Capacity Bounds for Parallel Optical Wireless Channels

Anas Chaaban, Zouheir Rezki, and Mohamed-Slim Alouini

Abstract

A system consisting of parallel optical wireless channels with a total average intensity constraint is studied. Capacity upper and lower bounds for this system are derived. Under perfect channel-state information at the transmitter (CSIT), the bounds have to be optimized with respect to the power allocation over the parallel channels. The optimization of the lower bound is non-convex, however, the KKT conditions can be used to find a list of possible solutions one of which is optimal. The optimal solution can then be found by an exhaustive search algorithm, which is computationally expensive. To overcome this, we propose low-complexity power allocation algorithms which are nearly optimal. The optimized capacity lower bound nearly coincides with the capacity at high SNR. Without CSIT, our capacity bounds lead to upper and lower bounds on the outage probability. The outage probability bounds meet at high SNR. The system with average and peak intensity constraints is also discussed.

I. Introduction

One of the most challenging problems facing today's wireless communications is spectrum scarcity. Due to this problem, it is becoming extremely difficult to cope with the increasing demand for high data-rates. This has motivated researchers to explore new frequency bands for wireless communications. One example is communication using an optical carrier, which is commonly known as optical wireless communication (OWC).

Due to the above challenge, OWC has witnessed increasing research attention recently. The main focus was towards systems employing intensity-modulation and direct-detection (IM-DD), due to the practical simplicity and low cost of IM-DD. For instance, [1], [2] studied constellation design for

The authors are with the Division of Computer, Electrical, and Mathematical Sciences and Engineering (CEMSE) at King Abdullah University of Science and Technology (KAUST), Thuwal, Saudi Arabia. Email: {anas.chaaban,zouheir.rezki,slim.alouini}@kaust.edu.sa.

This work is supported in part by King Abdulaziz City of Science and Technology (KACST) under grant AT-34-145.

IM-DD systems, and [3]–[5] studied the performance of OFDM. Additionally, [6] studied achievable rates using discrete input distributions, [7] investigated the performance of on-off keying, [8] analyzed the outage performance, and [9]–[11] considered multi-hop systems. For more details about the recent advances in this field, the reader is referred to [12], [13] and references therein.

The transmission rate and outage probability performance of OWC is significantly affected by the distance between the transmitter and the receiver, and by weather conditions. These effects can be mitigated by employing multiple transmit and receive apertures (multiple lasers and photo-diodes, e.g.), leading to an optical MIMO channel. The optical MIMO channel was studied in [14]–[18]. In such a MIMO channel, if transmit-receive pairs are aligned and non-interfering (using narrow laser beams, e.g.), we get a system of parallel channels. Otherwise, if transmit-receive pairs interfere with each other, then we do not a system of parallel channels. However, the corresponding MIMO channel can be transformed –with some loss– to a set of parallel channels by channel inversion at the receiver [19]. Note also that IM-DD systems employing RGB color-frequency modulation [20], [21] with little or no cross-talk between colors can also be modeled as a system of parallel channels. Thus, it is important to study the performance of parallel OWC channels employing IM-DD, in order to reach a better understanding of the above systems.

In this paper, we approach this problem from an information-theoretic perspective. In particular, we study the performance of a system of parallel channels measured by its capacity. We consider N parallel IM-DD channels with a total average intensity constraint. The capacity for $N = 1$ was studied in [22]–[25], where capacity upper and lower bounds were derived. We adapt those bounds to obtain capacity upper and lower bounds for $N > 1$. Those bounds have to be optimized with respect to power allocation over the channels, which is the main new ingredient. Such an optimization is important for enhancing the system performance.

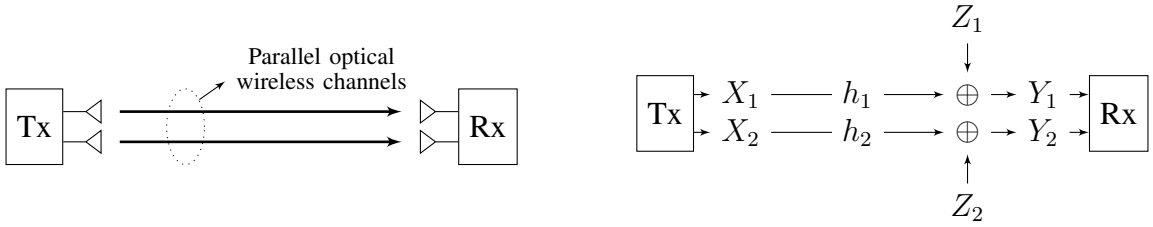
At first, we assume that channel state information (CSI) is available at the transmitter. Note that the availability of CSI at the transmitter (CSIT) is not a strong assumption in OWC, since the coherence time of OWC channels is much larger than the symbol duration [8]. Thus, channel estimation at the receiver and feedback of CSI from the receiver to the transmitter can be achieved in negligible time without considerably affecting the overall system performance. In a full-duplex system, this channel state can be estimated directly at the transmitter due to channel reciprocity [12]. We focus on optimizing the rate achieved by using an exponentially distributed input over each of the channels. Note that an exponentially distributed input is optimal at high SNR for $N = 1$. The resulting optimization problem for $N > 1$ is non-convex, contrary to its coherent-detection

counterpart where water-filling is the optimal solution [26]. We formulate the power allocation problem and then describe its solution by using the KKT conditions [27]. This leads to a list of potential candidates, one of which is the optimal power allocation. The list size is exponential in the number of channels N, which renders finding the optimal solution difficult (exponential complexity). Instead of searching over this list, we propose a simple (linear complexity) algorithm which almost always finds the optimal power allocation. The main idea of this algorithm stems from the behavior of the candidate solutions at moderate SNR. Their behavior allows us to eliminate many candidates, thus reducing complexity. We propose another variant of this algorithm with slightly lower complexity. Consequently, we obtain a simple capacity lower bound, which we prove to be asymptotically tight at high SNR.

Then, we consider a system with no CSIT. In this case, the capacity in the strict Shannon sense is zero, since for any non-zero rate, there is non-zero probability that the channel can not support this rate. In this case, the performance of the channel is reflected by its outage probability [28], i.e., the probability that the channel can not support the desired target rate. Note that optimal power allocation in a system of parallel channels is not possible in this case, due to the absence of CSIT. A feasible strategy however is to distribute the power equally over the N channels. Using this allocation, we derive an outage probability upper bound. An outage probability lower bound can be obtained by using the derived capacity upper bound for a system with CSIT. We show that the outage probability bounds coincide at high SNR, leading to the high-SNR outage-capacity of the system.

Finally, we consider the impact of a peak intensity constraint on each of the N transmitters. For this case, if CSI is available at the transmitter, we derive capacity lower bounds and upper bounds. We show that if the total average constraint is larger than $N/2$ times the peak constraint, then equal allocation is optimal. Otherwise, we propose a power allocation algorithm which yields a good performance. Moreover, we show that this performance is nearly optimal at high SNR, where the gap between the upper and lower bounds becomes negligible. In the absence of CSIT, we derive bounds on the outage probability, which are fairly tight at high SNR as well.

The rest of the paper is organized as follows. Sec. II presents the system model, and Sec. III presents some preliminaries. The channel with only an average constraint and with CSIT is discussed in Sec. IV. The case with no CSIT is studied in Sec. V. The impact of a peak intensity constraint is discussed in Sec. VI, and the paper is concluded in Sec. VII.



(a) An optical wireless communication system consisting of two transmit apertures (Lasers e.g.) and two detectors. The total average optical intensity is constrained by $\mathcal{E}$.

(b) Parallel Gaussian point-to-point channels. Here X_i is the optical intensity, h_i is a channel gain, and Z_i is Gaussian noise. The optical intensity is constrained by $\mathbb{E}_{X_1}[X_1] + \mathbb{E}_{X_2}[X_2] \leq \mathcal{E}$.

Fig. 1: A system consisting of two parallel optical wireless channels can be modeled as two Gaussian point-to-point channels.

II. Model

We consider an optical wireless communications (OWC) system consisting of a set of N parallel channels, each of which is an IM-DD point-to-point (P2P) channel (see Fig. 1). The transmit signal over channel $i \in \mathcal{N} = \{1, \cdots, N\}$ is $X_i \in \mathbb{R}_+$ where the non-negativity of X_i follows since X_i denotes the light intensity over the i-th channel. Due to practical and safety considerations, the signal X_i also has to satisfy a total average intensity constraint given by $\sum_{i\in\mathcal{N}} \mathcal{E}_i \leq \mathcal{E}$ where $\mathcal{E}_i = \mathbb{E}_{X_i}[X_i]$.[1] This condition can be written as $\|\boldsymbol{\mathcal{E}}\|_1 \leq \mathcal{E}$ where $\|\boldsymbol{\mathcal{E}}\|_1$ is the ℓ_1-norm of $\boldsymbol{\mathcal{E}} = (\mathcal{E}_1, \cdots, \mathcal{E}_N)$.

In addition to the average constraint, the input signal X_i might need to satisfy a peak intensity constraint given by $X_i \leq \mathcal{A}$. For simplicity of exposition, we ignore the peak constraint at fist and focus on the case with $\mathcal{A} = \infty$. We discuss that case $\mathcal{A} < \infty$ at the end of the paper.

The received signal over channel i is

$$Y_i = h_i X_i + Z_i, \tag{1}$$

where the noise Z_i is i.i.d. (over time) Gaussian with zero mean and unit variance, and $h_i \in \mathbb{R}_+$ represents the channel scaling factor. We assume without loss of generality that $h_1 \geq h_2 \geq \cdots \geq h_N$, and we define $\boldsymbol{h} = (h_1, \cdots, h_N)$. We assume that $\boldsymbol{h}$ retains the same value during a transmission block, which is a realistic assumption since optical channels fade very slowly in comparison to the symbol duration.

[1]Throughout the paper, we use $\mathbb{E}_X[\cdot]$ to denote the expectation with respect to the distribution $p(x)$ of a random variable X.

We are interested in studying the capacity $C(\boldsymbol{h}, \mathcal{E})$ of this system, defined as follows. Let $n \in \mathbb{N}_+$ be the code-length used by the transmitter. Also, let $P_e^{(n)}$ be the error probability, i.e., the probability that the receiver decodes a different codeword than the one sent by the transmitter. The capacity $C(\boldsymbol{h}, \mathcal{E})$ is the highest rate R, for which there exists a coding scheme that delivers R nats of information per transmission to the receiver, while satisfying $P_e^{(n)} \to 0$ as $n \to \infty$.

We study $C(\boldsymbol{h}, \mathcal{E})$ under two considerations. First, we consider the case of perfect CSIT, where the channel state $\boldsymbol{h}$ is known at the transmitter. Second, we consider the case of no CSIT. Before we proceed, let us review some bounds on the capacity of the IM-DD P2P channel.

III. Preliminaries

Capacity upper bounds and achievable rates (capacity lower bounds) for the IM-DD P2P channel ($N = 1$) have been derived in [6], [22]–[24]. Next, we present some bounds that will be needed in the sequel. Those bounds will be stated for a P2P channel with input X satisfying $\mathbb{E}[X] \leq \mathcal{E}$ and output

$$Y = hX + Z$$

where Z is Gaussian with zero mean and unit variance and $h \in \mathbb{R}_+$.

Achievable rates for this channel have been derived in [6], [22], [24]. In this paper, we are interested in achievable rates of the form $\frac{1}{2}\log(1 + c^2\mathcal{E}^2)$ for some constant c.[2] This captures the behavior of achievable rates in [22], [24]. For instance, when X follows an exponential distribution [22], the achievable rate is given by

$$r(h, \mathcal{E}) = \frac{1}{2}\log\left(1 + \frac{eh^2\mathcal{E}^2}{2\pi}\right). \tag{2}$$

In this case, $c = h\sqrt{\frac{e}{2\pi}}$. Thus, $C(h, \mathcal{E}) \geq r(h, \mathcal{E})$.

Several upper bounds on the capacity of this channel exist. We restrict ourselves to the upper bound $C(h, \mathcal{E}) \leq \overline{r}(h, \mathcal{E})$ where

$$\overline{r}(h, \mathcal{E}) = \sup_{\delta \in [0,1]} b(h, \mathcal{E}, \delta), \tag{3}$$

and $b(h, \mathcal{E}, \delta)$ is given in (4) at the top of the page. This upper bound was given in [24], and is tight at high SNR ($\mathcal{E} \gg 1$).

These two bounds will be used in what follows. Next, we consider a system of N parallel channels with prefect CSIT.

[2]All logarithms in this paper are natural logarithms.

$$b(h,\mathcal{E},\delta)=\frac{\delta}{2}\log\left(\frac{eh^2\mathcal{E}^2}{2\pi}\right)-\log\left((1-\delta)^{1-\delta}\delta^{\frac{3\delta}{2}}\right). \tag{4}$$

IV. Parallel Channels with CSIT

Consider a system consisting of N parallel channels. In this section, we derive bounds on the capacity of this system.

A. Capacity Bounds

First, we note that the capacity of the system is equal to the sum of the capacities of the N channels $C(h_i,\mathcal{E}_i)$, optimized over the set of feasible power allocations $\boldsymbol{\mathcal{E}}$. That is,

$$C(\boldsymbol{h},\mathcal{E})=\max_{\boldsymbol{\mathcal{E}}}\sum_{i\in\mathcal{N}}C(h_i,\mathcal{E}_i), \tag{5}$$

where $\boldsymbol{\mathcal{E}}\in\mathbb{R}_+^N$ such that $\|\boldsymbol{\mathcal{E}}\|_1\leq\mathcal{E}$. This is proved as follows.

The capacity $C(\boldsymbol{h},\mathcal{E})$ is given by [29]

$$C(\boldsymbol{h},\mathcal{E})=\max_{p(x_1,\cdots,x_N)\in\mathcal{P}}I(X_1,\cdots,X_N;Y_1,\cdots,Y_N), \tag{6}$$

where $\mathcal{P}$ is the set of distributions of $(X_1,\cdots,X_N)\in\mathbb{R}_+^N$ satisfying $\sum_{i\in\mathcal{N}}\mathbb{E}_{X_i}[X_i]\leq\mathcal{E}$. This can be written as

$$C(\boldsymbol{h},\mathcal{E})=\max_{\boldsymbol{\mathcal{E}}\in\mathcal{S}}\max_{p(x_1,\cdots,x_N)\in\mathcal{Q}(\boldsymbol{\mathcal{E}})}I(X_1,\cdots,X_N;Y_1,\cdots,Y_N). \tag{7}$$

where $\mathcal{S}=\{\boldsymbol{\mathcal{E}}\in\mathbb{R}_+^N|\|\boldsymbol{\mathcal{E}}\|_1\leq\mathcal{E}\}$, and $\mathcal{Q}(\boldsymbol{\mathcal{E}})$ is the subset of $\mathcal{P}$ satisfying $\mathbb{E}_{X_i}[X_i]=\mathcal{E}_i$. Using similar steps as [29, Sec. 9.4], we can write

$$I(X_1,\cdots,X_N;Y_1,\cdots,Y_N)\leq\sum_{i\in\mathcal{N}}I(X_i;Y_i). \tag{8}$$

Furthermore,

$$\max_{p(x_1,\cdots,x_N)\in\mathcal{Q}(\boldsymbol{\mathcal{E}})}\sum_{i\in\mathcal{N}}I(X_i;Y_i)\leq\sum_{i\in\mathcal{N}}\max_{p(x_i)\in\mathcal{Q}_i(\mathcal{E}_i)}I(X_i;Y_i) \tag{9}$$

$$=\sum_{i\in\mathcal{N}}C(h_i,\mathcal{E}_i), \tag{10}$$

where $\mathcal{Q}_i(\mathcal{E}_i)$ is the set of distributions of $X_i\geq 0$ satisfying $\mathbb{E}_{X_i}[X_i]=\mathcal{E}_i$. Therefore,

$$C(\boldsymbol{h},\mathcal{E})\leq\max_{\boldsymbol{\mathcal{E}}\in\mathcal{S}}\sum_{i\in\mathcal{N}}C(h_i,\mathcal{E}_i). \tag{11}$$

Since this upper bound is achievable by coding independently over each channel, we conclude that

$$C(\boldsymbol{h}, \mathcal{E}) = \max_{\boldsymbol{\mathcal{E}} \in \mathcal{S}} \sum_{i \in \mathcal{N}} C(h_i, \mathcal{E}_i). \tag{12}$$

Unfortunately, $C(h_i, \mathcal{E}_i)$ is not known in closed form. However, we can bound it by the bounds given in Sec. III. This leads to the following statements.

Theorem 1: The capacity $C(\boldsymbol{h}, \mathcal{E})$ of a system of N parallel channels is bounded by $R(\boldsymbol{h}, \mathcal{E}) \leq C(\boldsymbol{h}, \mathcal{E}) \leq \overline{C}(\boldsymbol{h}, \mathcal{E})$ where

$$R(\boldsymbol{h}, \mathcal{E}) = \max_{\boldsymbol{\mathcal{E}} \in \mathcal{S}} \sum_{i \in \mathcal{N}} r(h_i, \mathcal{E}_i), \tag{13}$$

$$\overline{C}(\boldsymbol{h}, \mathcal{E}) = \max_{\boldsymbol{\mathcal{E}} \in \mathcal{S}} \sum_{i \in \mathcal{N}} \overline{r}(h_i, \mathcal{E}_i), \tag{14}$$

$\mathcal{S} = \{\boldsymbol{\mathcal{E}} \in \mathbb{R}_+^N | \|\boldsymbol{\mathcal{E}}\|_1 \leq \mathcal{E}\}$, and $r(h_i, \mathcal{E}_i)$ and $\overline{r}(h_i, \mathcal{E}_i)$ are given in (2) and (3), respectively.

Proof: Using (12) and the bounds $r(h_i, \mathcal{E}_i)$ and $\overline{r}(h_i, \mathcal{E}_i)$ are given in (2) and (3). ■

The upper bound (14) will be maximized numerically in what follows. The lower bound (13) is achievable by independent coding over each channel, using an exponentially distributed input with the given power allocation $\mathcal{E}_i$. To maximize this lower bound, it is required to find a resource allocation $\boldsymbol{\mathcal{E}}$, which distributes the available resource $\mathcal{E}$ optimally over a subset (or all) of the N channels. Note that under perfect CSIT, this optimization can be performed by the transmitter. The rest of this section focuses on this optimization.

B. Power Allocation

As stated earlier, we focus on achievable rates of the form $\frac{1}{2}\log(1 + c^2\mathcal{E}^2)$. Thus, we consider a system of N parallel channels, where the achievable rate over channel $i \in \mathcal{N}$ is $\frac{1}{2}\log(1 + c_i^2\mathcal{E}_i^2)$ for some c_i, where $\mathcal{E}_i$ is the average optical intensity of channel i (as in (13)).

The maximal achievable rate is given by the solution of

$$\begin{aligned} \max_{\boldsymbol{\mathcal{E}}} \quad & f(\boldsymbol{\mathcal{E}}) \\ \text{s.t.} \quad & \boldsymbol{\mathcal{E}} \in \mathbb{R}_+^N, \ \|\boldsymbol{\mathcal{E}}\|_1 \leq \mathcal{E}, \end{aligned} \tag{15}$$

where

$$f(\boldsymbol{\mathcal{E}}) = \frac{1}{2} \sum_{i \in \mathcal{N}} \log\left(1 + c_i^2 \mathcal{E}_i^2\right). \tag{16}$$

Note that $f(\boldsymbol{\mathcal{E}})$ is non-concave in $\mathcal{E}_i$, which renders the problem different from the standard water-filling power-allocation problem [29, Sec. 9.4]. Nevertheless, the KKT conditions [27] can be used to simplify the solution of this problem. The Lagrangian associated with the problem is given by

$$L = -f(\boldsymbol{\mathcal{E}}) + \lambda \left(\|\boldsymbol{\mathcal{E}}\|_1 - \mathcal{E} \right) - \sum_{i \in \mathcal{N}} \lambda_i \mathcal{E}_i, \tag{17}$$

and the corresponding KKT conditions are

$$\begin{aligned} &\mathcal{E}_i, \lambda, \lambda_i \geq 0, \quad \lambda_i \mathcal{E}_i = 0, \quad \frac{\partial L}{\partial \mathcal{E}_i} = 0, \ \forall i \in \mathcal{N} \\ &\|\boldsymbol{\mathcal{E}}\|_1 \leq \mathcal{E}, \quad \lambda \left(\|\boldsymbol{\mathcal{E}}\|_1 - \mathcal{E} \right) = 0. \end{aligned} \tag{18}$$

By solving $\frac{\partial L}{\partial \mathcal{E}_i} = 0$ for λ_i and substituting in $\lambda_i \mathcal{E}_i = 0$ we get

$$\underbrace{\left(\lambda - \frac{c_i^2 \mathcal{E}_i}{1 + c_i^2 \mathcal{E}_i^2} \right)}_{= \lambda_i} \mathcal{E}_i = 0. \tag{19}$$

This equality holds if $\mathcal{E}_i = 0$. Otherwise, if $\mathcal{E}_i > 0$, then this equality necessitates $\lambda_i = 0$. This yields

$$\mathcal{E}_i(\lambda) \in \begin{cases} \{0, \mathcal{E}_i^+(\lambda), \mathcal{E}_i^-(\lambda)\}, & \lambda \leq \frac{c_i}{2} \\ \{0\}, & \text{otherwise,} \end{cases} \tag{20}$$

where

$$\mathcal{E}_i^+(\lambda) = \frac{1}{2\lambda} + \sqrt{\frac{1}{4\lambda^2} - \frac{1}{c_i^2}}, \tag{21}$$

$$\mathcal{E}_i^-(\lambda) = \frac{1}{2\lambda} - \sqrt{\frac{1}{4\lambda^2} - \frac{1}{c_i^2}}. \tag{22}$$

The optimal solution belongs to the set of candidate solutions $\boldsymbol{\mathcal{E}}(\lambda) = (\mathcal{E}_1(\lambda), \cdots, \mathcal{E}_N(\lambda))$ which satisfy the power constraint $\|\boldsymbol{\mathcal{E}}(\lambda)\|_1 \leq \mathcal{E}$ with equality. This set contains multiple candidates (according to (20)) corresponding to different values of λ. The optimal solution is the candidate which yields the highest rate. Thus, we can think of (20) as a recipe which reduces the set of candidate solutions to a finite set.

To obtain the set of candidate solutions, we consider $i \in \mathcal{N}$ and $\lambda \in \left(\frac{c_{i+1}}{2}, \frac{c_i}{2} \right]$,[3] and a corresponding power allocation $\boldsymbol{\mathcal{E}}(\lambda)$ satisfying (20). For all possibilities (3^i of them), we find $\lambda \in \left(\frac{c_{i+1}}{2}, \frac{c_i}{2} \right]$ for which $\|\boldsymbol{\mathcal{E}}(\lambda)\|_1 = \mathcal{E}$. If a solution λ^* exists, we include $\boldsymbol{\mathcal{E}}(\lambda^*)$ in our set of candidates. We repeat this $\forall i \in \mathcal{N}$. Finally, among the set of candidates, we choose the one which maximizes the objective function.

[3] we formally define $c_{N+1} = 0$

1) Example: Consider $N = 2$. For $i = 1$, we have $\lambda \in \left(\frac{c_2}{2}, \frac{c_1}{2}\right]$, and therefore, we consider the following possibilities $\mathcal{T}_1 = \{(0,0), (\mathcal{E}_1^+(\lambda), 0), (\mathcal{E}_1^-(\lambda), 0)\}$. Among those possibilities, we have to find the ones which have an ℓ_1-norm equal to $\mathcal{E}$, for some $\lambda \in \left(\frac{c_2}{2}, \frac{c_1}{2}\right]$. Clearly, $(0,0)$ is excluded. Then, we solve $\|(\mathcal{E}_1^+(\lambda), 0)\|_1 = \mathcal{E}$ for λ. If a solution λ^* exists in $\left(\frac{c_2}{2}, \frac{c_1}{2}\right]$, we include $(\mathcal{E}_1^+(\lambda^*), 0)$ in the set of candidate solutions $\mathcal{U}$. We apply a similar step for $(\mathcal{E}_1^-(\lambda), 0)$. Then, we consider $i = 2$. In this case, we have $\lambda \in \left(0, \frac{c_2}{2}\right]$, and therefore, we consider the possibilities $\mathcal{T}_1 \cup \mathcal{T}_2$ where

$$\mathcal{T}_2 = \{(0, \mathcal{E}_2^+(\lambda)), (0, \mathcal{E}_2^-(\lambda)), (\mathcal{E}_1^+(\lambda), \mathcal{E}_2^+(\lambda)), (\mathcal{E}_1^-(\lambda), \mathcal{E}_2^+(\lambda)), (\mathcal{E}_1^+(\lambda), \mathcal{E}_2^-(\lambda)), (\mathcal{E}_1^-(\lambda), \mathcal{E}_2^-(\lambda))\}.$$

For each $\boldsymbol{\mathcal{E}}(\lambda) \in \mathcal{T}_1 \cup \mathcal{T}_2$, we apply the same procedure. That is, we solve $\|\boldsymbol{\mathcal{E}}(\lambda)\|_1 = \mathcal{E}$ for λ. If a solution λ^* exists in $\left(0, \frac{c_2}{2}\right]$, we include $\boldsymbol{\mathcal{E}}(\lambda^*)$ in the set $\mathcal{U}$. Finally, the solution of the optimization problem is obtained as $\max_{\boldsymbol{\mathcal{E}} \in \mathcal{U}} f(\boldsymbol{\mathcal{E}})$, and the optimal allocation as $\arg\max_{\boldsymbol{\mathcal{E}} \in \mathcal{U}} f(\boldsymbol{\mathcal{E}})$.

Alternatively, we can plug $\boldsymbol{\mathcal{E}}(\lambda)$ in the Lagrangian L (17) (with $\lambda_i = 0$ if $\mathcal{E}_i \neq 0$), and minimize with respect to λ. Since the dual function $\min_{\mathcal{E}_i} L$ is convex in λ, we can find the λ which minimizes the duality gap by a gradient-decent algorithm over $\lambda \in \left(\frac{c_{i+1}}{2}, \frac{c_i}{2}\right]$. If the obtained λ leads to an allocation satisfying $\|\boldsymbol{\mathcal{E}}(\lambda)\|_1 = \mathcal{E}$, we add $\boldsymbol{\mathcal{E}}(\lambda)$ to the list of candidates. Finally, among the set of candidates, we choose the one which maximizes the objective function.

Let us now study the behavior of the optimal solution at high, low, and moderate SNR.

2) High SNR: At high SNR, i.e., if $c_i^2 \mathcal{E}^2 \gg 1 \ \forall i \in \mathcal{N}$, then (19) becomes

$$\left(\lambda - \frac{1}{\mathcal{E}_i}\right) \mathcal{E}_i = 0. \tag{23}$$

This leads to either $\mathcal{E}_i = 0$ or $\mathcal{E}_i = \frac{1}{\lambda}$. In other words, the optimal solutions leads to inactive channels where $\mathcal{E}_i = 0$, and active channels where $\mathcal{E}_i = \frac{1}{\lambda}$. Next, we show that as $\mathcal{E}$ grows, the optimal solution activates all channels.

If all channels are active, then $\mathcal{E}_i(\lambda) = \frac{1}{\lambda}$ for all $i \in \mathcal{N}$. The constraint $\|\boldsymbol{\mathcal{E}}(\lambda)\|_1 = \mathcal{E}$ is satisfied with equality if $\lambda = \frac{N}{\mathcal{E}}$. Thus, $\mathcal{E}_i(\lambda) = \frac{\mathcal{E}}{N}$ for all $i \in \mathcal{N}$. On the other hand, consider another allocation where channels $\mathcal{N}_a = \{1, \cdots, N_a\}$ are activated, and the remaining channels $\mathcal{N} \setminus \mathcal{N}_a$ are not. In this

case, $\mathcal{E}_i = \frac{\mathcal{E}}{N_a}$ for $i \in \mathcal{N}_a$. Call this allocation $\boldsymbol{\mathcal{E}}_a$. The achievable rate for the first solution satisfies

$$f\left(\frac{\mathcal{E}}{N}, \cdots, \frac{\mathcal{E}}{N}\right) = \frac{1}{2}\sum_{i\in\mathcal{N}} \log\left(1 + c_i^2 \frac{\mathcal{E}^2}{N^2}\right) \tag{24}$$

$$\geq \frac{1}{2}\sum_{i\in\mathcal{N}} \log\left(1 + c_i^2 \frac{\mathcal{E}^2}{N_a^2}\right) + \frac{N}{2}\log\left(\frac{N_a^2}{N^2}\right) \tag{25}$$

$$= \frac{1}{2}\sum_{i\in\mathcal{N}_a} \log\left(1 + c_i^2 \frac{\mathcal{E}^2}{N_a^2}\right) + \frac{1}{2}\sum_{i\in\mathcal{N}\setminus\mathcal{N}_a} \log\left(1 + c_i^2 \frac{\mathcal{E}^2}{N_a^2}\right) + \frac{N}{2}\log\left(\frac{N_a^2}{N^2}\right) \tag{26}$$

$$= f(\boldsymbol{\mathcal{E}}_a) + \frac{1}{2}\sum_{i\in\mathcal{N}\setminus\mathcal{N}_a} \log\left(1 + c_i^2 \frac{\mathcal{E}^2}{N_a^2}\right) + \frac{N}{2}\log\left(\frac{N_a^2}{N^2}\right) \tag{27}$$

This lower bound on $f\left(\frac{\mathcal{E}}{N}, \cdots, \frac{\mathcal{E}}{N}\right)$ eventually becomes larger than $f(\boldsymbol{\mathcal{E}}_a)$ as $\mathcal{E}$ increases. Thus, the optimal solutions at high SNR is to allocate $\mathcal{E}$ equally over all the N channels.

3) Low SNR: On the other hand, at low SNR where $c_i^2\mathcal{E}^2 \ll 1\ \forall i \in \mathcal{N}$, (19) becomes $(\lambda - c_i^2\mathcal{E}_i)\mathcal{E}_i = 0$ by neglecting $c_i^2\mathcal{E}_i^2$ in the denominator. The set of candidate solutions becomes the set of $\boldsymbol{\mathcal{E}}(\lambda)$ with $\mathcal{E}_i(\lambda) \in \{0, \frac{\lambda}{c_i^2}\}\ \forall i \in \mathcal{N}$, from which the best $\boldsymbol{\mathcal{E}}(\lambda)$ has to be chosen. In this case, we can show that the optimal solution has $\mathcal{E}_1 = \mathcal{E}$ and $\mathcal{E}_i = 0$ for $i = 2, \cdots, N$.

This can be shown as follows. The achievable rate $f(\boldsymbol{\mathcal{E}})$ satisfies

$$\frac{1}{2}\sum_{i\in\mathcal{N}} \log(1 + c_i^2\mathcal{E}_i^2) \leq \frac{N}{2}\log\left(1 + \frac{1}{N}\sum_{i\in\mathcal{N}} c_i^2\mathcal{E}_i^2\right) \tag{28}$$

$$\leq \frac{N}{2}\log\left(1 + \frac{c_1^2}{N}\sum_{i\in\mathcal{N}} \mathcal{E}_i^2\right) \tag{29}$$

$$\leq \frac{N}{2}\log\left(1 + \frac{c_1^2}{N}\left(\sum_{i\in\mathcal{N}} \mathcal{E}_i\right)^2\right) \tag{30}$$

$$\leq \frac{N}{2}\log\left(1 + \frac{c_1^2}{N}\mathcal{E}^2\right) \tag{31}$$

$$\leq \frac{c_1^2\mathcal{E}^2}{2}, \tag{32}$$

where the first step follows by Jensen's inequality, the second since $c_i \leq c_1$, the third since $\sum \mathcal{E}_i^2 \leq (\sum \mathcal{E}_i)^2$, the forth since $\sum_{i\in\mathcal{N}} \mathcal{E}_i \leq \mathcal{E}$, and the last by $\log(1+x) \leq x$. This upper bound is asymptotically achievable at low SNR by allocating the whole power to the strongest channel c_1^2, since $\frac{1}{2}\log(1 + c_1^2\mathcal{E}^2) \to \frac{c_1^2\mathcal{E}^2}{2}$ as $\mathcal{E}$ decreases.

This solution can also be demonstrated geometrically as follows. At low SNR, the objective function $f(\boldsymbol{\mathcal{E}})$ of problem (15) becomes $\sum_{i\in\mathcal{N}} \frac{c_i^2\mathcal{E}_i^2}{2}$ by using the approximation $\log(1+x) \approx x$ for small x. This function describes an elliptic paraboloid symmetric with respect to the $\mathcal{E}_i$-axes. The

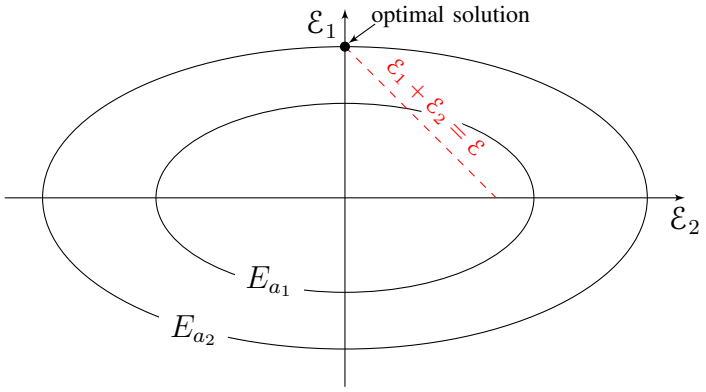


Fig. 2: For $N=2$, the level-sets of the low-SNR objective function are ellipses E_a. The constraints $\boldsymbol{\mathcal{E}}=(\mathcal{E}_1,\mathcal{E}_2)\in\mathbb{R}_+^2$ and $\|\boldsymbol{\mathcal{E}}\|_1=\mathcal{E}$ define a line segment L. The optimal solution lies at the intersection of L with the largest E_a, i.e., at $\boldsymbol{\mathcal{E}}=(\mathcal{E},0)$.

solutions of $\sum_{i\in\mathcal{N}}\frac{c_i^2\mathcal{E}_i^2}{2}=a$ describe an elliposoid E_a also symmetric with respect to the $\mathcal{E}_i$-axes, where a is the value of the objective function over E_a (see Fig. 2). Clearly, a larger a leads to a larger ellipsoid. On the other hand, our optimal power allocation must lie on the hyperplane described by $\boldsymbol{\mathcal{E}}\in\mathbb{R}_+^N$, $\|\boldsymbol{\mathcal{E}}\|_1=\mathcal{E}$. Thus, the optimal solution lies at the intersection of this hyperplane with the largest ellipsoid. It is easy to see that the intersection with the largest E_a lies at the $\mathcal{E}_i$-axis corresponding to the largest c_i, i.e., c_1. Thus, the optimal solution at low SNR is $\boldsymbol{\mathcal{E}}=(\mathcal{E},0,\cdots,0)$, where all the power is allocated to the strongest channel.

4) Moderate SNR: Unfortunately, while the solutions at high and low SNR are simple, the solution at moderate SNR is not as simple. In general, at moderate SNR, multiple channels have to be activated according to the general solution (20). The complexity of finding this solution grows exponentially with the number of channels. To simplify the solution, we propose an algorithm based on the following analysis of the $N=2$ case.

For $N=2$, there are eight candidate solutions among which one is optimal. Namely, the candidates are the feasible power allocations among $(\mathcal{E}_1^+(\lambda),0)$, $(\mathcal{E}_1^-(\lambda),0)$, $(0,\mathcal{E}_2^+(\lambda))$, $(0,\mathcal{E}_2^-(\lambda))$, $(\mathcal{E}_1^+(\lambda),\mathcal{E}_2^+(\lambda))$, $(\mathcal{E}_1^+(\lambda),\mathcal{E}_2^-(\lambda))$, $(\mathcal{E}_1^-(\lambda),\mathcal{E}_2^+(\lambda))$, and $(\mathcal{E}_1^-(\lambda),\mathcal{E}_2^-(\lambda))$. Recall that a power allocation is feasible if there exist $\lambda\geq 0$ for which this power allocation satisfies the power constraint with equality.

Let us plot the trajectories of those solutions as a function of λ. Fig. 3 shows the trajectories for a channel with $(c_1,c_2)=(1,0.8)$. On this figure, an allocation is optimal if $\mathcal{E}_1+\mathcal{E}_2=\mathcal{E}$. Therefore, the optimal solution must lie at the intersection of one of the trajectories with the line defined by

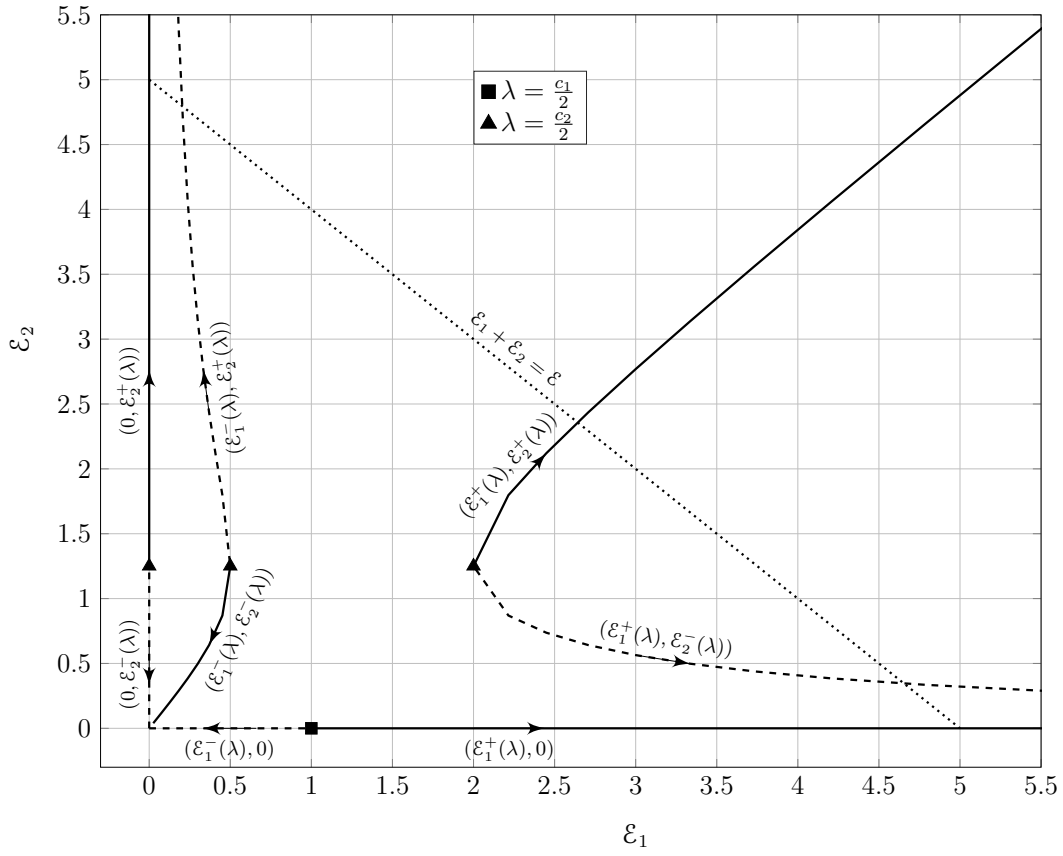


Fig. 3: Trajectories of candidate solutions of (15) as a function of λ. The arrows show the direction of decrease of λ. Here, we have chosen $\mathcal{E} = 5$.

$\mathcal{E}_1 + \mathcal{E}_2 = \mathcal{E}$ (the dotted line in Fig. 3). Note from this figure that there might be multiple such intersections. Next, we proceed to reduce the number of possibilities.

Let us start analyzing Fig. 3 from the black square which corresponds to $\lambda = \frac{c_1}{2}$ where $\mathcal{E}_1^+(\lambda) = \mathcal{E}_1^-(\lambda) = \frac{1}{2\lambda} = \frac{1}{c_1}$ and $\mathcal{E}_2 = 0$. As λ decreases, $\mathcal{E}_1^+(\lambda)$ increases while $\mathcal{E}_1^-(\lambda)$ decreases. If $\frac{1}{c_1} \leq \mathcal{E}$, then the solution $(\mathcal{E}_1^+(\lambda), 0)$ is feasible, otherwise, then $(\mathcal{E}_1^-(\lambda), 0)$ is feasible. Note however that the latter case corresponds to $c_1\mathcal{E} \leq 1$ which is close to the low SNR regime. Since we are focusing on the moderate SNR regime here, we omit the candidate $(\mathcal{E}_1^-(\lambda), 0)$. Thus, at this point, our list of candidates contains $(\mathcal{E}_1^+(\lambda), 0)$ only.

At $\lambda = \frac{c_2}{2}$, $\mathcal{E}_2^+(\lambda)$ and $\mathcal{E}_2^-(\lambda)$ become active and we reach the points marked by triangles in Fig. 3. Now, we have the additional possibilities $(0, \mathcal{E}_2^+(\lambda))$, $(0, \mathcal{E}_2^-(\lambda))$, $(\mathcal{E}_1^+(\lambda), \mathcal{E}_2^+(\lambda))$, $(\mathcal{E}_1^+(\lambda), \mathcal{E}_2^-(\lambda))$, $(\mathcal{E}_1^-(\lambda), \mathcal{E}_2^+(\lambda))$, and $(\mathcal{E}_1^-(\lambda), \mathcal{E}_2^-(\lambda))$. Since $(\mathcal{E}_1^+(\lambda), \mathcal{E}_2^-(\lambda))$ and $(\mathcal{E}_1^-(\lambda), \mathcal{E}_2^+(\lambda))$ converge respectively to $(\mathcal{E}_1^+(\lambda), 0)$ and $(0, \mathcal{E}_2^+(\lambda))$ as λ decreases, we omit these two possibilities. On the other

hand, since the objective function evaluated at $(0, \mathcal{E}_2^+(\lambda))$ and $(0, \mathcal{E}_2^-(\lambda))$ for a given λ is smaller than that at $(\mathcal{E}_1^+(\lambda), 0)$ and $(\mathcal{E}_1^-(\lambda), 0)$ respectively, we omit these two possibilities too.

The remaining candidates at this point are $(\mathcal{E}_1^+(\lambda), 0)$, $(\mathcal{E}_1^+(\lambda), \mathcal{E}_2^+(\lambda))$, and $(\mathcal{E}_1^-(\lambda), \mathcal{E}_2^-(\lambda))$. Finally, since the last candidate $(\mathcal{E}_1^-(\lambda), \mathcal{E}_2^-(\lambda))$ satisfies $\mathcal{E}_1^-(\lambda) + \mathcal{E}_2^-(\lambda) \leq \frac{2}{c_2}$, and since it approaches $(0, 0)$ as λ decreases, then this solution is feasible if $c_2\mathcal{E} \leq 2$ which implies that we are close to low SNR. Thus, we omit this candidate, leaving only two candidates in the list of candidate solutions: $(\mathcal{E}_1^+(\lambda), 0)$ and $(\mathcal{E}_1^+(\lambda), \mathcal{E}_2^+(\lambda))$. We choose the candidate which yields larger $f(\boldsymbol{\mathcal{E}})$. Keep in mind that we have to select λ so that the resulting allocation $\boldsymbol{\mathcal{E}}$ satisfies $\|\boldsymbol{\mathcal{E}}\|_1 = \mathcal{E}$.

Based on this, we define

$$\boldsymbol{A}_i(\lambda) = (A_{i1}(\lambda), A_{i2}(\lambda), \cdots, A_{iN}(\lambda)), \tag{33}$$

where $A_{ij}(\lambda) = \mathcal{E}_i^+(\lambda)$ given in (21) if $j \leq i$, and $A_{ij}(\lambda) = 0$ otherwise. Note that $\boldsymbol{A}_i(\lambda)$ is real if $\lambda \in \left(0, \frac{c_i}{2}\right]$. We propose Algorithm 1 for finding a good solution for (15). This algorithm jointly activates channels and allocates powers to them, and therefore, we call it the 'joint activation/power allocation' (JAPA) algorithm. For this algorithm, we use $\frac{1}{0} = \infty$ and allow λ to be 0.

This algorithm activates components of $\boldsymbol{\mathcal{E}}$ starting with $\mathcal{E}_1$ followed by $\mathcal{E}_2$ and so on. Each time, an optimal λ is found, which leads to satisfying the power constraint with equality. As long as such λ exists, and as long as it leads to a higher achievable rate, the algorithm activates one more component of $\boldsymbol{\mathcal{E}}$. Note that this algorithm searches over at most N candidates ($\boldsymbol{A}_i(\lambda)$, $i \in \mathcal{N}$), contrary to the exhaustive search algorithm which searches over 3^N candidates. Thus, this algorithm reduces the complexity from $O(3^N)$ to $O(N)$. Step 10 in this algorithm can be solved using bisection or the Newton-Raphson method, and might be executed up to $N - 1$ times.

To reduce the number of executions of step 10, we can split the algorithm into two stages. First we find the channels that should be activated. Then, we find λ^* leading to a solution that satisfies the constraint with equality. To this end, we define $\boldsymbol{B}(\lambda) = (B_1(\lambda), \cdots, B_N(\lambda))$ where

$$B_i(\lambda) = \begin{cases} \mathcal{E}_i^+(\lambda), & \lambda \leq \frac{c_i}{2} \\ 0, & \text{otherwise.} \end{cases} \tag{34}$$

Note that $\|\boldsymbol{B}(\lambda)\|_1$ is decreasing in λ. Thus, Algorithm 2 can be used to find a good solution for (15). We call this algorithm the 'successive activation/power allocation' (SAPA) algorithm. Similar to the JAPA algorithm, we allow $\lambda = 0$, and use $\frac{1}{0} = \infty$.

Steps 8 to 12 find the number of channels to be activated. A new channel is activated as long as $\boldsymbol{B}(\lambda)$ evaluated at the boundaries of $\left[\frac{c_i}{2}, \frac{c_{i-1}}{2}\right]$ is smaller than $\mathcal{E}$. In this case, $\boldsymbol{B}(\lambda)$ has to be

Algorithm 1 JAPA for parallel OWC

1: **function** JAPA($c_1, \cdots, c_{N+1}$, $\mathcal{E}$)
2: $i \leftarrow 1$
3: $\boldsymbol{\mathcal{E}}_b \leftarrow (\mathcal{E}, 0, \cdots, 0)$
4: $R_a \leftarrow 0$
5: $R_b \leftarrow f(\boldsymbol{\mathcal{E}}_b)$
6: **while** $R_b > R_a$ **do**
7: $R_a \leftarrow R_b$
8: $\boldsymbol{\mathcal{E}}_a \leftarrow \boldsymbol{\mathcal{E}}_b$
9: $i \leftarrow i + 1$
10: **if** $\exists \lambda^* \in \left[0, \frac{c_i}{2}\right]$ so that $\|\boldsymbol{A}_i(\lambda^*)\|_1 = \mathcal{E}$ **then**
11: $\boldsymbol{\mathcal{E}}_b \leftarrow \boldsymbol{A}_i(\lambda^*)$
12: $R_b \leftarrow f(\boldsymbol{\mathcal{E}}_b)$
13: **else**
14: $R_b \leftarrow 0$
15: **end if**
16: **end while**
17: **return** $\mathcal{E}_a$
18: **end function**

increased, which is done by reducing λ below $\frac{c_i}{2}$ which in turn activates channel i. After determining the number of active channels, step 13 finds λ^* for which the constraint is satisfied with equality. Similar to the JAPA algorithm, this algorithm also searches over at most N candidates ($\boldsymbol{B}_i(\lambda)$, $i \in \mathcal{N}$), and thus also has complexity $O(N)$. The main difference is that, contrary to the JAPA algorithm, step 13 (step 10 in Algorithm 1) is executed only once.

Note that due to the definition of $\boldsymbol{B}(\lambda)$, we can not reduce λ below $\frac{c_i}{2}$ without activating the i-th channel. This is the main difference with respect to the JAPA algorithm which allows λ to go below $\frac{c_i}{2}$ while channel i is inactive, thus allowing more flexibility at the expense of higher complexity. This additional flexibility makes the JAPA algorithm slightly superior to the SAPA algorithm.

Remark 1: At high SNR, step 8 in Algorithm 2 will always be true. Thus, the algorithm will activate all channels. After activating all channels, step 13 finds $\lambda^* \in \left[0, \frac{c_N}{2}\right]$ which yields $\|\boldsymbol{B}(\lambda^*)\|_1 = \mathcal{E}$. It is easy to see that all $B_i(\lambda)$ will approach $\frac{\mathcal{E}}{N}$ as $\mathcal{E}$ grows. Thus, the SAPA algorithm is optimal at

Algorithm 2 SAPA for Parallel OWC

function SAPA($c_1, \cdots, c_{N+1}$, $\mathcal{E}$)
  $\boldsymbol{\mathcal{E}}_h \leftarrow \boldsymbol{B}(\infty)$
  $\boldsymbol{\mathcal{E}}_l \leftarrow \boldsymbol{B}(\frac{c_1}{2})$
  $i \leftarrow 1$
  if $\|\boldsymbol{\mathcal{E}}_l\|_1 > \mathcal{E}$ **then**
    $\boldsymbol{\mathcal{E}} \leftarrow (\mathcal{E}, 0, \cdots, 0)$
  else
    while $(\|\boldsymbol{\mathcal{E}}_h\|_1 - \mathcal{E})(\|\boldsymbol{\mathcal{E}}_l\|_1 - \mathcal{E}) \geq 0 \wedge i < N+1$ **do**
      $i \leftarrow i + 1$
      $\boldsymbol{\mathcal{E}}_h \leftarrow \boldsymbol{\mathcal{E}}_l$
      $\boldsymbol{\mathcal{E}}_l \leftarrow \boldsymbol{B}(\frac{c_i}{2})$
    end while
    Find $\lambda^* \in \left[0, \frac{c_{i-1}}{2}\right]$ so that $\|\boldsymbol{B}(\lambda^*)\|_1 = \mathcal{E}$
    $\boldsymbol{\mathcal{E}} \leftarrow \boldsymbol{B}(\lambda^*)$
  end if
  return $\mathcal{E}$
end function

high SNR. This is also true for the JAPA algorithm, since it is more accurate.

Fig. 4 shows an evaluation of the algorithms for a system with $N = 4$ channels. The performance is averaged over 10^2 realizations of the channel, which follows a log-normal distribution[4] on h_i [8]

$$f(h_i) = \frac{1}{h_i\sqrt{2\pi\sigma_R^2}} \exp\left(-\frac{(\log(h_i) + \frac{\sigma_R^2}{2})^2}{2\sigma_R^2}\right), \tag{35}$$

with Rytov variance σ_R^2 assumed to be 1. Five average achievable rates are plotted. The solid line corresponds to exhaustive search, where the optimal solution is found from the set of all possible candidate solutions (as in Sec. IV-B1). This set contains 3^N elements $\boldsymbol{\mathcal{E}}$ whose i-th component can be $\mathcal{E}_i^+$, $\mathcal{E}_i^-$, or 0 (20), and for each element, the λ that leads to $\|\boldsymbol{\mathcal{E}}\|_1 = \mathcal{E}$ is found. The achievable rates corresponding to the power allocation obtained by the JAPA and the SAPA algorithms are shown using square and $\times$ markers, respectively. Those rates almost coincide with the exhaustive search

[4]Recall that the log-normal distribution is a good model for OWC channels under weak turbulence conditions [12].

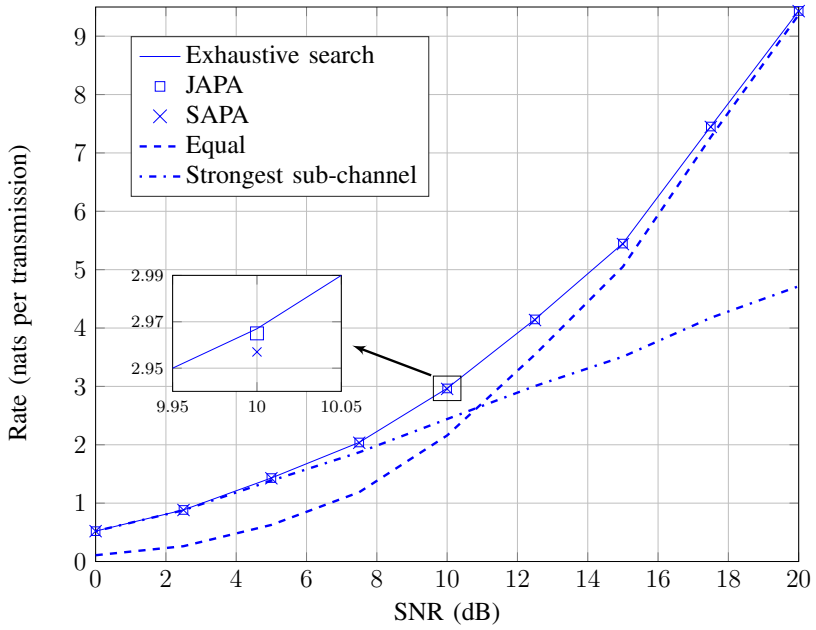


Fig. 4: The achievable rate as a function of SNR ($\mathcal{E}/\sigma$) for a system with 4 channels. The power allocation over the channels is performed using exhaustive search, the JAPA algorithm, and the SAPA algorithm. The achievable rate with equal power allocation and allocating all power to the strongest channel are plotted for comparison.

solution, which indicates that the proposed algorithms almost always find the optimal solution. Based on this plot, the SAPA algorithm is recommended for finding a near optimal power allocation since it has lower complexity. The achievable rates by using equal power allocation over the channels and by allocating all power to the strongest channel (largest c_i) are also shown. These plots confirm the high and low SNR solutions discussed in Sec. IV-B2 and Sec. IV-B3, respectively.

Since the SAPA algorithm delivers a solution which is very close to optimal, we restrict ourselves to this algorithm from this point on. Using this algorithm, we can state the following.

Proposition 1: The capacity $C(\boldsymbol{h}, \mathcal{E})$ of a system of N parallel channels is lower bounded by the achievable rate

$$\tilde{R}(\boldsymbol{h}, \mathcal{E}) = \sum_{i \in \mathcal{N}} r(h_i, \mathcal{E}_i), \tag{36}$$

where $r(h, \mathcal{E})$ is defined in (2), and where $(\mathcal{E}_1, \cdots, \mathcal{E}_N)$ is the output of the SAPA algorithm with $c_i = h_i \sqrt{\frac{e}{2\pi}}$.

This proposition supplements Theorem 1, since it simplifies the lower bound $R(\boldsymbol{h}, \mathcal{E})$ (13). In fact, when averaged over multiple channel realizations, the two lower bounds $R(\boldsymbol{h}, \mathcal{E})$ and $\tilde{R}(\boldsymbol{h}, \mathcal{E})$ are

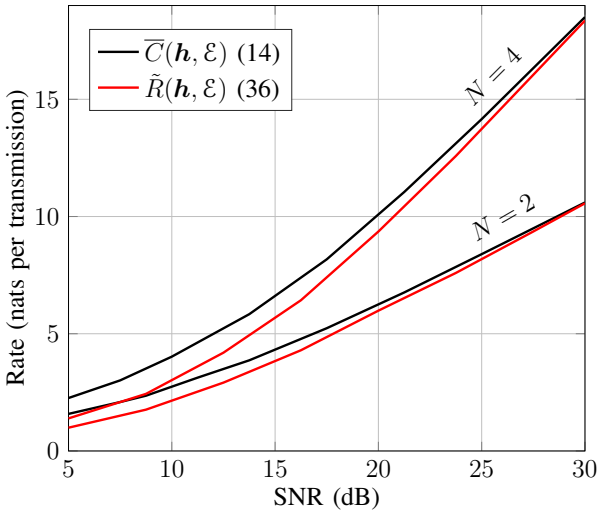


Fig. 5: Capacity upper and lower bounds as a function of SNR ($\mathcal{E}/\sigma$) for a system with CSIT, with $N = 2$ and $N = 4$ and an average constraint $\mathcal{E}$.

almost equal as seen in Fig. 4.

C. Evaluation

We consider a system with $N = 2$ and another with $N = 4$ channels, where h_i follows a log-normal distribution according to (35). The achievable rate $\tilde{R}(\boldsymbol{h}, \mathcal{E})$ of this system, averaged over 10^3 channel realizations, is shown in Fig. 5. The upper bound given in Theorem 1 is also plotted. It can be seen that this upper bound is tight at high SNR. This indicates that the SAPA algorithm applied with an exponential input distribution ($c_i = h_i\sqrt{\frac{e}{2\pi}}$) is nearly optimal at high SNR.

In fact, this can be easily proved by noting that the upper bound $\overline{r}(h, \mathcal{E})$ converges to $\frac{1}{2}\log\left(\frac{eh^2\mathcal{E}^2}{2\pi}\right)$ for large $\mathcal{E}$. Thus, the upper bound in Theorem 1 satisfies

$$\overline{C}(\boldsymbol{h},\mathcal{E}) = \max_{\mathcal{E}\in\mathcal{S}} \sum_{i\in\mathcal{N}} \frac{1}{2}\log(c_i^2\mathcal{E}_i^2) + \epsilon_{\mathcal{E}} \tag{37}$$

$$= \sum_{i\in\mathcal{N}} \frac{1}{2}\log\left(\frac{c_i^2\mathcal{E}^2}{N^2}\right) + \epsilon_{\mathcal{E}}, \tag{38}$$

where $\epsilon_{\mathcal{E}} \to 0$ as $\mathcal{E} \to \infty$, and where the last step follows since the maximization in (37) is similar to (15), and hence has a similar solution, which is $\mathcal{E}_i = \frac{\mathcal{E}}{N}$ at high SNR. Thus, by using $c_i = h_i\sqrt{\frac{e}{2\pi}}$, we can state the following corollary.

Corollary 1: The capacity of a system of N parallel channels with average constraint $\mathcal{E}$ satisfies

$$\lim_{\mathcal{E}\to\infty}\left[C(\boldsymbol{h},\mathcal{E})-\sum_{i\in\mathcal{N}}\frac{1}{2}\log\left(\frac{eh_i^2\mathcal{E}^2}{2\pi N^2}\right)\right]=0. \tag{39}$$

V. Parallel Channels without CSIT

If no CSIT is available, the transmitter does not know $\boldsymbol{h}$, and consequently, does not know $C(\boldsymbol{h},\mathcal{E})$. Consequently, for any $R>0$, there is a non-zero probability that $R>C(\boldsymbol{h},\mathcal{E})$, and hence reliable communication can not be guaranteed. In this case, the notion of outage becomes important [28]. For a given target rate R_0, and a given transmit strategy defined by an input distribution $p(x_1,\cdots,x_N)$, the outage probability is defined as $\mathbb{P}\left\{I(X_1,\cdots,X_N;Y_1,\cdots,Y_N)<R_0\right\}$. The optimal outage probability is thus

$$P_o(\mathcal{E})=\min_{p(x_1,\cdots,x_N)\in\mathcal{P}}\mathbb{P}\left\{I(X_1,\cdots,X_N;Y_1,\cdots,Y_N)<R_0\right\}, \tag{40}$$

where $\mathcal{P}$ is the set of distributions of $(X_1,\cdots,X_N)\in\mathbb{R}_+^N$ satisfying $\sum_{i\in\mathcal{N}}\mathbb{E}_{X_i}[X_i]\le\mathcal{E}$. Note that here, the CSIT can be exploited in the minimization, i.e., the optimal distribution depends on $\boldsymbol{h}$.

An upper bound on the outage probability can be obtained by fixing $p(x_1,\cdots,x_N)$. One convenient choice is to fix $(X_1,\cdots,X_N)$ to be independent and exponentially distributed. Using this strategy and distributing the optical intensity $\mathcal{E}$ equally over the N channels leads to

$$P_o(\mathcal{E})\le\mathbb{P}\left\{R'(\boldsymbol{h},\mathcal{E})<R_0\right\}, \tag{41}$$

where

$$R'(\boldsymbol{h},\mathcal{E})=\sum_{i\in\mathcal{N}}r\left(h_i,\frac{\mathcal{E}}{N}\right), \tag{42}$$

and $r(h,\mathcal{E})$ is defined in (2).

On the other hand, the outage probability is lower bounded by

$$P_o(\mathcal{E})\ge\mathbb{P}\left\{\overline{C}(\boldsymbol{h},\mathcal{E})<R_0\right\}, \tag{43}$$

where $\overline{C}(\boldsymbol{h},\mathcal{E})$ is given in Theorem 1. This follows since the capacity of the channel with no CSIT is upper bounded by that with CSIT, for a given channel state. Therefore, $I(X_1,\cdots,X_N;Y_1,\cdots,Y_N)\le\overline{C}(\boldsymbol{h},\mathcal{E})$, leading to $\mathbb{P}\left\{I(X_1,\cdots,X_N;Y_1,\cdots,Y_N)<R_0\right\}\ge\mathbb{P}\left\{\overline{C}(\boldsymbol{h},\mathcal{E})<R_0\right\}$ for any distribution $p(x_1,\cdots,x_N)\in\mathcal{P}$. This consequently leads to the lower bound (43).

The outage probability upper and lower bounds above do not coincide in general. However, they coincide at high SNR. To show this, we note that similar to $R'(\boldsymbol{h},\mathcal{E})$, $\overline{C}(\boldsymbol{h},\mathcal{E})$ also converges to

$\sum_{i\in\mathcal{N}} \frac{1}{2}\log\left(\frac{eh_i^2\mathcal{E}^2}{2\pi N^2}\right)$ at high SNR, as shown in (37)-(38). Consequently, (41) and (43) are tight in this case. This leads to the following statement.

Corollary 2: The outage probability of a system of N parallel channels with a total average constraint $\mathcal{E}$ and with no CSIT satisfies

$$\lim_{\mathcal{E}\to\infty}\left[P_o(\mathcal{E}) - \mathbb{P}\left\{\sum_{i\in\mathcal{N}}\frac{1}{2}\log\left(\frac{eh_i^2\mathcal{E}^2}{2\pi N^2}\right) < R_0\right\}\right] = 0. \tag{44}$$

Thus, reliable communication at a rate R_0 with outage probability $P_o(\mathcal{E})$ is possible at high SNR if (44) is satisfied. Furthermore, there is no coding scheme with rate R_0 which can achieve a lower outage probability than $\mathbb{P}\left\{\sum_{i\in\mathcal{N}}\frac{1}{2}\log\left(\frac{eh_i^2\mathcal{E}^2}{2\pi N^2}\right) < R_0\right\}$ at high SNR. This leads to the high-SNR ϵ-outage capacity C_ϵ, defined as the highest rate that can be achieved at an outage probability less than ϵ. Namely, at high SNR, C_ϵ is the largest $R\in\mathbb{R}_+$ such that

$$\mathbb{P}\left\{\sum_{i\in\mathcal{N}}\frac{1}{2}\log\left(\frac{eh_i^2\mathcal{E}^2}{2\pi N^2}\right) < R\right\} \le \epsilon.$$

VI. Both Average and Peak Constraints

Here, we discuss the case with a peak intensity constraint $\mathcal{A}$ in addition to the average constraint $\mathcal{E}$. We denote the capacity in this case by $C_{\mathcal{A}}(\boldsymbol{h},\mathcal{E})$ for a given channel state $\boldsymbol{h}$.

A. Bounds for $N = 1$

Upper bounds on the capacity $C_{\mathcal{A}}(h,\mathcal{E})$ in this case and achievable rates were given in [22]–[24]. For instance, the following achievable rate was given by Lapidoth *et al.* in [22]

$$r_l(h,\mathcal{E},\mathcal{A}) \begin{cases} \frac{1}{2}\log\left(1+\frac{h^2\mathcal{E}^2e^{2\alpha\mu^*}}{2\pi e\alpha^2}\left(\frac{1-e^{-\mu^*}}{\mu^*}\right)^2\right), & \frac{\mathcal{E}}{\mathcal{A}} \le \frac{1}{2} \\ \frac{1}{2}\log\left(1+\frac{h^2\mathcal{A}^2}{2\pi e}\right), & \frac{\mathcal{E}}{\mathcal{A}} > \frac{1}{2} \end{cases} \tag{45}$$

where μ^* is the solution of $\frac{1}{\mu^*} - \frac{e^{-\mu^*}}{1-e^{-\mu^*}} = \alpha$. Thus, $C_{\mathcal{A}}(h,\mathcal{E}) \ge r_l(h,\mathcal{E},\mathcal{A})$.

Another lower bound was given in [24] achievable if X follows a truncated-Gaussian distribution

$$\tilde{g}_{\mu,\nu}(x) = \eta g_{\mu,\nu}(x)$$

over $[0,\mathcal{A}]$ where $\eta = (G_{\mu,\nu}(\mathcal{A}) - G_{\mu,\nu}(0))^{-1}$ and $g_{\mu,\nu}(x)$ and $G_{\mu,\nu}(x)$ denote the Gaussian distribution with mean μ and variance ν^2 and its cumulative-distribution function, respectively. The mean and variance of $\tilde{g}_{\mu,\nu}(x)$ are given by

$$\tilde{\mu} = \nu^2[\tilde{g}_{\mu,\nu}(0) - \tilde{g}_{\mu,\nu}(\mathcal{A})] + \mu \tag{46}$$

$$\tilde{\nu}^2 = \nu^2\left[1 - \mathcal{A}\tilde{g}_{\mu,\nu}(\mathcal{A}) - \tilde{\mu}\left(\tilde{g}_{\mu,\nu}(0) - \tilde{g}_{\mu,\nu}(\mathcal{A})\right)\right]. \tag{47}$$

The achievable rate for a given μ and ν, so that $\tilde{\mu} \leq \mathcal{E}$, is given by

$$\tilde{r}_t(h, \mathcal{A}, \mu, \nu) = \frac{1}{2} \log \left(\frac{\nu^2}{\tilde{\nu}^2} + h^2 \nu^2 \right) - \phi, \tag{48}$$

where $\phi = \log(\eta) + \frac{1}{2} \left((\mathcal{A} - \mu) \tilde{g}_{\mu,\nu}(\mathcal{A}) + \mu \tilde{g}_{\mu,\nu}(0) \right)$. Thus, $C_{\mathcal{A}}(h, \mathcal{E})$ is lower bounded by

$$r_t(h, \mathcal{E}, \mathcal{A}) = \max_{\substack{\mu,\nu \\ \tilde{\mu} \leq \mathcal{E}}} \tilde{r}_t(h, \mathcal{A}, \mu, \nu). \tag{49}$$

As shown in [24], $r_t(h, \mathcal{E}, \mathcal{A})$ is close to capacity at high SNR. Furthermore, the rate $r_t(h, \mathcal{E}, \mathcal{A})$ satisfies $r_t(h, \mathcal{E}, \mathcal{A}) \geq r_{ts}(h, \mathcal{E}, \mathcal{A})$ [30] where the simplified truncated-Gaussian lower bound $r_{ts}(h, \mathcal{E}, \mathcal{A})$ is given by

$$r_{ts}(h, \mathcal{E}, \mathcal{A}) = \frac{1}{2} \log \left(1 + h^2 \min \left\{ \frac{\mathcal{E}^2}{9}, \frac{\mathcal{A}^2}{36} \right\} \right), \tag{50}$$

by choosing $\tilde{\mu} = \min\{\mathcal{E}, \mathcal{A}/2\}$ and $\nu = \frac{\mu}{3}$ and plugging in (48). Although this simplification sacrifices the near optimality of the achievable rate at high SNR, the simplified rate captures the behavior of the channel capacity at high SNR.

Note that the capacity upper bound $\overline{r}(h, \mathcal{E})$ given in (3) also holds in this case. Another upper bound was given in [24] as

$$\overline{\overline{r}}(h, \mathcal{A}) = \sup_{\delta \in [0,1]} \left[\frac{\delta}{2} \log \left(\frac{h^2 \mathcal{A}^2}{2\pi e} \right) - \log \left((1-\delta)^{\frac{3(1-\delta)}{2}} \delta^{\delta} \right) \right]. \tag{51}$$

Therefore, we can write

$$C_{\mathcal{A}}(h, \mathcal{E}) \leq \min\{\overline{r}(h, \mathcal{E}), \overline{\overline{r}}(h, \mathcal{A})\}. \tag{52}$$

B. Parallel Channels with CSIT

1) Capacity Bounds: Capacity lower bounds for this case can be derived by optimizing the achievable rates (45), (49), and (50) over the parallel channels with respect to power allocation. This is formally stated as follows.

Theorem 2: The capacity $C_{\mathcal{A}}(\boldsymbol{h}, \mathcal{E})$ of a system of N parallel channels with an average constraint $\mathcal{E}$ and a peak constraint $\mathcal{A}$ is lower bounded by the achievable rates

$$R_a(\boldsymbol{h}, \mathcal{E}, \mathcal{A}) = \max_{\boldsymbol{\mathcal{E}} \in \mathcal{S}_{\mathcal{A}}} \sum_{i \in \mathcal{N}} r_a(h_i, \mathcal{E}_i, \mathcal{A}), \quad a \in \{l, t, ts\}, \tag{53}$$

where $r_a(h_i, \mathcal{E}_i, \mathcal{A})$, $a \in \{l, t, ts\}$, is defined in (45), (49) and (50), respectively, and where $\mathcal{S}_{\mathcal{A}} = \{\boldsymbol{\mathcal{E}} \in (0, \mathcal{A}/2]^N | \|\boldsymbol{\mathcal{E}}\|_1 \leq \mathcal{E}\}$.

Proof: Similar to (6)-(12), we can show that $C_{\mathcal{A}}(\boldsymbol{h},\mathcal{E}) = \max_{\boldsymbol{\mathcal{E}}\in\mathcal{S}_{\mathcal{A}}} \sum_{i\in\mathcal{N}} C_{\mathcal{A}}(h_i,\mathcal{E}_i)$. Then, by using (50), (49) and (45) we can obtain the given lower bounds. Note that here, we restrict $\mathcal{E}_i$ to $(0,\mathcal{A}/2]$ since increasing $\mathcal{E}_i$ beyond $\mathcal{A}/2$ does not increase $C_{\mathcal{A}}(h_i,\mathcal{E}_i)$ as shown in [22]. ■

Using $\overline{r}(h_i,\mathcal{E}_i)$ and $\overline{\overline{r}}(h_i,\mathcal{A})$, we can obtain the following capacity upper bound.

Theorem 3: The capacity $C_{\mathcal{A}}(\boldsymbol{h},\mathcal{E})$ of a system of N parallel channels with an average constraint $\mathcal{E}$ and a peak constraint $\mathcal{A}$ is upper bounded by

$$\overline{C}_{\mathcal{A}}(\boldsymbol{h},\mathcal{E}) = \max_{\boldsymbol{\mathcal{E}}\in\mathcal{S}_{\mathcal{A}}} \sum_{i\in\mathcal{N}} \min\{\overline{r}(h_i,\mathcal{E}_i), \overline{\overline{r}}(h_i,\mathcal{A})\}, \tag{54}$$

where $\overline{r}(h_i,\mathcal{E}_i)$ and $\overline{\overline{r}}(h_i,\mathcal{A})$ are defined in (3) and (51), respectively, and where $\mathcal{S}_{\mathcal{A}}$ is as defined in Theorem 2.

Proof: Follows from $C_{\mathcal{A}}(\boldsymbol{h},\mathcal{E}) = \max_{\boldsymbol{\mathcal{E}}\in\mathcal{S}_{\mathcal{A}}} \sum_{i\in\mathcal{N}} C_{\mathcal{A}}(h_i,\mathcal{E}_i)$ and (52). ■

For $\mathcal{E} > \frac{N\mathcal{A}}{2}$, the following theorem holds.

Theorem 4: The capacity $C_{\mathcal{A}}(\boldsymbol{h},\mathcal{E})$ of a system of N parallel channels with an average constraint $\mathcal{E}$ and a peak constraint $\mathcal{A}$ satisfying $\mathcal{E} > \frac{N\mathcal{A}}{2}$ satisfies

$$C_{\mathcal{A}}(\boldsymbol{h},\mathcal{E}) = \sum_{i\in\mathcal{N}} C_{\mathcal{A}}(h_i,\mathcal{A}/2), \tag{55}$$

and

$$\lim_{\mathcal{A}\to\infty} \left[C_{\mathcal{A}}(\boldsymbol{h},\mathcal{E}) - \sum_{i\in\mathcal{N}} \frac{1}{2}\log\left(\frac{h_i^2\mathcal{A}^2}{2\pi e}\right)\right] = 0. \tag{56}$$

Proof: The first part follows from $C_{\mathcal{A}}(\boldsymbol{h},\mathcal{E}) = \max_{\boldsymbol{\mathcal{E}}\in\mathcal{S}_{\mathcal{A}}} \sum_{i\in\mathcal{N}} C_{\mathcal{A}}(h_i,\mathcal{E}_i)$ and since $C_{\mathcal{A}}(h_i,\mathcal{E}_i)$ is maximized if $\mathcal{E}_i = \mathcal{A}/2$ as shown in [22]. Note that choosing $\mathcal{E}_i = \frac{\mathcal{A}}{2}$ for all $i \in \mathcal{N}$ is a valid power allocation since $\mathcal{E} > \frac{N\mathcal{A}}{2}$. The second part is proved as follows. From Theorem 2, we have $R_l(\boldsymbol{h},\mathcal{E},\mathcal{A}) \geq \sum_{i\in\mathcal{N}} r_l(h_i,\mathcal{A}/2,\mathcal{A})$. Substituting in (45) leads to the achievable rate $\sum_{i\in\mathcal{N}} \frac{1}{2}\log\left(1+\frac{h_i^2\mathcal{A}^2}{2\pi e}\right)$ which converges to $\sum_{i\in\mathcal{N}} \frac{1}{2}\log\left(\frac{h_i^2\mathcal{A}^2}{2\pi e}\right)$ as $\mathcal{A}$ grows. On the other hand, it holds that $\overline{C}_{\mathcal{A}}(\boldsymbol{h},\mathcal{E}) \leq \sum_{i\in\mathcal{N}} \overline{\overline{r}}(h_i,\mathcal{A})$. But $\overline{\overline{r}}(h_i,\mathcal{A})$ also converges to $\frac{1}{2}\log\left(\frac{h_i^2\mathcal{A}^2}{2\pi e}\right)$ as $\mathcal{A}$ grows, which proves the second part of the theorem. ■

Next, we discuss power allocation for the case

$$\mathcal{E} \leq \frac{N\mathcal{A}}{2}. \tag{57}$$

2) Power Allocation: The achievable rates $r_l(h,\mathcal{E},\mathcal{A})$ (45) and $r_t(h,\mathcal{E},\mathcal{A})$ 49 can not be written in the form $\frac{1}{2}\log(1+c^2\mathcal{E}^2)$. Hence, the power allocation algorithms discussed previously do not apply immediately to $R_l(\boldsymbol{h},\mathcal{E},\mathcal{A})$ and $R_t(\boldsymbol{h},\mathcal{E},\mathcal{A})$ given in Theorem 2. However, those algorithms can be applied for $R_{ts}(\boldsymbol{h},\mathcal{E},\mathcal{A})$, since given $\mathcal{E}_i \leq \frac{\mathcal{A}}{2}$, $r_{ts}(h,\mathcal{E},\mathcal{A})$ becomes $\frac{1}{2}\log(1+h_i^2\mathcal{E}^2/9)$.

For this reason, we focus on $R_{ts}(\boldsymbol{h}, \mathcal{E}, \mathcal{A})$ to obtain a power allocation. The power allocation problem can be stated as

$$\begin{aligned} \max_{\boldsymbol{\mathcal{E}}} \quad & f(\boldsymbol{\mathcal{E}}) \\ \text{s.t.} \quad & \boldsymbol{\mathcal{E}} \in [0, \mathcal{A}/2]^N, \ \|\boldsymbol{\mathcal{E}}\|_1 \leq \mathcal{E}, \end{aligned} \tag{58}$$

where $f(\boldsymbol{\mathcal{E}}) = \frac{1}{2}\sum_{i\in\mathcal{N}} \log\left(1 + c_i^2 \mathcal{E}_i^2\right)$ with $c_i = h_i/3$.

This problem can be approached similar to problem (15). The main difference is that $\mathcal{E}_i^+(\lambda)$ and $\mathcal{E}_i^-(\lambda)$ have to be replaced with $\min\left\{\mathcal{E}_i^+(\lambda), \frac{\mathcal{A}}{2}\right\}$ and $\min\left\{\mathcal{E}_i^-(\lambda), \frac{\mathcal{A}}{2}\right\}$, respectively. Indeed, exhaustive search over solutions of this form can be applied to find the optimal solution. Instead of exhaustive search which has exponential complexity in N, the JAPA algorithm can also be used to find a close-to-optimal solution after some modification. The modified algorithm is given below in Algorithm 3.

Algorithm 3 JAPA-P for peak-constrained parallel OWC

function JAPA-P($c_1, \cdots, c_{N+1}$, $\mathcal{E}$)
  $i \leftarrow 1$
  $\tilde{\mathcal{E}} \leftarrow \mathcal{E}$
  while $i < N+1$ **do**
    $\hat{\boldsymbol{\mathcal{E}}} \leftarrow$JAPA($c_i, \cdots, c_{N+1}, \tilde{\mathcal{E}}$) ▷ or SAPA($c_i, \cdots, c_{N+1}, \tilde{\mathcal{E}}$)
    $\boldsymbol{\mathcal{E}} \leftarrow [\boldsymbol{\mathcal{E}}(1:i-1), \hat{\boldsymbol{\mathcal{E}}}(i:N)]$ ▷ $\boldsymbol{x}(a:b) = (x_a, x_{a+1}, \cdots, x_b)$
    if $\mathcal{E}_i \leq \frac{\mathcal{A}}{2}$ **then**
      $i \leftarrow N+1$
    else
      while $\mathcal{E}_i > \frac{\mathcal{A}}{2}$ **do**
        $\mathcal{E}_i \leftarrow \frac{\mathcal{A}}{2}$
        $i \leftarrow i+1$
      end while
      $\tilde{\mathcal{E}} \leftarrow \mathcal{E} - (i-1)\frac{\mathcal{A}}{2}$
    end if
  end while
  return $\mathcal{E}$
end function

This algorithm first releases the peak constraint and finds a power allocation satisfying $\|\boldsymbol{\mathcal{E}}\|_1 = \mathcal{E}$ using the JAPA algorithm (step 5). Assume that this algorithm leads to $\mathcal{E}_j > \frac{\mathcal{A}}{2}$, $j \leq i$. The algorithm then reintroduces the peak constraint by reducing every $\mathcal{E}_j$, $j = 1, \cdots, i$ to $\frac{\mathcal{A}}{2}$ (step 11). This completes the allocation for the first i channels. After doing this, the residual average intensity is calculated (step 14), and the process is repeated for allocating this residual intensity to channels $i+1, \cdots, N$. This is repeated until all channels have been considered.

In this algorithm, step 5 which has complexity $O(N)$ might be called up to N times. Thus, the overall complexity is $O(N^2)$. Notice that we can call the SAPA algorithm in step 5 instead of the JAPA algorithm. We call this variant SAPA-P.

Fig. 6 shows the achievable rate $\frac{1}{2}\sum_{i\in\mathcal{N}} \log\left(1 + \frac{h_i^2 \mathcal{E}_i^2}{9}\right)$, averaged over 10^3 log-normally distributed h_i (35). The power allocation is obtained using exhaustive search, JAPA-P, SAPA-P, and equal power allocation across all channels, i.e., $\mathcal{E}_i = \frac{\mathcal{E}}{N}$ (which is permissible due to (57)). We also plot the rate achievable by activating the strongest channels. Since setting $\mathcal{E}_1 = \mathcal{E}$ is not generally feasible, we distribute $\mathcal{E}$ successively in chunks of $\frac{\mathcal{A}}{2}$ starting with the strongest channel. That is, we set $\mathcal{E}_i = \frac{\mathcal{A}}{2}$ for $i = 1, \cdots, N_s$ where $N_s = \left\lfloor \frac{\mathcal{E}}{\mathcal{A}/2} \right\rfloor$ channels, $\mathcal{E}_{N_s+1} = \mathcal{E} - \frac{N_s \mathcal{A}}{2}$, and $\mathcal{E}_i = 0$ for $i = N_s + 2, \cdots, N$. This variant is denoted 'strongest channels' in Fig. 6. Note that power allocation becomes more important for lower ratios of $\mathcal{E}/\mathcal{A}$. When this ratio is closer to $\frac{N}{2}$, we will have 'almost' enough power to allocate $\mathcal{E}_i = \frac{A}{2}$ to each channel. That is why the plots in Fig. 6b are very close.

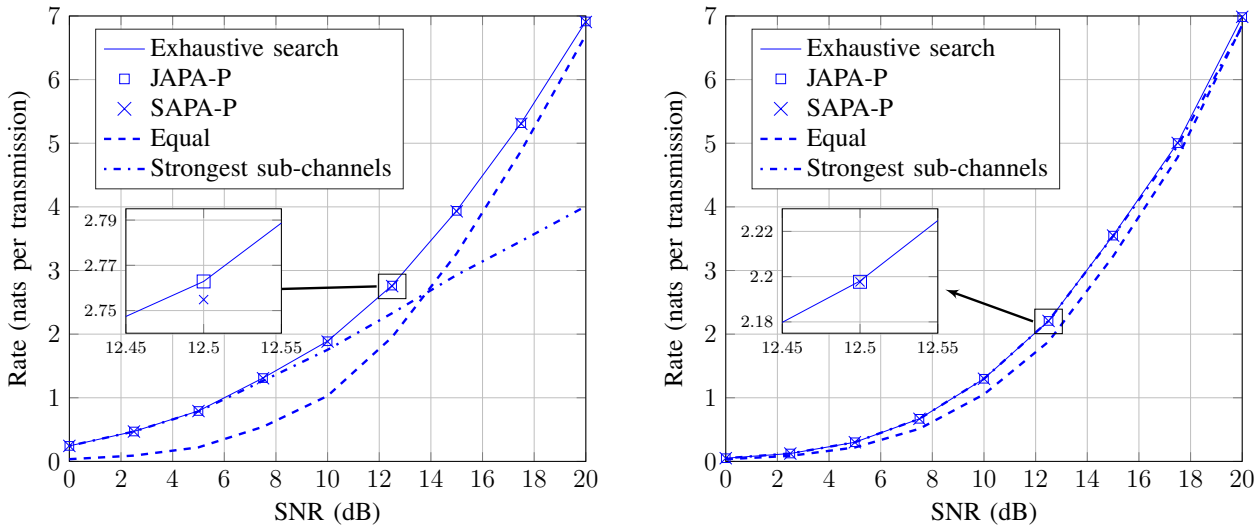
The resulting power allocation from JAPA-P or SAPA-P can be used in conjunction with Theorem 2 to obtain the following achievable rates.

Proposition 2: The capacity $C_{\mathcal{A}}(\boldsymbol{h}, \mathcal{E})$ of a system of N parallel channels with an average constraint $\mathcal{E}$ and a peak constraint $\mathcal{A}$ is lower bounded by the achievable rates

$$\tilde{R}_a(\boldsymbol{h}, \mathcal{E}, \mathcal{A}) = \sum_{i\in\mathcal{N}} r_a(h_i, \mathcal{E}_i, \mathcal{A}), \quad a \in \{l, t, ts\}, \tag{59}$$

where $r_a(h, \mathcal{E}_i, \mathcal{A})$, $a \in \{l, t, ts\}$ is defined in (45), (49), and (50), respectively, and where $(\mathcal{E}_1, \cdots, \mathcal{E}_N)$ is the output of the SAPA-P algorithm with $c_i = h_i/3$.

3) Evaluation: The lower bounds in Proposition 2 are shown in Fig. 7 for log-normally distributed h_i (35), where the achievable rate is averaged over 10^3 channel realizations. Although that rate $\tilde{R}_{ts}(\boldsymbol{h}, \mathcal{E}, \mathcal{A})$ is not optimal, it serves as an intermediate step to obtain a good power allocation for $\tilde{R}_t(\boldsymbol{h}, \mathcal{E}, \mathcal{A})$ and $\tilde{R}_l(\boldsymbol{h}, \mathcal{E}, \mathcal{A})$. The latter two achievable rates are nearly optimal at high SNR as shown in the figure. Namely, the gap between them and the upper bound $\overline{C}_{\mathcal{A}}(\boldsymbol{h}, \mathcal{E})$ converges to a constant as $\mathcal{E}$ increases. This can be shown as follows.



(a) $\mathcal{E} = \frac{N\mathcal{A}}{10}$ (b) $\mathcal{E} = \frac{2N\mathcal{A}}{5}$

Fig. 6: The achievable rate as a function of SNR ($\mathcal{E}/\sigma$) for a peak- and average-constrained system with $N = 4$ channels.

Consider the lower bound $\tilde{R}_t(\boldsymbol{h}, \mathcal{E}, \mathcal{A})$. We know from Sec. IV-B4 that under an average constraint only, the SAPA algorithm finds the optimal solution $\mathcal{E}_i = \frac{\mathcal{E}}{N}$ at high SNR. Thus, steps 5 and 6 in Algorithm 3 give $\mathcal{E}_i = \frac{\mathcal{E}}{N}$ (which is feasible by (57)). Thus, by Proposition 2, $\tilde{R}_t(\boldsymbol{h}, \mathcal{E}, \mathcal{A}) = \sum_{i\in\mathcal{N}} r_t(h_i, \mathcal{E}/N, \mathcal{A})$ is achievable at high SNR. From [24], we know that $r_t(h_i, \mathcal{E}/N, \mathcal{A})$ is within 0.1 nats at most of the upper bound $\min\{\overline{r}(h_i, \mathcal{E}/N), \overline{\overline{r}}(h_i, \mathcal{A})\}$ at high SNR. This upper bound can be written as

$$\min\left\{\frac{1}{2}\log\left(\frac{eh_i^2\mathcal{E}^2}{2\pi N^2}\right), \frac{1}{2}\log\left(\frac{h_i^2\mathcal{A}^2}{2\pi e}\right)\right\}.$$

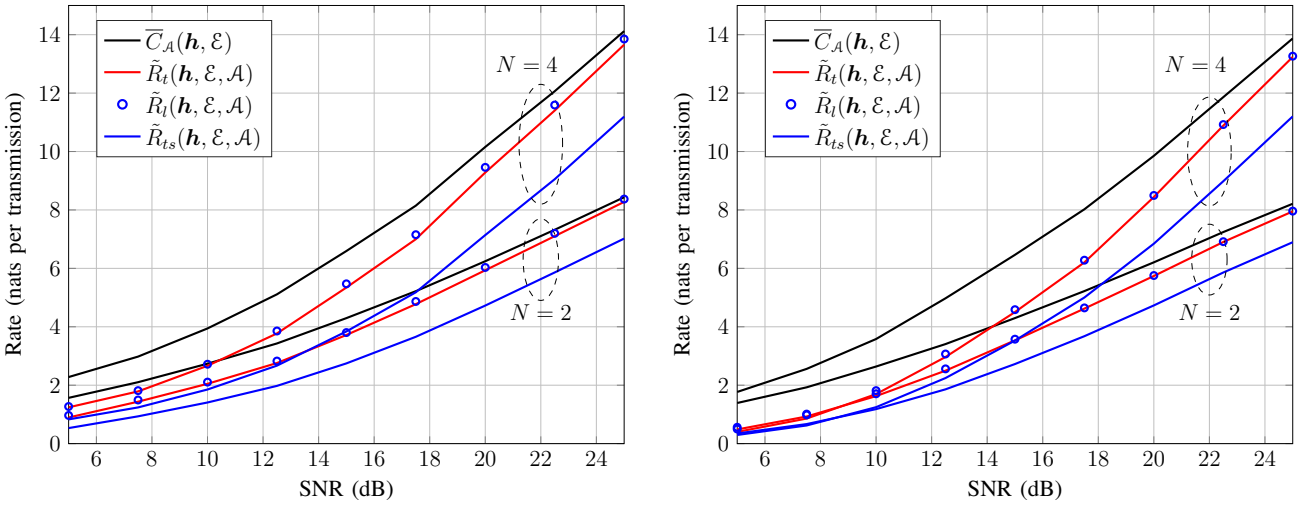
By applying this over all channels, we get the high-SNR achievable rate

$$\tilde{R}_t(\boldsymbol{h}, \mathcal{E}, \mathcal{A}) = \sum_{i\in\mathcal{N}} \min\left\{\frac{1}{2}\log\left(\frac{eh_i^2\mathcal{E}^2}{2\pi N^2}\right), \frac{1}{2}\log\left(\frac{h_i^2\mathcal{A}^2}{2\pi e}\right)\right\} - 0.1N. \tag{60}$$

On the other hand, consider the upper bound $\overline{C}_{\mathcal{A}}(\boldsymbol{h}, \mathcal{E})$ given in Theorem 3. At high SNR, this upper bound satisfies

$$\overline{C}_{\mathcal{A}}(\boldsymbol{h}, \mathcal{E}) \leq \max_{\boldsymbol{\mathcal{E}}\in\mathcal{S}} \sum_{i\in\mathcal{N}} \overline{r}(h_i, \mathcal{E}_i) \tag{61}$$

$$= \sum_{i\in\mathcal{N}} \frac{1}{2}\log\left(\frac{eh_i^2\mathcal{E}^2}{2\pi N^2}\right), \tag{62}$$



(a) $\mathcal{E}=\frac{N\mathcal{A}}{10}$ (b) $\mathcal{E}=\frac{2N\mathcal{A}}{5}$

Fig. 7: Achievable rates as a function of SNR ($\mathcal{E}/\sigma$) for a system with $N=2$ and $N=4$ and with both average and peak constraints. The upper bound $\overline{C}_{\mathcal{A}}(\boldsymbol{h},\mathcal{E})$ is from Theorem (3), and the lower bounds $\tilde{R}_a(\boldsymbol{h},\mathcal{E},\mathcal{A})$, $a\in\{l,t,ts\}$ are given in (59).

which follows by (37)-(38), and

$$\overline{C}_{\mathcal{A}}(\boldsymbol{h},\mathcal{E}) \le \max_{\mathcal{E}\in\mathcal{S}} \sum_{i\in\mathcal{N}} \overline{\overline{r}}(h_i,\mathcal{A}) \tag{63}$$

$$= \sum_{i\in\mathcal{N}} \frac{1}{2}\log\left(\frac{h_i^2\mathcal{A}^2}{2\pi e}\right). \tag{64}$$

By comparing (60), (62), and (64), we can see that

$$\overline{C}_{\mathcal{A}}(\boldsymbol{h},\mathcal{E}) - \tilde{R}_t(\boldsymbol{h},\mathcal{E},\mathcal{A}) \le 0.1N \text{ nats} \tag{65}$$

at high SNR, leading to the following corollary.

Corollary 3: The capacity of a system of N parallel channels with an average constraint $\mathcal{E}$ and a peak constraint $\mathcal{A}$, where $\mathcal{E}\le\frac{N\mathcal{A}}{2}$ satisfies

$$\lim_{\mathcal{E}\to\infty}\left[C_{\mathcal{A}}(\boldsymbol{h},\mathcal{E}) - \sum_{i\in\mathcal{N}}\frac{1}{2}\log\left(\frac{eh_i^2\mathcal{E}^2}{2\pi N^2}\right)\right] \le 0.1N \text{ nats} \tag{66}$$

C. Parallel Channels without CSIT

In the absence of CSIT, we study the outage probability of the system as in Sec. V. An outage capacity upper bound can be obtained by fixing a specific distribution over each channel, and using

equal power allocation. This leads to the outage probability upper bounds

$$P_o(\mathcal{E}) \leq \mathbb{P}\left\{R_l'(\boldsymbol{h}, \mathcal{E}) < R_0\right\}, \tag{67}$$

$$P_o(\mathcal{E}) \leq \mathbb{P}\left\{R_t'(\boldsymbol{h}, \mathcal{E}) < R_0\right\}, \tag{68}$$

$$P_o(\mathcal{E}) \leq \mathbb{P}\left\{R_{ts}'(\boldsymbol{h}, \mathcal{E}) < R_0\right\}, \tag{69}$$

where

$$R_a(\boldsymbol{h}, \mathcal{E}) = \sum_{i \in \mathcal{N}} r_a(h_i, \mathcal{E}/N, \mathcal{A}), \quad a \in \{l, t, ts\}, \tag{70}$$

and $r_a(h_i, \mathcal{E}_i, \mathcal{A})$, $a \in \{l, t, ts\}$ is defined in (45), (49), (50), and respectively.

On the other hand, the outage probability without CSIT is lower bounded by that with CSIT. Thus, the outage probability lower bound (43) also holds in this case, i.e.,

$$P_o(\mathcal{E}) \geq \mathbb{P}\left\{\overline{C}_{\mathcal{A}}(\boldsymbol{h}, \mathcal{E}) < R_0\right\}, \tag{71}$$

where $\overline{C}_{\mathcal{A}}(\boldsymbol{h}, \mathcal{E})$ is given in Theorem 3.

The upper bound (70) and the lower bound (71) are fairly tight at high SNR. Namely, at high SNR, $R_t'(\boldsymbol{h}, \mathcal{E}, \mathcal{A})$ can be written similar to $\tilde{R}_t(\boldsymbol{h}, \mathcal{E}, \mathcal{A})$ (60), whose gap to $\overline{C}_{\mathcal{A}}(\boldsymbol{h}, \mathcal{E})$ is at most $0.1N$ at high SNR (65). Consequently, at high SNR,

$$P_o(\mathcal{E}) \leq \mathbb{P}\left\{\sum_{i \in \mathcal{N}} \frac{1}{2} \log\left(\frac{e h_i^2 \mathcal{E}^2}{2\pi N^2}\right) - 0.1N < R_0\right\}, \tag{72}$$

$$P_o(\mathcal{E}) \geq \mathbb{P}\left\{\sum_{i \in \mathcal{N}} \frac{1}{2} \log\left(\frac{e h_i^2 \mathcal{E}^2}{2\pi N^2}\right) < R_0\right\}. \tag{73}$$

By neglecting $0.1N$ for large $\mathcal{E}$, we conclude that Corollary 2 also holds under a peak constraint.

VII. Conclusion

We have studied a system of N parallel OWC channels employing IM-DD, under a total average intensity constraint. Optimally distributing the available resources over the N channels is important for enhancing the overall system performance when channel-state information (CSI) is available at the transmitter. However, under an exponential input distribution, which is optimal for $N = 1$ at high SNR, the optimization problem for $N > 1$ is non-convex. We have proposed simple algorithms which almost always find the optimal resource allocation. The optimal power allocation at moderate SNR is different from water-filling. However, it is similar to water-filling at high and low SNR. Namely, at high SNR, power is distributed equally over the N channels, while at low SNR it is assigned to

the strongest channel. As a result, we obtained capacity upper and lower bounds, which are fairly close at moderate SNR and nearly tight at high SNR. Without CSI at the transmitter, since reliable communication is not feasible, we provide bounds on the outage probability which are also tight at high SNR. We have also obtained similar results for a system with both average and peak intensity constraints. An interesting extension of this work would be to investigate how these bounds can be approached practically, by using adaptive modulation/coding at a target error/outage probability e.g., as in [31], [32].

REFERENCES

[1] J. Karout, E. Agrell, K. Szczerba, , and M. Karlsson, "Optimizing constellations for single-subcarrier intensity-modulated optical Systems," *IEEE Trans. on Info. Theory*, vol. 58, no. 7, pp. 4645–4659, Apr. 2012.

[2] R. Drost and B. Sadler, "Constellation design for channel precompensation in multi-wavelength visible light communications," *IEEE Transactions on Communications*, vol. 62, no. 6, pp. 1995–2005, June 2014.

[3] S. Dissanayake and J. Armstrong, "Comparison of ACO-OFDM, DCO-OFDM and ADO-OFDM in IM/DD systems," *Journal of Lightwave Technology*, vol. 31, no. 7, pp. 1063–1072, April 2013.

[4] X. Li, J. Vucic, V. Jungnickel, and J. Armstrong, "On the capacity of intensity-modulated direct-detection systems and the information rate of ACO-OFDM for indoor optical wireless applications," *IEEE Transactions on Communications*, vol. 60, no. 3, pp. 799–809, March 2012.

[5] R. You and J. Kahn, "Upper-bounding the capacity of optical IM/DD channels with multiple-subcarrier modulation and fixed bias using trigonometric moment space method," *IEEE Transactions on Information Theory*, vol. 48, no. 2, pp. 514–523, Feb 2002.

[6] A. A. Farid and S. Hranilovic, "Channel capacity and non-uniform signalling for free-space optical intensity channels," *IEEE Journal on Selected Areas in Communications*, vol. 27, no. 9, pp. 1–12, Dec. 2009.

[7] A. García-Zambrana, C. Castillo-Vázquez, and B. Castillo-Vázquez, "On the capacity of FSO links over Gamma-Gamma atmospheric turbulence channels using OOK signaling," *EURASIP J. Wirel. Commun. Netw.*, Jan. 2010.

[8] A. A. Farid and S. Hranilovic, "Outage capacity optimization for free-space optical links with pointing errors," *IEEE/OSA Journal of Lightwave Technology*, vol. 25, no. 7, pp. 1702–1710, Jul. 2007.

[9] M. A. Kashani and M. Uysal, "Outage performance and diversity gain analysis of free-space optical multi-hop parallel relaying," *IEEE/OSA Journal of Optical Communications and Networking*, vol. 5, no. 8, pp. 901–909, Aug. 2013.

[10] J. Akella, M. Yuksel, and S. Kalyanaraman, "Error analysis of multi-hop free-space optical communications," in *Proc. of IEEE International Conference on Communications (ICC)*, May. 2005, pp. 1777–1781.

[11] S. Kazemlou, S. Hranilovic, and S. Kumar, "All-optical multihop free-space optical communication systems," *Journal of Lightwave Technology*, vol. 29, no. 18, pp. 2663–2669, Jun. 2011.

[12] M. A. Khalighi and M. Uysal, "Survey on free space optical communications: A communication theory perspective," *IEEE Communication Surveys and Tutorials*, vol. 16, no. 4, pp. 2231–2258, 4th quarter 2014.

[13] H. Elgala, R. Mesleh, and H. Haas, "Indoor optical wireless communication: Potential and state-of-the-art," *IEEE Comm. Magazine*, vol. 49, no. 9, pp. 56–62, Sep. 2011.

[14] P. Butala, H. Elgala, and T. Little, "SVD-VLC: A novel capacity maximizing VLC MIMO system architecture under illumination constraints," in *IEEE Globecom Workshops*, Dec 2013, pp. 1087–1092.

[15] H. Kazemi, Z. Mostaani, M. Uysal, and Z. Ghassemlooy, "Outage performance of MIMO FSO systems in Gamma-Gamma fading channels," in *Proc. of IEEE 18th European Conference on Network and Optical Communications*, Graz, Austria, Jul. 2013.

[16] A. A. Farid and S. Hranilovic, "Diversity gain and outage probability for MIMO free-space optical links with misalignment," *IEEE Trans. on Communications*, vol. 60, no. 2, pp. 479–487, Feb. 2012.

[17] T. Tsiftsis, H. Sandalidis, G. Karagiannidis, and M. Uysal, "Optical wireless links with spatial diversity over strong atmospheric turbulence channels," *IEEE Transactions on Wireless Communications*, vol. 8, no. 2, pp. 951–957, Feb 2009.

[18] M. Riediger, R. Schober, and L. Lampe, "Fast multiple-symbol detection for free-space optical communications," *IEEE Transactions on Communications*, vol. 57, no. 4, pp. 1119–1128, April 2009.

[19] L. Zeng, D. O'brien, H. Minh, G. Faulkner, K. Lee, D. Jung, Y. Oh, and E. T. Won, "High data rate multiple input multiple output (MIMO) optical wireless communications using white led lighting," *IEEE Journal on Selected Areas in Communications*, vol. 27, no. 9, pp. 1654–1662, December 2009.

[20] Q. Gao, R. Wang, Z. Xu, and Y. Hua, "DC-informative joint color-frequency modulation for visible light communications," *Journal of Lightwave Technology*, vol. 33, no. 11, pp. 2181–2188, June 2015.

[21] E. Monteiro and S. Hranilovic, "Design and implementation of color-shift keying for visible light communications," *Journal of Lightwave Technology*, vol. 32, no. 10, pp. 2053–2060, May 2014.

[22] A. Lapidoth, S. M. Moser, and M. Wigger, "On the capacity of free-space optical intensity channels," *IEEE Trans. on Info. Theory*, vol. 55, no. 10, pp. 4449–4461, Oct. 2009.

[23] A. A. Farid and S. Hranilovic, "Capacity bounds for wireless optical intensity channels with Gaussian noise," *IEEE Trans. on Info. Theory*, vol. 56, no. 12, pp. 6066–6077, Dec. 2010.

[24] A. Chaaban, J.-M. Morvan, and M.-S. Alouini, "Free-space optical communications: Capacity bounds, approximations, and a new sphere-packing perspective," *KAUST Technical Report http://hdl.handle.net/10754/552096*, Apr. 2015.

[25] J.-B. Wang, Q.-S. Hu, J. Wang, M. Chen, and J.-Y. Wang, "Tight bounds on channel capacity for dimmable visible light communications," *Journal of Lightwave Technology*, vol. 31, no. 23, pp. 3771–3779, Dec. 2013.

[26] S. Haas, "Capacity of and coding for multiple-aperture, wireless, optical communications," Ph.D. dissertation, Massachusetts Institute of Technology, Cambridge, Massachusetts, USA, 2003.

[27] S. Boyd and L. Vandenberghe, *Convex Optimization*. New York, NY, USA: Cambridge University Press, 2004.

[28] D. Tse and P. Viswanath, *Fundamentals of Wireless Communications*. Cambridge University Press, 2005.

[29] T. Cover and J. Thomas, *Elements of Information Theory (Second Edition)*. John Wiley and Sons, Inc., 2006.

[30] A. Chaaban, Z. Rezki, and M.-S. Alouini, "On the capacity region of the intensity-modulation direct-detection optical broadcast channel," *KAUST Technical Report, http://hdl.handle.net/10754/565823*, Aug. 2015.

[31] M. Karimi and M. Uysal, "Novel adaptive transmission algorithms for free-space optical links," *IEEE Transactions on Communications*, vol. 60, no. 12, pp. 3808–3815, December 2012.

[32] I. Djordjevic, "Adaptive modulation and coding for free-space optical channels," *IEEE/OSA Journal of Optical Communications and Networking*, vol. 2, no. 5, pp. 221–229, May 2010.