On the Capacity of MIMO Optical Wireless Channels

Longguang Li[§], Stefan M. Moser[*†], Ligong Wang[‡], and Michèle Wigger[§]
[*]Signal and Information Processing Lab, ETH Zürich, Switzerland
[†]Department of Electrical and Computer Engineering, National Chiao Tung University, Hsinchu, Taiwan
[‡]ETIS—Université Paris Seine, Université de Cergy-Pontoise, ENSEA, CNRS, Cergy-Pontoise, France
[§]LTCI, Telecom ParisTech, Université Paris-Saclay, 75013 Paris, France

***Abstract*—This paper investigates the capacity of the MIMO free-space optical intensity channel under a per input-antenna peak-power constraint and a total average-power constraint over all input antennas. Our work considers the setup with more transmit than receive antennas, and characterizes capacity as an alternative optimization problem over the distribution of the input vector times the channel matrix. This alternative capacity expression is then used to obtain upper and lower bounds on the capacity, which match asymptotically in the high signal-to-noise ratio (SNR) regime.**

I. Introduction and Channel Model

Consider a wireless optical *intensity-modulation direct-detection (IM-DD)* system where the transmitter is equipped with n_T LEDs (or LDs) and the receiver with n_R photodetectors. The photodetectors receive a superposition of the signals sent by the LEDs, and we assume that the crosstalk between the LEDs is constant. Hence, the channel output is given by

$$\mathbf{Y} = \mathsf{H}\mathbf{x} + \mathbf{Z}, \tag{1}$$

where $\mathbf{x} = (x_1, \ldots, x_{n_\mathrm{T}})^\intercal$ denotes the n_T-dimensional channel input vector, $\mathbf{Z}$ the n_R-dimensional noise vector with independent standard Gaussian entries, and H the deterministic n_R-by-n_T channel matrix, which we also write in the form

$$\mathsf{H} = [\mathbf{h}_1, \mathbf{h}_2, \ldots, \mathbf{h}_{n_\mathrm{T}}]. \tag{2}$$

The channel inputs correspond to optical intensities sent by the LEDs, hence they are nonnegative:

$$x_k \in \mathbb{R}_0^+, \quad k = 1, \ldots, n_\mathrm{T}. \tag{3}$$

We assume the inputs are subject to a peak-power (peak-intensity) and an average-power (average-intensity) constraint:

$$\Pr\big[X_k > \mathsf{A}\big] = 0, \quad \forall\, k \in \{1, \ldots, n_\mathrm{T}\}, \tag{4a}$$

$$\mathsf{E}\big[\|\mathbf{X}\|_1\big] \le \mathsf{E}, \tag{4b}$$

for some fixed parameters $\mathsf{A}, \mathsf{E} > 0$, and where $\|\cdot\|_1$ denotes the L_1-norm. Note that the average-power constraint is on the expectation of the channel input and not on its square. Also note that A describes the maximum power of each single LED, while E describes the allowed total average power of all LEDs together. We denote the ratio between the allowed average power and the allowed peak power by α:

$$\alpha \triangleq \frac{\mathsf{E}}{\mathsf{A}}, \tag{5}$$

where $0 < \alpha \le n_\mathrm{T}$. For $\alpha \ge \frac{n_\mathrm{T}}{2}$, by symmetry of the setup, the average-power constraint is inactive. Thus, when $\alpha \ge \frac{n_\mathrm{T}}{2}$, the channel essentially reduces to the case where there is only a peak-power constraint.

In previous works, bounds on the capacity of the above-described wireless optical channel were derived for the single-antenna case ($n_\mathrm{T} = n_\mathrm{R} = 1$) [1]–[4], the no-crosstalk case (H is diagonal) [5], the multi-input single-output (MISO) case ($n_\mathrm{T} > 1$ and $n_\mathrm{R} = 1$) [6], [7], the case with a full-rank square H ($n_\mathrm{T} = n_\mathrm{R} > 1$) [8], [9], and the case with a full-column-rank H [8].

In this paper we focus on the MIMO setup with more transmit than receive antennas:

$$n_\mathrm{T} > n_\mathrm{R} > 1. \tag{6}$$

Inspired by our results on the MISO channel [7], we find the most energy-efficient signaling method for the setup (6) (Lemma 1 ahead). In the MISO case, antennas can be ordered according to increasing channel gains, and the optimal signaling strategy is to rely as much as possible on stronger antennas. In other words, if an antenna is used for active signaling in a channel use, then all stronger antennas should transmit at full power A and all weaker antennas should be silenced. In the MIMO case, $n_\mathrm{T} > n_\mathrm{R}$, it is not clear how to order the transmit antennas. Nevertheless, we show that it is optimal to restrict to signaling methods that, for each channel use, choose n_R transmit antennas as "active" signaling antennas, and set the remaining antennas to either full power A or to 0 according to a given rule. The "active" antennas can then be used in the same way as suggested in earlier works for MIMO channels that have a square full-rank channel matrix [8].

We introduce some further notation. We define

$$\mathcal{U} \triangleq \big\{\mathcal{I} \subseteq \{1, \ldots, n_\mathrm{T}\} \colon |\mathcal{I}| = n_\mathrm{R}\big\}, \tag{7}$$

and for every $\mathcal{I} = \{i_1, \ldots, i_{n_\mathrm{R}}\} \in \mathcal{U}$ with $i_1 < \cdots < i_{n_\mathrm{R}}$, we define

$$\mathsf{H}_\mathcal{I} \triangleq \big[\mathbf{h}_{i_1}, \ldots, \mathbf{h}_{i_{n_\mathrm{R}}}\big]. \tag{8}$$

Throughout this paper, we assume that every $\mathsf{H}_\mathcal{I}$, $\mathcal{I} \in \mathcal{U}$, is of full rank n_R, i.e., that any n_R column vectors in H are linearly independent. To simplify derivations, we further assume that for all $\mathcal{I} \in \mathcal{U}$ and $j \in \{1, \ldots, n_\mathrm{T}\} \setminus \mathcal{I}$,

$$\mathbf{1}_{n_\mathrm{R}}^\intercal \mathsf{H}_\mathcal{I}^{-1} \mathbf{h}_j \ne 1, \tag{9}$$

where $\mathbf{1}_{n_\mathsf{R}}$ denotes the n_R-dimensional vector $(1,\dots,1)^\mathsf{T}$.

II. Capacity and Minimum-Energy Signaling

The capacity of the channel (1) is [10]

$$\mathsf{C}_\mathsf{H}(\mathsf{A},\mathsf{E}) = \sup_{Q_\mathbf{X}} \mathrm{I}(\mathbf{X};\mathbf{Y}) \tag{10}$$

where the supremum is over all distributions $Q_\mathbf{X}$ on $\mathbf{X}$ satisfying (3) and (4). Define

$$\bar{\mathbf{X}} \triangleq \mathsf{H}\mathbf{X} \tag{11}$$

and notice that $\bar{\mathbf{X}}$ takes value in the zonotope [11]

$$\mathcal{R}(\mathsf{H}) \triangleq \left\{ \sum_{k=1}^{n_\mathsf{T}} \delta_k \mathbf{h}_k \colon \delta_1,\dots,\delta_{n_\mathsf{T}} \in [0,\mathsf{A}] \right\}, \tag{12}$$

which has volume [12]

$$\mathsf{V}_\mathsf{H} \triangleq \sum_{\mathcal{I}\in\mathcal{U}} \det \mathsf{H}_\mathcal{I}. \tag{13}$$

Define for each $\mathcal{I}\in\mathcal{U}$ the parallelepiped

$$\mathcal{D}_\mathcal{I} \triangleq \left\{ \sum_{i\in\mathcal{I}} \delta_i \mathbf{h}_i \colon \delta_i \in [0,\mathsf{A}], \quad \forall i \in \mathcal{I} \right\}; \tag{14}$$

the coefficient

$$s_\mathcal{I} \triangleq \sum_{j\in\{1,\dots,n_\mathsf{T}\}\setminus\mathcal{I}} \mathbb{1}\big\{ \mathbf{1}_{n_\mathsf{R}}^\mathsf{T} \mathsf{H}_\mathcal{I}^{-1}\mathbf{h}_j > 1 \big\}; \tag{15}$$

and the vector

$$\mathbf{v}_\mathcal{I} \triangleq \mathsf{A} \sum_{j\in\{1,\dots,n_\mathsf{T}\}\setminus\mathcal{I}} \mathbb{1}\big\{ \mathbf{1}_{n_\mathsf{R}}^\mathsf{T} \mathsf{H}_\mathcal{I}^{-1}\mathbf{h}_j > 1 \big\}\mathbf{h}_j; \tag{16}$$

where $\mathbb{1}\{\cdot\}$ denotes the indicator function.

Lemma 1:

1) For any $\mathcal{I},\mathcal{J}\in\mathcal{U}$, $\mathcal{I}\neq\mathcal{J}$, the intersection of $\mathbf{v}_\mathcal{I}+\mathcal{D}_\mathcal{I}$ and $\mathbf{v}_\mathcal{J}+\mathcal{D}_\mathcal{J}$ has Lebesgue measure zero.
2) The union

$$\bigcup_{\mathcal{I}\in\mathcal{U}} \mathbf{v}_\mathcal{I}+\mathcal{D}_\mathcal{I} = \mathcal{R}(\mathsf{H}). \tag{17}$$

3) For any $\mathcal{I}\in\mathcal{U}$, any point in $\mathbf{v}_\mathcal{I}+\mathcal{D}_\mathcal{I}$ is achieved with minimum total input power by an input vector $\mathbf{x}$ satisfying, for every $j\in\{1,\dots,n_\mathsf{T}\}\setminus\mathcal{I}$,

$$x_j = \mathsf{A}\cdot\mathbb{1}\big\{ \mathbf{1}_{n_\mathsf{R}}^\mathsf{T} \mathsf{H}_\mathcal{I}^{-1}\mathbf{h}_j > 1 \big\}. \tag{18}$$

Proof: Omitted. ■

Figure 1 shows the partitionings of $\mathcal{R}(\mathsf{H})$ into the union (17) for two examples of channel matrices H.

Using the above lemma, we can characterize the capacity $\mathsf{C}_\mathsf{H}(\mathsf{A},\mathsf{E})$ as an optimization problem over $\bar{\mathbf{X}}$.

Proposition 2: The capacity of the MIMO optical intensity channel is given by

$$\mathsf{C}_\mathsf{H}(\mathsf{A},\mathsf{E}) = \sup_{Q_{\bar{\mathbf{X}}}} \mathrm{I}(\bar{\mathbf{X}};\mathbf{Y}), \tag{19}$$

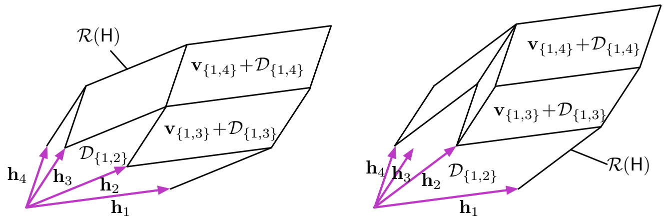

Fig. 1. Partitioning of $\mathcal{R}(\mathsf{H})$ into the union (17) for two 4-by-2 MIMO examples. The example on the left is for $\mathsf{H}=[7,5,2,1;1,2,2.9,3]$ and the example on the right for $\mathsf{H}=[7,5,2,1;1,3,2.9,3]$.

where the supremum is over all distributions[1] $Q_{\bar{\mathbf{X}}}$ over $\mathcal{R}(\mathsf{H})$ subject to the power constraint

$$\mathsf{E}_U\big[\mathsf{A}s_U + \big\|\mathsf{H}_U^{-1}\big(\mathsf{E}\big[\bar{\mathbf{X}}\big|U\big]-\mathbf{v}_U\big)\big\|_1\big] \le \alpha\mathsf{A}, \tag{20}$$

where U is a random variable taking value in $\mathcal{U}$ such that $U=\mathcal{I}$ if, and only if, $\bar{\mathbf{X}}$ lies in the interior of $\mathbf{v}_\mathcal{I}+\mathcal{D}_\mathcal{I}$.

Proof: Notice that $\bar{\mathbf{X}}$ is a function of $\mathbf{X}$ and that the Markov chain $\mathbf{X} \multimap\!\!- \bar{\mathbf{X}} \multimap\!\!- \mathbf{Y}$ holds. Therefore, $\mathrm{I}(\bar{\mathbf{X}};\mathbf{Y}) = \mathrm{I}(\mathbf{X};\mathbf{Y})$. By the last part of Lemma 1, $\mathcal{R}(\mathsf{H})$ can be decomposed into the shifted parallelepipeds $\{\mathbf{v}_\mathcal{I}+\mathcal{D}_\mathcal{I}\}$, whereas, by the first part of Lemma 1, the minimum energy required to achieve an image point $\bar{\mathbf{x}}$ in a given parallelepiped $\mathbf{v}_\mathcal{I}+\mathcal{D}_\mathcal{I}$ is:

$$\mathsf{A}s_\mathcal{I} + \|\mathsf{H}_\mathcal{I}^{-1}(\bar{\mathbf{x}}-\mathbf{v}_\mathcal{I})\|_1. \tag{21}$$

The proposition follows by taking expectation over (21) ■

III. Capacity Results

This section presents upper and lower bounds, as well as asymptotic expressions for the capacity.

Let $\mathbf{q}$ be a probability vector on $\mathcal{U}$ with entries

$$q_\mathcal{I} \triangleq \frac{\det \mathsf{H}_\mathcal{I}}{\mathsf{V}_\mathsf{H}}, \quad \mathcal{I}\in\mathcal{U}. \tag{22}$$

Further define

$$\alpha_{\text{th}} \triangleq \frac{n_\mathsf{R}}{2} + \sum_{\mathcal{I}\in\mathcal{U}} s_\mathcal{I} q_\mathcal{I}. \tag{23}$$

We shall see that the threshold α_{th} determines whether $\bar{\mathbf{X}}$ can be made uniform over $\mathcal{R}(\mathsf{H})$ or not.

A. Lower Bounds

Theorem 3: If $\alpha \ge \alpha_{\text{th}}$, then

$$\mathsf{C}_\mathsf{H}(\mathsf{A},\alpha\mathsf{A}) \ge \frac{1}{2}\log\left(1+\frac{\mathsf{A}^{2n_\mathsf{R}}\mathsf{V}_\mathsf{H}^2}{(2\pi e)^{n_\mathsf{R}}}\right). \tag{24}$$

If $\alpha < \alpha_{\text{th}}$, then

$$\mathsf{C}_\mathsf{H}(\mathsf{A},\alpha\mathsf{A}) \ge \frac{1}{2}\log\left(1+\frac{\mathsf{A}^{2n_\mathsf{R}}\mathsf{V}_\mathsf{H}^2}{(2\pi e)^{n_\mathsf{R}}}\cdot e^{2\nu}\right), \tag{25}$$

[1] To make the statement simpler, we further require that $\bar{\mathbf{X}}$ *not* be on the surface of any $\mathbf{v}_\mathcal{I}+\mathcal{D}_\mathcal{I}$. This does not cause any loss of optimality, thanks to the supremum in (19).

where

$$\nu \triangleq \sup_{\lambda \in \left(\max\{0, \frac{n_R}{2}+\alpha-\alpha_{\text{th}}\}, \min\{\frac{n_R}{2}, \alpha\}\right)} \left\{ n_R \left(1 - \log \frac{\mu}{1-e^{-\mu}} - \frac{\mu\, e^{-\mu}}{1-e^{-\mu}}\right) - \inf_{\mathbf{p}} D(\mathbf{p}\|\mathbf{q}) \right\}, \tag{26}$$

where μ is the unique solution to the following equation:

$$\frac{1}{\mu} - \frac{e^{-\mu}}{1-e^{-\mu}} = \frac{\lambda}{n_R}, \tag{27}$$

and where the infimum is over all probability vectors $\mathbf{p}$ on $\mathcal{U}$ such that

$$\sum_{\mathcal{I}\in\mathcal{U}} p_{\mathcal{I}} s_{\mathcal{I}} = \alpha - \lambda. \tag{28}$$

B. Upper Bounds

Theorem 4: Ignoring the average-power constraint (4b),

$$C_H(A, \alpha A) \le n_R \log\left(1 + \frac{A}{\sqrt{2\pi e}}\right) + \log V_H. \tag{29}$$

Theorem 5: If $\alpha < \alpha_{\text{th}}$,

$$\begin{aligned} C_H(A,\alpha A) \le \sup_{\mathbf{p}} \inf_{\mu>0} \Bigg\{ & n_R \log\left(1 + \frac{A}{\sqrt{2\pi e}} \frac{1-e^{-\mu}}{\mu}\right) + \log V_H \\ & + \frac{\mu n_R}{A\sqrt{2\pi}}\left(1 - e^{-\frac{A^2}{2}}\right) \\ & + \mu\left(\alpha - \sum_{\mathcal{I}\in\mathcal{U}} p_{\mathcal{I}} s_{\mathcal{I}}\right) - D(\mathbf{p}\|\mathbf{q}) \Bigg\}, \end{aligned} \tag{30}$$

where the supremum is over all probability vectors $\mathbf{p}$ on $\mathcal{U}$.

Theorem 6: If $\alpha < \alpha_{\text{th}}$,

$$\begin{aligned} C_H(A,\alpha A) \le \sup_{\mathbf{p}} \inf_{\delta,\mu>0} \Bigg\{ & n_R \log\left(A \cdot \frac{e^{\frac{\mu\delta}{A}} - e^{-\mu(1+\frac{\delta}{A})}}{\sqrt{2\pi}\mu(1-2\,\mathcal{Q}(\delta))}\right) + \log V_H \\ & - \frac{n_R}{2} + n_R\, \mathcal{Q}(\delta) + \frac{\delta}{\sqrt{2\pi}} e^{-\frac{-\delta^2}{2}} \\ & + \frac{\mu n_R}{A\sqrt{2\pi}}\left(e^{-\frac{\delta^2}{2}} - e^{-\frac{(A+\delta)^2}{2}}\right) \\ & + \mu\left(\alpha - \sum_{\mathcal{I}\in\mathcal{U}} p_{\mathcal{I}} s_{\mathcal{I}}\right) - D(\mathbf{p}\|\mathbf{q}) \Bigg\}, \end{aligned} \tag{31}$$

where $\mathcal{Q}(\cdot)$ denotes the Gaussian $\mathcal{Q}$-function and the supremum is over all probability vectors $\mathbf{p}$ on $\mathcal{U}$.

C. Asymptotic High-SNR Capacity Expressions

Theorem 7: If $\alpha \ge \alpha_{\text{th}}$,

$$\lim_{A\to\infty} \left\{ C_H(A,\alpha A) - n_R \log A \right\} = \frac{1}{2}\log\left(\frac{V_H^2}{(2\pi e)^{n_R}}\right). \tag{32}$$

If $\alpha < \alpha_{\text{th}}$,

$$\begin{aligned} \lim_{A\to\infty} & \left\{ C_H(A,\alpha A) - n_R \log A \right\} \\ &= \frac{1}{2}\log\left(\frac{V_H^2}{(2\pi e)^{n_R}}\right) \\ &\quad + \sup_{\lambda \in \left(\max\{0, \frac{n_R}{2}+\alpha-\alpha_{\text{th}}\}, \min\{\frac{n_R}{2}, \alpha\}\right)} \Bigg\{ 1 \\ &\quad - \log\frac{\mu}{1-e^{-\mu}} - \frac{\mu\, e^{-\mu}}{1-e^{-\mu}} - \inf_{\mathbf{p}} D(\mathbf{p}\|\mathbf{q}) \Bigg\} \end{aligned} \tag{33}$$

where μ and $\mathbf{p}$ are the same as in Theorem 3.

Figure 2 depicts the derived lower and upper bounds for an example with $\alpha < \alpha_{\text{th}}$. The bounds match asymptotically at high SNR.

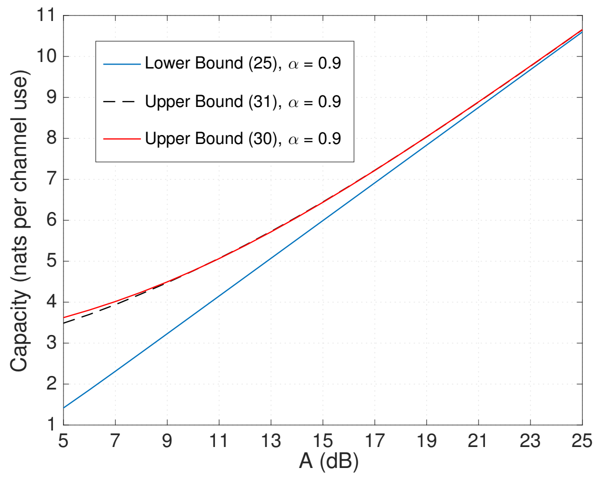

Fig. 2. Bounds on capacity of 3×2 MIMO channel with channel matrix $H = [1, 1.5, 3; 2, 2, 1]$, and average-to-peak power ratio $\alpha = 0.9$. Note that the threshold of the channel is $\alpha_{\text{th}} = 1.476$.

IV. Proofs

A. Derivation of Lower Bounds

For any choice of the random vector $\bar{\mathbf{X}}$ over $\mathcal{R}(H)$ that satisfies (20), the following holds:

$$\begin{aligned} C_H(A,\alpha A) &\ge I(\bar{\mathbf{X}}; \bar{\mathbf{X}} + \mathbf{Z}) & (34)\\ &= h(\bar{\mathbf{X}} + \mathbf{Z}) - h(\mathbf{Z}) & (35)\\ &\ge \frac{1}{2}\log\left(e^{2\,h(\bar{\mathbf{X}})} + e^{2\,h(\mathbf{Z})}\right) - h(\mathbf{Z}) & (36)\\ &= \frac{1}{2}\log\left(1 + \frac{e^{2\,h(\bar{\mathbf{X}})}}{(2\pi e)^{n_R}}\right) & (37)\\ &= \frac{1}{2}\log\left(1 + \frac{e^{2\,h(\bar{\mathbf{X}}|U)}}{(2\pi e)^{n_R}} \cdot e^{2\,H(U)}\right), & (38) \end{aligned}$$

where (36) holds by the Entropy Power Inequality [13] and (38) holds because

$$h(\bar{\mathbf{X}}) = I(\bar{\mathbf{X}}; U) + h(\bar{\mathbf{X}}|U) = H(U) + h(\bar{\mathbf{X}}|U), \tag{39}$$

where we used that $H(U|\bar{\mathbf{X}}) = 0$.

The lower bounds in this paper are then obtained from (37) and (38). When $\alpha < \alpha_{\text{th}}$, we choose $\lambda \in \big(\max\{0, \frac{n_{\mathsf{R}}}{2} + \alpha - \alpha_{\text{th}}\}, \min\{\frac{n_{\mathsf{R}}}{2}, \alpha\}\big)$, the probability vector $\mathbf{p}$ such that it satisfies (28), and μ as the unique solution to (27). Then we apply (38) and choose for each $\mathcal{I}$ the distribution $Q_{\bar{\mathbf{X}}|U=\mathcal{I}}$ as the n_{R}-dimensional product cut-exponential distribution rotated by the matrix $\mathsf{H}_{\mathcal{I}}$:

$$Q_{\bar{\mathbf{X}}|U=\mathcal{I}}(\bar{\mathbf{x}}) = \frac{1}{\mathsf{A}^{n_{\mathsf{R}}} \det \mathsf{H}_{\mathcal{I}}} \cdot \left(\frac{\mu}{1-e^{-\mu}}\right)^{n_{\mathsf{R}}} e^{\frac{-\mu \|\mathsf{H}_{\mathcal{I}}^{-1}(\bar{\mathbf{x}} - \mathsf{A}\mathbf{v}_{\mathcal{I}})\|_1}{\mathsf{A}}}. \tag{40}$$

This is the entropy-maximizing distribution under a total average-power constraint on $\mathsf{H}_{\mathcal{I}}^{-1}\bar{\mathbf{X}}$. The resulting expression is finally optimized over the distribution Q_U.

When $\alpha \geq \alpha_{\text{th}}$, then we apply (37) and choose $\bar{\mathbf{X}}$ uniform over $\mathcal{R}(\mathsf{H})$. It can be verified that this choice satisfies the power constraint whenever $\alpha \geq \alpha_{\text{th}}$.

B. Derivation of Upper Bounds

Let $\bar{\mathbf{X}}^\star$ be a maximizer in (19). Then,

$$\begin{aligned} \mathsf{C}_{\mathsf{H}}(\mathsf{A}, \mathsf{E}) &= \mathrm{I}(\bar{\mathbf{X}}^\star; \bar{\mathbf{X}}^\star + \mathbf{Z}) && (41)\\ &\leq \mathrm{I}(\bar{\mathbf{X}}^\star; \bar{\mathbf{X}}^\star + \mathbf{Z}, U^\star) && (42)\\ &\leq \mathrm{H}(U^\star) + \mathrm{I}(\bar{\mathbf{X}}^\star; \bar{\mathbf{X}}^\star + \mathbf{Z}|U^\star). && (43)\end{aligned}$$

Moreover, for each subset $\mathcal{I} \in \mathcal{U}$ of size $|\mathcal{I}| = n_{\mathsf{R}}$:

$$\begin{aligned} &\mathrm{I}(\bar{\mathbf{X}}^\star; \bar{\mathbf{X}}^\star + \mathbf{Z}|U^\star = \mathcal{I})\\ &\quad= \mathrm{I}(\bar{\mathbf{X}}^\star; (\bar{\mathbf{X}}^\star - \mathbf{v}_{\mathcal{I}}) + \mathbf{Z}|U^\star = \mathcal{I}) && (44)\\ &\quad= \mathrm{I}(\bar{\mathbf{X}}^\star; \mathsf{H}_{\mathcal{I}}^{-1}(\bar{\mathbf{X}}^\star - \mathbf{v}_{\mathcal{I}}) + \mathsf{H}_{\mathcal{I}}^{-1}\mathbf{Z}|U^\star = \mathcal{I}) && (45)\\ &\quad= \mathrm{I}(\bar{\mathbf{X}}_{\mathcal{I}}^\star; \bar{\mathbf{X}}_{\mathcal{I}}^\star + \mathbf{Z}_{\mathcal{I}}|U^\star = \mathcal{I}), && (46)\end{aligned}$$

where we defined

$$\mathbf{Z}_{\mathcal{I}} \triangleq \mathsf{H}_{\mathcal{I}}^{-1}\mathbf{Z}, \tag{47}$$

$$\bar{\mathbf{X}}_{\mathcal{I}}^\star \triangleq \mathsf{H}_{\mathcal{I}}^{-1}(\bar{\mathbf{X}}^\star - \mathbf{v}_{\mathcal{I}}). \tag{48}$$

To further bound the term in (46), we then use the duality upper-bounding technique with a product output distribution

$$R_{\mathcal{I}}(\mathbf{y}_{\mathcal{I}}) = \prod_{\ell=1}^{n_{\mathsf{R}}} R_{\mathcal{I},\ell}(y_{\mathcal{I},\ell}). \tag{49}$$

Denoting by $W_{\mathcal{I}}(\cdot|\bar{\mathbf{X}}_{\mathcal{I}}^\star)$ the transition law of the MIMO channel $\bar{\mathbf{X}}_{\mathcal{I}}^\star \mapsto \mathbf{Y}_{\mathcal{I}} \triangleq (\bar{\mathbf{X}}_{\mathcal{I}}^\star + \mathbf{Z}_{\mathcal{I}})$ and by $W_{\mathcal{I},\ell}(\cdot|\bar{X}_{\mathcal{I},\ell}^\star)$ the marginal transition law of its ℓ-th component, we have:

$$\begin{aligned} &\mathrm{I}(\bar{\mathbf{X}}_{\mathcal{I}}^\star; \bar{\mathbf{X}}_{\mathcal{I}}^\star + \mathbf{Z}_{\mathcal{I}}|U^\star = \mathcal{I})\\ &\leq \sup_{Q_{\bar{\mathbf{X}}_{\mathcal{I}}^\star}} \mathsf{E}_{\bar{\mathbf{X}}_{\mathcal{I}}^\star|U^\star=\mathcal{I}}\big[\mathsf{D}\big(W_{\mathcal{I}}(\cdot|\bar{\mathbf{X}}_{\mathcal{I}}^\star)\big\|R_{\mathcal{I}}(\cdot)\big)\big] && (50)\\ &\leq -\inf_{Q_{\bar{\mathbf{X}}_{\mathcal{I}}^\star}} \mathsf{h}\big(\bar{\mathbf{X}}_{\mathcal{I}}^\star + \mathbf{Z}_{\mathcal{I}}\big|\bar{\mathbf{X}}_{\mathcal{I}}^\star, U^\star = \mathcal{I}\big)\\ &\quad -\inf_{Q_{\bar{\mathbf{X}}_{\mathcal{I}}^\star}} \mathsf{E}_{\bar{\mathbf{X}}_{\mathcal{I}}^\star|U^\star=\mathcal{I}}\left[\sum_{\ell=1}^{n_{\mathsf{R}}} \mathsf{E}_{W_{\mathcal{I}}(\mathbf{Y}_{\mathcal{I}}|\bar{\mathbf{X}}_{\mathcal{I}}^\star)}[\log R_{\mathcal{I},\ell}(Y_{\mathcal{I},\ell})]\right] && (51)\\ &= \frac{1}{2}\log\big((2\pi e)^{n_{\mathsf{R}}} \det \mathsf{H}_{\mathcal{I}}\mathsf{H}_{\mathcal{I}}^{\mathsf{T}}\big)\\ &\quad -\sum_{\ell=1}^{n_{\mathsf{R}}} \inf_{Q_{\bar{X}_{\mathcal{I},\ell}^\star}} \mathsf{E}_{Q_{\bar{X}_{\mathcal{I},\ell}^\star|U^\star=\mathcal{I}}}\Big[\mathsf{E}_{W_{\mathcal{I},\ell}(Y_{\mathcal{I},\ell}|\bar{X}_{\mathcal{I},\ell}^\star)}[\log R_{\mathcal{I},\ell}(Y_{\mathcal{I},\ell})]\Big], && (52)\end{aligned}$$

where the last equality holds because

$$\begin{aligned} &\mathsf{h}(\bar{\mathbf{X}}_{\mathcal{I}}^\star + \mathbf{Z}_{\mathcal{I}}|\bar{\mathbf{X}}_{\mathcal{I}}^\star, U^\star = \mathcal{I})\\ &\quad= \mathsf{h}(\mathbf{Z}_{\mathcal{I}}) && (53)\\ &\quad= -\frac{1}{2}\log\big((2\pi e)^{n_{\mathsf{R}}} \det \mathsf{H}_{\mathcal{I}}\mathsf{H}_{\mathcal{I}}^{\mathsf{T}}\big), && (54)\end{aligned}$$

and by the total law of expectation which allows to eliminate all transition laws except for $W_{\mathcal{I},\ell}(\cdot|\cdot)$.

The upper bounds in this paper are then obtained by combining (43) with (46) and (52), by optimizing over the probability vector $\mathbf{p}$ on $\mathcal{U}$, and by picking appropriate choices for the output distribution $R_{\mathcal{I},\ell}(\cdot)$.

To prove Theorem 4, we choose

$$R_{\mathcal{I},\ell}(y) = \begin{cases} \frac{1}{\mathsf{A}+\sqrt{2\pi e}} & \text{if } y \in [0, \mathsf{A}],\\ \frac{1}{\mathsf{A}+\sqrt{2\pi e}}\, e^{-\frac{y^2}{2}} & \text{otherwise,}\end{cases} \tag{55}$$

which yields, irrespective of the value of $\bar{x}_{\mathcal{I},\ell}^\star$:

$$-\mathsf{E}_{W_{\mathcal{I},\ell}(Y_{\mathcal{I},\ell}|\bar{x}_{\mathcal{I},\ell}^\star)}[\log R_{\mathcal{I},\ell}(Y_{\mathcal{I},\ell})] \leq \log\big(\mathsf{A} + \sqrt{2\pi e}\big). \tag{56}$$

In this case, the optimizing Q_U^* is given by the probability vector $\mathbf{p}$ over $\mathcal{U}$ with components $\{p_{\mathcal{I}} = q_{\mathcal{I}},\ \mathcal{I} \in \mathcal{U}\}$.

To prove Theorem 5, we choose

$$R_{\mathcal{I},\ell}(y) = \begin{cases} \frac{1}{\sqrt{2\pi e}+\frac{\mathsf{A}}{\mu}\cdot(1-e^{-\mu})} \cdot e^{-\frac{\mu y}{\mathsf{A}}} & \text{if } y \in [0, \mathsf{A}],\\ \frac{1}{\sqrt{2\pi e}+\frac{\mathsf{A}}{\mu}\cdot(1-e^{-\mu})} \cdot e^{-\frac{y^2}{2}} & \text{otherwise.}\end{cases} \tag{57}$$

Following the steps in [14, Appendix B], we obtain:

$$\begin{aligned} &-\mathsf{E}_{Q_{\bar{X}_{\mathcal{I},\ell}^\star|U^\star=\mathcal{I}}}\Big[\mathsf{E}_{W_{\mathcal{I},\ell}(Y_{\mathcal{I},\ell}|\bar{X}_{\mathcal{I},\ell}^\star)}[\log R_{\mathcal{I},\ell}(Y_{\mathcal{I},\ell})]\Big]\\ &\quad\leq \log\left(\sqrt{2\pi e} + \frac{\mathsf{A}}{\mu}(1-e^{-\mu})\right)\\ &\qquad + \frac{\mu}{\mathsf{A}\sqrt{2\pi}}\Big(1 - e^{-\frac{\mathsf{A}^2}{2}}\Big) + \frac{\mu}{\mathsf{A}}\,\mathsf{E}_{Q_{\bar{X}_{\mathcal{I},\ell}^\star|U^\star=\mathcal{I}}}\big[\bar{X}_{\mathcal{I},\ell}^\star\big]. && (58)\end{aligned}$$

The proof is concluded by combining (58) with (43), (46), and (52), and by noting that

$$\begin{aligned} &\mathsf{E}_{Q_{U^\star}}\left[\frac{\mu}{\mathsf{A}}\sum_{i=1}^{n_{\mathsf{R}}} \mathsf{E}_{Q_{\bar{X}_{\mathcal{I},\ell}^\star|U^\star=\mathcal{I}}}\big[\bar{X}_{\mathcal{I},\ell}^\star\big]\right]\\ &\quad= \mu\,\mathsf{E}_{Q_{U^\star}}\big[\big\|\mathsf{H}_U^{-1}\big(\mathsf{E}\big[\bar{\mathbf{X}}\,\big|\,U\big] - \mathbf{v}_U\big)\big\|_1\big] && (59)\\ &\quad\leq \mu\big(\alpha - \mathsf{E}_{Q_{U^\star}}[s_{U^\star}]\big) && (60)\end{aligned}$$

where (59) follows by (48), and (60) by (20).

To prove Theorem 6, we choose

$$R_{\mathcal{I},\ell}(y) = \begin{cases} \frac{1}{\sqrt{2\pi}}\, e^{-\frac{y^2}{2}} & \text{if } y \in (-\infty, -\delta),\\ \frac{\mu}{\mathsf{A}} \cdot \frac{1-2\,\mathcal{Q}(\delta)}{e^{\frac{\mu\delta}{\mathsf{A}}} - e^{-\mu(1+\frac{\delta}{\mathsf{A}})}}\, e^{-\frac{\mu y}{\mathsf{A}}} & \text{if } y \in [-\delta, \mathsf{A}+\delta],\\ \frac{1}{\sqrt{2\pi}}\, e^{-\frac{(y-\mathsf{A})^2}{2}} & \text{if } y \in (\mathsf{A}+\delta, +\infty).\end{cases} \tag{61}$$

Following the steps in the proof of [3, Theorem 3], we obtain:

$$\begin{aligned}
&-\mathsf{E}_{Q_{\bar{X}^\star_{\mathcal{I},\ell}|U^\star=\mathcal{I}}}\Big[\mathsf{E}_{W_{\mathcal{I},\ell}(Y_{\mathcal{I},\ell}|\bar{X}^\star_{\mathcal{I},\ell})}[\log R_{\mathcal{I},\ell}(Y_{\mathcal{I},\ell})]\Big] \\
&\le \log\left(\mathsf{A}\cdot\frac{e^{\frac{\mu\delta}{\mathsf{A}}}-e^{-\mu(1+\frac{\delta}{\mathsf{A}})}}{\mu(1-2\,\mathcal{Q}(\delta))}\right)+\frac{\delta}{\sqrt{2\pi}}\,e^{-\frac{-\delta^2}{2}}+\mathcal{Q}(\delta) \\
&\quad+\frac{\mu}{\mathsf{A}\sqrt{2\pi}}\left(e^{-\frac{\delta^2}{2}}-e^{-\frac{(\mathsf{A}+\delta)^2}{2}}\right)+\frac{\mu}{\mathsf{A}}\,\mathsf{E}_{Q_{\bar{X}^\star_{\mathcal{I},\ell}|U^\star=\mathcal{I}}}\big[\bar{X}^\star_{\mathcal{I},\ell}\big].
\end{aligned} \tag{62}$$

The proof is concluded by combining (58) with (43), (46), (52), and (60).

C. Derivation of Asymptotic High-SNR Capacity

The proof for the case $\alpha \ge \alpha_{\text{th}}$ is straightforward and omitted.

For $\alpha < \alpha_{\text{th}}$, the fact that the LHS of (33) cannot be smaller than its RHS follows directly from (25). In the following, we use inequality (30) to show the reverse direction. We have:

$$\begin{aligned}
&\mathsf{C}_{\mathsf{H}}(\mathsf{A},\alpha\mathsf{A})-n_{\text{R}}\log\mathsf{A} \\
&\le \sup_{\mathbf{p}}\inf_{\mu>0}\Bigg\{-\mathsf{D}(\mathbf{p}\|\mathbf{q})+\log\mathsf{V}_{\mathsf{H}}+n_{\text{R}}\log\frac{1-e^{-\mu}}{\mu} \\
&\qquad -\frac{n_{\text{R}}}{2}\log 2\pi e+\mu\left(\alpha-\sum_{\mathcal{I}\in\mathcal{U}}p_{\mathcal{I}}s_{\mathcal{I}}\right)+o(\mathsf{A})\Bigg\}.
\end{aligned} \tag{63}$$

Let

$$\lambda = \alpha - \sum_{\mathcal{I}\in\mathcal{U}} p_{\mathcal{I}}s_{\mathcal{I}}, \tag{64}$$

and choose μ^* to be the unique solution of equation (27) in μ. Then proceeding from (63), we obtain:

$$\begin{aligned}
&\mathsf{C}_{\mathsf{H}}(\mathsf{A},\alpha\mathsf{A})-n_{\text{R}}\log\mathsf{A} \\
&\le \sup_{\mathbf{p}}\Bigg\{\frac{1}{2}\log\left(\frac{\mathsf{V}_{\mathsf{H}}^2}{(2\pi e)^{n_{\text{R}}}}\right)-\mathsf{D}(\mathbf{p}\|\mathbf{q})+n_{\text{R}}\log\frac{1-e^{-\mu^*}}{\mu^*} \\
&\qquad +\mu^*\lambda+o(\mathsf{A})\Bigg\} &(65)\\
&= \sup_{\mathbf{p}}\Bigg\{\frac{1}{2}\log\left(\frac{\mathsf{V}_{\mathsf{H}}^2}{(2\pi e)^{n_{\text{R}}}}\right)-\mathsf{D}(\mathbf{p}\|\mathbf{q}) \\
&\qquad +n_{\text{R}}\left(1-\log\frac{\mu^*}{1-e^{-\mu^*}}-\frac{\mu\,e^{-\mu^*}}{1-e^{-\mu^*}}\right)+o(\mathsf{A})\Bigg\}. &(66)
\end{aligned}$$

The proof is concluded by taking $\mathsf{A}\to\infty$.

V. Conclusion

In this paper, we derive upper and lower bounds on the capacity of the MIMO free-space optical intensity channel when the transmitter has more antennas than the receiver ($n_{\text{T}} > n_{\text{R}}$). In our model, channel inputs are subject to an individual peak-power constraint for each antenna and an average sum-power constraint over all antennas. The bounds match asymptotically in the high-SNR regime and show that the high-SNR asymptotic capacity saturates in the total average power $\alpha\mathsf{A}$ for all $\alpha \ge \alpha_{\text{th}}$. The reason is that this threshold suffices to attain a uniform distribution over the image set $\mathcal{R}(\mathsf{H})$ of the channel matrix H.

To derive our capacity bounds, we provide an alternative expression for this capacity. It is based on the insight that the optimal (most energy-efficient) signaling strategy is to choose a set $\mathcal{I}\in\mathcal{U}$, set all inputs $\{x_j : j \in \{1,\dots,n_{\text{T}}\}\setminus\mathcal{I}\}$ either to 0 or to full power A according to the rule in (3), and signal with the n_{R} antennas in $\mathcal{D}_{\mathcal{I}}$ as for a full-rank $n_{\text{R}}\times n_{\text{R}}$ channel matrix $\mathsf{H}_{\mathcal{I}}$ [8]. In the MISO case, this means that if a given antenna is used for signaling, all stronger antennas need to send at full power A [7].

Acknowledgment

The work of L. Li and M. Wigger was supported by the ERC under grant agreement 715111.

References

[1] Michèle A. Wigger, "Bounds on the capacity of free-space optical intensity channels," Master's thesis, Signal and Inf. Proc. Lab., ETH Zürich, Switzerland, Mar. 2003.

[2] Andrew L. McKellips, "Simple tight bounds on capacity for the peak-limited discrete-time channel," in *Proc. IEEE Int. Symp. Inf. Theory*, Chicago, IL, USA, Jun. 27 – Jul. 2, 2004, p. 348.

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

[4] Andrew Thangaraj, Gerhard Kramer, and Georg Böcherer, "Capacity bounds for discrete-time, amplitude-constrained, additive white Gaussian noise channels," *IEEE Trans. Inf. Theory*, vol. 63, no. 7, pp. 4172–4182, Jul. 2017.

[5] Anas Chaaban, Zouheir Rezki, and Mohamed-Slim 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.

[6] Stefan M. Moser, Ligong Wang, and Michèle Wigger, "Asymptotic high-SNR capacity of MISO optical intensity channels," in *Proc. IEEE Inf. Theory Workshop*, Kaohsiung, Taiwan, Nov. 6–10, 2017, pp. 86–90.

[7] Stefan M. Moser, Ligong Wang, and Michèle Wigger, "Capacity results on multiple-input single-output wireless optical channels," 2018, to app. in *IEEE Trans. Inf. Theory*. Available: http://moser-isi.ethz.ch/publications.html

[8] Stefan M. Moser, Michail Mylonakis, Ligong Wang, and Michèle Wigger, "Asymptotic capacity results for MIMO wireless optical communication," in *Proc. IEEE Int. Symp. Inf. Theory*, Aachen, Germany, Jun. 25–30, 2017, pp. 536–540.

[9] Anas Chaaban, Zouheir Rezki, and Mohamed-Slim Alouini, "MIMO intensity-modulation channels: Capacity bounds and high SNR characterization," in *Proc. IEEE Int. Conf. Commun.*, Paris, France, May 21–25, 2017, pp. 1–6.

[10] Claude E. Shannon, "A mathematical theory of communication," *Bell Syst. Tech. J.*, vol. 27, pp. 379–423 and 623–656, Jul. and Oct. 1948.

[11] Matthias Beck and Sinai Robins, *Computing the Continuous Discretely*. Springer Verlag, 2015.

[12] Ram Zamir and Meir Feder, "On the volume of the Minkowski sum of line sets and the Entropy-Power Inequality," *IEEE Trans. Inf. Theory*, vol. 44, no. 7, pp. 3039–3043, Nov. 1998.

[13] Thomas M. Cover and Joy A. Thomas, *Elements of Information Theory*, 2nd ed. New York, NY, USA: Wiley, 2006.

[14] Longguang Li, Stefan M. Moser, Ligong Wang, and Michèle Wigger, "The MISO free-space optical channel at low and moderate SNR," in *Proc. Annu. Conf. Inf. Sci. Syst.*, Princeton, NJ, USA, Mar. 21–23, 2018, pp. 1–6.