Low-SNR Asymptotic Capacity of MIMO Optical Intensity Channels with Peak and Average Constraints

Anas Chaaban, *Senior Member, IEEE*, Zouheir Rezki, *Senior Member, IEEE*, and Mohamed-Slim Alouini, *Fellow, IEEE*

Abstract—The low-SNR asymptotic capacity of the multiple-input multiple-output (MIMO) optical intensity channel is studied under both average and peak intensity constraints. We focus on low SNR, which can be modeled as the scenario where both constraints proportionally vanish, or where the peak constraint is held constant while the average constraint vanishes. A capacity upper bound is derived, and is shown to be tight at low SNR under both scenarios. The capacity achieving input distribution at low SNR is shown to be a maximally-correlated vector-binary input distribution. Consequently, the low-SNR capacity of the channel is characterized. As a byproduct, it is shown that for a channel with peak intensity constraints only, or with peak intensity constraints and individual (per aperture) average intensity constraints, a simple scheme composed of coded on-off keying, spatial repetition, and maximum-ratio combining is optimal at low SNR.

Index Terms—Intensity modulation; MIMO; low-SNR capacity; maximally-correlated binary; OOK; repetition; MRC

I. Introduction

Intensity modulation is a simple transmission scheme that is favorable in applications such as optical wireless communications (OWC) [2]. In this context, a transmitter modulates light intensity, and a receiver uses a photo-detector to detect the signal. The resulting scheme is known as intensity-modulation direct-detection (IM-DD).

IM-DD imposes a nonnegativity constraint on the transmit signal. In addition to this, the transmit signal is commonly constrained by average and peak constraints due to safety and practical considerations. At the receiver side, the signal is disturbed by several sources of noise, such as ambient light and thermal noise. The resulting channel can be modeled as a Poisson channel [3] or an additive Gaussian noise channel [4], [5] depending on the regime of operation, the latter being more common in regimes with high received signal power. In this

A. Chaaban is with the School of Engineering, University of British Columbia, Kelowna, BC Canada V1V 1V7. Email: anas.chaaban@ubc.ca. His work was supported by the Natural Sciences and Engineering Research Council of Canada (NSERC).

Z. Rezki is with the Electrical and Computer Engineering Department, University of Idaho, Moscow, ID 83844 USA. Email: zrezki@uidaho.edu. His work is supported in part by the Qatar National Research Fund (a member of Qatar Foundation) under Grant NPRP 9-077-2-036. The statements made herein are solely the responsibility of the authors.

M.-S. Alouini is with the Division of Computer, Electrical, and Mathematical Sciences and Engineering (CEMSE) at King Abdullah University of Science and Technology (KAUST), Thuwal 23955-6900, Saudi Arabia. Email: slim.alouini@kaust.edu.sa.

Part of the results of this paper will appear in ISIT 2018 [1].

paper, we focus on the IM-DD channel with input-independent Gaussian noise [4], which is a suitable model for OWC with strong ambient light and/or thermal noise [6].

Due to the above constraints, the capacity of the Gaussian IM-DD channel is not the same as the classical Gaussian channel used for modeling radio-frequency (RF) communications. Studying the capacity of this channel is important for understanding the fundamental limits of OWC, which has witnessed increasing attention recently, see [2], [7]–[10] and references therein. In this context, the capacity of the single-input single-output (SISO) IM-DD channel has been studied in [4], [11], [12]. The SISO IM-DD channel with input-dependent (relative intensity) Gaussian noise has been studied in [5]. The capacity of multi-user SISO IM-DD channels has been studied in [13], [14]. The performance of transmission schemes in IM-DD OWC in terms of error and outage probability has been investigated in [15]–[19], to name a few.

Similar to RF communication, one can also realize MIMO transmission in OWC using multiple light sources and detectors [20] or multiple colors in visible-light communication [19]. If the transmit-receive aperture pairs do not interfere, then we obtain a system of parallel channels [21]. Otherwise, we get a general MIMO IM-DD channel which has been studied in [20], [22]–[24] where various communication schemes were analyzed. Recently, the capacity of the Gaussian MIMO IM-DD channel has been studied in [25]–[27], where capacity bounds have been derived (for any SNR), and the asymptotically high-SNR capacity has been characterized for the case where the number of detectors is more than the number of transmit apertures. However, these works did not consider the low-SNR asymptotic capacity of the MIMO IM-DD channel. This can arise if the receiver experiences strong noise, or if the average and/or peak intensity constraints at the transmitter are small.

Studying the low-SNR asymptotic capacity is not new in itself. To name a few, this has been studied earlier in [28] which provides capacity expressions under "weak" input signals, in [29] which focuses on low-SNR second order asymptotics of channel capacity, and in [30] which discusses low SNR asymptotics using the relation between mutual information and minimum mean square error (MMSE) over Gaussian channels. In this context, the contribution of this work is characterizing the low-SNR asymptotic capacity of MIMO IM-DD channels and shedding light on the capacity achieving input in this regime. It is worth to note that the low-SNR

asymptotic capacity for multiple-input single-output (MISO) IM-DD channels has been characterized recently, in a parallel and independent work, in [31].

We start by deriving a capacity upper bound for any SNR using the "Gaussian maximizes entropy" principle. To prove its asymptotic tightness at low SNR, we inspire from earlier works asserting the optimality of binary inputs at low SNR, such as [4], [32], [33]. We show that the asymptotically low-SNR optimal input in the MIMO IM-DD context is a *maximally correlated vector-binary distributed input*. For instance, with 3 transmit apertures, a peak constraint $\mathcal{A}$, and average constraints $\mathcal{E}_1 \geq \mathcal{E}_2 \geq \mathcal{E}_3$, respectively, the resulting optimal input alphabet is

$$\{(0,0,0),(\mathcal{A},0,0),(\mathcal{A},\mathcal{A},0),(\mathcal{A},\mathcal{A},\mathcal{A})\}.$$

Thus, transmit apertures are either ON or OFF with a specific structure. Namely, an aperture is OFF if an aperture with a larger average intensity is OFF, and can be ON or OFF otherwise. This input has the largest pair-wise correlation among vector-binary distributions with the given peak and average constraints, and hence the name (a formal definition is given in Sec. III). The achievable rate is derived using results in [28]. This leads to the low-SNR asymptotic capacity under a total average intensity constraint for two cases: (i) proportionally vanishing average and peak constraints, and (ii) fixed peak constraint and vanishing average constraint. As a byproduct, we obtain the low-SNR asymptotic capacity for a channel with per aperture average intensity constraints under the two cases above. We also provide an alternative proof of this result based on the relation between mutual information and MMSE over a Gaussian channel [30].

We note that the optimal distribution above bears resemblance to the structure of the high-SNR capacity-achieving distribution for MISO IM-DD channels discovered recently in [34], where one transmit aperture must be ON in order for the remaining ones to be activated. Moreover, our result for proportionally vanishing constraints reproduces the low-SNR asymptotic capacity of the MISO channel given independently in [31, Proposition 12].

In addition to the interesting optimal input distribution above, we conclude the following. Under individual average intensity constraints, we show that the asymptotically optimal scheme at low SNR consists of *coded on-off keying (OOK), spatial repetition, and maximum-ratio combining (MRC)*. In other words, the transmitter encodes a single binary stream, which is sent bit-by-bit repeated at all transmit apertures. Thus, for the example above, the input distribution consists of symbols in $\{(0,0,0),(\mathcal{A},\mathcal{A},\mathcal{A})\}$. The receiver performs MRC [35] to combine the received signals of the detectors into a single stream which is then decoded. This simple scheme also applies for the channel under peak constraints only.

Next, we introduce the channel model in Sec. II. Then, we state the main results of the paper in Sec. III which we support with numerical evaluations. The detailed derivation of the results is given in Sec. IV and the paper is concluded in Sec. V.

Throughout the paper, we denote scalars, vectors, and matrices using normal font, boldface lower case, and boldface upper case letters, respectively. The distinction between constants and random quantities will be clear from the context. We use $\boldsymbol{I}_M$ and $\boldsymbol{1}_M$ to denote the $M \times M$ identity matrix and the M-dimensional all-ones vector, respectively. The ℓ_2-norm of $\boldsymbol{x}$ is denoted $\|\boldsymbol{x}\|$, its transpose denoted $\boldsymbol{x}^\mathsf{T}$, and the trace of $\boldsymbol{X}$ is denoted $\mathrm{Tr}(\boldsymbol{X})$. The differential entropy and mutual information are denoted $\mathsf{h}(\cdot)$ and $\mathsf{I}(\cdot;\cdot)$, respectively. We write $\boldsymbol{x} \sim P_{\boldsymbol{x}}(\boldsymbol{x})$ to indicate that $\boldsymbol{x}$ is distributed according to $P_{\boldsymbol{x}}(\boldsymbol{x})$, and use $\mathrm{Bern}(a)$ to denote a Bernoulli distribution with $\mathrm{Prob}(x=1)=a$. We denote the expectation with respect to $P_{\boldsymbol{x}}(\boldsymbol{x})$ as $\mathbb{E}_{P_{\boldsymbol{x}}}[\cdot]$. We write $o(x)$ to denote a function $f(x)$ satisfying $\lim_{x\to 0} \frac{f(x)}{x} = 0$ and $o_x(1)$ to denote a function $f(x)$ satisfying $\lim_{x\to 0} f(x) = 0$.

II. CHANNEL MODEL

Consider an OWC system comprising M transmit and N receive apertures, employing IM-DD. Denote the light intensity of the ith transmitter by x_i, and the received signal at the jth receiver by y_j. The received vector $\boldsymbol{y} = (y_1, \cdots, y_N)^\mathsf{T}$ can be expressed in terms of the input vector $\boldsymbol{x} = (x_1, \cdots, x_M)^\mathsf{T}$ as (Fig. 1)

$$\boldsymbol{y} = \boldsymbol{H}\boldsymbol{x} + \boldsymbol{z}, \tag{1}$$

where $\boldsymbol{z} = (z_1, \cdots, z_N)^\mathsf{T}$ is Gaussian noise with zero mean and covariance matrix $\sigma^2 \boldsymbol{I}_N$, independent and identically distributed (i.i.d.) over time, and $\boldsymbol{H} \in \mathbb{R}_+^{N \times M}$ is a constant matrix with elements $h_{j,i} \geq 0$ denoting the channel gain from transmitter i to receiver j. We denote the ith column of $\boldsymbol{H}$ by $\boldsymbol{h}_i$. The Gaussian noise combines thermal noise which is well-modeled as Gaussian, and shot-noise which has impact when its intensity is large in which case it is well-approximated as Gaussian. The resulting Gaussian channel model is common in the OWC context [4], [36].

The transmit signal x_i is a random variable which satisfies

$$0 \leq x_i \leq \mathcal{A}, \tag{2}$$

i.e., a peak intensity constraint.[1] Additionally, $\boldsymbol{x}$ is subjected to a sum average intensity constraint

$$\sum_{i=1}^{M} \mathcal{E}_i \leq \mathcal{E} \tag{3}$$

for some $\mathcal{E} \in (0, M\mathcal{A})$, where $(\mathcal{E}_1, \ldots, \mathcal{E}_M)^\mathsf{T} \triangleq \boldsymbol{\mathcal{E}} \in [0, \mathcal{A}]^M$ denotes $\mathbb{E}_{P_{\boldsymbol{x}}}[\boldsymbol{x}]$.[2] Note that $\mathcal{E} \geq M\mathcal{A}$ leads to an inactive average constraint.

We denote the channel capacity by $C_\sigma(\mathcal{A}, \mathcal{E})$, which is defined in the standard information-theoretic sense, and is given by $C_\sigma(\mathcal{A}, \mathcal{E}) = \max_{P_{\boldsymbol{x}}} \mathsf{I}(\boldsymbol{x}; \boldsymbol{y})$ [38]. We study $C_\sigma(\mathcal{A}, \mathcal{E})$ at asymptotically low SNR, i.e., where the signal intensity is small *relative* to noise power. Low SNR can be defined in two ways as follows:

---

[1]The case $0 \leq x_i \leq \mathcal{A}_i$ can be transformed to one with a common peak constraint $\mathcal{A}$ by normalization.

[2]Throughout the paper, we use the measure-theoretic definition of the expected-value with respect to a probability measure $P_{\boldsymbol{x}}$, i.e., $\mathbb{E}_{P_{\boldsymbol{x}}}[f(\boldsymbol{x})] = \int_{[0,\mathcal{A}]^M} f(\boldsymbol{x}) P_{\boldsymbol{x}}(d\boldsymbol{x})$ [37, Sec. 21]. Using this definition, the mean and covariance matrix of $\boldsymbol{x}$ exist and have finite components due to the nonnegativity of $\boldsymbol{x}$ and its bounded support.
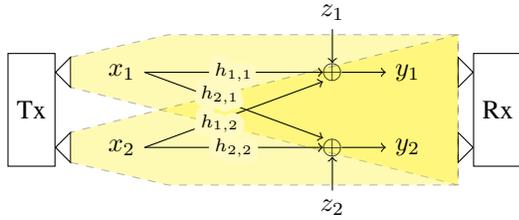
Fig. 1: A MIMO optical wireless communication system: $x_i \geq 0$ is the optical intensity, $h_{i,j} \geq 0$ is a channel gain, and z_i is Gaussian noise.

1) $\frac{\mathcal{A}}{\sigma} \to 0$ and $\frac{\mathcal{E}}{\sigma} \to 0$ proportionally. This will be written as $\frac{\mathcal{A}}{\sigma} \to 0$ with $\mathcal{E} = \alpha \mathcal{A}$ for $\alpha \in (0, M)$ (if $\alpha \geq M$, then the average constraint is redundant). We define the low-SNR asymptotic capacity in this case as $C_{\sigma,0}^{[1]}(\mathcal{A}, \alpha\mathcal{A})$ which satisfies

$$\lim_{\frac{\mathcal{A}}{\sigma} \to 0} \frac{C_\sigma(\mathcal{A}, \alpha\mathcal{A})}{C_{\sigma,1}^{[1]}(\mathcal{A}, \alpha\mathcal{A})} = 1. \qquad (4)$$

2) $\frac{\mathcal{E}}{\sigma} \to 0$ while $\frac{\mathcal{A}}{\sigma}$ is held fixed. We define the low-SNR asymptotic capacity then as $C_{\sigma,0}^{[2]}(\mathcal{A}, \mathcal{E})$ which satisfies

$$\lim_{\frac{\mathcal{E}}{\sigma} \to 0} \frac{C_\sigma(\mathcal{A}, \mathcal{E})}{C_{\sigma,0}^{[2]}(\mathcal{A}, \mathcal{E})} = 1. \qquad (5)$$

We ignore the third case where $\frac{\mathcal{E}}{\sigma}$ is held fixed while $\frac{\mathcal{A}}{\sigma} \to 0$ since this is similar to the first case with $\alpha = M$. Namely, in this case, the corresponding average constraint $\mathcal{E}$ becomes redundant as $\frac{\mathcal{A}}{\sigma}$ decreases.

Another scenario of interest is when $\boldsymbol{x}$ has individual average intensity constraints, i.e., $\mathcal{E}_i \leq \mathcal{E}_{\text{ind}}$ with $\mathcal{E}_{\text{ind}} \in (0, \mathcal{A})$. The low-SNR asymptotic capacities in this case are denoted as $C_{\sigma,0}^{[1],\text{ind}}(\mathcal{A}, \alpha_{\text{ind}}\mathcal{A})$ for $\frac{\mathcal{A}}{\sigma} \to 0$ where $\mathcal{E}_{\text{ind}} = \alpha_{\text{ind}}\mathcal{A}$ and $\alpha_{\text{ind}} \in (0, 1)$, and as $C_{\sigma,0}^{[2],\text{ind}}(\mathcal{A}, \mathcal{E}_{\text{ind}})$ for $\frac{\mathcal{E}_{\text{ind}}}{\sigma} \to 0$ with fixed $\frac{\mathcal{A}}{\sigma}$. These asymptotic capacities follow as a byproduct of $C_{\sigma,0}^{[1]}(\mathcal{A}, \alpha\mathcal{A})$ and $C_{\sigma,0}^{[2]}(\mathcal{A}, \mathcal{E})$. The main results are given next.

III. MAIN RESULTS

Before presenting the low-SNR asymptotic capacity results, we start by defining the *maximally-correlated M-variate binary distribution*, in addition to introducing capacity bounds that hold for any SNR.

Definition 1 (Maximally Correlated M-Variate Binary Distribution): Fix $\boldsymbol{\mathcal{E}} \in [0, \mathcal{A}]^M$ and let $(a_1, \ldots, a_M)$ denote the indexes of the components of $\boldsymbol{\mathcal{E}}$ sorted in descending order,[3] and $\boldsymbol{u}_i = (u_{i,1}, \ldots, u_{i,M})^\mathsf{T}$, $i = 1, \ldots, M$, satisfy $u_{i,j} = \mathcal{A}$ for $j \in \{a_1, \ldots, a_i\}$ and $u_{i,j} = 0$ otherwise. We say that $\boldsymbol{x} \in [0, \mathcal{A}]^M$ with mean $\boldsymbol{\mathcal{E}}$ follows a maximally-correlated M-variate binary distribution if $\boldsymbol{x} \sim P_{\boldsymbol{x},\boldsymbol{\mathcal{E}}}^*(\boldsymbol{x})$ with

$$P_{\boldsymbol{x},\boldsymbol{\mathcal{E}}}^*(\boldsymbol{x}) = \left(1 - \frac{\mathcal{E}_{a_1}}{\mathcal{A}}\right)\delta(\boldsymbol{x}) + \sum_{i=1}^{M} \frac{\mathcal{E}_{a_i} - \mathcal{E}_{a_{i+1}}}{\mathcal{A}}\delta(\boldsymbol{x} - \boldsymbol{u}_i)$$

[3]Ties are resolved using a coin flip.

where $\delta(\cdot)$ is the Dirac delta and $\mathcal{E}_{a_{M+1}} = 0$.

Such a distribution has the largest pairwise covariance among all distributions of $\boldsymbol{x} \in \{0, \mathcal{A}\}^M$ with mean $\boldsymbol{\mathcal{E}}$ given by $\min\{\mathcal{E}_i, \mathcal{E}_j\}(\mathcal{A} - \max\{\mathcal{E}_i, \mathcal{E}_j\})$ (see Appendix A). For $\boldsymbol{x} \in \{0, \mathcal{A}\}^M$ with mean $\boldsymbol{\mathcal{E}}$, the variances of x_i and x_j are fixed, and maximum covariance implies maximum correlation. Thus, $P_{\boldsymbol{x},\boldsymbol{\mathcal{E}}}^*(\boldsymbol{x})$ has maximum pairwise correlation, and hence the name. The general structure of $P_{\boldsymbol{x},\boldsymbol{\mathcal{E}}}^*(\boldsymbol{x})$ is shown in Table I. Using $P_{\boldsymbol{x},\boldsymbol{\mathcal{E}}}^*(\boldsymbol{x})$, the achievable rate is given by

$$R_\sigma^*(\mathcal{A}, \boldsymbol{\mathcal{E}}) \triangleq \mathsf{I}(\boldsymbol{x}; \boldsymbol{y})|_{\boldsymbol{x} \sim P_{\boldsymbol{x},\boldsymbol{\mathcal{E}}}^*(\boldsymbol{x})}. \qquad (6)$$

This distribution forms the basis for proving the achievability results throughout the paper. The following lemma forms the basis for converse results.

Lemma 1: The capacity $C_\sigma(\mathcal{A}, \mathcal{E})$ satisfies

$$C_\sigma(\mathcal{A}, \mathcal{E}) \leq \max_{\boldsymbol{Q} \in \mathcal{Q}} \frac{1}{2} \log|\sigma^{-2}\boldsymbol{H}\boldsymbol{Q}\boldsymbol{H}^\mathsf{T} + \boldsymbol{I}_N|, \qquad (7)$$

where $\mathcal{Q}$ is the set of covariance matrices of $\boldsymbol{x} \in [0, \mathcal{A}]^M$ with $\sum_{i=1}^M \mathcal{E}_i \leq \mathcal{E}$.

Proof: The proof is based on the "Gaussian maximizes entropy under a covariance constraint" principle, and is detailed in Sec. IV-A. ∎

This upper bound extends [4, (11)] and [31, (43)] to the MIMO channel. The optimization problem in Lemma 1 is convex. First, $\log|\sigma^{-2}\boldsymbol{H}\boldsymbol{Q}\boldsymbol{H}^\mathsf{T} + \boldsymbol{I}_N|$ is concave with respect to positive semi-definite $\boldsymbol{Q}$. Moreover, $\mathcal{Q}$ is a convex set since if $\boldsymbol{Q}_1, \boldsymbol{Q}_2 \in \mathcal{Q}$ corresponding to distributions $P_{\boldsymbol{x}}^{[1]}(\boldsymbol{x})$ and $P_{\boldsymbol{x}}^{[2]}(\boldsymbol{x})$, respectively, with $\sum_{i=1}^M \mathcal{E}_i \leq \mathcal{E}$, then $\mu\boldsymbol{Q}_1 + (1-\mu)\boldsymbol{Q}_2 = \boldsymbol{Q}_3 \in \mathcal{Q}$ since there exists a distribution with covariance matrix $\boldsymbol{Q}_3$ and $\sum_{i=1}^M \mathcal{E}_i \leq \mathcal{E}$, namely, $\mu P_{\boldsymbol{x}}^{[1]}(\boldsymbol{x}) + (1-\mu)P_{\boldsymbol{x}}^{[2]}(\boldsymbol{x})$. However, evaluating the bound in Lemma 1 is difficult due to the lack of a full description of $\mathcal{Q}$.

Nevertheless, since $\sigma^{-2}\boldsymbol{H}\boldsymbol{Q}\boldsymbol{H}^\mathsf{T} + \boldsymbol{I}_N$ is positive definite, we have that

$$\log|\sigma^{-2}\boldsymbol{H}\boldsymbol{Q}\boldsymbol{H}^\mathsf{T} + \boldsymbol{I}_N| \leq \operatorname{Tr}(\sigma^{-2}\boldsymbol{H}\boldsymbol{Q}\boldsymbol{H}^\mathsf{T}). \qquad (8)$$

This simplifies the upper bound in (7) and also the optimization involved, and leads to a capacity upper bound that is easily computable.

Theorem 1: The capacity of the MIMO IM-DD channel with $\boldsymbol{x} \in [0, \mathcal{A}]^M$ and $\sum_{i=1}^M \mathcal{E}_i \leq \mathcal{E}$ is bounded by

$$C_\sigma(\mathcal{A}, \mathcal{E}) \leq \max_{\boldsymbol{\mathcal{E}} \in \mathcal{S}} \phi_\sigma(\mathcal{A}, \boldsymbol{\mathcal{E}}) \qquad (9)$$

where $\mathcal{S} = \{\boldsymbol{\mathcal{E}} \in [0, \mathcal{A}]^M \mid \sum_{i=1}^M \mathcal{E}_i \leq \mathcal{E}\}$ and

$$\phi_\sigma(\mathcal{A}, \boldsymbol{\mathcal{E}}) = \frac{1}{2\sigma^2} \sum_{i=1}^M \sum_{j=1}^M \boldsymbol{h}_i^\mathsf{T}\boldsymbol{h}_j \min\{\mathcal{E}_i, \mathcal{E}_j\}(\mathcal{A} - \max\{\mathcal{E}_i, \mathcal{E}_j\}). \qquad (10)$$

Proof: This bound is based on Lemma 1, (8), and showing that $P_{\boldsymbol{x},\boldsymbol{\mathcal{E}}}^*(\boldsymbol{x})$ defined in Definition 1 maximizes the right-hand side of (8) with respect to $\boldsymbol{Q}$ for a given $\boldsymbol{\mathcal{E}}$. Details are given in Sec. IV-B. ∎

Remark 1: We note that the optimization in (9) is convex. In particular, the Hessian of the function $\phi_\sigma(\mathcal{A}, \boldsymbol{\mathcal{E}})$ with respect

TABLE I: Maximally correlated distribution of $\boldsymbol{x} \in \{0, \mathcal{A}\}^M$ with $\mathcal{E}_1 \geq \mathcal{E}_2 \geq \cdots \geq \mathcal{E}_M$, showing only realizations with nonzero probability.

$\boldsymbol{x}^{\mathsf{T}}$	$(0,0,0,\ldots,0,0)$	$(\mathcal{A},0,0,\ldots,0,0)$	$(\mathcal{A},\mathcal{A},0,\ldots,0,0)$	$\cdots$	$(\mathcal{A},\mathcal{A},\mathcal{A},\ldots,\mathcal{A},0)$	$(\mathcal{A},\mathcal{A},\mathcal{A},\ldots,\mathcal{A},\mathcal{A})$
$P_{\boldsymbol{x},\boldsymbol{\mathcal{E}}}^*(\boldsymbol{x})$	$1 - \frac{\mathcal{E}_1}{\mathcal{A}}$	$\frac{\mathcal{E}_1 - \mathcal{E}_2}{\mathcal{A}}$	$\frac{\mathcal{E}_2 - \mathcal{E}_3}{\mathcal{A}}$	$\cdots$	$\frac{\mathcal{E}_{M-1} - \mathcal{E}_M}{\mathcal{A}}$	$\frac{\mathcal{E}_M}{\mathcal{A}}$

to $\boldsymbol{\mathcal{E}}$ is $\frac{-1}{\sigma^2} \boldsymbol{H}^{\mathsf{T}} \boldsymbol{H}$ which is negative semidefinite. Thus, the intensity allocation can be solved using standard solvers [39].

The bound in Theorem 1 is asymptotically tight at low SNR leading to the low-SNR asymptotic capacity as given next. We start with the case of proportional average and peak SNRs.

A. Proportional Average and Peak SNRs

In this case, $\mathcal{E} = \alpha\mathcal{A}$, and asymptotically low SNR is defined as $\frac{\mathcal{A}}{\sigma} \to 0$ which implies $\frac{\mathcal{E}}{\sigma} \to 0$. The low-SNR asymptotic capacity in this case is given next.

Theorem 2 (Proportional SNRs, $\frac{\mathcal{A}}{\sigma} \to 0$): The low-SNR asymptotic capacity of the MIMO IM-DD channel with $\boldsymbol{x} \in [0, \mathcal{A}]^M$ and $\sum_{i=1}^M \mathcal{E}_i \leq \mathcal{E} = \alpha\mathcal{A}$ is given by

$$C_{\sigma,0}^{[1]}(\mathcal{A}, \alpha\mathcal{A}) = \frac{\mathcal{A}^2}{2\sigma^2}\omega(\boldsymbol{H}, \alpha) \qquad (11)$$

achievable using a maximally-correlated M-variate binary input distribution, where

$$\omega(\boldsymbol{H}, \alpha) = \max_{\boldsymbol{\alpha} \in \bar{\mathcal{S}}} \sum_{i=1}^M \sum_{j=1}^M \boldsymbol{h}_i^{\mathsf{T}} \boldsymbol{h}_j \min\{\alpha_i, \alpha_j\}(1 - \max\{\alpha_i, \alpha_j\}), \qquad (12)$$

and $\bar{\mathcal{S}} = \{\boldsymbol{\alpha} \in [0,1]^M \mid \sum_{i=1}^M \alpha_i \leq \alpha\}$.

Proof: The converse follows by rewriting the upper bound in Theorem 1 as $\frac{\mathcal{A}^2}{2\sigma^2}\omega(\boldsymbol{H}, \alpha)$. The achievability follows using results in [28] to evaluate $R_\sigma^*(\mathcal{A}, \boldsymbol{\mathcal{E}})$ at asymptotically low SNR and is detailed in Sec. IV-C. Another proof that relies on the relation between the mutual information and the minimum mean-square error over a Gaussian channel [30] is given in Appendix B. ∎

This shows that the low-SNR asymptotic capacity $C_{\sigma,0}^{[1]}(\mathcal{A}, \alpha\mathcal{A})$ scales linearly with $\frac{\mathcal{A}^2}{\sigma^2}$ with a slope $\omega(\boldsymbol{H}, \alpha)$. This slope can be obtained using standard convex optimization solvers, since (12) is a convex problem as the Hessian of its objective function is $-2\boldsymbol{H}^{\mathsf{T}}\boldsymbol{H}$ which is negative semidefinite.

Remark 2: The low-SNR asymptotic capacity of the MISO case given in [31, Proposition 12] can be obtained from Theorem 2 as a special case when $N = 1$.

Note that equal intensity allocation, i.e., $\boldsymbol{\alpha} = \frac{\alpha}{M}\mathbf{1}_M$ is generally not the optimal solution of (12). For instance, if $\|\boldsymbol{h}_1\| \ll \|\boldsymbol{h}_i\|$ for all $i > 1$, then one can see that it is better to not rely on aperture 1 in the transmission but rather set $\alpha_1 \ll \alpha_i$ in (12). Nevertheless, in some cases, $\boldsymbol{\alpha} = \frac{\alpha}{M}\mathbf{1}_M$ is the optimal solution of (12). One such case is when $\alpha \geq \frac{M}{2}$. In this case, the average intensity constraint is redundant, and the channel is similar to one with peak constraints only. The optimal solution of (12) is then $\alpha_i = \frac{1}{2} \; \forall i$, leading to the following corollary.

Corollary 1: If $\alpha \geq \frac{M}{2}$, then the low-SNR asymptotic capacity is given by

$$C_{\sigma,0}^{[1]}(\mathcal{A}, \alpha\mathcal{A}) = C_{\sigma,0}^{[1]}(\mathcal{A}, M\mathcal{A}) = \frac{\mathcal{A}^2}{8\sigma^2}\|\boldsymbol{H}\mathbf{1}_M\|^2. \qquad (13)$$

Proof: Note that $\min\{\alpha_i, \alpha_j\}(1 - \max\{\alpha_i, \alpha_j\}) \leq \max\{\alpha_i, \alpha_j\}(1 - \max\{\alpha_i, \alpha_j\}) \leq \frac{1}{4}$. This upper bound is achieved using $\alpha_i = \frac{1}{2} \; \forall i$ which satisfies $\sum_{i=1}^M \alpha_i \leq \alpha$ if $\alpha \geq \frac{M}{2}$. Using this in (12) and noting that $\sum_{i=1}^M \sum_{j=1}^M \boldsymbol{h}_i^{\mathsf{T}} \boldsymbol{h}_j = \|\boldsymbol{H}\mathbf{1}_M\|^2$ concludes the proof. ∎

Corollary 1 shows that using *coded OOK, spatial repetition, and MRC* is optimal at asymptotically low-SNR in the MIMO IM-DD channel with peak constraints only. To see this, let $\boldsymbol{x} = \mathcal{A}\mathbf{1}_M s_p$ where $s_p \sim \text{Bern}(p)$ (OOK and spatial repetition), and let the receiver apply MRC to obtain $\frac{\mathbf{1}_M^{\mathsf{T}} \boldsymbol{H}^{\mathsf{T}}}{\|\boldsymbol{H}\mathbf{1}_M\|}\boldsymbol{y}$ from which it decodes s_p. The resulting achievable rate is

$$R_\sigma^{\text{orm}}(\mathcal{A}, p) = \mathsf{I}\left(s_p; \mathcal{A}\|\boldsymbol{H}\mathbf{1}_M\|s_p + \frac{\mathbf{1}_M^{\mathsf{T}}\boldsymbol{H}^{\mathsf{T}}}{\|\boldsymbol{H}\mathbf{1}_M\|}\boldsymbol{z}\right). \qquad (14)$$

This is the achievable rate using OOK over a SISO IM-DD channel with a peak constraint $\mathcal{A}$, channel gain $\|\boldsymbol{H}\mathbf{1}_M\|$, and noise variance σ^2. Setting $p = \frac{1}{2}$, and using similar analysis as in Sec. IV-C, this achievable rate can be written as $R_\sigma^{\text{orm}}(\mathcal{A}, \frac{1}{2}) = (1 + o_{\frac{\mathcal{A}}{\sigma}}(1))\frac{\mathcal{A}^2}{8\sigma^2}\|\boldsymbol{H}\mathbf{1}_M\|^2$. This coincides with $C_{\sigma,0}^{[1]}(\mathcal{A}, M\mathcal{A})$ given in Corollary 1 at asymptotically low SNR.

Fig. 2 shows the upper bound given in Theorem 1 versus $\frac{\mathcal{A}}{\sigma}$ for an exemplary MIMO channel with $M = N = 3$, and

$$\boldsymbol{H} = \begin{bmatrix} 1 & 0.4 & 0.2 \\ 0.4 & 1 & 0.4 \\ 0.2 & 0.4 & 1 \end{bmatrix}. \qquad (15)$$

In Fig. 2a, we set $\alpha = 0.2M$ and compare the upper bound with the rate achieved using a maximally-correlated M-variate binary distribution, i.e., $\max_{\boldsymbol{\mathcal{E}}} R_\sigma^*(\mathcal{A}, \boldsymbol{\mathcal{E}})$ defined in (6). The figure clearly shows that the two curves coincide with the low-SNR asymptotic capacity as SNR approaches zero verifying Theorem 2. Fig. 2b shows the same performance under $\alpha = \frac{M}{2}$, which is equivalent to a channel with peak constraints only. Additionally, Fig. 2b shows $R_\sigma^{\text{orm}}(\mathcal{A}, p)$ given in (14), which is the achievable rate using OOK, spatial repetition, and MRC. This plot verifies Corollary 1.
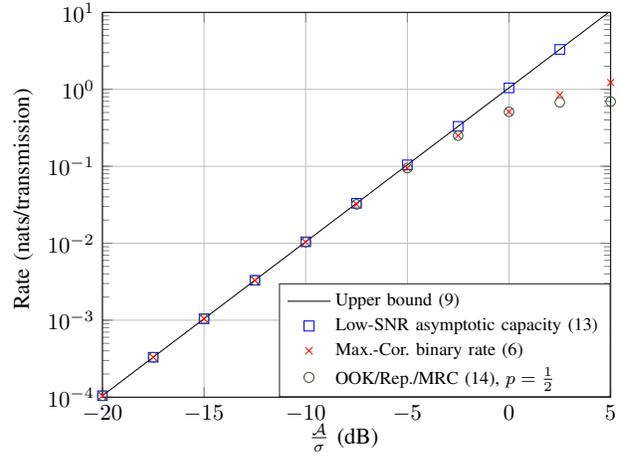
Similar statements can be derived under a fixed peak SNR, although the low-SNR asymptotic capacity can be simplified further as given next.

B. Fixed Peak SNR

Here, we consider the asymptotic low SNR regime where $\frac{\mathcal{A}}{\sigma}$ is fixed and $\frac{\mathcal{E}}{\sigma} \to 0$. In this case, using OOK with $p = \frac{\mathcal{E}}{M\mathcal{A}}$, spatial repetition, and MRC leads to the achievable
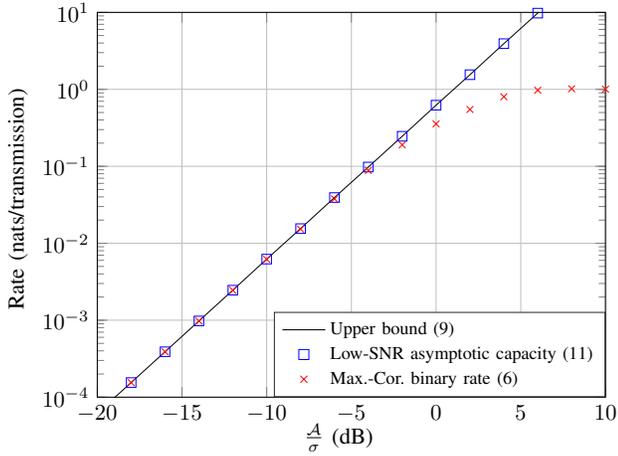
(a) $\alpha = 0.2M$.

(b) $\alpha = 0.5M$. (equivalently, peak constraint only).

Fig. 2: Achievable rates and upper bounds versus $\frac{A}{\sigma}$ for a 3×3 MIMO IM-DD channel with $\boldsymbol{H}$ as given in (15) and $\mathcal{E} = \alpha \mathcal{A}$.
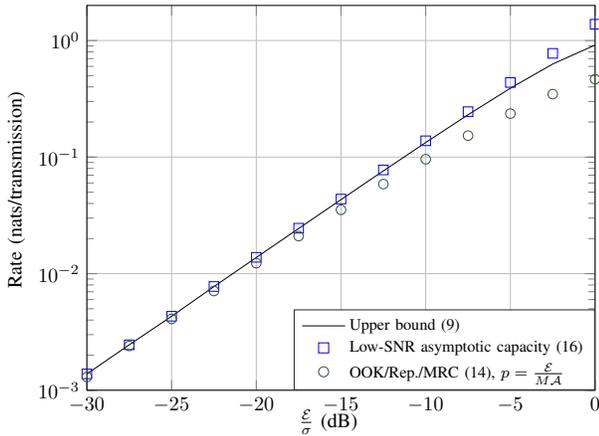

Fig. 3: Achievable rates and upper bounds versus $\frac{\mathcal{E}}{\sigma}$ for a 3×3 MIMO IM-DD channel with $\boldsymbol{H}$ as given in (15) and fixed $\frac{A}{\sigma} = 1$.

rate $R_\sigma^{\text{orm}}\left(\mathcal{A}, \frac{\mathcal{E}}{M\mathcal{A}}\right)$ (14), which coincides with the low-SNR asymptotic capacity as stated next.

Theorem 3 (Fixed Peak SNR, $\frac{\mathcal{E}}{\sigma} \to 0$): The low-SNR asymptotic capacity of the MIMO IM-DD channel with $\boldsymbol{x} \in [0, \mathcal{A}]^M$ and $\sum_{i=1}^M \mathcal{E}_i \leq \mathcal{E}$, with fixed $\frac{A}{\sigma}$ and $\frac{\mathcal{E}}{\sigma} \to 0$, is given by

$$C_{\sigma,0}^{[2]}(\mathcal{A}, \mathcal{E}) = \frac{\mathcal{A}\mathcal{E}}{2M\sigma^2}\|\boldsymbol{H}\mathbf{1}_M\|^2, \qquad (16)$$

achievable using OOK, spatial repetition coding, and MRC.

Proof: The statement follows using Theorem 1, applying results in [28] to (14), and using the definition of $C_{\sigma,0}^{[2]}(\mathcal{A}, \mathcal{E})$ given in Sec. II. Details are given in Sec. IV-D. ∎
This shows that the low-SNR asymptotic capacity $C_{\sigma,0}^{[2]}(\mathcal{A}, \mathcal{E})$ scales linearly with $\frac{\mathcal{E}}{\sigma}$ with a slope of $\frac{A}{2M\sigma}\|\boldsymbol{H}\mathbf{1}_M\|^2$.
Fig. 3 shows the upper bound given in Theorem 1 versus $\frac{\mathcal{E}}{\sigma}$ for an exemplary MIMO channel with $M = N = 3$, $\boldsymbol{H}$ as given in (15), and $\frac{A}{\sigma} = 1$. The figure also shows the low-SNR asymptotic capacity in Theorem 3 and the achievable rate using OOK, spatial repetition coding, and

MRC $R_\sigma^{\text{orm}}\left(\mathcal{A}, \frac{\mathcal{E}}{M\mathcal{A}}\right)$ (14). All three converge as $\frac{\mathcal{E}}{\sigma}$ decreases, which demonstrates Theorem 3.

A byproduct of these results is the low-SNR capacity of the channel under individual average intensity constraints. This is given next.

C. Individual Average Constraints

The function $\phi_\sigma(\mathcal{A}, \boldsymbol{\mathcal{E}})$ can be interpreted as the low-SNR asymptotic capacity of the channel subject to the constraints $\boldsymbol{x} \in [0, \mathcal{A}]^M$ and $\mathbb{E}_{P_{\boldsymbol{x}}}[\boldsymbol{x}] = \boldsymbol{\mathcal{E}}$. This can be used to obtain the following corollaries on the low-SNR asymptotic capacity under individual average intensity constraints.

Corollary 2 (Proportional SNRs, $\frac{A}{\sigma} \to 0$): The low-SNR asymptotic capacity of the MIMO IM-DD channel with $\boldsymbol{x} \in [0, \mathcal{A}]^M$, $\mathcal{E}_i \leq \mathcal{E}_{\text{ind}} = \alpha_{\text{ind}}\mathcal{A}$ for $i \in \{1, \ldots, M\}$, and $\alpha_{\text{ind}} \in (0, 1)$ is

$$C_{\sigma,0}^{[1],\text{ind}}(\mathcal{A}, \alpha_{\text{ind}}\mathcal{A}) = \frac{\mathcal{A}^2}{2\sigma^2}\bar{\alpha}_{\text{ind}}(1 - \bar{\alpha}_{\text{ind}})\|\boldsymbol{H}\mathbf{1}_M\|^2, \quad (17)$$

where $\bar{\alpha}_{\text{ind}} = \min\{\alpha_{\text{ind}}, \frac{1}{2}\}$.

Proof: This is obtained similar to Theorem 2, with the set of feasible $\boldsymbol{\alpha}$ replaced with $[0, \alpha_{\text{ind}}]^M$. This leads to the result as detailed in Sec. IV-E1. ∎
Here, if $\alpha_{\text{ind}} \geq \frac{1}{2}$, then the average constraint is redundant, and the low-SNR asymptotic capacity coincides with that given in Corollary 1.
For a fixed peak SNR, we have the following.
Corollary 3 (Fixed Peak SNR, $\frac{\mathcal{E}_{\text{ind}}}{\sigma} \to 0$): The low-SNR asymptotic capacity of the MIMO IM-DD channel with $\boldsymbol{x} \in [0, \mathcal{A}]^M$ and $\mathcal{E}_i \leq \mathcal{E}_{\text{ind}}$ with fixed $\frac{A}{\sigma}$ and $\frac{\mathcal{E}_{\text{ind}}}{\sigma} \to 0$ is

$$C_{\sigma,0}^{[2],\text{ind}}(\mathcal{A}, \mathcal{E}_{\text{ind}}) = \frac{\mathcal{A}\mathcal{E}_{\text{ind}}}{2\sigma^2}\|\boldsymbol{H}\mathbf{1}_M\|^2. \qquad (18)$$

Proof: This can be proved similar to Theorem 3. See Sec. IV-E2 for details. ∎
Both these corollaries show that *coded OOK, spatial repetition, and MRC are optimal at asymptotically low SNR under individual average constraints*. In particular, the achievable rate $R_\sigma^{\text{orm}}(\mathcal{A}, \bar{\alpha}_{\text{ind}})$ (14) can be expressed as $(1 +$
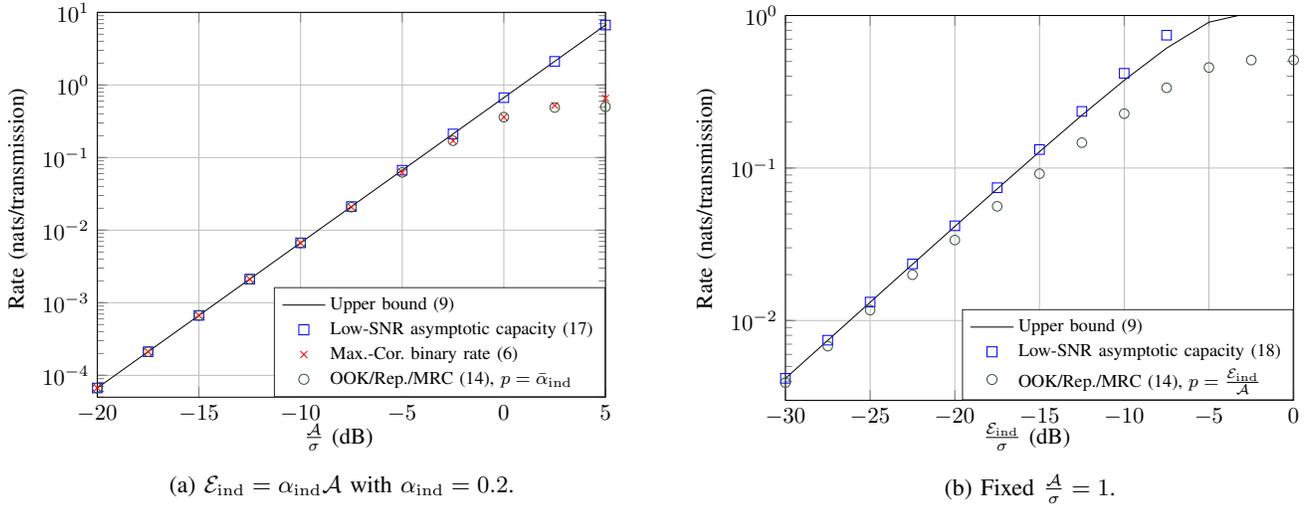
(a) $\mathcal{E}_{\mathrm{ind}} = \alpha_{\mathrm{ind}}\mathcal{A}$ with $\alpha_{\mathrm{ind}} = 0.2$.

(b) Fixed $\frac{\mathcal{A}}{\sigma} = 1$.

Fig. 4: Achievable rates and upper bounds versus SNR for a 3×3 MIMO IM-DD channel with $\boldsymbol{H}$ as given in (15).

$o_{\frac{\mathcal{A}}{\sigma}}(1)\frac{\mathcal{A}^2}{2\sigma^2}\bar{\alpha}_{\mathrm{ind}}(1-\bar{\alpha}_{\mathrm{ind}})\|\boldsymbol{H}\boldsymbol{1}_M\|^2$ using [28]. This coincides with $C_{\sigma,0}^{[1],\mathrm{ind}}(\mathcal{A}, \alpha_{\mathrm{ind}}\mathcal{A})$ in Corollary 2 at asymptotically low SNR. Similarly, the achievable rate $R_\sigma^{\mathrm{orm}}\left(\mathcal{A}, \frac{\mathcal{E}_{\mathrm{ind}}}{\mathcal{A}}\right)$ can be written as $(1 + o_{\frac{\mathcal{E}_{\mathrm{ind}}}{\sigma}}(1))\frac{\mathcal{E}_{\mathrm{ind}}\mathcal{A}}{2\sigma^2}\|\boldsymbol{H}\boldsymbol{1}_M\|^2$, which coincides with $C_{\sigma,0}^{[2],\mathrm{ind}}(\mathcal{A}, \mathcal{E}_{\mathrm{ind}})$ in Corollary 3 at asymptotically low SNR.

Fig. 4a shows the upper bound in Theorem 1 for an exemplary MIMO channel with $M = N = 3$ and $\boldsymbol{H}$ as given in (15), with $\alpha_{\mathrm{ind}} = 0.2$. We compare this with the achievable rate using coded OOK, spatial repetition, and MRC, and using a maximally-correlated M-variate binary distribution given by $\max_{\boldsymbol{\mathcal{E}} \in (0, \min\{\mathcal{E}_{\mathrm{ind}}, \frac{\mathcal{A}}{2}\}]^M} R_\sigma^*(\mathcal{A}, \boldsymbol{\mathcal{E}})$ (6). Note that the three converge to the low-SNR asymptotic capacity given in Corollary 2 as $\frac{\mathcal{A}}{\sigma}$ decreases. Fig. 4b shows a comparison between the upper bound in theorem 1 with the achievable rate using OOK, repetition coding, and MRC for the same channel with fixed $\frac{\mathcal{A}}{\sigma} = 1$. The two converge with the low-SNR asymptotic capacity in Corollary 3 as $\frac{\mathcal{E}_{\mathrm{ind}}}{\sigma} \to 0$ which confirms the above discussion.

Next, we give the proofs of these results.

IV. CAPACITY ANALYSIS

A. Proof of Lemma 1

We start by representing $C_\sigma(\mathcal{A}, \mathcal{E})$ as

$$C_\sigma(\mathcal{A}, \mathcal{E}) = \max_{P_{\boldsymbol{x}} \in \mathcal{P}} \mathsf{I}(\boldsymbol{x}; \boldsymbol{y}) \quad (19)$$

$$= \max_{P_{\boldsymbol{x}} \in \mathcal{P}} \mathsf{h}(\boldsymbol{H}\boldsymbol{x} + \boldsymbol{z}) - \mathsf{h}(\boldsymbol{z}), \quad (20)$$

where $\mathcal{P}$ is the collection of all probability measures $P_{\boldsymbol{x}}$ of $\boldsymbol{x} \in [0, \mathcal{A}]^M$ satisfying $\sum_{i=1}^M \mathcal{E}_i \leq \mathcal{E}$. Let $\mathcal{P}_{\boldsymbol{Q}}$ be the subset of $\mathcal{P}$ with covariance matrix $\boldsymbol{Q}$. Then, we can write $\mathcal{P} = \cup_{\boldsymbol{Q} \in \mathcal{Q}} \mathcal{P}_{\boldsymbol{Q}}$ where $\mathcal{Q}$ is the collection of covariance matrices of $P_{\boldsymbol{x}} \in \mathcal{P}$. For $P_{\boldsymbol{x}} \in \mathcal{P}_{\boldsymbol{Q}}$, we have that $\boldsymbol{H}\boldsymbol{x} + \boldsymbol{z}$ has a covariance matrix $\boldsymbol{H}\boldsymbol{Q}\boldsymbol{H}^\mathsf{T} + \sigma^2 \boldsymbol{I}_N$, and thus, $\mathsf{h}(\boldsymbol{H}\boldsymbol{x} + \boldsymbol{z}) \leq \frac{1}{2}\log\left((2\pi e)^N |\boldsymbol{H}\boldsymbol{Q}\boldsymbol{H}^\mathsf{T} + \sigma^2 \boldsymbol{I}_N|\right)$ since the Gaussian distribution maximizes the differential entropy under a covariance

constraint [38]. Moreover, $\mathsf{h}(\boldsymbol{z}) = \frac{1}{2}\log\left((2\pi e)^N |\sigma^2 \boldsymbol{I}_N|\right)$. Hence,

$$C_\sigma(\mathcal{A}, \mathcal{E}) \leq \max_{\boldsymbol{Q} \in \mathcal{Q}} \frac{1}{2}\log|\sigma^{-2}\boldsymbol{H}\boldsymbol{Q}\boldsymbol{H}^\mathsf{T} + \boldsymbol{I}_N|, \quad (21)$$

which proves the statement of Lemma 1.

B. Proof of Theorem 1

By Lemma 1, we have

$$C_\sigma(\mathcal{A}, \mathcal{E}) \leq \max_{\boldsymbol{Q} \in \mathcal{Q}} \frac{1}{2}\log|\sigma^{-2}\boldsymbol{H}\boldsymbol{Q}\boldsymbol{H}^\mathsf{T} + \boldsymbol{I}_N|. \quad (22)$$

Since $\log|\boldsymbol{A}| \leq \mathrm{Tr}(\boldsymbol{A} - \boldsymbol{I}_N)$ for a positive definite matrix $\boldsymbol{A} \in \mathbb{R}^{N \times N}$, and since $\sigma^{-2}\boldsymbol{H}\boldsymbol{Q}\boldsymbol{H}^\mathsf{T} + \boldsymbol{I}_N$ is positive definite, then we can write

$$C_\sigma(\mathcal{A}, \mathcal{E}) \leq \max_{\boldsymbol{Q} \in \mathcal{Q}} \frac{1}{2}\mathrm{Tr}(\sigma^{-2}\boldsymbol{H}\boldsymbol{Q}\boldsymbol{H}^\mathsf{T}) = \max_{\boldsymbol{Q} \in \mathcal{Q}} \frac{1}{2\sigma^2}\mathrm{Tr}(\boldsymbol{H}\boldsymbol{Q}\boldsymbol{H}^\mathsf{T}). \quad (23)$$

Let us write $\mathcal{Q} = \cup_{\boldsymbol{\mathcal{E}} \in \mathcal{S}} \mathcal{Q}_{\boldsymbol{\mathcal{E}}}$ where $\mathcal{S} = \{\boldsymbol{\mathcal{E}} \in [0, \mathcal{A}]^M | \sum_{i=1}^M \mathcal{E}_i \leq \mathcal{E}\}$ and $\mathcal{Q}_{\boldsymbol{\mathcal{E}}}$ is the subset of $\mathcal{Q}$ corresponding to distributions with mean $\boldsymbol{\mathcal{E}}$. Then, we can write

$$C_\sigma(\mathcal{A}, \mathcal{E}) \leq \max_{\boldsymbol{\mathcal{E}} \in \mathcal{S}} \max_{\boldsymbol{Q} \in \mathcal{Q}_{\boldsymbol{\mathcal{E}}}} \frac{1}{2\sigma^2}\mathrm{Tr}(\boldsymbol{H}\boldsymbol{Q}\boldsymbol{H}^\mathsf{T}). \quad (24)$$

Now the proof proceeds along the following steps. First, we show that the inner maximization is solved by a $\boldsymbol{Q}$ corresponding to $\boldsymbol{x} \in \{0, \mathcal{A}\}^M$, i.e., M-variate binary. For a given $\boldsymbol{\mathcal{E}}$, this determines the diagonal of $\boldsymbol{Q}$. Then, we show that its off-diagonal components are equal to the maximum covariance between binary random variables with the predefined mean $\boldsymbol{\mathcal{E}}$. This allows us to upper bound capacity using a simple expression.

Consider any $P_{\boldsymbol{x}}$ with a covariance matrix $\boldsymbol{Q} \in \mathcal{Q}_{\boldsymbol{\mathcal{E}}}$, and define $\boldsymbol{G} = \boldsymbol{H}^\mathsf{T}\boldsymbol{H}$. Then,

$$\mathrm{Tr}(\boldsymbol{H}\boldsymbol{Q}\boldsymbol{H}^\mathsf{T}) = \mathrm{Tr}(\boldsymbol{G}\boldsymbol{Q}) \quad (25)$$

$$= \sum_{i=1}^M \sum_{j=1}^M g_{i,j} q_{j,i} \quad (26)$$

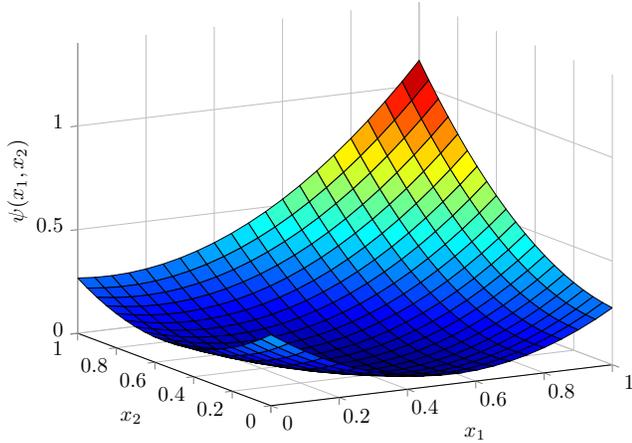$$= \mathbb{E}_{P_{\boldsymbol{x}}}[\psi(\boldsymbol{x})], \quad (27)$$

Fig. 5: The function $\psi(\boldsymbol{x})$ for $M = 2$, $\mathcal{A} = 1$, $\boldsymbol{\mathcal{E}} = [0.4,\ 0.3]^{\mathsf{T}}$ and $\boldsymbol{H} = [\boldsymbol{h}_1,\ \boldsymbol{h}_2]$ with $\boldsymbol{h}_1 = [1,\ 0.2]^{\mathsf{T}}$ and $\boldsymbol{h}_2 = [0.2,\ 0.8]^{\mathsf{T}}$. Note that for any $P_{\boldsymbol{x}}$, there exists a binary probability measure $P_{\boldsymbol{x}}^{\mathrm{b}}(\boldsymbol{x})$ for $\boldsymbol{x} \in \{0, \mathcal{A}\}^2$ with the same mean, for which $\mathbb{E}_{P_{\boldsymbol{x}}}[\psi(\boldsymbol{x})] \leq \mathbb{E}_{P_{\boldsymbol{x}}^{\mathrm{b}}}[\psi(\boldsymbol{x})]$. This $P_{\boldsymbol{x}}^{\mathrm{b}}$ can be obtained from $P_{\boldsymbol{x}}$ by 'concentrating' the probability mass at the vertices of $[0, \mathcal{A}]^2$ while preserving the mean as detailed in Appendix C.

where $q_{j,i}$ is the component of $\boldsymbol{Q}$ in the jth row and i^{th} column, $g_{j,i}$ is defined similarly with respect to $\boldsymbol{G}$, and

$$\psi(\boldsymbol{x}) = \sum_{i=1}^{M} \sum_{j=1}^{M} g_{i,j}(x_i - \mathcal{E}_i)(x_j - \mathcal{E}_j), \tag{28}$$

where we used the definition of covariance. Due to the nonnegativity of $h_{i,j}$ and hence also $g_{i,j}$, the function $\psi(\boldsymbol{x})$ is a convex elliptic paraboloid with minimum at $\boldsymbol{x} = \boldsymbol{\mathcal{E}}$. For fixed $(x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_M)$, the function $\psi(\boldsymbol{x})$ increases monotonically as x_i deviates from $\mathcal{E}_i$ towards the boundaries $x_i = 0$ and $x_i = \mathcal{A}$. This depicted in Fig. 5 for a special example with $M = 2$. Due to this, we have the following.

Proposition 1: For any $P_{\boldsymbol{x}}$, we can construct a probability measure $P_{\boldsymbol{x}}^{\mathrm{b}}(\boldsymbol{x})$ over the support $\{0, \mathcal{A}\}^M$ so that $\mathbb{E}_{P_{\boldsymbol{x}}^{\mathrm{b}}}[\boldsymbol{x}] = \mathbb{E}_{P_{\boldsymbol{x}}}[\boldsymbol{x}]$ and $\mathbb{E}_{P_{\boldsymbol{x}}}[\psi(\boldsymbol{x})] \leq \mathbb{E}_{P_{\boldsymbol{x}}^{\mathrm{b}}}[\psi(\boldsymbol{x})]$.

Proof: This construction is detailed in Appendix C. ∎

Consequently, we can replace the maximization with respect to $\boldsymbol{Q} \in \mathcal{Q}_{\boldsymbol{\mathcal{E}}}$ with a maximization with respect to $\boldsymbol{Q} \in \mathcal{Q}_{\boldsymbol{\mathcal{E}}}^{\mathrm{b}}$, where $\mathcal{Q}_{\boldsymbol{\mathcal{E}}}^{\mathrm{b}}$ is the set of covariance matrices of $\boldsymbol{x} \in \{0, \mathcal{A}\}^M$ with $\mathbb{E}_{P_{\boldsymbol{x}}}[\boldsymbol{x}] = \boldsymbol{\mathcal{E}}$. Thus,

$$C_\sigma(\mathcal{A}, \mathcal{E}) \leq \max_{\boldsymbol{\mathcal{E}} \in \mathcal{S}} \max_{\boldsymbol{Q} \in \mathcal{Q}_{\boldsymbol{\mathcal{E}}}^{\mathrm{b}}} \frac{1}{2\sigma^2} \mathrm{Tr}(\boldsymbol{G}\boldsymbol{Q}). \tag{29}$$

Since $\boldsymbol{Q} \in \mathcal{Q}_{\boldsymbol{\mathcal{E}}}^{\mathrm{b}}$, then $q_{i,i} = \mathcal{E}_i(\mathcal{A} - \mathcal{E}_i)$. To obtain the off-diagonal elements of $\boldsymbol{Q}$, we use Lemma 2 given in Appendix A through which we can write $q_{i,j} \leq \min\{\mathcal{E}_i, \mathcal{E}_j\}(\mathcal{A} - \max\{\mathcal{E}_i, \mathcal{E}_j\})$. At this point, we know that $\boldsymbol{Q} \in \mathcal{Q}_{\boldsymbol{\mathcal{E}}}^{\mathrm{b}}$ satisfies

$$q_{i,i} = \mathcal{E}_i(\mathcal{A} - \mathcal{E}_i) \tag{30a}$$
$$q_{i,j} \leq \min\{\mathcal{E}_i, \mathcal{E}_j\}(\mathcal{A} - \max\{\mathcal{E}_i, \mathcal{E}_j\}). \tag{30b}$$

Noting that $\mathrm{Tr}(\boldsymbol{G}\boldsymbol{Q})$ is increasing in $q_{i,j}$ ($g_{i,j} \geq 0\ \ \forall i, j$), we conclude that the optimal $\boldsymbol{Q}$ should satisfy (30b) with equality,

i.e., the optimal $\boldsymbol{Q}$ corresponds to a maximally-correlated M-variate binary distribution. This leads to

$$C_\sigma(\mathcal{A}, \mathcal{E}) \tag{31}$$
$$\leq \max_{\boldsymbol{\mathcal{E}} \in \mathcal{S}} \frac{1}{2\sigma^2} \sum_{i=1}^{M} \sum_{j=1}^{M} g_{i,j} \min\{\mathcal{E}_i, \mathcal{E}_j\}(\mathcal{A} - \max\{\mathcal{E}_i, \mathcal{E}_j\}).$$

Since $g_{i,j} = \sum_{k=1}^{N} h_{k,i}h_{k,j} = \boldsymbol{h}_i^{\mathsf{T}}\boldsymbol{h}_j$, this proves the upper bound in Theorem 1.

Remark 3: In [31, Lemma 8] which considers the MISO case, it was shown that an input distribution that satisfies the structure of $P_{\boldsymbol{x}, \boldsymbol{\mathcal{E}}}^*(\boldsymbol{x})$ maximizes the variance of $\boldsymbol{h}^{\mathsf{T}}\boldsymbol{x}$ where $\boldsymbol{h}$ is the channel vector. The derivation herein shows that $P_{\boldsymbol{x}, \boldsymbol{\mathcal{E}}}^*(\boldsymbol{x})$ maximizes $\mathbb{E}_{P_{\boldsymbol{x}}}[\mathrm{Tr}(\boldsymbol{H}\boldsymbol{x}\boldsymbol{x}^{\mathsf{T}}\boldsymbol{H}^{\mathsf{T}})]$ (see (24)) for a given $\boldsymbol{\mathcal{E}}$, which reproduces [31, Lemma 8] as a special case when $N = 1$.

C. Proof of Theorem 2

We have the upper bound in (31) which can be rewritten as $\overline{C}_\sigma^{[1]}(\mathcal{A}, \alpha\mathcal{A}) = \frac{\mathcal{A}^2}{2\sigma^2}\omega(\boldsymbol{H}, \alpha)$ where

$$\omega(\boldsymbol{H}, \alpha) = \max_{\boldsymbol{\alpha} \in \bar{\mathcal{S}}} \sum_{i=1}^{M} \sum_{j=1}^{M} g_{i,j} \min\{\alpha_i, \alpha_j\}(1 - \max\{\alpha_i, \alpha_j\}),$$

and $\bar{\mathcal{S}} = \{\boldsymbol{\alpha} \in [0,1]^M \mid \sum_{i=1}^{M} \alpha_i \leq \alpha\}$. To show that this upper bound coincides with capacity at asymptotically low-SNR, we show the achievability of $\overline{C}_\sigma^{[1]}(\mathcal{A}, \alpha\mathcal{A})(1 + o_{\frac{\mathcal{A}}{\sigma}}(1))$ using a maximally-correlated M-variate binary input as follows.

Fix $\boldsymbol{\alpha} \in \bar{\mathcal{S}}$, and choose $\boldsymbol{x}$ to be distributed according to a maximally-correlated M-variate binary input distribution $P_{\boldsymbol{x}, \boldsymbol{\alpha}\mathcal{A}}^*(\boldsymbol{x})$ as defined in Definition 1. This $\boldsymbol{x}$ has a covariance matrix $\boldsymbol{Q}^*$ with

$$q_{i,j}^* = \mathcal{A}^2 \min\{\alpha_i, \alpha_j\}(1 - \max\{\alpha_i, \alpha_j\}). \tag{32}$$

Now let us scale the channel output $\boldsymbol{y}$ by $\frac{1}{\sigma}$ to obtain $\tilde{\boldsymbol{y}} = \boldsymbol{H}\tilde{\boldsymbol{x}} + \tilde{\boldsymbol{z}}$ where $\tilde{\boldsymbol{z}}$ is Gaussian with zero mean and covariance matrix $\boldsymbol{I}_N$, and $\tilde{\boldsymbol{x}} = \frac{\boldsymbol{x}}{\sigma}$. This channel is equivalent to the original channel in terms of its capacity. Moreover, $\tilde{\boldsymbol{x}}$ has covariance matrix $\tilde{\boldsymbol{Q}}^* = \frac{1}{\sigma^2}\boldsymbol{Q}^*$ with components $\tilde{q}_{i,j}^*$. This channel satisfies conditions A–F in [28], and hence [28, Theorem 1] applies. Hence, the achievable rate is given by

$$R_\sigma^*(\mathcal{A}, \boldsymbol{\alpha}\mathcal{A}) = \mathsf{I}(\tilde{\boldsymbol{x}}; \tilde{\boldsymbol{y}}) \tag{33}$$
$$= \frac{1}{2} \sum_{i=1}^{M} \sum_{j=1}^{M} \iota_{i,j} \tilde{q}_{i,j}^* + o(\mathrm{Tr}(\tilde{\boldsymbol{Q}}^*)), \tag{34}$$

where $\iota_{i,j}$ is the Fisher-information-type integral

$$\iota_{i,j} = \int_{\mathbb{R}^N} \frac{1}{P_{\tilde{\boldsymbol{y}}|\tilde{\boldsymbol{x}}}(\tilde{\boldsymbol{y}}|\tilde{\boldsymbol{x}})} \frac{\partial P_{\tilde{\boldsymbol{y}}|\tilde{\boldsymbol{x}}}(\tilde{\boldsymbol{y}}|\tilde{\boldsymbol{x}})}{\partial \tilde{x}_i} \frac{\partial P_{\tilde{\boldsymbol{y}}|\tilde{\boldsymbol{x}}}(\tilde{\boldsymbol{y}}|\tilde{\boldsymbol{x}})}{\partial \tilde{x}_j} d\tilde{\boldsymbol{y}}, \tag{35}$$

and $P_{\tilde{\boldsymbol{y}}|\tilde{\boldsymbol{x}}}(\tilde{\boldsymbol{y}}|\tilde{\boldsymbol{x}})$ is the distribution of $\tilde{\boldsymbol{y}}$ given $\tilde{\boldsymbol{x}}$ which is $(2\pi)^{-\frac{N}{2}} e^{-\frac{1}{2}\|\tilde{\boldsymbol{y}} - \boldsymbol{H}\tilde{\boldsymbol{x}}\|^2}$. Substituting in $\iota_{i,j}$, we can show that $\iota_{i,j} = \sum_{k=1}^{N} h_{k,i}h_{k,j} = g_{i,j}$ independent of $\tilde{\boldsymbol{x}}$ (see details in Appendix D). Therefore,

$$R_\sigma^*(\mathcal{A}, \boldsymbol{\alpha}\mathcal{A}) = \frac{1}{2} \sum_{i=1}^{M} \sum_{j=1}^{M} g_{i,j}\tilde{q}_{i,j}^* + o(\mathrm{Tr}(\tilde{\boldsymbol{Q}}^*)). \tag{36}$$

Note that

$$\sum_{i=1}^{M}\sum_{j=1}^{M} g_{i,j}\tilde{q}_{i,j}^{*} \geq \sum_{i=1}^{M} g_{i,i}\tilde{q}_{i,i}^{*} \geq \min_{i} g_{i,i}\mathrm{Tr}(\tilde{\boldsymbol{Q}}^{*}). \quad (37)$$

Thus,

$$0 \leq \frac{1}{\frac{1}{2}\sum_{i=1}^{M}\sum_{j=1}^{M} g_{i,j}\tilde{q}_{i,j}^{*}} \leq \frac{1}{\frac{1}{2}\min_{i} g_{i,i}\mathrm{Tr}(\tilde{\boldsymbol{Q}}^{*})}, \quad (38)$$

This implies that

$$0 \leq \left| \frac{o(\mathrm{Tr}(\tilde{\boldsymbol{Q}}^{*}))}{\frac{1}{2}\sum_{i=1}^{M}\sum_{j=1}^{M} g_{i,j}\tilde{q}_{i,j}^{*}} \right| \leq \left| \frac{o(\mathrm{Tr}(\tilde{\boldsymbol{Q}}^{*}))}{\frac{1}{2}\min_{i} g_{i,i}\mathrm{Tr}(\tilde{\boldsymbol{Q}}^{*})} \right|. \quad (39)$$

Now since $\frac{\mathcal{A}}{\sigma} \to 0$ implies that $\mathrm{Tr}(\tilde{\boldsymbol{Q}}^{*}) \to 0$, then $\frac{o(\mathrm{Tr}(\tilde{\boldsymbol{Q}}^{*}))}{\mathrm{Tr}(\tilde{\boldsymbol{Q}}^{*})} \to 0$ as $\frac{\mathcal{A}}{\sigma} \to 0$, and the upper bound above approaches zero as $\frac{\mathcal{A}}{\sigma} \to 0$. Therefore, we have that $\lim_{\frac{\mathcal{A}}{\sigma}\to 0} \frac{o(\mathrm{Tr}(\tilde{\boldsymbol{Q}}^{*}))}{\frac{1}{2}\sum_{i=1}^{M}\sum_{j=1}^{M} g_{i,j}\tilde{q}_{i,j}^{*}} = 0$, and we can write

$$R_{\sigma}^{*}(\mathcal{A}, \boldsymbol{\alpha}\mathcal{A})$$
$$= \left(\frac{1}{2}\sum_{i=1}^{M}\sum_{j=1}^{M} g_{i,j}\tilde{q}_{i,j}^{*} \right)(1 + o_{\frac{\mathcal{A}}{\sigma}}(1)) \quad (40)$$
$$= \left(\frac{\mathcal{A}^2}{2\sigma^2}\sum_{i=1}^{M}\sum_{j=1}^{M} g_{i,j}\min\{\alpha_i,\alpha_j\}(1-\max\{\alpha_i,\alpha_j\}) \right)$$
$$\times (1 + o_{\frac{\mathcal{A}}{\sigma}}(1)). \quad (41)$$

Maximizing with respect to $\boldsymbol{\alpha} \in \bar{\mathcal{S}}$ leads to the achievable rate $R^{[1]}(\mathcal{A}, \alpha\mathcal{A}) = \overline{C}_{\sigma}^{[1]}(\mathcal{A}, \alpha\mathcal{A})(1 + o_{\frac{\mathcal{A}}{\sigma}}(1))$. Noting that $\lim_{\frac{\mathcal{A}}{\sigma}\to 0} \frac{\overline{C}_{\sigma}^{[1]}(\mathcal{A},\alpha\mathcal{A})}{R^{[1]}(\mathcal{A},\alpha\mathcal{A})} = 1$ concludes the proof of Theorem 2.

D. Proof of Theorem 3

By Theorem 1, we have $C_{\sigma}(\mathcal{A}, \mathcal{E}) \leq \max_{\boldsymbol{\mathcal{E}}\in\mathcal{S}} \phi_{\sigma}(\mathcal{A}, \boldsymbol{\mathcal{E}})$. For any $\boldsymbol{\mathcal{E}} \in \mathcal{S}$, we have

$$\phi_{\sigma}(\mathcal{A}, \boldsymbol{\mathcal{E}}) \leq \frac{\mathcal{A}}{2\sigma^2}\sum_{i=1}^{M}\sum_{j=1}^{M} \boldsymbol{h}_i^{\mathsf{T}}\boldsymbol{h}_j \min\{\mathcal{E}_i,\mathcal{E}_j\} \quad (42)$$
$$\leq \frac{\mathcal{A}}{2\sigma^2}\max_{i=1,\ldots,M} \mathcal{E}_i \sum_{i=1}^{M}\sum_{j=1}^{M} \boldsymbol{h}_i^{\mathsf{T}}\boldsymbol{h}_j \quad (43)$$
$$\leq \frac{\mathcal{A}\mathcal{E}}{2M\sigma^2}\sum_{i=1}^{M}\sum_{j=1}^{M} \boldsymbol{h}_i^{\mathsf{T}}\boldsymbol{h}_j \quad (44)$$
$$= \frac{\mathcal{A}\mathcal{E}}{2M\sigma^2}\|\boldsymbol{H}\boldsymbol{1}_M\|^2. \quad (45)$$

Thus, $C_{\sigma}(\mathcal{A}, \mathcal{E}) \leq \overline{C}_{\sigma}^{[2]}(\mathcal{A}, \mathcal{E}) \triangleq \frac{\mathcal{A}\mathcal{E}}{2M\sigma^2}\|\boldsymbol{H}\boldsymbol{1}_M\|^2$.

On the other hand, using OOK, spatial repetition, and MRC achieves the rate $R_{\sigma}^{\mathrm{orm}}\left(\mathcal{A}, \frac{\mathcal{E}}{M\mathcal{A}}\right)$ as defined in (14). Using [28, Theorem 1] this can be written as

$$R_{\sigma}^{\mathrm{orm}}\left(\mathcal{A}, \frac{\mathcal{E}}{M\mathcal{A}}\right) = \frac{\mathcal{A}^2}{2\sigma^2}\|\boldsymbol{H}\boldsymbol{1}_M\|^2 \frac{\mathcal{E}}{M\mathcal{A}}\left(1 - \frac{\mathcal{E}}{M\mathcal{A}}\right)$$
$$+ o\left(\frac{\mathcal{E}}{M\mathcal{A}}\left(1 - \frac{\mathcal{E}}{M\mathcal{A}}\right) \right). \quad (46)$$

Here, $\frac{\mathcal{E}}{M\mathcal{A}}\left(1 - \frac{\mathcal{E}}{M\mathcal{A}}\right)$ is the variance of the OOK variable $S \sim \mathrm{Bern}\left(\frac{\mathcal{E}}{M\mathcal{A}}\right)$. Thus,

$$R_{\sigma}^{\mathrm{orm}}\left(\mathcal{A}, \frac{\mathcal{E}}{M\mathcal{A}}\right) \quad (47)$$
$$= \frac{\mathcal{A}\mathcal{E}}{2M\sigma^2}\|\boldsymbol{H}\boldsymbol{1}_M\|^2 \left(1 - \frac{\mathcal{E}}{M\mathcal{A}} + \frac{o\left(\frac{\mathcal{E}}{M\mathcal{A}}\left(1 - \frac{\mathcal{E}}{M\mathcal{A}}\right)\right)}{\frac{\mathcal{A}\mathcal{E}}{2M\sigma^2}\|\boldsymbol{H}\boldsymbol{1}_M\|^2} \right).$$

But $\frac{\mathcal{E}}{M\mathcal{A}}$ is $o_{\frac{\mathcal{E}}{\sigma}}(1)$. Moreover, similar to (37)–(39), we can show that $\frac{o\left(\frac{\mathcal{E}}{M\mathcal{A}}\left(1-\frac{\mathcal{E}}{M\mathcal{A}}\right)\right)}{\frac{\mathcal{A}\mathcal{E}}{2M\sigma^2}\|\boldsymbol{H}\boldsymbol{1}_M\|^2}$ is also $o_{\frac{\mathcal{E}}{\sigma}}(1)$. Thus, this achievable rate can be written as

$$R_{\sigma}^{\mathrm{orm}}\left(\mathcal{A}, \frac{\mathcal{E}}{M\mathcal{A}}\right) = \frac{\mathcal{A}\mathcal{E}}{2M\sigma^2}\|\boldsymbol{H}\boldsymbol{1}_M\|^2 \left(1 + o_{\frac{\mathcal{E}}{\sigma}}(1)\right). \quad (48)$$

Since $\lim_{\frac{\mathcal{E}}{\sigma}\to 0} \frac{\overline{C}_{\sigma}^{[2]}(\mathcal{A},\mathcal{E})}{R_{\sigma}^{\mathrm{orm}}\left(\mathcal{A},\frac{\mathcal{E}}{M\mathcal{A}}\right)} = 1$, this proves Theorem 3.

E. Low-SNR Asymptotic Capacity Under Individual Average Constraints

We consider proportional SNRs first, where $\mathcal{E}_{\mathrm{ind}} = \alpha_{\mathrm{ind}}\mathcal{A}$ and $\alpha_{\mathrm{ind}} \in (0,1)$.

1) Proportional Constraints: Similar to Theorem 2, the low-SNR asymptotic capacity in this case evaluates to

$$C_{\sigma,0}^{[1],\mathrm{ind}}(\mathcal{A}, \alpha_{\mathrm{ind}}\mathcal{A}) \quad (49)$$
$$= \frac{\mathcal{A}^2}{2\sigma^2}\max_{\boldsymbol{\alpha}}\sum_{i=1}^{M}\sum_{j=1}^{M} \boldsymbol{h}_i^{\mathsf{T}}\boldsymbol{h}_j \min\{\alpha_i,\alpha_j\}(1-\max\{\alpha_i,\alpha_j\}),$$

where the maximization is over the set of feasible $\boldsymbol{\alpha}$ in $[0, \alpha_{\mathrm{ind}}]^M$ due to the individual average intensity constraint. Note that

$$\max_{\boldsymbol{\alpha}\in[0,\alpha_{\mathrm{ind}}]^M}\sum_{i=1}^{M}\sum_{j=1}^{M} \boldsymbol{h}_i^{\mathsf{T}}\boldsymbol{h}_j \min\{\alpha_i,\alpha_j\}(1-\max\{\alpha_i,\alpha_j\})$$
$$\leq \sum_{i=1}^{M}\sum_{j=1}^{M} \boldsymbol{h}_i^{\mathsf{T}}\boldsymbol{h}_j \max_{\substack{(\alpha_i,\alpha_j)\\ \in[0,\alpha_{\mathrm{ind}}]^2}} \min\{\alpha_i,\alpha_j\}(1-\max\{\alpha_i,\alpha_j\})$$
$$\leq \sum_{i=1}^{M}\sum_{j=1}^{M} \boldsymbol{h}_i^{\mathsf{T}}\boldsymbol{h}_j \max_{\alpha_{\min}\in[0,\alpha_{\mathrm{ind}}]} \alpha_{\min}(1-\alpha_{\min}) \quad (50)$$
$$= \sum_{i=1}^{M}\sum_{j=1}^{M} \boldsymbol{h}_i^{\mathsf{T}}\boldsymbol{h}_j \bar{\alpha}_{\mathrm{ind}}(1-\bar{\alpha}_{\mathrm{ind}}) \quad (51)$$
$$= \|\boldsymbol{H}\boldsymbol{1}_M\|^2 \bar{\alpha}_{\mathrm{ind}}(1-\bar{\alpha}_{\mathrm{ind}}), \quad (52)$$

where $\bar{\alpha}_{\mathrm{ind}} = \min\{\alpha_{\mathrm{ind}}, \frac{1}{2}\}$. But $\boldsymbol{\alpha} = \bar{\alpha}_{\mathrm{ind}}\boldsymbol{1}_M$ achieves this upper bound and is feasible in this case. Thus

$$C_{\sigma,0}^{[1],\mathrm{ind}}(\mathcal{A}, \alpha_{\mathrm{ind}}\mathcal{A}) = \frac{\mathcal{A}^2}{2\sigma^2}\bar{\alpha}_{\mathrm{ind}}(1-\bar{\alpha}_{\mathrm{ind}})\|\boldsymbol{H}\boldsymbol{1}_M\|^2. \quad (53)$$

This proves Corollary 2.

2) Fixed Peak SNR: By Theorem 1, we have $C_{\sigma}(\mathcal{A}, \mathcal{E}) \leq \max_{\boldsymbol{\mathcal{E}}\in[0,\mathcal{E}_{\mathrm{ind}}]^M} \phi_{\sigma}(\mathcal{A}, \boldsymbol{\mathcal{E}})$, where the set of feasible $\boldsymbol{\mathcal{E}}$ is replaced with $[0, \mathcal{E}_{\mathrm{ind}}]^M$ due to the individual intensity constraints. Similar to the derivation in Sec. IV-D, we can show that for $\boldsymbol{\mathcal{E}} \in [0, \mathcal{E}_{\mathrm{ind}}]^M$, we have

$$\phi_{\sigma}(\mathcal{A}, \boldsymbol{\mathcal{E}}) \leq \frac{\mathcal{A}\mathcal{E}_{\mathrm{ind}}}{2\sigma^2}\|\boldsymbol{H}\boldsymbol{1}_M\|^2. \quad (54)$$

Thus, $C_\sigma(\mathcal{A}, \mathcal{E}) \leq \frac{\mathcal{A}\mathcal{E}_{\text{ind}}}{2\sigma^2}\|\boldsymbol{H}\boldsymbol{1}_M\|^2$. On the other hand, the achievable rate $R_\sigma^{\text{orm}}(\mathcal{A}, \mathcal{E}_{\text{ind}})$ (14) becomes $\frac{\mathcal{A}\mathcal{E}_{\text{ind}}}{2\sigma^2}\|\boldsymbol{H}\boldsymbol{1}_M\|^2\left(1 + o_{\frac{\mathcal{E}}{\sigma}}(1)\right)$ which can be proved similar to (48). Since the ratio of the upper bound and the achievable rate converges to 1 as $\frac{\mathcal{E}_{\text{ind}}}{\sigma} \to 0$, this proves Corollary 3.

V. Conclusion

We studied the low-SNR asymptotic capacity of the optical intensity MIMO channel under average and peak intensity constraint, which is an important quantity in the context of optical wireless communications. We derived a capacity upper bound and proved its achievability at asymptotically low SNR. In this process, we have shown that the optimal input distribution is a maximally-correlated vector-binary distribution. We have also shown that if the transmitter is constrained by individual (per aperture) constraints or peak constraints only, then coded OOK and spatial repetition at the transmitter, and MRC at the receiver achieve the low-SNR asymptotic capacity of the channel. This leads to a simple practical scheme under this scenario. The results of this paper can be useful in optical wireless communication systems with strong noise due to ambient light e.g., or with low average intensity (relative to noise) due to light-dimming for instance.

Acknowledgment

Appendix A
Maximum Covariance of $(x_1, x_2) \in \{0, \mathcal{A}\}^2$

Lemma 2: The covariance of $(x_1, x_2)^\mathsf{T} \in \{0, \mathcal{A}\}^2$ satisfying $\mathbb{E}_{P_{x_1,x_2}}[(x_1, x_2)^\mathsf{T}] = (\mathcal{E}_1, \mathcal{E}_2)^\mathsf{T}$ and $\mathcal{E}_1 \geq \mathcal{E}_2$ is upper bounded by $\mathcal{E}_2(\mathcal{A} - \mathcal{E}_1)$, which is achievable if

$$P_{x_1,x_2}(x_1, x_2) \tag{55}$$
$$= \left(1 - \frac{\mathcal{E}_1}{\mathcal{A}}\right)\delta(x_1, x_2) + \frac{\mathcal{E}_1 - \mathcal{E}_2}{\mathcal{A}}\delta(x_1 - \mathcal{A}, x_2)$$
$$+ \frac{\mathcal{E}_2}{\mathcal{A}}\delta(x_1 - \mathcal{A}, x_2 - \mathcal{A}).$$

where $\delta(\cdot)$ is the Dirac delta.

Proof: We have that

$$q_{1,2} = \sum_{(x_1,x_2)^\mathsf{T} \in \{0,\mathcal{A}\}^2} P_{x_1,x_2}(x_1, x_2)(x_1 - \mathcal{E}_1)(x_2 - \mathcal{E}_2).$$

The maximum covariance can be obtained by solving

$$\max_{P_{x_1,x_2}(x_1,x_2)} \quad q_{1,2} \tag{56a}$$

$$\text{s.t.} \sum_{(x_1,x_2)^\mathsf{T} \in \{0,\mathcal{A}\}^2} P_{x_1,x_2}(x_1, x_2) = 1, \tag{56b}$$

$$p_{x_1,x_2}(x_1, x_2) \geq 0, \ \forall (x_1, x_2)^\mathsf{T} \in \{0, \mathcal{A}\}^2, \tag{56c}$$

$$P_{x_1,x_2}(\mathcal{A}, 0) + P_{x_1,x_2}(\mathcal{A}, \mathcal{A}) = \frac{\mathcal{E}_1}{\mathcal{A}}, \tag{56d}$$

$$P_{x_1,x_2}(0, \mathcal{A}) + P_{x_1,x_2}(\mathcal{A}, \mathcal{A}) = \frac{\mathcal{E}_2}{\mathcal{A}}. \tag{56e}$$

The last two constraints are implied by $\mathbb{E}_{P_{x_i}}[x_i] = \mathcal{E}_i$, $i \in \{1, 2\}$. This is a simple linear program, whose solution can be shown to be given by $P_{x_1,x_2}(0, 0) = 1 - \frac{\mathcal{E}_1}{\mathcal{A}}$, $P_{x_1,x_2}(0, \mathcal{A}) = 0$, $P_{x_1,x_2}(\mathcal{A}, 0) = \frac{\mathcal{E}_1 - \mathcal{E}_2}{\mathcal{A}}$, and $P_{x_1,x_2}(\mathcal{A}, \mathcal{A}) = \frac{\mathcal{E}_2}{\mathcal{A}}$. This leads to the statement. ∎

Appendix B
Alternative Proof of Theorem 2

Here, we provide a proof of Theorem 2 which relies on the relation between the mutual information and the minimum mean-square error (MMSE) over a Gaussian channel [30].

Let us rewrite our channel model in the following equivalent form

$$\bar{\boldsymbol{y}} = \sqrt{\text{snr}}\boldsymbol{H}\bar{\boldsymbol{x}} + \bar{\boldsymbol{z}}, \tag{57}$$

where $\bar{\boldsymbol{x}} \in [0, 1]^M$, $\bar{\boldsymbol{z}}$ is Gaussian with zero mean and covariance matrix $\boldsymbol{I}_N$, and $\text{snr} = \frac{A^2}{\sigma^2}$. For this Gaussian vector channel, which has the same capacity as the original one, since $\mathbb{E}_{P_{\bar{x}}}[\|\bar{\boldsymbol{x}}\|^2] < \infty$, it holds that [30, Theorem 2]

$$\frac{\mathrm{d}}{\mathrm{dsnr}}\mathsf{I}(\bar{\boldsymbol{x}}; \sqrt{\text{snr}}\boldsymbol{H}\bar{\boldsymbol{x}} + \bar{\boldsymbol{z}}) = \frac{1}{2}\mathsf{mmse}(\text{snr}), \tag{58}$$

where $\mathsf{mmse}(\text{snr})$ is the MMSE at a given snr given by $\mathbb{E}_{P_{\bar{x},\bar{y}}}\left[\|\boldsymbol{H}\bar{\boldsymbol{x}} - \boldsymbol{H}\mathbb{E}_{P_{\bar{x}|\bar{y}}}[\bar{\boldsymbol{x}}|\bar{\boldsymbol{y}}; \text{snr}]\|^2\right]$. This relation implies that

$$\mathsf{I}(\bar{\boldsymbol{x}}; \sqrt{\text{snr}}\boldsymbol{H}\bar{\boldsymbol{x}} + \bar{\boldsymbol{z}}) = \frac{1}{2}\int_0^{\text{snr}} \mathsf{mmse}(\gamma)\mathrm{d}\gamma. \tag{59}$$

Now we focus on low SNR, i.e., $\text{snr} \to 0$. Note that the following holds

$$\lim_{\text{snr} \to 0} \mathbb{E}_{P_{\bar{x}|\bar{y}}}[\bar{\boldsymbol{x}}|\bar{\boldsymbol{y}}; \text{snr}] = \mathbb{E}_{P_{\bar{x}}}[\bar{\boldsymbol{x}}], \tag{60}$$

by the continuity of the MMSE estimate. Thus, in this regime, we have

$$\mathsf{mmse}(\gamma) \approx \mathbb{E}_{P_{\bar{x}}}\left[\|\boldsymbol{H}\bar{\boldsymbol{x}} - \boldsymbol{H}\mathbb{E}_{P_{\bar{x}}}[\bar{\boldsymbol{x}}]\|^2\right] \tag{61}$$
$$= \text{Tr}\left(\boldsymbol{H}^\mathsf{T}\boldsymbol{H}\mathbb{E}_{P_{\bar{x}}}\left[(\bar{\boldsymbol{x}} - \mathbb{E}_{P_{\bar{x}}}[\bar{\boldsymbol{x}}])^\mathsf{T}(\bar{\boldsymbol{x}} - \mathbb{E}_{P_{\bar{x}}}[\bar{\boldsymbol{x}}])\right]\right)$$
$$= \text{Tr}\left(\boldsymbol{H}^\mathsf{T}\boldsymbol{H}\boldsymbol{Q}\right). \tag{62}$$

Here, we used $\approx$ to indicate an asymptotic equality as $\text{snr} \to 0$. Hence, the asymptotic equality

$$\mathsf{I}(\bar{\boldsymbol{x}}; \sqrt{\text{snr}}\boldsymbol{H}\bar{\boldsymbol{x}} + \bar{\boldsymbol{z}}) \approx \frac{\text{snr}}{2}\text{Tr}\left(\boldsymbol{H}^\mathsf{T}\boldsymbol{H}\boldsymbol{Q}\right) \tag{63}$$

holds as $\text{snr} \to 0$. To obtain capacity, we have to maximize the left-hand side with respect to the distribution of $\bar{\boldsymbol{x}}$, which is equivalent to maximizing the right-hand side with respect to $\boldsymbol{Q}$. The problem is similar to (24), and using similar techniques as in Sec. IV-B, we can show that for a given $\mathbb{E}_{P_{\bar{x}}}[\bar{\boldsymbol{x}}]$, a maximally-correlated M-variate binary distribution maximizes the above trace. It remains to maximize with respect to $\mathbb{E}_{P_{\bar{x}}}[\bar{\boldsymbol{x}}]$

so that $\sum_{i=1}^{M} \mathbb{E}_{P_{\bar{x}_i}}[\bar{x}_i] \leq \frac{\mathcal{E}}{\mathcal{A}}$. Denoting $\mathbb{E}_{P_{\bar{x}_i}}[x_i]$ by α_i leads to the following asymptotic result as $\mathsf{snr} \to 0$

$$\max_{P_{\bar{x}}(\bar{x})} \mathsf{I}(\bar{x}; \sqrt{\mathsf{snr}}\boldsymbol{H}\bar{x} + \bar{z}) \tag{64}$$

$$\approx \max_{\mathbb{E}_{P_{\bar{x}}}[\bar{x}]} \frac{\mathsf{snr}}{2} \sum_{i=1}^{M} \sum_{j=1}^{M} \boldsymbol{h}_i^{\mathsf{T}} \boldsymbol{h}_j \min\{\mathbb{E}_{P_{\bar{x}_i}}[\bar{x}_i], \mathbb{E}_{P_{\bar{x}_j}}[\bar{x}_j]\} \tag{65}$$
$$\times (1 - \max\{\mathbb{E}_{P_{\bar{x}_i}}[\bar{x}_i], \mathbb{E}_{P_{\bar{x}_j}}[\bar{x}_j]\})$$

$$= \frac{\mathcal{A}^2}{2\sigma^2} \max_{\boldsymbol{\alpha} \in \mathcal{S}} \sum_{i=1}^{M} \sum_{j=1}^{M} \boldsymbol{h}_i^{\mathsf{T}} \boldsymbol{h}_j \min\{\alpha_i, \alpha_j\}(1 - \max\{\alpha_i, \alpha_j\}).$$

This proves Theorem 2.

APPENDIX C
OPTIMALITY OF BINARY INPUTS

Herein, we show how we can produce from any $P_{\boldsymbol{x}}(x)$ with support $[0, \mathcal{A}]^M$ and $\mathbb{E}_{P_{\boldsymbol{x}}}[\boldsymbol{x}] = \boldsymbol{\mathcal{E}}$ another probability measure $P_{\boldsymbol{x}}^{\mathrm{b}}(\boldsymbol{x})$ with support $\{0, \mathcal{A}\}^M$ so that $\mathbb{E}_{P_{\boldsymbol{x}}^{\mathrm{b}}}[\boldsymbol{x}] = \boldsymbol{\mathcal{E}}$ and $\mathbb{E}_{P_{\boldsymbol{x}}^{\mathrm{b}}}[\psi(\boldsymbol{x})] \geq \mathbb{E}_{P_{\boldsymbol{x}}}[\psi(\boldsymbol{x})]$. We start by writing

$$\mathbb{E}_{P_{\boldsymbol{x}}}[\psi(\boldsymbol{x})]$$
$$= \sum_{i=1}^{M} \sum_{j=1}^{M} g_{i,j} \mathbb{E}_{P_{\boldsymbol{x}}}[(x_i - \mathcal{E}_i)(x_j - \mathcal{E}_j)] \tag{66}$$

$$= g_{1,1} \mathbb{E}_{P_{\boldsymbol{x}}}[(x_1 - \mathcal{E}_1)^2] + 2 \sum_{j=2}^{M} g_{1,j} \mathbb{E}_{P_{\boldsymbol{x}}}[(x_1 - \mathcal{E}_1)(x_j - \mathcal{E}_j)]$$
$$+ \sum_{i=2}^{M} \sum_{j=2}^{M} g_{i,j} \mathbb{E}_{P_{\boldsymbol{x}}}[(x_i - \mathcal{E}_i)(x_j - \mathcal{E}_j)], \tag{67}$$

where we used $g_{1,j} = g_{j,1}$ due to the symmetry of $\boldsymbol{G}$. Note that

$$\mathbb{E}_{P_{\boldsymbol{x}}}[(x_1 - \mathcal{E}_1)^2] \leq \mathbb{E}_{P_{\boldsymbol{x}}}[x_1 \mathcal{A} - 2\mathcal{E}_1 x_1 + \mathcal{E}_1^2] \tag{68}$$
$$= \mathcal{E}_1(\mathcal{A} - \mathcal{E}_1). \tag{69}$$

Denote by $P_{x_2,\ldots,x_M}$ the probability measure of $(x_2, \ldots, x_M)$ induced by $P_{\boldsymbol{x}}$. Now consider the probability measure $P_{\boldsymbol{x}}^{[1]} = P_{x_2,\ldots,x_M} P_{x_1|x_2,\ldots,x_M}^{[1]}$, where $P_{x_1|x_2,\ldots,x_M}^{[1]} = p_0^{[1]} \delta(x_1) + p_1^{[1]} \delta(x_1 - \mathcal{A})$ with $\delta(x)$ being the Dirac delta, and with $p_0^{[1]} = 1 - p_1^{[1]}$ and $p_1^{[1]} = \frac{1}{\mathcal{A}} \mathbb{E}_{P_{\boldsymbol{x}}}[x_1|x_2,\ldots,x_M]$. It is easy to see that $\mathbb{E}_{P_{\boldsymbol{x}}^{[1]}}[\boldsymbol{x}] = \boldsymbol{\mathcal{E}}$ so this preserves the mean. Moreover, $\mathbb{E}_{P_{\boldsymbol{x}}^{[1]}}[(x_1 - \mathcal{E}_1)^2] = \mathcal{E}_1(\mathcal{A} - \mathcal{E}_1)$. Furthermore, for all $j = 2, \ldots, M$, we have

$$\mathbb{E}_{P_{\boldsymbol{x}}^{[1]}}[(x_1 - \mathcal{E}_1)(x_j - \mathcal{E}_j)] = \mathbb{E}_{P_{\boldsymbol{x}}}[(x_1 - \mathcal{E}_1)(x_j - \mathcal{E}_j)]. \tag{70}$$

Consequently, since $g_{1,1} \geq 0$, we have $\mathbb{E}_{P_{\boldsymbol{x}}^{[1]}}[\psi(\boldsymbol{x})] \geq \mathbb{E}_{P_{\boldsymbol{x}}}[\psi(\boldsymbol{x})]$.

Repeating the same argument to produce $P_{\boldsymbol{x}}^{[i]}$ from $P_{\boldsymbol{x}}^{[i-1]}$, for all $i = 2, \ldots, M$ leads to a probability measure $P_{\boldsymbol{x}}^{[M]}$ with support $\{0, \mathcal{A}\}^M$, mean $\boldsymbol{\mathcal{E}}$, and $\mathbb{E}_{P_{\boldsymbol{x}}^{[M]}}[\psi(\boldsymbol{x})] \geq \mathbb{E}_{P_{\boldsymbol{x}}}[\psi(\boldsymbol{x})]$. This proves Proposition 1.

APPENDIX D
DERIVATION OF $\iota_{i,j}$ IN SEC. IV-C

We need to derive

$$\iota_{i,j} = \int_{\mathbb{R}^N} \frac{1}{P_{\tilde{\boldsymbol{y}}|\tilde{\boldsymbol{x}}}(\tilde{\boldsymbol{y}}|\tilde{\boldsymbol{x}})} \frac{\partial P_{\tilde{\boldsymbol{y}}|\tilde{\boldsymbol{x}}}(\tilde{\boldsymbol{y}}|\tilde{\boldsymbol{x}})}{\partial \tilde{x}_i} \frac{\partial P_{\tilde{\boldsymbol{y}}|\tilde{\boldsymbol{x}}}(\tilde{\boldsymbol{y}}|\tilde{\boldsymbol{x}})}{\partial \tilde{x}_j} d\tilde{\boldsymbol{y}}, \tag{71}$$

where $P_{\tilde{\boldsymbol{y}}|\tilde{\boldsymbol{x}}}(\tilde{\boldsymbol{y}}|\tilde{\boldsymbol{x}}) = (2\pi)^{-\frac{N}{2}} e^{-\frac{1}{2}\|\tilde{\boldsymbol{y}} - \boldsymbol{H}\tilde{\boldsymbol{x}}\|^2}$. Substituting in the above integral, we obtain

$$\iota_{i,j} = \int_{\mathbb{R}^N} (2\pi)^{\frac{N}{2}} e^{\frac{1}{2}\|\tilde{\boldsymbol{y}} - \boldsymbol{H}\tilde{\boldsymbol{x}}\|^2} \frac{\partial}{\partial \tilde{x}_i} \left((2\pi)^{-\frac{N}{2}} e^{-\frac{1}{2}\|\tilde{\boldsymbol{y}} - \boldsymbol{H}\tilde{\boldsymbol{x}}\|^2} \right)$$
$$\times \frac{\partial}{\partial \tilde{x}_j} \left((2\pi)^{-\frac{N}{2}} e^{-\frac{1}{2}\|\tilde{\boldsymbol{y}} - \boldsymbol{H}\tilde{\boldsymbol{x}}\|^2} \right) d\tilde{\boldsymbol{y}}$$
$$= \int_{\mathbb{R}^N} (2\pi)^{-\frac{N}{2}} e^{\frac{1}{2}\|\tilde{\boldsymbol{y}} - \boldsymbol{H}\tilde{\boldsymbol{x}}\|^2} \frac{\partial}{\partial \tilde{x}_i} \left(e^{-\frac{1}{2}\|\tilde{\boldsymbol{y}} - \boldsymbol{H}\tilde{\boldsymbol{x}}\|^2} \right)$$
$$\times \frac{\partial}{\partial \tilde{x}_j} \left(e^{-\frac{1}{2}\|\tilde{\boldsymbol{y}} - \boldsymbol{H}\tilde{\boldsymbol{x}}\|^2} \right) d\tilde{\boldsymbol{y}}. \tag{72}$$

But

$$\frac{\partial}{\partial \tilde{x}_i} \left(e^{-\frac{1}{2}\|\tilde{\boldsymbol{y}} - \boldsymbol{H}\tilde{\boldsymbol{x}}\|^2} \right)$$
$$= \frac{\partial}{\partial \tilde{x}_i} \left(e^{-\frac{1}{2}\sum_{n=1}^{N} (\tilde{y}_n - \sum_{m=1}^{M} h_{n,m}\tilde{x}_m)^2} \right) \tag{73}$$
$$= -\frac{1}{2} \left(\sum_{n=1}^{N} (-2) h_{n,i} \left(\tilde{y}_n - \sum_{m=1}^{M} h_{n,m}\tilde{x}_m \right) \right)$$
$$\times e^{-\frac{1}{2}\sum_{n=1}^{N}(\tilde{y}_n - \sum_{m=1}^{M} h_{n,m}\tilde{x}_m)^2} \tag{74}$$
$$= \left(\sum_{n=1}^{N} h_{n,i} \left(\tilde{y}_n - \sum_{m=1}^{M} h_{n,m}\tilde{x}_m \right) \right) e^{-\frac{1}{2}\|\tilde{\boldsymbol{y}} - \boldsymbol{H}\tilde{\boldsymbol{x}}\|^2}. \tag{75}$$

Moreover, $\tilde{\boldsymbol{y}} - \boldsymbol{H}\tilde{\boldsymbol{x}} = \tilde{\boldsymbol{z}}$, $\tilde{y}_n - \sum_{m=1}^{M} h_{n,m}\tilde{x}_m = \tilde{z}_n$, and $d\tilde{\boldsymbol{y}} = d\tilde{\boldsymbol{z}}$. Thus,

$$\iota_{i,j}$$
$$= \int_{\mathbb{R}^N} (2\pi)^{-\frac{N}{2}} e^{-\frac{1}{2}\|\tilde{\boldsymbol{z}}\|^2}$$
$$\times \left(\sum_{n_1=1}^{N} h_{n_1,i}\tilde{z}_{n_1} \right) \left(\sum_{n_2=1}^{N} h_{n_2,j}\tilde{z}_{n_2} \right) d\tilde{\boldsymbol{z}} \tag{76}$$
$$= \int_{\mathbb{R}^N} (2\pi)^{-\frac{N}{2}} e^{-\frac{1}{2}\|\tilde{\boldsymbol{z}}\|^2} \left(\sum_{n=1}^{N} h_{n,i}h_{n,j}\tilde{z}_n^2 \right) d\tilde{\boldsymbol{z}}$$
$$+ \int_{\mathbb{R}^N} (2\pi)^{-\frac{N}{2}} e^{-\frac{1}{2}\|\tilde{\boldsymbol{z}}\|^2}$$
$$\times \left(2 \sum_{n_1=1}^{N} \sum_{n_2=1, n_2 \neq n_1}^{N} h_{n_1,i}h_{n_2,j}\tilde{z}_{n_1}\tilde{z}_{n_2} \right) d\tilde{\boldsymbol{z}} \tag{77}$$
$$= \sum_{n=1}^{N} h_{n,i}h_{n,j} \int_{\mathbb{R}^N} (2\pi)^{-\frac{N}{2}} e^{-\frac{1}{2}\|\tilde{\boldsymbol{z}}\|^2} \tilde{z}_n^2 d\tilde{\boldsymbol{z}} \tag{78}$$
$$+ 2 \sum_{n_1=1}^{N} \sum_{\substack{n_2=1 \\ n_2 \neq n_1}}^{N} h_{n_1,i}h_{n_2,j} \int_{\mathbb{R}^N} (2\pi)^{-\frac{N}{2}} e^{-\frac{1}{2}\|\tilde{\boldsymbol{z}}\|^2} \tilde{z}_{n_1}\tilde{z}_{n_2} d\tilde{\boldsymbol{z}}.$$

Note that the first integral is the variance of $\tilde{z}_n$ (zero mean) which is 1. The second integral is the covariance of z_{n_1} and z_{n_2} which is 0 since the two are independent with zero mean. Thus, $\iota_{i,j} = \sum_{n=1}^{N} h_{n,i}h_{n,j}$ as given in Sec. IV-C.

REFERENCES

[1] A. Chaaban, Z. Rezki, and M.-S. Alouini, "MIMO Optical Intensity Channels with Per-Aperture Constraints: Low-SNR Capacity," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, 2018 (to appear).

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

[3] A. D. Wyner, "Capacity and error exponent for the direct detection photon channelParts I and II," *IEEE Trans. Inf. Theory*, vol. 34, no. 6, pp. 1462–1471, Nov. 1988.

[4] 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.

[5] S. M. Moser, "Capacity results of an optical intensity channel with input-dependent Gaussian noise," *IEEE Trans. Inf. Theory*, vol. 58, no. 1, pp. 207–223, Jan. 2012.

[6] J. M. Kahn and J. R. Barry, "Wireless infrared communications," *Proc. IEEE*, vol. 85, no. 2, pp. 265–298, Feb. 1997.

[7] H. Burchardt, N. Serafimovski, D. Tsonev, S. Videv, and H. Haas, "VLC: Beyond point-to-point communication," *IEEE Commun. Mag.*, vol. 52, no. 7, pp. 98–105, July 2014.

[8] P. Pathak, X. Fen, P. Hu, and P. Mohapatra, "Visible light communication, networking, and sensing: A survey, potential and challenges," *IEEE Commun. Surveys Tuts.*, vol. 17, no. 4, pp. 2047–2077, 4th quarter 2015.

[9] H. Haas, L. Yin, Y. Wang, and C. Chen, "What is LiFi?" *IEEE/OSA J. Lightw. Technol.*, vol. 34, no. 6, pp. 1533–1544, March 2016.

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

[11] 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.

[12] 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.

[13] 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.

[14] A. Chaaban, O. M. S. Al-Ebraheemy, T. Y. Al-Naffouri, and M. S. Alouini, "Capacity bounds for the Gaussian IM-DD optical multiple-access channel," *IEEE Trans. Wireless Commun.*, vol. 16, no. 5, pp. 3328–3340, May 2017.

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

[16] A. A. Farid and S. Hranilovic, "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.

[17] N. Letzepis and A. G. i Fàbregas, "Outage probability of the Gaussian MIMO free-space optical channel with PPM," *IEEE Trans. Commun.*, vol. 57, no. 12, pp. 3682–3690, Dec. 2009.

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

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

[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] A. Chaaban, Z. Rezki, and M.-S. Alouini, "Fundamental limits of parallel optical wireless channels: Capacity results and outage formulation," *IEEE Trans. Commun.*, vol. 65, no. 1, pp. 296–311, Jan. 2017.

[22] M. Safari, *MIMO Free-Space Optical Communication*. Springer International Publishing, 2016, pp. 231–253.

[23] K. Xu, H. Y. Yu, Y. J. Zhu, and H. B. Cai, "Channel-adaptive space-collaborative constellation design for MIMO VLC with fast maximum likelihood detection," *IEEE Access*, vol. 5, pp. 842–852, Mar. 2017.

[24] 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.

[25] A. Chaaban, Z. Rezki, and M.-S. Alouini, "MIMO intensity-modulation channels: Capacity bounds and high SNR characterization," in *Proc. IEEE Int. Conf. Commun. (ICC)*, Paris, France, May 2017, pp. 1–6.

[26] ——, "Capacity Bounds and High-SNR Capacity of MIMO Intensity-Modulation Optical Channels," *IEEE Trans. Wireless Commun.*, 2018 (to appear).

[27] S. M. Moser, M. Mylonakis, L. Wang, and M. Wigger, "Asymptotic capacity results for MIMO wireless optical communication," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Aachen, Germany, Jun. 2017, pp. 536–540.

[28] V. V. Prelov and E. C. van der Meulen, "An asymptotic expression for the information and capacity of a multidimensional channel with weak input signals," *IEEE Trans. Inf. Theory*, vol. 39, no. 5, pp. 1728–1735, Sep. 1993.

[29] V. V. Prelov and S. Verdú, "Second-order asymptotics of mutual information," *IEEE Trans. Inf. Theory*, vol. 50, no. 8, pp. 1567–1580, Aug 2004.

[30] D. Guo, S. Shamai, and S. Verdú, "Mutual information and minimum mean-square error in Gaussian channels," *IEEE Trans. Inf. Theory*, vol. 51, no. 4, pp. 1261–1282, April 2005.

[31] S. M. Moser, L. Wang, and M. A. Wigger, "Capacity results on multiple-input single-output wireless optical channels," *ArXiv abs/1710.09177*, October 2017.

[32] 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.

[33] J. G. Smith, "The information capacity of amplitude- and variance-constrained sclar Gaussian channels," *Information and Control*, vol. 18, no. 3, pp. 203–219, 1971.

[34] S. M. Moser, L. Wang, and M. Wigger, "Asymptotic high-SNR capacity of MISO optical intensity channels," in *Proc. IEEE Inf. Theory Workshop (ITW)*, Kaohsiung, Taiwan, Nov. 2017, pp. 6–10.

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

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

[37] P. Billingsley, *Probability and Measure*. John Wiley and Sons, Inc., 1995.

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

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

Anas Chaaban (S'09 - M'14 - SM'17) received the Maîtrise ès Sciences degree in electronics from Lebanese University, Lebanon, in 2006, the M.Sc. degree in communications technology and the 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. From 2008 to 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, University of Ulm, Germany, from 2009 to 2011, which relocated to the Ruhr-University of Bochum in 2011. He was a PostDoctoral Researcher with the Ruhr-University of Bochum from 2013 to 2014, and with King Abdullah University of Science and Technology from 2015 to 2017. He joined the School of Engineering at the University of British Columbia as an Assistant Professor in 2018. 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 in electrical engineering from École Polytechnique, Montreal, Québec, in 2008. After a few years of experience as a postdoctoral fellow and a research scientist at KAUST, he joined University of Idaho as an Assistant Professor in the ECE Department.

Mohamed-Slim Alouini (S94-M98-SM03-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.