordered version of the good set in the ascending order according to the ordering method, i.e., from the bad channels to the good ones; 2: Set $\{P_m\}_{m=1}^M = \{P'_m\}_{m=1}^M = \mathbf{0}$; 3: Set $\mathcal{S}' = \mathcal{S}$; 4: Let $r(P_m) = \sum_{m \in \mathcal{S}} f_m R_m^+(P_m)$; 5: while $r(P'_m) \geq r(P_m) \ \&\& \ \mathcal{S}' \neq \emptyset$ do 6: $\{P_m\}_{m=1}^M \leftarrow \{P'_m\}_{m=1}^M, \{P'_m\}_{m=1}^M \leftarrow \mathbf{0}, \mathcal{S} \leftarrow \mathcal{S}'$; 7: Run Algorithm 1 \rightarrow Update solution $\{P'_m\}_{m \in \mathcal{S}}$; 8: if $\exists m$ such that $R_m^+(P'_m) = 0$ then 9: $\tilde{\mathcal{S}} = \{m \in \mathcal{S} R_m(P'_m) > 0\}$; 10: $\{P'_m\}_{m \notin \tilde{\mathcal{S}}} = \mathbf{0}$; 11: Run Algorithm 1 with $\tilde{\mathcal{S}} \rightarrow$ Update $\{P'_m\}_{m \in \tilde{\mathcal{S}}}$; 12: end if 13: $\mathcal{S}' \leftarrow \mathcal{S} \setminus \{\text{1st channel state}\}$; 14: end while 15: $\{P_m^*\}_{m=1}^M = \{P_m\}_{m=1}^M$. | <p><i>DP1</i>:</p> <p>Set \mathcal{S} as the ordered version of $\{1, \dots, M\}$ in the ascending order according to the ordering method, i.e., from the bad channels to the good ones;</p> <p>Set $\mathbf{P} = \mathbf{P}' = \mathbf{0}$, where $\mathbf{P} = \{\mathbf{p}_1, \dots, \mathbf{p}_M\}$;</p> <p>Set $\mathcal{S}' = \mathcal{S}$;</p> <p>Let $r(\mathbf{P}) = \sum_{m \in \mathcal{S}} f_m R_m^+(\mathbf{p}_m)$;</p> <p>while $r(\mathbf{P}') \geq r(\mathbf{P}) \ \&\& \ \mathcal{S}' \neq \emptyset$ do</p> <p> $\mathbf{P} \leftarrow \mathbf{P}', \mathbf{P}' \leftarrow \mathbf{0}, \mathcal{S} \leftarrow \mathcal{S}'$;</p> <p> Solve <i>DP2</i> \rightarrow Update solution $\{\mathbf{p}'_m\}_{m \in \mathcal{S}}$;</p> <p> if $\exists m$ such that $R_m^+(\mathbf{p}'_m) = 0$ then</p> <p> $\tilde{\mathcal{S}} = \{m \in \mathcal{S} R_m(\mathbf{p}'_m) > 0\}$;</p> <p> $\{\mathbf{p}'_m\}_{m \notin \tilde{\mathcal{S}}} = \mathbf{0}$;</p> <p> Solve <i>DP2</i> with $\tilde{\mathcal{S}} \rightarrow$ Update $\{\mathbf{p}'_m\}_{m \in \tilde{\mathcal{S}}}$;</p> <p> end if</p> <p> $\mathcal{S}' \leftarrow \mathcal{S} \setminus \{\text{1st channel state}\}$;</p> <p>end while</p> <p>$\mathbf{P}^* = \mathbf{P}$.</p> |
|--|---|

Ordering method 1: The first ordering method is to order the channel states based on the point where $R_m(P_m) = 0$ in the symmetric power policy case, i.e., $P_m = \frac{\|\mathbf{a}\|^2 - 1}{\|\mathbf{h}_m\|^2 - (\|\mathbf{h}_m\|^2 \|\mathbf{a}\|^2 - (\mathbf{h}_m^T \mathbf{a})^2)}$. So we can order the channel states by the following criterion

$$O_m^{(1)} = \|\mathbf{h}_m\|^2 - (\|\mathbf{h}_m\|^2 \|\mathbf{a}\|^2 - (\mathbf{h}_m^T \mathbf{a})^2), \quad (27)$$

which indicates the goodness of the channel states when the power constraint is small. A larger $O_m^{(1)}$ means the rate of the channel state m needs a smaller power to be positive, which is better.

Ordering method 2: The second ordering method is to order the channel states based on the asymptotic value of $R_m(P_m)$ in the symmetric power policy case, i.e., $\lim_{P_m \rightarrow \infty} R_m(P_m) = \frac{1}{2} \log \frac{\|\mathbf{h}_m\|^2}{\|\mathbf{h}_m\|^2 \|\mathbf{a}\|^2 - (\mathbf{h}_m^T \mathbf{a})^2}$. So we can order the channel states by the following criterion

$$O_m^{(2)} = \frac{\|\mathbf{h}_m\|^2}{\|\mathbf{h}_m\|^2 \|\mathbf{a}\|^2 - (\mathbf{h}_m^T \mathbf{a})^2}. \quad (28)$$

It shows the goodness of the channel states when the power constraint is large. With a larger $O_m^{(2)}$, the channel state m has a larger rate asymptotically.

There is a clear intuition behind the above two ordering methods. The norm of the channel coefficient $\|\mathbf{h}_m\|^2$ indicates the strength of the channel where a larger norm corresponds to a better channel. The difference term $\|\mathbf{h}_m\|^2 \|\mathbf{a}\|^2 - (\mathbf{h}_m^T \mathbf{a})^2$ indicates how the coefficients \mathbf{a} is

”aligned” with the channel coefficients. Notice this term is always non-negative and is zero if and only if when \mathbf{h}_m and \mathbf{a} are collinear. Hence a larger difference corresponds to a worse channel. There is a tension between these two terms, and the two ordering methods captures this tension in either additively (method 1) or multiplicatively (method 2).

Apart from the algorithms mentioned before one can exhaustively evaluate all possible active sets, and solve the corresponding power allocation policies. Notice that this algorithm can guarantee the optimality for the symmetric power allocation problem $DP1^{(s)}$, while it can only guarantee a local minimum in the asymmetric power policy case. This is because even if an optimal active set is given, the corresponding $DP2$ is still non-convex, and a numerical solution is not guaranteed to a global minimum. The procedure is presented in Algorithm A3.

Algorithm A3 Exhaustive search

| | |
|--|--|
| $DP1^{(s)}$: 1: Identify all the possible active set $\mathcal{A}_1, \dots, \mathcal{A}_{2^M}$; 2: for $i = 1 : 2^M$ do 3: Set $\{P_m^{(i)}\}_{m=1}^M = \mathbf{0}$; 4: $\mathcal{S} = \mathcal{A}_i$; 5: Run Algorithm 1 \rightarrow Update solution $\{P_m^{(i)}\}_{m \in \mathcal{S}}$; 6: end for 7: $\{P_m^*\}_{m=1}^M = \max_{\{P_m^{(i)}\}_{m=1}^M} \sum_{m \in \mathcal{A}_i} f_m R_m^+(P_m^{(i)})$. | $DP1$: Identify all the possible active set $\mathcal{A}_1, \dots, \mathcal{A}_{2^M}$; for $i = 1 : 2^M$ do Set $\mathbf{P}^{(i)} = \mathbf{0}$; $\mathcal{S} = \mathcal{A}_i$; Numerically solve $DP2 \rightarrow$ Update solution $\{\mathbf{p}_m^{(i)}\}_{m \in \mathcal{S}}$; end for $\mathbf{P}^* = \max_{\mathbf{p}^{(i)}} \sum_{m \in \mathcal{A}_i} f_m R_m^+(\mathbf{p}_m^{(i)})$. |
|--|--|

The exhaustive search is not applicable for large M due to its computational complexity. As an informal argument, let us denote the computational complexity of solving $DP2^{(s)}$ (or $DP2$ in the general case) by c . Then the complexity of Algorithm A3 is given by $2^M c$, which increases exponentially with M . For other algorithms, A0 essentially requires no computation, while A1 has a complexity of $2c$. The complexity of Algorithm A2 depends on the number of the iterations it runs and the number of the ordering methods it applies. For each method, the complexity is at most $2Mc$, which increases linearly with M . Note that if the power constraint is large enough as shown in Theorem 2, the computational complexity is only c for the optimal solution in the symmetric policy case.

C. Numerical results for the discrete case

1) **Example 1:** We look at a two-user example ($L = 2$). Assume the channel states are distributed independently in the following Table I, and the computation coefficients are given by

$\mathbf{a} = (1 \ 1)^T$. Then the joint channel state distribution is given by Table II. There are totally four possible channel states ($M = 4$). According to Definition 4, the good set in this case is given by $\mathcal{G} = \{1, 2, 3, 4\}$.

Table I: Channel state distribution in Example 1

| | | | | | |
|-------|-----|-----|-------|-----|-----|
| h_1 | 1 | 3 | h_2 | 0.5 | 2 |
| p_1 | 0.6 | 0.4 | p_2 | 0.8 | 0.2 |

Table II: Joint channel state distribution in Example 1

| | | | | |
|------------------|----------|--------|----------|--------|
| \mathbf{h}_m^T | (1, 0.5) | (1, 2) | (3, 0.5) | (3, 2) |
| f_m | 0.48 | 0.12 | 0.32 | 0.08 |

Let us first consider the symmetric power allocation policy case. We will first derive \bar{P}_o from Theorem 2. We solve the equation $R'_m(x_m)x_m = R_m(x_m)$ for x_m for $m = 1, 2, 3, 4$, and then solve λ_m from $P_m^{(KK^T)}(\lambda_m) = x_m$. Then we take $\lambda_o = \min \lambda_m$. Thus, \bar{P}_o is finally given by $\sum_{m=1}^4 f_m P_m^{(KK^T)}(\lambda_o) = 2.09$.

Fig. 1 plots the expected achievable rate using the exhaustive search, the first iteration of Algorithm A1 and constant power policy with the power constraint larger than $\bar{P}_o = 2.09$. As discussed in the previous section, Algorithm A3 (exhaustive search) will achieve the optimal expected rate. It can be observed that the first iteration of Algorithm A1 ($\{P_m^{(1)}\}_{m=1}^M$) also achieves optimality, which is consistent with Theorem 2. Moreover, the average rate of Algorithm A0 has a constant gap below the optimality for all power constraints.

Fig. 2 shows the expected achievable rate for the power constraint no larger than \bar{P}_o . Algorithm A3 (exhaustive search) gives the optimal expected rate. It is clear that Algorithm A0 has the poorest performance with the trivial power allocation policy. Algorithm A1 is optimal for some power constraints but is in general sub-optimal. A sharp rate falling occurs from $\bar{P} = 0.8$ to 0.9 because the power allocation policy derived by Algorithm A1 is regarding the modified objective function without $(\cdot)^+$. However, the power policy is applied to the primal objective when calculating the expected rate. Finally, we run Algorithm A2 with both ordering methods. For any value of the power constraint (up to \bar{P}_o), at least one of the two ordering methods finds the optimal power allocation policy. By taking the larger rate of the two, Fig. 2 shows that our

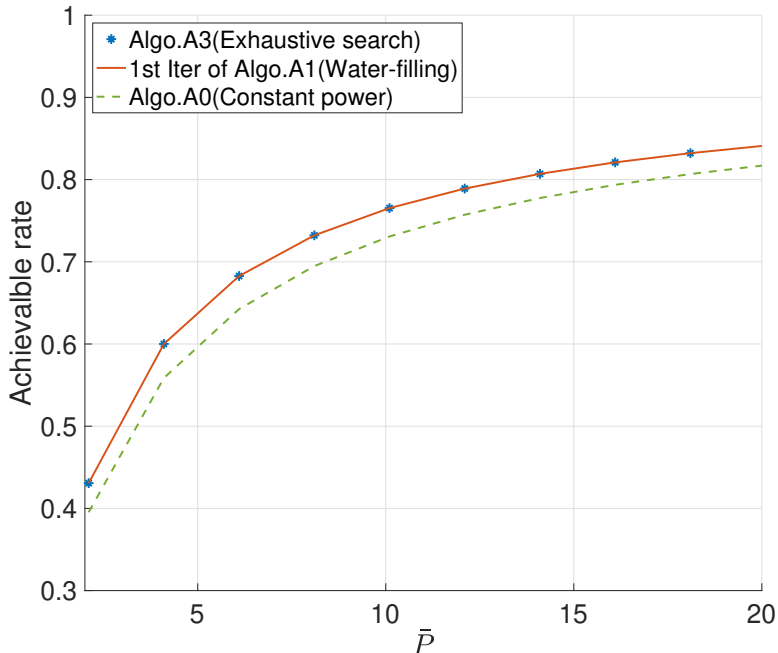


Fig. 1: Achievable rate in the symmetric power policy case with $\bar{P} > \bar{P}_o$ in Example 1. We observe that Algorithm A1 gives the same result as Algorithm A3 (hence optimal), as proven in Theorem 2.

proposed algorithm is optimal for this example for all values of power constraints. We also run an off-the-shelf algorithm (`fmincon` in MATLAB) and find that it has a similar performance to the water-filling algorithm (A1) which is suboptimal for some power constraints.

Fig. 3 shows an example for the general (asymmetric power allocation policy) case where the channel statistics are the same with the symmetric power policy example. We apply Algorithm A1, A2 and A3 for $DP1$, where both ordering methods are applied and the larger rate is taken. In the figure, we also plot the optimal result of the symmetric power policy case and the constant power case (Algorithm A0) for comparison. It can be observed that Algorithm A1 has a similar behaviour in asymmetric case compared to the symmetric case. It can match the performance of the exhaustive algorithm at some power constraint points while falls quite far at other points. However, it still results in a larger computation rate than the optimal symmetric policy. There are still sharp rate decreases in the curve of Algorithm A1 due to the same reason in Fig. 2. For the proposed algorithm (A2), it matches the rate derived from the exhaustive algorithm in all the power constraint regime in this example, which implies that the ordering methods may still work well for the general power allocation policy. We also observed that in

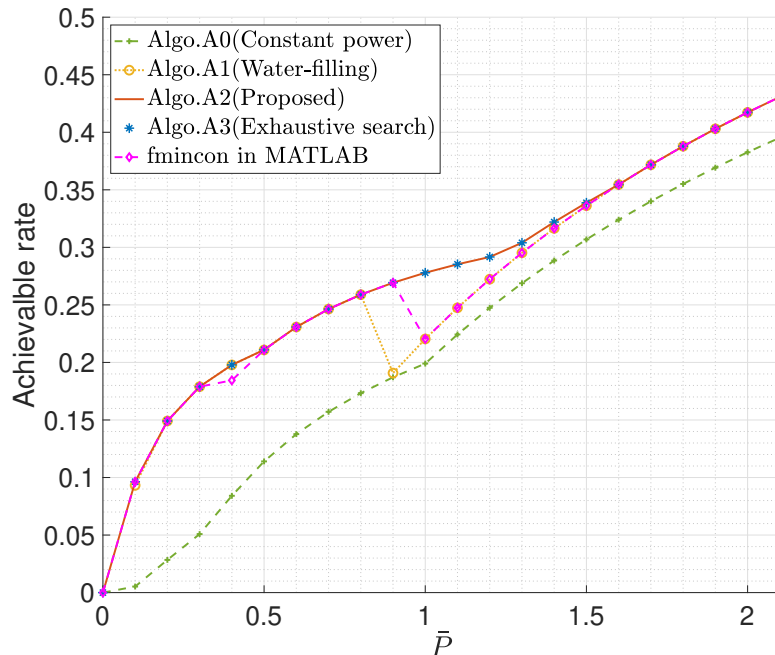


Fig. 2: Achievable rate in the symmetric power policy case with $\bar{P} \leq \bar{P}_o$ in Example 1. We observe that Algorithm A2 gives the same result as Algorithm A3 hence is optimal in this case.

the small power constraint regime, ordering method 2 is usually not optimal while in a larger power constraint regime, ordering method 1 will become suboptimal, which is consistent with our analysis in Sec.III. As the power constraint increases, both ordering methods achieve the optimal rate since all channel states become active. The off-the-shelf algorithm (fmincon in MATLAB) is also sub-optimal for some power constraints. In particular, for some small power constraints (for example, $\bar{P} = 0, 2, 0.4$), it has less rate than Algorithm A1 (water-filling algorithm).

2) **Example 2:** In this example, we will consider a two-user case, where each user has three possible channel states and their channel statistics are the same. The channel coefficients are given by 0.5, 1, 2.5. The probabilities are taken from the CDF of the Rayleigh distribution with unity scale parameter. The Probability mass function (PMF) of the channel coefficients for each user is given in Table. III. In this case, $M = 9$. The computation coefficients are still given by $\mathbf{a} = (1 \ 1)^T$. In this example, \bar{P}_o in Theorem 2 is given by 5.02. So we consider the power constraint regime $[0, 5]$. We run the power allocation algorithms A0 – A3 for both symmetric and asymmetric power policy scenarios and plot the average achievable rate against the power constraint in Fig 4. We only plot the results for $\bar{P} = [0.5, 3.5]$ because the average rates for

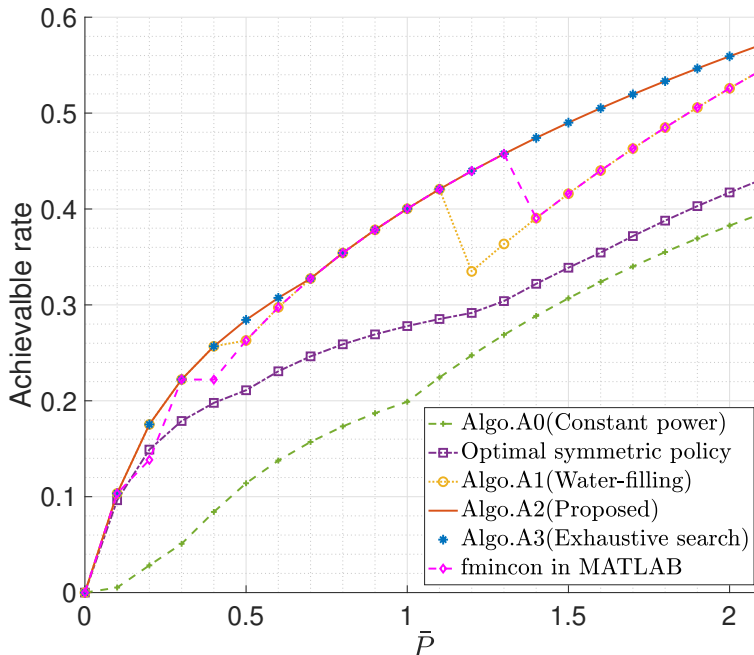


Fig. 3: Achievable rate in the asymmetric power policy case in Example 1. We observe that Algorithm A2 gives the same result as Algorithm A3 hence is optimal in this case.

Algorithms A0 – A3 are the same out of this range.

Table III: Channel state distribution in Example 2

| | | | |
|-----------|--------|--------|--------|
| $h_{1,2}$ | 0.5 | 1 | 2.5 |
| $p_{1,2}$ | 0.1175 | 0.2760 | 0.6065 |

In Fig. 4, when the power constraint is very small, for example $\bar{P} = 0.5$, the number of the active channels is expected to be small. So both symmetric and asymmetric policy scenarios have a similar average rate. When \bar{P} increases, as the number of the active channels also increases, the asymmetric policy gives a larger average rate than the symmetric policy. Regarding the performances of the different algorithms, Algorithm A3 is the best since it explores all the possible active sets. Algorithm A0 has a worst average rate with a constant gap below other algorithms. The differences between Algorithm A3, A1 and A2 are not very big. In general, Algorithm A1 is almost always below the exhaustive search except for very large power constraints, where the optimality is guaranteed by Theorem. 2. Algorithm A2 can achieve the

exhaustive search performance for some power constraints. Specifically, in the symmetric policy scenario, it finds an optimal active set for most power constraints and fails when the power constraint is between 1.25 and 2.25. In the asymmetric policy scenario, it finds an optimal active set only when the power constraint is around 1.5. In addition, Algorithm A2 gives better power policies than Algorithm A1 for almost all the power constraints in the symmetric policy scenario. It also provides a clear rate gain over Algorithm A1 when the power constraint is between 1 and 2.5 in the asymmetric power policy scenario.

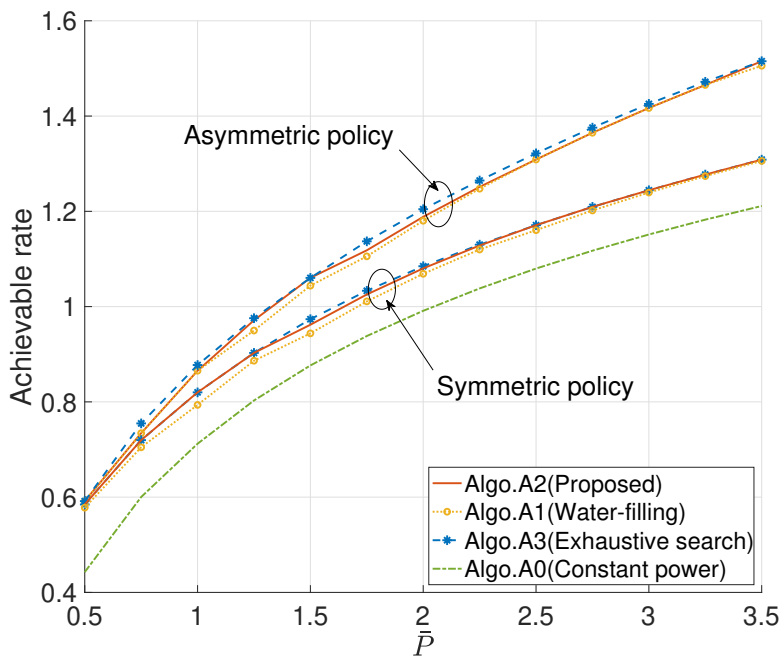


Fig. 4: Achievable rate against the power constraint in Example 2.

3) **Example 3:** In this example, we will consider a larger number of possible channel states, which is more practically relevant. Still in the two-user case with computation coefficients given by $\mathbf{a} = (1 \ 1)^T$, we assume each user has ten possible channel states. The channel coefficients are given by 0.25, 0.5, 0.75, 1, 1.25, 1.5, 2, 2.5, 3, 4. The probabilities are taken from the CDF of the Rayleigh distribution with unity scale parameter. The PMF of the channel coefficients for each user is given in Table IV. In this case, $M = 100$, and algorithm A3 is not applicable due to the complexity. Note that \bar{P}_o in Theorem 2 is given by 13.05 in this example. So we consider the power constraint regime $[0, 13]$. We run the power allocation algorithms A0 – A2 for both symmetric and asymmetric power policy scenarios and plot the average achievable rate against

the power constraint in Fig 5. We only plot the results for $\bar{P} = [0.5, 10]$ because the average rates for Algorithm A1 and Algorithm A2 are the same out of this range.

Table IV: Channel state distribution in Example 3

| $h_{1,2}$ | 0.25 | 0.5 | 0.75 | 1 | 1.25 | 1.5 | 2 | 2.5 | 3 | 4 |
|-----------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|
| $p_{1,2}$ | 0.0308 | 0.0867 | 0.1277 | 0.1483 | 0.1487 | 0.1332 | 0.1893 | 0.0914 | 0.0328 | 0.0111 |

In Fig 5, the asymmetric policy gives a larger average rate than the symmetric policy even at a low power constraint, for example $\bar{P} = 0.5$. Through all the power constraints, Algorithm A0 has a worst average rate with a constant gap below other algorithms. Although it looks like Algorithm A2 and Algorithm A1 have nearly the same average rates in the figure, we can find out the differences from their values. This can be clearly observed from the zoomed-in parts in Fig 5. If there is a rate difference between the algorithms, no matter how tiny it is, it indicates that different active sets are decided. Specifically, in the symmetric power policy scenario, Algorithm A2 gives better power policies than Algorithm A1 at all the considered power constraints. In the asymmetric power policy scenario, it finds better active sets than Algorithm A1 when the power constraint is between 4 and 8.5.

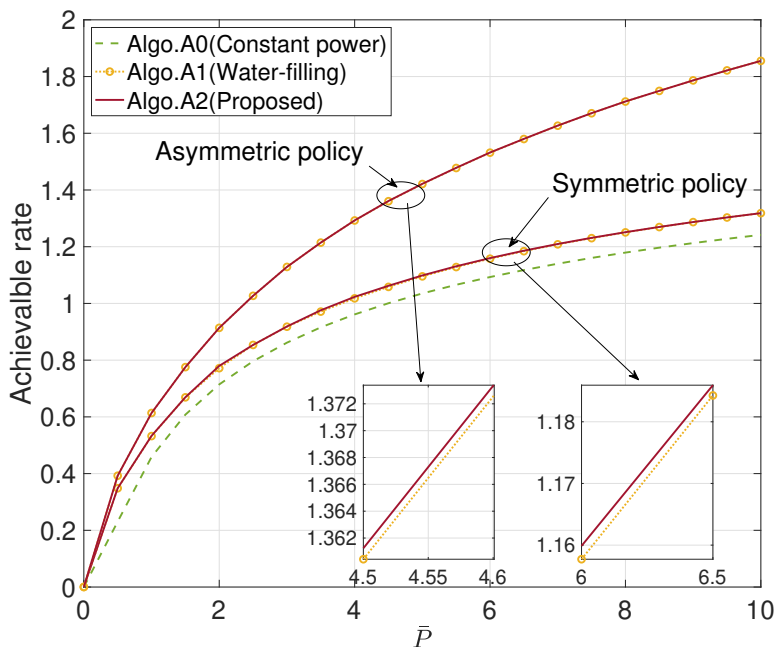


Fig. 5: Achievable rate against the power constraint in Example 3.

IV. CONTINUOUS CASE

Now we turn our attention to the case when the channel coefficients are modeled as continuous random variables. In this case, we do not have an active set of finite size as in the discrete case so the results from the previous section do not extend to the continuous case directly. However, by looking at the active “support” of the channel states, similar results could be established. Also notice that in practice, channel coefficients are always quantized so that we only need to deal with a finite number of possible channel states. In this sense, the results from the discrete case are more relevant. Nevertheless, we state a few results for the symmetric power policy in the continuous case.

A. Problem statement

Assume h_1, \dots, h_L are continuously distributed and independent with each other. Then the target problem in the symmetric power policy case is given by *CP1*.

$$\begin{aligned}
 CP1 : \quad & \max_{P(\mathbf{h})} \int_{\mathcal{D}(\mathbf{h})} g^+(\mathbf{h}, P(\mathbf{h})) f(\mathbf{h}) d\mathbf{h}, \\
 & s.t. \quad \int_{\mathcal{D}(\mathbf{h})} P(\mathbf{h}) f(\mathbf{h}) d\mathbf{h} \leq \bar{P}, \\
 & P(\mathbf{h}) \geq 0,
 \end{aligned} \tag{29}$$

where $g(\mathbf{h}, x) = \frac{1}{2} \log \frac{1 + \|\mathbf{h}\|^2 x}{\|\mathbf{a}\|^2 + \epsilon(\mathbf{h})x}$, $\epsilon(\mathbf{h}) = \|\mathbf{h}\|^2 \|\mathbf{a}\|^2 - (\mathbf{h}^T \mathbf{a})^2$, $\mathcal{D}(\mathbf{h})$ is the support domain of the channel state vector and $f(\mathbf{h})$ is the probability density function (PDF) of \mathbf{h} . Similar to the discrete case, we can define the active domain \mathcal{D}_A as $\mathcal{D}_A = \{\mathbf{h} \in \mathcal{D}(\mathbf{h}) | P(\mathbf{h}) > 0\}$, where $P(\mathbf{h})$ is any feasible power policy. Furthermore, the optimal active domain \mathcal{D}_A^* corresponding to any optimal power policy $P^*(\mathbf{h})$ is defined as $\mathcal{D}_A^* = \{\mathbf{h} \in \mathcal{D}(\mathbf{h}) | P^*(\mathbf{h}) > 0\}$. If \mathcal{D}_A^* is known, straightforwardly extending Lemma 2 to the continuous case, we can simplify *CP1* to *CP2* with its support domain $\mathcal{D}_S = \mathcal{D}_A^*$.

$$\begin{aligned}
 CP2 : \quad & \max_{P(\mathbf{h})} \int_{\mathcal{D}_S} g(\mathbf{h}, P(\mathbf{h})) f(\mathbf{h}) d\mathbf{h}, \\
 & s.t. \quad \int_{\mathcal{D}_S} P(\mathbf{h}) f(\mathbf{h}) d\mathbf{h} \leq \bar{P}, \\
 & P(\mathbf{h}) \geq 0.
 \end{aligned} \tag{30}$$

Note that $CP2$ is a convex problem, and its optimal solution can also be given by KKT conditions, which can be proven in a similar way as Lemma 3. Specifically, let us define

$$P(\mathbf{h}, \mu) = \begin{cases} \left(\frac{1}{\mu} - \frac{1}{\|\mathbf{h}\|^2}\right)^+, & \mathbf{h} \text{ is collinear with } \mathbf{a} \\ \left(\frac{-b(\mathbf{h}) + \sqrt{b(\mathbf{h})^2 - 4d(\mathbf{h})c(\mathbf{h}, \mu)}}{2d(\mathbf{h})}\right)^+, & \mathbf{h} \text{ is not collinear with } \mathbf{a} \end{cases}, \quad (31)$$

where $d(\mathbf{h}) = \|\mathbf{h}\|^2 \epsilon(\mathbf{h})$, $b(\mathbf{h}) = \|\mathbf{h}\|^2 \|\mathbf{a}\|^2 + \epsilon(\mathbf{h})$, $c(\mathbf{h}, \mu) = \|\mathbf{a}\|^2 - (\mathbf{h}^T \mathbf{a})^2 / \mu$ and $\mu > 0$. Then the optimal solution to $CP2$ can be given by

$$P^*(\mathbf{h}) = P(\mathbf{h}, \mu^*), \quad (32)$$

with

$$\int_{\mathcal{D}_S} P(\mathbf{h}, \mu^*) f(\mathbf{h}) d\mathbf{h} = \bar{P}. \quad (33)$$

Practically, bisection algorithm can be used to derive the optimal μ^* , and then the optimal solution is given by $P(\mathbf{h}, \mu^*)$. This is presented in Algorithm 2.

Algorithm 2 Optimal algorithm for $CP2$

- 1: Define $q(\mu) = \int_{\mathcal{D}_S} P(\mathbf{h}, \mu) f(\mathbf{h}) d\mathbf{h}$;
 - 2: Choose a small enough μ_l such that $q(\mu_l) > \bar{P}$;
 - 3: Choose a large enough μ_u such that $q(\mu_u) < \bar{P}$;
 - 4: $\mu \leftarrow \frac{\mu_l + \mu_u}{2}$;
 - 5: **while** $\bar{P} - q(\mu) > \epsilon$ **||** $\bar{P} - q(\mu) < 0$ **do** $\triangleright \epsilon$ is a small number, e.g. 10^{-3}
 - 6: **if** $\bar{P} - q(\mu) > 0$ **then**
 - 7: $\mu_u \leftarrow \mu$;
 - 8: **else**
 - 9: $\mu_l \leftarrow \mu$;
 - 10: **end if**
 - 11: $\mu \leftarrow \frac{\mu_l + \mu_u}{2}$;
 - 12: **end while**
 - 13: $P^*(\mathbf{h}) = P(\mathbf{h}, \mu)$.
-

To study the property of the optimal active domain in the continuous case, we define the good domain \mathcal{D}_G and the bad domain \mathcal{D}_B as $\mathcal{D}_G = \{\mathbf{h} \in \mathcal{D}(\mathbf{h}) \mid \|\mathbf{h}\|^2 > \epsilon(\mathbf{h})\}$ and $\mathcal{D}_B = \{\mathbf{h} \in \mathcal{D}(\mathbf{h}) \mid \|\mathbf{h}\|^2 \leq \epsilon(\mathbf{h})\}$ respectively. Using the same argument as in the discrete case, it is easy to see that the optimal domain does not contain \mathcal{D}_B , that is $\mathcal{D}_A^* \cap \mathcal{D}_B = \emptyset$. It is tempting to conjecture that similar to the discrete case (Theorem 2), there exists a threshold on the power constraint, above which the optimal domain \mathcal{D}_A^* is equal the good domain. However, we now show that this is not always the case in the continuous case.

Let us assume $\mathcal{D}_A^* = \mathcal{D}_G$. Then the optimal solution can be given by Algorithm 2 and its analytical form is given by $P(\mathbf{h}, \mu^*)$ in (31). Using the same argument as in Lemma 1 for the continuous case, it follows that

$$P(\mathbf{h}, \mu^*) > \frac{\|\mathbf{a}\|^2 - 1}{\|\mathbf{h}\|^2 - \epsilon(\mathbf{h})}, \forall \mathbf{h} \in \mathcal{D}_G, \quad (34)$$

Substituting (32) into (34), it holds that

$$\frac{1}{\mu^*} > \begin{cases} \frac{\|\mathbf{a}\|^2}{\|\mathbf{h}\|^2}, & \epsilon(\mathbf{h}) = 0, \\ \frac{(\mathbf{h}^T \mathbf{a})^2}{\|\mathbf{h}\|^2 - \epsilon(\mathbf{h})}, & \epsilon(\mathbf{h}) \neq 0. \end{cases} \quad (35)$$

This means when $\mathcal{D}_A^* = \mathcal{D}_G$, the above inequality holds for all $\mathbf{h} \in \mathcal{D}_G$. However, from the definition of the good domain $\mathcal{D}_G = \{\mathbf{h} \in \mathcal{D}(\mathbf{h}) \mid \|\mathbf{h}\|^2 > \epsilon(\mathbf{h})\}$, the inequality in (35) implies that $1/\mu^*$ cannot be bounded from above if there exists a sequence of $\mathbf{h} \in \mathcal{D}_G$ such that $\|\mathbf{h}\|^2$ can approach $\epsilon(\mathbf{h})$ asymptotically. On the other hand, an unbounded $1/\mu^*$ implies that the power constraint \bar{P} cannot be bounded from above as well. This is because of (33) and the fact that $P(\mathbf{h}, \mu)$ in (31) grows at least linearly with $\sqrt{1/\mu}$. This shows that in general there does not exist an upper bound on the power constraint, above which \mathcal{D}_A^* is equal to \mathcal{D}_G in the continuous case. In the following, we study the power allocation algorithms for the continuous case.

B. Power allocation algorithms for the continuous case

Unlike the discrete case, we cannot apply Algorithm A3 (exhaustive search) in the continuous case to derive the optimal solution. However, the algorithms A0 and A1 can be straightforwardly extended to the continuous case. Let us keep their captions as in the discrete case. Moreover, the proposed ordering methods can also be extended to the continuous case with the ordering metric defined by $O^{(1)} = \|\mathbf{h}\|^2 - (\|\mathbf{h}\|^2 \|\mathbf{a}\|^2 - (\mathbf{h}^T \mathbf{a})^2)$ and $O^{(2)} = \frac{\|\mathbf{h}\|^2}{\|\mathbf{h}\|^2 \|\mathbf{a}\|^2 - (\mathbf{h}^T \mathbf{a})^2}$. Unlike the discrete case, the channel states in the continuous case cannot be ordered into a finitely countable set. However, the intuition from the discrete case still carries over, namely, the larger the ordering criterion is, the better the channel is. So we will search the optimal active domain in the direction depending on the ordering criterion. In each iteration, the active domain will be shaped further along the direction where the ordering criterion is larger. This intuition is substantiated in Algorithm A4.

Algorithm A4 Proposed iterative algorithm for $CP1$

- 1: Define $r(P(\mathbf{h})) = \int_{\mathcal{D}_S} g(\mathbf{h}, P(\mathbf{h}))f(\mathbf{h})d\mathbf{h}$;
 - 2: $C_t = 0, s = 0.1$; $\triangleright C_t$ is the shaping parameter, s is the step size
 - 3: $P(\mathbf{h}) = P'(\mathbf{h}) = 0$;
 - 4: **while** $r(P'(\mathbf{h})) - r(P(\mathbf{h})) > \epsilon$ **do** $\triangleright \epsilon$ is a small number, e.g. 10^{-3}
 - 5: $\mathcal{D}_S \leftarrow \{\mathbf{h} | O(\mathbf{h}) > C_t\}$; $\triangleright O(\mathbf{h})$ is the ordering method, it can be either $O^{(1)}$ or $O^{(2)}$ in this case
 - 6: $P(\mathbf{h}) \leftarrow P'(\mathbf{h})$;
 - 7: Run Algorithm 2 \rightarrow Solution $P'(\mathbf{h})$;
 - 8: $C_t = C_t + s$;
 - 9: **end while**
 - 10: $P^*(\mathbf{h}) = P(\mathbf{h})$.
-

C. Numerical results for the continuous case

We consider a two-user case where $L = 2$ and $\mathbf{a} = (1 \ 1)^T$. Assume the channel state vector follow the Gaussian distribution with zero mean and unit variance. Since we assume non-negative real channel coefficients, the PDF is given by

$$f(h_1, h_2) = \frac{2}{\pi} e^{-\frac{h_1^2 + h_2^2}{2}}. \quad (36)$$

Fig. 6 shows the achievable computation rate under Algorithm A0, A1 and A4 with $O^{(1)}$. It can be observed that Algorithm A0 gives the worst expected rate, while the proposed iterative algorithm A4 has the best rate. However, it only outperforms Algorithm A1 with a small gain.

V. CONCLUSION

The power allocation problem for CF over fading channels is nontrivial due to its non-convexity and the $\log^+(\cdot)$ function in the rate expression. We investigate the properties of the optimal solutions. We notice that the classical water-filling like algorithm provides a sub-optimal solution with an implicit ordering methods for the channel states. To improve that, we propose two explicit ordering methods and an order-based iterative algorithm to determine the active set and find out the corresponding power policy. The proposed iterative algorithm gives a better rate than the classical algorithm while remains a linear complexity. It achieves optimality in certain examples, but in general gives good but suboptimal solutions. This possibly implies that the two ordering methods are not able to capture all the properties of the channel states. One way to improve the proposed algorithm is to find out a more judicious ordering method and feed it into the iterative algorithm frame. A potential ordering method is to combine the two proposed methods with different weights. In this case, the future research may focus on how to decide the weights.

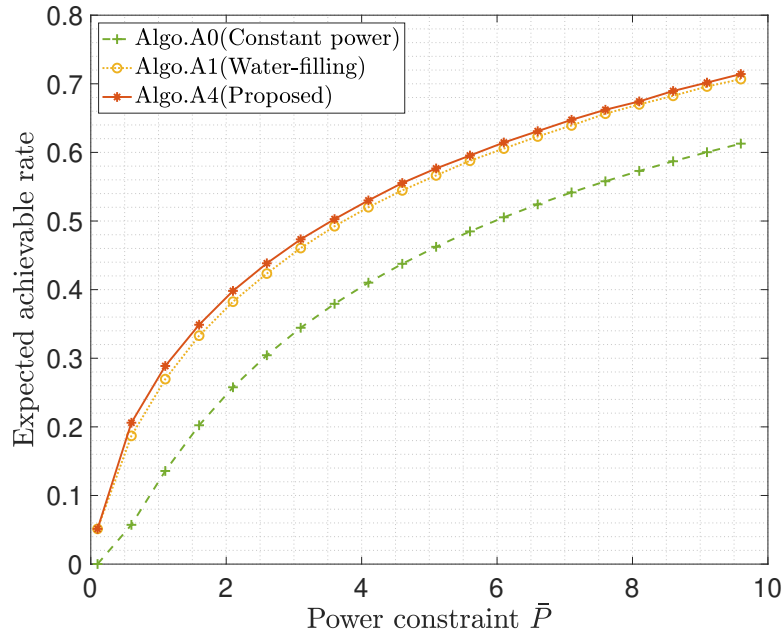


Fig. 6: Performance comparison for different algorithms in the continuous case

Moreover, the power constraint could be included in the ordering methods as well. It is possible that at different power constraints the ordering of the channel states are different.

REFERENCES

- [1] D. Tse and P. Viswanath, *Fundamentals of Wireless Communication*. Cambridge University Press, 2005.
- [2] R. Ahlswede, N. Cai, S.-Y. Li, and R. Yeung, “Network information flow,” *IEEE Transactions on Information Theory*, vol. 46, no. 4, pp. 1204–1216, 2000.
- [3] S. Zhang, S. C. Liew, and P. P. Lam, “Hot topic: Physical-layer network coding,” in *Proceedings of the 12th Annual International Conference on Mobile Computing and Networking*, ser. MobiCom ’06. New York, NY, USA: Association for Computing Machinery, 2006, p. 358–365. [Online]. Available: <https://doi.org/10.1145/1161089.1161129>
- [4] B. Nazer and M. Gastpar, “Compute-and-forward: Harnessing interference through structured codes,” *IEEE Transactions on Information Theory*, vol. 57, no. 10, pp. 6463–6486, 2011.
- [5] J. Zhu and M. Gastpar, “Gaussian multiple access via compute-and-forward,” *IEEE Transactions on Information Theory*, vol. 63, no. 5, pp. 2678–2695, 2017.
- [6] B. Nazer, V. R. Cadambe, V. Ntranos, and G. Caire, “Expanding the compute-and-forward framework: Unequal powers, signal levels, and multiple linear combinations,” *IEEE Transactions on Information Theory*, vol. 62, no. 9, pp. 4879–4909, 2016.
- [7] S.-N. Hong and G. Caire, “Compute-and-forward strategies for cooperative distributed antenna systems,” *IEEE Transactions on Information Theory*, vol. 59, no. 9, pp. 5227–5243, 2013.
- [8] O. Ordentlich and Y. Polyanskiy, “Low complexity schemes for the random access gaussian channel,” in *2017 IEEE International Symposium on Information Theory (ISIT)*, 2017, pp. 2528–2532.

- [9] J. Zhan, B. Nazer, U. Erez, and M. Gastpar, "Integer-forcing linear receivers," *IEEE Transactions on Information Theory*, vol. 60, no. 12, pp. 7661–7685, 2014.
- [10] W. He, B. Nazer, and S. Shamai Shitz, "Uplink-downlink duality for integer-forcing," *IEEE Transactions on Information Theory*, vol. 64, no. 3, pp. 1992–2011, 2018.
- [11] L. Zhang, J. Zhu, and J. Evans, "Compute-forward multiple access for gaussian fast fading channels," in *2024 IEEE International Symposium on Information Theory (ISIT)*, 2024, pp. 2999–3004.
- [12] L. Wei and W. Chen, "Compute-and-forward network coding design over multi-source multi-relay channels," *IEEE Transactions on Wireless Communications*, vol. 11, no. 9, pp. 3348–3357, 2012.
- [13] S. Sahraei and M. Gastpar, "Compute-and-forward: Finding the best equation," in *2014 52nd Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, 2014, pp. 227–233.
- [14] P. Tran and J. S. Lehnert, "Joint optimization of power allocation and integer coefficients of relay functions in compute-and-forward relay networks," in *2014 International Conference on Advanced Technologies for Communications (ATC 2014)*, 2014, pp. 669–674.
- [15] M. El Soussi, A. Zaidi, and L. Vandendorpe, "Compute-and-forward on a multiaccess relay channel: Coding and symmetric-rate optimization," *IEEE Transactions on Wireless Communications*, vol. 13, no. 4, pp. 1932–1947, 2014.
- [16] Z. Chen, P. Fan, and K. B. Letaief, "Compute-and-forward: Optimization over multisource–multirelay networks," *IEEE Transactions on Vehicular Technology*, vol. 64, no. 5, pp. 1806–1818, 2015.
- [17] 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.
- [18] A. El Gamal and Y.-H. Kim, *Network Information Theory*. Cambridge University Press, 2011.
- [19] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge University Press, 2004.