

Capacity Bounds for Power- and Band-Limited Optical Intensity Channels Corrupted by Gaussian Noise

Steve Hranilovic, *Member, IEEE*, and Frank R. Kschischang, *Senior Member, IEEE*

Abstract—We determine upper and lower bounds on the channel capacity of power- and bandwidth-constrained optical intensity channels corrupted by white Gaussian noise. These bounds are shown to converge asymptotically at high optical signal-to-noise ratios (SNRs). Unlike previous investigations on low-intensity Poisson photon counting channels, such as some fiber optic links, this channel model is realistic for indoor free space optical channels corrupted by intense ambient light. An upper bound on the capacity is found through a sphere-packing argument while a lower bound is computed through the maxentropic source distribution. The role of bandwidth is expressed by way of the effective dimension of the set of signals and, together with an average optical power constraint, is used to determine bounds on the spectral efficiency of time-disjoint optical intensity signaling schemes. The bounds show that, at high optical SNRs, pulse sets based on raised-quadrate amplitude modulation (QAM) and prolate spheroidal wave functions have larger achievable maximum spectral efficiencies than traditional rectangular pulse basis sets. This result can be considered as an extension of previous work on photon counting channels which closely model low optical intensity channels with rectangular pulse shapes.

Index Terms—Channel capacity, direct detection, free space, infrared communication, optical communications, optical intensity modulation, wireless.

I. INTRODUCTION

INVESTIGATIONS into the capacity of optical intensity channels have focused on channels in which the dominant noise source is quantum in nature. In these channels, the transmitted optical intensity is constant in discrete time intervals. The received signal is modeled by a Poisson-distributed count of the number of received photons in each discrete interval. In this work, we present capacity bounds for a fundamentally different optical intensity channel. The indoor free-space optical channel can be modeled as a linear channel with additive, white, signal independent, Gaussian noise [1]. Unlike previous

treatments, capacity bounds are computed for any time-disjoint modulation scheme.

The literature is replete with channel capacity results for the photon-counting channel. Gordon and Pierce demonstrated that in the absence of external noise the capacity of such photon-counting channels (in units of nats per photon) under an average optical power limit is unbounded [2], [3]. In fact, M -ary pulse-position modulation (PPM) can achieve arbitrarily small probability of error for any rate (in units of nats per photon) [3], [4]. Under the constraint of fixed peak optical power, Davis and Wyner determined the capacity of this channel and showed that binary level modulation schemes were capacity achieving [5], [6]. These channel capacity derivations, however, do not consider the bandwidth of the transmitted signal. McEliece demonstrated that schemes based on photon counting in discrete intervals require an exponential increase in bandwidth as a function of the rate (in nats per photon) for reliable communication [4]. In the case of rectangular pulse-amplitude modulated (PAM) signals confined to discrete time intervals of length T and with a given peak and average optical power, Shamai [7] showed that the capacity achieving input distribution is discrete with a finite number of levels increasing with T .

In indoor free-space optical communication systems, the dominant noise source arises due to ambient light, which produces high-intensity shot noise. Indeed, in some fiber and free-space links, front-end thermal noise due to the electronic preamplifier is often a dominant source of noise. In both of these cases, the noise process can be modeled as being additive, white, Gaussian distributed and dominant over the quantum fluctuations of the channel [1]. Unlike the photon-counting model of the fiber channel, free-space optical channels are intrinsically bandwidth limited due to the use of large inexpensive optoelectronic components and multipath distortion. Multidimensional, multilevel constellations have been proposed for the indoor free-space channel to achieve high data rates [8], [9]. A previous upper bound on the capacity of this channel was determined for the specific case of multicarrier systems where the average optical amplitude in each disjoint symbol interval is fixed [10].

In this work, we compute upper and lower bounds on the capacity of optical intensity channels corrupted by Gaussian noise in the case of time-disjoint signaling. This work is not confined to rectangular pulse modulation techniques, as in earlier photon-counting channel results, but represents all time-disjoint techniques in a common framework. Unlike the previous capacity bound derived for this channel, the bounds presented do

Manuscript received May 14, 2002; revised December 7, 2003. This work was supported in part by an Ontario Graduate Scholarship in Science and Technology as well as a Walter C. Sumner Memorial Fellowship. The material in this paper was presented in part at the 21st Biennial Symposium on Communications, Kingston, ON, Canada, June 2002 and at the Fortieth Annual Allerton Conference on Communication, Control, and Computing, Monticello, IL, October 2002.

S. Hranilovic is with the Department of Electrical and Computer Engineering, McMaster University, Hamilton, ON L8S 4K1, Canada (e-mail: hrnilovic@ece.mcmaster.ca).

F. R. Kschischang is with the Edward S. Rogers Sr. Department of Electrical and Computer Engineering, University of Toronto, Toronto, ON M5S 1A4, Canada (e-mail: frank@comm.utoronto.ca).

Communicated by G. Caire, Associate Editor for Communications.

Digital Object Identifier 10.1109/TIT.2004.826649

not assume a particular signaling set and allow for the average optical amplitude of each symbol to vary.

Section II-A outlines the salient features of the optical intensity channel and a signal space model suited to defining time-disjoint optical intensity schemes is presented in Section II-B. A bandwidth constraint is cast as a restriction on the number of available dimensions in Section II-D. Upper and lower bounds on the capacity of the channel are derived in Sections III and IV and the capacity is given in terms of a spectral efficiency. The bounds are then shown to converge at high optical signal-to-noise ratios (SNRs). The paper concludes by presenting some example schemes in Section V and presents the implications of the results and future directions in Section VI.

II. PRELIMINARIES

A. The Optical Intensity Channel

Indoor wireless optical communication channels transmit information by directly modulating the optical power of a laser or light-emitting diode (LED) light source. Due to cost as well as eye and skin safety considerations, incoherent optical sources are used as transmitters and inexpensive optical intensity detectors are used at the receiver. Whereas electrical channels allow for the modulation of both the amplitude and the phase of the underlying carrier, these inexpensive wireless optical systems are able to modulate only the intensity of the optical carrier. As a result, all transmitted signals are constrained to be non-negative. Additionally, the average optical power transmitted on such channels is the average amplitude of the transmitted signal, unlike the wired channel, where electrical energy is the mean-square value of the transmitted signal.

Let $x(t)$ be some optical intensity signal to be transmitted. The channel, which is composed of the multipath response of the room as well as the electrical characteristics of the optoelectronics, can be modeled by a linear conversion between optical and electrical domains [1]. The transmitted signal is corrupted by additive white Gaussian noise (AWGN). The basis for this approximation is that electronic noise as well as high intensity shot noise are the dominant noise sources and are both additive, white in the band of interest, and approximately Gaussian distributed by a central limit theorem argument [1]. Thus, the received electrical signal $y(t)$ can be written as

$$y(t) = rx(t) + z(t)$$

where $r > 0$ is the *responsivity* of the photodiode in units of amperes per watt and is the optoelectronic conversion factor. Without loss of generality, we set $r = 1$ for the remainder of this paper. The random signal $z(t)$, in units of amperes, is zero-mean AWGN.

Since the transmitted signal is an intensity, $x(t)$ must satisfy $\forall t x(t) \geq 0$. Due to eye and skin safety regulations, the average optical power is limited, and hence the average amplitude of $x(t)$ is limited. The received electrical signal $y(t)$, however, can assume negative amplitude values. This channel model applies not only to free-space optical channels but also to fiber-optic links with negligible dispersion and signal-independent AWGN.

B. Signal Space

The set of time-disjoint optical intensity signals can be represented in a signal space which represents both the nonnegativity constraint and average optical power cost geometrically [9]. Let $\mathbb{N} = \{1, 2, \dots, N\}$ be a finite index set and let $\Phi = \{\phi_n(t) : n \in \mathbb{N}\}$ be a set of orthonormal basis functions where $\phi_n(t) = 0$ for $t \notin [0, T)$, $T > 0$. By definition, the basis function $\phi_1(t)$ is set to be

$$\phi_1(t) = \begin{cases} 1/\sqrt{T}, & t \in [0, T) \\ 0, & \text{otherwise.} \end{cases} \quad (1)$$

This basis function contains the average amplitude of each symbol and, as a result, represents the average optical power of each symbol. Due to the orthogonality of the other basis functions, $\int_0^T \phi_n(t) dt = 0$ for $n > 1$. In this manner, the average optical power requirement is represented in a single dimension.

The *admissible region* [9] of an optical intensity modulation scheme is the set Υ of all points in the signal space which describe nonnegative pulses, or formally

$$\Upsilon = \left\{ (x_1, x_2, \dots, x_N) \in \mathbb{R}^N : (\forall t \in \mathbb{R}), \sum_{n \in \mathbb{N}} x_n \phi_n(t) \geq 0 \right\}.$$

It is useful to partition Υ into sets of points with a given average optical power. Define

$$\Upsilon_k = \{(v_1, v_2, \dots, v_N) \in \Upsilon : v_1 = k, k \in \mathbb{R}^+\}$$

as the set of all points in Υ with average optical power k/\sqrt{T} . It can be shown that $\Upsilon_k = k\Upsilon_1$, $k \geq 0$, since scaling by nonnegative factors always yields a signal in Υ . Thus, $\Upsilon = \cup_{k \geq 0} k\Upsilon_1$. By definition, since the cross section of Υ perpendicular to the ϕ_1 basis for a given k scales linearly with k , the boundary surface of Υ is a generalized N -cone [11, p. 341]. Furthermore, since all convex linear combinations of points in Υ represent nonnegative signals, Υ is convex. Thus, Υ is the convex hull of a generalized N -cone with vertex at the origin [9].

C. Problem Definition

Consider transmitting symbols formed as linear combinations of the N basis functions in the set Φ . Assuming a bank of N matched filters at the receiver, the channel can be modeled as the N -dimensional vector channel model

$$\mathbf{Y} = \mathbf{X} + \mathbf{Z}$$

where each term is an N -dimensional random vector with probability densities $f_{\mathbf{Y}}(\mathbf{y})$, $f_{\mathbf{X}}(\mathbf{x})$, and $f_{\mathbf{Z}}(\mathbf{z})$, respectively.

The noise vector \mathbf{Z} is an N -dimensional Gaussian random vector with independent components of mean zero and variance σ^2 per dimension.

The transmitted vector \mathbf{X} is selected subject to both a nonnegativity constraint and an average optical power constraint. Let \mathcal{F} denote the set of densities which simultaneously satisfy both of these constraints. For all $f_{\mathbf{X}}(\mathbf{x}) \in \mathcal{F}$, $f_{\mathbf{X}}(\mathbf{x}) = 0$ for

$\mathbf{x} \notin \Upsilon$, in order to satisfy the nonnegativity constraint. Additionally, in order to satisfy the average amplitude, i.e., average optical power constraint, every $f_{\mathbf{X}}(\mathbf{x}) \in \mathcal{F}$ must satisfy

$$P \geq \frac{1}{\sqrt{T}} \int_{\mathbf{x} \in \Upsilon} x_1 f_{\mathbf{X}}(\mathbf{x}) d\mathbf{x} \quad (2)$$

for given bounded values of $P > 0$ and $T > 0$. Notice that the average optical power of the signaling scheme is not completely described by the geometry of the constellation, as in conventional signal space models, but also depends on the symbol period. This is due to the fact that $\phi_1(t)$ is set to have unit electrical energy. This scaling is appropriate since detection of the signal is performed in the electrical domain where the orthonormal basis defined is appropriate [9].

The problem of finding the maximum rate at which reliable communication can take place was treated by Shannon [12] and can be posed in the context of this channel as

$$C_s(\Phi) = \max_{f_{\mathbf{X}}(\mathbf{x}) \in \mathcal{F}} I(\mathbf{X}; \mathbf{Y}) \quad (3)$$

in units of bits per channel use where $I(\mathbf{X}; \mathbf{Y})$ is the mutual information between \mathbf{X} and \mathbf{Y} . Notice that this is the maximum achievable rate when the transmitted waveforms are restricted to be linear combinations of Φ . This is analogous to the case of the band-limited electrical channel model where all transmitted signals are scaled versions of the the basis function $\text{sinc}(\pi t/T)$.

In what follows, bounds on the maximum rate for a fixed average optical power are computed for a family of channel models each determined by the choice of basis Φ .

D. Bandwidth Constraint

Previous results with rectangular pulses on the photon-counting channel demonstrated that the rate is unbounded if the average optical power is the only constraint [2], [3]. Indeed, it is possible to show that signaling with arbitrarily narrow rectangular pulses in a symbol interval of T at a given average optical power causes the capacity in (3) to tend to infinity.

However, previous work on the photon-counting channel also indicated that this unbounded rate necessarily comes at the price of an infinite bandwidth requirement [4]. It is clear that in order to have a consistent bound or notion of maximum rate for this channel, a bandwidth constraint must be placed on the set of signals transmitted.

The *Landau–Pollak* dimension [13] of the set of signals is used here to cast the spectral constraint as an effective number of dimensions provided by the channel. Let $\mathcal{L}^2[0, T]$ denote the set of all finite-energy signals with support contained in $[0, T]$. Define the $(1-\epsilon)$ -fractional energy bandwidth $W_\epsilon(x)$ of a transmitted symbol $x(t) \in \mathcal{L}^2[0, T]$ with Fourier transform $X(f)$ as

$$W_\epsilon(x) = \inf \left\{ W \in [0, \infty): \int_{-W}^W |X(f)|^2 df \geq (1-\epsilon) \int_{-\infty}^{\infty} |X(f)|^2 df \right\} \quad (4)$$

where $\epsilon \in (0, 1)$ is fixed to some value, typically 10^{-2} or 10^{-3} . This bandwidth measure quantifies the frequency concentration

of $x(t)$. In practical terms, $x(t)$ can be thought of as being effectively band limited to $W_\epsilon(x)$ hertz if ϵ is chosen so that the out-of-band energy is below the noise floor of the channel [14].

A dimension for the signal $x(t) \in \mathcal{L}^2[0, T]$ can be defined through the fractional energy bandwidth. Consider approximating $x(t) \in \mathcal{L}^2[0, T]$ as a linear combination of some orthonormal basis functions. For a given $W_\epsilon(x)$ and T , the best such basis, in the sense of minimizing the energy in the error of the approximation, is the family of prolate spheroidal wave functions $\varphi_n(f)$ [15]. The $\varphi_n(f)$ are functions strictly time limited to $[0, T]$ which have the maximum energy in $[-W_\epsilon(x), W_\epsilon(x)]$ of all unit energy functions [16]. The error in the approximation can be upper-bounded as [13]

$$\inf_{\{a_i\}} \int_{-\infty}^{\infty} \left| X(f) - \sum_{n=0}^{\lceil 2W_\epsilon(x)T \rceil} a_n \varphi_n(f) \right|^2 df < 12\epsilon^2. \quad (5)$$

In this sense, the signal $x(t)$ can be thought of as being indistinguishable from some linear combination of prolate spheroidal basis functions. It can then be said that $x(t)$ is essentially $2W_\epsilon(x)T$ -dimensional with the error in the approximation tending to zero as $\epsilon \rightarrow 0$.

Define

$$\kappa(\Phi) = \max_{\mathbf{x} \in \Upsilon} 2W_\epsilon \left(\sum_{n \in \mathbb{N}} x_n \phi_n(t) \right) T \quad (6)$$

as the effective dimension of the signal space associated with the optical intensity basis Φ . The definition of $\kappa(\Phi)$ is equivalent to fixing T and defining the available channel bandwidth W_{ch} as the largest bandwidth of any transmittable signal, i.e., $W_{\text{ch}} = \kappa(\Phi)/2T$. This channel bandwidth definition can be interpreted as ensuring that the channel is able to support the transmission of at most $\kappa(\Phi)$ dimensions per symbol. Since each transmitted symbol in the model is at most $\kappa(\Phi)$ -dimensional, the received symbols are uncorrupted by the channel, i.e., the received signals are indistinguishable from the transmitted signals in the sense of (5).

In practice, it may be difficult to compute $\kappa(\Phi)$ for a given basis set. The effective dimension $\kappa(\Phi)$ can be computed by fixing T and maximizing the bandwidth over all transmittable signals. Without loss of generality, consider an $x(t) \in \mathcal{L}^2[0, T]$ which is unit energy. The bandwidth of this signal, as defined in (4), can be approximated by expanding it into its components as

$$W_\epsilon(x) \approx \inf \left\{ W \in [0, \infty): \sum_{n \in \mathbb{N}} x_n^2 \int_{-W}^W |\phi_n^F(f)|^2 df \geq (1-\epsilon) \right\} \quad (7)$$

where $\phi_n^F(t)$ is the Fourier transform of $\phi_n(t)$ and $\sum_m x_m^2 = 1$. Thus, the fractional energy bandwidth of a signal is approximated as a weighted sum of the fractional energy contained in each of the basis functions. This approximation is valid for large W or equivalently for small ϵ , when the approximation

$$\int_{-W}^W \phi_m^F(f) \phi_n^{F*}(f) df \approx 0$$

is valid for $m \neq n$. This approximation is used in some of the example schemes presented in Section V.

III. UPPER BOUND ON CHANNEL CAPACITY

An upper bound on the capacity of a Gaussian-noise-corrupted channel can be obtained by considering a sphere-packing argument in the set of all received codewords while imposing an average optical power constraint. This analysis is done in the same spirit as Shannon's sphere-packing argument for channels subject to an average electrical power constraint [17]. Determining this bound requires that the volume of the set of received codewords be computed for a given average optical power limit.

A. Set of Transmitted Codewords

Consider transmitting a codeword \mathbf{x} formed from a series of L, N -dimensional symbols with a low probability of error. Geometrically, in order for \mathbf{x} to be transmittable, $\mathbf{x} \in \Upsilon^L$, where Υ^L is the L -fold Cartesian product of Υ with itself. For the sake of notation, group the coordinate values for each constituent constellation together to form $\mathbf{x} \in \Upsilon^L$ as

$$\mathbf{x} = (x_{1,1}, \dots, x_{N,1}, x_{1,2}, \dots, x_{N,2}, \dots, x_{1,L}, \dots, x_{N,L}).$$

Recall from Section II-B that Υ is the convex hull of a generalized cone parameterized by the Υ_k cross sections at $\phi_1 = k$. The Cartesian product Υ^L represents the set of transmittable codewords formed by the concatenation in time of L time-limited symbols. As a result, Υ^L represents a time-limited optical intensity scheme and must therefore be the convex hull of a generalized cone with vertex at the origin. In an similar fashion to (1), define the ϕ_1^{LN} basis vector as

$$\phi_1^{LN} = \frac{1}{\sqrt{L}} \underbrace{(1, 0, 0, \dots, 0, 1, 0, 0, \dots, 0, 1, 0, 0, \dots)}_{LN} \quad (8)$$

so that it represents the average optical power of each LN -dimensional codeword $\mathbf{x} \in \Upsilon^L$. The region Υ^L is then parameterized by cross sections for a given ϕ_1^{LN} coordinate value.

For a fixed symbol period T , assume that the average optical power of each transmitted codeword is limited to be at most P as defined in (2). In terms of the signal-space definition for Υ

$$\frac{1}{L} \sum_{k=1}^L x_{1,k} \leq \sqrt{TP} \quad (9)$$

where $x_{1,k}$ is the coordinate value in the ϕ_1 direction for each constituent symbol. The transmitted LN -dimensional vector \mathbf{x} is taken from the set $\Theta(\sqrt{TP}) = \Upsilon^L \cap \Psi(\sqrt{TP})$, where the shaping region $\Psi(\sqrt{TP})$ is a hyperplane defined so that the power constraint (9) is satisfied. The region $\Psi(\sqrt{TP})$ can be expressed in terms of the signal space as well as in terms of (8) as

$$\begin{aligned} \Psi(\sqrt{TP}) &= \left\{ \mathbf{x} \in \mathbb{R}^{LN} : \frac{1}{L} \sum_{k=1}^L x_{1,k} \leq \sqrt{TP} \right\} \\ &= \left\{ \mathbf{x} \in \mathbb{R}^{LN} : \langle \mathbf{x}, \phi_1^{LN} \rangle \leq \sqrt{LTP} \right\} \end{aligned} \quad (10)$$

where $\langle \cdot, \cdot \rangle$ is the Euclidean inner product.

B. Set of Received Codewords

When $\mathbf{x} \in \Theta(\sqrt{TP})$ is transmitted, the received vector \mathbf{Y} is normally distributed with mean \mathbf{x} and variance equal to the

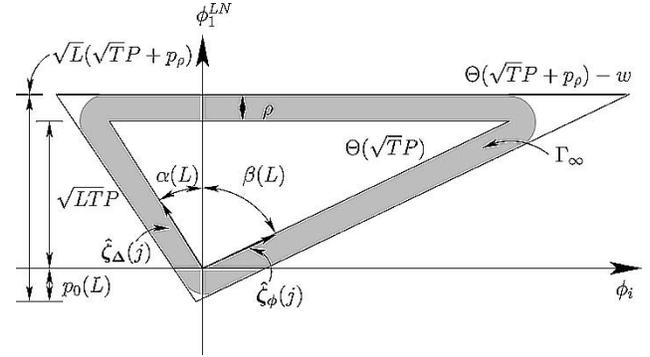


Fig. 1. Cross section of $\Theta(\sqrt{TP})$, Γ_∞ , and upper bound $\Theta(\sqrt{TP} + p_\rho)$.

noise variance σ^2 per dimension. Let Γ_{LN} denote the set of all possible received vectors. With high probability, \mathbf{Y} will lie near the surface of a sphere of radius ρ centered at \mathbf{x} where

$$\rho = \sqrt{LN\sigma^2}. \quad (11)$$

This assertion can be justified by viewing the components of the vector $\mathbf{Y} - \mathbf{x}$ as LN independent and identically distributed Gaussian random variables. By the weak law of large numbers, for every $\epsilon > 0$

$$\lim_{L \rightarrow \infty} \Pr \left\{ \sqrt{LN(\sigma^2 - \epsilon)} < \|\mathbf{Y} - \mathbf{x}\| < \sqrt{LN(\sigma^2 + \epsilon)} \right\} = 1. \quad (12)$$

Thus, the probability that \mathbf{Y} is not on a sphere of radius ρ about \mathbf{x} can be made arbitrarily small by increasing L . A codeword is decoded by assigning all vectors contained inside the sphere to the given codeword.

Define the region Γ_∞ as

$$\begin{aligned} \Gamma_\infty &= \left\{ \mathbf{x} + \mathbf{b} : \mathbf{x} \in \Theta(\sqrt{TP}), \mathbf{b} \in \rho B^{LN} \right\} \\ &= \Theta(\sqrt{TP}) \oplus \rho B^{LN} \end{aligned} \quad (13)$$

where the \oplus operation is the *Minkowski addition* of two sets [8, p. 126], and B^{LN} is the LN -dimensional unit ball. Since $\Theta(\sqrt{TP})$ is convex, Γ_∞ is termed the outer parallel body of $\Theta(\sqrt{TP})$ at a distance of ρ [18, p. 197] and is the set of all points with distance at most ρ from $\Theta(\sqrt{TP})$. Since (12) applies for all transmitted vectors, for large enough L the distance between any $\mathbf{y} \in \Gamma_{LN}$ and the corresponding $\mathbf{x} \in \Theta(\sqrt{TP})$ tends to ρ with probability arbitrarily close to one as L gets large. In other words, for large enough L , the probability that \mathbf{y} does not lie in Γ_∞ can be made arbitrarily small. Since the capacity calculations depend on the asymptotic behavior of Γ_{LN} in L , the properties of Γ_∞ must be determined in order to determine an upper bound. Fig. 1 presents an example two-dimensional cross section of $\Theta(\sqrt{TP})$ and the corresponding Γ_∞ region.

Clearly, $\Theta(\sqrt{TP}) \subset \Gamma_\infty$ since $\mathbf{0} \in B^{LN}$. Where the boundary of $\Theta(\sqrt{TP})$ is smooth, the boundary points of Γ_∞ are a subset of the points parallel to $\Theta(\sqrt{TP})$ at distance ρ away. Form the parallel extension of $\Theta(\sqrt{TP})$ as the region $\Theta(\sqrt{TP} + p_\rho) - w$, for some constants $w, p_\rho > 0$, as the set of points which are at most distance of ρ away from $\Theta(\sqrt{TP})$ whenever the boundary of $\Theta(\sqrt{TP})$ is smooth. At points of discontinuity, that is, in the "corners" of the bodies in question, the points in Γ_∞ lie inside the parallel extension of $\Theta(\sqrt{TP})$

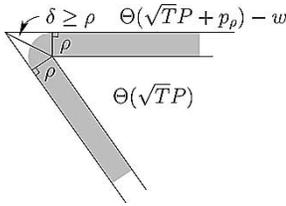


Fig. 2. Parallel extension of $\Theta(\sqrt{TP})$ at a distance ρ at point of discontinuity. By triangle inequality, $\delta \geq \rho$.

at a distance ρ away due to the triangle inequality, as illustrated in Fig. 2. In other words

$$\Theta(\sqrt{TP}) \subset \Gamma_\infty \subset \Theta(\sqrt{TP} + p_\rho) - w. \quad (14)$$

C. Upper Bound on the Volume of Γ_∞

Since Γ_∞ is the Minkowski sum of a convex set and an LN -ball, its volume can be computed explicitly by way of *Steiner's formula* [18, p. 210]. This approach, however, requires that all LN volume measures of $\Theta(\sqrt{TP})$ are known. In general, it is difficult to obtain these volume measures and one must resort to an upper bound on the volume of Γ_∞ .

Let $V(\cdot)$ evaluate to the volume of the region. Since all the regions are closed, an upper bound on $V(\Gamma_\infty)$ can be found using (14) to give

$$V(\Theta(\sqrt{TP} + p_\rho)) > V(\Gamma_\infty) > V(\Theta(\sqrt{TP})).$$

Thus, to find an upper bound on $V(\Gamma_\infty)$ it is necessary to compute an expression for $V(\Theta(q))$, for any $q > 0$, as well as estimating the parameter p_ρ .

1) *Computing $V(\Theta(q))$* : Since $\Theta(q)$ is the truncation of the L -fold Cartesian product of a generalized N -cone with itself, $V(\Theta(q))$ can be computed by exploiting the geometry of the region. Define

$$\Psi_1(q; L) = \left\{ (x_{1,1}, x_{1,2}, \dots, x_{1,L}) \in \mathbb{R}^L: \forall (k = 1 \dots L) x_{1,k} \geq 0, \sum_{k=1}^L x_{1,k} \leq Lq \right\} \quad (15)$$

as the projection of $\Psi(q)$ (10) into the space defined by the L , ϕ_1 basis vectors of Υ^L . If $\mathbf{x}_1 = (x_{1,1}, x_{2,1}, \dots, x_{L,1})$, the volume $V(\Theta(q))$ can be written using (15) as

$$\begin{aligned} V(\Theta(q)) &= \int_{\mathbf{z} \in \Upsilon^L \cap \Psi(\sqrt{TP})} d\mathbf{z} \\ &= \int_{\mathbf{x}_1 \in \Psi_1(q; L)} \int_{\mathbf{z}_1 \in \Upsilon_{x_{1,1}}} \int_{\mathbf{z}_2 \in \Upsilon_{x_{1,2}}} \\ &\quad \dots \int_{\mathbf{z}_L \in \Upsilon_{x_{1,L}}} d\mathbf{z}_L \dots d\mathbf{z}_2 d\mathbf{z}_1 d\mathbf{x}_1. \end{aligned}$$

As defined in Section II-B, the set of admissible points for each N -dimensional constituent constellation Υ is a generalized cone. As such, it is characterized by a single cross section in its ϕ_1 direction. Since each cross section for a given ϕ_1 coordinate value is directly similar to Υ_1 , $V(\Upsilon_k) = V(\Upsilon_1)k^{N-1}$ for $k \geq 0$ [9]. Thus, the integral for $V(\Theta(q))$ can be simplified by

performing the integration in each constituent constellation to yield

$$V(\Theta(q)) = \int_{\mathbf{x}_1 \in \Psi_1(q; L)} V(\Upsilon_1)^L \prod_{k=1}^L x_{k,1}^{N-1} d\mathbf{x}_1. \quad (16)$$

As shown the Appendix, this integral can be simplified to the form

$$V(\Theta(q)) = V(\Upsilon_1)^L \frac{(N-1)!^L}{(LN)!} (Lq)^{LN}. \quad (17)$$

2) *Asymptotic Value of p_ρ* : Since $\Theta(q)$ is a generalized cone, the $LN - 1$ -dimensional cross-sectional volume and the height is required for a given value in the ϕ_1^{LN} direction. In order to determine the value of p_ρ required to compute the volume of $\Theta(q)$, it is necessary to consider only a two-dimensional cross section of $\Theta(q)$ in a hyperplane containing ϕ_1^{LN} . Fig. 1 presents an example of just such a cross section with ϕ_1^{LN} as a coordinate axis. In what follows, the specific cross section taken to compute p_ρ is developed. Note that since ϕ_1^{LN} is taken as a coordinate axis a scaling factor of \sqrt{L} is present in this coordinate due to (10).

Define the set

$$\Delta_j = \{(x_1, x_2, \dots, x_{LN}) \in \Theta(\sqrt{TP}) : x_j = 0\}$$

where $j \in \{1, \dots, LN\}$. The set $\Delta_j \subset \Theta(\sqrt{TP})$ and can be thought of as a surface which has ϕ_j as a normal vector. The perpendicular distance between any vector in $\Theta(\sqrt{TP})$ and the corresponding one Γ_∞ is ρ by (13). So, the distance between Δ_j and Γ_∞ in the ϕ_j direction is ρ . Let $\zeta_\Delta(j)$ be the projection of ϕ_1^{LN} into Δ_j and $\zeta_\phi(j)$ be the projection of ϕ_1^{LN} onto ϕ_j . Taking a cross section of the Υ^L with a hyperplane containing $\zeta_\phi(j)$ and $\zeta_\Delta(j)$ gives Fig. 1.

The angles $\alpha(L)$ and $\beta(L)$ in Fig. 1 can be determined from the geometry of the region. Define unit vectors $\hat{\zeta}_\phi(j)$ and $\hat{\zeta}_\Delta(j)$ as unit vectors parallel to $\zeta_\phi(j)$ and $\zeta_\Delta(j)$, respectively. The desired angles can be found through the inner product as

$$\begin{aligned} \langle \hat{\zeta}_\Delta(j), \phi_1^{LN} \rangle &= \cos \alpha(L) \\ &= \begin{cases} \sqrt{\frac{L-1}{L}}, & j = kN + 1, k = 0, 1, \dots, (L-1) \\ 1, & \text{otherwise} \end{cases} \end{aligned}$$

and

$$\begin{aligned} \langle \hat{\zeta}_\phi(j), \phi_1^{LN} \rangle &= \cos \beta(L) \\ &= \begin{cases} \frac{1}{\sqrt{L}}, & j = kN + 1, k = 0, 1, \dots, (L-1) \\ 0, & \text{otherwise.} \end{cases} \end{aligned}$$

Note that $\alpha(L) + \beta(L) = \pi/2$ by the definition of $\zeta_\phi(j)$ and $\zeta_\Delta(j)$.

From Fig. 1, the value of p_ρ can be written in

$$p_\rho = \rho + p_0(L).$$

The value of $p_0(L)$ depends on L through $\alpha(L)$ and $\beta(L)$. However, in the upper bound for channel capacity in Section III-D it is only required to compute the asymptotic value of $p_0(L)$ for

large L . Therefore, only the asymptotic value of p_ρ needs to be computed. The angles in Fig. 1 approach the limits

$$\begin{aligned}\lim_{L \rightarrow \infty} \alpha(L) &= 0 \\ \lim_{L \rightarrow \infty} \beta(L) &= \frac{\pi}{2}.\end{aligned}$$

As a result, it is clear that as L grows the distance $p_0(L)$ must then behave asymptotically as

$$\lim_{L \rightarrow \infty} p_0(L) = \rho.$$

As a result, using (11), the asymptotically correct value for p_ρ as $L \rightarrow \infty$ then is

$$p_\rho = \frac{2\rho}{\sqrt{L}} = 2\sigma\sqrt{N}. \quad (18)$$

D. Upper Bound Computation

The channel capacity in bits per symbol for a given optical intensity basis set $C_s(\Phi)$ can be upper-bounded using the sphere-packing argument developed for electrical power constrained channels [17]. The maximum rate is upper-bounded by the asymptotic number of nonoverlapping spheres that can be packed in Γ_{LN} as L goes to infinity. Using the previously defined regions

$$\begin{aligned}C_s(\Phi) &\leq \lim_{L \rightarrow \infty} \frac{1}{L} \log_2 \frac{V(\Gamma_{LN})}{V(\rho B^{LN})} \\ &\leq \lim_{L \rightarrow \infty} \frac{1}{L} \log_2 \frac{V(\Theta(\sqrt{TP} + p_\rho))}{V(\rho B^{LN})}\end{aligned} \quad (19)$$

where the volume of the LN -ball can be written as

$$V(\rho B^{LN}) = \frac{\pi^{LN/2} \rho^{LN}}{(LN/2)!}.$$

Using (17) and (18), the number of reliably decodable code-words can be upper-bounded as

$$\begin{aligned}&\frac{V(\Theta(\sqrt{TP} + p_\rho))}{V(\rho B^{LN})} \\ &= \frac{(LN/2)!}{(LN)!} \frac{((L(\sqrt{TP} + 2\sigma\sqrt{N}))^N V(\Upsilon_1)(N-1)!)^L}{(LN\pi\sigma^2)^{LN/2}}\end{aligned}$$

Using Stirling's formula [19] to bound the factorial function $\sqrt{2\pi n} \left(\frac{n}{e}\right)^n \exp\left(\frac{1}{12n+1}\right) < n! < \sqrt{2\pi n} \left(\frac{n}{e}\right)^n \exp\left(\frac{1}{12n}\right)$ and simplifying the resulting expressions, the number of code-words can be upper-bounded as

$$\begin{aligned}&\frac{V(\Theta(\sqrt{TP} + p_\rho))}{V(\rho B^{LN})} < \frac{\exp f_\epsilon(L)}{\sqrt{2}} \\ &\times \left(\frac{(\sqrt{TP} + 2\sigma\sqrt{N})V(\Upsilon_1)^{1/N}(N-1)!^{1/N}}{N\sigma} \sqrt{\frac{e}{2\pi}} \right)^{LN}\end{aligned}$$

where $\lim_{L \rightarrow \infty} f_\epsilon(L) = 0$. Substituting into (19) and taking the limit as $L \rightarrow \infty$ yields the upper bound,

$$\begin{aligned}C_s(\Phi) &\leq N \log_2 \left[\left(\sqrt{T} \frac{P}{\sigma} + 2\sqrt{N} \right) \right. \\ &\times \left. \frac{V(\Upsilon_1)^{1/N}(N-1)!^{1/N}}{N} \sqrt{\frac{e}{2\pi}} \right] \text{ [bits/symbol]} \quad (20)\end{aligned}$$

for some symbol period T .

Notice that the upper bound on capacity $C_s(\Phi)$ depends on the pulse set chosen via the admissible region Υ . Section V-D

presents a discussion on the selection of pulse shapes to maximize the achievable rates under a bandwidth constraint and under differing SNRs.

A spectral constraint can be imposed on the capacity of the channel by expressing it as a maximum spectral efficiency in units of bits per second per hertz, $C_\eta(\Phi)$, as opposed to $C_s(\Phi)$ in (20), which is in units of bits per symbol. Spectral efficiency is a more appropriate measure of the rate since it combines important practical channel performance measures of data rate and bandwidth. The upper bound on channel capacity for a given Φ in (20) can be represented as a bound on the maximum spectral efficiency for a channel bandwidth of W_{ch} hertz using the effective dimension $\kappa(\Phi)$ (6) as

$$C_\eta(\Phi) \leq C_\eta^{\text{up}}(\Phi)$$

where

$$\begin{aligned}C_\eta^{\text{up}}(\Phi) &= \frac{2N}{\kappa(\Phi)} \log_2 \left[\left(\sqrt{\frac{\kappa(\Phi)}{2W_{\text{ch}}}} \frac{P}{\sigma} + 2\sqrt{N} \right) \right. \\ &\times \left. \frac{V(\Upsilon_1)^{1/N}(N-1)!^{1/N}}{N} \sqrt{\frac{e}{2\pi}} \right] \text{ [bits/s/Hz]}. \quad (21)\end{aligned}$$

Unlike the band-limited case, where the dimension of each basis signal is one, here the effective dimension of each signal in Υ must be computed. The factor $N/\kappa(\Phi)$ can be thought of as a measure of the dimensional efficiency of a given model since N represents the dimension of each transmitted signal while $\kappa(\Phi)$ is the maximum dimension of the set of signals determined by Φ using a $(1-\epsilon)$ -fractional energy bandwidth measure.

Having imposed a spectral constraint on the channel model, $C_\eta(\Phi)$ is invariant to the symbol period unlike $C_s(\Phi)$. For example, any rectangular pulse technique is a subset of the Cartesian product of one-dimensional rectangular PAM. For a given symbol period, consider forming another pulse technique by transmitting L time disjoint pulses per period each with width T/L seconds. In the limit, as pulse width goes to zero for a fixed average optical power, the capacity limit $C_s(\Phi)$ is unbounded, while imposing a bandwidth constraint yields an upper bound on $C_\eta(\Phi)$ in (21) which is unaffected by the width of the pulses chosen. Thus, by imposing a bandwidth constraint and choosing spectral efficiency as the measure of channel capacity a consistent measure of the maximum data rate of the channel is obtained.

IV. LOWER BOUND ON CHANNEL CAPACITY

A lower bound on the capacity of the optical intensity channel can be found by computing the mutual information between the channel input and output for any input distribution $f_{\mathbf{X}}(\mathbf{x}) \in \mathcal{F}$. In this work, a lower bound is computed by selecting the maximum entropy source distribution, $f_{\mathbf{X}}^*(\mathbf{x}) \in \mathcal{F}$. This distribution maximizes the differential entropy of the source subject to both the nonnegativity and average optical power constraints.

Due to the signal space definition, for a fixed T , the average optical power depends solely on the mean of the ϕ_1 coordinate value, as shown in (2). To find $f_{\mathbf{X}}^*(\mathbf{x})$, consider first the family of distributions $f_{\mathbf{X}}^*(\mathbf{x}; A) \in \mathcal{F}$ which maximize the entropy for any fixed average optical power $A \leq P$. By the maximum entropy principle, $f_{\mathbf{X}}^*(\mathbf{x}; A) = K \exp(-\lambda x_1)$, for $\mathbf{x} \in \Upsilon$ and

for some $K, \lambda > 0$ [20, pp. 266–268]. The constants K and λ can be found by solving the following:

$$\int_{\mathbf{x} \in \Upsilon} f_{\mathbf{X}}^*(\mathbf{x}; A) d\mathbf{x} = 1$$

$$\int_{\mathbf{x} \in \Upsilon} x_1 f_{\mathbf{X}}^*(\mathbf{x}; A) d\mathbf{x} = \sqrt{T}A$$

to yield

$$f_{\mathbf{X}}^*(\mathbf{x}; A) = \left(\frac{N}{\sqrt{T}A} \right)^N \frac{1}{V(\Upsilon_1)(N-1)!} \exp\left(-N \frac{x_1}{\sqrt{T}A}\right) \quad (22)$$

for every $\mathbf{x} = (x_1, x_2, \dots, x_N) \in \Upsilon$. For \mathbf{X} distributed as $f_{\mathbf{X}}^*(\mathbf{x}; A)$, the differential entropy of $f_{\mathbf{X}}^*(\mathbf{x}; A)$, $h(\mathbf{X}; A)$, can be computed as

$$h(\mathbf{X}; A) = - \int_{\mathbb{R}^N} f_{\mathbf{X}}^*(\mathbf{x}; A) \log_2 f_{\mathbf{X}}^*(\mathbf{x}; A) d\mathbf{x}$$

$$= N \log_2 \left(\sqrt{T}A \frac{V(\Upsilon_1)^{1/N} (N-1)!^{1/N} e}{N} \right). \quad (23)$$

Since this entropy is an increasing function of A

$$h(\mathbf{X}, P) = \max_{0 < A \leq P} h(\mathbf{X}, A)$$

and, hence, the maxentropic source distribution satisfies the average optical power constraint with equality, that is, $f_{\mathbf{X}}^*(\mathbf{x}) = f_{\mathbf{X}}^*(\mathbf{x}; P)$. Notice that $f_{\mathbf{X}}^*(\mathbf{x})$ is a function of solely the coordinate in the ϕ_1 direction which represents the average optical power of each symbol. The conditional distribution of a given $x_1 = k$ is uniform over all elements of Υ_k , which is entropy maximizing in the absence of constraints.

The lower bound on the capacity can be written as

$$C_s(\Phi) \geq I(\mathbf{X}^*; \mathbf{Y}) \quad (24)$$

where $\mathbf{X}^* \sim f_{\mathbf{X}}^*(\mathbf{x})$. Although the use of the maxentropic distribution provides a lower bound on the channel capacity, a closed-form solution is in general difficult to derive. A closed-form lower bound on the channel capacity can be found by expanding the mutual information as

$$I(\mathbf{X}^*; \mathbf{Y}) = h(\mathbf{X}^* + \mathbf{Z}) - h(\mathbf{Z}) \quad (25)$$

where $h(\cdot)$ evaluates to the differential entropy. A lower bound on $I(\mathbf{X}^*; \mathbf{Y})$ can be computed as

$$I(\mathbf{X}^*; \mathbf{Y}) = h(\mathbf{X}^* + \mathbf{Z}) - h(\mathbf{Z})$$

$$\geq h(\mathbf{X}^* + \mathbf{Z} | \mathbf{Z}) - h(\mathbf{Z}) \quad (26)$$

$$= h(\mathbf{X}^*) - h(\mathbf{Z}) \quad (27)$$

where (26) arises since conditioning reduces differential entropy and (27) since translation does not affect differential entropy.

Finally, using (23) with $A = P$ along with (24) and (27), and applying the spectral constraint (6) yields a lower bound on the spectral efficiency

$$C_\eta(\Phi) \geq C_\eta^{\text{low}}(\Phi)$$

where

$$C_\eta^{\text{low}}(\Phi) = \frac{2N}{\kappa(\Phi)} \log_2 \left(\sqrt{\frac{\kappa(\Phi) P}{2W_{\text{ch}} \sigma}} \right)$$

$$\times \frac{V(\Upsilon_1)^{1/N} (N-1)!^{1/N}}{N} \sqrt{\frac{e}{2\pi}} \quad [\text{bits/s/Hz}]. \quad (28)$$

TABLE I
COMPARISON OF MAIN CHARACTERISTICS OF EACH
OPTICAL INTENSITY MODEL

	Rectangular PAM	Raised-QAM	2-PSWF	3-PSWF
N	1	3	2	3
$\kappa(\Phi)$	20.572	27.038	20.572	20.572
$V(\Upsilon_1)$	1	$\pi/2 \approx 1.571$	1.914	2.173
$N/\kappa(\Phi)$ (%)	4.86	11.10	9.72	14.58

The asymptotic behavior of the bounds (21) and (28) can be found by computing the limit of the difference of the bounds. It is simple to show that for a given average optical power that

$$\lim_{\sigma \rightarrow 0} C_\eta^{\text{up}}(\Phi) - C_\eta^{\text{low}}(\Phi) = 0$$

that is, the upper and lower bounds on the spectral efficiency are asymptotically exact as optical SNR tends to infinity. Note that $C_\eta^{\text{low}}(\Phi)$ is tight only at high optical SNR and that numerical computation of $I(\mathbf{X}^*; \mathbf{Y})$ is required to provide a better low bound at low optical SNR. The relationship between $C_\eta(\Phi)$ and the capacity bounds derived can be summarized as

$$C_\eta^{\text{up}}(\Phi) \geq C_\eta(\Phi) \geq \frac{2}{\kappa(\Phi)} I(\mathbf{X}^*; \mathbf{Y}) \geq C_\eta^{\text{low}}(\Phi). \quad (29)$$

Section V computes the capacity bounds for a number of example basis sets.

V. EXAMPLES AND DISCUSSION

A. Rectangular PAM

Form an M -ary PAM scheme using the rectangular pulse shape of (1) so that $\Phi_{\text{PAM}} = \{\phi_1(t)\}$. The effective dimension of the scheme using a 0.99-fractional energy bandwidth ($\epsilon = 10^{-2}$) is $\kappa(\Phi_{\text{PAM}}) = 20.572$, and is computed by direct integration in frequency domain using a symbolic mathematics package [21]. The cross section Υ_1 in this case is a point and the volume of the cross section in this case is taken as 1 which allows all previous derivations to hold [9], and is presented in Table I.

Fig. 3 presents the upper and the lower bounds on $C_\eta(\Phi_{\text{PAM}})$ for the PAM scheme defined as well as spectral efficiency curves for discrete uniform 2-, 4-, 8-, and 16-point constellations versus optical SNR. The spectral efficiency curves for the uniformly distributed examples were computed numerically using the well known technique developed by Ungerboeck [22].

The upper bound on capacity is obtained by direct application of (21) to give

$$C_\eta(\Phi_{\text{PAM}}) \leq \frac{2}{\kappa(\Phi_{\text{PAM}})} \log_2 \left[\left(\sqrt{\frac{\kappa(\Phi_{\text{PAM}}) P}{2W \sigma}} + 2 \right) \sqrt{\frac{e}{2\pi}} \right].$$

The lower bound on capacity was determined first by computing $f_{\mathbf{Y}}(\mathbf{y})$, which takes the form

$$f_{\mathbf{Y}}(\mathbf{y}) = f_{\mathbf{X}}^*(\mathbf{y}) * f_{\mathbf{Z}}(\mathbf{y})$$

$$= \frac{1}{\sqrt{TP}} \left(1 - Q \left(\frac{y}{\sigma} - \frac{\sigma}{\sqrt{TP}} \right) \right) \exp \left(\frac{\sigma^2 - 2y}{2\sqrt{TP}} \right)$$

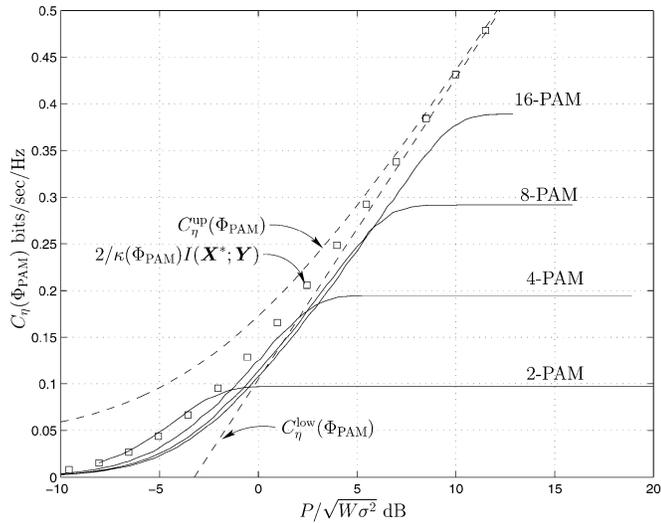


Fig. 3. Capacity bounds and mutual information curves for rectangular PAM schemes.

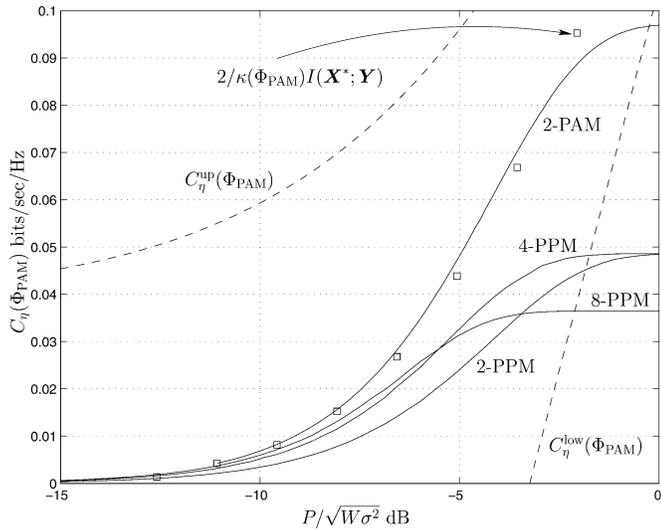


Fig. 4. Capacity bounds for rectangular PAM with mutual information curves for 2-PAM and N -PPM schemes at low optical SNR.

where

$$Q(x) \triangleq \frac{1}{\sqrt{2\pi}} \int_x^\infty \exp(-u^2/2) du.$$

Since $f_Y(\mathbf{y})$ does not have a closed form in this case, $I(\mathbf{X}^*; \mathbf{Y})$ is computed numerically for a number of SNRs using a mathematics software package [21] and plotted in Fig. 3. As expected, at high optical SNR, the lower and upper bounds on capacity approach one another. Notice also that for all SNRs considered that the relationship (29) holds in this example.

Fig. 4 shows the same bounds as Fig. 3 at low SNR. Spectral efficiency curves for a widely used coded rectangular PAM scheme called N -PPM are also presented. A 4-PPM modulation scheme is the standard signaling format for the Infrared Data Association 4-Mb/s wireless infrared optical link [23]. An N -PPM symbol is a block code defined over NT seconds in

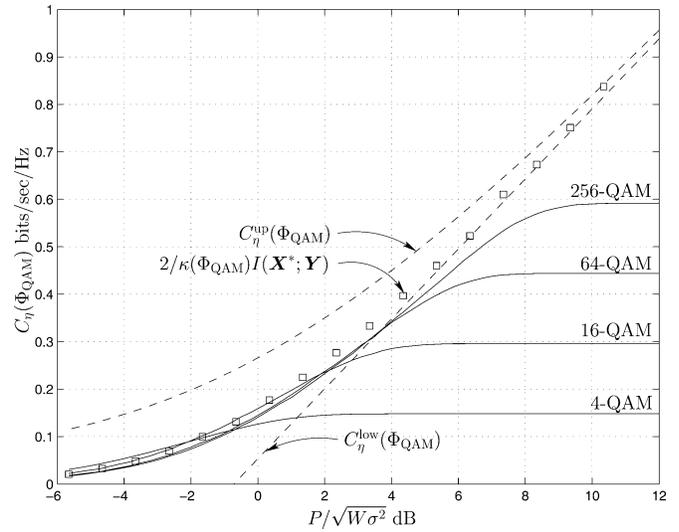


Fig. 5. Capacity bounds for raised-QAM schemes with mutual information curves for 4-, 16-, 64-, and 256-point constellations.

which the output intensity is zero except for a single T interval which assumes a fixed nonzero value. The mutual information curves for a channel model based on N -PPM are computed, using the same numerical techniques as in the PAM case, and plotted for $N = 2, 4, 8$ in Fig. 4. Note that PPM schemes approach the lower bound for capacity at low optical SNR. Also note that the PPM curves are in agreement with previous work which shows that higher cardinality PPM constellations asymptotically have better performance than lower size constellations at low SNR [4]. The lower bound $C_\eta^{\text{low}}(\Phi_{\text{PAM}})$ is very loose at these low optical SNRs. The numerical computation of mutual information in (25) is a better lower bound at low optical SNRs.

B. Raised Quadrature Amplitude Modulation (QAM)

An optical three-dimensional raised-QAM scheme can be defined by specifying $\Phi_{\text{QAM}} = \{\phi_1(t), \phi_2(t), \phi_3(t)\}$ where $\phi_1(t)$ is defined as in (1) and

$$\begin{aligned} \phi_2(t) &= \sqrt{\frac{2}{T}} \cos(2\pi t/T) \\ \phi_3(t) &= \sqrt{\frac{2}{T}} \sin(2\pi t/T) \end{aligned}$$

for $t \in [0, T)$ [9]. Fig. 5 presents a plot of the upper bound on capacity (21) for a three-dimensional raised-QAM scheme which takes the form

$$\begin{aligned} C_\eta(\Phi_{\text{QAM}}) &\leq \frac{6}{\kappa(\Phi_{\text{QAM}})} \log_2 \left[\left(\sqrt{\frac{\kappa(\Phi_{\text{QAM}}) P}{2W}} \frac{P}{\sigma} + 2\sqrt{3} \right) \sqrt{\frac{e}{18\pi^{1/3}}} \right]. \end{aligned}$$

The effective dimension was computed for $W_{0.01}$ using (7) and by noting that $|\phi_2^F(f)|^2 = |\phi_3^F(f)|^2$. Fig. 6 plots the value of $\kappa(\Phi_{\text{QAM}})$ found using (7) for unit energy signals versus the allowed coordinate value in the ϕ_1 basis. The effective dimension was computed by numerically integrating the Fourier transforms

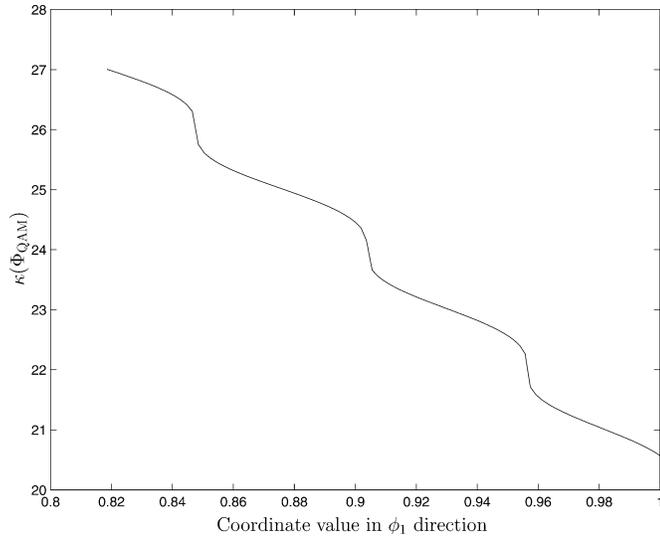


Fig. 6. Effective dimension $\kappa(\Phi_{\text{QAM}})$ as a function of coordinate value.

of the basis functions and scaling them according to (7) to yield $W_{0.01}$. Notice that the maximum value occurs when the selected vector is taken from the boundary points of Υ , i.e., by maximizing the coordinate value in the sinusoidal bases. The maximum effective dimension was then estimated to be $\kappa(\Phi_{\text{QAM}}) = 27.038$ and is achieved by points on the the boundary of Υ . The cross-sectional volume $V(\Upsilon_1) = \pi/2$ and is the area of a circular disc of radius $1/\sqrt{2}$ [9], and is presented in Table I.

As is the case with PAM, the lower bound on spectral efficiency via $I(\mathbf{X}^*; \mathbf{Y})$ must be computed numerically. Unfortunately, computation of $f_{\mathbf{X}}(\mathbf{y})$ is difficult and the $I(\mathbf{X}^*; \mathbf{Y})$ was computed using a discretized version of $f_{\mathbf{X}}^*(\mathbf{x})$ (22). A constellation of 1024 points was constructed by sampling $f_{\mathbf{X}}^*(\mathbf{x})$ at integer lattice points and scaling to have unit sum. The maximum rate using this source distribution was computed using the numerical technique presented by Ungerboeck [22] for a number of average optical powers. Again, the upper and lower bounds approach one another at high optical SNRs. Spectral efficiency curves for 4-, 16-, 64-, and 256-point uniform distributions were determined using the same numerical techniques as in the PAM case [22] and are also presented in Fig. 5.

C. Prolate Spheroidal Wave Function Bases

As discussed in Section II-D, for a given $2WT$ product, the prolate spheroidal wave functions are the time-limited functions with support in $[0, T]$ with maximum energy in the frequency band $[-W, W]$ of all unit energy functions [16]. In light of the bandwidth constraint imposed it seems natural to form an optical intensity signaling scheme based on this orthonormal family of functions.

Fig. 7(a) presents of plot of the $\varphi_0(t)$ and $\varphi_1(t)$ for $2WT = \kappa(\Phi_{\text{PAM}})$. These functions are approximated by generating highly oversampled discrete prolate spheroidal sequences in a popular numerical mathematics package [24]. It can be shown that in the limit as sampling rate goes to infinity the discrete sequence converges point-wise to the associated prolate spheroidal wave function [25].

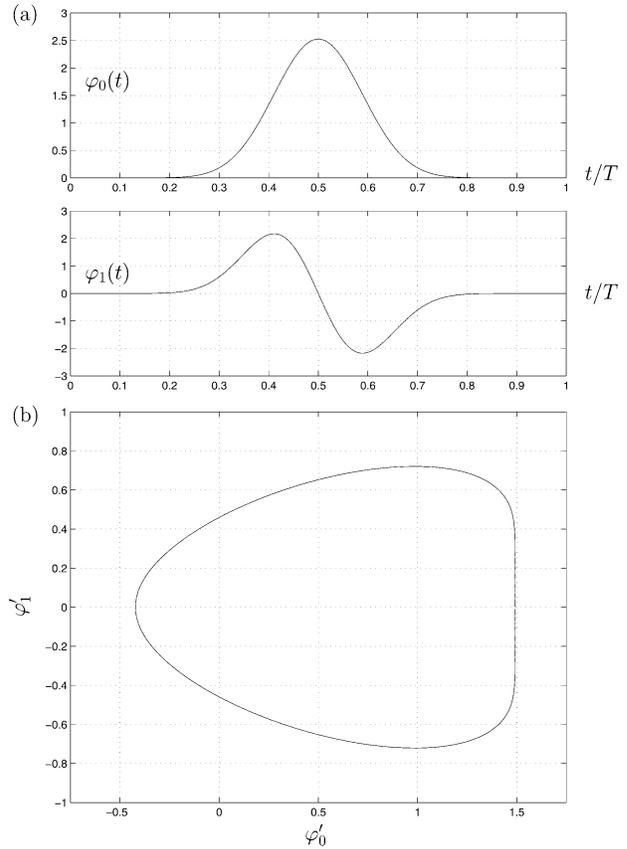


Fig. 7. (a) Prolate spheroidal wave functions $\varphi_0(t)$ and $\varphi_1(t)$ for $\kappa(\Phi_{\text{PAM}}) = 20.572$ and (b) the corresponding Υ_1 region.

Consider forming an optical intensity signal basis as discussed in Section II-B. The function $\phi_1(t)$ (1) is a basis function for every intensity signaling set with $2W_\epsilon(\phi_1)T = \kappa(\Phi_{\text{PAM}})$. Consider choosing the remaining basis functions so that $2W_\epsilon(\phi_n)T < \kappa(\Phi_{\text{PAM}})$ for $n > 1$. The effective dimension κ can be approximated by fixing T and maximizing the bandwidth over all transmittable signals via (7). For any nonnegative $x(t) \neq \phi_1(t)$, $W_\epsilon(x) \leq W_\epsilon(\phi_1)$ since every other basis function has a lower $(1 - \epsilon)$ -fractional energy bandwidth. In this case, the approximated effective dimension is achieved since $\phi_1(t)$ has the maximum bandwidth and is transmittable. As a result, the effective dimension of such optical intensity sets is $\kappa(\Phi_{\text{PAM}})$.

An N -prolate spheroidal wave function (PSWF) optical intensity model can be formed by performing a Gram-Schmidt orthogonalization procedure with $\phi_1(t)$ and $\varphi_n(t)$ for $n = 0, 1, \dots, N - 2$ at a time-bandwidth product of $2W_\epsilon(\phi_1)T$ to form the basis set Φ_{PSWF} . The basis functions for this model are then denoted $\phi_1(t)$ and $\varphi'_n(t)$.

In order to verify the value of $\kappa(\Phi_{\text{PSWF}})$, the time-bandwidth product for each basis function must be computed. In the case of 2- and 3-PSWF, $2W_\epsilon(\varphi'_n)T$ were computed numerically for $n = 0, 1$ and evaluated to 8.275 and 8.510, respectively. Since the time-bandwidth product of $\varphi'_n(t)$ is less than that of $\phi_1(t)$, $\kappa(\Phi_{\text{PSWF}}) = \kappa(\Phi_{\text{PAM}})$ for these models.

The $V(\Upsilon_1)$ for 2- and 3-PSWF were also computed numerically and are presented in Table I. In the case of 2-PSWF, the minimum and maximum value of $\varphi'_0(t)$ were estimated from

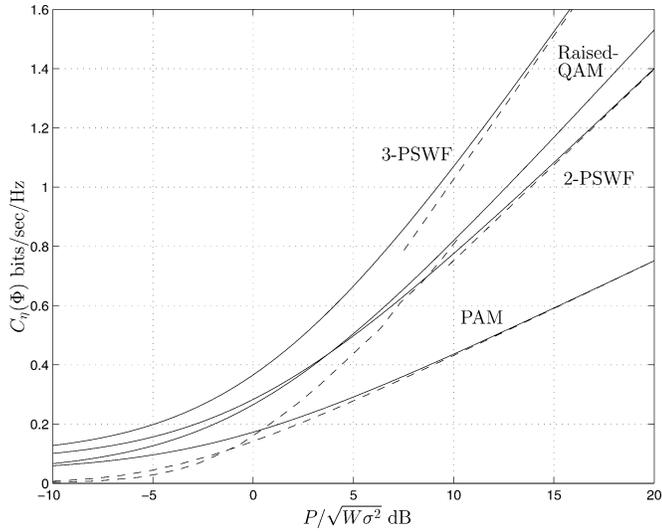


Fig. 8. Upper and lower bounds on the spectral efficiency of rectangular PAM, raised-QAM, and 2- and 3-PSWF (solid and dashed lines are upper and lower bounds, respectively).

the sampled waveform and their difference is taken as an estimate of $V(\Upsilon_1)$. Fig. 7(b) presents the boundary of Υ_1 for the 3-PSWF scheme. The values of the boundary of Υ_1 were estimated at 1000 points taken at a constant angular increment of $\Delta\theta = 2\pi/1000$. The resulting sequence of boundary points r_k were computed by increasing the size of a ray from the origin until the corresponding time-domain signal was no longer transmittable. This procedure was done at a number of increment steps until the boundary was known to within a tolerance of 10^{-3} units. The volume was then estimated using the following Riemann sum in polar coordinates:

$$V(\Upsilon_1) \approx \sum_k \frac{1}{2} r_k^2 \Delta\theta.$$

The upper and lower bounds on the spectral efficiency (21) and (28) were computed and are presented in Fig. 8 along with those for rectangular PAM and raised-QAM.

D. Discussion

Fig. 8 presents a comparison of the capacity bounds derived in Sections III and IV for the examples presented. Note that at high SNRs, bandwidth-efficient signaling schemes based on prolate spheroidal wave function basis or on the raised-QAM pulse set have more than twice the maximum spectral efficiency of the rectangular PAM techniques at a given SNR. At lower optical SNR, the derived bounds are loose and do not reveal any new insight. Indeed, at low SNR, when the available spectral efficiencies tend to zero, rectangular pulse techniques may be attractive due to their ease of implementation.

Table I presents a comparison of the parameters which define the capacity bounds for the examples considered. As indicated by $C_\eta(\Phi)$ in (21), for a given $\kappa(\Phi)$, the spectral efficiency bounds are maximized by the Φ which simultaneously maximize the dimensional efficiency $N/\kappa(\Phi)$ and cross-sectional volume $V(\Upsilon_1)$. In the examples, raised-QAM achieves higher spectral efficiencies by increasing the dimension of each symbol for a modest increase in the effective dimension. As a result, the

dimensional efficiency is improved along with the spectral efficiency. For 2- and 3-PSWF schemes, the effective dimension is fixed to be $\kappa(\Phi_{\text{PAM}})$ by definition. The additional bases improve the dimensional efficiency and thus allow for larger maximum spectral efficiencies. The role of $V(\Upsilon_1)$, although logarithmic in its impact on the spectral efficiency, is significant. Since $V(\Upsilon_1)$ depends on the amplitude characteristics of Φ , it is not clear that N -PSWF optical intensity schemes will necessarily support higher spectral efficiencies as N increases.

It is important to recognize that $C_\eta(\Phi)$ is the maximum achievable spectral efficiency for the given pulse set Φ . The basis set determines the set of transmittable signals over which a distribution is chosen to maximize the spectral efficiency. The effective dimension of the set of transmittable signals $\kappa(\Phi)$ is required to quantify the maximum achievable rate in terms of quantities of the continuous channel. The pulse set which maximizes $C_\eta(\Phi)$ for a given channel bandwidth, however, is still unknown. Some early work on the photon-counting channel demonstrated that narrow pulse position techniques were optimal pulse techniques in the sense of a given average distance measure [26], [27]. Capacity results for the photon-counting channel nearly exclusively assume that rectangular pulse techniques are employed and, as a result, do not generalize to other pulse shapes.

It should also be noted that the classical capacity result $C = W \log_2(1 + \text{SNR})$ in the electrical channel is really the maximum achievable rate for a strictly band-limited channel model and a defined bandwidth constraint. In this case, $\Phi = \{\text{sinc}(\pi t/T)\}$ and $\kappa(\Phi) = 1$ [17]. In other words, the capacity is computed for a given channel model, which includes the basis set Φ under an energy and bandwidth constraint as is done in a similar context in this work.

VI. CONCLUSION

We have derived asymptotically exact capacity bounds for the optical intensity channel with average optical power and bandwidth constraints in Gaussian noise. These results complement rather than contradict previous work on the Poisson photon-counting channel. The photon-counting channel can be viewed as an optical system operating at low optical power where the quantum nature of the photons dominates performance. Rectangular pulse techniques are uniquely considered in the photon-counting channel since the bandwidth of the channel is considered to be very large.

In this work, we treat a fundamentally different channel. Indoor free-space optical channels suffer from reduced bandwidth due to multipath distortion and from white Gaussian noise due to high background illumination. The derived capacity bounds are not restricted to rectangular pulse techniques, as in previous work, but are general and treat all time-disjoint optical intensity schemes. A bandwidth constraint is imposed on the set of signals that are transmitted by way of determining the effective dimension of the space of time-limited signals with a given fractional energy bandwidth. The derived capacity bounds demonstrate that optical intensity signaling schemes based on rectangular basis sets have significantly lower achievable maximum spectral efficiencies than bandwidth-efficient techniques. In partic-

ular, significant rate gains can be obtained by using raised-QAM or prolate spheroidal wave function pulses set over rectangular PAM at high optical SNRs.

This work, however, does not address the complexity of implementing encoders and decoders for the multidimensional optical signaling formats presented. In designing practical signaling strategies, the complexity of the codes required to approach the derived capacity may limit their application to lower rate optical intensity links. Wireless optical channels are one such class of low-pass optical intensity channel which may benefit from the application of spectrally efficient signaling schemes considered in this paper. The performance of coding strategies for a given pulse set, however, can now be compared versus the capacity bounds derived here.

An interesting long-standing open question which remains is the pulse set Φ which maximizes $C_\eta(\Phi)$ subject to the channel constraints. This work has not solved this problem but has suggested that the solution will require a pulse set which is dimensionally efficient while providing a large volume of admissible signals.

APPENDIX

This appendix presents a proof for the volume of the set of transmit codewords $V(\Theta(q))$ in Section III-C1. This development proceeds from (16) and shows how the closed-form (17) is derived.

Theorem 1: For $\mathbf{x} = (x_1, x_2, \dots, x_L)$

$$\int_{\mathbf{x} \in \Psi_1(q; L)} \prod_{k=1}^L x_k^{N-1} d\mathbf{x} = \frac{(N-1)!^L}{(LN)!} (Lq)^{LN}$$

for all integers $L > 0$ and for some fixed integer $N > 0$ and $\Psi_1(q; L)$ defined in (15).

Proof: By induction on L .

For $L = 1$, the above equality can be easily verified.

Assume the above equality holds for some $L > 1$. Now consider the left-hand side of the equality for $L + 1$

$$I(L+1) = \int_{\mathbf{x} \in \Psi_1(q; L+1)} \prod_{k=1}^{L+1} x_k^{N-1} d\mathbf{x}.$$

This integral can be rewritten by decomposing it into two integrals, one over L variables and the other over 1, i.e.,

$$I(L+1) = \int_0^{(L+1)q} x_{L+1}^{N-1} \int_{\mathbf{x} \in R} \prod_{k=1}^L x_k^{N-1} d\mathbf{x} dx_{L+1}$$

where

$$R = \Psi_1\left(\frac{1}{L}((L+1)q - x_{L+1}); L\right).$$

Applying the inductive hypothesis

$$I(L+1) = \int_0^{(L+1)q} x_{L+1}^{N-1} \frac{(N-1)!^L}{(LN)!} \times ((L+1)q - x_{L+1})^{LN} dx_{L+1}.$$

Let $x = x_{L+1}$ and $h = (L+1)q$ to simplify notation. Expanding the integrand using the binomial theorem and interchanging integration and summation gives

$$I(L+1) = \frac{(N-1)!^L}{(LN)!} \sum_{k=0}^{LN} \binom{LN}{k} (-1)^k h^{LN-k} \int_0^h x^{N-1+k} dx.$$

Evaluating the integral and simplifying

$$I(L+1) = \frac{(N-1)!^L}{(LN)!} h^{N(L+1)} \cdot S_{N,L} \quad (30)$$

where

$$S_{N,L} = \sum_{k \geq 0} \binom{LN}{k} \frac{(-1)^k}{N+k}.$$

A powerful technique for determining the closed-form solution of sums involving binomial coefficients involves the use of hypergeometric series [28, Sec. 5.5]. A *hypergeometric series* is one in which the first term is unity and the ratio of successive terms of the series is the quotient of polynomials. Consider $NS_{N,L}$ and let t_k be the k th term in this series. It is a simple matter to verify that $t_0 = 1$ and that the ratio of successive terms is

$$\frac{t_{k+1}}{t_k} = \frac{(k+N)(k-LN)}{(k+1)(k+N+1)}$$

which is the ratio of two polynomials each of degree 2. Thus, $NS_{N,L}$ is hypergeometric. In fact, $NS_{N,L}$ is a classical series known as a Gaussian hypergeometric series whose closed form is known. In the notation of hypergeometric series, $NS_{N,L}$ can be written as [28, eq. 5.92]

$$NS_{N,L} = {}_2F_1\left(\begin{matrix} N & -LN \\ N+1 & 1 \end{matrix}\right) = \frac{\Gamma(LN+1)\Gamma(N+1)}{\Gamma(LN+N+1)}.$$

Simplifying and using the identity $\Gamma(n+1) = n!$

$$S_{N,L} = \frac{(LN)!(N-1)!}{(N(L+1))!}.$$

Substituting into (30) and expanding h gives

$$I(L+1) = \frac{(N-1)!^{L+1}}{(N(L+1))!} ((L+1)q)^{N(L+1)}$$

which is the right-hand side evaluated at $L + 1$. \square

Therefore, the volume $V(\Theta(q))$ can be written as

$$V(\Theta(q)) = V(\Upsilon_1)^L \frac{(N-1)!^L}{(LN)!} (Lq)^{LN}$$

as presented in (16).

ACKNOWLEDGMENT

The authors would like to thank the reviewers for their insightful comments on the paper. They have contributed to the rigor and completeness of the work. Additionally, the authors would like to thank Dr. Terence H. Chan and Prof.

Adrian Nachman for their suggestions regarding the asymptotic behavior of the derived capacity bounds.

REFERENCES

- [1] J. M. Kahn and J. R. Barry, "Wireless infrared communications," *Proc. IEEE*, vol. 85, pp. 263–298, Feb. 1997.
- [2] J. P. Gordon, "Quantum effects in communication systems," *Proc. IRE*, vol. 50, pp. 1898–1908, Sept. 1962.
- [3] J. R. Pierce, "Optical channels: Practical limits with photon counting," *IEEE Trans. Commun.*, vol. COM-26, pp. 1819–1821, Dec. 1978.
- [4] R. J. McEliece, "Practical codes for photon communication," *IEEE Trans. Inform. Theory*, vol. IT-27, pp. 393–398, July 1981.
- [5] M. H. A. Davis, "Capacity and cutoff rate for poisson-type channels," *IEEE Trans. Inform. Theory*, vol. IT-26, pp. 710–715, Nov. 1980.
- [6] A. D. Wyner, "Capacity and error exponent for the direct detection photon channel—Parts I and II," *IEEE Trans. Inform. Theory*, vol. 34, pp. 1449–1471, Nov. 1988.
- [7] S. Shamai (Shitz), "Capacity of a pulse amplitude modulated direct detection photon channel," *Proc. Inst. Elec. Eng.*, vol. 137, no. 6, pp. 424–430, Dec. 1990.
- [8] S. Hranilovic and F. R. Kschischang, "Signal constellation design for optical intensity modulated channels," in *Proc. IEEE Int. Symp. Information Theory*, Washington, DC, June 2001, p. 275.
- [9] —, "Optical intensity-modulated direct detection channels: Signal space and lattice codes," *IEEE Trans. Inform. Theory*, vol. 49, pp. 1385–1399, June 2003.
- [10] R. You and J. M. Kahn, "Upper-bounding the capacity of optical IM/DD channels with multiple-subcarrier modulation and fixed bias using trigonometric moment space method," *IEEE Trans. Inform. Theory*, vol. 48, pp. 514–523, Feb. 2002.
- [11] A. Gray, *Modern Differential Geometry of Curves and Surfaces*. Boca Raton, FL: CRC, 1993.
- [12] C. E. Shannon, "A mathematical theory of communication," *Bell Syst. Tech. J.*, vol. 27, pp. 379–423, July/Oct. 1948.
- [13] H. J. Landau and H. O. Pollak, "Prolate spheroidal wave functions, Fourier analysis and uncertainty—III: The dimension of the space of essentially time- and band-limited signals," *Bell Syst. Tech. J.*, vol. 41, pp. 1295–1336, July 1962.
- [14] D. Slepian, "On bandwidth," *Proc. IEEE*, vol. 64, pp. 292–300, Mar. 1976.
- [15] D. Slepian and H. O. Pollak, "Prolate spheroidal wave functions, Fourier analysis and uncertainty—I," *Bell Syst. Tech. J.*, vol. 40, pp. 43–63, Jan. 1961.
- [16] H. J. Landau and H. O. Pollak, "Prolate spheroidal wave functions, Fourier analysis and uncertainty—II," *Bell Syst. Tech. J.*, vol. 40, pp. 65–84, Jan. 1961.
- [17] C. E. Shannon, "Communication in the presence of noise," *Proc. IRE*, vol. 37, pp. 10–21, Jan. 1949.
- [18] R. Schneider, *Convex Bodies: The Brunn-Minkowski Theory*, ser. Encyclopedia of Mathematics and its Applications. New York: Cambridge Univ. Press, 1993.
- [19] W. Feller, *An Introduction to Probability Theory and Its Applications*, 3rd ed. New York: Wiley, 1968, vol. 1.
- [20] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. New York: Wiley, 1991.
- [21] *Maple V Programming Guide : Release 5*, Springer-Verlag, New York, NY, 1997.
- [22] G. Ungerboeck, "Channel coding with multilevel/phase signals," *IEEE Trans. Inform. Theory*, vol. IT-28, pp. 55–67, Jan. 1982.
- [23] (2001) Infrared Data Association Serial Infrared Physical Layer Specification Version 1.4. I. D. Association. [Online] Available: www.irda.org
- [24] *MATLAB Version 5.3*, Prentice-Hall, Upper Saddle River, NJ, 1999.
- [25] D. Slepian, "Prolate spheroidal wave functions, Fourier analysis and uncertainty—V: The discrete case," *Bell Syst. Tech. J.*, vol. 57, pp. 1371–1430, May/June 1978.
- [26] B. Reiffen and H. Sherman, "An optimum demodulator for poisson processes: Photon source detectors," *Proc. IEEE*, vol. 51, pp. 1316–1320, Oct. 1963.
- [27] K. Abend, "Optimum photon detection," *IEEE Trans. Inform. Theory*, vol. IT-12, pp. 64–65, Jan. 1966.
- [28] R. L. Graham, D. E. Knuth, and O. Patashnik, *Concrete Mathematics: A Foundation for Computer Science*, 2nd ed. Reading, MA: Addison-Wesley, 1994.