Signaling for MISO Channels Under First- and Second-Moment Constraints

Shuai Ma*, Stefan M. Moser†‡, Ligong Wang§, and Michèle Wigger*
mashuai001@cumt.edu.cn, moser@isi.ee.ethz.ch, ligong.wang@ensea.fr, michele.wigger@telecom-paris.fr

*LTCI, Telecom Paris, IP Paris, 91120 Palaiseau, France
†Signal and Information Processing Lab, ETH Zurich, Switzerland
‡Institute of Communications Engineering at National Yang Ming Chiao Tung University, Hsinchu, Taiwan
§ETIS—CY Cergy Paris Univeristy, ENSEA, CNRS, Cergy-Pontoise, France

***Abstract*—Consider a multiple-input single-output system, where the nonnegative, peak-limited inputs $X_1, \ldots, X_{n_\mathrm{T}} \in [0, \mathsf{A}]$ are subject to first- and second-moment sum-constraints on all antennas. The paper characterizes all probability distributions that can be induced for the "channel image," which is given by the inner product of the input vector with a given channel vector. Key to this result is the description of input vectors that achieve a given deterministic channel image with the smallest energy, where "energy" of an input vector refers to a weighted sum of its one- and two-norms. Minimum-energy input vectors have an interesting structure: depending on the desired channel image, some of the weakest antennas are silenced, and the remaining antennas are chosen according to a shifted and amplitude-constrained beamforming rule.**

I. Introduction and Problem Setup

Consider a multiple-input single-output (MISO) antenna system of the form:

$$\bar{X} = \mathbf{h}^\mathsf{T}\mathbf{X}, \tag{1}$$

where $\mathbf{h} = (h_1, \ldots, h_{n_\mathrm{T}})^\mathsf{T}$ is a constant channel state vector and $\mathbf{X}$ an n_T-dimensional channel input vector $\mathbf{X} = (X_1, \ldots, X_{n_\mathrm{T}})^\mathsf{T}$.

The channel image $\bar{X}$ in (1) is useful in describing the input-output relation of MISO (optical or radio-frequency) wireless channels, where the channel output is modeled as $Y = \bar{X} + Z$, with Z being additive noise [1]–[4]. In radio-frequency communication, the inputs $X_1, \ldots, X_{n_\mathrm{T}}$ correspond to the modulated electromagnetic field, and as such both the inputs and the channel coefficients can take values over the entire real line $\mathbb{R}$. In this context, battery and power limitations impose a second-moment constraint $\mathsf{E}\left[\|\mathbf{X}\|_2^2\right] \leq \mathsf{P}$, and modulation schemes can be restricted without loss of optimality to beamforming input-vectors [4], [5]

$$\mathbf{x}_{\text{beamformer}}(\bar{x}) = \frac{\bar{x}}{\|\mathbf{h}\|_2^2} \cdot \mathbf{h}, \quad \bar{x} \in \mathbb{R}, \tag{2}$$

since $\mathbf{x}_{\text{beamformer}}(\bar{x})$ has the smallest two-norm $\|\mathbf{x}\|_2^2$ among all input vectors $\mathbf{x}$ inducing the same channel image $\bar{x} = \mathbf{h}^\mathsf{T}\mathbf{x}$. In particular, because $\bar{X}$ forms a sufficient statistic for $\mathbf{X}$ with respect to the output $Y = \bar{X} + Z$, the capacity-achieving input distribution for such a channel can be restricted to taking value only in the set of beamforming vectors like (2).

In intensity-modulated direct-detection (IM/DD) optical communication systems, the inputs are directly proportional to the intensity of the emitted light. They therefore cannot be negative: $X_1, \ldots, X_{n_\mathrm{T}} \geq 0$. The channel coefficients are also nonnegative in this case. To avoid degeneracy, we henceforth assume that no two coefficients are equal, and, without loss of generality, that they are ordered as

$$h_1 > h_2 > \cdots > h_{n_\mathrm{T}} > 0. \tag{3}$$

Furthermore, optical-power limitations are captured by the first-moment constraints $\mathsf{E}[\|\mathbf{X}\|_1] \leq \mathsf{E}$. The input vector with the smallest one-norm inducing a target channel image $\bar{x}$ is

$$\mathbf{x}_{\text{strongest}}(\bar{x}) = \left(\tfrac{\bar{x}}{h_1}, 0, \ldots, 0\right)^\mathsf{T}, \quad \bar{x} \in [0, \infty). \tag{4}$$

Therefore, modulation systems for first-moment constrained MISO additive noise channels as well as the capacity-achieving input distributions of these channels can be restricted in such a way that only the first input antenna is used.

Safety considerations and technical limitations often impose strict peak constraints $\mathsf{A} > 0$ on the inputs:

$$X_1, \ldots, X_{n_\mathrm{T}} \in [0, \mathsf{A}], \tag{5}$$

in which case the random channel image $\bar{X}$ takes value in the interval $\bar{\mathcal{X}} \triangleq [0, h_{\text{sum}}\mathsf{A}]$, for $h_{\text{sum}} \triangleq \sum_{k=1}^{n_\mathrm{T}} h_k$. Note that for the remainder of this paper, we will always assume that (5) must be satisfied. Under such peak constraints, the input vector $\mathbf{x}$ that induces a target channel image $\bar{x}$ with the smallest one-norm is [6], [7]

$$\mathbf{x}_{\text{peak}}(\bar{x}) = \left(\mathsf{A}, \ldots, \mathsf{A}, \tfrac{\bar{x} - \sum_{k=1}^{i-1}\mathsf{A}h_k}{h_i}, 0, \ldots, 0\right)^\mathsf{T}, \tag{6}$$

where the single input that is neither 0 nor A is at position $i \in \{1, \ldots, n_\mathrm{T}\}$ if $\bar{x} \in \left(\mathsf{A}\sum_{k=1}^{i-1} h_k, \mathsf{A}\sum_{k=1}^{i} h_k\right]$.

In this paper, we consider both first- and second-moment input constraints:

$$\mathsf{E}[\|\mathbf{X}\|_1] \leq \alpha_1\mathsf{A}, \tag{7a}$$

$$\mathsf{E}\left[\|\mathbf{X}\|_2^2\right] \leq \alpha_2\mathsf{A}^2, \tag{7b}$$

as encountered, e.g., in IM/DD systems with limitations on the optical power and the power consumed by the electronic circuit [8]–[12]. We characterize the set of all random channel images $\overline{X}$ that can be induced under these constraints. This characterization allows us to reduce the capacity calculation of an additive-noise MISO channel with input vector $\mathbf{X}$ under constraints (5) and (7) to a simpler optimization problem over the distribution of $\overline{X}$, a technique that was also applied in [6], [7] to channels without a second-moment constraint. A main step in our proof is to determine the "minimum-energy" input vectors $\mathbf{x}_{\min,\lambda}(\bar{x})$ that, among all input vectors inducing channel image $\bar{x}$, minimize the weighted sum-norm $\lambda\|\mathbf{x}\|_1 + (1-\lambda)\|\mathbf{x}\|_2^2$ for some $\lambda \in [0,1]$. For $\lambda = 1$, $\mathbf{x}_{\min,1}(\bar{x}) = \mathbf{x}_{\text{peak}}(\bar{x})$, as determined in [6]. A key result of the present paper is the minimum-energy solution for $\lambda \in [0,1)$: it sets a number of strongest antennas to the maximum value A, switches off a number of the weakest antennas, and applies "shifted beamforming" on the remaining antennas. The exact solution is given in Lemma 9 ahead.

The above minimum-energy solution can be used to simplify capacity calculation for additive-noise MISO channels with nonnegative inputs that satisfy peak, first-moment, and second-moment constraints. Capacity of such channels with peak or first-moment constraint, or both (but without a second-moment constraint), has been studied in many recent works [6], [7], [11], [13]–[24].

The remainder of this paper is arranged as follows. In Section II, we show that the problem of characterizing all achievable $\overline{X}$ can be simplified to a class of optimization problems for deterministic channel images $\bar{x} \in \overline{\mathcal{X}}$ and some parameter $\lambda \in [0,1]$. In Section III, we then present the solution to these optimization problems. Section IV combines these findings to present an explicit characterization of the set of channel images $\overline{X}$ that are achievable under constraints (5) and (7).

II. Pareto-Optimal Inputs

Since we have two power constraints, it is convenient to think of the first and second moments of a random vector as two utility functions, and introduce the notion of Pareto optimality.

Definition 1: A random vector $\mathbf{X}$ is said to be *Pareto optimal* if there exists no other random vector $\mathbf{X}'$ such that $\mathbf{h}^\mathsf{T}\mathbf{X}'$ has the same distribution as $\mathbf{h}^\mathsf{T}\mathbf{X}$, while

$$\mathsf{E}[\|\mathbf{X}'\|_1] \le \mathsf{E}[\|\mathbf{X}\|_1] \tag{8a}$$

$$\mathsf{E}\big[\|\mathbf{X}'\|_2^2\big] \le \mathsf{E}\big[\|\mathbf{X}\|_2^2\big] \tag{8b}$$

with at least one of the two inequalities in (8) being strict.

Lemma 2 below says that we can restrict our attention to Pareto-optimal input distributions. Lemma 3 characterizes these Pareto-optimal distributions as solutions to a class of optimization problems. The proofs of these two lemmas are elementary and therefore omitted.

Lemma 2: Consider a random channel image $\overline{X}$ with a desired distribution. If there exists an input vectors $\mathbf{X}$ satisfying (7) and inducing $\overline{X}$, then there must exist a Pareto-optimal input vector $\mathbf{X}^*$ satisfying (7) and inducing $\overline{X}$.

Lemma 3: If a random vector $\mathbf{X}^*$ is Pareto optimal, then there exists a $\lambda \in [0,1]$ such that $\mathbf{X}^*$ minimizes

$$\lambda \cdot \frac{\mathsf{E}[\|\mathbf{X}\|_1]}{\mathsf{A}} + (1-\lambda)\cdot\frac{\mathsf{E}\big[\|\mathbf{X}\|_2^2\big]}{\mathsf{A}^2} \tag{9}$$

among all $\mathbf{X}$ for which $\mathbf{h}^\mathsf{T}\mathbf{X}$ has the same distribution as $\mathbf{h}^\mathsf{T}\mathbf{X}^*$.

Lemma 4: Fix any $\lambda \in [0,1]$. For any $\bar{x} \in \overline{\mathcal{X}}$, denote

$$g(\bar{x},\lambda) \triangleq \min_{\mathbf{x}\colon \mathbf{h}^\mathsf{T}\mathbf{x}=\bar{x}} \left\{\lambda\cdot\frac{\|\mathbf{x}\|_1}{\mathsf{A}} + (1-\lambda)\cdot\frac{\|\mathbf{x}\|_2^2}{\mathsf{A}^2}\right\}. \tag{10}$$

For any target distribution for $\overline{X}$,

$$\begin{aligned}\min_{\mathbf{X}\colon \mathbf{h}^\mathsf{T}\mathbf{X}\sim\overline{X}} &\left\{\lambda\cdot\frac{\mathsf{E}[\|\mathbf{X}\|_1]}{\mathsf{A}} + (1-\lambda)\cdot\frac{\mathsf{E}\big[\|\mathbf{X}\|_2^2\big]}{\mathsf{A}^2}\right\} \\ &= \mathsf{E}\big[g(\overline{X},\lambda)\big]. \end{aligned} \tag{11}$$

Proof: For any $\mathbf{X}$ such that $\mathbf{h}^\mathsf{T}\mathbf{X} \sim \overline{X}$, we have

$$\begin{aligned}\lambda &\cdot \frac{\mathsf{E}[\|\mathbf{X}\|_1]}{\mathsf{A}} + (1-\lambda)\cdot\frac{\mathsf{E}\big[\|\mathbf{X}\|_2^2\big]}{\mathsf{A}^2} \\ &= \mathsf{E}\left[\lambda\cdot\frac{\|\mathbf{X}\|_1}{\mathsf{A}} + (1-\lambda)\cdot\frac{\|\mathbf{X}\|_2^2}{\mathsf{A}^2}\right] &(12)\\ &= \mathsf{E}\left[\mathsf{E}\left[\lambda\cdot\frac{\|\mathbf{X}\|_1}{\mathsf{A}} + (1-\lambda)\cdot\frac{\|\mathbf{X}\|_2^2}{\mathsf{A}^2}\,\middle|\,\mathbf{h}^\mathsf{T}\mathbf{X}=\overline{X}\right]\right] &(13)\\ &\ge \mathsf{E}\big[g(\overline{X},\lambda)\big]. &(14)\end{aligned}$$

Equality in the above is achieved by choosing, for every $\bar{x} \in \overline{\mathcal{X}}$, a corresponding input vector $\mathbf{x}$ that achieves $g(\bar{x},\lambda)$. ∎

Recall that our goal is to find out whether a desired distribution for $\overline{X}$ is achievable under the constraints (7) or not. Using Lemmas 2, 3, and 4, we see that this task can be simplified as follows. We first determine, for every $\lambda \in [0,1]$ and $\bar{x} \in \overline{\mathcal{X}}$, the input vectors that achieve the minimum in (10). (This can be done without knowing the target distribution for $\overline{X}$.) We refer to such input vectors as *minimum-energy signaling* (with respect to parameter λ). If a target distribution for $\overline{X}$ is achievable under (7), then there must exist some $\lambda \in [0,1]$ such that an $\mathbf{X}$ taking values only on minimum-energy signaling input vectors with respect to parameter λ induces $\overline{X}$ and satisfies (7). In fact, as we shall later see, such $\mathbf{X}$ is uniquely determined by $\overline{X}$ and λ. Hence, instead of considering distributions over $[0,\mathsf{A}]^{n_\mathrm{T}}$, we only need to examine these specific distributions for each λ.

III. Characterization of Minimum-Energy Signaling

In this section we characterize, for every $\lambda \in [0,1)$ and $\bar{x} \in \overline{\mathcal{X}}$, the solution to the following minimization problem:

$$\min_{\mathbf{x}\in[0,\mathsf{A}]^{n_\mathrm{T}}\colon \mathbf{h}^\mathsf{T}\mathbf{x}=\bar{x}} \left\{\lambda\cdot\frac{\|\mathbf{x}\|_1}{\mathsf{A}} + (1-\lambda)\cdot\frac{\|\mathbf{x}\|_2^2}{\mathsf{A}^2}\right\}. \tag{15}$$

(Recall that, for $\lambda = 1$, the optimization problem in (15) was solved in [6].)

Fix a $\lambda \in [0,1)$ and denote

$$\nu \triangleq \frac{\lambda}{1-\lambda}. \tag{16}$$

We introduce some indices and threshold values.

Definition 5: Set $\kappa_0 \triangleq 0$. For each index $\ell \in \{1,\dots,n_{\mathrm{T}}\}$, define the integer

$$\kappa_\ell \triangleq \max\biggl\{j \in \{\ell,\dots,n_{\mathrm{T}}\} \colon \frac{\nu}{2}\biggl(\frac{h_\ell}{h_j}-1\biggr) < 1\biggr\}; \tag{17}$$

define the point

$$t_\ell \triangleq \mathsf{A}\sum_{i=1}^{\ell} h_i + \mathsf{A}\sum_{i=\ell+1}^{\kappa_\ell}\biggl(\frac{h_i^2}{h_\ell}+\frac{\nu}{2}\biggl(\frac{h_i^2}{h_\ell}-h_i\biggr)\biggr); \tag{18}$$

and finally, for every $k \in \{\kappa_{\ell-1}+1,\dots,\kappa_\ell\}$, define

$$s_k \triangleq \mathsf{A}\sum_{i=1}^{\ell-1} h_i + \mathsf{A}\sum_{i=\ell}^{k-1}\frac{\nu}{2}\biggl(\frac{h_i^2}{h_k}-h_i\biggr). \tag{19}$$

Remark 6: Notice that κ_k, t_k, and s_k are all nondecreasing in $k \in \{1,\dots,n_{\mathrm{T}}\}$. Moreover, if $\kappa_\ell < \kappa_{\ell+1}$, then

$$s_{\kappa_\ell} \le t_\ell \le s_{\kappa_\ell+1}, \tag{20}$$

because

$$\begin{aligned}\biggl(\frac{h_i^2}{h_\ell}+\frac{\nu}{2}\biggl(\frac{h_i^2}{h_\ell}-h_i\biggr)\biggr)-\frac{\nu}{2}\biggl(\frac{h_i^2}{h_k}-h_i\biggr)\\ = \frac{h_i^2}{h_\ell}\underbrace{\biggl(1-\frac{\nu}{2}\biggl(\frac{h_\ell}{h_k}-1\biggr)\biggr)}_{\triangleq\Gamma},\end{aligned} \tag{21}$$

where Γ is positive for any $k \le \kappa_\ell$ and $i \in \{1,\dots,k\}$ and is negative for $k = \kappa_\ell + 1$ and $i \in \{1,\dots,k\}$. Thus we have an ordering of the form:

$$\begin{aligned}0 = s_1 \le \cdots \le s_{\kappa_1} \le t_1\\ \le s_{\kappa_1+1} \le \cdots \le s_{\kappa_2} \le t_2\\ \le s_{\kappa_2+1} \le \cdots \le s_{\kappa_3} \le t_3\\ \le \cdots \le s_{n_{\mathrm{T}}} \le \cdots \le t_{n_{\mathrm{T}}} = h_{\mathrm{sum}}\mathsf{A}.\end{aligned} \tag{22}$$

As will be shown later, for each $k \in \{1,\dots,n_{\mathrm{T}}\}$, s_k is the threshold on $\bar{x}$ at which the kth antenna should be switched on (i.e., the optimal choice should have $x_k = 0$ for $\bar{x} \le s_k$ and $x_k > 0$ for $\bar{x} > s_k$); and t_k is the threshold on $\bar{x}$ at which the kth antenna should be set to the maximum value A (i.e., the optimal choice should have $x_k = \mathsf{A}$ for $\bar{x} \ge t_k$). The integer κ_k indicates the number of antennas that should be switched on before the kth antenna is fixed to its maximum value A.

Definition 7: For any $\ell \in \{1,\dots,n_{\mathrm{T}}\}$ for which $\kappa_\ell > \kappa_{\ell-1}$, and for any $k \in \{\kappa_{\ell-1},\dots,\kappa_\ell\}$, define the subintervals

$$\mathcal{I}_{\ell,k} \triangleq \begin{cases}[t_{\ell-1}, s_{k+1}] & \text{if } k = \kappa_{\ell-1},\\ [s_k, s_{k+1}] & \text{if } \kappa_{\ell-1} < k < \kappa_\ell,\\ [s_k, t_\ell] & \text{if } k = \kappa_\ell.\end{cases} \tag{23}$$

For any $\ell \in \{2,\dots,n_{\mathrm{T}}\}$ for which $\kappa_\ell = \kappa_{\ell-1}$, define the subinterval

$$\mathcal{I}_{\ell,\emptyset} \triangleq [t_{\ell-1}, t_\ell]. \tag{24}$$

The next lemma follows immediately from Remark 6.

Lemma 8: The set of intervals

$$\begin{aligned}\mathcal{P} \triangleq \{\mathcal{I}_{\ell,\kappa_{\ell-1}+1},\dots,\mathcal{I}_{\ell,\kappa_\ell}\}_{\{\ell\colon \kappa_{\ell-1}<\kappa_\ell\}}\\ \cup \{\mathcal{I}_{\ell,\emptyset}\}_{\{\ell\colon \kappa_{\ell-1}=\kappa_\ell\}}\end{aligned} \tag{25}$$

overlap on a set of Lebesgue measure zero, and their union is $\overline{\mathcal{X}}$.

We can now present the solution to (15).

Lemma 9: Fix $\lambda \in [0,1)$. For any $\bar{x} \in \overline{\mathcal{X}}$, the optimization problem (15) has the following unique[1] solution $\mathbf{x}_{\min,\lambda}(\bar{x}) = (x_{\min,\lambda,1},\dots,x_{\min,\lambda,n_{\mathrm{T}}})^{\mathsf{T}}$. If $\bar{x}$ lies in an interval $\mathcal{I}_{\ell,k}$ as defined in (23), then the solution is given by

$$x_{\min,\lambda,i}(\bar{x}) = \mathsf{A}, \qquad i \le \ell-1, \tag{26a}$$

$$\begin{aligned}x_{\min,\lambda,i}(\bar{x}) = \Biggl(\bar{x}-\mathsf{A}\sum_{j=1}^{\ell-1}h_j\Biggr)\cdot\frac{h_i}{\sum_{j=\ell}^{k}h_j^2}\\ + \frac{\nu\mathsf{A}}{2}\left(\frac{\sum_{j=\ell}^{k}h_j}{\sum_{j=\ell}^{k}\frac{h_j^2}{h_i}}-1\right), \qquad i = \ell,\dots,k,\end{aligned} \tag{26b}$$

$$x_{\min,\lambda,i}(\bar{x}) = 0, \qquad i \ge k+1, \tag{26c}$$

where ν is defined in (16). Furthermore, the one- and two-norms of $\mathbf{x}_{\min,\lambda}(\bar{x})$ are given respectively by

$$\begin{aligned}&\|\mathbf{x}_{\min,\lambda}(\bar{x})\|_1 = m(\lambda,\bar{x})\\ &\triangleq (\ell-1)\mathsf{A}-(k-\ell+1)\nu\frac{\mathsf{A}}{2}\\ &\quad+\left(\frac{\bar{x}-\mathsf{A}\sum_{j=1}^{\ell-1}h_j+\nu\frac{\mathsf{A}}{2}\cdot\sum_{j=\ell}^{k}h_j}{\sum_{j=\ell}^{k}h_j^2}\right)\cdot\Biggl(\sum_{i=\ell}^{k}h_i\Biggr),\end{aligned} \tag{27}$$

and

$$\begin{aligned}&\|\mathbf{x}_{\min,\lambda}(\bar{x})\|_2^2 = v(\lambda,\bar{x})\\ &\triangleq (\ell-1)\mathsf{A}^2+(k-\ell+1)\nu^2\frac{\mathsf{A}^2}{4}\\ &\quad+\frac{\Bigl(\bar{x}-\mathsf{A}\sum_{j=1}^{\ell-1}h_j+\nu\frac{\mathsf{A}}{2}\cdot\sum_{j=\ell}^{k}h_j\Bigr)^2}{\sum_{j=\ell}^{k}h_j^2}\\ &\quad-\left(\frac{\bar{x}-\mathsf{A}\sum_{j=1}^{\ell-1}h_j+\nu\frac{\mathsf{A}}{2}\cdot\sum_{j=\ell}^{k}h_j}{\sum_{j=\ell}^{k}h_j^2}\right)\cdot\Biggl(\sum_{i=\ell}^{k}h_i\Biggr)\cdot\nu\mathsf{A}.\end{aligned} \tag{28}$$

If $\bar{x}$ lies in an interval $\mathcal{I}_{\ell,\emptyset}$ as defined in (24), then $\mathbf{x}_{\min,\lambda}(\bar{x})$ is given as in (26) with k replaced by κ_ℓ; its one- and two-norms are given by $m(\lambda,\bar{x})$ and $v(\lambda,\bar{x})$ as in (27) and (28), again with the parameter k replaced by κ_ℓ.

Proof: See the appendix. ■

Remark 10: Lemma 9 shows that $\mathbf{x}_{\min,\lambda}(\bar{x})$ sets the strongest $\ell-1$ antennas at the maximum value A and switches off the weakest $n_{\mathrm{T}}-k$ antennas; the remaining $k-\ell+1$ antennas are used in a shifted beamforming fashion, in the sense that the first term on the right-hand side of (26b) is

[1] Uniqueness relies on the strictness of the inequalities in (3).

the same as in standard beamforming subject to a second-moment constraint [4], while the second term there is a shifting constant that depends on the antenna index i.

For completeness, we also define $m(1,\cdot)$ and $v(1,\cdot)$ following (6).

Definition 11: For $i \in \{1,\dots,n_{\mathrm{T}}\}$ and for

$$\bar{x} \in \left[\mathsf{A}\sum_{k=1}^{i-1} h_k, \mathsf{A}\sum_{k=1}^{i} h_k\right], \tag{29}$$

we define

$$m(1,\bar{x}) \triangleq \sum_{k=1}^{i-1} \mathsf{A} + \frac{\bar{x} - \mathsf{A}\sum_{k=1}^{i-1} h_k}{h_i}, \tag{30}$$

$$v(1,\bar{x}) \triangleq \sum_{k=1}^{i-1} \mathsf{A}^2 + \left(\frac{\bar{x} - \mathsf{A}\sum_{k=1}^{i-1} h_k}{h_i}\right)^2. \tag{31}$$

IV. Characterization of All Achievable $\bar{X}$

Theorem 12: The target random variable $\bar{X}$ can be generated with a random input vector $\mathbf{X}$ satisfying (7) if, and only if, there exists a value $\lambda \in [0,1]$ such that the following two inequalities are satisfied:

$$\mathsf{E}_{\bar{X}}\big[m(\lambda,\bar{X})\big] \leq \alpha_1 \mathsf{A}, \tag{32a}$$

$$\mathsf{E}_{\bar{X}}\big[v(\lambda,\bar{X})\big] \leq \alpha_2 \mathsf{A}^2, \tag{32b}$$

where the functions $m(\cdot,\cdot)$ and $v(\cdot,\cdot)$ are defined in Lemma 9 and Definition 11.

Proof: The theorem follows immediately from Lemmas 2, 3, 4, and 9. ■

Example 13: Consider a MISO channel with $n_{\mathrm{T}} = 4$ input antennas and channel vector

$$\mathbf{h} = (4,3,2,1)^{\mathsf{T}}. \tag{33}$$

Figure 1 illustrates the set of (α_1,α_2)-pairs that permit the uniform distribution for $\bar{X} \in \bar{\mathcal{X}}$. The uniform distribution is optimal in the high-signal-to-noise-ratio (high-SNR) limit among all $\bar{X}$ taking values on $\bar{\mathcal{X}}$ without other constraints; see, e.g., [6]. Thus, we have identified the set of (α_1,α_2)-pairs for which the constraints (7) do not limit the capacity in the high-SNR limit. This set depends on the channel vector $\mathbf{h}$ and has a fundamentally different shape from the corresponding set for the single-input single-output channels, which is characterized by $\alpha_1 \geq 1/2$ and $\alpha_2 \geq 1/3$ and is rectangular [11].

Acknowledgments

M. Wigger acknowledges funding support from the European Research Council (ERC) under the European Union's Horizon 2020 program, grant agreement number 715111.

Appendix

In this appendix we prove Lemma 9. The optimization problem (15) is convex by the convexity of the domain set $\{\mathbf{x} \in [0,\mathsf{A}]^{n_{\mathrm{T}}} : \mathbf{h}^{\mathsf{T}}\mathbf{x} = \bar{x}\}$ and the convexity of the function

$$\mathbf{x} \mapsto \lambda \cdot \frac{\|\mathbf{x}\|_1}{\mathsf{A}} + (1-\lambda)\cdot\frac{\|\mathbf{x}\|_2^2}{\mathsf{A}^2}. \tag{34}$$

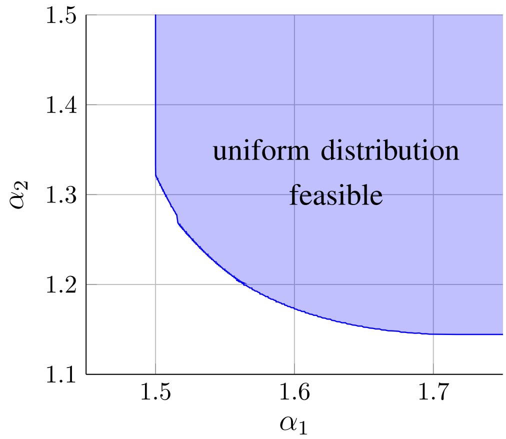

Fig. 1. The figure illustrates the set of all (α_1,α_2)-pairs that allow to induce a uniform distribution over $\bar{X}$ under a MISO channel vector $\mathbf{h} = (4,3,2,1)^{\mathsf{T}}$.

The solution can thus be found by inspecting the Karush–Kuhn–Tucker (KKT) conditions. The KKT conditions (after multiplication of the objective function by the constant $\frac{\mathsf{A}}{1-\lambda}$ and recalling $\nu = \frac{\lambda}{1-\lambda}$) are

$$\nu + \frac{2x_k}{\mathsf{A}} + \lambda h_k - v_k + \rho_k = 0, \quad k \in \{1,\dots,n_{\mathrm{T}}\}, \tag{35a}$$

$$\mathbf{h}^{\mathsf{T}}\mathbf{x} - \bar{x} = 0, \tag{35b}$$

$$-x_k \leq 0, \quad k \in \{1,\dots,n_{\mathrm{T}}\}, \tag{35c}$$

$$x_k - \mathsf{A} \leq 0, \quad k \in \{1,\dots,n_{\mathrm{T}}\}, \tag{35d}$$

$$v_k x_k = 0, \quad k \in \{1,\dots,n_{\mathrm{T}}\}, \tag{35e}$$

$$\rho_k(x_k - \mathsf{A}) = 0, \quad k \in \{1,\dots,n_{\mathrm{T}}\}, \tag{35f}$$

$$v_k, \rho_k \geq 0, \quad k \in \{1,\dots,n_{\mathrm{T}}\}. \tag{35g}$$

Notice first that only one of the two parameters v_k or ρ_k can be strictly positive, and when the solution $x_k^* \in (0,1)$, then both $v_k = \rho_k = 0$. Defining $\zeta_k \triangleq v_k - \rho_k$, we can summarize (35a) and (35e)–(35g) as:

$$\nu + \frac{2x_k}{\mathsf{A}} + \lambda h_k = \zeta_k, \qquad k \in \{1,\dots,n_{\mathrm{T}}\}, \tag{36}$$

where

1) $\zeta_k = 0$ when $x_k \in (0,\mathsf{A})$;
2) $\zeta_k \geq 0$ when $x_k = 0$; and
3) $\zeta_k \leq 0$ when $x_k = \mathsf{A}$.

We fix two indices $1 \leq \ell < k \leq n_{\mathrm{T}}$ and resolve (36) for λ. This yields:

$$x_\ell = \frac{h_\ell}{h_k}x_k + \frac{\mathsf{A}\nu}{2}\underbrace{\left(\frac{h_\ell}{h_k} - 1\right)}_{>0} + \left(\zeta_\ell - \frac{h_\ell}{h_k}\zeta_k\right)\frac{\mathsf{A}}{2},$$
$$1 \leq \ell < k \leq n_{\mathrm{T}}. \tag{37}$$

Now we study the implications of (37) and above statements 1)–3), where we distinguish the cases $x_k = 0$, $x_k = \mathsf{A}$, and $x_k \in (0,\mathsf{A})$.

Case 1: $x_k = 0$**:** In this case, $\zeta_k \geq 0$ and

$$x_\ell \leq \frac{\mathsf{A}\nu}{2}\left(\frac{h_\ell}{h_k} - 1\right) + \frac{\mathsf{A}\zeta_\ell}{2}. \tag{38}$$

Now, if $x_\ell = \mathsf{A}$ then $\zeta_\ell \le 0$, and thus $x_\ell \le \frac{\mathsf{A}\nu}{2}\left(\frac{h_\ell}{h_k} - 1\right)$. But since $x_\ell = \mathsf{A}$, this is only possible if

$$\frac{\nu}{2}\left(\frac{h_\ell}{h_k} - 1\right) \ge 1. \tag{39}$$

If $x_\ell \in (0, \mathsf{A})$, then $\zeta_\ell = 0$ and

$$x_\ell \le \frac{\mathsf{A}\nu}{2}\left(\frac{h_\ell}{h_k} - 1\right). \tag{40}$$

The case $x_\ell = 0$ is trivially possible and its analysis does not provide further insights.

Case 2: $x_k \in (0, \mathsf{A})$**:** In this case $\zeta_k = 0$ and

$$x_\ell = \frac{h_\ell}{h_k} x_k + \frac{\mathsf{A}\nu}{2}\left(\frac{h_\ell}{h_k} - 1\right) + \frac{\mathsf{A}\zeta_\ell}{2}. \tag{41}$$

Now, if $x_\ell = 0$, then $\zeta_\ell \ge 0$ and

$$0 \ge \frac{h_\ell}{h_k} x_k + \frac{\mathsf{A}\nu}{2}\underbrace{\left(\frac{h_\ell}{h_k} - 1\right)}_{>0}, \tag{42}$$

which is not possible because $\nu \le 0$ and $x_k \ge 0$.
If $x_\ell \in (0, \mathsf{A})$, then $\zeta_\ell = 0$ and

$$x_\ell = \frac{h_\ell}{h_k} x_k + \frac{\mathsf{A}\nu}{2}\left(\frac{h_\ell}{h_k} - 1\right). \tag{43}$$

If $x_\ell = \mathsf{A}$, then $\zeta_\ell \le 0$ and

$$\mathsf{A} \le \frac{h_\ell}{h_k} x_k + \frac{\mathsf{A}\nu}{2}\left(\frac{h_\ell}{h_k} - 1\right). \tag{44}$$

Case 3: $x_k = \mathsf{A}$**:** In this case $\zeta_k \le 0$ and

$$x_\ell \ge \frac{h_\ell}{h_k}\mathsf{A} + \frac{\mathsf{A}\nu}{2}\left(\frac{h_\ell}{h_k} - 1\right) + \frac{\mathsf{A}\zeta_\ell}{2}. \tag{45}$$

Now, if $x_\ell \in [0, \mathsf{A})$, then $\zeta_\ell \ge 0$ and

$$x_\ell \ge \frac{h_\ell}{h_k}\mathsf{A} + \frac{\mathsf{A}\nu}{2}\left(\frac{h_\ell}{h_k} - 1\right) > \mathsf{A}, \tag{46}$$

where the second inequality holds because $h_\ell > h_k$ and $\nu > 0$. Since this is not possible, we necessarily have $x_\ell = \mathsf{A}$ when $x_k = \mathsf{A}$.

We summarize our findings on the optimal solution $x_1^*, \ldots, x_{n_\mathrm{T}}^*$, where we include also the conditions $x_1^*, \ldots, x_{n_\mathrm{T}}^* \in [0, \mathsf{A}]$. For any pair of indices $1 \le \ell < k \le n_\mathrm{T}$:

- If $x_k^* = 0$, then $x_\ell^* \in \left[0, \mathsf{A}\cdot\min\left\{1, \frac{\nu}{2}\left(\frac{h_\ell}{h_k} - 1\right)\right\}\right]$.
- If $x_k^* \in \left(0, \mathsf{A}\frac{h_k}{h_\ell}\left(1 - \frac{\nu}{2}\left(\frac{h_\ell}{h_k} - 1\right)\right)\right]$, then $x_\ell^* = \frac{h_\ell}{h_k} x_k^* + \frac{\mathsf{A}\nu}{2}\left(\frac{h_\ell}{h_k} - 1\right)$.
- If $x_k^* \in \left[\mathsf{A}\frac{h_k}{h_\ell}\left(1 - \frac{\nu}{2}\left(\frac{h_\ell}{h_k} - 1\right)\right), \mathsf{A}\right]$, then $x_\ell^* = \mathsf{A}$.

Combining these findings for all antennas $1 \le \ell < k \le n_\mathrm{T}$, we can conclude that the antennas are turned on one after the other for increasing values of $\bar{x}$ starting from the strongest antenna 1 until the weakest antenna n_T; all active antennas follow a shifted beamforming solution; and antennas are kept at maximum power A for all subsequent $\bar{x}$-values once the beamforming solution achieves this value and this (for smaller values of $\bar{x}$) happens first for antennas $1, 2, \ldots$ and finally for the last antenna n_T. We next determine the thresholds on $\bar{x}$

$$s_1 < \cdots < s_{n_\mathrm{T}} \tag{47}$$

when antennas $1, \ldots, n_\mathrm{T}$ are switched on, and the thresholds on $\bar{x}$

$$t_1 < \cdots < t_{n_\mathrm{T}} \tag{48}$$

when they are fixed to the maximum value A. Trivially,

$$s_1 = 0 \qquad \text{and} \qquad t_{n_\mathrm{T}} = h_{\mathrm{sum}}\mathsf{A}. \tag{49}$$

Determining the other values is more involved. In particular, the order of the various points depends on the specific values of the antenna channel gains. From the orderings in (47) and (48), we can conclude that the interval $\overline{\mathcal{X}} = [0, h_{\mathrm{sum}}\mathsf{A}]$ of all possible $\bar{x}$-values is divided into subintervals which are characterized by two indices $\ell, k \in \{1, \ldots, n_\mathrm{T}\}$ with $\ell \le k$ in the sense that for any $\bar{x}$ in this subinterval all antennas $1, \ldots, \ell - 1$ are set to the maximum value A and all antennas $k+1, \ldots, n_\mathrm{T}$ are set to 0. The $\bar{x}$-values in this interval are induced by a minimum-energy vector $\mathbf{x}_{\min,\lambda}$ of the form

$$x_{\min,\lambda,1} = \cdots = x_{\min,\lambda,\ell-1} = \mathsf{A}, \tag{50a}$$

$$x_{\min,\lambda,i} = \frac{h_i}{h_k} x_{\min,\lambda,k} + \frac{\mathsf{A}\nu}{2}\left(\frac{h_i}{h_k} - 1\right), \quad i = \ell, \ldots, k, \tag{50b}$$

$$x_{\min,\lambda,k+1} = \ldots = x^*_{\min,\lambda,n_\mathrm{T}} = 0. \tag{50c}$$

By the definition of $\bar{x}$,

$$\bar{x} = \sum_{i=1}^{n_\mathrm{T}} h_i x_{\min,\lambda,i} \tag{51}$$

$$= \mathsf{A}\sum_{j=1}^{\ell-1} h_j + \sum_{i=\ell}^{k}\left(\frac{h_i^2}{h_k} x_{\min,\lambda,k} + \frac{\mathsf{A}\nu}{2}\left(\frac{h_i^2}{h_k} - h_i\right)\right) \tag{52}$$

and thus

$$x_{\min,\lambda,k} = \frac{\bar{x} - \mathsf{A}\sum_{j=1}^{\ell-1} h_j + \frac{\mathsf{A}\nu}{2}\cdot\sum_{j=\ell}^{k} h_j}{\sum_{j=\ell}^{k} \frac{h_j^2}{h_k}} - \frac{\mathsf{A}\nu}{2} \tag{53}$$

and for $i = \ell, \ldots, k$ the solution in (26b) follows. We conclude that the minimum-energy solution $\mathbf{x}_{\min,\lambda}(\bar{x})$ is unique for any $\lambda \in [0, 1)$.

We next determine the end-point of the interval. The interval ends when either $x_\ell^* = \mathsf{A}$, in which case the ℓ-th antenna has to be frozen to its maximum value A and the end point of the interval is $\bar{x} = t_\ell$, or when

$$x_i^* = \frac{\mathsf{A}\nu}{2}\left(\frac{h_i}{h_{k+1}} - 1\right), \qquad i = \ell, \ldots, k, \tag{54}$$

in which case this interval ends with $\bar{x} = s_{k+1}$ and the $(k+1)$-th antenna will be used in the next-following interval. (Notice that by (50b), all the conditions in (54), for $i = \ell, \ldots, k$, are equivalent.) Which of the two events happens first, $\bar{x} = t_\ell$ or $\bar{x} = s_{k+1}$, depends on whether A is smaller or larger than $\frac{\mathsf{A}\nu}{2}\left(\frac{h_i}{h_{k+1}} - 1\right)$, which leads to the definitions of indices κ_ℓ.

REFERENCES

[1] Z. Ghassemlooy, L. N. Alves, S. Zvánovec, and M.-A. Khalighi (Eds.). “Visible Light Communications: Theory and Applications (1st ed.)," CRC Press, 2016.

[2] S. Arnon, J. Barry, G.K. Karagiannidis, R. Schober and M. Uysal (Eds.), “Advanced Optical Wireless Communication Systems," Cambridge University Press, 2012.

[3] E. Biglieri, A. R. Calderbank, A. Constantinides, A.J. Goldsmith, A. Paulraj, H. V. Poor, “MIMO Wireless Communications," Cambridge University Press, 2007.

[4] İ. E. Telatar, “Capacity of multi-antenna Gaussian channels,” *Eur. Trans. Telecommun.*, vol. 10, no. 6, pp. 585–595, Nov.–Dec. 1999.

[5] P. Algoet and J. Cioffi, “The capacity of a channel with Gaussian noise and intersymbol interference," in *Proc. IEEE Int. Symp. Inf. Theory*, Budapest, Hungary, June 1991, p.1 6.

[6] S. M. Moser, L. Wang, and M. Wigger, “Capacity results on multiple-input single-output wireless optical channels,” *IEEE Trans. Inf. Theory*, vol. 64, no. 11, pp. 6954–6966, Nov. 2018.

[7] L. Li, S. M. Moser, L. Wang, and M. Wigger, “On the capacity of MIMO optical wireless channels,” *IEEE Trans. Inf. Theory*, vol. 66, no. 9, pp. 5660–5682, Sept. 2020.

[8] X. Ling, J. Wang, X. Liang, Z. Ding, and C. Zhao, “Offset and power optimization for DCO-OFDM in visible light communication systems,” *IEEE Trans. Signal Process.*, vol. 64, no. 2, pp. 349–363, Jan. 2016.

[9] X. Huang, J. Shi, J. Li, Y. Wang, and N. Chi, “A Gb/s VLC transmission using hardware preequalization circuit,” *IEEE Photon. Technol. Lett.*, vol. 27, no. 18, pp. 1915–1918, Sep. 2015.

[10] F. Che, L. Wu, B. Hussain, X. Li, and C. P. Yue, “A fully integrated IEEE 802.15.7 visible light communication transmitter with on-chip 8-W 85% efficiency boost LED driver,” *J. Lightw. Technol.*, vol. 34, no. 10, pp. 2419–2430, May, 2016.

[11] S. Ma and M. Wigger, “First- and second-moment constrained Gaussian channels,” in *Proc. IEEE Int. Symp. Inf. Theory*, Melbourne, Australia, Jul. 12–20, 2021, pp. 432–437.

[12] S. Ma, H. Li, Y. He, R. Yang, S. Lu, W. Cao, and S. Li, “Capacity bounds and interference management for interference channel in visible light communication networks,” *IEEE Trans. Wireless Commun.*, vol. 18, no. 1, pp. 182–193, Jan. 2019.

[13] J. G. Smith,“The information capacity of amplitude- and variance-constrained scalar Gaussian channels," *Inf. Contr.*, vol. 18, no. 3, pp. 203–219, Feb. 1971.

[14] T. H. Chan, S. Hranilovic, and F. R. Kschischang, “Capacity-achieving probability measure for conditionally Gaussian channels with bounded inputs,” *IEEE Trans. Inf. Theory*, vol. 51, no. 6, pp. 2073–2088, June 2005.

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

[16] A. A. Farid and S. Hranilovic, “Upper and lower bounds on the capacity of wireless optical intensity channels,” in *Proc. IEEE Int. Symp. Inf. Theory*, Nice, France, June 24–30, 2007.

[17] A. Dytso, M. Goldenbaum, S. Shamai (Shitz), and H. V. Poor, “Upper and lower bounds on the capacity of amplitude-constrained MIMO channels,” in Proc. *IEEE Global Commun. Conf. (GLOBECOM)*, Singapore, Dec. 2017, pp. 1–6.

[18] A. Chaaban, Z. Rezki, and M.-S. Alouini, “Capacity bounds and high-SNR capacity of MIMO intensity-modulation optical channels,” *IEEE Trans. Wireless Commun.*, vol. 17, no. 5, pp. 3003–3017, May 2018.

[19] A. Chaaban, Z. Rezki, and M.-S. Alouini, “Low-SNR asymptotic capacityof MIMO optical intensity channels with peak and average constraints,” *IEEE Trans. Commun.*, vol. 66, no. 10, pp. 4694–4705, Oct. 2018

[20] A. Dytso, M. Al, H. V. Poor, and S. Shamai (Shitz), “On the capacity of the peak power constrained vector Gaussian channel: An estimation theoretic perspective,” *IEEE Trans. Inf. Theory*, vol. 65, no. 6, pp. 3907–3921, 2019.

[21] R.-H. Chen, L. Li, J. Zhang, W. Zhang, and J. Zhou, “ On the capacity of MISO optical intensity channels with per-antenna intensity constraints." Available: https://arxiv.org/abs/2111.11637.

[22] R.-H. Chen, L. Li, J. Zhang, and L. Li, “Capacity results for MIMO optical wireless communication with per-antenna intensity constraints." Available: https://arxiv.org/abs/2112.02220.

[23] K. Keykhosravi, E. Agrell, M. Secondini, and M. Karlsson, “When to use optical amplification in moncoherent transmission: An information-theoretic approach," in *IEEE Transactions on Communications*, vol. 68, no. 4, pp. 2438–2445, April 2020.

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