On-Off Keying Communication Over Optical Channels With Crosstalk

Hongchao Zhou, *Member, IEEE*, Yuval Kochman, *Member, IEEE*, and Gregory W. Wornell, *Fellow, IEEE*

***Abstract*—We investigate the fundamental limits of communication over optical on-off-keying channels with crosstalk, where a light pulse may span over multiple time slots or spatial pixels, and the receiver is equipped with single-photon detectors. First, we analyze achievable rates of communication over such channels, and observe that increasing transmission power (expected number of photons emitted per slot or pixel) does not necessarily lead to higher rates. Under simple but reasonable models, the highest rates are often achieved in a low-photon regime, with an average of 3 to 7 photons received in each slot or pixel. We further characterize the tradeoff between information rate and photon efficiency (in terms of the expected number of bits transmitted per photon) in the presence of crosstalk. Finally, we develop guidelines for slot length and pixel size selection for different application scenarios. Our analysis reveals that optimum optical-communication systems do not minimize the level of crosstalk.**

***Index Terms*—Optical communication, on-off-keying modulation, crosstalk, capacity, photon efficiency.**

I. Introduction

OPTICAL communication has long played a critical role in enabling high-speed, long-distance, point-to-point information delivery. While there has been much emphasis on optical fiber, there is a growing demand for efficient optical wireless technology for ground, space-based, and underwater systems in a host of emerging applications. Such systems generally seek to exploit all available degrees of freedom, including time, wavelength, polarization, and space, through efficient multiplexing. In practice, crosstalk between degrees of freedom is a key source of interference limiting performance. For example, in systems employing time-division multiplexing, photons in an emitted light pulse can arrive in multiple slots (depending on the slot length), causing temporal crosstalk. Such temporal crosstalk arises from a variety of sources, including, e.g., electronics jitter [15], [22] and dispersion in the medium [4]. Similarly, due to, e.g., backscatter [17], in systems using wavelength-division multiplexing, photons from a channel can leak into another unless the channels are widely spaced in wavelength. Additionally, in multi-spatial-mode communication systems, crosstalk between adjacent spatial modes at the photon detector array can arise due to, e.g., lack of orthogonality of the modes, or the presence of turbulence [6], [18], [36].

In this paper, we analyze optical communication over channels in which there are both photon losses and crosstalk, the latter of which can be one-dimensional or two-dimensional depending on the deployed multiplexing technique. Keeping in mind implementation considerations, we restrict our attention to an architecture based on on-off keying (OOK) and single-photon detection (SPD). With OOK, coherent light is pulsed on or off in each slot (or pixel), where the number of emitted photons in the slot is Poisson distributed with mean λ, the photon transmission density, which is controlled through choosing the corresponding transmission power. With SPD, receivers can detect the presence or absence of incoming photons in a slot (or pixel), and can be realized via photomultiplier tubes, avalanche photodiodes, or superconducting nanowires. The simplicity of this architecture makes it particularly suitable for many optical communication systems, such as multi-mode communication systems where many transmitters and detectors need to be implemented, and image array detection in the weak light scenario. Furthermore, in the regime where the transmitted energy is not too high this architecture is known to be nearly optimal (without crosstalk).

Poisson-type channels with inter-symbol interference were recently studied in [1] and [12] for the applications of non-line of sight optical wireless communication and molecular communication. Different from the OOK/SPD channel model considered in the present paper, they assume that the number of photons or molecules arriving at time instance t can be accurately detected, hence resulting in very different channel behaviors. The capacities of OOK/SPD systems without crosstalk have been well studied [5], [19], [26]. However, results on performance in the presence of crosstalk are quite limited. For example, [15] quantifies the rates achievable by pulse-position modulation (PPM) with soft-decision decoding in the case of 1-D crosstalk. But such rates are generally far from the capacity of such channels. Even fewer studies have investigated the impact of crosstalk in the 2-D case. While crosstalk does not necessarily arise in optical communication systems operating in a regime relatively far from fundamental limits, crosstalk does emerge as a critical issue if one uses single-photon detectors and seeks to approach the communication limit.

Beginning with a simple but useful model whereby crosstalk occurs only between neighboring slots or pixels, we quantify the capacity of such channels, the impact of practical code constraints, and how to choose power levels that yield certain

Manuscript received May 28, 2014; revised November 10, 2014; accepted April 11, 2015. Date of publication May 15, 2015; date of current version August 17, 2015. This work was supported in part by AFOSR under Grant. No. FA9550-11-1-0183, by the DARPA InPho program under Contract No. HR0011-10-C-0159, and by ISF under Grant No. 956/12. This paper was presented in part at the 2014 IEEE International Symposium on Information Theory.

H. Zhou and G. W. Wornell are with the Research Laboratory of Electronics, Massachusetts Institute of Technology, Cambridge, MA 02139 USA (e-mail: hongchao@mit.edu; gww@mit.edu).

Y. Kochman is with School of Computer Science and Engineering, Hebrew University, Jerusalem 91904, Israel (e-mail: yuvalko@cs.huji.ac.il).

Digital Object Identifier 10.1109/JSAC.2015.2432521

photon densities. This crosstalk model is broadly applicable not only to system-inherent crosstalk (such as jitter crosstalk), but also to important turbulence-induced crosstalk and other dispersive weather-related phenomena. Many practical channels can be well approximated and analyzed based on this crosstalk model, even when some photons appear in non-neighboring slots or pixels. We also analyze the photon efficiency of such systems, which is important when energy is a critical resource. Among other results, we quantify the tradeoff between the information rate and photon efficiency in the presence of crosstalk in the energy-efficient regime.

Finally, we consider the selection of slot/pixel size in such systems. Given the physical crosstalk mechanisms, decreasing the slot/pixel size increases the number of degrees of freedom for communication, but also increases the interference between these degrees of freedom due to the greater inter-slot/pixel crosstalk. As an illustration of the associated analysis, we discuss how to select a good slot/pixel size for Gaussian pulses in both 1-D and 2-D systems whose performance we can characterize.

In this paper, we use the following notations:

λ	transmission density, the expected number of emitted photons in each signal slot;
p_c	crosstalk probability, for a photon appearing in neighboring slots;
$C(\lambda, p_c)$	channel capacity given by the maximum number of bits sent per slot;
$R(\lambda, p_c)$	information rate given by the number of bits transmitted per slot;
ϵ	the expected number of emitted photons per slot, with $\epsilon = p_x\lambda$;
$\tilde{C}(\epsilon)$	photon efficiency defined by the maximum number of bits sent per photon.

An outline of the paper is as follows. Section II defines the models and analysis framework of the paper. Sections III and IV develop bounds on rates achievable over optical OOK channels with 1-D or 2-D crosstalk, respectively, as a function of the transmission density. Section V analyzes the photon efficiency in the presence of crosstalk in the low-photon regime, and Section VI discusses the issue of slot length (and pixel size) selection in some practical scenarios of interest. Finally, Section VII contains some concluding remarks.

II. Channel and System Models, and Performance Measures

In this section, we define the basic channel and system models of interest, and the associated performance measures to be used in our analysis.

A. Channel Model

In (1-D) OOK, the encoder prepares a sequence of binary inputs $\mathbf{x} = x_1^m = (x_1, x_2, \ldots, x_m) \in \{0, 1\}^m$, where $x_i = 1$ represents that a coherent pulse of (average) λ photons is sent in slot i. Given a pulse sent in slot k, the probability that a photon from this pulse arrives in slot $k + i$ is given by the crosstalk coefficient p_i. The number of photons arriving in slot j is a Poisson random variable k_j with mean

$$\bar{k}_j = \lambda\eta \sum_{i=1}^{m} x_i p_{j-i},$$

where $\eta \leq 1$ denotes the transmissivity of the channel, i.e., the probability of a photon arriving at the receiver. Since η affects a simple scaling of λ, we can, without loss of generality, set $\eta = 1$ in our analysis. With SPD, the channel outputs $\mathbf{y} = y_1^m = (y_1, y_2, \ldots, y_m)$, where y_j is the output in the jth slot, are also binary. Specifically, $y_j = 1$ if $k_j > 0$; otherwise, $y_j = 0$.

The crosstalk coefficients are strongly influenced by the slot size. In particular, we have

$$p_i = \int_{(i-1/2)\ell}^{(i+1/2)\ell} f(z)\mathrm{d}z. \tag{1}$$

where ℓ denotes the slot length, and $f(z)$ denotes the probability density function of the position of a photon in a received pulse centered at $z = 0$.

In systems with suitably long slots, there is effectively only crosstalk between adjacent slots, in which case we have (assuming symmetry) $p_1 = p_{-1} = p_c/2$ and $p_0 = 1 - p_c$, where p_c is the (single) crosstalk parameter. Thus the output photon count is conditionally Poisson with mean

$$\bar{k}_i = \left[\frac{p_c}{2}(x_{i-1} + x_{i+1}) + (1 - p_c)x_i\right]\lambda, \tag{2}$$

where we let $x_0 = 0$ and $x_{m+1} = 0$.

Extensions to the 2-D case of an $m \times m$ pixel array are straightforward. We use $\mathbf{X} = \{x_{i,j}\} \in \{0, 1\}^{m\times m}$ to denote the input array, where $x_{i,j}$ is the bit at the ith row and jth column, and $\mathbf{Y} = \{y_{i,j}\} \in \{0, 1\}^{m\times m}$ to denote the output array. A correspondingly simple but useful adjacent crosstalk model in this case is one with four-neighbor interference. Specifically, the number of photons received at pixel (i, j) is a Poisson random variable $k_{i,j}$ with mean

$$\bar{k}_{i,j} = \left[\frac{p_c}{4}(x_{i-1,j} + x_{i+1,j} + x_{i,j-1} + x_{i,j+1}) + (1 - p_c)x_{i,j}\right]\lambda, \tag{3}$$

where $x_{0,j} = 0$, $x_{m+1,j} = 0$, $x_{i,0} = 0$, $x_{i,m+1} = 0$ for all $1 \leq i, j \leq m$.

B. Channel Capacity and Information Rates

For a given photon transmission density λ, the maximum number of bits that can be transmitted per slot in the 1-D case is given by the OOK/SPD capacity

$$C(\lambda) = \lim_{m\to\infty} \max_{P_{\mathbf{x}}} \frac{1}{m} I(\mathbf{x}; \mathbf{y}), \tag{4}$$

where $\mathbf{x}, \mathbf{y} \in \{0, 1\}^m$ are the input sequence and output sequence, $P_{\mathbf{x}}$ is a distribution over $\mathbf{x}$, and $I(\mathbf{x}; \mathbf{y})$ denotes the mutual information. Analogously, in the 2-D case we have

$$C(\lambda) = \lim_{m\to\infty} \max_{P_{\mathbf{X}}} \frac{1}{m^2} I(\mathbf{X}; \mathbf{Y}).$$

Furthermore, the maximum capacity of a channel is $C^{\max} = \max_{\lambda \geq 0} C(\lambda)$. We will often make the dependence of capacity on the crosstalk explicit in our notation, using, e.g., $C(\lambda, p_c)$ for the capacity in the case of our nearest-neighbor crosstalk model.

In practice, when the optimizing input distribution in (4) is biased and/or correlated, it can be difficult to construct capacity-approaching error-correcting codes for these channels. In such cases, we are also interested in the performance of existing codes optimized for simpler input distributions—specifically, independent, identically distributed (i.i.d.) distributions, uniform i.i.d. distributions, and those corresponding to the use of PPM. Accordingly, in addition to capacity we analyze the associated *i.i.d. information rate* $R_{\text{iid}}(\lambda, p_c)$, *symmetric information rate* $R_{\text{sym}}(\lambda, p_c)$, and *PPM information rate* $R_{\text{PPM}}(\lambda, p_c)$, respectively.

C. Photon Efficiency

While capacity measures the limits of bandwidth efficiency, the limits of energy efficiency are given by the photon efficiency of the channel, which in the 1-D case is the maximum number of bits per photon that can be sent (or received, since without loss of generality we have taken $\eta = 1$), i.e.,

$$\tilde{C}(\epsilon) \triangleq \max_{P_{\mathbf{x}}:\, p_x \lambda \leq \epsilon} \frac{I(\mathbf{x}; \mathbf{y})}{m\epsilon},$$

$$\text{with} \quad p_x \triangleq \frac{1}{m} \sum_{i=1}^{m} \Pr\{x_i = 1\},$$

and corresponds to normalizing the rate per slot by the number of photons per slot. The photon efficiency varies with ϵ, the maximum number of photons available, on average, for transmission in a slot. The extension to the 2-D case is straightforward.

III. Channels With 1-D Crosstalk

In this section, we derive bounds on the capacities and information rates of 1-D channels described by (2), for which

$$P_{\mathbf{y}|\mathbf{x}}(y_1^m | x_1^m) = \prod_{i=1}^{m} P\left(y_i | x_{i-1}^{i+1}\right) \tag{5}$$

with $x_0 = x_{m+1} = 0$. Eq. (5) describes a class of inter-symbol interference (ISI) channels, or more generally, channels with memory. The computation of capacities of channels with memory has attracted much attention in the information theory literature. However, for most such channels with memory, computing the exact capacity is difficult, and thus much of the emphasis has been put on developing bounds. It is this bounding approach we apply to the channel (5).

A general approach to obtaining a lower bound on capacity is to calculate the information rate of the channel when its input is an kth-order Markov chain. By optimizing the distribution of the input Markov chain and increasing k, one can get an increasingly tight lower bound on the channel capacity [8], [16], [34]. For a given Markov chain input, an efficient simulation-based method for computing the information rates of channels with memory is developed in [2], [3], [27]. This method requires the generation of very long output realizations from the channel model. Approaches for upper bounding the capacity of channels with memory are developed in [33], [37].

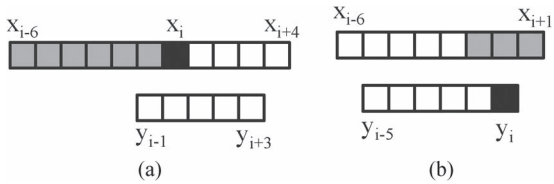


Fig. 1. Bits for computing the bounds on the capacity of channels with 1-D crosstalk. (a) The lower bound is computed by optimizing over kth-order Markov input distributions on x_{i-k}^{i+d} with $k = 6$ and parameter $d = 4$. (b) The upper bound is computed by optimizing over stationary and reflection-symmetric input distributions on x_{i-t-1}^{i+1} with $t = 5$.

A. Lower Bound on Capacity

Following [8], [16], [34], we compute the lower bound of the capacity (5) by optimizing a kth order stationary Markov chain input. Specifically, the information rate $R(\Psi)$ of the channel as a function of the transition probability $\Psi \in \mathbb{R}^k$ of the Markov chain input is a lower bound.

Proceeding, via the chain rule of mutual information we have, for any input distribution,

$$I(\mathbf{x}; \mathbf{y}) = \sum_{i=1}^{m} \left[H\left(x_i | x_1^{i-1}\right) - H\left(x_i | y_1^m, x_1^{i-1}\right) \right], \tag{6}$$

and for the subset of the inputs as in Fig. 1(a), we have, due to the Markov structure of the input,

$$\begin{aligned} H(x_i | x_1^{i-1}) &= H\left(x_i | x_{i-k}^{i-1}\right), \\ H\left(x_i | y_1^m, x_1^{i-1}\right) &\leq H\left(x_i | x_1^{i-1}, y_{i-1}^m\right) \leq H\left(x_i | x_{i-k}^{i-1}, y_{i-1}^{i-1+d}\right), \end{aligned} \tag{7}$$

where k is the order of the input Markov chain, and d is a suitably small integer. Moreover, when m is sufficiently large, we can further ignore the edge effects. Thus, we obtain the bound

$$\begin{aligned} C(\lambda, p_c) &\geq C^{-}(\lambda, p_c) \\ &\triangleq \max_{\mathbf{0} \leq \Psi \leq \mathbf{1}} \left[H\left(x_i | x_{i-k}^{i-1}\right) - H\left(x_i | x_{i-k}^{i-1}, y_{i-1}^{i-1+d}\right) \right] \end{aligned} \tag{8}$$

for $k \leq i \leq m - d + 1$, where we note that the bound does not depend on the chosen value of i in this range.

To evaluate the quantity in brackets in (8), we require the probability distribution on x_{i-k}^{i+d}, i.e., $P(x_{i-k}^{i+d})$, which is obtained as follows. Given the transition probability Ψ, we compute the stationary distribution of the input as follows. We treat k consecutive input bits as a super-symbol and express the input sequence as a first-order Markov process over an alphabet of size 2^k with transition matrix $\mathbf{A}(\Psi)$. The stationary distribution $\mathbf{u}(\Psi)$ is the solution to $\mathbf{u}(\Psi)\mathbf{A}(\Psi) = \mathbf{u}(\Psi)$, from which $P(x_{i-k}^{i+d})$ is computed. Although the bracketed expression is in general not concave, convex-programming techniques can nevertheless be applied to obtain a local maximum as the lower bound. Our numerical experiments suggest that the local maximum may

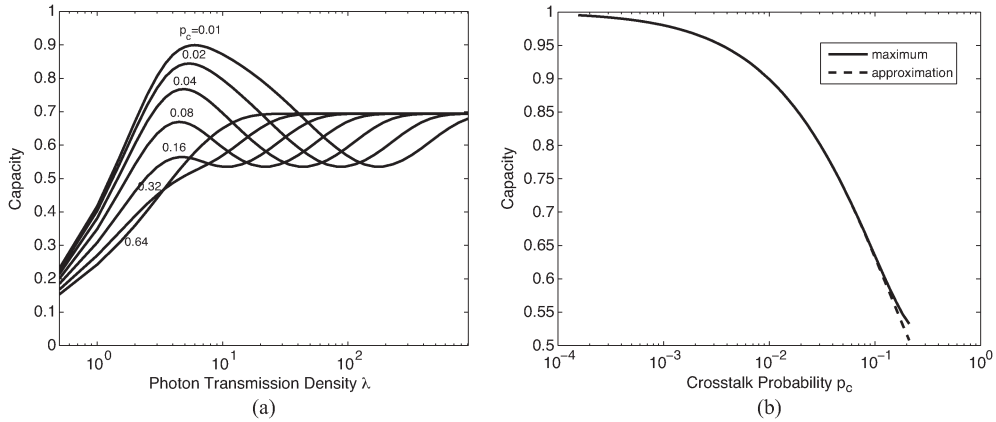


Fig. 2. (a) Capacities of optical channels with 1-D crosstalk. (b) Maximum capacities and their approximations based on (11) for different p_c.

be the global maximum when using the initialization $\Psi_0 = [1/2, 1/2, \ldots, 1/2]$.

B. Upper Bound on Capacity

To compute an upper bound on the capacity of an optical channel with 1-D crosstalk, we first show the following.

Lemma 1: For the OOK/SPD channel model (5), there exists a stationary and reflection-symmetric input distribution that achieves the capacity of the channel, i.e., a distribution $P_{\mathbf{x}}$ such that

$$P_{\mathbf{x}}\left(x_1^{m-1}\right) = P_{\mathbf{x}}\left(x_2^{m}\right),$$
$$P_{\mathbf{x}}(x_1, x_2, \ldots, x_m) = P_{\mathbf{x}}(x_m, x_{m-1}, \ldots, x_1).$$

Proof: Since [8] establishes that a stationary input distribution $P_{\mathbf{x}}$ achieves the capacity for a stationary channel with memory, we need only focus on the reflection-symmetry property. Accordingly, let $P'_{\mathbf{x}}$ be the reflection of the optimal stationary input distribution $P_{\mathbf{x}}$, i.e., $P'_{\mathbf{x}}(x_1, x_2, \ldots, x_m) = P_{\mathbf{x}}(x_m, x_{m-1}, \ldots, x_1)$. Clearly $P'_{\mathbf{x}}$ is a stationary input distribution that also achieves the capacity, and $u \triangleq (P_{\mathbf{x}} + P'_{\mathbf{x}})/2$ is both stationary and reflection-symmetric.

But $I_{P_{\mathbf{x}}}(\mathbf{x}; \mathbf{y})$, the mutual information for input distribution is $P_{\mathbf{x}}$, is a concave function of $P_{\mathbf{x}}$, i.e.,

$$I_u(\mathbf{x}; \mathbf{y}) \geq \frac{1}{2} I_{P_{\mathbf{x}}}(\mathbf{x}; \mathbf{y}) + \frac{1}{2} I_{P'_{\mathbf{x}}}(\mathbf{x}; \mathbf{y}),$$

so the stationary and reflection-symmetric distribution u must also achieve capacity. □

According to Lemma 1,

$$C(\lambda, p_c) = \lim_{m\to\infty} \max_{P_{\mathbf{x}}} \frac{1}{m} I(\mathbf{x}; \mathbf{y})$$

s.t. $P_{\mathbf{x}}$ is a stationary and reference-symmetric distribution. To obtain our upper bound, we use

$$I(\mathbf{x}; \mathbf{y}) = \sum_{i=1}^{m} \left[H\left(y_i | y_1^{i-1}\right) - H\left(y_i | y_1^{i-1}, x_1^m\right) \right], \qquad (9)$$

and restrict our attention to a subset of inputs parameterized by t as depicted in Fig. 1(b). In particular, for $t < i \leq m-1$ we have

$$H\left(y_i | y_1^{i-1}\right) \leq H\left(y_i | y_{i-t}^{i-1}\right),$$
$$H\left(y_i | y_1^{i-1}, x_1^m\right) = H\left(y_i | x_{i-1}^{i+1}\right),$$

and since we can ignore edge effects for m sufficiently large, we obtain

$$C(\lambda, p_c) \leq C^+(\lambda, p_c) \triangleq \max_{\varphi} \left[H\left(y_i | y_{i-t}^{i-1}\right) - H\left(y_i | x_{i-1}^{i+1}\right) \right] \qquad (10)$$

for $t < i \leq m-1$, where φ is a stationary and reflection-symmetric distribution on x_{i-t-1}^{i+1}. Note that φ is the marginal distribution of $P_{\mathbf{x}}$, and this constraint on φ is a relaxation of the constraint that $P_{\mathbf{x}}$ is a stationary and reference-symmetric distribution.

The optimization in (10) is straightforward to carry out using convex-programming techniques. Indeed, we have the following result, whose proof is in Appendix A.

Lemma 2: For some t and $t < i \leq m-1$, let φ be a distribution on x_{i-t-1}^{i+1}, then $R^+(\varphi) = H(y_i|y_{i-t}^{i-1}) - H(y_i|x_{i-1}^{i+1})$ is concave in φ.

Furthermore, the constraints on φ are linear. Specifically, the stationarity constraint is $\varphi(x_{i-t-1}^{i}) = \varphi(x_{i-t}^{i+1})$, where both the left and the right terms are marginal distributions (and thus linear transformations) of φ. The linear constraint that φ is reflection-symmetric is

$$\varphi(x_{i-t-1}, x_{i-t}, \ldots, x_{i+1}) = \varphi(x_{i+1}, x_i, \ldots, x_{i-t-1}).$$

And the remaining constraints $\sum_i \varphi_i = 1$ and $0 \leq \varphi_i \leq 1$ are also linear.

C. Capacities

As depicted in Fig. 2, (8) and (10) provide very tight bounds on the capacity of the channel (5) when we choose, e.g., $k = 6$, $d = 4$, and $t = 5$. Fig. 2(a) shows the resulting capacities for a wide range of λ and values of p_c chosen according to the geometric sequence $0.01, 0.02, 0.04, \ldots, 0.64$.

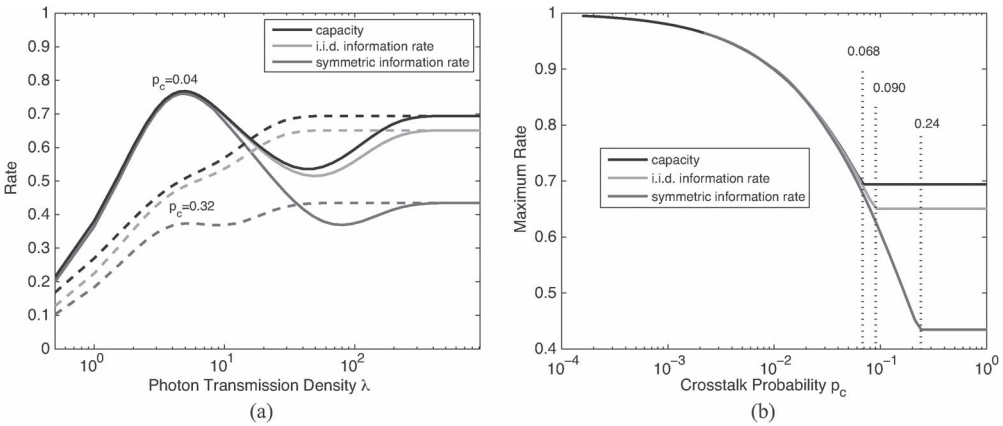


Fig. 3. (a) Information rates of optical channels with 1-D crosstalk. (b) Maximum capacity and maximum information rates of optical channels with 1-D crosstalk.

Note that as λ changes, the channel capacities for different values of p_c exhibit qualitatively different behaviors. In particular, there is a threshold $p^* \cong 0.22$ such that when $p_c \geq p^*$, the capacity is a nondecreasing function of λ, while for $p_c < p^*$, the capacity has two local maxima, one at a finite transmission density λ^* that can be well-approximated by

$$\lambda^* \simeq -2\log_{10}(p_c) + 2, \tag{11}$$

and the other at infinite transmission density. The local maximum at finite transmission density is larger than the second one when $p_c \leq p^{**} \cong 0.068$.

The figure also reflects the following behavior, whose proof is provided in Appendix B.

Proposition 3: The capacity of the channel (5) satisfies $C(\lambda, p_c) \to C_\infty = \log\gamma \cong 0.6942$ as $\lambda \to \infty$, where γ is the largest real root of the polynomial $x^4 - 2x^3 + x^2 - 1$.

We observe that the rate of convergence of capacity to C_∞ depends on p_c, with larger p_c resulting in faster convergence rates. Note, too, that for given a transmission density λ, larger values of p_c do not necessarily result in lower channel capacities.

Fig. 2(b) demonstrates the excellent quality of the approximation of the first local maximum capacity based on (11). Fig. 2(b) also demonstrates that even small amounts of crosstalk lead to substantial performance degradation. For example, $p_c \cong 0.01$ leads to a loss in maximum capacity of roughly 10%.

D. Information Rates

Fig. 3(a) compares the channel capacity $C(\lambda, p_c)$, the i.i.d. information rate $R_{\text{iid}}(\lambda, p_c)$, and the symmetric (uniform i.i.d.) information rate $R_{\text{sym}}(\lambda, p_c)$ when $p_c = 0.04$ and 0.32. As λ increases, $R_{\text{iid}}(\lambda, p_c) \to R^*_{\text{iid}} \cong 0.65$, $R_{\text{sym}}(\lambda, p_c) \to R^*_{\text{sym}} \cong 0.43$, independent of p_c. Moreover, $R_{\text{iid}}(\lambda, p_c)$ is much closer than $R_{\text{sym}}(\lambda, p_c)$ to $C(\lambda, p_c)$, implying that in practice we can use an i.i.d. input distribution to achieve near-capacity performance. As λ decreases, the gaps between both the information rates and $C(\lambda, p_c)$ decrease. In particular, when p_c is small, the maxima of both $R_{\text{iid}}(\lambda, p_c)$ and $R_{\text{sym}}(\lambda, p_c)$ (where λ is not large) are very close to the maximum of $C(\lambda, p_c)$.

Fig. 3(b) further plots the maximum i.i.d. information rate $R^{\max}_{\text{iid}}(p_c) = \max_{\lambda \geq 0} R_{\text{iid}}(\lambda, p_c)$ and the maximum symmetric information rate $R^{\max}_{\text{sym}}(p_c) = \max_{\lambda \geq 0} R_{\text{sym}}(\lambda, p_c)$, comparing them to $C^{\max}(p_c)$. Evidently, $R^{\max}_{\text{iid}}(p_c) \cong C^{\max}(p_c)$ when $p_c \leq p^{**}$. Note that there is a small gap between $R^{\max}_{\text{sym}}(p_c)$ and $C^{\max}(p_c)$ but this gap vanishes quickly with decreasing p_c. And for $R^{\max}_{\text{iid}}(p_c)$ and $R^{\max}_{\text{sym}}(p_c)$, there are turning point thresholds $p^*_{\text{iid}} \cong 0.090$ and $p^*_{\text{sym}} \cong 0.24$, respectively, analogous to p^* for $C^{\max}(p_c)$. Hence, for systems with sufficiently small p_c, near optimum performance can be achieved with low transmission density even when restricting attention to uniform i.i.d. codebooks.

The effect of crosstalk on the $C^{\max}(p_c)$ can be analyzed when p_c is sufficiently small, since $C^{\max}(p_c) \cong R^{\max}_{\text{sym}}(p_c)$ in this regime. In particular, we obtain the following theorem, whose proof is provided in Appendix C.

Theorem 4: For the channel of (5),

$$R^{\max}_{\text{sym}}(p_c) = 1 - \frac{\log_2 e}{4} p_c (\log p_c)^2 + o\left(p_c (\log p_c)^2\right). \tag{12}$$

IV. Channels With 2-D Crosstalk

In this section, we extend our results to the case of 2-D crosstalk for our simple four-neighbor model

$$P_{\mathbf{Y}|\mathbf{X}}(y|x) = \prod_{i=1}^{m}\prod_{j=1}^{m} P(y_{i,j}|x_{\mathcal{N}(i,j)}),$$
$$\mathcal{N}(i,j) = \{(i,j), (i-1,j), (i+1,j), (i,j-1), (i,j+1)\}. \tag{13}$$

The calculation of capacity is more involved in the 2-D case. Finite-state ISI channels in 2-D have been studied mostly in the context of magnetic and optical recording devices. For example, [7], [25], [24] introduce methods to compute upper bounds on the symmetric information rate. Another approach that applies simulation-based methodologies, the "generalized belief propagation" algorithm is developed in [29]. Due to the high computational complexity, tight capacity bounds are only obtained for some special constraints. Specifically, lower bounds based for bit-stuffing encoders or tilling encoders are presented and analyzed in, e.g., [11], [13], [28], [31], and upper bounds are provided, e.g., in [10], [32].

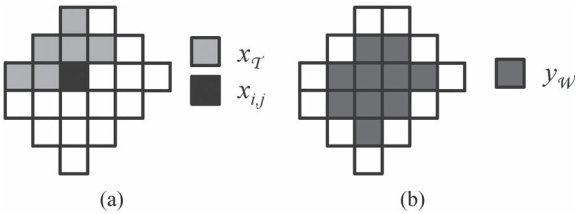


Fig. 4. Regions for calculating the lower bound $R^-(\lambda, p_c)$, where the black square is for $x_{i,j}$, the light gray squares of the input are for $x_{\mathcal{T}}$, and the dark gray squares of the output are for $y_{\mathcal{W}}$. (a) Input. (b) Output.

A. Lower Bound on Capacity

Our lower bound is obtained based on computing an achievable i.i.d. information rate. Specifically, we can bound the i.i.d. information rate according to

$$R(\lambda, p_c) = \lim_{m\to\infty} \frac{I(\mathbf{X};\mathbf{Y})}{m^2} \tag{14}$$

$$= H(x_{i,j}) - \lim_{m\to\infty} \frac{1}{m^2} \sum_{i,j=1}^{m} H\left(x_{i,j}|x^{i,j-1}, \mathbf{Y}\right) \tag{15}$$

$$\geq H(x_{i,j}) - H(x_{i,j}|x_{\mathcal{T}}, y_{\mathcal{W}}). \tag{16}$$

To obtain (15), we use (6) and the i.i.d. input distribution, where $x^{i,j-1}$ consists of all indices (i', j') such that either $i' < i$, or $i' = i$ and $j' < j$. To obtain (16), we increase the entropy by reducing the conditions. The subsets $\mathcal{T}$ and $\mathcal{W}$ are finite neighborhoods of (i, j), where $\mathcal{T}$ is a subset of that used for $x^{i,j-1}$; we ignore edge effects in the limit $m \to \infty$. Hence we have

$$C(\lambda, p_c) \geq R^-(\lambda, p_c) \triangleq \max_{0\leq p_x\leq 1} \left[H(x_{i,j}) - H(x_{i,j}|x_{\mathcal{T}}, y_{\mathcal{W}})\right] \tag{17}$$

where p_x is the parameter of i.i.d. input distribution. Suitable regions that render (17) tractable are depicted in Fig. 4.

B. Upper Bound on Capacity

The capacity-achieving input distributions for (13) have special properties we will exploit to obtain an upper bound. Towards this goal, we begin with definitions that are natural extensions of their 1-D counterparts.

Definition 1: A 2-D distribution φ is *stationary* if and only if $\varphi(x_{\mathcal{U}}) = \varphi(x_{\mathcal{U}'})$ for any two sets of indices such that $\mathcal{U}'$ is a shift of $\mathcal{U}$, i.e., $\mathcal{U}' = \{(i+\alpha, j+\beta) : (i, j) \in \mathcal{U}\}$ for two integers α, β. Furthermore, a stationary distribution φ is *reflection-symmetric* if and only if $\varphi(x_{\mathcal{U}}) = \varphi(x_{\mathcal{U}'})$ for any sets of indices $\mathcal{U}$ and $\mathcal{U}'$ that satisfy one of the following conditions:

1) $\mathcal{U}'$ is a horizontal reflection of $\mathcal{U}$, i.e., $\mathcal{U}' = \{(\alpha - i, \beta + j) : (i, j) \in \mathcal{U}\}$ for some α, β;
2) $\mathcal{U}'$ is a vertical reflection of $\mathcal{U}$, i.e., $\mathcal{U}' = \{(\alpha + i, \beta - j) : (i, j) \in \mathcal{U}\}$ for some α, β;
3) $\mathcal{U}'$ is a diagonal reflection of $\mathcal{U}$, i.e., $\mathcal{U}' = \{(\alpha + j, \beta + i) : (i, j) \in \mathcal{U}\}$ for some α, β.

In [32], it is shown there exists a stationary and reflection-symmetric input distribution that achieves the capacity under some special 2-D constraints. In the following theorem, whose proof we provide in Appendix D, we show that this is also true for the channel (13).

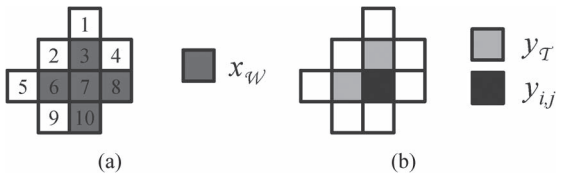


Fig. 5. Regions for calculating the upper bound $C^+(\lambda, p_c)$, where the black square of the output is for $y_{i,j}$, the light gray squares of the output are for $y_{\mathcal{T}}$, and the dark gray squares of the input are for $x_{\mathcal{W}}$. (a) Input. (b) Output.

Theorem 5: There exists a stationary reflection-symmetric input distribution that achieves the capacity of the channel (13).

Analogous to the 1-D case, we obtain an explicit upper bound on capacity by optimizing over finite-memory distributions, yielding

$$C(\lambda, p_c) \leq C^+(\lambda, p_c) \triangleq \max_{\varphi} \left[H(y_{i,j}|y_{\mathcal{T}}) - H(y_{i,j}|y_{\mathcal{T}}, x_{\mathcal{W}})\right] \tag{18}$$

where the optimization of φ is over stationary and reflection-symmetric distributions on a suitable region $\mathcal{S}$ that includes $\mathcal{T}$, $\mathcal{W}$ and (i, j), and $\mathcal{T}$ is also a subset of indices used for $y^{i,j-1}$. Using an argument similar to that used to obtain Lemma 2, it can be proved that the quantity in brackets in (18) is concave in φ. Furthermore, the stationary and reflection-symmetric constraints on φ are linear. Hence, (18) can be computed via convex programming. As the number of variables in this problem is exponential in $|\mathcal{S}|$, we use the small regions depicted in Fig. 5. In the figure, we label all the input bits as $1, 2, \ldots, 10$ from left to right and then from top to bottom, then due to the stationary requirement, it has linear constraints

$$\varphi\left(x_1^4 x_7^8\right) = \varphi\left(x_2 x_5^7 x_9^{10}\right), \quad \varphi\left(x_2^3 x_5^7 x_9\right) = \varphi\left(x_3^4 x_6^8 x_{10}\right);$$

and due to the reflection-symmetric requirement, it has linear constraints

$$\varphi\left(x_1^4 x_6^8 x_{10}\right) = \varphi\left(x_{10} x_6^8 x_2^4 x_1\right),$$

$$\varphi\left(x_2^3 x_5^{10}\right) = \varphi(x_3 x_2 x_8 x_7 x_6 x_5 x_{10} x_9),$$

$$\varphi(x_3 x_6 x_4 x_7 x_9 x_8 x_{10}) = \varphi(x_8 x_{10} x_4 x_7 x_9 x_3 x_6).$$

To obtain upper bounds on $R_{\text{iid}}(\lambda, p_c)$ and $R_{\text{sym}}(\lambda, p_c)$, we specialize φ in (18) to the case of an i.i.d. and uniform i.i.d. distribution, respectively. In these cases, sharper bounds are possible since the comparatively larger regions $\mathcal{S}$, $\mathcal{T}$ and $\mathcal{W}$ depicted in Fig. 6 can be used. Indeed, due to the independence of all the input bits, the joint distribution of $y_{i,j}, y_{\mathcal{T}}, x_{\mathcal{W}}$ can be computed efficiently by iteration.

C. Capacities

Fig. 7 depicts the upper and lower bounds on the capacity of optical channels with 2-D crosstalk as a function of λ. We see that when p_c and λ are small, the upper and lower bounds are tight. But the gap increases with p_c and λ.

The capacity of (13) evidently exhibits behavior similar to the 1-D case. When p_c is small, $C^{\max}(p_c)$ is achieved at a low transmission density λ^*, which can be well-approximated by (11).

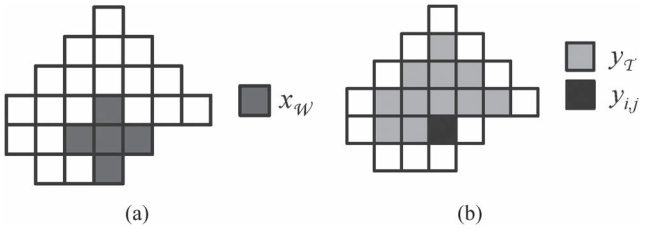


Fig. 6. Regions for computing an upper bound on information rates for 2-D crosstalk, where the black square of the output is for $y_{i,j}$, the light gray squares of the output are for $y_{\mathcal{T}}$, and the dark gray squares of the input are for $x_{\mathcal{W}}$. (a) Input. (b) Output.

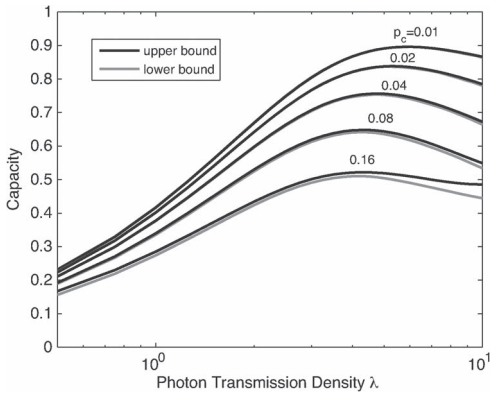


Fig. 7. Upper and lower bounds of the capacities with 2-D crosstalk for different p_c.

Proposition 6: The capacity of an optical OOK channel with 2-D crosstalk converges to C_∞, which is independent of p_c when $p_c > 0$, as the transmission density λ goes to infinity.

For an optical channel with 2-D crosstalk, when the transmission density $\lambda \to \infty$, if $x_{i,j} = 1$, then $y_{i,j} = y_{i-1,j} = y_{i+1,j} = y_{i,j-1} = y_{i,j+1} = 1$. So each one in the input array produces a '+' shape in the output array. The output array $\mathbf{Y} \in \{0, 1\}^{m\times m}$ can be treated as a 2-D constrained code, as a union of a collection of '+' shapes. Obviously, the resulting channel is independent of p_c when $p_c > 0$, so is its capacity. With numerical techniques of computing the bounds on the capacity of some 2-D constraints, we obtain that $0.5639 \le C_\infty \le 0.6126$. Details of the computation are provided in Appendix E.

D. Information Rates

Fig. 8(a) shows the lower and upper bounds on the i.i.d. information rate and the symmetric information rate when $p_c = 0.04$ or 0.3. From this figure, we observe that the bounds on the i.i.d. information rate are tight when λ is not large. In general, given λ and p_c, the optimal i.i.d. input distribution is more efficient than the uniform input distribution for achieving high information rate, especially when p_c or λ is large. When the transmission density goes to infinity, the i.i.d. information rate converges to a constant between 0.5639 and 0.6126; at the same time, the symmetric information rate converges to a constant between 0.1804 and 0.1855.

According to Fig. 7, when both p_c and λ are not large, the lower bound on the i.i.d. information rate (which is also used as the lower bound on the capacity) is very close to the upper bound on the capacity. When $\lambda \to \infty$ (no matter what the value p_c is), Theorem 6 showed that the upper bound on the capacity is 0.6126, which is close to the lower bound 0.5639 for the i.i.d. information rate with $\lambda \to \infty$. It implies that when $\lambda \to \infty$, the i.i.d. information rate is near the capacity. Both the evidences support our intuition that the optimal i.i.d. input distribution can achieve information rate very close to the capacity for optical channels with 2-D crosstalk. Fig. 8(b) further compares the maximum i.i.d. information rate and the maximum symmetric information rate with the optimal transmission density λ when $\lambda \le 10$. We see that when p_c is small, the maximum symmetric information rate is very close to the maximum i.i.d. information rate, implying that one can use regular linear codes such as LDPC codes for the error-correcting purpose when crosstalk is small.

V. The Photon Efficiency

In this section, we investigate the tradeoff between the information rate and photon efficiency of optical OOK channels with crosstalk when the expected number of photons transmitted per slot/pixel ϵ is small.

A. Background: Photon Efficiency Without Crosstalk

Without crosstalk, the information rate with an i.i.d. input distribution is just that of a Z channel with input distribution $\Pr\{x = 1\} = p_x$ and transition probability $\Pr\{y = 1|x = 1\} = 1 - \exp\{-\lambda\} \triangleq \mu(\lambda)$, whence the information rate $R(p_x, \lambda) = H(p_x\mu(\lambda)) - p_x H(\mu(\lambda))$, where $H(q)$ is the entropy of a Bernoulli distribution with probability q. This rate should be optimized keeping $p_x\lambda = \epsilon$ fixed. In [9], the tradeoff between PPM information rate and photon efficiency was investigated. We are interested in the small-ϵ behavior. As it turns out, in that limit the optimal choice is given by [18]

$$p_x^* = \frac{\epsilon}{2}\log\frac{1}{\epsilon}, \quad \lambda^* = \frac{\epsilon}{p_x^*}. \tag{19}$$

Substituting, the photon efficiency can be shown to be[1]

$$\begin{aligned}\tilde{C}(\epsilon) &= \left(1 - \frac{\lambda^*}{2}\right)\log_2\frac{1}{\epsilon} - \log_2\frac{1}{\lambda^*} + o(1)\\ &= \log_2\frac{1}{\epsilon} - \log_2\ln\frac{1}{\epsilon} - \log_2 e + 1 + o(1).\end{aligned}$$

While the efficiency is unbounded as ϵ decreases, the rate $r = \epsilon\tilde{C}(\epsilon)$ goes to zero.

We note that in the high-efficiency regime the resulting channel is highly skewed, both in terms of input distribution and in terms of crossover probability. Thus, the task of coding may be very difficult. This can be solved by replacing the i.i.d. binary codebook by pulse-position modulation (PPM), an idea already exploited in [21], [23]: the input sequence consists of "frames" of length $b = \lceil 1/p_x^* \rceil$, where each frame includes exactly one pulse, whose position is uniformly chosen inside the frame.[2] This scheme converts the channel to an b-ary erasure channel.

[1] See, e.g., [35] and the references therein for a finer-grained analysis that takes into account dark current.

[2] If the blocklength m is not divisible by b, then we ignore the remainder.

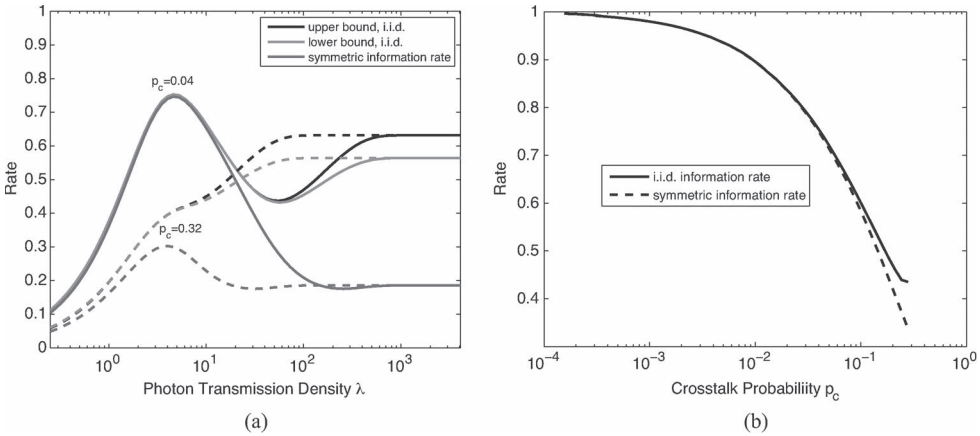


Fig. 8. (a) Information rates of optical channels with 2-D crosstalk. (b) Maximum information rates for 2-D crosstalk with different p_c.

By computing the capacity of this erasure channel, one finds that the photon efficiency of the PPM scheme is given by

$$\tilde{C}_{\mathrm{PPM}}(\epsilon) = \mu(b\epsilon) \cdot \frac{\log_2 b}{b\epsilon}, \tag{20}$$

which behaves the same as the expression we derived for OOK, i.e., the loss of efficiency is only $o(1)$. The resulting large-alphabet erasure channel is much like a packet-erasure channel encountered in Internet applications, and good off-the-shelf codes are available.

B. Photon Efficiency With Crosstalk

We have the following result for a crosstalk distribution p over a finite range $[-a, \cdots, 0, \cdots, a]$.

Proposition 7: For an i.i.d. input distribution, the OOK efficiency with crosstalk is given by

$$\tilde{C}(\epsilon) = \log_2 \frac{1}{\epsilon} - \log_2 \ln \frac{1}{\epsilon} - \log_2 e + 1 - H + o(1),$$
$$H \triangleq -\sum_{i=-a}^{a} p_i \log_2 p_i.$$

with H denoting the entropy of the crosstalk.

This result, reflecting an efficiency loss of $H + o(1)$ with respect to the crosstalk-free case, can be shown in the following way. Take the input to be i.i.d. with p_x^* (19). As we assumed that the crosstalk has a finite range, the probability of a specific slot to be in the range of two input pulses decays strongly, and does not effect the terms of interest. Thus, the output entropy remains the same as in the crosstalk-free case; the conditional entropy, in turn, is reduced per each received photon by H, the entropy of the location of detection. Since we only consider an i.i.d. distribution in this limit, the result can be extended in a straightforward manner to the 2-D case. It is highly plausible that such a distribution is optimal (to the order of interest), since the inter-pulse effects are very small; however, we do not formally prove this.

Now, consider the use of PPM in this setting. There is a single pulse per frame, and since λ is low, with very high probability there is at most one detection. Thus, if a frame does not result in an erasure, it will result in a detection in one slot. The effect of crosstalk is thus reduced to a slot error, making the task of coding much easier. The resulting photon efficiency is given by

$$\tilde{C}_{\mathrm{PPM}}(\epsilon) = \mu(b\epsilon) \cdot \frac{\log_2 b - H}{b\epsilon} + o(1), \tag{21}$$

where the correction term is due to the probability for more than one detection. Analyzing this expression reveals that PPM attains the optimum expression of Proposition 7.

VI. Selection of Slot/Pixel Size

So far we have considered a discrete model, where the relevant space is partitioned into slots or pixels. However, in many cases the system designer has control over the slot size, which affects the crosstalk distribution p as shown in (1). In this section we consider the selection of this parameter.

We note in advance that while performance can only increase as the slot/pixel size is decreased, in general such a decrease is accompanied by a corresponding increase in system complexity, in terms of hardware and/or computation. Accordingly, in our analysis, we restrict our attention to relatively simple architectures, and develop good slot/pixel sizes choices for them. Moreover, we are able to provide performance guarantees for such optimized schemes. We emphasize in advance, however, that our analysis does not preclude the possiblity of alternative achitectures that may be practical and efficient in comparatively short slot/pixel size regimes.

A. Non-Adjacent Crosstalk

In the continuous 1-D model, when the slot length is very small, we face a challenge in calculating the channel capacity: crosstalk may happen between non-adjacent slots, unlike the model we analyzed. Specifically, let $f(z)$ be the probability density function of the position of a photon in a received pulse centered at $z = 0$, given a slot length ℓ, the probability of having crosstalk is given by

$$p_{\mathrm{c}}(\ell) = 1 - p_0 = 1 - \int_{-\ell/2}^{\ell/2} f(z)\,\mathrm{d}z,$$

and the probability of having non-adjacent crosstalk is given by

$$p_{\mathrm{n}}(\ell) = 1 - p_{-1} - p_0 - p_1 = 1 - \int_{-3\ell/2}^{3\ell/2} f(z)\,\mathrm{d}z.$$

When the crosstalk probability is monotonously decreasing and the slot length is not too small, the non-adjacent crosstalk is low, but may not be negligible. We solve the issue by considering the adjacent crosstalk as the first-order approximation, while treating the non-adjacent crosstalk as additional random noise, which introduces a penalty term. Due to the practical concern of constructing error-correcting codes, we focus on uniform i.i.d. input distributions. Thus, we define a penalized symmetric information rate

$$R^{-}_{\mathrm{sym}}(\ell) \triangleq \max_{\lambda} \left[R_{\mathrm{sym}}(\lambda, p_{\mathrm{c}}(\ell)) - \Delta(\lambda, p_{\mathrm{n}}(\ell)) \right] \tag{22}$$

where $R_{\mathrm{sym}}(\lambda, p_{\mathrm{c}}(\ell))$ is the symmetric information rate with transmission density λ and crosstalk probability $p_{\mathrm{c}}(\ell)$ and

$$\Delta(\lambda, p_{\mathrm{n}}(\ell)) = \frac{1}{2} H\left(\min\left(1 - e^{-p_{\mathrm{n}}(\ell)\lambda/2}, 1/2 \right) \right)$$

is the penalty term. We are interested in how the slot length ℓ affects the rate $R^{-}_{\mathrm{sym}}(\ell)$, and based on which we further study the selection of ℓ with different application requirements in the following subsections.

Lemma 8: $R^{-}_{\mathrm{sym}}(\ell)$ is a lower bound on the symmetric information rate with slot length ℓ.

Proof: For the bound, we only need to upper-bound the penalty term caused by non-adjacent crosstalk. When the input sequence is stationary and uniformly distributed, the expected number of photons falling into a specific slot caused by non-adjacent crosstalk is $\lambda_n(\ell) = p_{\mathrm{n}}(\ell) p_x \lambda = p_{\mathrm{n}}(\ell)\lambda/2$. The probability for a slot having at least one photon from non-adjacent slots is thus $1 - \exp\{-\lambda_n(\ell)\}$, yielding the desired bound. □

It is also interesting to consider the case where the photon budget is limited. Let $\tilde{C}(\epsilon, \ell)$ be the photon efficiency when the average number of photons per slot is ϵ and the slot length is ℓ. Evaluating $\tilde{C}(\epsilon, \ell)$ in the limit of small ϵ. That is, since for any positive ℓ, the probability to send a pulse p_x is small in the limit of small ϵ. Consequently, each detector can only have detections originating from a single beam. As a direct corollary of Proposition 7 we have $\tilde{C}(\epsilon, \ell) = \log_2(1/\epsilon) - \log_2 \ln(1/\epsilon) - \log_2 e + 1 - H(\ell) + o(1)$, where $H(\ell)$ is the entropy of the crosstalk probabilities p_i.

B. Application: Slot-Length for Maximizing Throughput

In optical communications with time-division multiplexing or wavelength-division multiplexing techniques, it is desired to maximize the channel throughput, i.e., the number of available bits transmitted per second (or per Hz). For our models, the channel throughput can be written as the ratio between the information rate (i.e., the number of bits transmitted per slot) and the slot length (either in seconds or Hz). Intuitively, as slot size shrinks, one can get more slots for transmitting information, but meanwhile, the crosstalk probability increases and hence the capacity of each slot may decrease, and also, it is much more difficult to design efficient error-correcting codes. In order to optimize the performance of optical communications systems, it is important to understand the relation between the channel throughput and the selected slot length, as well as the respective crosstalk probability.

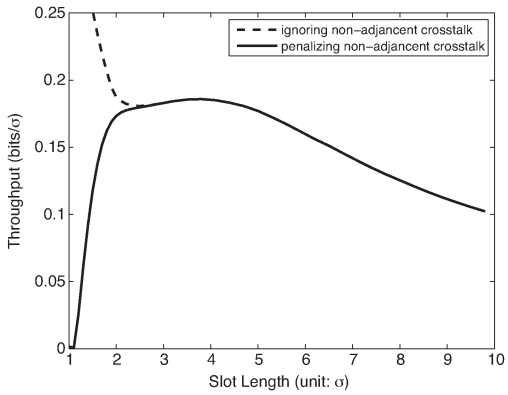


Fig. 9. The relation between the channel throughput and the slot length for Gaussian pulses with variance σ^2.

We first consider a simple model in which the positions z of the received photons in a pulse are independent and Gaussian distributed with mean at the middle of the target slot (position $z = 0$) and variance σ^2, for instance, $\sigma = 200$ ps. With this model, the crosstalk probability is given by $p_{\mathrm{c}}(\ell) = \Pr(|z| \geq \ell/2)$, and the probability of having non-adjacent crosstalk is $p_{\mathrm{n}}(\ell) = \Pr(|z| \geq 3\ell/2)$.

Fig. 9 shows achievable channel throughput $r(\ell)$ based on the lower bound $r^{-}(\ell) = R^{-}_{\mathrm{sym}}(\ell)/\ell$, which penalizes all non-adjacent-slot crosstalk. As a comparison, an upper bound on the channel throughput of a system with the uniform i.i.d. input distribution is derived by simplify ignoring all non-adjacent-slot crosstalk, see the dashed curve in the figure. Since it is difficult to handle non-adjacent crosstalk in practical applications, we are in particular interested in the local maximum in the figure as a feasible operating point. We see that a local optimal slot length that maximizes $r^{-}(\ell)$ is about $\ell = 4\sigma$, which is 800 ps when σ is 200 ps. In this case, the crosstalk probability is about $p_{\mathrm{c}} = 0.06$, the optimal transmission density is $\lambda = 4.4$ photons per slot, and the throughput of a single channel is about 220 Mb/s. We also observe that the channel throughput is not very sensitive to the slot length around the maximum, so in practical systems, we can choose the slot length slightly larger than 4σ, e.g., 5σ, for improving the coding efficiency (the crosstalk probability is further reduced) and reducing the computational cost. Typically, for Gaussian-like pulses, an optimum optical-communication system works with a low transmission density and a little crosstalk.

But for some rare occasions where light pulses have sharp boundaries, a system may work better with a high transmission density and a big crosstalk probability. Let's consider an extreme ideal model: the photons in each light pulse are uniformly distributed in an interval of width $w = 1000$ ps when they arrive at the receiver. In this case, the maximum throughput is achieved with slot length $\ell = w/3 = 333.3$ ps and a very high transmission density λ. According to Lemma 3, this maximum throughput is equal to $0.6942/\ell = 2.08/w = 2.08$ Gb/s. As a comparison, if we choose $\ell = w$, then it leads to a channel without crosstalk, but the respective throughput is reduced to $1/w = 1$ Gb/s.

In the energy-constrained case where photon efficiency is of primary interest, stronger results are possible. In this case, the

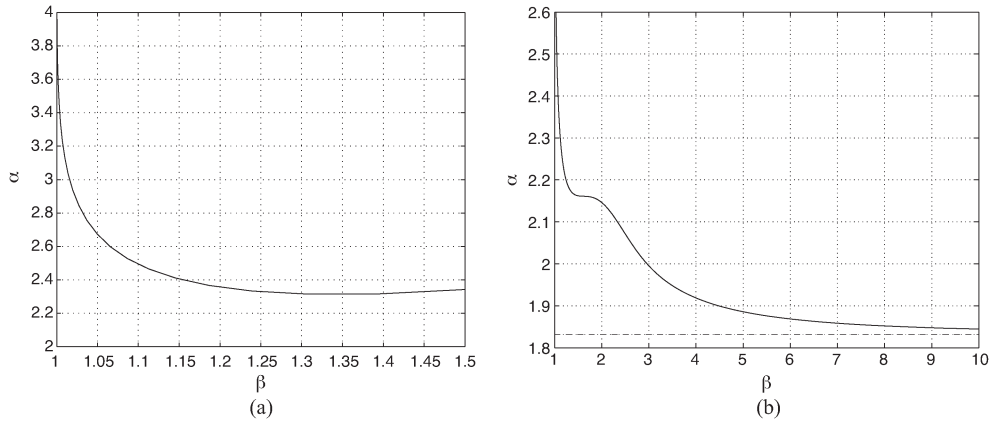


Fig. 10. (a) Minimum Fresnel-number product redundancy α as a function of the detector number redundancy β. (b) Minimum Fresnel-number product redundancy α as a function of the detector number redundancy β when the photon budget is tight. Dashed line depicts the asymptotic value α_0.

lower bounds are tight for all slot lengths, and thus the full spectrum of tradeoffs can be evaluated. However, to avoid unnecessary duplication of exposition, we omit this development here and instead provide in the next section as part of a 2-D application.

C. Application: Pixel Size in Free-Space Communications With Multiple Spatial Modes

We now consider a free-space communication scenario, where the required rate cannot be supported by a single optical mode, thus multiple modes are required. We extend the analysis applied so far, by explicitly taking into account the cost of using many detectors. Thus, we considerate the balance between performance in terms of bits/area and bits/mode. We note that in principle, one could work with orthogonal modes, avoiding crosstalk altogether, however these are undesired from a practical point of view. Thus, we quantify the loss (in terms of are and of number of modes) of a simple system, compared to orthogonal modes.

Assume that one needs to transmit N bits per channel use (defined in the time domain, that we keep discrete). In principle, this can be done using N OOK/SPD transmitter-detector pairs, each one communicating 1 bit per channel use, avoiding crosstalk altogether, as long as spatially orthogonal modes are utilized. The optimal "packing" of modes is given by prolate spheroidal function analysis [30]. The maximal number of orthogonal modes that a link can carry is given by the *Fresnel-number product* $\Gamma \triangleq A_{\mathrm{t}}A_{\mathrm{r}}/(\nu L)^2$, where A_{t} and A_{r} are the transmit and receive apertures, respectively, ν is the carrier wavelength and L is the distance between transmitter and receiver (see, e.g., [38]).

Unfortunately, both the generation of orthogonal modes and their detection are very hard tasks. Furthermore, they are severely distorted by atmospheric phenomena such as dispersion. A practical design would thus use non-orthogonal modes. We analyze a simple scheme where N sources of Gaussian beams are multiplexed and transmitted through a common aperture, towards an array of square detectors of edge ℓ each. Each beam is directed towards the center of a unique detector. In the beam, the displacements of photons from the beam center are independent and distributed according to circularly-symmetric zero-mean Gaussian with variance σ^2 (see, e.g., [20]). The beam width is related to the geometry of the system via:

$$\sigma^2 = \frac{\pi}{16} \cdot \frac{(\lambda L)^2}{A_t} = \frac{\pi}{16} \cdot \frac{A_r}{\Gamma}.$$

Given the problem geometry and the choice of pixel edge ℓ we have exactly the discrete model analyzed in the previous sections, thus we can find the per-pixel capacity C. We can use this number to quantify the system parameters as follows.

1) Number of pixels. In order to convey N bits per channel use, we will need $N/C \triangleq \beta^2 N$ pixels. The parameter $\beta > 1$ reflects the increase in the required hardware (transmitter-detector pairs).
2) Aperture. Since each pixel require an area of ℓ^2, the Fresnel number needed is

$$\Gamma = \frac{\pi A_r}{16\sigma^2} = \pi\beta^2 \left(\frac{\ell}{4\sigma}\right)^2 N \triangleq \alpha^2 N. \qquad (23)$$

The parameter $\alpha > 1$ reflects the increase in Fresnel number, which can be achieved, e.g., by increasing the transmit and apertures each by factor α.

An achievable tradeoff between the penalty factors α and β can now be easily evaluated, substituting a 2-D Gaussian distribution in $R^{-}_{\mathrm{sym}}(\ell)$ defined in (22). Fig. 10(a) shows that the relation between the factors α and β, where the minimal value of α is achieved at $\ell = 4\sigma$. A working point that seems to be good for practical purposes is $(\alpha, \beta) = (2.37, 1.19)$, which is achieved at $\ell \cong 4.5\sigma$. Note that σ is proportional to the distance L between transmitter and receiver, assume that $\sigma = 10$ mm and $\ell = 45$ mm, then the crosstalk probability is 4.83%. At this point, the radii of the transmit and receive apertures need to be multiplied by a factor 1.54 each, comparing to the ideal orthogonal system, and the number of detectors needs to be 42% more comparing to the same.

We now return to the setting where the photon budget is tight. As we saw, in this limit it is nearly optimal to use a PPM scheme. In terms of the multiple spatial-mode system considered in this section, this serves to save hardware: rather than having a transmitter per spatial mode, we can have one transmitter per b modes, where b is the size of PPM symbol,

and a suitable mechanism to dynamically direct the beam to the center of the required detector. Taking the high-efficiency limit also simplifies the derivation of the α-β tradeoff and qualitatively changes the results, as follows.

Suppose that we are to send B bits per use of the aperture, at $\tilde{R}$ bits per photon. To that purpose we use $\tilde{N}$ detectors, operating at $\tilde{\epsilon}$ photons per detector. The parameters can be found by solving:

$$\tilde{\epsilon}\tilde{R}\tilde{N} = B, \qquad \tilde{R} = \tilde{C}(\epsilon, \ell).$$

Now, let N and ϵ be the same parameters without crosstalk. Then,

$$\beta^2 = \frac{\tilde{N}}{N} = \frac{\epsilon}{\tilde{\epsilon}}.$$

In the 2-D case the axes decompose, the rate-loss due to crosstalk is $2H(\ell)$ instead of $H(\ell)$, where $H(\ell)$ is the entropy of 1-D quantized Gaussian noise with variance σ^2. Since the photon efficiency is held fixed, we have by Proposition 7:

$$\log_2 \frac{1}{\epsilon} - \log_2 \ln \frac{1}{\epsilon} = \log_2 \frac{1}{\tilde{\epsilon}} - \log_2 \ln \frac{1}{\tilde{\epsilon}} - 2H(\ell)$$

which, in the limit of high efficiency, gives $\beta = 2^{H(\ell)}$. Given this value of β, α can be found using (23). Scanning over all $\ell > 0$, one finds the optimal tradeoff, depicted in Fig. 10(b). The rather strange curve is due to the shape of the Gaussian distribution. We can analytically consider the tradeoff in the two extremes. If $\alpha \ll \beta$, Z is a high-resolution version of Gaussian noise, thus

$$H(Z_\ell) \cong \frac{1}{2}\log_2(2\pi e\sigma^2) - \log_2 \ell = \log_2\left(\frac{\pi}{2}\sqrt{\frac{e}{2}}\cdot\frac{\beta}{\alpha}\right)$$

thus (23) yields $\alpha = \frac{\pi}{2}\sqrt{\frac{e}{2}} \triangleq \alpha_0 \cong 1.83$. This is a minimum value for α at any β, meaning that there must be a loss in the Fresnel-number product with respect to orthogonal modes, even if we use many small detectors. If, to the contrary, $\alpha \gg \beta$, then Z may have very low entropy. Thus in the limit of a very large aperture (almost orthogonal beams) we can have β close to one, representing no increase in the number of detectors, as expected. Fig. 10(b) depicts a numerical evaluation of the tradeoff. A working point that seems to be good for practical purposes is $(\alpha, \beta) = (2.15, 1.4)$, since any improvement in one of the parameters beyond this point will incur a large penalty in the other. At this point, the radii of the transmit and receive apertures need to be multiplied by a factor 1.5 each, comparing to the ideal orthogonal system, and the number of detectors need to be doubled comparing to the same. This is achieved with pixel edge $\ell \cong 4.5\sigma$. Again, if $\sigma = 10$ mm, then we can set $\ell = 45$ mm that leads to a crosstalk probability 4.83%.

VII. Concluding Remarks

In this paper, we studied communication over optical OOK channels with 1-D or 2-D crosstalk, and developed mathematical analysis tools that can be used to help system engineers to choose key parameters (such as transmission power, input distribution, slot length) for optimizing overall performance of optical communication systems. We concentrated on simple, practical schemes, thus assumed transmission of coherent pulses with on-off keying, and single-photon detectors. With these choices, we demonstrated non-trivial tradeoffs: the optimal transmission density may be finite, and a good choice for the slot size will allow some amount of crosstalk. We also considered the same tradeoffs in the case of a low photon budget, and showed that in this limit the analysis simplifies and we can have yet sharper results.

Our derivations correspond to a simple system design, which admits practical error-correcting codes. Still, the efficient design of such codes for optical OOK channels with crosstalk, the interplay between code design and the system parameters, and the experimental validation of the modeling assumptions in various applications, are left for further research.

Appendix

A. Proof of Lemma 2

The conditional probability $P(y_i|x_{i-1}^{i+1})$ only depends on the channel. Hence, for any $a \in \{0,1\}^3$, $H(y_i|x_{i-1}^{i+1} = a)$ is a constant. In this case, $H(y_i|x_{i-1}^{i+1}) = \sum_a P(x_{i-1}^{i+1} = a)H(y_i|x_{i-1}^{i+1} = a)$ is a linear function of $P(x_{i-1}^{i+1})$. Also, because $P(x_{i-1}^{i+1})$ is a marginal distribution of φ, it can written as a linear transformation of φ. Hence, $H(y_i|x_{i-1}^{i+1})$ is a linear function of φ, and we only need to prove that $H(y_i|y_{i-t}^{i-1})$ is a concave function of φ.

Note that

$$H\left(y_i|y_{i-t}^{i-1}\right) = -\sum_{y_{i-t}^{i}} P\left(y_{i-t}^{i}\right)\log_2\frac{P\left(y_{i-t}^{i}\right)}{P\left(y_{i-t}^{i-1}\right)}.$$

in which $P(y_{i-t}^{i})$ is a linear transformation of φ. It is sufficient to show that $P(y_{i-t}^{i})\log_2\frac{P(y_{i-t}^{i})}{P(y_{i-t}^{i-1})}$ is a convex function of $P(y_{i-t}^{i})$.

Let $P_1(y_{i-t}^{i})$ and $P_2(y_{i-t}^{i})$ be two arbitrary distributions on y_{i-t}^{i}. Assume that $P(y_{i-t}^{i}) = \epsilon P_1(y_{i-t}^{i}) + (1-\epsilon)P_2(y_{i-t}^{i})$ for an arbitrary ϵ with $0 \le \epsilon \le 1$. Then $P(y_{i-t}^{i-1}) = \epsilon P_1(y_{i-t}^{i-1}) + (1-\epsilon)P_2(y_{i-t}^{i-2})$.

According to the log-sum inequality,

$$P\left(y_{i-t}^{i}\right)\log_2\frac{P\left(y_{i-t}^{i}\right)}{P\left(y_{i-t}^{i-1}\right)} \le \epsilon P_1\left(y_{i-t}^{i}\right)\log_2\frac{P_1\left(y_{i-t}^{i}\right)}{P_1\left(y_{i-t}^{i-1}\right)} + (1-\epsilon)P_2\left(y_{i-t}^{i}\right)\log_2\frac{P_2\left(y_{i-t}^{i}\right)}{P_2\left(y_{i-t}^{i-1}\right)}.$$

It shows that $P(y_{i-t}^{i})\log_2\frac{P(y_{i-t}^{i})}{P(y_{i-t}^{i-1})}$ is a convex function of $P(y_{i-t}^{i})$. Finally, we can claim that $R^+(\varphi)$ is a concave function of φ. □

B. Proof of Proposition 3

When $p_c > 0$ and $\lambda \to \infty$, the optical channel with 1-D crosstalk can be simplified: if $x_i = 1$ then $y_{i-1} = y_i = y_{i+1} = 1$. In this case, the capacity of the channel can be written as

$$C_\infty = \lim_{m\to\infty}\frac{\log_2 N_m}{m},$$

where N_m is the total number of feasible patterns for $\mathbf{y} \in \{0,1\}^m$. It is easy to see that the capacity is independent of the crosstalk probability p_c when $p_c > 0$. Actually, we can consider $\mathbf{y}$ as a constrained code where the number of consecutive 1's is at least three.

We let H_m be the number of feasible patterns for $\mathbf{y}' = \mathbf{y}111$ with $\mathbf{y} \in \{0,1\}^m$ ($\mathbf{y}$ followed by three ones) that satisfies the above constraint. We find the following iterative relation between N_m and H_m,

$$N_m = H_{m-3} + N_{m-1}, \tag{24}$$

in which, H_{m-3} is the number of feasible patterns for $\mathbf{y}$ if $\mathbf{y}$ ends with 1, and N_{m-1} is the number of feasible patterns for $\mathbf{y}$ if $\mathbf{y}$ ends with 0. Similarly, we also get

$$H_{m-3} = H_{m-4} + N_{m-4}. \tag{25}$$

Substituting (25) into (24) leads to $N_m - 2N_{m-1} + N_{m-2} - N_{m-4} = 0$.

Because γ is the biggest real root of the equation $x^4 - 2x^3 + x^2 - 1 = 0$,

$$\lim_{m\to\infty} \frac{\log_2 N_m}{m} = \log_2 \gamma.$$

Hence, the asymptotic capacity of an optical channel with 1-D crosstalk is $C_\infty = \log_2 \gamma \cong 0.6942$. □

C. Proof of Theorem 4

According to (7), a lower bound on the symmetric information rate is

$$R^-(\lambda, p_c) = H(x_i) - H(x_i|y_i) = H(y_i) - H(y_i|x_i). \tag{26}$$

According to the properties of the optical channels, we have

$$P(y_i = 0|x_i = 1) = \frac{1}{4}e^{-\lambda} + \frac{1}{2}e^{-\lambda(1-\frac{p_c}{2})} + \frac{1}{4}e^{-\lambda(1-p_c)},$$

$$P(y_i = 0|x_i = 0) = \frac{1}{4} + \frac{1}{2}e^{-\lambda\frac{p_c}{2}} + \frac{1}{4}e^{-\lambda p_c}.$$

Now, we write $P(y_i = 0|x_i = 1)$ as p_{10}, and write $P(y_i = 0|x_i = 0)$ as p_{00}. Based on (26), we have

$$R^-(\lambda, p_c) = h\left(\frac{1}{2}p_{10} + \frac{1}{2}p_{00}\right) - \frac{1}{2}h(p_{10}) - \frac{1}{2}h(p_{00}),$$

with $h(\cdot)$ denoting, as before, the entropy of a Bernoulli distribution.

Note that when x is a small value, $h(x) = h(1-x) = -x\log_2 x + o(x\log x)$ and $h(x + 1/2) = 1 - o(x)$.

Let $\lambda' = \log(1/p_c)$. As p_c becomes small enough, based on the approximations above, we can get that

$$R^-(\lambda', p_c) = 1 - \frac{\log_2 e}{4}p_c\left(\log\frac{1}{p_c}\right)^2 + o\left(p_c\left(\log\frac{1}{p_c}\right)^2\right), \tag{27}$$

which is a lower bound of the maximum symmetric information rate $R^{\max}_{\text{sym}}(p_c)$.

According to (8), an upper bound on the symmetric information rate is $R^+(\lambda, p_c) = H(y_i) - H(y_i|x_{i-1}^{i+1})$.

Note that $H(y_i) = h((p_{10} + p_{00})/2)$. As $p_c \to 0$, $H(y_i) \to h((1 + e^{-\lambda})/2)$. When $p_c \ll 1$, in order to make $H(y_i)$ to converge to 1, we need $\lambda \gg 1$, and meanwhile, $\lambda p_c \ll 1$.

For the second term $H(y_i|x_{i-1}^{i+1})$, we have

$$H\left(y_i|x_{i-1}^{i+1}\right) = \frac{1}{8}h(e^{-\lambda}) + \frac{2}{8}h\left(e^{-\lambda(1-\frac{p_c}{2})}\right) + \frac{1}{8}h\left(e^{-\lambda(1-p_c)}\right) + \frac{2}{8}h\left(e^{-\lambda\frac{p_c}{2}}\right) + \frac{1}{8}h\left(e^{-\lambda p_c}\right),$$

and we show that when $\lambda \gg 1$ and $\lambda p_c \ll 1$,

$$H\left(y_i|x_{i-1}^{i+1}\right) \geq \frac{\log_2 e}{4}p_c\left(\log\frac{1}{p_c}\right)^2 + o\left(p_c\left(\log\frac{1}{p_c}\right)^2\right). \tag{28}$$

Let λ' be a value satisfying $e^{-\lambda'} = \lambda' p_c$, from which we can get that

$$\lambda' = \log\frac{1}{p_c} + o\left(\log\frac{1}{p_c}\right).$$

When $\lambda < \lambda'$ and $\lambda \gg 1$,

$$H\left(y_i|x_{i-1}^{i+1}\right) \geq \frac{1}{8}h\left(e^{-\lambda'}\right) + \frac{2}{8}h\left(e^{-\lambda'(1-\frac{p_c}{2})}\right) + \frac{1}{8}h\left(e^{-\lambda'(1-p_c)}\right) = \frac{\log_2 e}{2}\lambda'^2 p_c + o(\lambda'^2 p_c) \tag{29}$$

When $\lambda \geq \lambda'$ and $\lambda p_c \ll 1$,

$$H\left(y_i|x_{i-1}^{i+1}\right) \geq \frac{2}{8}h\left(e^{-\lambda'\frac{p_c}{2}}\right) + \frac{1}{8}h\left(e^{-\lambda' p_c}\right) = \frac{\log_2 e}{4}\lambda'^2 p_c + o\left(\lambda'^2 p_c\right) \tag{30}$$

Based on (29) and (30), we can see that (28) holds. As a result,

$$R^+(\lambda, p_c) = H(y_i) - H\left(y_i|x_{i-1}^{i+1}\right) \leq 1 - \frac{\log_2 e}{4}p_c\left(\log\frac{1}{p_c}\right)^2 + o\left(p_c\left(\log\frac{1}{p_c}\right)^2\right). \tag{31}$$

Comparing (31) with (27), we see that the upper and lower bounds on the maximum symmetric information rate match. □

D. Proof of Theorem 5

We first prove the following weaker claim.

Lemma 9: There exists a stationary input distribution that achieves the capacity of the optical OOK channels with 2-D crosstalk.

Proof: Let φ be an optimal input distribution that achieves the capacity of an optical channel with 2-D crosstalk. Assume that $x \in \{0,1\}^{m\times m}$ is an assignment of the array. We define another distribution $\sigma_{\alpha,\beta}(\varphi)$ with $0 \leq \alpha, \beta < m$ as follow:

$$\sigma_{\alpha,\beta}(\varphi)(x) = \varphi(\sigma_{\alpha,\beta}(x)). \tag{32}$$

Here, $\sigma_{\alpha,\beta}(x)$ is a shift of the $m \times m$ matrix x by α rows and β column, i.e., $\sigma_{\alpha,\beta}(x)$ equals

$$\begin{pmatrix} x_{\alpha+1,\beta+1} & \cdots & x_{\alpha+1,m} & x_{\alpha+1,1} & \cdots & x_{\alpha+1,\beta} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ x_{m,\beta+1} & \cdots & x_{m,m} & x_{m,1} & \cdots & x_{m,\beta} \\ x_{1,\beta+1} & \cdots & x_{1,m} & x_{1,1} & \cdots & x_{1,\beta} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ x_{\alpha,\beta+1} & \cdots & x_{\alpha,m} & x_{\alpha,1} & \cdots & x_{\alpha,\beta} \end{pmatrix}.$$

Since φ is optimal,

$$C(\lambda, p_c) = \lim_{m\to\infty} \frac{(\mathbf{X}(\varphi); \mathbf{Y}(\varphi))}{m^2}.$$

It can be proved that

$$I(\mathbf{X}(\varphi); \mathbf{Y}(\varphi)) \le I(\mathbf{X}(\sigma_{\alpha,\beta}(\varphi)); \mathbf{Y}(\sigma_{\alpha,\beta}(\varphi))) + 8m - 16.$$

As a result, we can get

$$C(\lambda, p_c) = \lim_{m\to\infty} \frac{I\left[\mathbf{X}(\sigma_{\alpha,\beta}(\varphi)); \mathbf{Y}(\sigma_{\alpha,\beta}(\varphi))\right]}{m^2},$$

showing that the shifted distribution $\sigma_{\alpha,\beta}(\varphi)$ is asymptotically optimal.

Now, define a new input distribution μ, such that

$$u = \frac{1}{m^2} \sum_{\alpha,\beta=0}^{m-1} \sigma_{\alpha,\beta}(\varphi).$$

It can be proved that μ is stationary (the proof is emitted). Since mutual information is a concave function of the input distribution, we can also prove that μ can achieve the capacity of optical channels with 2-D crosstalk asymptotically. □

Now we turn to our proof of Theorem 5. We define a new distribution $\sigma_h(\varphi)$, as a horizontal reflection of φ, such that $\sigma_h(\varphi(x)) = \varphi(\sigma_h(x))$, where $\sigma_h(x)$ is the horizontal reflection of x, i.e., $\sigma_h(x)_{i,j} = x_{m+1-i,j}$ for all $1 \le i, j \le m$.

Let μ be the average of φ and $\sigma_h(\varphi)$, i.e., $\mu = (\varphi + \sigma_h(\varphi))/2$. Then μ is a stationary and horizontally reflection-symmetric input distribution that achieves the capacity of the optical channel with 2-D crosstalk asymptotically.

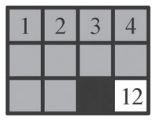
Using the same technique, we can further prove that there exists a stationary and reflection-symmetric (horizontally, vertically and diagonally) input distribution that that achieves the capacity of the optical channel with 2-D crosstalk asymptotically. □

E. Bounds of C_∞ for 2-D Crosstalk

The lower bound is computed based on i.i.d. input distribution, as described in Section IV-A. The upper bound is computed as follows, based on the convex-programming technique described in [32]. It is tighter than that derived based on the method described in Section IV-B.

The basic idea is based on the conclusion that when the array (i.e., m) is infinitely large, there exits an optimal distribution φ on the output array $\mathbf{Y}$, which is stationary and reflection-symmetric.

We consider the projection of the optimal distribution on a 3×4 sub-array T, denoted by φ_T, and we label all the nodes by $1, 2, \ldots$ from left to right and then from top to bottom as follows:



The channel capacity satisfies $C_\infty \le \max_{\varphi_T} H(y_{11}|y_1^{10})$ [32] such that φ_T is a stationary, reflection-symmetric and *feasible* distribution.

In the optimization problem, the objective function is a concave function of φ_T. The condition that φ_T is a stationary, reflection-symmetric and feasible distribution can be written as a group of linear constraints:

1) φ_T is stationary, which implies that $\varphi_T(y_1^8) = \varphi_T(y_5^{12})$ and $\varphi_T(y_1^3 y_5^7 y_9^{11}) = \varphi_T(y_2^4 y_6^8 y_{10}^{12})$;
2) φ_T is reflection-symmetric, which implies that
$$\varphi_T\left(y_1^{12}\right) = \varphi_T(y_4, y_3, y_2, y_1, \ldots, y_{12}, y_{11}, y_{10}, y_9) = \varphi_T\left(y_9^{12} y_5^8 y_1^4\right),$$
$$\varphi_T\left(y_1^3 y_5^7 y_9^{11}\right) = \varphi_T(y_{11} y_7 y_3 y_{10} y_6 y_2 y_9 y_5 y_1);$$
3) φ_T is feasible (it satisfies the 2-D constraint), i.e., if y_1^{12} is not a feasible assignment then $\varphi_T(y_1^{12}) = 0$. All feasible assignments for y_1^{12} can be enumerated by brute-force search;
4) φ_T is a distribution, hence $0 \le \varphi_{Ti}$ and $\sum_i \varphi_{Ti} = 1$.

Finally, the upper bound 0.6126 can be obtained based on convex programming. □

REFERENCES

[1] H. Arjmandi, G. Aminian, A. Gohari, M. N. Kenari, and U. Mitra, "Capacity of diffusion based molecular communication networks over LIT-Poisson channels," arXiv:1410.3988v2 [cs.IT], 2014.
[2] D. M. Arnold and H. A. Loeliger, "On the information rate of binary input channels with memory," in *Proc. IEEE Int. Conf. Commun.*, Helsinki, Finland, Jun. 2001, vol. 9, pp. 2692–2695.
[3] D. M. Arnold, H. A. Loeliger, P. O. Vontobel, A. Kavcic, and W. Zeng, "Simulation-based computation of information rates for channels with memory," *IEEE Trans. Inf. Theory*, vol. 52, no. 8, pp. 3498–3508, Aug. 2006.
[4] S. Arnon, "Underwater optical wireless communication network," *Optical Engineering*, vol. 49, no. 015001, Jan. 2010.
[5] I. Bar-David, "Communication under the Poisson regime," *IEEE Trans. Inf. Theory*, vol. 15, pp. 31–37, Jan. 1969.
[6] N. Chandrasekaran and J. H. Shapiro, "Turbulence-induced crosstalk in multiple-spatial-mode optical communication," presented at the CLEO: Science Innovations, Optical Society of America, San Jose, CA, USA, 2012.
[7] J. Chen and P. H. Siegel, "On the symmetric information rate of two dimensional finite state ISI channels," *IEEE Trans. Inf. Theory*, vol. 52, no. 1, pp. 227–236, Jan. 2006.
[8] J. Chen and P. H. Siegel, "Markov processes asymptotically achieve the capacity of finite state intersymbol interference channels," *IEEE Trans. Inf. Theory*, vol. 54, no. 3, pp. 1295–1303, Mar. 2008.
[9] S. Dolinar, K. M. Birnbaum, B. I. Erkmen, and B. Moision, "On approaching the ultimate limits of photon-efficient and bandwidth-efficient optical communication," in *Proc. ICSOS*, May 2011, pp. 269–278.
[10] S. Forchhammer and J. Justesen, "Bounds on the capacity of constrained two-dimensional codes," *IEEE Trans. Inf. Theory*, vol. 46, no. 7, pp. 2659–2666, Nov. 2000.
[11] S. Forchhammer and T. V. Laursen, "Entropy of bit-stuffing-induced measures for two-dimensional checkerboard constraints," *IEEE Trans. Inf. Theory*, vol. 53, no. 4, pp. 1537–1546, Apr. 2007.

[12] C. Gong and Z. Xu, "Particle stream channel modeling and estimation for non-line of sight optical wireless communication," in *Proc. IEEE GLOBECOM*, Dec. 2014, pp. 2114–2118.
[13] S. Halevy, J. Chen, R. M. Roth, P. H. Siegel, and J. K. Wolf, "Improved bit-stuffing bounds on two-dimensional constraints," *IEEE Trans. Inf. Theory*, vol. 50, no. 5, pp. 824–838, May 2004.
[14] W. Hirt and J. L. Massey, "Capacity of the discrete-time Gaussian channel with intersymbol interference," *IEEE Trans. Inf. Theory*, vol. 34, no. 3, pp. 380–388, May 1988.
[15] A. L. Kachelmyer and D. M. Boroson, "Efficiency penalty of photon-counting with timing jitter," presented at the Optical Engineering and Applications, International Society for Optics and Photonics, Bellingham, WC, USA, 2007.
[16] A. Kavcic, "On the capacity of Markov sources over noisy channels," in *Proc. IEEE GLOBECOM*, San Antonio, TX, USA, Nov. 2001, pp. 2997–3001.
[17] D. Kedar and S. Arnon, "Backscattering-induced crosstalk in WDM optical wireless communication," *J. Lightw. Technol.*, vol. 23, no. 6, pp. 2023–2030, Jun. 2005.
[18] Y. Kochman and G. W. Wornell, "On high-efficiency optical communication and key distribution," in *Proc. ITA*, 2012, pp. 172–179.
[19] A. Lapidoth and S. M. Moser, "On the capacity of the discretetime Poisson channel," *IEEE Trans. Inf. Theory*, vol. 55, no. 1, pp. 303–322, Jan. 2009.
[20] L. Mandel and E. Wolf, *Optical Coherence and Quantum Optics*. Cambridge, U.K.: Cambridge Univ. Press, 1995.
[21] J. L. Massey, "Capacity, cutoff rate, and coding for direct-detection optical channel," *IEEE Trans. Commun.*, vol. 29, no. 11, pp. 1615–1621, Nov. 1981.
[22] B. Moision and W. Farr, "Communication limits due to photon detector jitter," *IEEE Photon. Technol. Lett.*, vol. 20, no. 9, pp. 388–396, May 2008.
[23] J. Pierce, "Optical channels: Practical limits with photon counting," *IEEE Trans. Commun.*, vol. 26, no. 12, pp. 1819–1821, Dec. 1978.
[24] J. B. Soriaga, P. H. Siegel, J. K. Wolf, and M. Marrow, "On achievable rates of multistage decoding two-dimensional ISI channels," in *Proc. IEEE Int. Symp. Inf. Theory*, 2005, pp. 1348–1352.
[25] P. H. Siegel, "Information-theoretic limits of two-dimensional optical recording channels," in *Proc. IEEE Opt. Data Storage Topical Meet.*, Montreal, QC, Canada, Apr. 2006, pp. 165–167.
[26] S. Shamai and A. Lapidoth, "Bounds on the capacity of a spectrally constrained Poisson channel," *IEEE Trans. Inf. Theory*, vol. 39, no. 1, pp. 19–29, Jan. 1993.
[27] V. Sharma and S. K. Singh, "Entropy and channel capacity in the regenerative setup with applications to Markov channels," in *Proc. IEEE Int. Symp. Inf. Theory*, Washington, DC, USA, Jun. 2001, p. 283.
[28] A. Sharov and R. M. Roth, "Two-dimensional constrained coding based on tiling," *IEEE Trans. Inf. Theory*, vol. 56, no. 4, pp. 1800–1807, 2010.
[29] O. Shental *et al.*, "Discrete-input two-dimensional Gaussian channels with memory: Estimation and information via graphical models and statistical mechanics," *IEEE Trans. Inf. Theory*, vol. 54, no. 4, pp. 1500–1513, Apr. 2008.
[30] D. Slepian, "Prolate spheroidal wave functions, Fourier analysis and uncertainty-IV: Extension to many dimensions; generalized prolate spheroidal functions," *Bell Syst. Tech. J.*, vol. 43, no. 6, pp. 3009–3057, Nov. 1964.
[31] I. Tal and R. M. Roth, "Bounds on the rate of 2-D bit-stuffing encoders," *IEEE Trans. Inf. Theory*, vol. 56, no. 6, pp. 2561–2567, Jun. 2010.
[32] I. Tal, and R. M. Roth, "Convex programming upper bounds on the capacity of 2-D constraints," *IEEE Trans. Inf. Theory*, vol. 57, no. 1, pp. 381–391, Jan. 2011.
[33] P. O. Vontobel and D. M. Arnold, "An upper bound on the capacity of channels with memory and constraint input," in *Proc. IEEE Inf. Theory Workshop*, Cairns, QLD, Australia, Sep. 2001, pp. 147–149.
[34] P. O. Vontobel, A. Kavcic, D. M. Arnold, and H. A. Loeliger, "A generalization of the Blahut Arimoto algorithm to finite-state channels," *IEEE Trans. Inf. Theory*, vol. 54, no. 5, pp. 1887–1918, May 2008.
[35] L. Wang and G. W. Wornell, "A refined analysis of the Poisson channel in the high-photon-efficiency regime," in *Proc. ITW*, Lausanne, Switzerland, pp. 587–591, Sep. 2012.
[36] P. J. Winzer and G. J. Foschini, "MIMO capacities and outage probabilities in spatially multiplexed optical transport systems," *Opt. Exp.*, vol. 19, no. 17, pp. 16680–16696, 2011.
[37] S. Yang, A. Kavcic, and S. Tatikonda, "Feedback capacity of finite-state machine channels," *IEEE Trans. Inf. Theory*, vol. 51, no. 3, pp. 799–810, Mar. 2005.
[38] H. P. Yuen, and J. H. Shapiro, "Optical communication with two-photon coherent states-Part I: Quantum-state propagation and quantum-noise," *IEEE Trans. Inf. Theory*, vol. 24, no. 5, pp. 657–668, 1978.

Hongchao Zhou (S'10–M'12) received the B.Sc. degree in physics and mathematics and M.Sc. degree in control science and engineering from Tsinghua University, Beijing, China, in 2006 and 2008, respectively, and the M.Sc. and Ph.D. degrees in electrical engineering from the California Institute of Technology, Pasadena, CA, USA, in 2009 and 2012, respectively.

He is a Postdoctoral Researcher in the Research Laboratory of Electronics, Massachusetts Institute of Technology, Cambridge, MA, USA. His current interests include information theory and randomness, data storage, information security, and stochastic biological networks. He is a recipient of the 2013 Charles Wilts Prize for the best doctoral thesis in electrical engineering at the California Institute of Technology.

Yuval Kochman (S'06–M'09) received the B.Sc. (cum laude), M.Sc. (cum laude), and Ph.D. degrees from Tel Aviv University in 1993, 2003, and 2010, respectively, all in electrical engineering. During 2009–2011, he was a Postdoctoral Associate with the Signals, Information and Algorithms Laboratory, Massachusetts Institute of Technology (MIT), Cambridge, MA, USA. Since 2012, he has been with the School of Computer Science and Engineering, Hebrew University of Jerusalem. Outside academia, he has worked in the areas of radar and digital communications. His research interests include information theory, communications and signal processing.

Gregory W. Wornell (S'83–M'91–SM'00–F'04) received the B.A.Sc. degree from the University of British Columbia, Canada, and the S.M. and Ph.D. degrees from the Massachusetts Institute of Technology, all in electrical engineering and computer science, in 1985, 1987 and 1991, respectively.

Since 1991, he has been on the faculty at MIT, where he is the Sumitomo Professor of Engineering in the Department of Electrical Engineering and Computer Science. At MIT, he leads the Signals, Information, and Algorithms Laboratory within the Research Laboratory of Electronics. He is also Chair of Graduate Area I (Information and System Science, Electronic and Photonic Systems, and Physical Science and Nanotechnology) within the EECS department's doctoral program. He has held visiting appointments at the former AT&T Bell Laboratories, Murray Hill, NJ, USA, the University of California, Berkeley, CA, USA, and Hewlett-Packard Laboratories, Palo Alto, CA.

His research interests and publications span the areas of signal processing, statistical inference, digital communication, and information theory, and include algorithms and architectures for wireless and sensor networks, multimedia applications, imaging systems, and aspects of computational biology and neuroscience. He has been involved in the Signal Processing and Information Theory societies of the IEEE in a variety of capacities, and maintains a number of close industrial relationships and activities. He has won a number of awards for both his research and teaching.