Fundamental Limits of Parallel Optical Wireless Channels: Capacity Results and Outage Formulation

Anas Chaaban, *Member, IEEE*, Zouheir Rezki, *Senior Member, IEEE*, and Mohamed-Slim Alouini, *Fellow, IEEE*
(Invited Paper)

Abstract—Multi-channel (MC) optical wireless communication (OWC) systems employing wave-division multiplexing for outdoors free-space optical communications, or multi-user time-division multiple access for indoors visible-light communications, e.g., can be modeled as parallel channels. Multi-input multi-output OWC systems can also be transformed, possibly with some performance loss, to parallel channels using pre-/post-coding. Studying the performance of such MC-OWC systems requires characterizing the capacity of the underlying parallel channels. In this paper, upper and lower bounds on the capacity of constant parallel OWC channels with a total average intensity constraint are derived. Then, the paper focuses on finding intensity allocations that maximize the lower bounds given channel-state information at the transmitter (CSIT). Due to its nonconvexity, the KKT conditions are used to describe a list of candidate allocations. Instead searching exhaustively for the best solution, low-complexity near-optimal algorithms are proposed. The resulting optimized lower bound nearly coincides with capacity at high signal-to-noise ratio (SNR). Under a quasi-static channel model and in the absence of CSIT, outage probability upper and lower bounds are derived. Those bounds also meet at high SNR, thus characterizing the outage capacity in this regime. Finally, the results are extended to a system with both average and peak intensity constraints.

Index Terms—Optical wireless; intensity-modulation; parallel channels, intensity allocation; capacity bounds; outage probability.

I. Introduction

One of the most challenging problems facing today's wireless communications is spectrum scarcity. This makes meeting the increasing demand for high data-rates extremely difficult, given the exhausted radio frequency (RF) spectrum. This motivated researcher to explore new frequency bands for wireless communications, such as the millimeter wave band [2] and the optical spectrum [3], [4].

Due to the above challenge, optical wireless communication (OWC) has witnessed increasing research attention recently. Of particular interest are systems employing intensity-modulation and direct-detection (IM-DD) due to their practical simplicity. For instance, [5], [6] studied constellation design

A. Chaaban and M.-S. Alouini 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,slim.alouini}@kaust.edu.sa.

Z. Rezki was with the Division of CEMSE at KAUST, and is now with the Department of Electrical and Computer Engineering, University of Idaho, Moscow, ID, USA. Email: zrezki@uidaho.edu.

Parts of the results in this paper were presented in ICC 2016 [1].

and [7]–[9] studied OFDM for IM-DD systems. Information-theoretic achievable rates in IM-DD channels have been studied in [10] which focused on on-off keying, and in [11] which considered higher order modulation. Outage performance of IM-DD systems [12]–[14] as well as multi-hop systems [15]–[17] have also been considered in the literature. For more recent advances in this field, the reader is referred to [3], [18], [19] and references therein.

OWC is significantly affected by distance, weather, and transmitter-receiver alignment, which drastically impacts the communication rates. To mitigate these effects, multi-channel (MC) techniques can be used. In outdoors OWC e.g., one can employ multiple laser transmitters and multiple photo-detectors to form an optical multiple-input multiple-output (MIMO) channel. This can also be applied in indoors OWC using multiple LEDs and detectors [20]. The optical MIMO channel was studied in [13], [21]–[24]. In such a MIMO system, cross-talk between signals from different transmitters occurs. Systems with little or no cross-talk can be well approximated as parallel channels. This can be the case, e.g., if laser-detector pairs are perfectly aligned, or in indoors MIMO OWC if the transmitters and receivers are properly spaced [20]. On the other hand, if cross-talk occurs, the MIMO channel can also be transformed to parallel channels by some processing techniques. For instance, one can apply channel inversion [25] and ignore the induced noise correlation at the receiver, or apply DC-biased SVD transmission [21]. Although those techniques might incur some performance loss, they are practically appealing due to their simplicity. That being said, the achievable performance using these techniques depends on the underlying system of parallel channels.

Another OWC system that can be modeled as parallel channels is a wave-division multiplexing (WDM) system, in which several carriers with different wavelengths are multiplexed to one beam [26], [27]. One form of WDM proposed for indoors visible-light communication is color-frequency modulation [28], [29], where LEDs of different colors are used as transmitters, and color filters are used at the receiver. Yet another example is a multi-user time-division multiple-access (TDMA) OWC system [30]. The overall system is a set of parallel channels, each to one of the receivers.

Therefore, for a better understanding of the performance of the above systems, it is important to study the performance of parallel OWC channels, which is the main topic of this paper. To this end, we consider N parallel IM-DD channels, where N light sources are used to communicate with a receiver with N detectors. The emitted light is constrained by a total

	RF parallel channels	IM-DD parallel channels
Inputs	$X_i \in \mathbb{C},\ i \in \{1,\dots,N\}$	$X_i \in \mathbb{R},\ i \in \{1,\dots,N\}$
Constraints	$\sum\limits_{i=1}^{N} \underbrace{\mathbb{E}_{X_i}[|X_i|^2]}_{P_i} \leq P$	$\sum\limits_{i=1}^{N} \underbrace{\mathbb{E}_{X_i}[X_i]}_{\mathcal{E}_i} \leq \mathcal{E},\ X_i \geq 0$
Outputs	$Y_i = h_i X_i + Z_i$ Z_i Complex-Gaussian $h_i \in \mathbb{C}$	$Y_i = h_i X_i + Z_i$ Z_i Real-Gaussian $h_i \in \mathbb{R}_+$
Capacity	$\max\limits_{P_i} \sum\limits_{i=1}^{N} \log(1+|h_i|^2 P_i)$	$\max\limits_{\mathcal{E}_i} \sum\limits_{i=1}^{N} \max\limits_{p(x_i)} I(X_i; Y_i)$
Allocation	Convex problem, Water-filling [36]	Nonconvex

TABLE I: Comparison of RF and IM-DD OWC parallel channels.

average intensity constraint. We focus on the Gaussian channel model with input-independent noise [31, Ch. 7]. The special case $N = 1$ reduces to a single IM-DD channel, whose capacity was studied in [32]–[35]. Here, we develop capacity upper and lower bounds for the case $N > 1$ when the channel is constant and hence assuming perfect channel state information at the transmitter (CSIT). The main additional ingredient in comparison to [32]–[35] is that the bounds have to be optimized with respect to intensity allocation across the channels, which is important for enhancing performance.

Given CSIT, we focus on optimizing the rate achieved by using an exponentially distributed input (a capacity lower bound) over each of the N channels. This lower bound is tight at high SNR for $N = 1$ as shown in [32]. The resulting optimization for $N > 1$ is nonconvex, contrary to its coherent-detection (RF or optical) counterpart where water-filling is the optimal solution [37], [38]. Table I highlights the main differences between RF and IM-DD parallel channels. To solve this optimization, we describe its solution by using the Karush-Kuhn-Tucker (KKT) conditions [39], leading to a list of candidates solutions. Generally, equal intensity allocation is optimal at high SNR, activating the strongest channel is optimal at low SNR, and using a subset of channels is optimal at moderate SNR. The list size is exponential in the number of channels N, rendering finding the optimal solution at moderate SNR impractical. For instance, for a 4×4 MIMO system [20], there would be $3^4 = 81$ candidate solutions. Instead of searching over this list, we propose a simple (linear complexity) algorithm which finds a near-optimal intensity allocation for maximizing the capacity lower bound. The main idea of this algorithm stems from the behavior of the candidate solutions at moderate optical intensity, which allows eliminating many candidates thus reducing complexity. We also propose a less accurate algorithm with slightly lower complexity. This leads to a simple capacity lower bound, which we prove to be asymptotically tight at high SNR, where SNR is defined as the ratio of the average optical intensity to the noise standard deviation.

The availability of CSIT is not a strong assumption in OWC, whose coherence time is typically much larger than the symbol duration [12], [18]. Thus, estimation and feedback of the channel state can be achieved in negligible time without considerably affecting performance. Furthermore, in a full-

duplex system, CSI can be estimated directly at the transmitter due to channel reciprocity [18]. However, in the absence of such a feedback mechanism, and under channel fading, availability of CSIT can not be assumed. For such a system with no CSIT, and under a quasi-static fading channel model [38, Sec. 5.4.1], the capacity in the strict Shannon sense is zero. Performance in this case is captured by the outage probability [38], i.e., the probability that the channel cannot support a desired target rate. Intensity allocation is not possible in this case due to the absence of CSIT. A feasible strategy however is equal intensity allocation across the N channels. Using this allocation, we formulate an outage probability upper bound and lower bound, and we show that the bounds coincide at high SNR, leading again to the high-SNR outage-capacity of the system.

Finally, since practical systems typically have a peak intensity constraint at the transmitter, we consider the impact of this constraint on the results. Assuming constant channels with CSIT, we derive capacity lower and upper bounds, and we show that equal intensity allocation maximizes the lower bound if the total average constraint is larger than $N/2$ times the peak constraint. Otherwise, we describe the optimal solution for maximizing the lower bound, which leads to equal allocation at high SNR and activating a subset of the strongest channels at low SNR. We also propose an intensity allocation algorithm which yields a near-optimal solution at moderate SNR. Moreover, we show that this solution is nearly capacity achieving at high SNR, where the gap between the capacity upper and lower bounds becomes negligible. In the absence of CSIT and under a quasi-static fading channel model, we formulate bounds on the outage probability, which are fairly tight at high SNR as well. The contributions of the paper are summarized in Table II.

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 quasi-static channel without 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.

II. SYSTEM MODEL AND MOTIVATION

A. System Model

Consider an OWC system with N light sources and N detectors. Each source-detector pair forms a point-to-point (P2P) channel and employs an IM-DD scheme. This system is modeled as N parallel channels as shown in Fig. 1. The transmit intensity over channel $i \in \mathcal{N} = \{1,\dots,N\}$ is $X_i \in \mathbb{R}_+$. Due to practical and safety considerations, X_i 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 $\|\mathcal{E}\|_1 \leq \mathcal{E}$ where $\|\mathcal{E}\|_1$ is the ℓ_1-norm of $\mathcal{E} = (\mathcal{E}_1,\dots,\mathcal{E}_N)$. Additionally, a peak intensity constraint might be imposed, i.e., $X_i \leq \mathcal{A}$. For simplicity of exposition,

---

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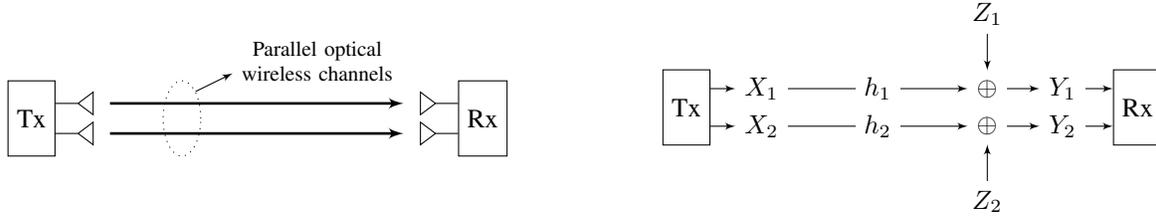
(a) An OWC system with two transmit apertures (Lasers e.g.) and two detectors. The total average optical intensity is constrained by $\mathcal{E}$.

(b) Parallel Gaussian channels. Here h_i is a channel gain, Z_i is Gaussian noise, and $X_i \geq 0$ is the optical intensity constrained by $\mathbb{E}_{X_1}[X_1] + \mathbb{E}_{X_2}[X_2] \leq \mathcal{E}$.

Fig. 1: A system of parallel OWC channels can be modeled as Gaussian point-to-point channels.

Constraint		Total-average	Total-average and peak
Constant channels	Capacity bounds	✓	✓
	Lower bound maximization (optimal solution)	✓	✓
	High-SNR optimal allocation	Equal	Equal
	Low-SNR optimal allocation	Strongest channel only	Strongest $\left\lfloor \frac{\mathcal{E}}{\mathcal{A}/2} \right\rfloor + 1$ channels
	Allocation algorithms	✓	✓
	High-SNR capacity	✓	within a small constant
Quasi-static	Outage probability bounds	✓	✓
	Outage capacity	✓	–

TABLE II: Contributions for parallel channels with a total average constraint $\mathcal{E}$ and a peak constraint per channel $\mathcal{A}$. 'Optimal allocation' refers to optimality with respect to the lower bound maximization.

we ignore the peak constraint at first and focus on the case with $\mathcal{A} = \infty$. We discuss the case $\mathcal{A} < \infty$ in Sec. VI.

The received signal over channel i is

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

where $h_i \in \mathbb{R}_+$ represents the channel scaling factor, the noise Z_i is Gaussian with zero mean and variance $\sigma^2 = 1$, and Z_i and Z_j are independent $\forall j \neq i$. The noise instances (over time) are also independent and identically distributed (i.i.d.). We assume without loss of generality that

$$h_1 \geq h_2 \geq \ldots \geq h_N, \tag{2}$$

and we define $\boldsymbol{h} = (h_1, \ldots, h_N)$. We further assume that $\boldsymbol{h}$ is constant throughout a transmission block, which is a realistic assumption in optical channels whose variation is very slow in comparison to the symbol duration.

Two cases are considered regarding $\boldsymbol{h}$: Either it is constant and known at the transmitter (perfect CSIT), or it is changing between transmission blocks and unknown at the transmitter (no CSIT). CSIT acquisition requires the existence of a channel from the receiver to the transmitter. This can be an RF or an optical channel. In a hybrid RF/optical system, CSIT can be obtained by sending a pilot signal which is used to form a channel estimate at the receiver, which is in turn fed back

to the transmitter through an RF feedback channel [40]. In a full-duplex OWC system, CSIT can be estimated directly at the transmitter due to channel reciprocity, by sending a pilot signal from the receiver to the transmitter [18]. Due to the slow channel variation, CSIT acquisition overhead can be neglected. If such a CSIT acquisition mechanism is not available, then the system is said to have no CSIT.

For the perfect CSIT case, we are interested in the capacity $C(\boldsymbol{h}, \mathcal{E})$ of the system for a given $\boldsymbol{h}$, defined as follows. Let $n \in \mathbb{N}_+$ be the code-length used by the transmitter. Also, let $P_e^{(n)}$ be 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 $n \to \infty$.

For the no CSIT case, we are interested in the outage probability of the system, when the transmitter sends at a constant rate over a quasi-static fading channel [38, Sec. 5.4.1].

B. Motivation

Studying parallel channels is self-motivated since parallel channels arise in WDM OWC systems and multi-user TDMA OWC systems. Another instance is MIMO OWC systems as shown in the following example.

A Motivating Example: Consider a MIMO OWC system with N light sources and N detectors. Such a system can be captured by the following input-output equation

$$\boldsymbol{b} = \boldsymbol{T}\boldsymbol{a} + \boldsymbol{n}, \tag{3}$$

where $\boldsymbol{n} \in \mathbb{R}^N$ is i.i.d. Gaussian noise with zero mean and covariance matrix $\boldsymbol{I}$ (identity matrix), $\boldsymbol{a} = [a_1, \ldots, a_N]^t \in \mathbb{R}_+^N$ is the transmit signal (intensity), and $\boldsymbol{b} \in \mathbb{R}^N$ is the received signal. The (i, j) element of the channel matrix $\boldsymbol{T} \in \mathbb{R}_+^{N \times N}$ represents the channel scaling factor from the jth light source to the ith detector.

Consider a simple receiver architecture as in [25], which employs a decorrelator and decodes each signal a_i separately (see [38, Chapter 3] for an RF equivalent). The received signal after the decorrelator is

$$\tilde{\boldsymbol{b}} = \boldsymbol{T}^{-1}\boldsymbol{b} = \boldsymbol{a} + \boldsymbol{T}^{-1}\boldsymbol{n}. \tag{4}$$

The result is a set of N parallel channels

$$\tilde{b}_i = a_i + \tilde{n}_i, \quad i \in \mathcal{N}, \tag{5}$$

where $\tilde{n}_i = \boldsymbol{t}_i^t \boldsymbol{n}$ and $\boldsymbol{t}_i^t$ is the ith row of $\boldsymbol{T}^{-1}$. The noises are now correlated since $\mathbb{E}_{\tilde{N}_i, \tilde{N}_j}[\tilde{n}_i \tilde{n}_j] = \frac{\boldsymbol{t}_i \boldsymbol{t}_j^t}{\|\boldsymbol{t}_i\| \|\boldsymbol{t}_j\|} \neq 0$. This correlation can be small if $\boldsymbol{T}$ is close to diagonal such as [28, eq. (11)], or large otherwise [20, eq. (14)]. In the former case, modeling the MIMO channel as parallel channels is a good approximation. This is not the case in the latter. However, one can still treat (5) as parallel channels in this case by ignoring the noise correlation, which induces some performance loss. To reduce this loss, one can also exploit noise correlation by estimating $\tilde{n}_2$ from $\tilde{b}_1 - a_1$ after decoding a_1, and subtracting this estimate from $\tilde{b}_2$ before decoding a_2, and then estimating $\tilde{n}_3$ from $(\tilde{b}_1 - a_1, \tilde{b}_2 - a_2)$ and so on. The result after this operation is also parallel channels similar to (5), but with uncorrelated noises whose variances are modified accordingly.

Thus, WDM systems, multi-user TDMA systems, and MIMO systems as in this example motivate the study of the parallel IM-DD channels. Next, we study parallel channels starting with the perfect CSIT case. But before that, let us review some bounds on the capacity of the IM-DD P2P channel.

III. PRELIMINARIES

Consider a P2P channel $Y = hX + Z$ where $X \in \mathbb{R}_+$ satisfies $\mathbb{E}_X[X] \leq \mathcal{E}$, Z is real Gaussian with zero mean and unit variance, and $h \in \mathbb{R}_+$. Capacity upper bounds and achievable rates (capacity lower bounds) for the IM-DD P2P channel have been derived in [11], [32]–[34]. In this paper, we are interested in achievable rates of the form

$$\frac{1}{2} \log(1 + c^2 \mathcal{E}^2) \tag{7}$$

for some constant c.[2] This captures the behavior of achievable rates with exponentially distributed X [32] or truncated-Gaussian distributed X [34]. For instance, when X follows an exponential distribution [32], we obtain the capacity lower bound

$$C(h, \mathcal{E}) \geq r(h, \mathcal{E}) = \frac{1}{2} \log \left(1 + \frac{e h^2 \mathcal{E}^2}{2\pi} \right). \tag{8}$$

In this case, $c = h \sqrt{\frac{e}{2\pi}}$.

Several upper bounds on the capacity of this channel exist. We restrict ourselves to the upper bound

$$C(h, \mathcal{E}) \leq \overline{r}(h, \mathcal{E}) = \inf_{\beta, \delta > 0} b(h, \mathcal{E}, \beta, \delta), \tag{9}$$

where $b(h, \mathcal{E}, \beta, \delta)$ is given in (6) at the top of the page. This upper bound was given in [32], 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.

IV. PARALLEL CHANNELS WITH CSIT

A. Capacity Bounds Given $\boldsymbol{h}$

To bound the capacity of the system for a given $\boldsymbol{h}$, we first note it is equal to the sum of the capacities of the N channels

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

$C(h_i, \mathcal{E}_i)$, optimized over the set of feasible power allocations $\mathcal{E}$. That is,

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

where $\boldsymbol{\mathcal{E}} \in \mathbb{R}_+^N$ such that $\|\boldsymbol{\mathcal{E}}\|_1 \leq \mathcal{E}$. This is proved as in [36], repeated here for completeness.

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

$$C(\boldsymbol{h}, \mathcal{E}) = \max_{p(x_1, \ldots, x_N) \in \mathcal{P}} I(X_1, \ldots, X_N; Y_1, \ldots, Y_N) \tag{11}$$

$$= \max_{\boldsymbol{\mathcal{E}} \in \mathcal{S}} \max_{p(x_1, \ldots, x_N) \in \mathcal{Q}(\boldsymbol{\mathcal{E}})} I(X_1, \ldots, X_N; Y_1, \ldots, Y_N), \tag{12}$$

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

$$\mathcal{S} = \{\boldsymbol{\mathcal{E}} \in \mathbb{R}_+^N | \|\boldsymbol{\mathcal{E}}\|_1 \leq \mathcal{E}\}, \tag{13}$$

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 [36, Sec. 9.4], we can write

$$\max_{p(x_1, \ldots, x_N) \in \mathcal{Q}(\boldsymbol{\mathcal{E}})} I(X_1, \ldots, X_N; Y_1, \ldots, Y_N)$$

$$\leq \max_{p(x_1, \ldots, x_N) \in \mathcal{Q}(\boldsymbol{\mathcal{E}})} \sum_{i \in \mathcal{N}} I(X_i; Y_i) \tag{14}$$

$$\leq \sum_{i \in \mathcal{N}} \max_{p(x_i) \in \mathcal{Q}_i(\mathcal{E}_i)} I(X_i; Y_i) \tag{15}$$

$$= \sum_{i \in \mathcal{N}} C(h_i, \mathcal{E}_i), \tag{16}$$

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

$$C(\boldsymbol{h}, \mathcal{E}) \leq \max_{\boldsymbol{\mathcal{E}} \in \mathcal{S}} \sum_{i \in \mathcal{N}} C(h_i, \mathcal{E}_i). \tag{17}$$

Since this upper bound is achievable by coding independently over each channel, we obtain (10).

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

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

$$R(\boldsymbol{h}, \mathcal{E}) = \max_{\boldsymbol{\mathcal{E}} \in \mathcal{S}} \sum_{i \in \mathcal{N}} r(h_i, \mathcal{E}_i), \tag{18}$$

$$\overline{C}(\boldsymbol{h}, \mathcal{E}) = \max_{\boldsymbol{\mathcal{E}} \in \mathcal{S}} \sum_{i \in \mathcal{N}} \overline{r}(h_i, \mathcal{E}_i), \tag{19}$$

and $\mathcal{S}$, $r(h_i, \mathcal{E}_i)$, and $\overline{r}(h_i, \mathcal{E}_i)$ are given in (13), (8) and (9), respectively.

Proof: The statement follows by combining (8), (9), and (10). ∎

The upper bound (19) will be maximized numerically in what follows. The lower bound (18) is achievable by coding independently across the channels with an exponentially distributed input with intensity $\mathcal{E}_i^*$ for channel i, where $\mathcal{E}_i^*$ is the intensity allocation that maximizes the lower bound $\sum_{i \in \mathcal{N}} r(h_i, \mathcal{E}_i)$. The rest of this section deals with this maximization. Note that under perfect CSIT, this optimization can be performed by the transmitter.

$$b(h, \mathcal{E}, \beta, \delta) = \log \left[\frac{\beta e^{-\frac{\delta^2}{2}}}{\sqrt{2\pi e}} + \frac{\mathcal{Q}(\delta)}{\sqrt{e}} \right] + \frac{\mathcal{Q}(\delta)}{2} + \frac{\delta^2}{2} - \frac{\mathcal{Q}(\delta + h\mathcal{E})}{2} + \frac{\delta + h\mathcal{E}}{\beta} + \frac{e^{-\frac{\delta^2}{2}}}{\sqrt{2\pi}} \left[\frac{1}{\beta} + \frac{\delta}{2} \right]. \tag{6}$$

B. Intensity Allocation Given h

Consider N parallel channels, where the achievable rate over channel $i \in \mathcal{N}$ with optical intensity $\mathcal{E}_i$ is $\frac{1}{2}\log(1 + c_i^2 \mathcal{E}_i^2)$ for some c_i satisfying

$$c_1 \geq c_2 \geq \ldots \geq c_N. \tag{20}$$

This is e.g. the case in (18), where

$$c_i = h_i \sqrt{\frac{e}{2\pi}}. \tag{21}$$

The maximal achievable rate using this scheme[3] is given by the solution of

$$\max_{\mathcal{E}} \quad f(\mathcal{E}) \tag{22}$$
$$\text{s.t.} \quad \mathcal{E} \in \mathbb{R}_+^N, \ \|\mathcal{E}\|_1 \leq \mathcal{E},$$

where[4]

$$f(\mathcal{E}) = \frac{1}{2} \sum_{i \in \mathcal{N}} \log \left(1 + c_i^2 \mathcal{E}_i^2\right). \tag{23}$$

Note that $f(\mathcal{E})$ is nonconcave in $\mathcal{E}_i$, which renders the problem different from the standard water-filling power-allocation problem [36, Sec. 9.4]. Nevertheless, the optimal solution of (22) necessarily satisfies the KKT conditions [39]. Those conditions can thus be used to simplify problem (22).

The Lagrangian associated with the problem is given by

$$L = -f(\mathcal{E}) + \lambda \left(\|\mathcal{E}\|_1 - \mathcal{E} \right) - \sum_{i \in \mathcal{N}} \lambda_i \mathcal{E}_i, \tag{24}$$

and the corresponding KKT conditions are

$$\mathcal{E}_i, \lambda, \lambda_i \geq 0, \quad \lambda_i \mathcal{E}_i = 0, \quad \frac{\partial L}{\partial \mathcal{E}_i} = 0, \ \forall i \in \mathcal{N}$$
$$\|\mathcal{E}\|_1 \leq \mathcal{E}, \quad \lambda (\|\mathcal{E}\|_1 - \mathcal{E}) = 0. \tag{25}$$

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

$$\underbrace{\left(\lambda - \frac{c_i^2 \mathcal{E}_i}{1 + c_i^2 \mathcal{E}_i^2} \right)}_{=\lambda_i} \mathcal{E}_i = 0. \tag{26}$$

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

$$\mathcal{E}_i(\lambda) \in \begin{cases} \{0, \mathcal{E}_i^+(\lambda), \mathcal{E}_i^-(\lambda)\}, & \lambda \leq \frac{c_i}{2} \\ \{0\}, & \text{otherwise}, \end{cases} \tag{27}$$

where

$$\mathcal{E}_i^+(\lambda) = \frac{1}{2\lambda} + \sqrt{\frac{1}{4\lambda^2} - \frac{1}{c_i^2}}, \tag{28}$$

$$\mathcal{E}_i^-(\lambda) = \frac{1}{2\lambda} - \sqrt{\frac{1}{4\lambda^2} - \frac{1}{c_i^2}}. \tag{29}$$

The optimal solution of problem (22) belongs to the set of candidate solutions given by $\mathcal{E}(\lambda) = (\mathcal{E}_1(\lambda), \ldots, \mathcal{E}_N(\lambda))$ which satisfy the intensity constraint with equality, i.e., $\|\mathcal{E}(\lambda)\|_1 = \mathcal{E}$. This set contains multiple candidates (up to 3^N of them according to (27)) corresponding to different values of λ, the best of which is the one which yields the highest rate.

Since $\mathcal{E}_i(\lambda) = 0$ if $\lambda > \frac{c_i}{2}$, to obtain the set of candidate solutions, one can do the following for each $i \in \mathcal{N}$:

1) For all $\mathcal{E}(\lambda) = (\mathcal{E}_1(\lambda), \ldots, \mathcal{E}_i(\lambda), 0, \ldots, 0)$ with $\mathcal{E}_j(\lambda) \in \{0, \mathcal{E}_j^+(\lambda), \mathcal{E}_j^-(\lambda)\}$, $j = 1, \ldots, i$, $(3^i$ possibilities), find $\lambda^* \in \left(\frac{c_{i+1}}{2}, \frac{c_i}{2} \right]$ so that[5] $\|\mathcal{E}(\lambda^*)\|_1 = \mathcal{E}$.
2) If such λ^* exists, include the allocation $\mathcal{E}(\lambda^*)$ in the set of candidates.

Finally, among the set of candidates, the optimal solution is the one which maximizes the objective function.

1) Example: Consider $N = 2$. For $i = 1$, we consider $\lambda \in \left(\frac{c_2}{2}, \frac{c_1}{2} \right]$, and we consider the following possibilities[6] $\mathcal{T}_1 = \{(0,0), (+, 0), (-, 0)\}$. We search for the ones having an ℓ_1-norm equal to $\mathcal{E}$, for some $\lambda \in \left(\frac{c_2}{2}, \frac{c_1}{2} \right]$. 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 consider $\lambda \in \left(0, \frac{c_2}{2} \right]$, and the possibilities $\mathcal{T}_1 \cup \mathcal{T}_2$ where

$$\mathcal{T}_2 = \{(0, +), (0, -), (+, +), (-, +), (+, -), (-, -)\}.$$

For each $\mathcal{E}(\lambda) \in \mathcal{T}_1 \cup \mathcal{T}_2$, we solve $\|\mathcal{E}(\lambda)\|_1 = \mathcal{E}$ for λ. If a solution λ^* exists in $\left(0, \frac{c_2}{2} \right]$, we include $\mathcal{E}(\lambda^*)$ in the set $\mathcal{U}$.

Alternatively, we can plug $\mathcal{E}(\lambda)$ in the Lagrangian L (24) (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 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 $\|\mathcal{E}(\lambda)\|_1 = \mathcal{E}$, we add $\mathcal{E}(\lambda)$ to the list of candidates.

Finally, the optimal allocation of problem (22) is obtained as $\arg\max_{\mathcal{E} \in \mathcal{U}} f(\mathcal{E})$. Next, we study the behavior of the optimal solution at high, low, and moderate SNR.

---

[3]Keep in mind that this is an achievable rate. The obtained solution after optimization is not necessarily the capacity of the channel, but rather a capacity lower bound.

[4]A similar problem was treated recently in [41], however, without giving the general solution which is provided here.

[5]We formally define $c_{N+1} = 0$.

[6]For brevity, occasionally we will use '+' and '−' to denote $\mathcal{E}_1^+(\lambda)$ and $\mathcal{E}_1^-(\lambda)$, respectively.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCOMM.2016.2621743, IEEE Transactions on Communications

6

2) High SNR: At high SNR so that $c_i^2 \mathcal{E}^2 \gg 1 \; \forall i \in \mathcal{N}$, (26) becomes

$$\left(\lambda - \frac{1}{\mathcal{E}_i}\right)\mathcal{E}_i = 0. \tag{30}$$

This leads to either $\mathcal{E}_i = 0$ (inactive channel) or $\mathcal{E}_i = \frac{1}{\lambda}$ (active channel). Next, we show that as $\mathcal{E}$ grows, the solution of (22) allocates intensity equally to all channels.

Proposition 1: The solution of problem (22) converges to $\mathcal{E}_i = \frac{\mathcal{E}}{N} \; \forall i \in \mathcal{N}$ as $\mathcal{E} \to \infty$.

Proof: Consider an allocation where channels $\mathcal{N}_a = \{1, \ldots, N_a\}$ are activated ($N > N_a > 1$) and the rest are not. At high SNR, active channels have equal intensity (30), and hence $\mathcal{E}_i = \frac{\mathcal{E}}{N_a}$ for $i \in \mathcal{N}_a$ and $\mathcal{E}_i = 0$ otherwise. Call this allocation $\mathcal{E}_a$. We compare this allocation with $\mathcal{E}_i(\lambda) = \frac{1}{\lambda} = \frac{\mathcal{E}}{N}$ for all $i \in \mathcal{N}$, which satisfies $\|\mathcal{E}(\lambda)\|_1 = \mathcal{E}$. The achievable rate of the latter satisfies

$$f\left(\frac{\mathcal{E}}{N}, \ldots, \frac{\mathcal{E}}{N}\right)$$

$$= \frac{1}{2}\sum_{i \in \mathcal{N}}\log\left(1 + c_i^2 \frac{\mathcal{E}^2}{N^2}\right) \tag{31}$$

$$\geq \frac{1}{2}\sum_{i \in \mathcal{N}}\log\left(1 + c_i^2 \frac{\mathcal{E}^2}{N_a^2}\right) + \frac{N}{2}\log\left(\frac{N_a^2}{N^2}\right) \tag{32}$$

$$= f(\mathcal{E}_a) + \frac{1}{2}\sum_{i \in \mathcal{N}\setminus\mathcal{N}_a}\log\left(1 + c_i^2 \frac{\mathcal{E}^2}{N_a^2}\right) + \frac{N}{2}\log\left(\frac{N_a^2}{N^2}\right) \tag{33}$$

This lower bound on $f\left(\frac{\mathcal{E}}{N}, \ldots, \frac{\mathcal{E}}{N}\right)$ eventually becomes larger than $f(\mathcal{E}_a)$ as $\mathcal{E}$ increases. This proves the proposition. ∎

3) Low SNR: At low SNR so that $c_i^2 \mathcal{E}^2 \ll 1 \; \forall i \in \mathcal{N}$, (26) 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 $\mathcal{E}(\lambda)$ with $\mathcal{E}_i(\lambda) \in \{0, \frac{\lambda}{c_i^2}\} \; \forall i \in \mathcal{N}$, from which the best $\mathcal{E}(\lambda)$ has to be chosen. In this case, we have the following statement.

Proposition 2: The solution of problem (22) converges to $\mathcal{E}_1 = \mathcal{E}$ and $\mathcal{E}_i = 0$ for $i = 2, \ldots, N$ as $\mathcal{E} \to 0$.

Proof: At low SNR we have

$$\frac{1}{2}\sum_{i \in \mathcal{N}}\log(1 + c_i^2 \mathcal{E}_i^2) \leq \frac{N}{2}\log\left(1 + \frac{1}{N}\sum_{i \in \mathcal{N}}c_i^2 \mathcal{E}_i^2\right) \tag{34}$$
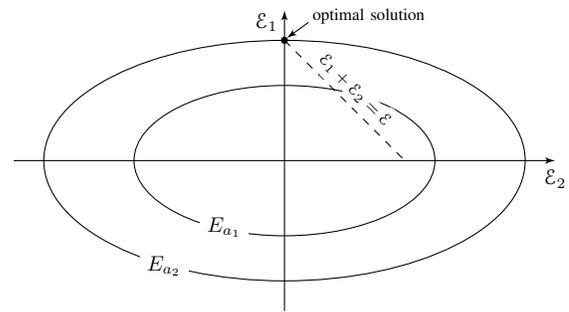
$$\leq \frac{N}{2}\log\left(1 + \frac{c_1^2}{N}\sum_{i \in \mathcal{N}}\mathcal{E}_i^2\right) \tag{35}$$

$$\leq \frac{N}{2}\log\left(1 + \frac{c_1^2}{N}\left(\sum_{i \in \mathcal{N}}\mathcal{E}_i\right)^2\right) \tag{36}$$

$$\leq \frac{N}{2}\log\left(1 + \frac{c_1^2}{N}\mathcal{E}^2\right) \tag{37}$$

$$\leq \frac{c_1^2 \mathcal{E}^2}{2}, \tag{38}$$

where the inequalities follow using Jensen's inequality, $c_i \leq c_1$, $\sum \mathcal{E}_i^2 \leq (\sum \mathcal{E}_i)^2$, $\sum_{i \in \mathcal{N}} \mathcal{E}_i \leq \mathcal{E}$, and $\log(1+x) \leq x$, respectively. This upper bound is achievable at low SNR by setting $\mathcal{E}_1 = \mathcal{E}$, 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, which concludes the proof. ∎

(a) For $N = 2$, the level-sets of the low-SNR objective function are ellipses E_a. The constraints $\mathcal{E} = (\mathcal{E}_1, \mathcal{E}_2) \in \mathbb{R}_+^2$ and $\|\mathcal{E}\|_1 = \mathcal{E}$ define a line segment L. The optimal solution lies at the intersection of L with the largest E_a, i.e., at $\mathcal{E} = (\mathcal{E}, 0)$.
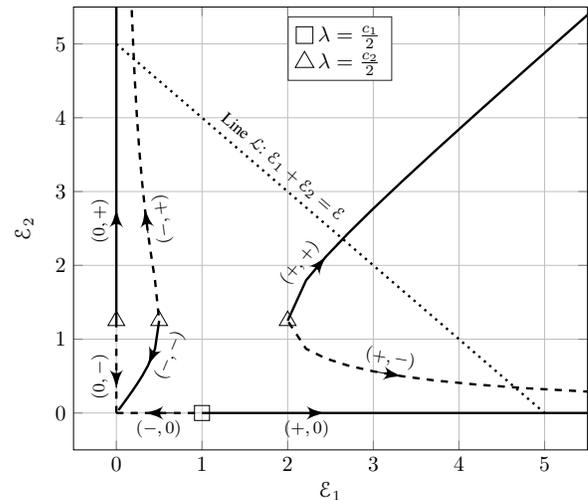
(b) Trajectories of candidate solutions of (22) as λ decreases. Here, we have chosen $\mathcal{E} = 5$. For clarity, we denote $\mathcal{E}_i^+(\lambda)$ by $+$ and $\mathcal{E}_i^-(\lambda)$ by $-$.

Fig. 2: A graphical illustration of solutions at low and moderate SNR.

This solution can also be demonstrated geometrically as shown in Fig. 2a. At low SNR, the objective function of (22) becomes $\sum_{i \in \mathcal{N}} \frac{c_i^2 \mathcal{E}_i^2}{2}$ 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 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. Clearly, a larger a leads to a larger ellipsoid. On the other hand, the optimal power allocation must lie on the hyperplane described by $\mathcal{E} \in \mathbb{R}_+^N$ and $\|\mathcal{E}\|_1 = \mathcal{E}$. 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, leading to the optimality of $\mathcal{E} = (\mathcal{E}, 0, \ldots, 0)$ at low SNR.

4) Moderate SNR: Unfortunately, while the solutions at high and low SNR are simple, the solution at moderate SNR is not. In general, multiple channels have to be activated according to the general solution (27). 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 allocations among $(+, 0)$, $(-, 0)$, $(0, +)$, $(0, -)$, $(+, +)$, $(+, -)$, $(-, +)$, and $(-, -)$, where we denoted $\mathcal{E}_1^+(\lambda)$ and $\mathcal{E}_1^-(\lambda)$. Recall that an allocation is feasible if there exist $\lambda \geq 0$ for which this allocation satisfies the total intensity constraint with equality.

The trajectories of those candidates are plotted in Fig. 2b as a function of λ for $(c_1, c_2) = (1, 0.8)$. Since the objective function is increasing in $\mathcal{E}_i$, the optimal solution must satisfy the total intensity constraint with equality. Hence, an allocation is optimal if it lies at the intersection of one of the trajectories with the line $\mathcal{L}$ defined by $\mathcal{E}_1 + \mathcal{E}_2 = \mathcal{E}$. There might be multiple such intersections. Next, we proceed to reduce the number of possibilities.

Consider the square at $(\mathcal{E}_1, \mathcal{E}_2) = (1, 0)$ in Fig. 2b, corresponding to $\lambda = \frac{c_1}{2}$ where $\mathcal{E}_1 = \mathcal{E}_1^+(\lambda) = \mathcal{E}_1^-(\lambda) = \frac{1}{2\lambda} = \frac{1}{c_1}$ 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 $(+, 0)$ intersects with $\mathcal{L}$, otherwise, then $(-, 0)$ does. Note however that the latter case corresponds to $c_1 \mathcal{E} \leq 1$ which is close to low SNR. Since we are focusing on moderate SNR here, we omit the candidate $(-, 0)$. Thus, at this point, our list of candidates contains $(+, 0)$ only.

At $\lambda = \frac{c_2}{2}$, channel 2 becomes active and we reach the points marked by triangles in Fig. 2b. Now, we have the additional possibilities $(0, +)$, $(0, -)$, $(+, +)$, $(+, -)$, $(-, +)$, and $(-, -)$. Since $(+, -)$ and $(-, +)$ converge respectively to $(+, 0)$ and $(0, +)$ as λ decreases, we omit these two possibilities. On the other hand, since the objective function evaluated at $(0, +)$ and $(0, -)$ for a given λ is smaller than that at $(+, 0)$ and $(-, 0)$ respectively, we omit these two possibilities too. Finally, since $(-, -)$ satisfies $\mathcal{E}_1^-(\lambda) + \mathcal{E}_2^-(\lambda) \leq \frac{2}{c_2}$ which approaches $(0, 0)$ as λ decreases, then this trajectory intersects $\mathcal{L}$ if $c_2 \mathcal{E} \leq 2$ which is close to low SNR. Thus, we omit this candidate.

The remaining candidates are $(+, 0)$ and $(+, +)$. We choose the one which yields larger $f(\mathcal{E})$. One should keep in mind that we have to select λ so that the resulting allocation $\mathcal{E}$ satisfies $\|\mathcal{E}\|_1 = \mathcal{E}$.

Based on this, define

$$\boldsymbol{A}_i(\lambda) = (A_{i1}(\lambda), A_{i2}(\lambda), \ldots, A_{iN}(\lambda)), \qquad (39)$$

where $A_{ij}(\lambda) = \mathcal{E}_i^+(\lambda)$ given in (28) for $j \leq i$ and $A_{ij}(\lambda) = 0$ otherwise. This $\boldsymbol{A}_i(\lambda)$ captures trajectories of the form $(+, 0, 0, \ldots, 0)$, $(+, +, 0, 0, \ldots, 0)$, etc. Note that $\boldsymbol{A}_i(\lambda)$ is real if $\lambda \in (0, \frac{c_i}{2}]$. Algorithm 1 can be used for finding a good solution for (22). This algorithm jointly activates channels and allocates powers to them, and therefore, we call it the 'joint activation/allocation' (JA) algorithm. For this algorithm, we use the convention $\frac{1}{0} = \infty$ and allow λ to be 0.

Algorithm 1 activates components of $\boldsymbol{\mathcal{E}}$ successively, starting by setting $\mathcal{E}_1 = \mathcal{E}_1^+(\lambda)$, followed by setting $\mathcal{E}_2 = \mathcal{E}_2^+(\lambda)$ and so on. Each time, an optimal λ which leads to $\|\boldsymbol{\mathcal{E}}\|_1 = \mathcal{E}$ is found. As long as such λ exists, and as long as it leads to a higher achievable rate, the algorithm activates one more component of $\boldsymbol{\mathcal{E}}$. Following the graphical illustration of Fig.

Algorithm 1 JA - Joint activation/allocation for parallel IM-DD channels

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

2b, this process can be interpreted as starting with the solution at the intersection of $\|\boldsymbol{\mathcal{E}}\|_1 = \mathcal{E}$ with $(+, 0, 0, \ldots, 0)$, and comparing it with the one at the intersection with $(+, +, 0, 0, \ldots, 0)$. If the first solution is better, we stop the algorithm. If not, we compare the second solution with the intersection with $(+, +, +, 0, 0, \ldots, 0)$, and so on. 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 5 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 5, we can split the algorithm into two stages. First we find the channels that should be activated, then we find λ^* leading to $\|\boldsymbol{\mathcal{E}}\|_1 = \mathcal{E}$. To this end, we define $\boldsymbol{B}(\lambda) = (B_1(\lambda), \ldots, B_N(\lambda))$ where

$$B_i(\lambda) = \begin{cases} \mathcal{E}_i^+(\lambda), & \lambda \leq \frac{c_i}{2} \\ 0, & \text{otherwise.} \end{cases} \qquad (40)$$

Note that $\|\boldsymbol{B}(\lambda)\|_1$ increases as λ decreases. A good solution for (22) can thus be found by evaluating $\boldsymbol{B}(\lambda^{[i]})$ at $\lambda^{[i]} = \frac{c_i}{2}$, $i = 1, \ldots, N$, which is a decreasing sequence in i. Once $\|\boldsymbol{B}(\lambda^{[i]})\|_1$ becomes greater than $\mathcal{E}$, we stop, and $i - 1$ determines the number of channels to be activated. Then, given those activated channels, we obtain the intensity allocation by finding $\lambda \in \left[0 \frac{c_{i-1}}{2}\right]$ which yields $\|\boldsymbol{B}(\lambda)\|_1 = \mathcal{E}$. This procedure is described in Algorithm 2. We call this algorithm the 'successive activation/allocation' (SA) algorithm. Similar to the JA algorithm, we allow $\lambda = 0$, and use $\frac{1}{0} = \infty$.

Steps 6 to 8 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 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 9 finds λ^* for which the constraint is satisfied with equality. Similar to the JA 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 JA algorithm, step 9 (step 5 in Algorithm 1) is executed only once.

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

---

Algorithm 2 SA - Successive activation/allocation for parallel IM-DD channels

1: **function** SA$(c_1, \ldots, c_{N+1}, \mathcal{E})$
2:     $\mathcal{E}_h \leftarrow \boldsymbol{B}(\infty); \mathcal{E}_l \leftarrow \boldsymbol{B}(\frac{c_1}{2}); i \leftarrow 1;$
3:     **if** $\|\mathcal{E}_l\|_1 > \mathcal{E}$ **then**
4:         $\mathcal{E} \leftarrow (\mathcal{E}, 0, \ldots, 0)$
5:     **else**
6:         **while** $(\|\mathcal{E}_h\|_1 - \mathcal{E})(\|\mathcal{E}_l\|_1 - \mathcal{E}) \geq 0$
                $\wedge\ i < N + 1$ **do**
7:             $i \leftarrow i + 1; \mathcal{E}_h \leftarrow \mathcal{E}_l; \mathcal{E}_l \leftarrow \boldsymbol{B}(\frac{c_i}{2});$
8:         **end while**
9:         Find $\lambda^* \in \left[0, \frac{c_{i-1}}{2}\right]$ so that $\|\boldsymbol{B}(\lambda^*)\|_1 = \mathcal{E}$
10:       $\mathcal{E} \leftarrow \boldsymbol{B}(\lambda^*)$
11:     **end if**
12:     **return** $\mathcal{E}$
13: **end function**

---

with respect to the JA 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 JA algorithm slightly superior to the SA algorithm.

Remark 1: At high SNR, step 6 in Algorithm 2 will always be true. Thus, the algorithm will activate all channels. After activating all channels, step 9 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 SA algorithm is optimal at high SNR. This is also true for the JA algorithm, since it is more accurate.

C. Evaluation

1) Given $\boldsymbol{h}$: Fig. 3a shows an evaluation of the algorithms for a system with $N = 4$ channels and with $\boldsymbol{h} = [1, .7, .3, .1]$ and $c_i = h_i\sqrt{\frac{e}{2\pi}}$. Five achievable rates are plotted corresponding to an exponential input distribution and an intensity allocation using (i) exhaustive search where the optimal solution is found from the set of all 3^N possible candidate solutions ($\mathcal{E}_i \in \{0, \mathcal{E}_i^+, \mathcal{E}_i^-\}$ (27) as in Sec. IV-B1), (ii) JA algorithm, (iii) SA algorithm, (iv) equal allocation, and (v) activating the strongest channel only. Note that the achievable rates with JA and SA coincide with exhaustive search, and that the achievable rates with equal allocation and activating the strongest channel confirm the high and low SNR solutions given in Propositions 1 and 2, respectively.

Since the SA algorithm delivers a good solution and requires less computation, we propose to use it for intensity allocation, leading to the following proposition.

Proposition 3: The capacity $C(\boldsymbol{h}, \mathcal{E})$ is lower bounded by the achievable rate

$$\tilde{R}(\boldsymbol{h}, \mathcal{E}) = \sum_{i \in \mathcal{N}} r(h_i, \mathcal{E}_i), \qquad (41)$$

where $r(h, \mathcal{E})$ is defined in (8), and where $(\mathcal{E}_1, \ldots, \mathcal{E}_N)$ is the output of the SA algorithm with $c_i = h_i\sqrt{\frac{e}{2\pi}}$.

This proposition supplements Theorem 1 since it simplifies the lower bound $R(\boldsymbol{h}, \mathcal{E})$ (18).

Fig. 3b shows a comparison of the SA solution $\tilde{R}(\boldsymbol{h}, \mathcal{E})$ with the capacity upper bound $\overline{C}(\boldsymbol{h}, \mathcal{E})$. It can be observed

that the achievable rate and the upper bound approach each other as SNR increases, suggesting that using an exponential input with the SA intensity allocation is optimal at high SNR. In fact, this can be proved leading to the following corollary.

Corollary 1: The capacity $C(\boldsymbol{h}, \mathcal{E})$ satisfies

$$\lim_{\mathcal{E} \to \infty} \left[C(\boldsymbol{h}, \mathcal{E}) - \sum_{i \in \mathcal{N}} \frac{1}{2} \log\left(\frac{eh_i^2\mathcal{E}^2}{2\pi N^2}\right) \right] = 0. \qquad (42)$$

Proof: The summand $\overline{r}(h, \mathcal{E})$ in the upper bound (19) converges to $\frac{1}{2} \log\left(\frac{eh^2\mathcal{E}^2}{2\pi}\right)$ for large $\mathcal{E}$. This leads to

$$\overline{C}(\boldsymbol{h}, \mathcal{E}) = \max_{\mathcal{E} \in \mathcal{S}} \sum_{i \in \mathcal{N}} \frac{1}{2} \log(c_i^2\mathcal{E}_i^2) + \epsilon_{\mathcal{E}} \qquad (43)$$

$$= \sum_{i \in \mathcal{N}} \frac{1}{2} \log\left(\frac{c_i^2\mathcal{E}^2}{N^2}\right) + \epsilon_{\mathcal{E}}, \qquad (44)$$

where $\epsilon_{\mathcal{E}} \to 0$ as $\mathcal{E} \to \infty$, and where the last step follows since the maximization in (43) is similar to (22), and hence has a similar solution, which is $\mathcal{E}_i = \frac{\mathcal{E}}{N}$ at high SNR. By using $c_i = h_i\sqrt{\frac{e}{2\pi}}$, we obtain the statement of the corollary. ∎

Defining the 'pre-log' of the channel as $\lim_{\mathcal{E} \to \infty} \frac{C(\boldsymbol{h}, \mathcal{E})}{\log(\mathcal{E})}$, Corollary 1 shows that this pre-log is N. This pre-log defines the capacity scaling law at high SNR; capacity scales as $N\log(\mathcal{E})$ and hence, doubling $\mathcal{E}$ increases capacity by $N\log(2)$ nats (N bits).

2) Block-fading $\boldsymbol{h}$: In Fig. 3a, the algorithm finds the optimal solution. One can expect that the algorithms lead to a suboptimal solution in some cases. In particular, the algorithms ignores allocations involving $\mathcal{E}_i^-$ for some i, although those might be optimal. However, excluding those solutions only slightly reduces the rate, as per the discussion on Fig. 2b. To examine this aspect, it is instructive to evaluate the average performance for block fading channels.

To this end, we consider a block fading free-space optical (FSO) system with $N = 4$ in clear and foggy weather. That is, we assume that the channel maintains the same value during a transmission block, and changes to an independent channel realization in the next block. We use the statistics given in [12] combining path loss, turbulence, and pointing errors. Thus,

$$h_i = \rho h_{i,\ell} h_{i,\mathsf{a}} h_{i,\mathsf{p}}, \qquad (45)$$

where ρ is the detector responsivity, $h_{i,\ell}$ is the path loss, $h_{i,\mathsf{a}}$ is the atmospheric turbulence, and $h_{i,\mathsf{p}}$ accounts for the pointing error. The atmospheric turbulence is assumed log-normal, which models weak turbulence conditions [18]. Thus, $h_{i,\mathsf{a}}$ is distributed according to

$$f_{\mathsf{a}}(h_{i,\mathsf{a}}) = \frac{1}{h_{i,\mathsf{a}}\sqrt{2\pi\sigma_R^2}} \exp\left(-\frac{(\log(h_{i,\mathsf{a}}) + \frac{\sigma_R^2}{2})^2}{2\sigma_R^2}\right), \quad (46)$$

where σ_R^2 is the Rytov variance. The pointing error contribution is given by $h_{i,\mathsf{p}} = A_0 \exp\left(-\frac{2r^2}{w_{z_{eq}}^2}\right)$, where r is the distance from the beam center to the receiver aperture center, $A_0 = (\mathrm{erf}(v))^2$, $w_{z_{eq}}^2 = w_z^2 \frac{\sqrt{\pi}\mathrm{erf}(v)}{2v\exp(-v^2)}$, $v = \frac{\sqrt{\pi}a}{\sqrt{2}w_z}$, a is the receiver aperture radius, and w_z is the beam radius at the
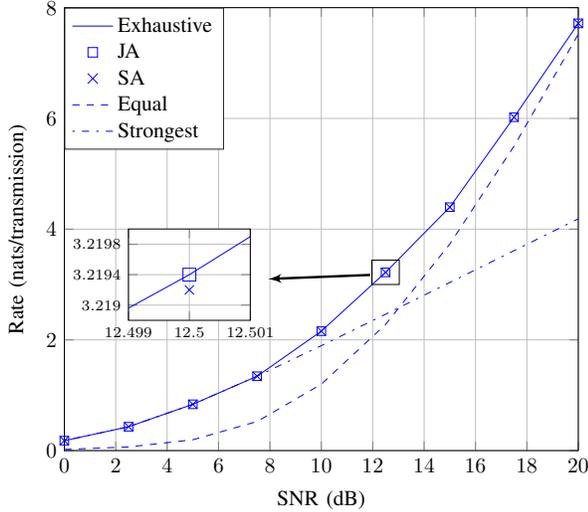
(a) Performance compared to exhaustive search.
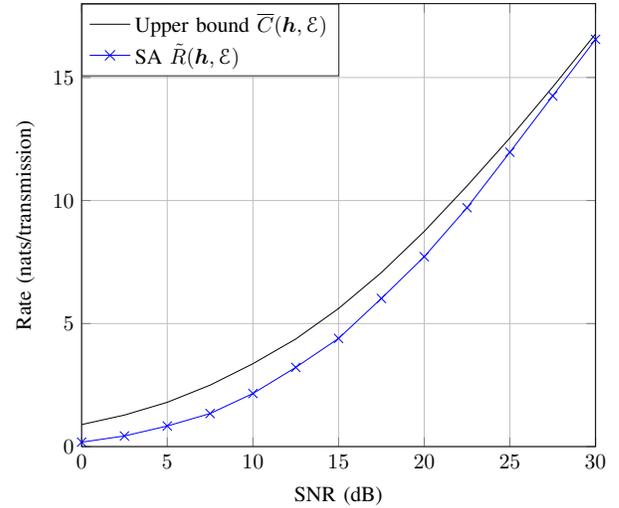
(b) Performance compared to the upper bound.

Fig. 3: Performance as a function of SNR ($\mathcal{E}/\sigma$) for a system with 4 channels and $\boldsymbol{h} = [1, .7, .3, .1]$ with intensity allocation by either exhaustive search, the JA algorithm, the SA algorithm, equal allocation, or activating the strongest channel.

Parameter	Symbol	Value	
		Clear	Light fog
Path-loss (clear weather)	$h_{i,\ell}$	0.9	0.008
Rytov variance	σ_R^2	1	0.1
Detector responsivity	ρ	0.5	
Beam radius at 1km	w_z	2.5 m	
Receiver aperture radius	a	10 cm	
Jitter variance	σ_s^2	0.09 m^2	

TABLE III: FSO System Paramters

receiver. The distance r follows a Rayleigh distribution [42], [43], hence

$$f_{\mathsf{p}}(r) = \frac{r}{\sigma_s^2} \exp\left(\frac{-r^2}{2\sigma_s^2}\right) \qquad (47)$$

with σ_s^2 being the jitter variance at the receiver. The parameters used for the numerical evaluation are taken from [12] for a transmitter-receiver distance of 1km, and are given in Table III. The transmitter/receiver apertures are assumed to be sufficiently separated so that inter-channel interference is negligible.[7]

The achievable rate using exponentially distributed input, i.e., $c_i = h_i\sqrt{\frac{e}{2\pi}}$, averaged over 10^3 channel realization is shown in Fig. 4 versus SNR.[8] The rates achieved using the JA and SA algorithms almost coincide with the exhaustive search solution, which indicates that the proposed algorithms almost always find the optimal solution (see [1, Table I]).

---

[7]If inter-channel interference is strong, a MIMO channel should be treated instead. The resulting MIMO channel can still be 'parallelized' (possibly reducing capacity) using standard MIMO detection techniques such as channel inversion, matched filtering, MMSE receivers, with or without successive interference cancellation [38]. The proposed intensity allocation can then be applied over the obtained parallel channels.

[8]Since we consider a block fading channel, this average rate is achievable by coding independently over each fading block.

V. QUASI-STATIC FADING PARALLEL CHANNELS WITHOUT CSIT

Under quasi-static fading, if no CSIT is available and the transmitter sends at rate $R_0 > 0$ and a total average intensity $\mathcal{E}$ per block, there is nonzero probability that $R_0 > C(\boldsymbol{h}, \mathcal{E})$, and hence reliable communication can not be guaranteed. In this case, the notion of outage becomes important [38]. For a given target rate R_0, a total average intensity per block $\mathcal{E}$, and a given transmit strategy defined by an input distribution $p(x_1, \ldots, x_N)$ in $\mathcal{P}$ (defined below (12)), the outage probability is defined as $\mathbb{P}\{I(X_1, \ldots, X_N; Y_1, \ldots, Y_N) < R_0\}$. The minimum outage probability is thus

$$P_o(\mathcal{E}, R_0) \qquad (48)$$
$$= \inf_{p(x_1, \ldots, x_N) \in \mathcal{P}} \mathbb{P}\{I(X_1, \ldots, X_N; Y_1, \ldots, Y_N) < R_0\}.$$

An upper bound on $P_o(\mathcal{E}, R_0)$ can be obtained by fixing $p(x_1, \ldots, x_N)$. One convenient choice is to choose $(X_1, \ldots, X_N)$ independent exponentially distributed, with equal optical intensity $\mathcal{E}/N$. This leads to

$$P_o(\mathcal{E}, R_0) \le \mathbb{P}\{R'(\boldsymbol{h}, \mathcal{E}) < R_0\}, \qquad (49)$$

where

$$R'(\boldsymbol{h}, \mathcal{E}) = \sum_{i \in \mathcal{N}} r\left(h_i, \frac{\mathcal{E}}{N}\right), \qquad (50)$$

and $r(h, \mathcal{E})$ is defined in (8).

On the other hand, $P_o(\mathcal{E}, R_0)$ is lower bounded by

$$P_o(\mathcal{E}, R_0) \ge \mathbb{P}\{\overline{C}(\boldsymbol{h}, \mathcal{E}) < R_0\}, \qquad (51)$$

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, $I(X_1, \ldots, X_N; Y_1, \ldots, Y_N) \le \overline{C}(\boldsymbol{h}, \mathcal{E})$, leading to $\mathbb{P}\{I(X_1, \ldots, X_N; Y_1, \ldots, Y_N) < R_0\} \ge \mathbb{P}\{\overline{C}(\boldsymbol{h}, \mathcal{E}) < R_0\}$ for any distribution $p(x_1, \ldots, x_N) \in \mathcal{P}$.
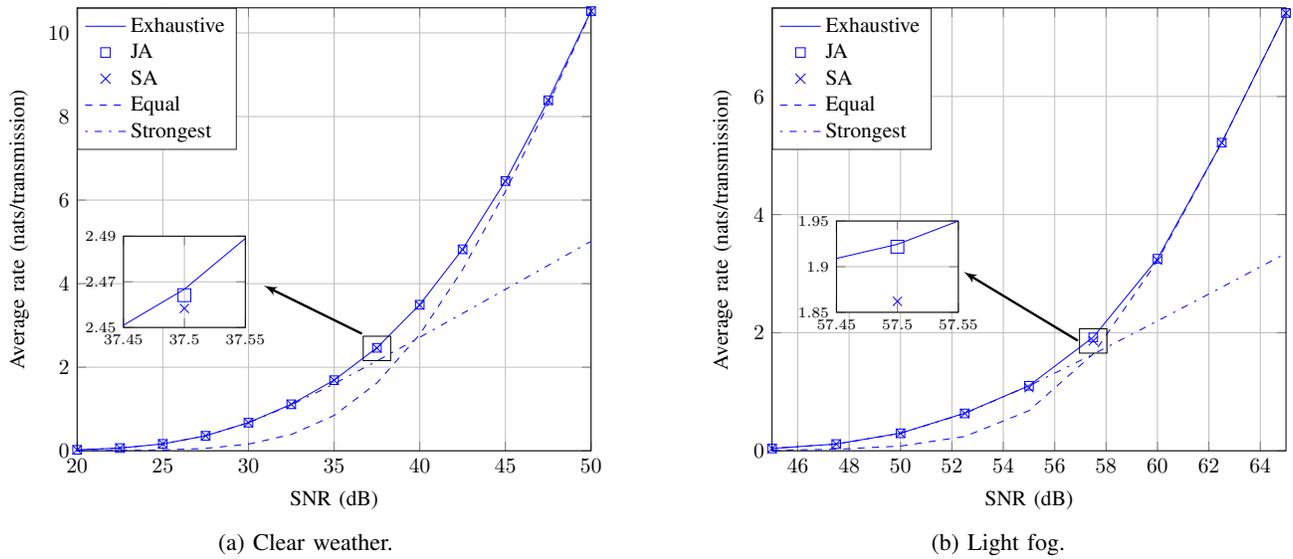
(a) Clear weather.

(b) Light fog.

Fig. 4: The achievable rate as a function of SNR ($\mathcal{E}/\sigma$) for a block fading FSO system with $N = 4$ channels as in (45).

The bounds (49) and (51) do not coincide in general. However, they coincide at high SNR, since 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 (43)–(44). Consequently, (49) and (51) are tight in this case, leading to the following statement.

Corollary 2: The minimum outage probability $P_o(\mathcal{E}, R_0)$ satisfies

$$\lim_{\mathcal{E} \to \infty} \left[P_o(\mathcal{E}, R_0) - \overline{P}_o(\mathcal{E}, R_0) \right] = 0, \qquad (52)$$

where

$$\overline{P}_o(\mathcal{E}, R_0) = \mathbb{P}\left\{ \sum_{i \in \mathcal{N}} \frac{1}{2} \log\left(\frac{eh_i^2 \mathcal{E}^2}{2\pi N^2} \right) < R_0 \right\}. \qquad (53)$$

Thus, reliable communication at rate R_0 with outage probability $\overline{P}_o(\mathcal{E}, R_0)$ is possible at high SNR if (53) is satisfied. Furthermore, there is no scheme with rate R_0 which can achieve lower outage probability than $\overline{P}_o(\mathcal{E}, R_0)$ at high SNR. This leads to the high-SNR ϵ-outage capacity C_ϵ, defined as the highest achievable rate at an outage probability $\leq \epsilon$.

Corollary 3: The ϵ-outage capacity C_ϵ of a system of N IM-DD parallel channels with total intensity constraint $\mathcal{E}$ converges at high SNR to

$$R_\epsilon = \max\left\{ R \in \mathbb{R}_+ \,\middle|\, \overline{P}_o(\mathcal{E}, R) \leq \epsilon \right\}. \qquad (54)$$

Fig. 5 shows ϵ-outage capacity bounds versus SNR for an FSO channel with $N = 2$ and the channel statistics given in (45) (clear weather). The upper bound is obtained by finding the largest R so that $\mathbb{P}\left\{ \overline{C}(\boldsymbol{h}, \mathcal{E}) < R \right\} \leq \epsilon$. The lower bounds are obtained similarly, with $\overline{C}(\boldsymbol{h}, \mathcal{E})$ replaced by $\sum_{i \in \mathcal{N}} r(h_i, \mathcal{E}_i)$ with $\mathcal{E}_i$ obtained using the SA algorithm or using equal intensity allocation. This latter is equivalent to R_ϵ given in Corollary 3 at high SNR. The bounds coincide at high SNR which confirms Corollary 3. By comparing with the rate bounds for the perfect CSIT case, we can see that in the absence of CSIT, achieving a rate of 8 nats/transmission e.g. requires $\approx$ 3.5dB more SNR for $\epsilon = 0.1$, and $\approx$ 7dB more
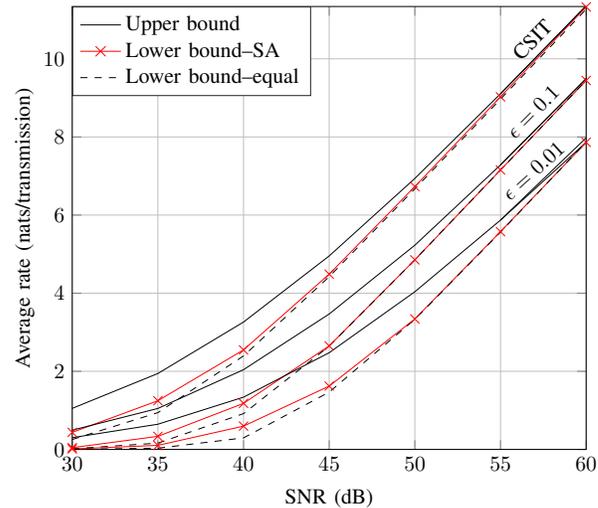

Fig. 5: Outage capacity bounds (outage probability ϵ) as a function of SNR ($\mathcal{E}/\sigma$) for a block fading FSO system with $N = 2$ and channels as in (45) under clear weather. Note that the lower bound with equal intensity allocation coincides with R_ϵ in (54).

SNR for $\epsilon = 0.01$. It is observed numerically, that this SNR cost decreases as N increases. In fact, it can be shown that this cost vanishes as $N \to \infty$. In this case, it holds that

$$\overline{P}_o(\mathcal{E}, R_0) = \mathbb{P}\left\{ \sum_{i \in \mathcal{N}} \frac{1}{2} \log\left(\frac{eh_i^2 \mathcal{E}^2}{2\pi N^2} \right) < R_0 \right\} \qquad (55)$$

$$= \mathbb{P}\left\{ N\frac{1}{N} \sum_{i \in \mathcal{N}} \frac{1}{2} \log\left(\frac{eh_i^2 \mathcal{E}^2}{2\pi N^2} \right) < R_0 \right\} \qquad (56)$$

$$\approx \mathbb{P}\left\{ N\mathbb{E}_h\left[\frac{1}{2} \log\left(\frac{eh^2 \mathcal{E}^2}{2\pi N^2} \right) \right] < R_0 \right\}, \qquad (57)$$

where the approximation becomes tight as N increases. Thus, by Corollary 3, the outage capacity approaches

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCOMM.2016.2621743, IEEE Transactions on Communications

11

$N\mathbb{E}_h\left[\frac{1}{2}\log\left(\frac{eh^2\mathcal{E}^2}{2\pi N^2}\right)\right]$ as $N \to \infty$, which is nothing but the high SNR capacity of the system with CSIT and with large N by Corollary 1. This can be interpreted as a diversity gain.

VI. IMPACT OF PEAK CONSTRAINTS

Here, we discuss the case with a peak intensity constraint $X_i \leq \mathcal{A}$ 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 and lower bounds on the capacity $C_{\mathcal{A}}(h, \mathcal{E})$ in this case were given in [32]–[34]. For instance, the following achievable rate (capacity lower bound) was given in [32]

$$C_{\mathcal{A}}(h, \mathcal{E}) \geq r_1(h, \mathcal{E}, \mathcal{A}) \tag{58}$$

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

where μ^* is the solution of $\frac{1}{\mu^*} - \frac{e^{-\mu^*}}{1-e^{-\mu^*}} = \alpha$. Here, we use a subscript 1 to indicate that this is a lower bound by Lapidoth *et al.* in [32].

Another lower bound was given in [34] achievable if X follows a truncated-Gaussian distribution

$$\tilde{g}_{\mu,\nu}(x) = \eta g_{\mu,\nu}(x) \tag{59}$$

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

$$\tilde{\mu} = \nu^2[\tilde{g}_{\mu,\nu}(0) - \tilde{g}_{\mu,\nu}(\mathcal{A})] + \mu \tag{60}$$

$$\tilde{\nu}^2 = \nu^2\left[1 - \mathcal{A}\tilde{g}_{\mu,\nu}(\mathcal{A}) - \tilde{\mu}\left(\tilde{g}_{\mu,\nu}(0) - \tilde{g}_{\mu,\nu}(\mathcal{A})\right)\right]. \tag{61}$$

For given μ and ν so that $\tilde{\mu} \leq \mathcal{E}$, the rate

$$\tilde{r}_{\text{t}}(h, \mathcal{A}, \mu, \nu) = \frac{1}{2}\log\left(\frac{\nu^2}{\tilde{\nu}^2} + h^2\nu^2\right) - \phi, \tag{62}$$

(subscript t for truncated-Gaussian) is achievable, 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,

$$C_{\mathcal{A}}(h, \mathcal{E}) \geq r_{\text{t}}(h, \mathcal{E}, \mathcal{A}) \triangleq \max_{\substack{\mu,\nu \\ \tilde{\mu} \leq \mathcal{E}}} \tilde{r}_{\text{t}}(h, \mathcal{A}, \mu, \nu). \tag{63}$$

As shown in [34], $r_{\text{t}}(h, \mathcal{E}, \mathcal{A})$ is close to capacity at high SNR. Furthermore, the rate $r_{\text{t}}(h, \mathcal{E}, \mathcal{A})$ satisfies $r_{\text{t}}(h, \mathcal{E}, \mathcal{A}) \geq r_{\text{ts}}(h, \mathcal{E}, \mathcal{A})$ [44] where the simplified truncated-Gaussian lower bound $r_{\text{ts}}(h, \mathcal{E}, \mathcal{A})$ is given by

$$r_{\text{ts}}(h, \mathcal{E}, \mathcal{A}) = \frac{1}{2}\log\left(1 + h^2\min\left\{\frac{\mathcal{E}^2}{9}, \frac{\mathcal{A}^2}{36}\right\}\right), \tag{64}$$

(subscript ts for simplified truncated-Gaussian) by choosing $\tilde{\mu} = \min\{\mathcal{E}, \mathcal{A}/2\}$ and $\nu = \frac{\mu}{3}$ and plugging in (62). Although this simplification sacrifices the near optimality of the achievable rate at high SNR, the simplified rate captures the scaling behavior of the channel capacity at high SNR.

Note that the capacity upper bound $\bar{r}(h, \mathcal{E})$ given in (9) also holds in this case. Another upper bound was given in [34] as

$$\bar{\bar{r}}(h, \mathcal{A}) = \sup_{\delta \in [0,1]}\left[\frac{\delta}{2}\log\left(\frac{h^2\mathcal{A}^2}{2\pi e}\right) - \log\left((1-\delta)^{\frac{3(1-\delta)}{2}}\delta^\delta\right)\right]. \tag{65}$$

Therefore, we can write

$$C_{\mathcal{A}}(h, \mathcal{E}) \leq \min\{\bar{r}(h, \mathcal{E}), \bar{\bar{r}}(h, \mathcal{A})\}. \tag{66}$$

B. Parallel Channels with CSIT

1) Capacity Bounds Given $\boldsymbol{h}$: Capacity lower bounds for this case can be derived by using (58), (63), and (64) and optimizing the intensity allocation over the parallel channels. This is formally stated as follows.

Theorem 2: The capacity $C_{\mathcal{A}}(\boldsymbol{h}, \mathcal{E})$ is lower bounded by the achievable rates

$$R_{\text{a}}(\boldsymbol{h}, \mathcal{E}, \mathcal{A}) = \max_{\mathcal{E} \in \mathcal{S}_{\mathcal{A}}}\sum_{i \in \mathcal{N}} r_{\text{a}}(h_i, \mathcal{E}_i, \mathcal{A}), \quad \text{a} \in \{1, \text{t}, \text{ts}\}, \tag{67}$$

where $r_{\text{a}}(h_i, \mathcal{E}_i, \mathcal{A})$, $\text{a} \in \{1, \text{t}, \text{ts}\}$ is defined in (58), (63) and (64), respectively, and where $\mathcal{S}_{\mathcal{A}} = \{\mathcal{E} \in (0, \mathcal{A}/2]^N | \|\mathcal{E}\|_1 \leq \mathcal{E}\}$.

Proof: Similar to (10)–(11), we can show that $C_{\mathcal{A}}(\boldsymbol{h}, \mathcal{E}) = \max_{\mathcal{E} \in \mathcal{S}_{\mathcal{A}}}\sum_{i \in \mathcal{N}} C_{\mathcal{A}}(h_i, \mathcal{E}_i)$. Then, by using (58), (63), and (64), 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 [32]. ∎

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

Theorem 3: The capacity $C_{\mathcal{A}}(\boldsymbol{h}, \mathcal{E})$ is upper bounded by

$$\overline{C}_{\mathcal{A}}(\boldsymbol{h}, \mathcal{E}) = \max_{\mathcal{E} \in \mathcal{S}_{\mathcal{A}}}\sum_{i \in \mathcal{N}}\min\{\bar{r}(h_i, \mathcal{E}_i), \bar{\bar{r}}(h_i, \mathcal{A})\}, \tag{68}$$

where $\bar{r}(h_i, \mathcal{E}_i)$ and $\bar{\bar{r}}(h_i, \mathcal{A})$ are defined in (9) and (65), respectively, and where $\mathcal{S}_{\mathcal{A}}$ is as defined in Theorem 2.

Proof: The statement follows from $C_{\mathcal{A}}(\boldsymbol{h}, \mathcal{E}) = \max_{\mathcal{E} \in \mathcal{S}_{\mathcal{A}}}\sum_{i \in \mathcal{N}} C_{\mathcal{A}}(h_i, \mathcal{E}_i)$ and (66). ∎

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

Theorem 4: The capacity $C_{\mathcal{A}}(\boldsymbol{h}, \mathcal{E})$ with $\mathcal{E} > \frac{N\mathcal{A}}{2}$ satisfies

$$C_{\mathcal{A}}(\boldsymbol{h}, \mathcal{E}) = \sum_{i \in \mathcal{N}} C_{\mathcal{A}}(h_i, \mathcal{A}/2), \tag{69}$$

and

$$\lim_{\mathcal{A} \to \infty}\left[C_{\mathcal{A}}(\boldsymbol{h}, \mathcal{E}) - \sum_{i \in \mathcal{N}}\frac{1}{2}\log\left(\frac{h_i^2\mathcal{A}^2}{2\pi e}\right)\right] = 0. \tag{70}$$

Proof: The first part follows from $C_{\mathcal{A}}(\boldsymbol{h}, \mathcal{E}) = \max_{\mathcal{E} \in \mathcal{S}_{\mathcal{A}}}\sum_{i \in \mathcal{N}} C_{\mathcal{A}}(h_i, \mathcal{E}_i)$ and since $C_{\mathcal{A}}(h_i, \mathcal{E}_i)$ is maximized if $\mathcal{E}_i = \mathcal{A}/2$ as shown in [32]. 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 $R_1(\boldsymbol{h}, \mathcal{E}, \mathcal{A}) \geq \sum_{i \in \mathcal{N}} r_1(h_i, \mathcal{A}/2, \mathcal{A})$. Substituting in (58) 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 intensity allocation for the case

$$\mathcal{E} \leq \frac{N\mathcal{A}}{2}. \tag{71}$$

2) Intensity Allocation Given $\boldsymbol{h}$: The achievable rates $r_{\mathtt{l}}(h, \mathcal{E}, \mathcal{A})$ (58) and $r_{\mathtt{t}}(h, \mathcal{E}, \mathcal{A})$ 63 can not be written in the form $\frac{1}{2} \log(1 + c^2 \mathcal{E}^2)$. Hence, the allocation algorithms discussed previously do not apply immediately to $R_{\mathtt{l}}(\boldsymbol{h}, \mathcal{E}, \mathcal{A})$ and $R_{\mathtt{t}}(\boldsymbol{h}, \mathcal{E}, \mathcal{A})$ given in Theorem 2. However, those algorithms can be applied for $R_{\mathtt{ts}}(\boldsymbol{h}, \mathcal{E}, \mathcal{A})$, since given $\mathcal{E}_i \leq \frac{\mathcal{A}}{2}$, $r_{\mathtt{ts}}(h, \mathcal{E}, \mathcal{A})$ becomes $\frac{1}{2} \log(1 + h_i^2 \mathcal{E}^2/9)$.

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

$$\max_{\mathcal{E}} \quad f(\mathcal{E}) \tag{72}$$
$$\text{s.t.} \quad \mathcal{E} \in [0, \mathcal{A}/2]^N, \ \|\mathcal{E}\|_1 \leq \mathcal{E},$$

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

$$c_i = \frac{h_i}{3}. \tag{73}$$

This problem can be approached similar to (22), except that $\mathcal{E}_i^+(\lambda)$ and $\mathcal{E}_i^-(\lambda)$ have to be replaced with their truncated versions $\min \left\{ \mathcal{E}_i^+(\lambda), \frac{\mathcal{A}}{2} \right\}$ and $\min \left\{ \mathcal{E}_i^-(\lambda), \frac{\mathcal{A}}{2} \right\}$, respectively.[9] It can be demonstrated similar to the case without a peak constraint (Sec. IV-B2) that equal intensity allocation, i.e., $\mathcal{E}_i = \min\{\frac{\mathcal{E}}{N}, \frac{\mathcal{A}}{2}\}$, is optimal in problem (72) at high SNR. On the other hand, the optimal solution for (72) at low SNR (small $\mathcal{E}$ and $\mathcal{A}$) satisfies

$$\mathcal{E}_i = \frac{\mathcal{A}}{2}, \ i = 1, \ldots, N_s, \tag{74a}$$
$$\mathcal{E}_{N_s+1} = \mathcal{E} - N_s \frac{\mathcal{A}}{2}, \tag{74b}$$
$$\mathcal{E}_i = 0, \ i = N_s + 2, \ldots, N, \tag{74c}$$

where $N_s = \left\lfloor \frac{\mathcal{E}}{\mathcal{A}/2} \right\rfloor$.[10] This can be proved similar to Proposition 2. Indeed, exhaustive search can be applied to find the optimal solution of (72) for general SNR. Instead of exhaustive search which has exponential complexity in N, the JA algorithm provides a near-optimal solution of (72) after some modification as given in Algorithm 3.

In this algorithm, we use $\boldsymbol{x}(a : b)$ to denote the vector $(x_a, x_{a+1}, \ldots, x_b)$. This algorithm first releases the peak constraint and finds an allocation satisfying $\|\mathcal{E}\|_1 = \mathcal{E}$ using the JA algorithm (step 4). If this allocation leads to $\mathcal{E}_j \leq \frac{\mathcal{A}}{2}$ for all $i \in \mathcal{N}$, the algorithm is terminated (step 7). Otherwise, the algorithm reintroduces the peak constraint by reducing every $\mathcal{E}_j > \frac{\mathcal{A}}{2}$ to $\frac{\mathcal{A}}{2}$ (step 10). If the involved channels are channels $j = 1, \ldots, i$, then this completes the allocation for the first i channels. The residual average intensity is calculated afterwards (step 12), and the process is repeated for allocating this residual intensity to channels $i+1, \ldots, N$. This is repeated until all channels have been considered.

---

[9]Similar to truncated water-filling in RF channels [45].

[10]$\lfloor x \rfloor$ is defined as the largest integer smaller than or equal to x.

Algorithm 3 JA-P - Joint activation/allocation for peak-constrained parallel IM-DD channels

```
1: function JA-P(c_1, ..., c_{N+1}, E)
2:     i ← 1; Ẽ ← E;
3:     while i < N + 1 do
4:         Ê ← JA(c_i, ..., c_{N+1}, Ẽ);
5:         E ← [E(1 : i − 1), Ê(i : N)];
6:         if E_i ≤ A/2 then
7:             i ← N + 1
8:         else
9:             while E_i > A/2 do
10:                E_i ← A/2; i ← i + 1;
11:            end while
12:            Ẽ ← E − (i − 1)A/2
13:        end if
14:    end while
15:    return E
16: end function
```

In this algorithm, step 4 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 SA algorithm in step 4 (i.e., $\hat{\mathcal{E}} \leftarrow \text{SA}(c_i, \ldots, c_{N+1}, \tilde{\mathcal{E}})$ in step 4) instead of the JA algorithm. We call this variant SA-P.

3) Evaluation:

a) Fixed $\boldsymbol{h}$: Fig. 6a shows the achievable rate $\frac{1}{2} \sum_{i \in \mathcal{N}} \log \left(1 + \frac{h_i^2 \mathcal{E}_i^2}{9} \right)$, for an exemplary channel with $N = 4$, $\boldsymbol{h} = [1, .7, .3, .1]$, and $\mathcal{E} = N\mathcal{A}/10$. The figure shows the performance of intensity allocation by exhaustive search, JA-P, SA-P, equal allocation, or allocating the intensity to the strongest $N_s + 1$ channels as in (74). The performance of both JA-P and SA-P nearly coincides with exhaustive search in this case. Thus, we focus on SA-P from this point on for its simplicity. The allocation from SA-P can be used in conjunction with Theorem 2 to obtain the following achievable rates.

Proposition 4: The capacity $C_{\mathcal{A}}(\boldsymbol{h}, \mathcal{E})$ of a parallel IM-DD channel with $\mathcal{E} \leq \frac{N\mathcal{A}}{2}$ is lower bounded by the achievable rates

$$\tilde{R}_{\mathtt{a}}(\boldsymbol{h}, \mathcal{E}, \mathcal{A}) = \sum_{i \in \mathcal{N}} r_{\mathtt{a}}(h_i, \mathcal{E}_i, \mathcal{A}), \quad \mathtt{a} \in \{\mathtt{l}, \mathtt{t}, \mathtt{ts}\}, \tag{75}$$

where $r_{\mathtt{a}}(h, \mathcal{E}_i, \mathcal{A})$ with $\mathtt{a} \in \{\mathtt{l}, \mathtt{t}, \mathtt{ts}\}$ is defined in (58), (63), and (64), respectively, and where $(\mathcal{E}_1, \ldots, \mathcal{E}_N)$ is the output of the SA-P algorithm with $c_i = h_i/3$.

The lower bounds in Proposition 4 are shown in Fig. 6b for the same exemplary channel. Here, $\tilde{R}_{\mathtt{ts}}(\boldsymbol{h}, \mathcal{E}, \mathcal{A})$ serves as an intermediate step to obtain a good allocation for $\tilde{R}_{\mathtt{t}}(\boldsymbol{h}, \mathcal{E}, \mathcal{A})$ and $\tilde{R}_{\mathtt{l}}(\boldsymbol{h}, \mathcal{E}, \mathcal{A})$. The jumps in the achievable rate that can be seen in this figure correspond to values of $\mathcal{E}$ where an additional channel is activated. 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 small constant as $\mathcal{E}$ increases. This can be shown as follows.

Consider the lower bound $\tilde{R}_{\mathtt{t}}(\boldsymbol{h}, \mathcal{E}, \mathcal{A})$. We know from Sec. IV-B4 that under an average constraint only, the SA
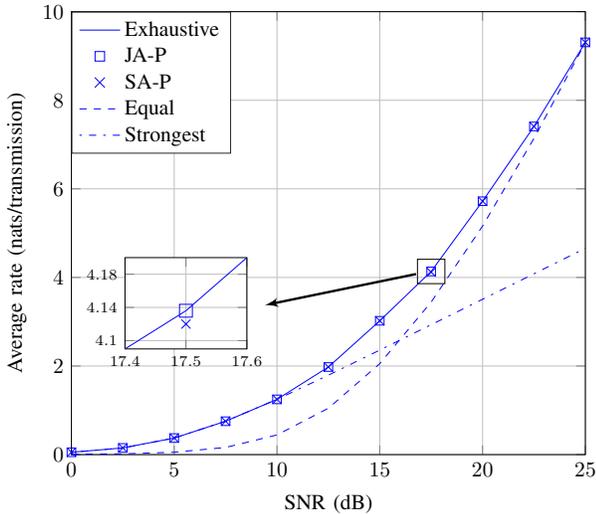
(a) Performance in comparison to exhaustive search.
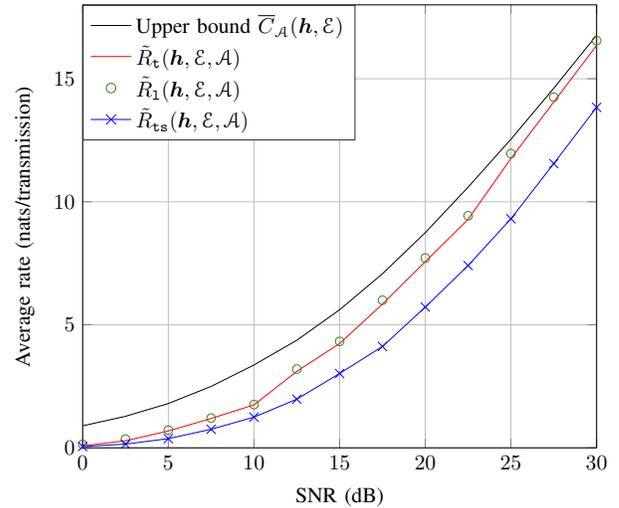
(b) Performance in comparison to the upper bound.

Fig. 6: Performance as a function of SNR ($\mathcal{E}/\sigma$) for a peak- and average-constrained system with $N = 4$, $\boldsymbol{h} = [1, .7, .3, .1]$, and $\mathcal{E} = N\mathcal{A}/10$.

algorithm finds the optimal solution $\mathcal{E}_i = \frac{\mathcal{E}}{N}$ at high SNR. Thus, steps 4 and 5 in Algorithm 3 give $\mathcal{E}_i = \frac{\mathcal{E}}{N}$ (which is feasible by (71)). By Proposition 4, $\tilde{R}_{\mathtt{t}}(\boldsymbol{h}, \mathcal{E}, \mathcal{A}) = \sum_{i \in \mathcal{N}} r_{\mathtt{t}}(h_i, \mathcal{E}/N, \mathcal{A})$ is achievable at high SNR. From [34], we know that $r_{\mathtt{t}}(h_i, \mathcal{E}/N, \mathcal{A})$ 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

$$\min\left\{\frac{1}{2}\log\left(\frac{eh_i^2\mathcal{E}^2}{2\pi N^2}\right), \frac{1}{2}\log\left(\frac{h_i^2\mathcal{A}^2}{2\pi e}\right)\right\}. \quad (76)$$

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

$$\tilde{R}_{\mathtt{t}}(\boldsymbol{h}, \mathcal{E}, \mathcal{A}) = \sum_{i \in \mathcal{N}} \min\left\{\frac{1}{2}\log\left(\frac{eh_i^2\mathcal{E}^2}{2\pi N^2}\right), \frac{1}{2}\log\left(\frac{h_i^2\mathcal{A}^2}{2\pi e}\right)\right\}$$
$$- 0.1N. \quad (77)$$

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

$$\overline{C}_{\mathcal{A}}(\boldsymbol{h}, \mathcal{E}) \leq \max_{\mathcal{E} \in \mathcal{S}} \sum_{i \in \mathcal{N}} \overline{r}(h_i, \mathcal{E}_i) \quad (78)$$

$$= \sum_{i \in \mathcal{N}} \frac{1}{2}\log\left(\frac{eh_i^2\mathcal{E}^2}{2\pi N^2}\right), \quad (79)$$

which follows by (43)–(44), and

$$\overline{C}_{\mathcal{A}}(\boldsymbol{h}, \mathcal{E}) \leq \max_{\mathcal{E} \in \mathcal{S}} \sum_{i \in \mathcal{N}} \overline{\overline{r}}(h_i, \mathcal{A}) \quad (80)$$

$$= \sum_{i \in \mathcal{N}} \frac{1}{2}\log\left(\frac{h_i^2\mathcal{A}^2}{2\pi e}\right). \quad (81)$$

By comparing (77), (79), and (81), we can see that

$$\overline{C}_{\mathcal{A}}(\boldsymbol{h}, \mathcal{E}) - \tilde{R}_{\mathtt{t}}(\boldsymbol{h}, \mathcal{E}, \mathcal{A}) \leq 0.1N \text{ nats} \quad (82)$$

at high SNR, leading to the following corollary.

Corollary 4: The capacity $C_{\mathcal{A}}(\boldsymbol{h}, \mathcal{E})$ with $\mathcal{E} \leq \frac{N\mathcal{A}}{2}$ satisfies

$$\lim_{\mathcal{E} \to \infty} \left[C_{\mathcal{A}}(\boldsymbol{h}, \mathcal{E}) - \sum_{i \in \mathcal{N}} \frac{1}{2}\log\left(\frac{eh_i^2\mathcal{E}^2}{2\pi N^2}\right)\right] \leq 0.1N \text{ nats} \quad (83)$$

Let us define the pre-log of the channel under average and peak constraints as $\lim_{a \to \infty} \frac{C_{\mathcal{A}}(\boldsymbol{h}, \mathcal{E})}{\log(a)}$ where $a = \min\{\mathcal{E}, \mathcal{A}\}$. By Theorem 4 and Corollary 4, the pre-log of the channel is N similar to the case with an average constraint only.

b) Block Fading $\boldsymbol{h}$: Similar to the case with no peak constraint, proposed algorithms might not always lead to the optimal solution since they exclude solutions of the form $\mathcal{E}_i^-(\lambda)$. To show that this only slightly affects performance, we consider block fading channels. Fig. 7 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 realizations of the FSO channel given in (45), with intensity allocation by exhaustive search, JA-P, SA-P, equal allocation, or allocating the intensity to the strongest $N_s + 1$ channels as in (74). It can be seen that the performance of the proposed algorithms nearly coincides with exhaustive search on average. Note from this figure that intensity allocation becomes more important for lower ratios of $\mathcal{E}/\mathcal{A}$. When this ratio is closer to $\frac{N}{2}$, we will have enough intensity to allocate $\mathcal{E}_i = \frac{\mathcal{A}}{2}$ to all channels. That is why the plots in Fig. 7b are very close.

C. Quasi-static Fading Parallel Channels without CSIT

In the absence of CSIT, we study the minimum outage probability $P_o(\mathcal{E}, \mathcal{A}, R_0)$ of the system when transmitting at rate R_0 and total average intensity per block $\mathcal{E}$, as in Sec. V. For $\mathcal{E} > \frac{N\mathcal{A}}{2}$, due to Theorem 4, we have that the minimum outage probability at high SNR converges to

$$P_o(\mathcal{E}, \mathcal{A}, R_0) = \mathbb{P}\left\{\sum_{i \in \mathcal{N}} \frac{1}{2}\log\left(\frac{h_i^2\mathcal{A}^2}{2\pi e}\right) < R_0\right\}. \quad (84)$$
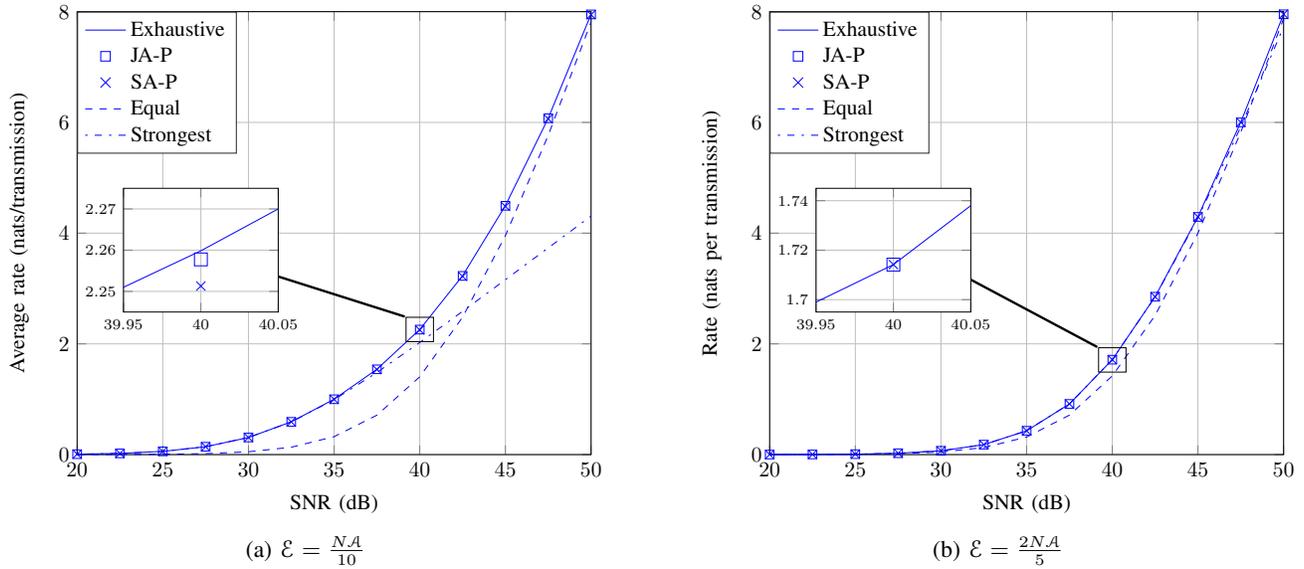
(a) $\mathcal{E} = \frac{N\mathcal{A}}{10}$

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

Fig. 7: The achievable rate as a function of SNR ($\mathcal{E}/\sigma$) for a peak- and average-constrained fading FSO system with $N = 4$ and channels as in (45) under clear weather.

On the other hand, for $\mathcal{E} \leq \frac{N\mathcal{A}}{2}$, an upper bound on $P_o(\mathcal{E}, \mathcal{A}, R_0)$ can be obtained by fixing a specific distribution over each channel, and using equal power allocation. This leads to

$$P_o(\mathcal{E}, \mathcal{A}, R_0) \leq \mathbb{P}\left\{R'_1(\boldsymbol{h}, \mathcal{E}, \mathcal{A}) < R_0\right\}, \tag{85}$$

$$P_o(\mathcal{E}, \mathcal{A}, R_0) \leq \mathbb{P}\left\{R'_t(\boldsymbol{h}, \mathcal{E}, \mathcal{A}) < R_0\right\}, \tag{86}$$

$$P_o(\mathcal{E}, \mathcal{A}, R_0) \leq \mathbb{P}\left\{R'_{ts}(\boldsymbol{h}, \mathcal{E}, \mathcal{A}) < R_0\right\}, \tag{87}$$

where

$$R'_a(\boldsymbol{h}, \mathcal{E}, \mathcal{A}) = \sum_{i \in \mathcal{N}} r_a(h_i, \mathcal{E}/N, \mathcal{A}), \quad a \in \{1, t, ts\}, \tag{88}$$

and $r_a(h_i, \mathcal{E}_i, \mathcal{A})$, $a \in \{1, t, ts\}$ is defined in (58), (63), and (64), respectively.

Moreover, when the transmitter transmits at rate R_0 and total average intensity per block $\mathcal{E}$, $P_o(\mathcal{E}, \mathcal{A}, R_0)$ is lower bounded by

$$P_o(\mathcal{E}, \mathcal{A}, R_0) \geq \mathbb{P}\left\{\overline{C}_{\mathcal{A}}(\boldsymbol{h}, \mathcal{E}) < R_0\right\}, \tag{89}$$

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

The upper bound (86) and the lower bound (89) are fairly tight at high SNR. Namely, at high SNR, $R'_t(\boldsymbol{h}, \mathcal{E}, \mathcal{A})$ can be written similar to $\tilde{R}_t(\boldsymbol{h}, \mathcal{E}, \mathcal{A})$ (77), whose gap to $\overline{C}_{\mathcal{A}}(\boldsymbol{h}, \mathcal{E})$ is at most $0.1N$ at high SNR (82). Consequently, at high SNR,

$$P_o(\mathcal{E}, \mathcal{A}, R_0) \leq \mathbb{P}\left\{\sum_{i \in \mathcal{N}} \frac{1}{2}\log\left(\frac{eh_i^2\mathcal{E}^2}{2\pi N^2}\right) - 0.1N < R_0\right\}, \tag{90}$$

$$P_o(\mathcal{E}, \mathcal{A}, R_0) \geq \mathbb{P}\left\{\sum_{i \in \mathcal{N}} \frac{1}{2}\log\left(\frac{eh_i^2\mathcal{E}^2}{2\pi N^2}\right) < R_0\right\}. \tag{91}$$

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

VII. CONCLUSION

We studied a system of N parallel OWC channels employing IM-DD, under a total average intensity constraint. We provided capacity bounds for this system. Optimal allocation of the permissible total intensity over the N channels is important for enhancing the overall system performance under availability of CSIT. For a constant channel, and under an exponential input distribution, we obtain a capacity lower bound whose maximization is nonconvex, contrary to water-filling power allocation. We described the optimal intensity allocation for this lower bound, and showed that it allocates intensity equally at high SNR, and activates the strongest channel at low SNR. Finding the optimal allocation at moderate SNR is of high complexity. To simplify this process, we proposed simple linear-complexity algorithms which find a near-optimal allocation, and outperform equal allocation and allocating all intensity 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. If CSIT is not available and under a quasi-static fading channel model, we provided bounds on the minimum outage probability which are also tight at high SNR. We also characterized the outage capacity at high SNR. For an FSO system with $N = 2$ in clear weather with log-normal fading and Rayleigh pointing error, we showed numerically that the absence of CSIT costs about 3.5dB and 7dB for an outage probability of 0.1 and 0.01, respectively, compared to a system with CSIT at high SNR. We also showed that this cost vanishes as $N \to \infty$, demonstrating a diversity gain. Finally, we extended the results to parallel channels 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 [40], [46]. Another interesting extension is to investigate the capacity of parallel IM-DD channels with

correlated noises, and the ergodic capacity under block fading.

REFERENCES

[1] A. Chaaban, Z. Rezki, and M.-S. Alouini, "Capacity bounds for parallel optical wireless channels," in *Proc. of IEEE Int. Conf. Commun. (ICC)*, Kuala Lumpur, Malaysia, May 2016.

[2] T. S. Rappaport, S. Shu, R. Mayzus, Z. Hang, Y. Azar, K. Wang, G. N. Wong, J. K. Schulz, M. Samimi, and F. Gutierrez, "Millimeter wave mobile communications for 5G cellular: It will work!" *IEEE Access*, vol. 1, pp. 335–349, May 2013.

[3] 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.

[4] S. Hranilovic and F. R. Kschischang, "Capacity bounds for power- and band-limited optical intensity channels corrupted by Gaussian noise," *IEEE Trans. Inf. Theory*, vol. 50, no. 5, pp. 784–795, May 2004.

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

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

[7] S. D. Dissanayake and J. Armstrong, "Comparison of ACO-OFDM, DCO-OFDM and ADO-OFDM in IM/DD systems," *J. Lightw. Technol.*, vol. 31, no. 7, pp. 1063–1072, Apr. 2013.

[8] 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 Trans. Commun.*, vol. 60, no. 3, pp. 799–809, Mar. 2012.

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

[10] 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.*, vol. 2010, no. 64, pp. 1–11, Jan. 2010.

[11] A. A. Farid and S. Hranilovic, "Channel capacity and non-uniform signalling for free-space optical intensity channels," *IEEE J. Sel. Areas Commun.*, vol. 27, no. 9, pp. 1–12, Dec. 2009.

[12] ——, "Outage capacity optimization for free-space optical links with pointing errors," *IEEE/OSA J. Lightw. Technol.*, vol. 25, no. 7, pp. 1702–1710, Jul. 2007.

[13] ——, "Diversity gain and outage probability for MIMO free-space optical links with misalignment," *IEEE Trans. Commun.*, vol. 60, no. 2, pp. 479–487, Feb. 2012.

[14] M. Khalighi, N. Schwartz, N. Aitamer, and S. Bourennane, "Fading reduction by aperture averaging and spatial diversity in optical wireless systems," *IEEE/OSA J. Optical Commun. and Networking*, vol. 1, no. 6, pp. 580–593, Nov. 2009.

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

[16] J. Akella, M. Yuksel, and S. Kalyanaraman, "Error analysis of multi-hop free-space optical communications," in *Proc. of IEEE Int. Conf. Commun. (ICC)*, May. 2005, pp. 1777–1781.

[17] S. Kazemlou, S. Hranilovic, and S. Kumar, "All-optical multihop free-space optical communication systems," *J. Lightw. Technol.*, vol. 29, no. 18, pp. 2663–2669, Jun. 2011.

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

[19] D. K. Borah, A. C. Boucouvalas, C. C. Davis, S. Hranilovic, and K. Yiannopoulos, "A review of communication-oriented optical wireless systems," *EURASIP J. Wireless Commun. and Networking*, vol. 2012, no. 1, pp. 1–28, 2012.

[20] T. Fath and H. Haas, "Performance comparison of MIMO techniques for optical wireless communications in indoor environments," *IEEE Trans. Commun.*, vol. 61, no. 2, pp. 733–742, Feb. 2013.

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

[22] 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 Conf. on Network and Optical Commun.*, Graz, Austria, July 2013, pp. 275–280.

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

[24] M. L. B. Riediger, R. Schober, and L. Lampe, "Fast multiple-symbol detection for free-space optical communications," *IEEE Trans. Commun.*, vol. 57, no. 4, pp. 1119–1128, Apr. 2009.

[25] 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 J. Sel. Areas Commun.*, vol. 27, no. 9, pp. 1654–1662, Dec. 2009.

[26] E. Ciaramella, Y. Arimoto, G. Contestabile, M. Presi, A. D'Errico, V. Guarino, and M. Matsumoto, "1.28 terabit/s (32x40 gbit/s) WDM transmission system for free space optical communications," *IEEE J. Sel. Areas Commun.*, vol. 27, no. 9, pp. 1639–645, Dec. 2009.

[27] K. Wang, A. Nirmalathas, C. Lim, and E. Skafidas, "4×12.5 Gb/s WDM optical wireless communication system for indoor applications," *J. Lightw. Technol.*, vol. 29, no. 13, pp. 1988–1996, July 2011.

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

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

[30] A. M. Abdelhady, O. Amin, A. Chaaban, and M.-S. Alouini, "Downlink resource allocation for multichannel TDMA visible light communications," in *Proc. of the IEEE Global Conf. on Signal and Inf. Process. (GlobalSIP)*, Washington, D.C., USA, Dec. 2016.

[31] S. Arnon, J. Barry, G. Karagiannidis, R. Schober, and M. Uysal, Eds., *Advanced Optical Wireless Communication Systems*. Cambridge University Press, 2012.

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

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

[34] A. Chaaban, J.-M. Morvan, and M.-S. Alouini, "Free-space optical communications: Capacity bounds, approximations, and a new sphere-packing perspective," *IEEE Trans. Commun.*, vol. 64, no. 3, pp. 1176–1191, Mar. 2016.

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

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

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

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

[39] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge University Press, 2004.

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

[41] H. Zhou, S. Mao, and P. Agrawal, "Optical power allocation for adaptive transmissions in wavelength-division multiplexing free space optical networks," *Digital Commun. and Networks*, vol. 1, no. 3, pp. 171–180, 2015.

[42] X. Liu, "Free-space optics optimization models for building sway and atmospheric interference using variable wavelength," *IEEE Trans. Commun.*, vol. 57, no. 2, pp. 492–498, Feb. 2009.

[43] S. Arnon, "Effects of atmospheric turbulence and building sway on optical wireless-communication systems," *Opt. Lett.*, vol. 28, no. 2, pp. 129–131, Jan 2003.

[44] A. Chaaban, Z. Rezki, and M.-S. Alouini, "On the capacity of the intensity-modulation direct-detection optical broadcast channel," *IEEE Trans. Wireless Commun.*, vol. 15, no. 5, pp. 3114–3130, May 2016.

[45] K. D. Nguyen, A. G. I. Fabregas, and L. K. Rasmussen, "Power allocation for block-fading channels with arbitrary input constellations," *IEEE Trans. Wireless Commun.*, vol. 8, no. 5, pp. 2514–2523, May 2009.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCOMM.2016.2621743, IEEE Transactions on Communications

16

[46] M. Karimi and M. Uysal, "Novel adaptive transmission algorithms for free-space optical links," *IEEE Trans. Commun.*, vol. 60, no. 12, pp. 3808–3815, Dec. 2012.

Anas Chaaban (S'09 - M'14) received his Maîtrise ès Sciences degree in electronics from the Lebanese University, Lebanon, in 2006. He received his M.Sc. degree in communications technology and his Dr.-Ing. (Ph.D.) degree in Electrical Engineering and Information Technology from the University of Ulm and the Ruhr-University of Bochum, Germany, in 2009 and 2013, respectively. During 2008-2009, he was with the Daimler AG research group on machine vision, Ulm, Germany. He was a Research Assistant with the Emmy- Noether Research Group on Wireless Networks at the University of Ulm, Germany, during 2009-2011, which relocated to Ruhr-Universität Bochum, Germany, in 2011. He was a postdoctoral researcher at the Ruhr-Universität Bochum, Germany, in 2013-2014, and joined King Abdullah University of Science and Technology as a postdoctoral researcher in 2015. His research interests are in the areas of information theory and wireless communications.

Zouheir Rezki (S'01-M'08-SM'13) was born in Casablanca, Morocco. He received the Diplôme d'Ingénieur degree from the École Nationale de l'Industrie Minérale (ENIM), Rabat, Morocco, in 1994, the M.Eng. degree from École de Technologie Supérieure, Montreal, Québec, Canada, in 2003, and the Ph.D. degree from École Polytechnique, Montreal, Québec, in 2008, all in electrical engineering. From October 2008 to September 2009, he was a postdoctoral research fellow with Data Communications Group, Department of Electrical and Computer Engineering, University of British Columbia. Then, has been a Senior Research Scientist at King Abdullah University of Science and Technology (KAUST), Saudi Arabia until June 2916. He joined University of Idaho as a Faculty Member on August 2016. His research interests include: performance limits of communication systems, physical-layer security, cognitive and sensor networks and low-complexity detection algorithms.

Mohamed-Slim Alouini (S'94, M'98, SM'03, F09) was born in Tunis, Tunisia. He received the Ph.D. degree in Electrical Engineering from the California Institute of Technology (Caltech), Pasadena, CA, USA, in 1998. He served as a faculty member in the University of Minnesota, Minneapolis, MN, USA, then in the Texas A&M University at Qatar, Education City, Doha, Qatar before joining King Abdullah University of Science and Technology (KAUST), Thuwal, Makkah Province, Saudi Arabia as a Professor of Electrical Engineering in 2009. His current research interests include the modeling, design, and performance analysis of wireless communication systems.